Recommendations

- Huawei Learning Website
  - http://learning.huawei.com/en
- Huawei e-Learning
  - https://ilearningx.huawei.com/portal/#/portal/ebg/51
- Huawei Certification
  - http://support.huawei.com/learning/NavigationAction!createNavi?navId=_31&lang=en
- Find Training
  - http://support.huawei.com/learning/NavigationAction!createNavi?navId=_trainingsearch&lang=en

More Information

- Huawei learning APP

Huawei Training & Certification

版权所有© 2018 华为技术有限公司

Overview of the IT Infrastructure Layer in Cloud DCs

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Objectives

- On completion of this course, you will be able to know:
  - Cloud DC architecture
  - Huawei server basics
  - Huawei storage device basics
  - Huawei switch basics

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

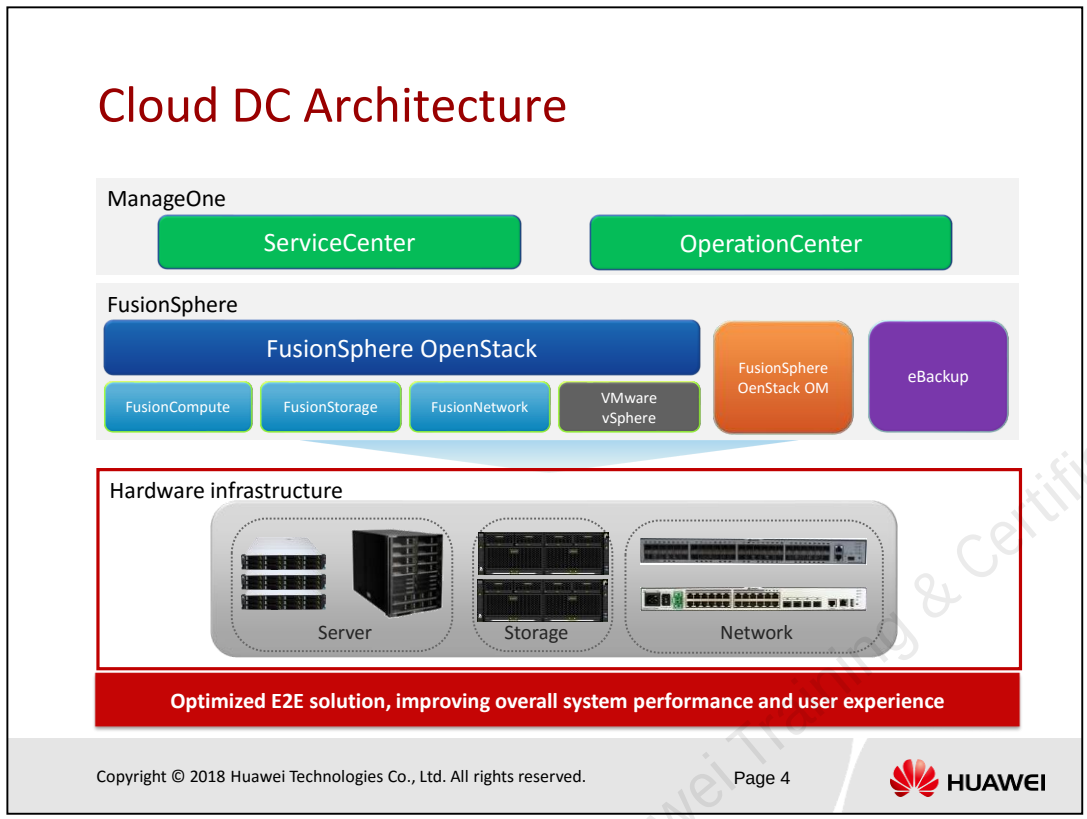
Huawei Training & Certification
Huawei Training & Certification

Contents

1. **Cloud DC Architecture**
2. Server Introduction
3. Storage Device Introduction
4. Switch Introduction

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification



- FusionCompute is a cloud operating system that virtualizes hardware resources and centrally manages virtual resources, service resources, and user resources. FusionCompute virtualizes computing, storage, and network resources using virtual computing, virtual storage, and virtual networking technologies.
- FusionStorage is distributed storage software that has both storage and computing capabilities. It can be deployed on general x86 servers to provide block storage by combining local disks on all the servers into a virtual storage resource pool.
- FusionNetwork is solution designed for evolving traditional virtual switching to forward-thinking software-defined networking (SDN). FusionNetwork leverages Virtual Extensible LAN (VXLAN) L2 tunnel encapsulation technologies and Huawei SDN controllers to automatically deploy and configure SDN networks, control service level agreement (SLA), and isolate and layer multiple tenants.
- FusionSphere OpenStack is enhanced and reinforced by Huawei for enterprises based on Juno of the OpenStack community. FusionSphere OpenStack centrally schedules and manages virtual computing, storage, and network resources over unified RESTful APIs. It reduces the operating expense (OPEX) and provides high system security and reliability, helping enterprises build secure and energy-saving cloud DCs.
- FusionSphere OpenStack OM, the O&M component of FusionSphere OpenStack, monitors and manages FusionSphere hardware and software and supports infrastructure O&M. It provides an O&M portal for internal O&M personnel to perform operations and management.
- BCManager eBackup software provides E2E backup solutions for all virtualization applications.
- Huawei chooses hardware that perfectly meets the requirements of cloud computing on the CPU, memory, storage space, storage IOPS, and network bandwidth to eliminate performance bottlenecks and avoid a hardware resource waste.

Huawei Training & Certification

Contents

1. Cloud DC Architecture
2. **Server Introduction**
   - E9000
     - RH2288 V3
3. Storage Device Introduction
4. Switch Introduction

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

E9000 Positioning

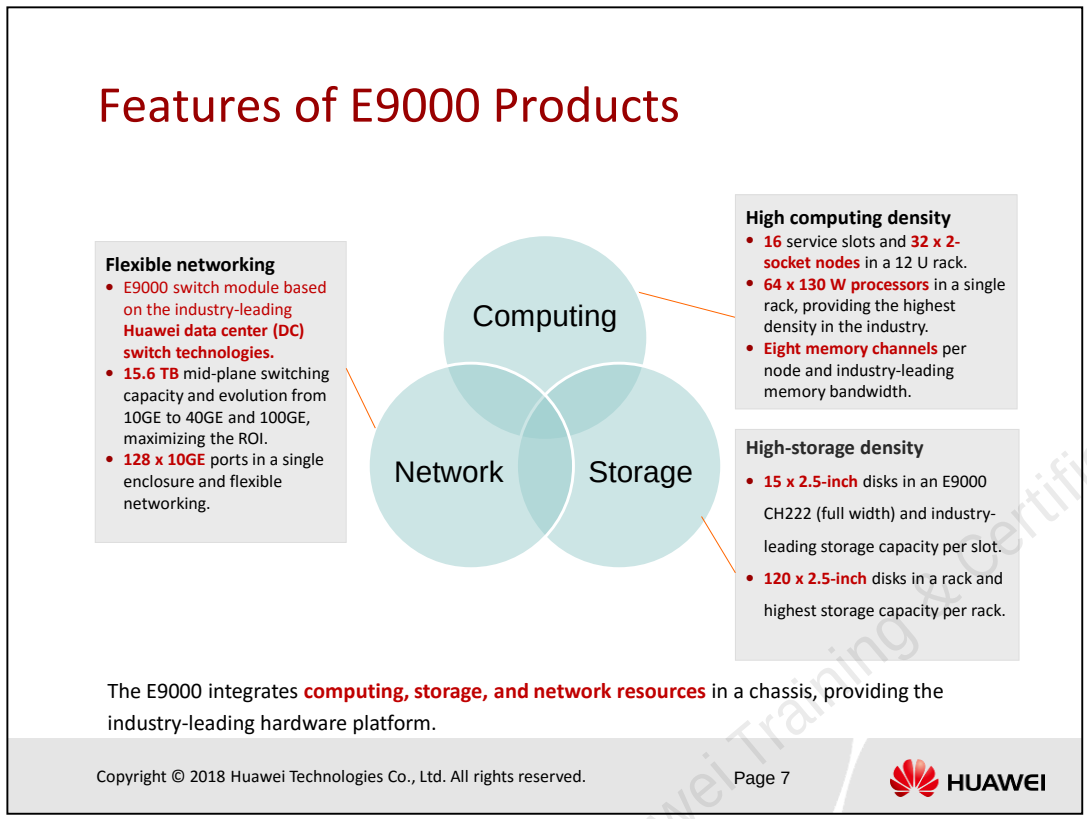


The E9000 is a high-end computing platform.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The Huawei E9000 blade server is a high-performance, high-end enterprise server designed for elastic computing and telecom computing and integrates computing, storage, and networking resources to support high-end core applications of carriers and enterprises. The E9000 is an industry-leading hardware computing platform that features reliability, availability, and serviceability (RAS), high computing density, energy saving, high mid-plane bandwidth, intelligent management and service, elastic configuration and flexible expansion for computing and storage resources, low network latency, and network acceleration. These features improve competitiveness of the E9000 server.
  - Enhances competitiveness for telecom carrier services and enterprise software services by providing equivalent quality and service capability of midrange computers.
  - Provides the same competitiveness as cost-effective servers for carrier's universal services and Internet services.
  - Provides the switching capability of high bandwidth and short latency for services that require the convergence of computing and data or media.
- The E9000 meets various workload requirements in the following typical scenarios:
  - Cloud computing: The E9000 uses high-performance CPUs and compute nodes with large-capacity memory to meet requirements for flexible virtual machine (VM) deployment. The built-in storage nodes with large capacity and low power consumption provide high-throughput shared storage to meet requirements for elastic computing applications.
  - HPC: The E9000 applies to HPC scenarios and provides InfiniBand switching capability with high performance and short latency.

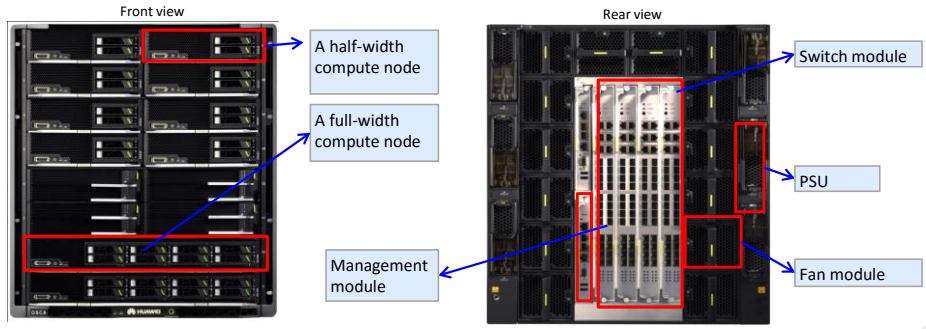
Huawei Training & Certification - Huawei Training & Certification



- The E9000, as a universal service processing platform, provides an industry-leading hardware infrastructure by integrating computing, storage, and networking resources in a 12 U chassis.
- High-performance compute node
  - Supports full series of Intel® x86 CPUs to meet requirements for CPU evolution in the future.
  - Supports compute nodes of various specifications: 2-socket, 4-socket, half-width, and full-width compute nodes.
  - Supports various network ports on mezz modules and acceleration cards to meet demands for high-performance processing.
- Built-in large-capacity storage
  - A half-width storage resource expansion module is integrated with a half-width compute node to form a full-width compute node, which integrates storage and computing resources.
  - Supports 15 x 2.5-inch Serial Attached SCSI (SAS) or Serial Advanced Technology Attachment (SATA) disks or solid-state drives (SSDs). SCSI refers to Small Computer System Interface.
  - Supports disk installation through removable drive trays and hot swap.
  - Supports power-off protection for the write cache.

Huawei Training & Certification

E9000 Hardware Overview



- 12 U: installed in a standard 19-inch cabinet and with the dimensions (H x W x D) 530.2 mm x 442 mm x 840 mm.
- 16 half-width compute nodes or 8 full-width compute nodes, implementing flexible configuration.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

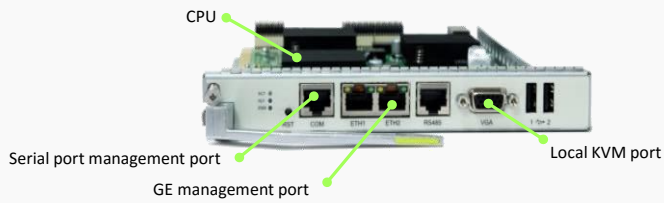
- The E9000 has the following functions and features:
  - The E9000 chassis houses eight full-width compute nodes or 16 half-width compute nodes for flexible slot configuration.
  - A half-width slot supports a compute node with two sockets and 24 DIMMs or two sets of two sockets and eight DIMMs.
  - A full-width slot supports a compute node with four sockets and 48 DIMMs.
  - A chassis provides the computing density of 64 processors or 768 cores with a maximum memory capacity of 12 TB.
  - The maximum mid-plane switch capacity reaches 15.6 Tbit/s.
  - Four switch modules (in two pairs) support Ethernet, Fibre Channel (FC), Fibre Channel over Ethernet (FCoE), and InfiniBand (IB) switching protocols and provide I/O ports on panels.
  - Each chassis is configured with two MM910s in active/standby mode. MM910s support active/standby switchover, hot swap, and the protocols of IPMI v2.0, Simple Network Management Protocol (SNMP) v3, SNMP trap v1, SNMP trap v2c, SNMP trap v3, Secure Sockets Layer (SSL), and Secure Shell (SSH). MM910s are used to manage the devices in the chassis.
  - A liquid crystal display (LCD) in the chassis front panel is used to set basic chassis information or monitor the running status of the chassis and the components in the chassis.

Huawei Training & Certification

E9000 Management Module

Management module

Ports and components on MM910



Description

MM910 description

- The MM910 manages E9000 blade server chassis and complies with IPMI v2.0 specifications. It provides functions including remote startup, shutdown, reset, logging, hardware monitoring, SOL, KVM over IP, virtual media, fan monitoring, and PSU monitoring. It also supports 1+1 redundancy.
- The MM910 provides local KVM ports for server management.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Each chassis is configured with two MM910s in active/standby mode. They support active/standby switchover and hot swap.

Huawei Training & Certification

E9000 Compute Nodes

Compute node	CH121 half-width compute node	CH222 full-width storage expansion compute node	CH240 full-width compute node	CH242 full-width compute node
				
Highlights	• **Super high performance:** a maximum of 2 x 8-core E5-2690 CPUs. • **Super high memory:** 24 x DIMM at 1.5 times the usual height and a maximum of 768 GB memory.	• **Super high memory:** a maximum of 24 x DIMMs at 1.5 times the usual height and a maximum of 768 GB memory. • **Super large storage:** 15 x 2.5-inch disks and suitable for big data processing and distributed computing.	• **Super high performance:** a maximum of four E5-4600 series CPUs. • **Super high memory:** 48 x DIMMs at 1.5 times the usual height and a maximum of 1.5 TB memory. • **Super large storage:** 8 x 2.5-inch disks and suitable for database applications that have high requirements on performance and capacity.	• **Super high performance:** a maximum of four E7-4800 series CPUs. • **Super high memory:** 32 x DIMMs at 1.5 times the usual height and a maximum of 1.0 TB memory. • **Super large storage:** 8 x 2.5-inch disks with the maximum capacity of 8 TB and suitable for database applications that have high requirements on performance and capacity.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Select a compute node as required.

Huawei Training & Certification

E9000 Switch Modules

Switch module

Switch module		
CX310	CX311	CX611
		
10GE converged switch module 16 x 10GE uplink ports 32 x 10GE downlink ports	10GE/FCoE converged switch module 16 x 10GE and 8 x 8 Gbit/s FC uplink ports 32 x 10GE downlink ports	IB QDR/FDR switch module 18 x QDR/FDR IB uplink ports 16 x QDR/FDR IB downlink ports

Highlights

- Support DC TRILL large layer 2 networks.
- Support DCB no-packet loss Ethernet to bear FCoE and iSCSI.
- Support a converged network, virtual paths, and flexible Ethernet and FC configuration.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

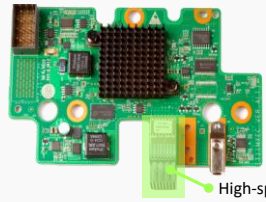
Huawei Training & Certification

- The E9000 switch modules use industry-leading Huawei DC switch technologies, which provide the mid-plane switching capacity of 15.6 TB, and support the evolution from 10GE to 40GE and 100GE. Each E9000 chassis provides 128 x 10GE upstream ports and supports various interfaces types, such as Ethernet, IB, and FC.
- The switch modules are installed in the rear slot of the E9000 chassis and connected to compute nodes, storage nodes, and management modules through the mid-plane. The switch modules exchange internal data packets and control management packets to provide high-speed data transmission.
- IB QDR/FDR and InfiniBand quad data rate/fourteen data rate
- A transparent interconnection of lots of links (TRILL) network functions as the high-speed bus in a DC and has many advantages. It can meet DC service requirements in the cloud computing era, such as random VM migration, non-blocking, low-delay data forwarding, multi-tenant, and large network scale.
- Data center bridging (DCB) enhances traditional Ethernet to prevent packet loss on Ethernet networks. The enhanced Ethernet is called lossless Ethernet, which, as the name indicates, prevents packet loss on Ethernet networks. Fibre Channel over Ethernet (FCoE) is a network convergence technology, which enables traditional storage FC network protocols to run in Ethernet networks to avoid packet loss during transmission. Actually, FCoE is an FC protocol that runs on the Ethernet networks and packet loss is not allowed. To avoid packet loss, the DCB enhancement Ethernet technology is introduced. Therefore, Ethernet and new storage FCoE protocols can run in the Ethernet networks. That is why the network is called a converged network.

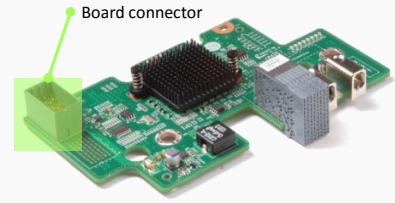
Mezz Modules on E9000 Compute Nodes

Mezz module

Front view



Axial view



Functions and models

Category	Model	Pluggable Mezz Slot	Specifications
10GE mezz module	MZ510	Mezz1/2	2 x 10GE CNA daughter cards
	MZ512	Mezz1/2	2 x (2 x 10GE) CNA daughter cards

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The mezz module is used for I/O expansion and supports NICs and SAS cards.

Huawei Training & Certification

Contents

1. Cloud DC Architecture
2. **Server Introduction**
   - E9000
   - RH2288 V3
3. Storage Device Introduction
4. Switch Introduction

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

RH2288H Server

RH2288H

Front view



Rear view



Form Factor	2U 2-Socket Rack Server
Number of CPUs	One or two
CPU model	Intel® Xeon® E5-2600/E5-2600 v2 series processors
DIMM slot	24 x DIMM slots for DDR3 RDIMMs/LRDIMMs and a maximum of 768 GB
Number of hard disks	8 or 26 x 2.5-inch SAS or SATA disks or SSDs, or 12 x 3.5-inch SAS or SATA disks and 2 x 2.5-inch SAS or SATA disks or SSDs
RAID supported	RAID 0, 1, 10, 5, 50, 6, and 60 with a RAID cache of 512 MB or 1 GB, optional battery or supercapacitor, and 3/24 hour power-off protection
PCIe expansion capacity	A maximum of seven PCIe slots
LOM	Two or four GE ports, or two 10GE ports
PSU	Two hot-swappable 460 W or 800 W AC PSUs or 800 W –48 V DC PSUs in N+1 redundancy mode
Fan module	Four hot-swappable counter-rotating fan modules in N+1 redundancy mode
Dimensions (H x W x D)	447 mm × 740 mm × 87.5 mm

Highlights

- **Super high performance:** a maximum of 2 x 8-core E5-2600 CPUs
- **High memory capacity:** 24 x DIMMs with a maximum of 768 GB memory
- **Flexible and openness:** standard PCIe expansion slots for a variety of PCIe cards

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The RH2288H supports HPC, databases, virtualization, basic enterprise applications, and telecommunication service applications thanks to its outstanding computing performance, large storage capacity, low energy consumption, high reliability, and ease of deployment and management.

Huawei Training & Certification

Contents

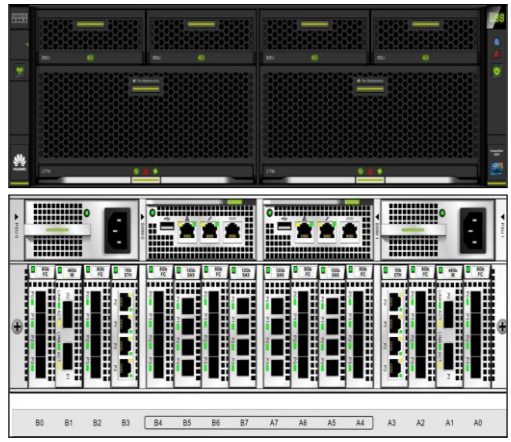
1. Cloud DC Architecture
2. Server Introduction
3. **Storage Device Introduction**
4. Switch Introduction

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

OceanStor 5600/5800 V3



- High performance:
  - PCIe 3.0 high-speed bus and SAS 3.0 I/O channel
- High scalability:
  - Hot-swappable I/O interface module
  - 4 x pluggable interface cards and 2 x onboard interface cards in a 2 U chassis
  - 16 x pluggable interface cards in a 3 U chassis
  - 24 x pluggable interface cards in a 6 U chassis
- High reliability:
  - Redundancy design for all components
  - Built-in BBU and data coffer
  - Various data protection technologies
- Energy saving:
  - Intelligent CPU frequency adjustment
  - Precise fan speed regulation

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Huawei V3 series storage devices use the latest technologies to provide the following features:
  - Use brand-new hardware architecture to provide industry-leading performance and specifications, support 16 Gbit/s FC and 10 Gbit/s FCoE host ports, and provide up to 40 GB/s of system bandwidth, 4 TB cache capacity, and 8 PB storage capacity.
  - SAN and NAS are converged to provide elastic storage, simplify service deployment, improve storage resource utilization, and reduce total cost of ownership (TCO).
  - Powerful scalability enables up to eight controllers and linear growth of performance and capacity. Loads are balanced among controllers that serve as hot backup for each other, achieving higher reliability. Resources are centrally stored, simplifying resource management.

Huawei Training & Certification

OceanStor 5600/5800 V3 Controller Enclosure



Control module
- Dual controllers
- Mainstream server platform
- Automatic frequency adjustment, reducing power consumption
- Built-in fan module (Fan modules are integrated in control modules and can be replaced independently)

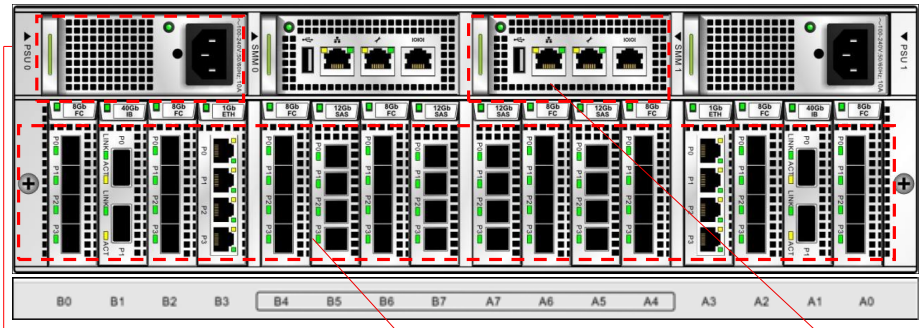
BBU module
- 5600 V3: 1+1 redundancy (the remaining two positions with BBU filler modules)
- 5800 V3: 2+1 redundancy (the remaining one position with a BBU filler module)
- DC/AC power-off protection

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Storage 5600 V3/5800 V3 uses disk and controller separation architecture (3 U independent engine). Controller modules and BBUs support redundancy and provide high availability.

Huawei Training & Certification

OceanStor 5600/5800 V3 Controller Enclosure



PSU
- 1+1 redundancy
- Conversion efficiency of up to 94%
- 240 V HVDC

Interface module
- 16 x interface module slots
- A maximum of 1250 disks
- Hot swap
- Various interface types: 16 Gbit/s FC, 12 Gbit/s SAS, GE, 10GE TOE, 10GE FCoE, and 8 Gbit/s FC

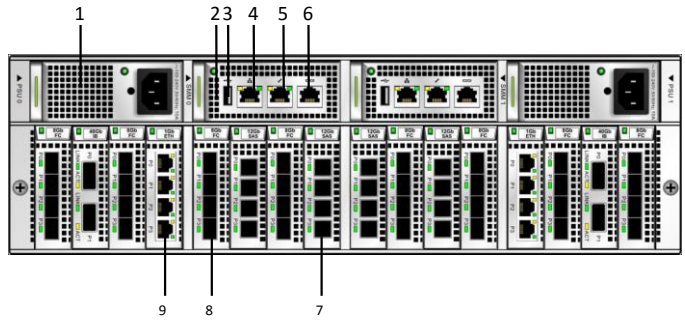
Management module
- 1+1 redundancy
- Hot swap
- Scale-out interconnection with a heartbeat mechanism

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Controller enclosures can be managed in a cascading manner.

Huawei Training & Certification

OceanStor 5600/5800 V3 Ports



1	PSU	2	Management module	3	USB (reserved)	4	Management network port
5	Maintenance network port	6	Serial port	7	12 Gbit/s mini SAS HD	8	8 Gbit/s FC port
9	GE port						

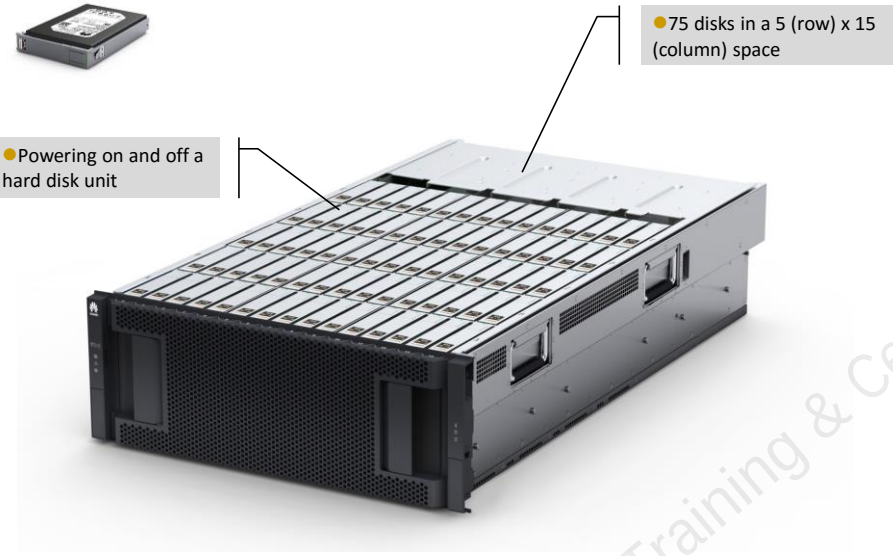
Notes:
1. One management network port can manage controllers A and B, that is, management networks are in fully-connected mode.
2. Serial ports on SMM 0 correspond to controller B and serial ports on SMM 1 correspond to controller A.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- All ports use a modular design, facilitating replacement and upgrade.

Huawei Training & Certification

4 U 3.5-Inch High-Density Disk Enclosure

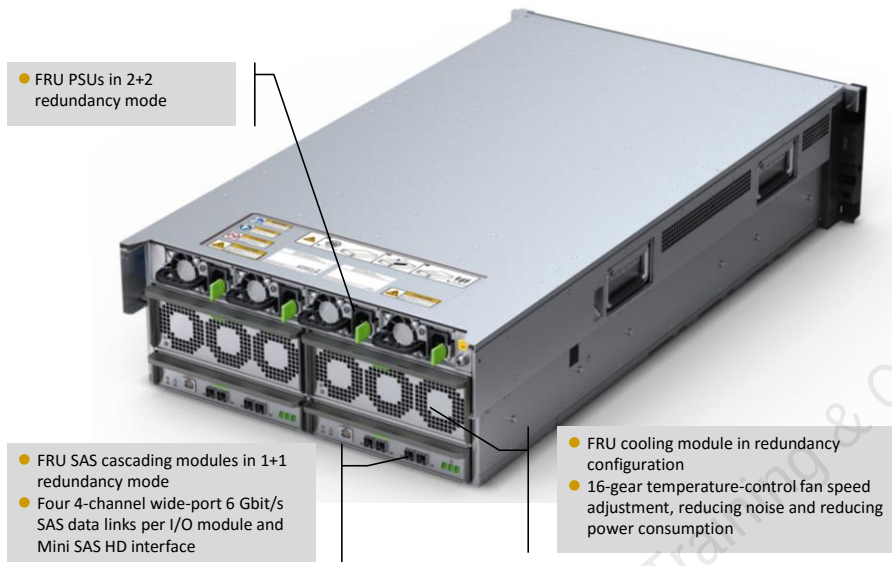


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Disk enclosures connect to controller enclosures to provide disk capacity for the controller enclosures.

Huawei Training & Certification

4 U 3.5-Inch High-Density Disk Enclosure



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Disk enclosures connect to controller enclosures using Mini SAS HD cables or are cascaded to other disk enclosures to implement capacity expansion.

Huawei Training & Certification

Contents

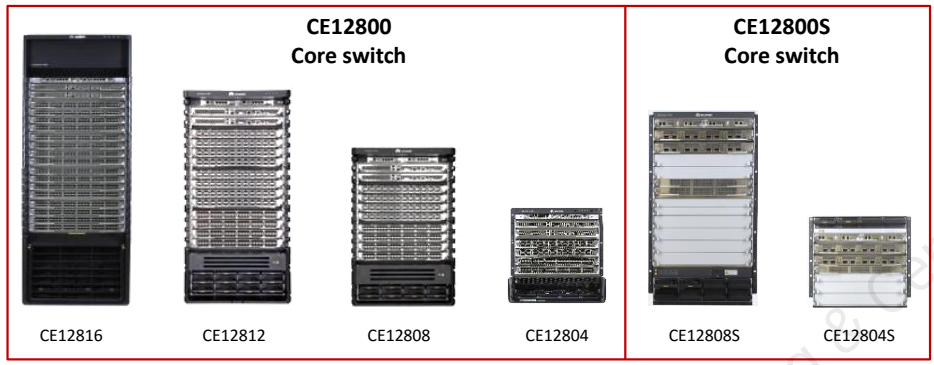
1. Cloud DC Architecture
2. Server Introduction
3. Storage Device Introduction
4. **Switch Introduction**

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

CE Series Modular Configuration Switches



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Huawei CloudEngine 12800 (CE12800 for short) series switches are Huawei's next-generation core switches for DC applications. By adopting a new architecture, CE12800 can be widely used at the aggregation and core layers of various DCs and can also serve as high-performance switching devices in campus networking.
- The CE12800 series switches include six models: CE12804, CE12808, CE12812, CE12816, CE12804S, and CE12808S. CE12804S and CE12808S are new models added in V100R005C00.
- CE12800 switches can be deployed at the core or aggregation layer in DCs. Servers connect to Top of Rack (TOR)/End of Row (EOR) switches through GE or 10GE uplinks, and TOR/EOR switches are aggregated to the CE12800 core switches through 10GE or 40GE uplinks.

Huawei Training & Certification

CE Series Fixed Configuration Switches

100GE/25GE switch



CE8860-4C-EI

40GE switch



CE7850-32Q-EI

10GE FCoE switches



CE6850U-48S6Q-HI



CE6850U-24S2Q-HI

10GE access switches



CE6851-48S6Q-HI



CE6850-48T6Q-HI



CE6810-48S4Q-LI



CE6810-32T16S4Q-LI



CE6810-48S-LI



CE6810-24S2Q-LI

GE access switches



CE5855-48T4S2Q-EI



CE5855-24T4S2Q-EI



CE5850-48T4S2Q-HI



CE5850-48T4S2Q-EI



CE5810-24T4S-EI



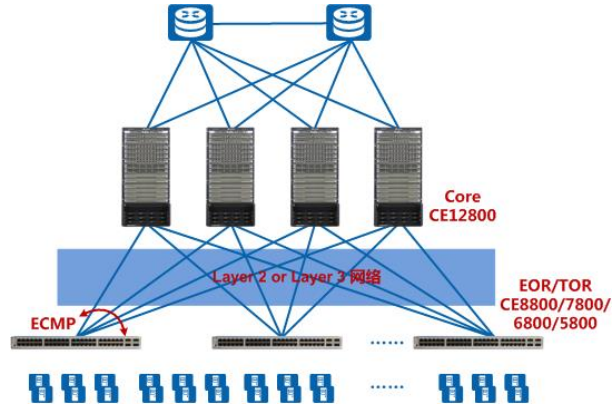
CE5810-48T4S-EI

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Core and Aggregation Layers in DCs: Fat-Tree Networking



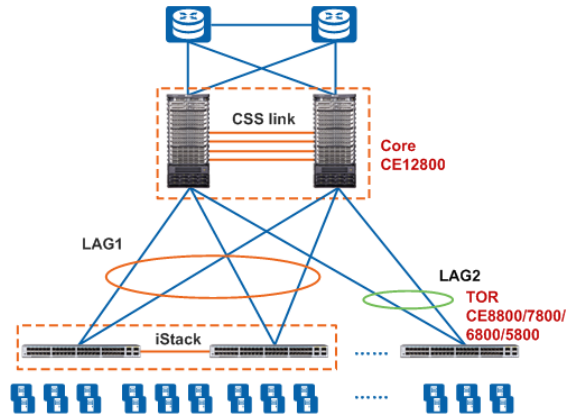
- In the fat-tree networking model, servers are dual-homed or single-homed to TOR switches through GE/10GE/25GE ports. TOR switches connect to multiple (two or four) core switching planes through 10GE/25GE/40GE uplinks. Each core switching plane has one core switch, which connects to egress routers through 10GE/25GE/40GE/100GE uplinks. In specific scenarios, core switches can also function as egress routers.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- North-to-south traffic between external clients and internal servers predominates in the traffic model of a traditional DC. A DC egress has a strict bandwidth limit. Therefore, traditional DCs often use a three-layer networking pattern consisting of access, aggregation, and core layers. In this pattern, the network has a high convergence ratio.
- The wide use of 10GE servers and VM load balancers rapidly increases the volume of east-to-west traffic in DCs. The traditional three-layer networking architecture cannot meet the requirements of the new traffic model. Therefore, the fat-tree networking based on the CLOS architecture was developed to reduce the convergence ratio.
- A Layer 2 or Layer 3 network can be deployed between TOR switches and core switches according to service requirements and network design. If a Layer 3 network is deployed, the IGP (OSPF for example) protocol can run between TOR and core switches. TOR switches are uplink distributed to different core switches based on IP addresses through ECMP.
- If a Layer 2 network is deployed, Multiple Spanning Tree Protocol (MSTP) can run between TOR and core switches to prevent loops. However, MSTP brings risks of link congestion, bandwidth waste, and slow convergence. For improved link efficiency and reliability of the DC network, the Transparent Interconnection of Lots of Links (TRILL) protocol can be used to build a large-scale Layer 2 loop-free network. TOR switches use uplink ports to connect to different core switches through ECMP.

Huawei Training & Certification

Core and Aggregation Layers in DCs: CSS Networking



- Servers are single-homed or dual-homed to TOR switches through GE/10GE ports. You can configure TOR switches one by one or configure multiple TOR switches to form a stack using the iStack function. TOR switches (or the TOR stack) connect to the CSS through the LAG consisting of 10GE or 40GE links. The CSS has two CE12800 switches, which establish a CSS link between service ports (10GE or 40GE ports) to provide inter-chassis connections for horizontal traffic of core switches.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- CSS virtualizes multiple physical core switches into one logical core switch. This virtualization combines multiple independent links used when TOR switches are dual-homed or multi-homed to multiple devices into a link aggregation group (LAG) of a logical device. Subsequently, TOR switches and the logical core switch establish a point-to-point connection. In this situation, MSTP and other complex protocols do not need to be deployed. Compared to fat-tree, multi-plane networking and TRILL networking, CSS networking is a flexible, lightweight deployment mode, which applies to small- and medium-sized DCs.
- In CSS networking, TOR switches (or TOR stack) connect to the CSS like two switches connect to each other. Therefore, a Layer 2 or Layer 3 network can be flexibly deployed.

Huawei Training & Certification

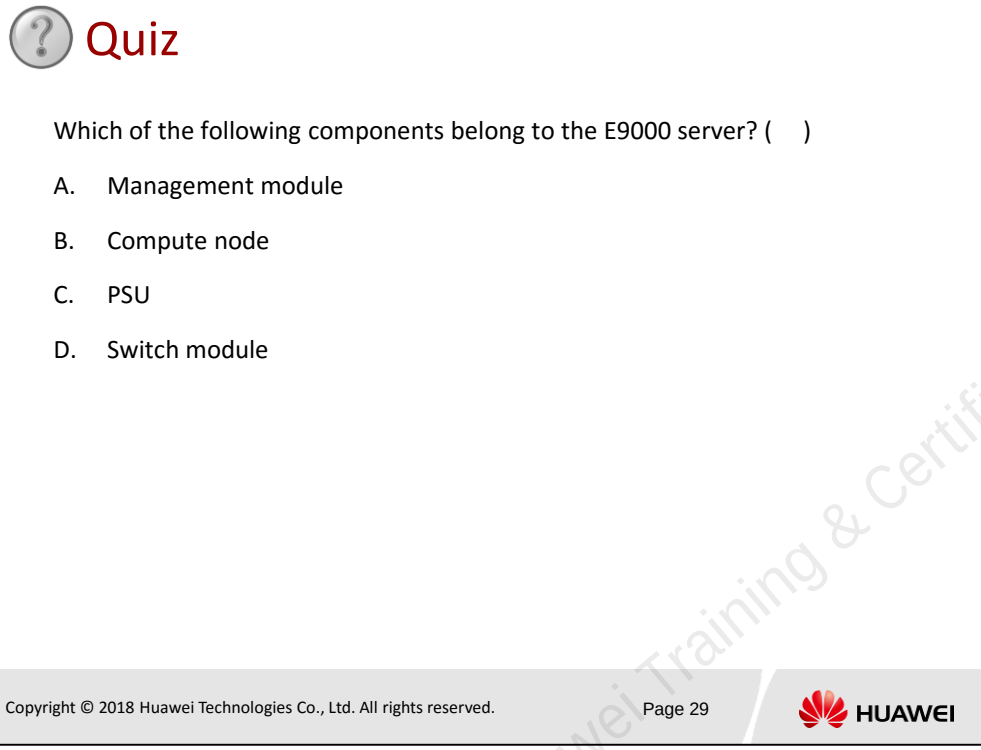
Summary

- Cloud DC architecture
- Server introduction
- Storage device introduction
- Switch introduction

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification



- Answer: ABCD

Huawei Training & Certification
Huawei Training & Certification

Recommendations

- Huawei Learning Website
  - http://support.huawei.com/learning/en/newindex.html
- Huawei Support Case Library
  - http://support.huawei.com/enterprise/servicecenter?lang=en

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

Stack Feature of CE Series Switches

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Foreword

- A single switch cannot meet requirements of the increasing access volume of the data center and the network reliability. To meet the forwarding requirement of the data center, the stacking technology is introduced.
- A stack consists of multiple switches that support the stacking function, and is equivalent to a logical switch. The switches are connected by stack cables to forward packets as a logical switch. An IP address is provided to users to manage and maintain the stack.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Objectives

- Upon completion of this course, you will be able to know:
- Stack working mechanism
- Implementation on CE switches
- Differences between Implementations on TOR and CE12800
- Configurations on TOR
- Configurations on CE12800

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

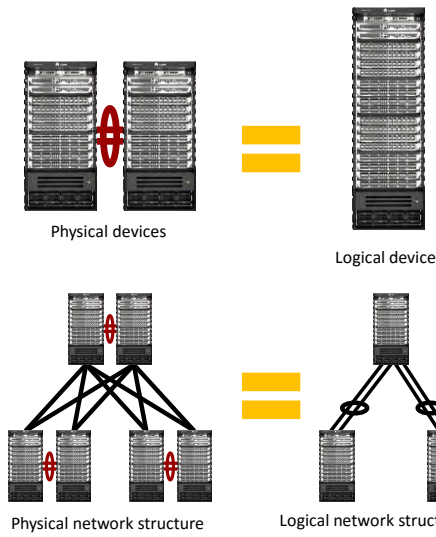
1. **Stack Overview**
2. Implementation Mechanism
3. Applications
4. Data Forwarding and Failover
5. Basic Configurations

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

What Is CSS?



CSS definition

Multiple switches are connected with stack cables to form a logical switch.

CSS features

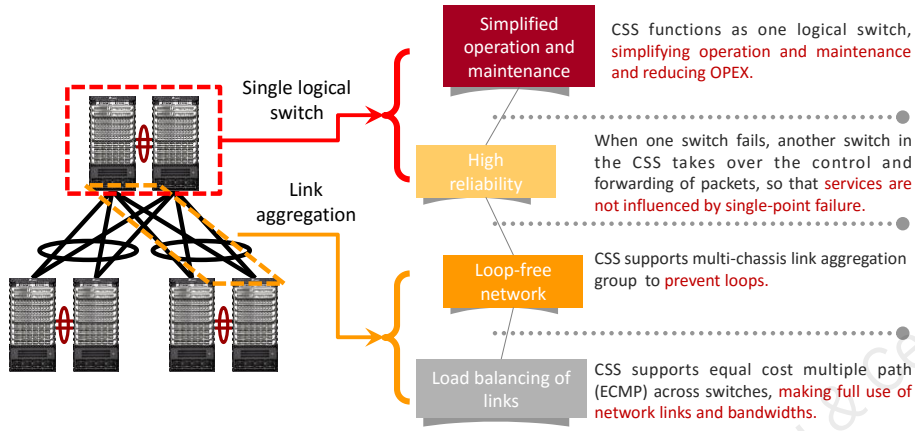
- Many-to-one virtualization: Multiple switches are virtualized into one logical switch that has only one control plane and provides unified management.
- Unified forwarding plane: CSS uses a unified forwarding plane that shares and synchronizes forwarding information.
- Multi-chassis link aggregation group : Links between an upstream and a downstream CSS are aggregated to one Eth-trunk link.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Cluster Switch System (CSS)
  - Multiple switches are connected with stack cables to form a logical switch with high performance.
- CSS features
  - Many-to-one virtualization: Multiple switches are virtualized into one logical switch that has only one control plane and provides unified management.
  - Unified forwarding plane: CSS uses a unified forwarding plane that shares and synchronizes forwarding information.
  - Multi-chassis link aggregation group : Links between an upstream and a downstream CSS are aggregated to one Eth-trunk link.

Huawei Training & Certification

What Can CSS Bring to Us?



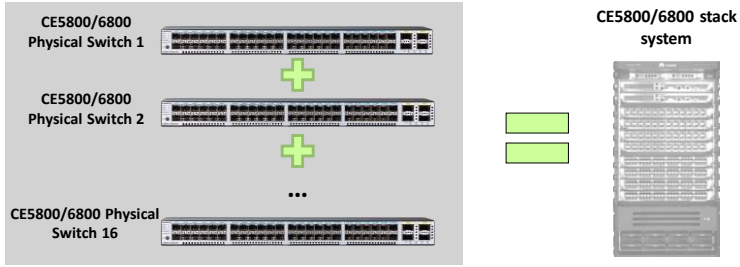
Preferred solution for small- and medium-range data centers to build easy-to-manage and efficient network

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- CSS virtualizes multiple switches into one logical device and supports inter-chassis link aggregation. It simplifies network topology and greatly improves network performance:
  - Simplified operation and maintenance: CSS functions as one logical switch, simplifying operation and maintenance and reducing OPEX.
  - High reliability: When one switch fails, another switch in the CSS takes over the control and forwarding of packets, so that services are not influenced by single-point failure.
  - Loop-free network: CSS supports multi-chassis link aggregation group to prevent loops.
  - Load balancing of links: CSS supports equal cost multiple path (ECMP) across switches, making full use of network links and bandwidths.
- CSS simplifies network architecture, improves forwarding performance, and does not lead to the loss of any network functions. CSS has all functions of the physical switches in the cluster and provides better performance. CSS gains wide customer recognition and becomes a preferred solution for simple and efficient network deployment.

Huawei Training & Certification

IStack Classification and Features



- iStack of box switches

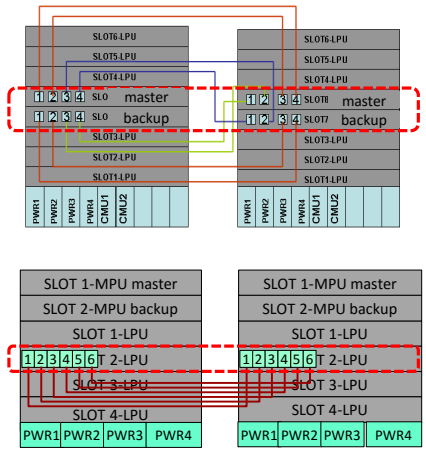
- **Deployment**: Box switches are connected by high-speed cooper cables or optical fibers to form a ring or chain. The stack ports must provide 10 Gbit/s or higher transmission speed.
- **Benefit**: No dedicated stack card or stack cable is required. Bandwidth is improved and loops are prevented.
- **Typical products:** CE5800，CE6800 and CE7800

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Multiple CE5800/6800/7800 switches form a stack system, which is equivalent to a chassis switch.

Huawei Training & Certification

CSS Classification and Features



- CSS using dedicated line cards
  - Deployment method: Switches that use dedicated line cards are connected by dedicated cables.
  - Advantages: Does not occupy service slots and increases bandwidth using high-speed interfaces.
  - Typical product: S9300
- CSS using service line cards
  - Deployment method: Switches that use service line cards are connected by standard cables.
  - Advantages: Does not need to buy extra components, allows switches far from each other to form a cluster, and increases bandwidth using trunk interfaces.
  - Typical product: CE12800

CSS using service line cards shines with greatly improved switching performance

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- CSS has gone through two development stages.
- At the early stage, switches set up a CSS using dedicated line cards. Physical switches are interconnected by dedicated line cards and stack cables. Because the dedicated line cards do not occupy slots for service line cards, they do not deteriorate forwarding performance of the entire system. Meanwhile, interfaces on dedicated line cards have a higher transmission rate than interfaces on service line cards, improving bandwidth for CSS interconnection.
- At the later stage, the physical switches set up a CSS using service line cards and are interconnected by standard cables. Although several service interfaces are used for interconnection, this mode has a slight impact on the forwarding performance of the entire system because the forwarding capability of switches has greatly improved. The use of standard service interfaces and standard cables allows switches far from each other to form a cluster, making CSS deployment flexible. Moreover, the interconnected interfaces support link aggregation, which allows flexible expansion of interconnection bandwidth.

Huawei Training & Certification

Contents

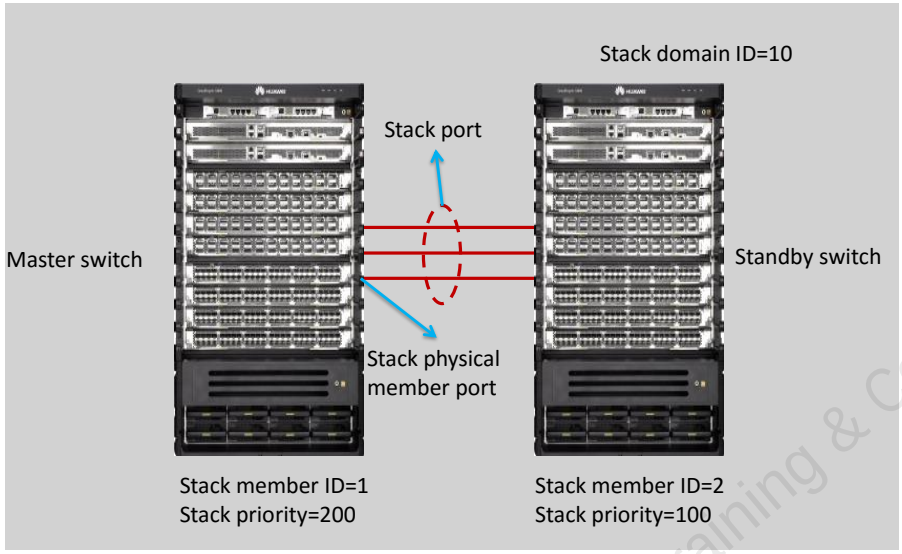
1. Stack Overview
2. **Implementation Mechanism**
   - CE12800 and CE12800S
   - CE5800, CE6800 and CE7800
3. Applications
4. Data Forwarding and Failover
5. Basic Configurations

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

CSS Basic Concepts

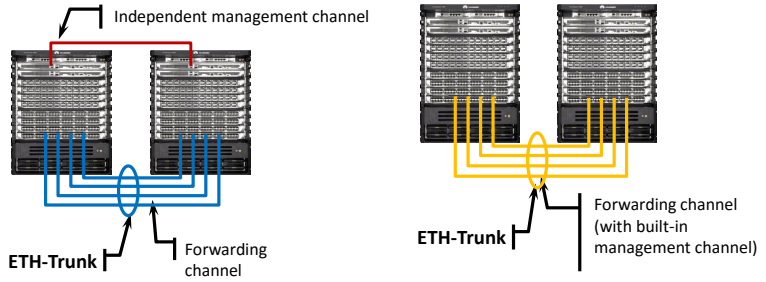


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Roles
- Switches that have joined a stack are member switches. Each member switch in a stack plays one of the following roles:
  - Master switch
    - The master switch manages the entire stack. A stack has only one master switch.
  - Standby switch
    - The standby switch is a backup of the master switch. When the master switch fails, the standby switch takes over all services from the master switch. A stack has only one standby switch.
- Stack domain
  - After switches are connected using stack links and set up a stack, they form a stack domain. Multiple stacks can be deployed on a network to support various applications. These stacks are identified by their domain IDs.
- Stack member ID
  - Stack member IDs are used to identify and manage member switches in a stack. Each member switch in a stack has a unique member ID.

Huawei Training & Certification

CSS Physical Structure



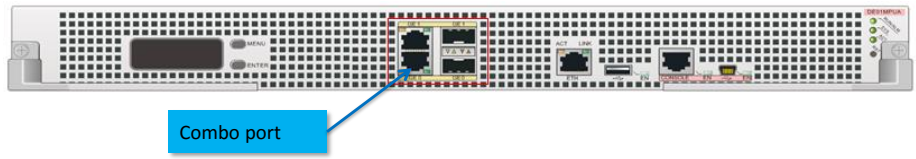
- Flexible CSS architecture
  - Flexible combination: CSS allows CE12800s to form a cluster, even if they are different models.
  - Two modes: The management channel supports two deployment modes: inband and outband.
  - Multiple bandwidths: The forwarding channel supports 10GE interfaces, 40GE interfaces, and 100GE interfaces.
- CSS high-speed interconnection
  - Aggregates 32 interfaces into a CSS forwarding channel.
  - The forwarding channel supports 100 Gbit/s forwarding speed on a single interface.
  - The unidirectional bandwidth of the CSS as a whole can reach 3.2Tbit/s.

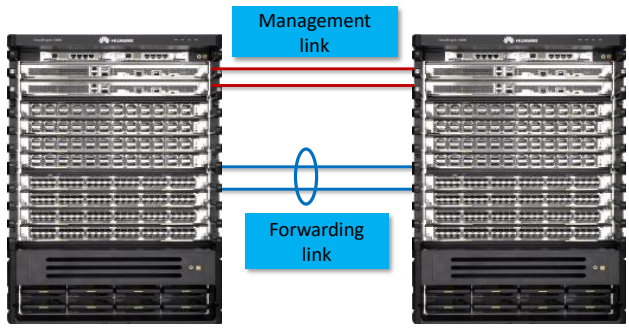
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

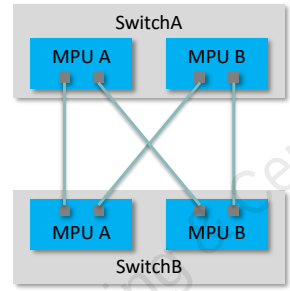
- Concepts
  - Management channel: It is a channel for data exchange between the management plane and control plane. It is responsible for device management, route synchronization, and status synchronization between CSS switches.
  - Forwarding channel: It is a channel for data exchange between the forwarding planes of CSS switches. It is responsible for packet forwarding and MAC address synchronization.
- Advantages
  - Outband Management Channel: The GE interfaces on the MPUs of two switches are connected to function as the management channel. It supports outband management, provides high reliability, and does not occupy service resources.
  - Inband Management Channel: The management channel is deployed within the forwarding channel. It simplifies network deployment, saves cabling costs, and allows switches far from each other to form a cluster.

Huawei Training & Certification

Management Port







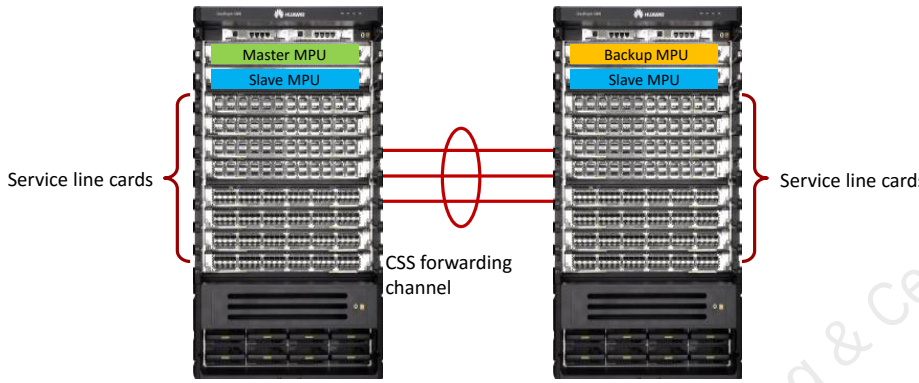
SIP Port

SIP port connections

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Duties of Each Component

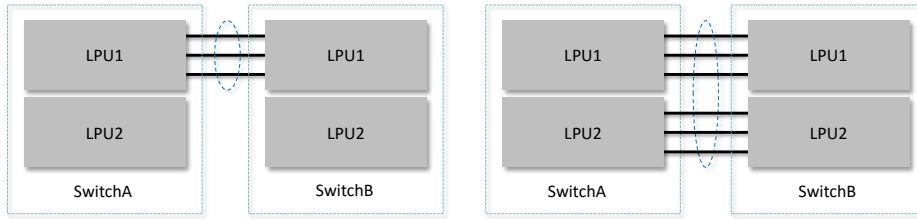


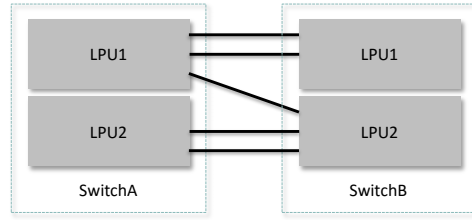
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- MPU
  - Master MPU: Manages the CSS.
  - Backup MPU: As a hot backup to the master MPU, the backup MPU manages the CSS when the master MPU fails.
  - Slave MPU: As a cold backup to the MPU, one slave MPU automatically becomes the backup MPU when the master or backup MPU fails.
- CSS forwarding channel
  - Be responsible for service forwarding and data synchronization.
  - All links between service line cards are aggregated to an Eth-trunk link.
  - The forwarding channel can be set up between interfaces in different slots.
- Service line cards
  - The service line cards provide access and forwarding services.
  - The interfaces of CSS forwarding channel cannot be used for service access.

Huawei Training & Certification

Service Port Connections



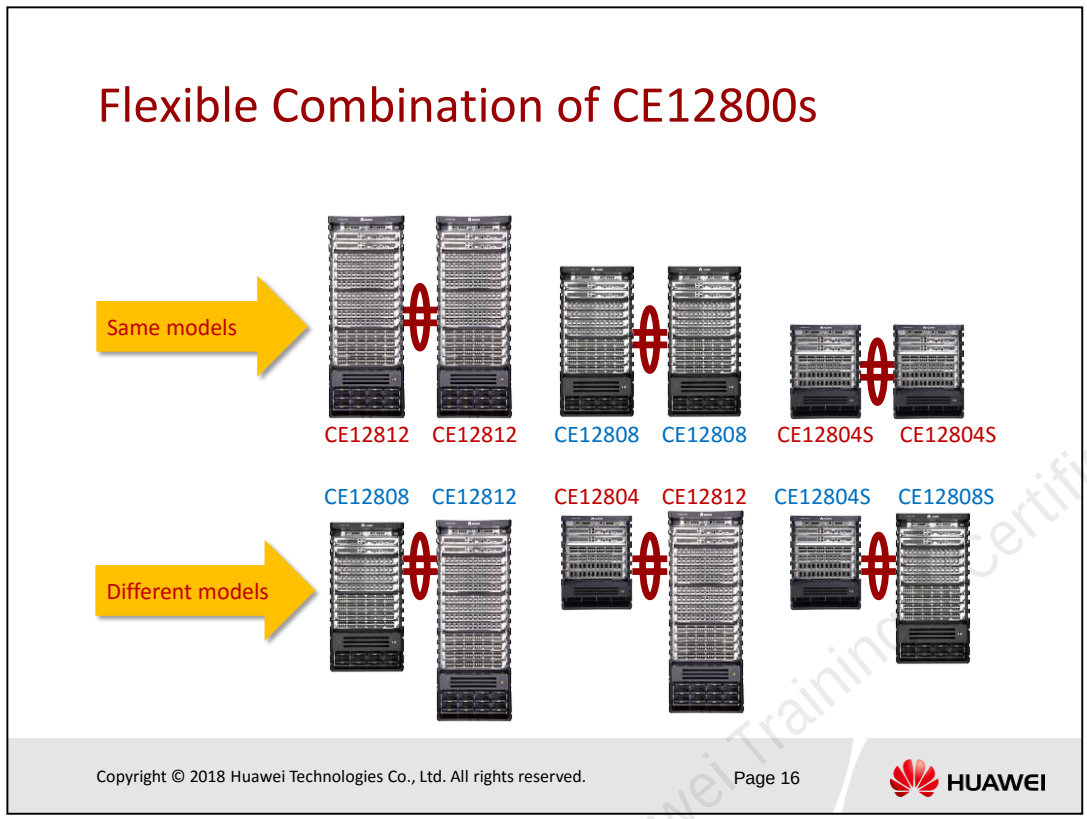


Cross connection mode

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- A logical stack port can contain physical member ports on the same LPU or different LPUs. A maximum of 32 physical member ports can be added to a stack port to improve stack link bandwidth and reliability. There are two networking forms according to the location of physical member ports, as shown in the figure.
  - 1+1 networking: Physical member ports are located on one LPU. Connect the two LPUs of the two switches to form a stack.
  - N+M networking (N ≥ 2, M ≥ 2): Physical member ports are located on multiple LPUs, and stack links of different LPUs back up each other.
- NOTE: The N+M networking is more reliable and is recommended.
- In N+M networking, physical member ports on an LPU of the local switch can be connected to multiple LPUs of the peer switch. That is, the cross connection mode is supported.

Huawei Training & Certification



- CE12800 switches of different models can set up a stack. For example, a CE12804 switch and a CE12808 can set up a stack.
- CE12800S switches of different models can set up a stack. For example, a CE12804S switch and a CE12808S can set up a stack.
- CE12800 and CE12800S switches cannot set up a stack.

Huawei Training & Certification

Software and Hardware Requirements (1)

Item	Requirement	Remarks
Switch model	CE12804, CE12808, CE12812, CE12816 CE12804S, CE12808S	•CE12800 switches of different models can set up a stack. For example, a CE12804 switch and a CE12808 can set up a stack. •CE12800S switches of different models can set up a stack. For example, a CE12804S switch and a CE12808S can set up a stack. •CE12800 and CE12800S switches cannot set up a stack.
Number of member switches	2	-
Number of MPUs in each switch	At least one MPU in each switch	It is recommended that you install two MPUs in each switch to improve system reliability.
Type of ports used for stack connection	10GE optical ports 10GE electrical ports 40GE optical ports 100GE optical ports	•10GE optical ports can be used for stack connection only when they have 10GE optical modules installed. If the GE optical or copper modules are installed on 10GE optical ports, the ports cannot be used to set up a stack. •10GE optical ports derived from 40GE optical ports can be used for stack connection. If 40GE optical ports have been configured as the stack physical member ports, they cannot be split. •10GE and 40GE optical ports derived from 100GE optical ports can be used for stack connection.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Software and Hardware Requirements (2)

Number of physical member ports in a stack port	1-32	•100GE, 40GE and 10GE optical ports cannot be added to the same stack port. •10GE optical ports and 10GE electrical ports can be added to the same stack port. •It is recommended that you add at least two physical member ports to a stack port to improve stack link bandwidth and reliability.
•High-speed cables used for stack connection	•10GE optical port •SFP+ to SFP+ high-speed cable •SFP+ to SFP+ high-speed cables are available in lengths of 1 m, 3 m, 5 m, 7 m, and 10 m, among which 7 m and 10 m SFP+ to SFP+ high-speed cables are active cables. •40GE optical port •QSFP+ to QSFP+ high-speed cable •QSFP+ to QSFP+ high-speed cables are available in lengths of 1 m, 3 m, and 5 m. •100GE optical port •CXP to CXP high-speed cable •CXP to CXP high-speed cables are available in lengths of 1.5 m and 3 m.	You can connect stack member switches using high-speed cables or optical modules and fibers.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Software and Hardware Requirements (3)

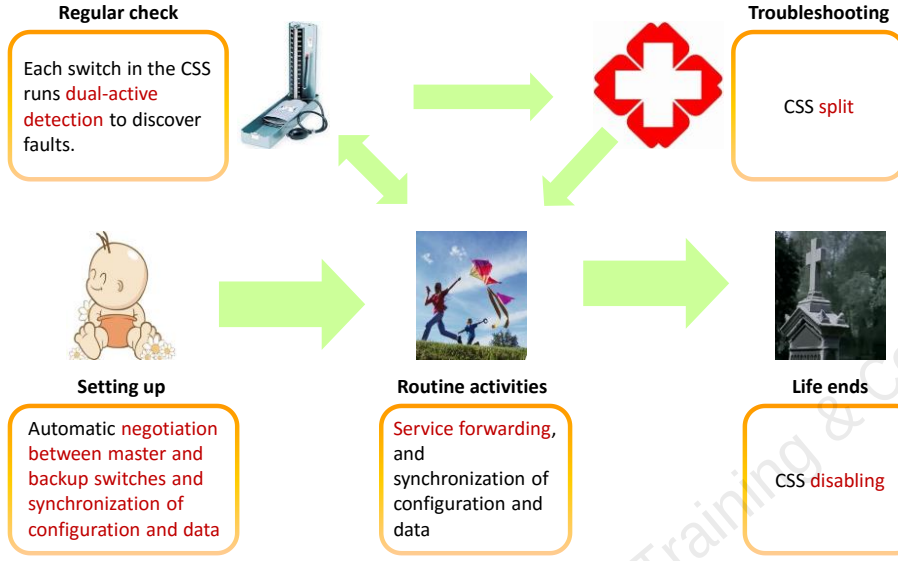
Optical modules used for stack connection	•10GE optical port •10GE SFP/SFP+ optical module •40GE optical port •40GE QSFP optical module The required optical fibers are determined by the optical modules you select.	When SFP/SFP+ optical modules are used, the maximum single-hop transmission distance is up to 80 km. When QSFP+ optical modules are used, the maximum single-hop transmission distance is up to 10 km. (Single-hop transmission distance means the transmission distance between two devices without a relay device between them.)
Network cables used for stack connection	10GE electrical ports use Category 6 network cables that comply with IEEE 802.3an.	The maximum single-hop transmission distance of a network cable is 100 m.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

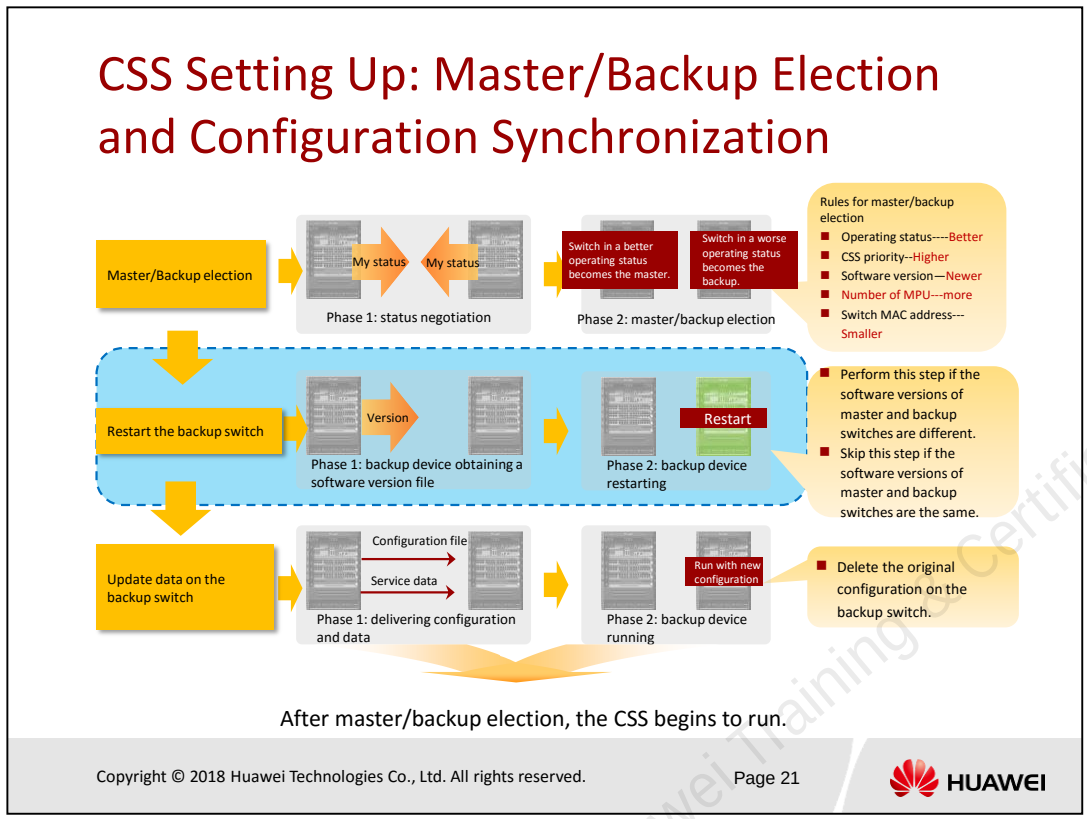
CSS Life Cycle



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification



- After two switches are connected using stack cables and configured with required stack parameters, they can set up a stack.
- Role Election
  - After a stack is set up, member switches exchange stack competition packets to elect a master switch. The switch that wins the competition becomes the master switch and manages the entire stack. The other switch becomes the standby switch and works as a backup of the master switch.
  - The two switches compare the following items in the listed order to elect the master switch (the election ends when a winning switch is found):
    - Running status: The switch that completes startup and enters stack running state first becomes the master switch.
    - Stack priority: The switch with a higher stack priority becomes the master switch.
    - Software version: The switch running a later software version becomes the master switch.
    - Number of Main Processing Units (MPUs): The switch with two MPUs is preferred over the switch with only one MPU.

Huawei Training & Certification

Inter-Device Link Aggregation

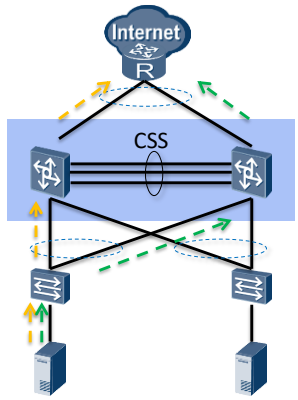


Figure 1, Data forwarding before a link failure

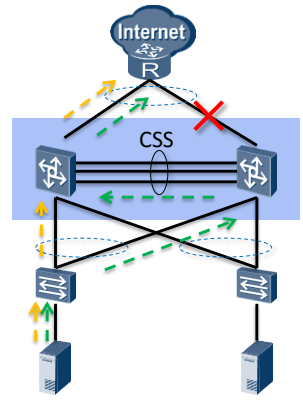


Figure 2, Data forwarding after a link failure

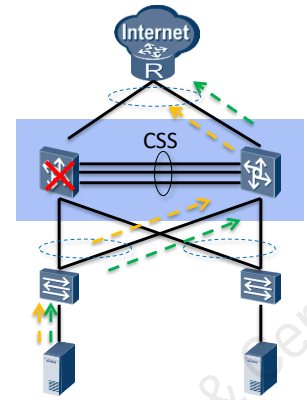


Figure 3, Data forwarding after a device failure

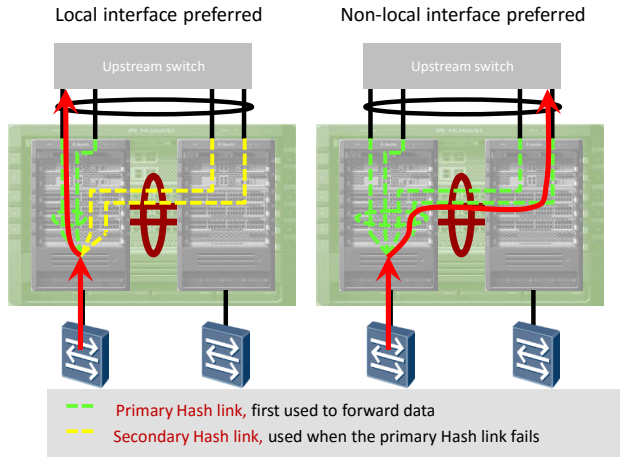
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification

- A stack supports inter-device link aggregation (Eth-Trunk). That is, Ethernet ports on different member switches can be bound to one Eth-Trunk. The Eth-Trunk link still works when a member switch or a member link in the Eth-Trunk fails, ensuring reliable data transmission. Inter-device link aggregation prevents single-point failures in a stack and greatly improves network reliability.
- As shown in Figure 2, traffic sent to the core device on the network is equally distributed to member links of an Eth-Trunk set up between the stack member switches. When an Eth-Trunk member link fails, traffic on this link is distributed to the other link. This link backup mechanism improves network reliability.
- As shown in Figure 3, when a member switch in the stack fails, traffic is switched to the Eth-Trunk member link on the other member switch. This device backup mechanism improves network reliability.

Local Interface Preferred



- Forwarding local packets first
  - **Packet forwarding**: CSS prefers to forward packets through a trunk interface on a local device.
  - **Advantage**: Reduces the burden of the forwarding channel and the forwarding delay.

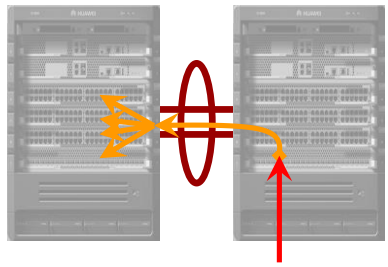
Fine-grained service processing and highly efficient resource utilization

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

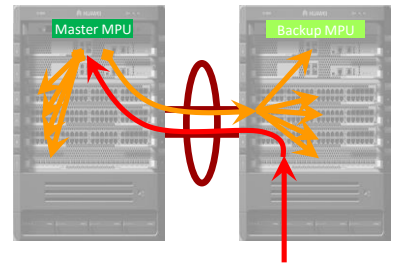
- When an inter-device Eth-Trunk is configured in a stack, the stack uses the hash algorithm to select outbound interfaces in the Eth-Trunk. Therefore, traffic received on a member switch may be forwarded through the other member switch. Inter-device forwarding consumes bandwidth on stack links. As bandwidth provided by a stack cable is limited, this forwarding mode increases loads on stack cables and reduces forwarding efficiency. Local preferential forwarding can solve this problem. This feature ensures that traffic reaching the local switch is preferentially forwarded through a local interface. If the local outbound interface fails, traffic is forwarded through an interface on the other member switch.
- NOTE: This function is only valid for known unicast packets, and is invalid for unknown unicast packets, broadcast packets and multicast packets.

Huawei Training & Certification

Configuration and Data Synchronization



- Data synchronization without the control plane involved
  - synchronization: Data synchronization between line cards of the same switch needs no involvement of the forwarding channel. The line cards on both master and backup switches can trigger data synchronization. Data synchronization between line cards of different switches needs involvement of the forwarding channel.
  - Typical service: MAC entry synchronization



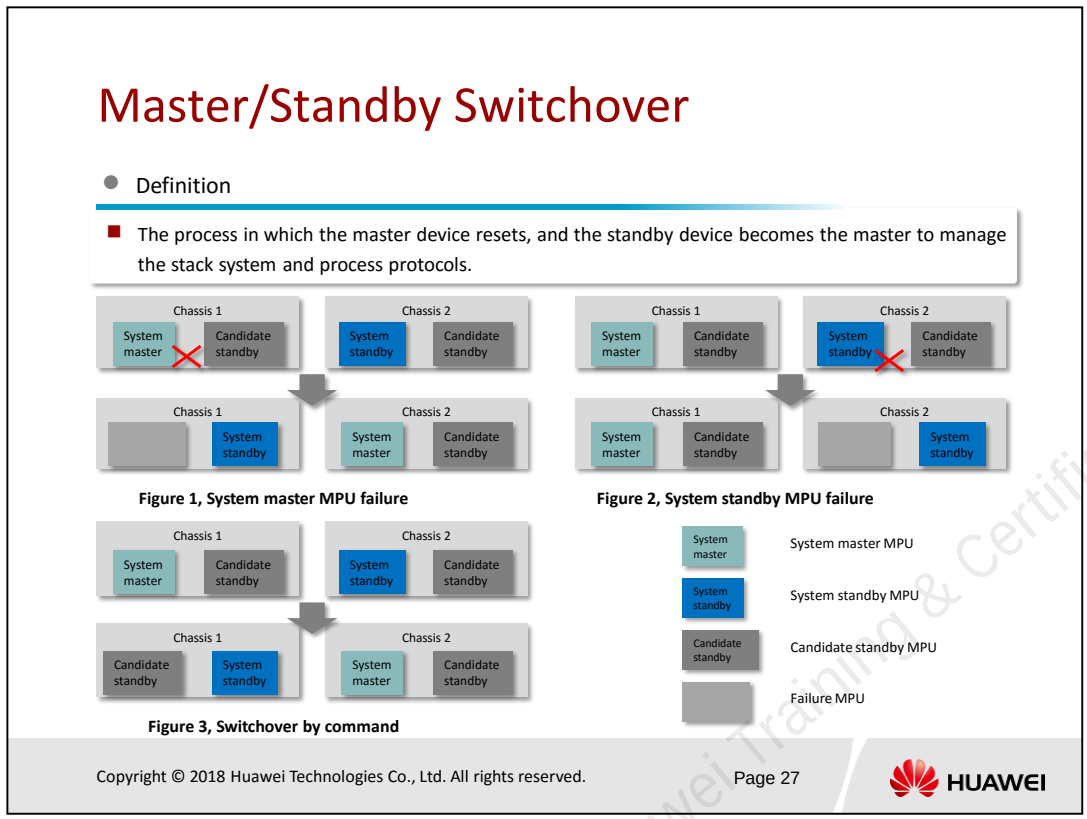
- Data synchronization with the control plane involved
  - synchronization: Data is sent to the master switch, processes on the control plane by the master MPU, and then delivered to all line cards and the backup MPU. Data synchronization between line cards of different switches needs involvement of the management channel.
  - Typical service: Synchronization of the routing table and ARP table

Real time synchronization of configuration Forwarding information in the system as a whole

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

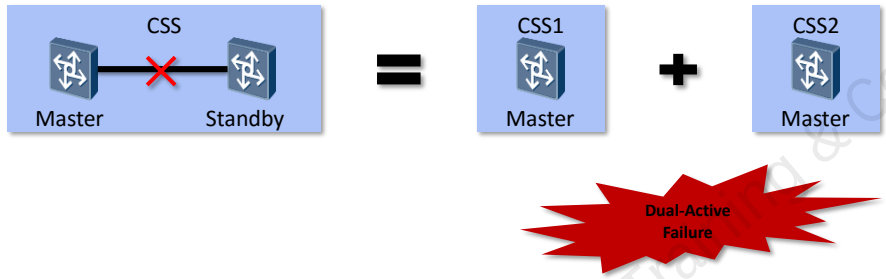


- Master/Standby Switchover Triggered by an MPU Failure
- Roles in a stack may change if an MPU in the stack fails.
- Figure 1 shows the changes of roles in a stack after the system master MPU fails.
  - The original standby switch becomes the master switch, and the original system standby MPU becomes the system master MPU.
  - The original master switch becomes the standby switch.
  - The standby MPU of the original master switch becomes the system standby MPU and synchronizes data with the system master MPU.
- Figure 2 shows the changes of roles in a stack after the system standby MPU fails.
  - The master and standby switches retain their roles.
  - The standby MPU of the standby switch becomes the system standby MPU and synchronizes data with the system master MPU.
- Failures of candidate standby MPUs do not cause any change of roles in the stack.
- Master/Standby Switchover Triggered by Commands
- Figure 3 shows the changes of roles in a stack after a master/standby switchover is triggered by a command.
- The original standby switch becomes the master switch, and the original

Huawei Training & Certification

Stack Split

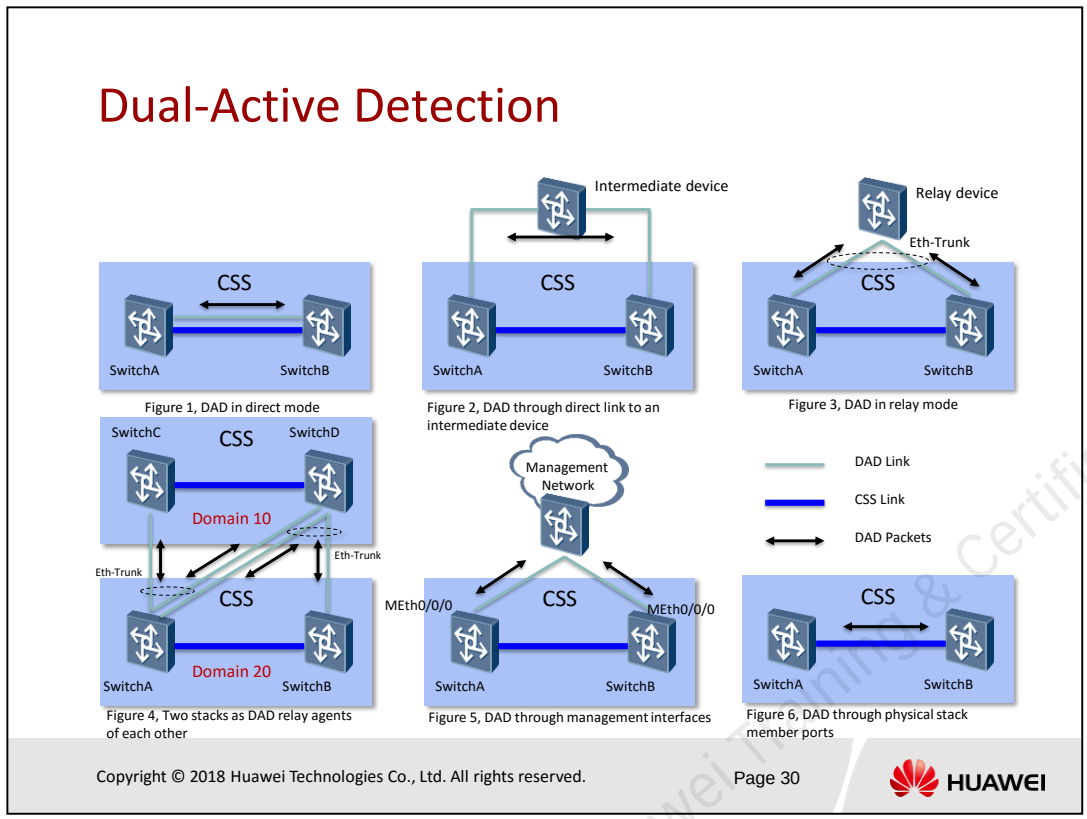
- After a stack is set up, the master and standby switches periodically send heartbeat packets to maintain the stack state. If communication between the two switches is interrupted due to failures of stack cables or MPUs or power failure or restart of a switch, the stack splits into two standalone switches, as shown in



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- After a stack splits, the two switches use the same global configuration if they are running normally. In this case, the two switches use the same IP address and MAC address to communicate with other network devices. The address conflict causes a communication failure on the entire network. Dual-active detection (DAD) can be configured to ensure that only one master switch exists after the stack splits.

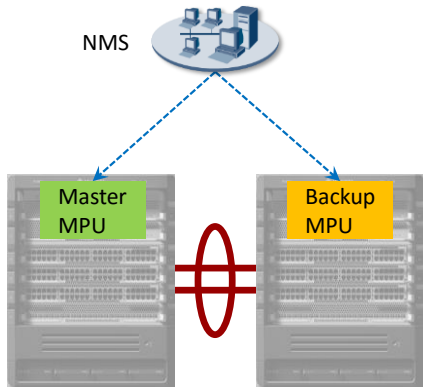
Huawei Training & Certification



- Dual-active detection (DAD) is a protocol that can detect stack split and dual-active situations and take recovery actions to minimize impact of a stack split on services.
- DAD can be implemented in the following modes:
- Direct mode
  - In direct mode, DAD is performed through dedicated direct links between member switches, as shown in Figure 1.
  - The direct detection links can also be connected through an intermediate device, as shown in Figure 2. In direct mode, DAD packets are bridge protocol data units (BPDUs), so the intermediate device must be configured to transparently transmit BPDUs. For details on the configuration method, see Configuring Interface-based Layer 2 Protocol Transparent Transmission in the CloudEngine 12800 Series Switches Configuration Guide - Ethernet.
- Proxy mode
  - In relay mode, DAD detection is performed through an inter-device Eth-Trunk link connected to a relay agent, as shown in Figure 3. The DAD proxy function must be enabled on the relay agent. Compared with the direct mode, the relay mode does not require additional interfaces because the Eth-Trunk interface can

Huawei Training & Certification, Huawei Training & Certification

CSS Disabling



- CSS disabling

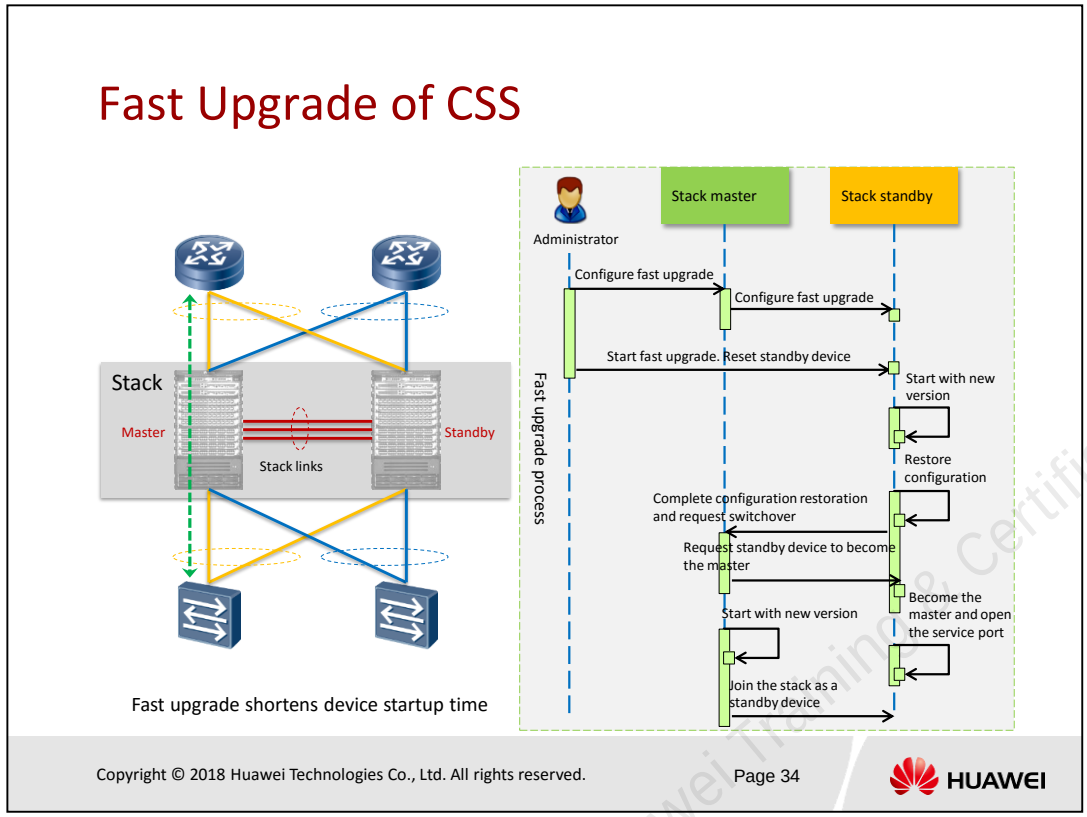
1. The operator disables CSS on the master and backup switches.
2. The operator restarts the master and backup switches.
3. After the two switches are started, they become independent to each other. The operator configures each of them separately and manages them as independent devices.

Two switches run independently after CSS is disabled

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification



- A stack can be upgraded using the traditional upgrade method (specify the next-startup files and restart the entire stack) or the fast upgrade or in-service software upgrade (ISSU) function.
- These stack upgrade methods are described as follows: Traditional upgrade method: You need to specify next-startup files and restart the entire stack. This method causes service interruption in a long time and is therefore not applicable to scenarios requiring short service interruption time.
- Fast upgrade: This upgrade method provides a mechanism to minimize the service interruption time during software upgrade of stack member switches, reducing impact of the upgrade on services.
- NOTE: It is recommended that the upstream and downstream devices be connected to the stack through Eth-Trunk links to reduce the traffic interruption time during an upgrade.
- Figure shows traffic forwarding during a fast upgrade. First, the standby switch restarts with the new system software. Data traffic is forwarded by the master switch in this period. After the standby switch is upgraded, it becomes the master switch and starts to forward data traffic. Then the original master switch restarts with the new system software. After the upgrade is complete, the original master switch becomes the backup switch in the stack.

Huawei Training & Certification

Contents

1. Stack Overview
2. **Implementation Mechanism**
   - CE12800 and CE12800S
   - CE5800, CE6800 and CE7800
3. Applications
4. Data Forwarding and Failover
5. Basic Configurations

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Software and Hardware Requirements

Device Model	Number of member switches	Software version	Ports That Support Stack
CE5810EI	9	V100R002C00 and later versions	10GE optical ports
CE5850EI	16	V100R001C00 and later versions	10GE optical ports
			40GE optical ports
CE5850HI	9	V100R003C00 and later versions	10GE optical ports
			40GE optical ports
CE6810EI	16	V100R003C00 and later versions	10GE optical ports
			40GE optical ports
CE6810LI	16	V100R003C10 and later versions	10GE optical ports
			40GE optical ports
CE6850EI	16	V100R001C00 and later versions	10GE optical ports
			10GE electrical ports
			40GE optical ports
			NOTE:
			10GE electrical ports can be used for stack connection in V100R002 and later versions.
CE6850HI	16	V100R005C00 and later versions	10GE optical ports
			40GE optical ports
CE7850EI	16	V100R003C00 and later versions	40GE optical ports

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Among the four 10GE optical ports of a CE5810-48T4S-EI, 10GE1/0/1 and 10GE1/0/2 must be added to the same stack port, and 10GE1/0/3 and 10GE1/0/4 must be added to another stack port. There is no such restriction on other models.
- NOTE:
- Switches of different series cannot set up a stack. For example, CE6850EI and CE5850EI switches cannot set up a stack; CE5850EI and CE5850HI switches cannot set up a stack.
- A stack port can contain a maximum of 16 physical member ports. The number of physical member ports varies according to the device model. Devices of different models have different numbers of physical member ports.
- 10GE and 40GE ports cannot be added to the same stack port.
- 10GE optical ports derived from 40GE optical ports on devices except CE7850EI can be used for stack connection. If 40GE optical ports have been configured as the stack physical member ports, they cannot be split.
- On the CE6850EI, CE6810EI, CE6810EI and CE5850EI, if a 10GE port with port number in the range 4N+1 to 4N+4 (N = 0, 1, 2, 3...) is configured as physical member port, all the ports with port numbers 4N+1 to 4N+4 are configured as physical member ports. There is no such restriction on other models.

Huawei Training & Certification

Requirements for Stack Cables

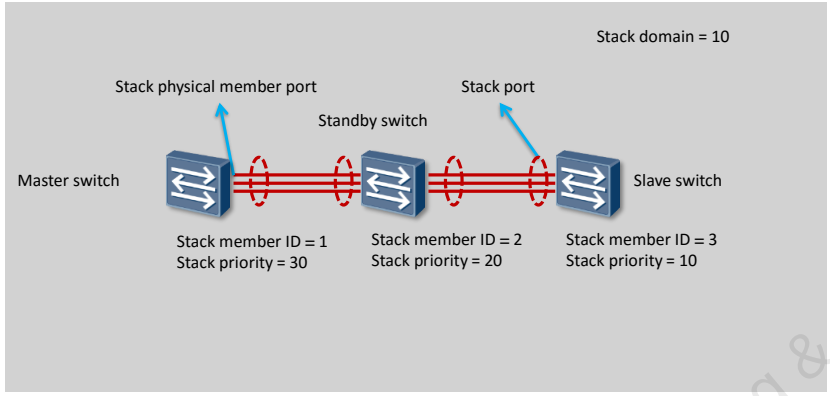
Port Type	High-Speed Cable	Optical Module	Network Cable
10GE optical port	SFP+ to SFP+ high-speed cable	10GE SFP/SFP+ optical module	-
	SFP+ to SFP+ high-speed cables are available in lengths of 1 m, 3 m, 5 m, 7 m, and 10 m, among which 7 m and 10 m SFP+ to SFP+ high-speed cables are active cables.	The required optical fibers are determined by the optical modules you select.	
10GE electrical port	-	-	10GBASE-T network cable
40GE optical port	QSFP+ to QSFP+ high-speed cable	40GE QSFP optical module	-
	QSFP+ to QSFP+ high-speed cables are available in lengths of 1 m, 3 m, and 5 m.	The required optical fibers are determined by the optical modules you select.	

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- NOTE:
- 10GE optical ports can be used for stack connection only when they have 10GE optical modules installed. If the GE optical or copper modules are installed on 10GE optical ports, the ports cannot be used to set up a stack.
- The maximum single-hop transmission of SFP/SFP+ optical modules is up to 80 km. The maximum single-hop transmission of QSFP optical modules is up to 10 km.

Huawei Training & Certification

IStack Basic Concepts



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

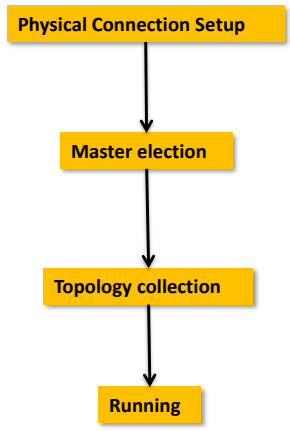
HUAWEI

Huawei Training & Certification

- Roles
- Switches that have joined a stack are member switches. Each member switch in a stack plays one of the following roles:
  - Master switch. The master switch manages the entire stack. A stack has only one master switch.
  - Standby switch. The standby switch is a backup of the master switch. When the master switch fails, the standby switch takes over all services from the master switch. A stack has only one standby switch.
  - Slave switch. A slave switch forwards service traffic. The more slave switches in a stack, the higher forwarding performance the stack can provide. Apart from the master and standby switches, all the other switches in a stack are slave switches.
- Stack domain
  - After switches are connected using stack links and set up a stack, they form a stack domain. Multiple stacks can be deployed on a network to support various applications. These stacks are identified by their domain IDs.
- Stack member ID
  - Stack member IDs are used to identify and manage member

Stack Setup Progress

- A stack is set up after the following stages:



When multiple switches are connected in a specific topology according to network requirements, a stack network is established.

Member switches exchange stack competition packets and elect a master switch according to master election rules.

The master switch collects information about all the member switches and calculates the topology. If some member switches have the same stack member ID, the master switch reassigns stack member IDs to the member switches.

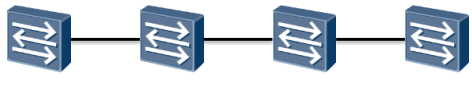
The master switch synchronizes the topology of the entire stack to all the member switches and selects a standby switch.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

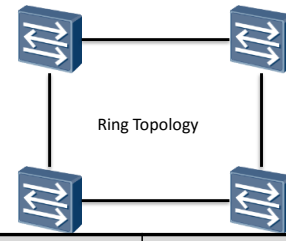
HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Physical Connection Setup



Chain Topology



Ring Topology

Topology	Advantage	Disadvantage	Usage Scenario
Chain topology	Supports long-distance stacking because the first and last member switches do not need to be connected by a physical link.	Low reliability: If any stack link fails, the stack splits. Low stack link bandwidth efficiency: The entire stack relies on a single path.	A chain topology is recommended when member devices are far from one another and a ring topology is difficult to deploy.
Ring topology	High reliability: If a stack link fails, the topology changes from ring to chain, and the stack can still function normally. High link bandwidth efficiency: Data can be forwarded along the shortest path.	The first and last member switches need to be connected by a physical link, so this topology is not appropriate for long-distance stacking.	A ring topology is recommended when member switches are located near one another, because this topology has higher reliability and link utilization.

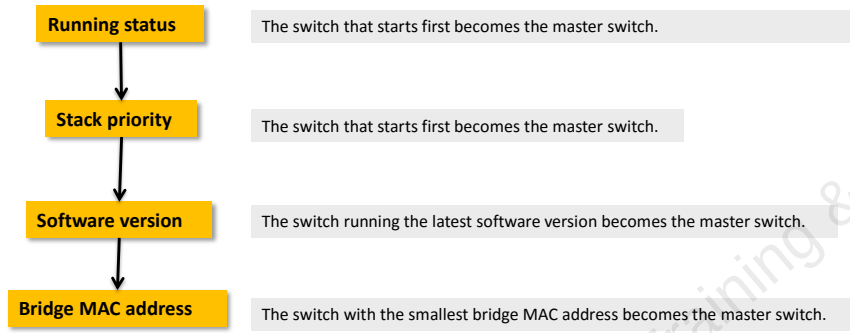
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Role Election

- After a stack is set up, member switches exchange stack competition packets to elect a master switch. The member switches compare the following items in the listed order to elect the master switch (the election ends when a winning switch is found):



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- During the delivery of a device, 16 MAC addresses are allocated to the device, among which the smallest MAC address becomes the bridge MAC address.

Huawei Training & Certification

Topology Collection and Stable Running

- Topology Collection
- After a master switch is elected, it collects information about all the member switches and calculates the topology. If some member switches have the same stack member ID, the master switch reassigns stack member IDs to the member switches.
- Stable Running
- After the master switch completes topology calculation, it synchronizes the topology of the entire stack to all the member switches and selects a standby switch. The master switch compares the following items of member switches in the listed order to select the standby switch (the election ends when a winning switch is found): Stack priority:
  - The switch with the highest stack priority becomes the standby switch.
  - MAC address: The switch with the smallest MAC address becomes the standby switch.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

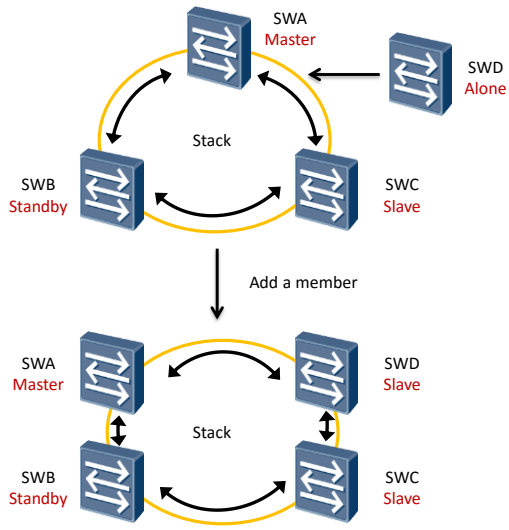
- Software Version Synchronization
- A stack supports software version synchronization among the member switches. The member switches do not have to run the same software version, and they can set up a stack as long as their software versions are compatible with each another. If software version running on a member switch is different from that on the master switch, the member switch downloads the system software from the master switch, restarts with the new system software, and rejoins the stack.
- Configuration File Synchronization
- A stack uses a strict configuration file synchronization mechanism to ensure that the member switches work like one device.
- When a stack is set up, each member switch starts with its own configuration file. After switches start, the standby and slave switches combine their stack configurations into the configuration file of the master switch to form the configuration file of the stack system.
- When the stack is running normally, the master switch manages the entire stack, and synchronizes configurations made by users to the other switches in real time to maintain configuration consistency on all the member switches.
- The configuration file synchronization mechanism ensures that the member switches save the same configuration file. If the master switch fails, other member switches can provide

Huawei Training & Certification
Huawei Training & Certification

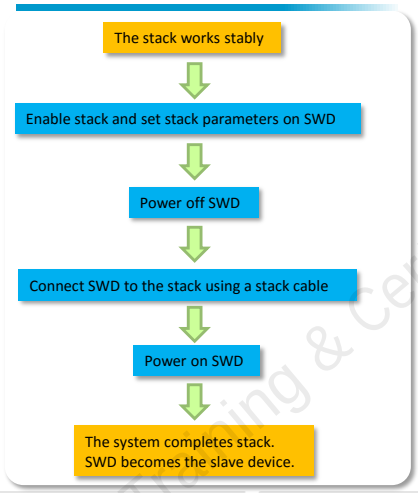
services using the same configuration file.

Huawei Training & Certification
Huawei Training & Certification

Joining a Stack



Process of joining a stack

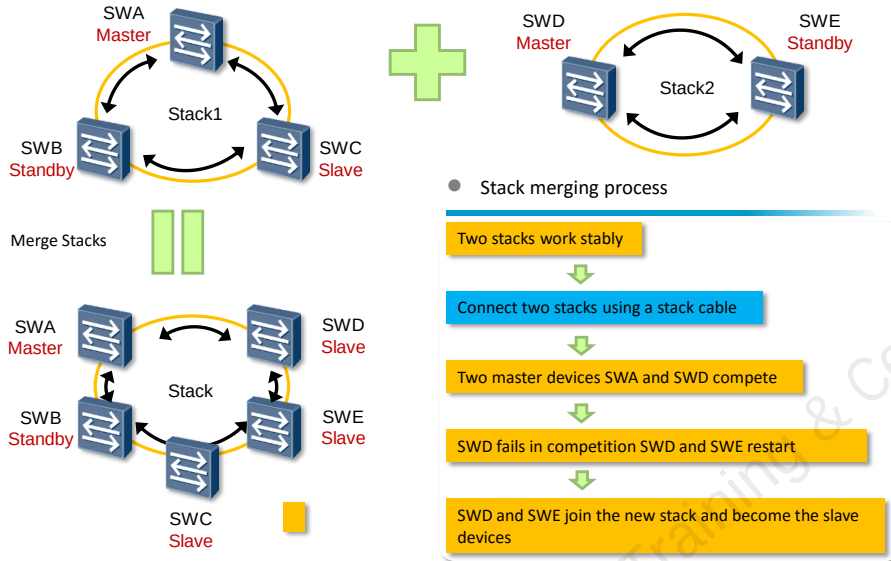


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The domain ID of SWD must be the same as the domain ID of the stack.
- It is recommended that you set the member ID for SWD before you configure the stack. If you set the member ID during stack configuration, SWD needs to start twice, which is time-consuming.

Huawei Training & Certification

Stack Merging



- Stack merging process

Two stacks work stably

↓

Connect two stacks using a stack cable

↓

Two master devices SWA and SWD compete

↓

SWD fails in competition SWD and SWE restart

↓

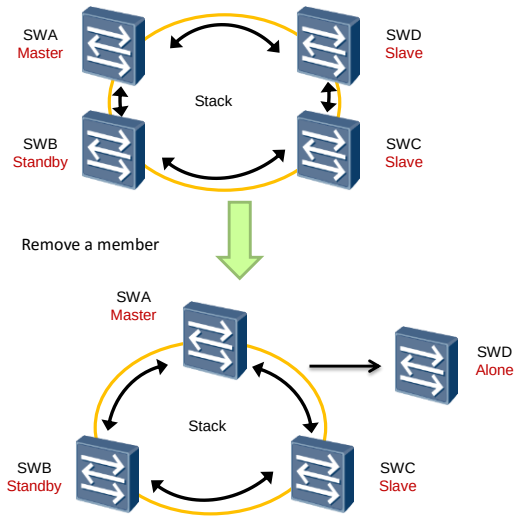
SWD and SWE join the new stack and become the slave devices

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Leaving a Stack



- Conditions triggering a device to leave a stack

① Removing the stack cable
② Shutting down the stack port or physical member interface
③ Restarting the stack member
④ Member device fault

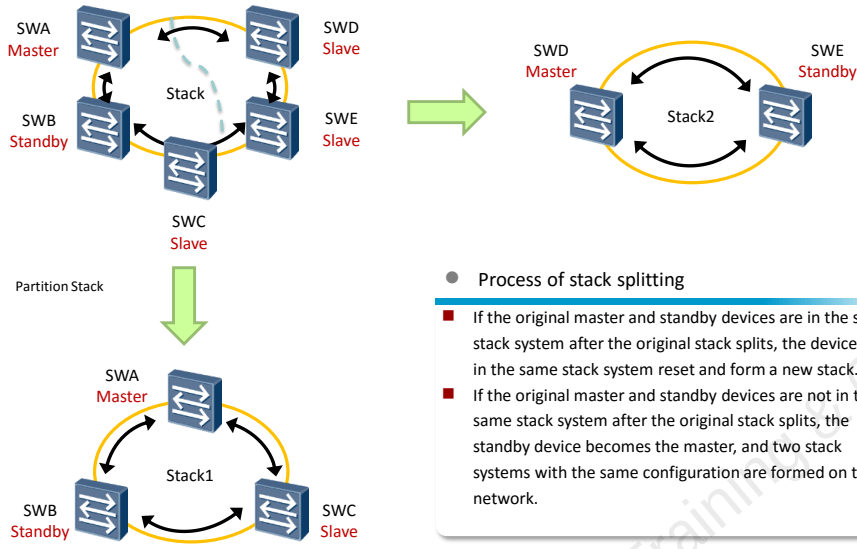
- Process of leaving a stack

- If the master device leaves the stack, the standby device becomes the master. The stack system updates the topology and continues to work.
- If the standby device leaves the stack, a new standby device is selected. The stack system updates the topology and continues to work.
- If a slave device leaves the stack, the stack system updates the topology and continues to work.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Stack Splitting



- Process of stack splitting
  - If the original master and standby devices are in the same stack system after the original stack splits, the devices not in the same stack system reset and form a new stack.
  - If the original master and standby devices are not in the same stack system after the original stack splits, the standby device becomes the master, and two stack systems with the same configuration are formed on the network.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Stack Upgrade

Upgrade Method	Description and Application Scenarios
Traditional upgrade method	You need to specify next-startup files and restart the entire stack. This method causes service interruption in a long time and is therefore not applicable to scenarios requiring short service interruption time.
Fast upgrade	This upgrade method provides a mechanism to minimize the service interruption time during software upgrade of stack member switches, reducing impact of the upgrade on services.
ISSU upgrade	The ISSU function completes an upgrade through a card switchover. This upgrade method ensures higher reliability and shorter service interruption time.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- NOTE:
  - Only the stack containing two member switches supports fast upgrade.
  - If 40GE high-speed cables are used to connect stack member devices, you cannot fast upgrade switches from versions earlier than V100R003C00 to later versions.
  - It is recommended that the upstream and downstream devices be connected to the stack through Eth-Trunk links to reduce the traffic interruption time during an upgrade.
- When the stack system been fast upgrading, First, the standby switch restarts with the new system software. Data traffic is forwarded by the master switch in this period. After the standby switch is upgraded, it becomes the master switch and starts to forward data traffic. Then the original master switch restarts with the new system software. After the upgrade is complete, the original master switch becomes the backup switch in the stack.
- If the upgrade fails due to a stack link failure, the system software rolls back to the original version.
- For more information about ISSU upgrade, see ISSU Configuration in the CloudEngine 7800&6800&5800 Series Switches Configuration Guide - Basic Configurations.

Huawei Training & Certification

Contents

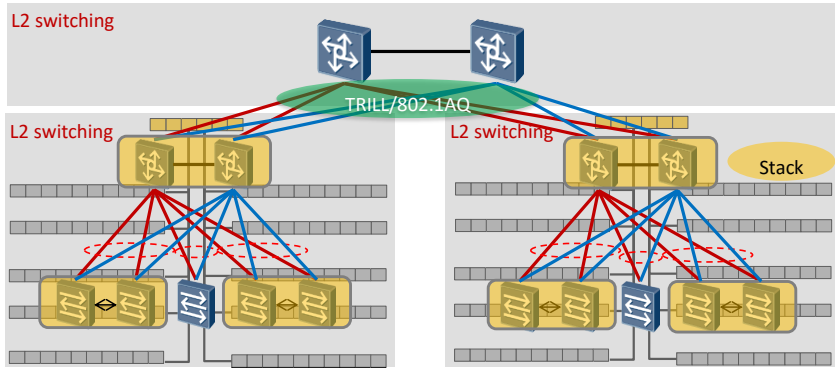
1. Stack Overview
2. Implementation Mechanism
3. **Applications**
4. Data Forwarding and Failover
5. Basic Configurations

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Typical Application



Intra-area networking solution

- Based on aggregation switches and TORs
- Simply network structure and easy maintenance

Inter-area networking solution

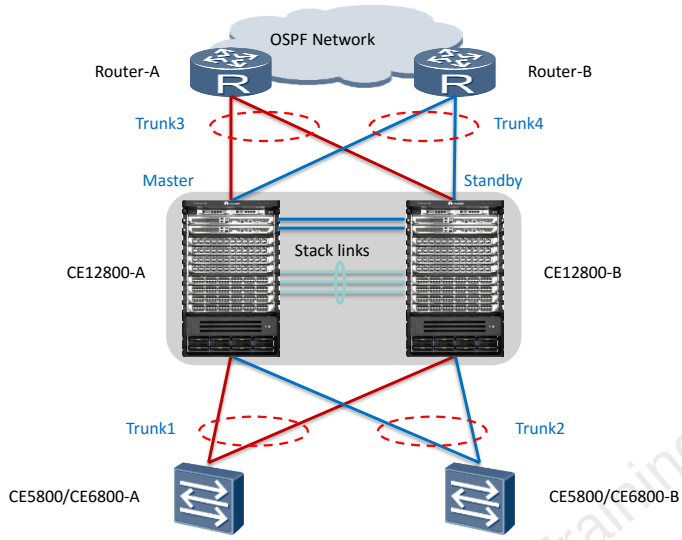
- TRILL/802.1aq based on aggregation switches and core switches
- Good extensibility

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Back-to-Back Stack

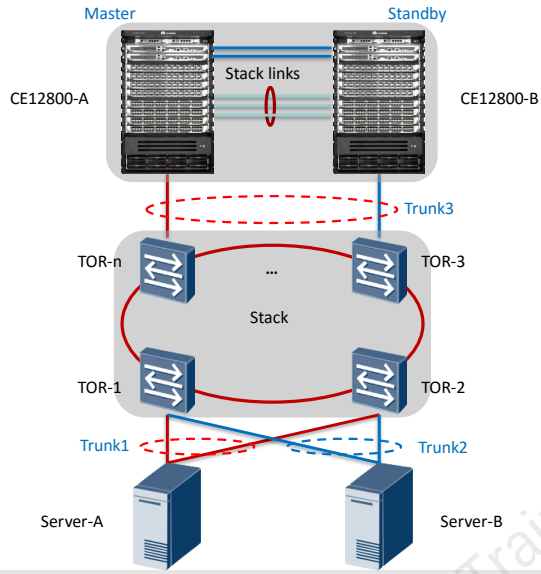


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Multiple physical switches form a logical switch through the stack technology. The stack system has only one control plane. (All physical switches use the same system MAC address, IP address, and global configuration.)
- Inter-chassis trunk: The traffic received by downstream ports on CE5800/CE6800-A can be sent out by any port in trunk 1.
- If the trunk hash algorithm selects the upstream physical port of the local switch (CE12800-A), traffic is forwarded through CE12800-A. If the trunk hash algorithm selects the upstream physical port of CE12800-B, traffic is forwarded to CE12800-B through the stack link, and then sent out through the upstream port on CE12800-B.
- With preferred local forwarding enabled, traffic received from CE5800/CE6800-A is sent out through the upstream physical port of CE12800-A.

Huawei Training & Certification

Ring Stack

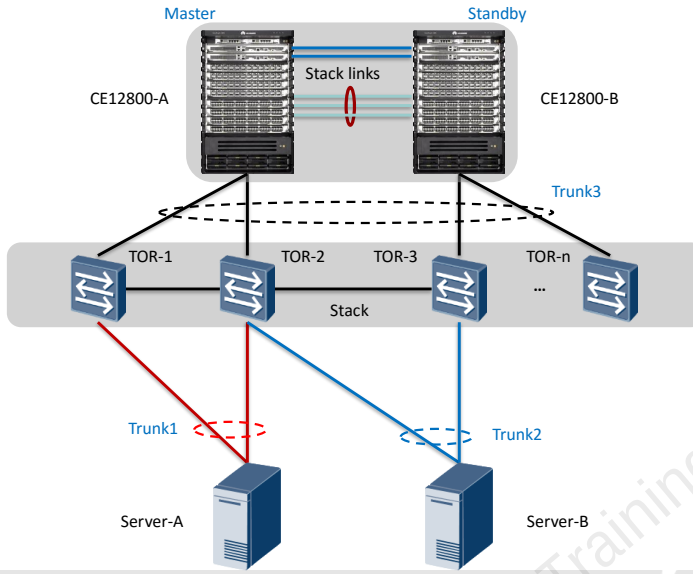


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Trunk 3 can consist of single links or mesh connections.
- Compared with single links, mesh connections provide higher bandwidth and reliability. Mesh connections are applicable to the scenario where link resources are sufficient and high forwarding performance is required.
- Single links reduce the number of required links without degrading reliability. It is applicable to the scenario where link resources and cables are limited.

Huawei Training & Certification

Chain Stack

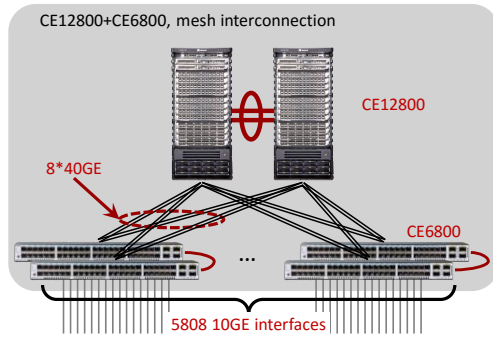


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Trunk 3 can consist of single links or mesh connections.
- Compared with single links, mesh connections provide higher bandwidth and reliability. Mesh connections are applicable to the scenario where link resources are sufficient and high forwarding performance is required.
- Single links reduce the number of required links without degrading reliability. It is applicable to the scenario where link resources and cables are limited.

Huawei Training & Certification

CSS Deployment Example



High reliability: Mesh networking

Large forwarding capability: 41 Tbit/s bidirectional forwarding

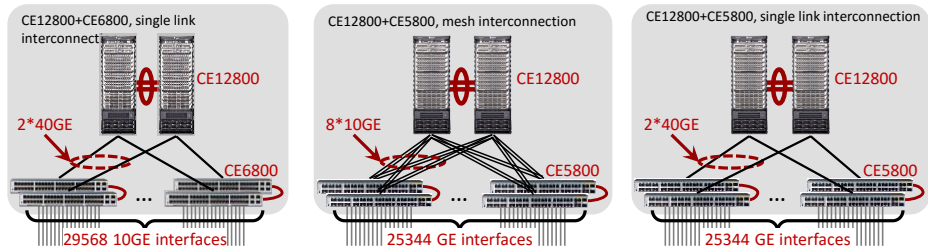
High-density interfaces:5808 x 10GE interfaces

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The figure shows an example of deploying CSS on CE12800 switches. Two CE12812 switches form a CSS. The CE6800 stack is connected to the CE12800 stack through 40GE ports. CE12800 and CE6800 switches are mesh connected and set up a trunk through the inter-chassis links. The entire network provides up to 5808 downstream 10GE ports.
- Number of TOR switches (calculated in terms of physical TOR)
  - 24 (40GE interface/slot) x 11 (slot/switch) x 2 (switch)/4 (interface/TOR) = 132 TORs
- Access capability
  - Number of 10GE interfaces: 44 (10GE/TOR) x 132 (TOR) = 5808 interfaces
- Oversubscription ratio:
  - (4 x 40GE) : (44 x 10GE)= 1 : 2.75

Huawei Training & Certification

CSS Deployment Example



Characteristics

- Single link interconnection: Conserves interfaces and simplifies cable layout.
- 40GE interconnection: Improves forwarding capability.
- Number of TORs: 528
- High-density access: 29568 x 10GE interfaces
- Oversubscription ratio: 1:14

Characteristics

- Mesh interconnection: Improves reliability and forwarding capability.
- Number of TORs: 528
- High-density access: 25344 x GE interfaces
- Oversubscription ratio: 1:1.2

Characteristics

- Single link interconnection: Conserves interfaces and simplifies cable layout.
- 40GE interconnection: Improves forwarding capability.
- Number of TORs: 528
- High-density access: 25344 x GE interfaces
- Oversubscription ratio: 1:1.2

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- As shown in the figure, a CE12800 CSS connect to CE5800 or CE6800 switches to form a network providing high-density GE or 10GE ports. Two CE12812 switches form a CSS, which connects to a CE5800 or CE6800 stack through 10GE ports. The entire network provides up to 25344 GE ports. Different combinations of upstream links (single links or mesh connections) can be selected according to service requirements.
- Because CE12800 switches have high forwarding performance, the networking capability of CSS formed by CE12800s is greatly improved, which allows deployment on networks of different scales.

Huawei Training & Certification

Contents

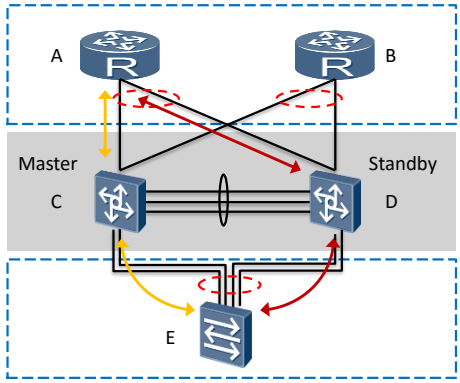
1. Stack Overview
2. Implementation Mechanism
3. Applications
4. **Data Forwarding and Failover**
5. Basic Configurations

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Unicast Traffic Forwarding Path



Unicast traffic forwarding path in a stack

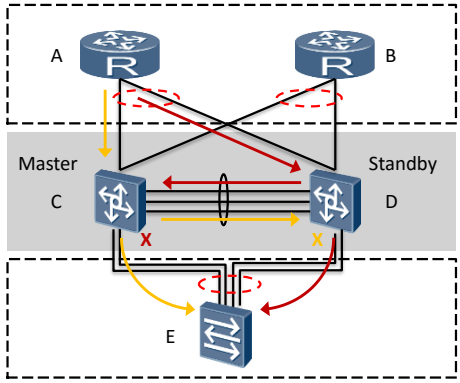
- Traffic forwarding path from A to E
  - A forwards traffic to member switches C and D using the hash algorithm through trunk.
  - C forwards traffic to E. The traffic does not pass stack links.
  - The processing method on D is the same as that on C.
- Traffic forwarding path from E to A
  - E forwards traffic to member switches C and D using the hash algorithm through trunk.
  - C forwards traffic to A. The traffic does not pass stack links.
  - The processing method on D is the same as that on C.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Multicast Traffic Forwarding Path



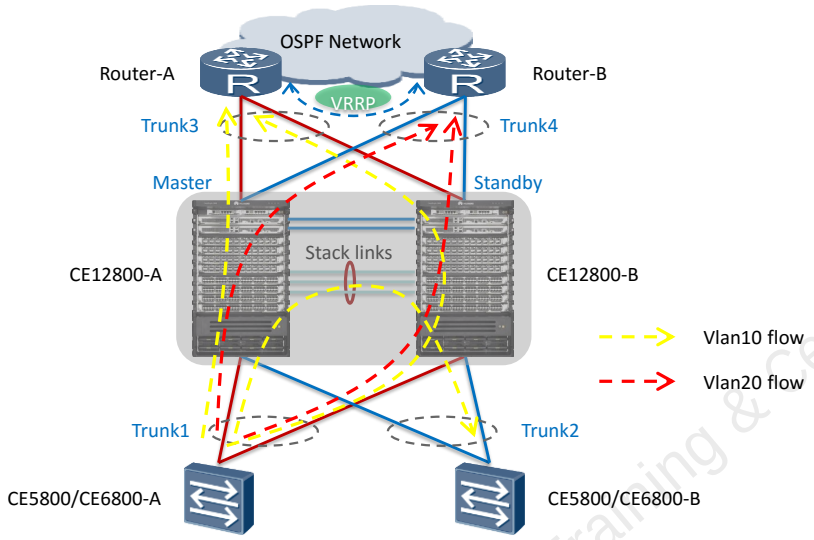
Multicast traffic forwarding path in a stack

- Multicast traffic forwarding path from A to E
  - A forwards multicast flows 1 and 2 to member switches C and D using hash algorithm
  - C copies flow 1 to E and to D through stack links. The multicast flows from D to E and from C to E are pruned on the outbound interfaces of the trunk; therefore, only one flow reaches E.
  - The processing of multicast flow 2 on D is the same as that of flow 1.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Stack Network and Data Forwarding (1)

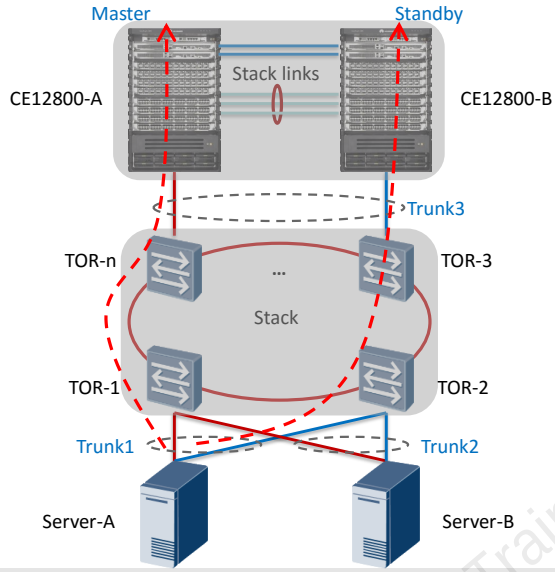


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- As shown in the figure, traffic sent by CE5800/CE6800-A is from VLAN 10 and VLAN 20. Router A and Router B run VRRP. The VRRP master of VLAN 10 is deployed on Router A, and the VRRP master of VLAN 20 is deployed on Router B.
- Traffic from CE5800/CE6800-A is sent to the two upstream links in Trunk 1 through hash algorithm. The CE12800 stack uses preferred local forwarding. Traffic from VLAN 10 is forwarded along the path in yellow line, and traffic from VLAN 20 is forwarded along the path in red line.
- Traffic forwarding process on CE5800/CE6800-B is similar to that on CE5800/CE6800-A.

Huawei Training & Certification

Stack Network and Data Forwarding (2)

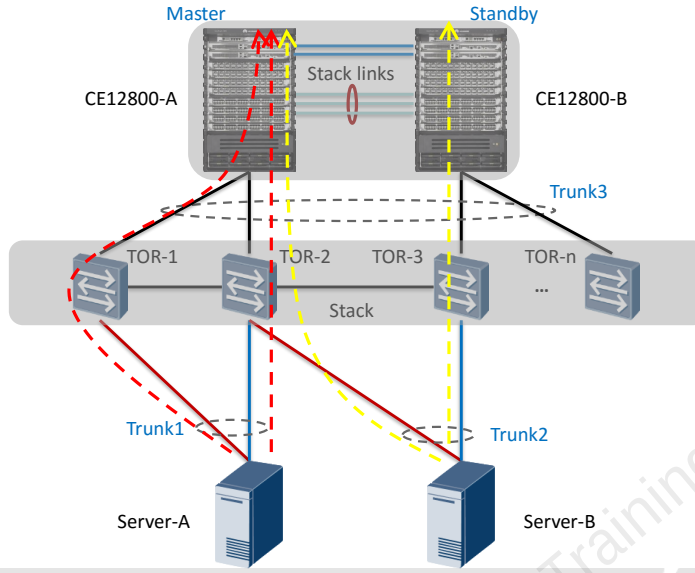


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- In the TOR stack, traffic from servers is sent to the physical links in the trunk through the hash algorithm, and the traffic is forwarded according to the preferred local forwarding principle.

Huawei Training & Certification

Stack Network and Data Forwarding (3)

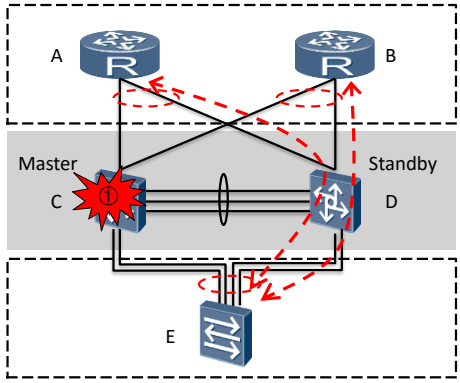


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- In the TOR stack, traffic from servers is sent to the physical links in the trunk through the hash algorithm, and the traffic is forwarded according to the preferred local forwarding principle.

Huawei Training & Certification

Stack Fault Scenario (1)



- Master device is faulty

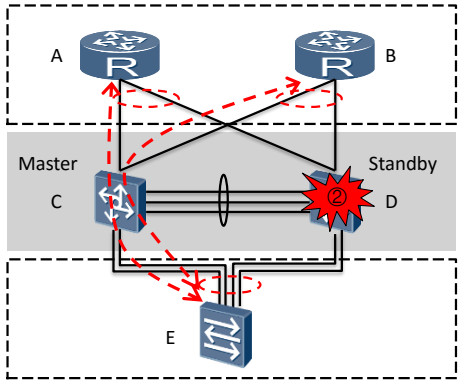
- Standby device D perceives the fault of master device C, so D upgrades itself to master.
- The trunk interface connecting to the stack perceives the fault of the link connecting to the master device, so the trunk interface updates the member interface table. The traffic them does not pass the trunk connecting to the master device.
- Traffic passing D is not affected.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Stack Fault Scenario (2)



- Standby device is faulty

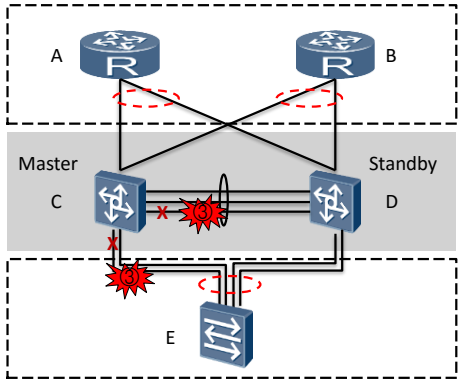
- The trunk interface connecting to the stack perceives the fault of the link connecting to the master device, so the trunk interface updates the member interface table. The traffic then does not pass the trunk connecting to the master device.
- Traffic passing D is not affected.
- An active/standby switchover is not required.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Stack Fault Scenario (3)

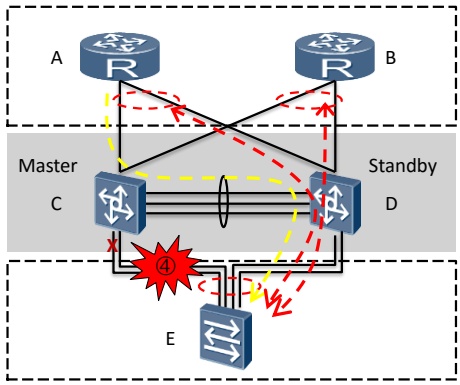


- Some member links in a Trunk link are faulty
  - When a link down failure is detected using the port down or stack link detection packet, the stack removes the faulty link from the trunk according to the trunk protocol.
  - The packets are forwarded by the faulty link, and the traffic that does not pass stack links is not affected.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Stack Fault Scenario (4)



- All member links in the downstream trunk are faulty

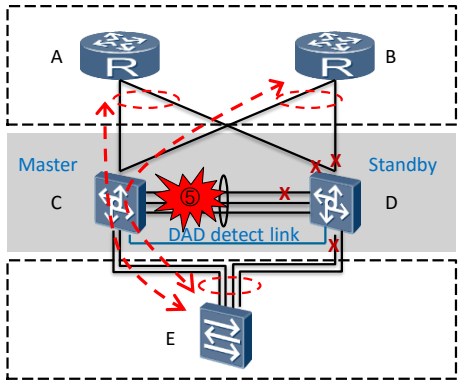
- When a trunk link down failure is detected using the port down or stack link detection packet, the stack does not forward traffic through the faulty links.
- Master device C receives traffic from the upstream link and forwards traffic to standby device D through stack links. D then forwards traffic to E.
- E detects the faults of member links, and updates the member interface table. The traffic then does not pass the links connected to C.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Stack Fault Scenario (5)



- All links between stack members are faulty

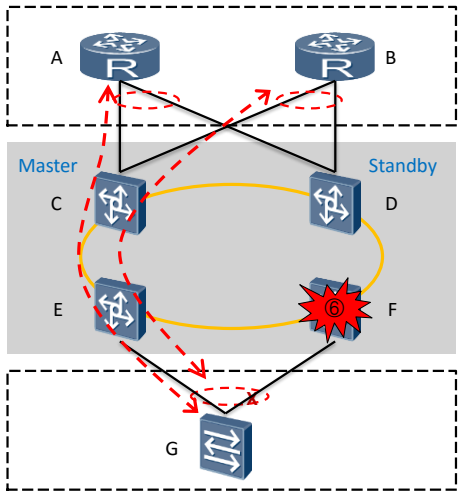
- The standby device becomes the master, so two master devices exist. The master devices shut down all ports.
- After the fault is rectified, all traffic from the access device is forwarded by the new master device.
- The new standby device shuts down all ports except the reserved port and dual-active detection ports, and enters the Recovery state to stop forwarding service packets.
- After the stack links are recovered, the device in Recovery state restarts, and restores the closed service ports to Up. The stack system is recovered.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Stack Fault Scenario (6)



- Member switch connected by a single link or a link is faulty

- If the slave device is faulty, the stack updates the topology, deletes the faulty device and path, and continues to work.
- If the standby device is faulty, a slave device is elected as the new standby device, and the stack updates the topology and continues to work.
- If the active device is faulty, the standby device upgrades to the master device, the stack elects a new standby device, updates the topology, and continues to work.
- If a stack link is faulty, the device roles are not changed, and the stack updates the topology, deletes the faulty link, and continues to work.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- In the preceding scenario, if the software is suspended, the device does not reset, and the stack links need to be detected using heartbeat packets, the packet discarding duration depends on the heartbeat timeout period. The heartbeat timeout period between stack members is 10 seconds.

Huawei Training & Certification

Contents

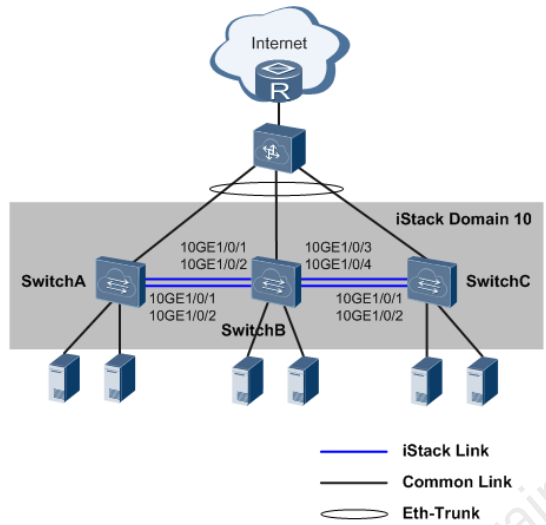
1. Stack Overview
2. Implementation Mechanism
3. Applications
4. Data Forwarding and Failover
5. **Basic Configurations**

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Example for Establishing a Stack in a Chain Topology



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- As shown in the figure, SwitchA, SwitchB, and SwitchC need to set up a stack in a chain topology. For stack connections, add 10GE1/0/1 and 10GE1/0/2 of SwitchA to stack port 1/1, 10GE1/0/1 and 10GE1/0/2 of SwitchB to stack port 1/1, 10GE1/0/3 and 10GE1/0/4 of SwitchB to stack port 1/2, and 10GE1/0/1 and 10GE1/0/2 of SwitchC to stack port 1/1.
- The configuration roadmap is as follows: To enable the switches to set up a stack and distinguish the stack from other stacks on the network, configure stack parameters for the switches, including the stack member ID, priority, and stack domain ID.
- To enable data packets to be forwarded between stack member switches, configure stack ports on the switches. Multiple physical member ports can be added to a stack port to improve stack link bandwidth and reliability.
- To make the configuration effective and ensure that the stack can be set up successfully, save the stack configuration, restart the switches, connect the switches using stack cables.

Huawei Training & Certification

Configure Stack Parameters

- On SwitchA, set the stack member ID to 1, stack priority to 150, and stack domain ID to 10. By default, the stack member ID of a switch is 1. In this example, SwitchA retains the default stack member ID 1, and you do not configure this parameter.

```
[~SwitchA]stack
[~SwitchA-stack] stack member 1 priority 150
[*SwitchA-stack] stack member 1 domain 10
```

- On SwitchB, set the stack member ID to 2 and stack domain ID to 10.

```
[~SwitchB]stack
[~SwitchB-stack] stack member 1 renumber 2 inherit-config
Warning: The stack configuration of member ID 1 will be inherited to member ID 2
after the device resets. Continue? [Y/N]: y
[*SwitchB-stack] stack member 1 domain 10
```

- On SwitchC, set the stack member ID to 3 and stack domain ID to 10.

```
[~SwitchC]stack
[~SwitchC-stack] stack member 1 renumber 3 inherit-config
[*SwitchC-stack] stack member 1 domain 10
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- NOTE:
- When you run the stack renumber new-member-id [inherit-config] command to change the stack member ID: If you specify inherit-config, the switch inherits the original stack configuration after it restarts, including the stack priority, stack domain ID, and physical member port configuration. The common service port configuration will be lost. For example, the port split configuration will be lost, but the split configuration of physical member ports is inherited.
- If you do not specify inherit-config, the current stack configuration and common service port configuration are lost after the switch restarts, and the configuration related to the new stack member ID in the configuration file takes effect.
- If the configuration file contains offline configuration for the new stack member ID, do not specify inherit-config in the command.

Huawei Training & Certification

Configure Stack Ports

- On SwitchA, add 10GE1/0/1 and 10GE1/0/2 to stack port 1/1.

```
[~SwitchA] interface stack-port 1/1
[*SwitchA-Stack-Port1/1] port member-group interface 10ge 1/0/1 to 1/0/2
Warning: The interface(s) (10GE1/0/1-1/0/4) will be converted to stack mode. [Y/N]: y
[*SwitchA-Stack-Port1/1] quit
```

- On SwitchB, add 10GE1/0/1 and 10GE1/0/2 to stack port 1/1, and add 10GE1/0/3 and 10GE1/0/4 to stack port 1/2.

```
[~SwitchB] interface stack-port 1/1
[*SwitchB-Stack-Port1/1] port member-group interface 10ge 1/0/1 to 1/0/2
Warning: The interface(s) (10GE1/0/1-1/0/4) will be converted to stack mode. [Y/N]: y
[*SwitchB-Stack-Port1/1] quit
[*SwitchB] interface stack-port 1/2
[*SwitchB-Stack-Port1/2] port member-group interface 10ge 1/0/3 to 1/0/4
[*SwitchB-Stack-Port1/2] quit
```

- On SwitchC, add 10GE1/0/1 and 10GE1/0/2 to stack port 1/1.

```
[~SwitchC] interface stack-port 1/1
[*SwitchC-Stack-Port1/1] port member-group interface 10ge 1/0/1 to 1/0/2
Warning: The interface(s) (10GE1/0/1-1/0/4) will be converted to stack mode. [Y/N]: y
[*SwitchC-Stack-Port1/1] quit
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Check the Stack Configuration

- Run the display stack configuration command on the switches to check whether the configuration is the same as expected.

```
<SwitchA>display stack configuration
Oper : Operation
Conf : Configuration
*    : Offline configuration
Isolated Port: The port is in stack mode, but does not belong to any Stack-Port
Attribute Configuration:
----------------------------------------
 MemberID      Domain         Priority
Oper(Conf)   Oper(Conf)      Oper(Conf)
----------------------------------------
1(1)         --(10)          100(150)
----------------------------------------
Stack-Port Configuration:
--------------------------------------------------------------------------------
Stack-Port           Member Ports
--------------------------------------------------------------------------------
Stack-Port1/1        10GE1/0/1     10GE1/0/2
--------------------------------------------------------------------------------
Isolated Port Configuration:
10GE1/0/3      10GE1/0/4
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Save the configurations of SwitchA, SwitchB, and SwitchC, and restart the switches.
- Connect the switches using stack cables to set up a stack.

Huawei Training & Certification

Verify the Configuration

- Check information about the stack member switches.

```
<SwitchA> display stack
--------------------------------------------------------------------------
MemberID Role     MAC              Priority   DeviceType
Description
--------------------------------------------------------------------------
1        Master   0004-9f31-d520   150        CE6850-48S4Q-EI
2        Standby  0004-9f62-1f40   100        CE6850-48S4Q-EI
3        Slave    0004-9f69-a391   100        CE6850-48S4Q-EI
--------------------------------------------------------------------------
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Verify the Configuration

- Display stack topology information.

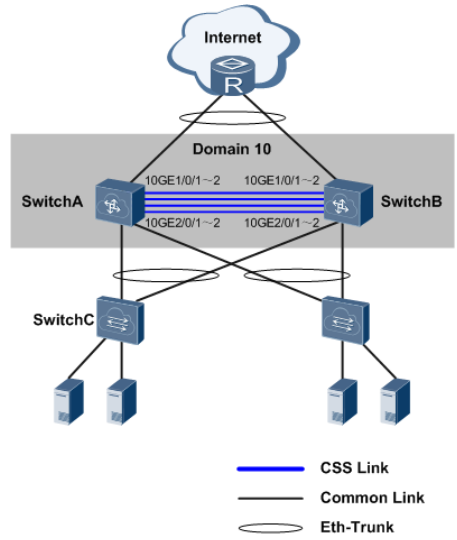
```
<SwitchA>display stack topology
Stack Topology:
-------------------------------------------
            Stack-Port 1      Stack-Port 2
MemberID   Status Neighbor   Status Neighbor
-------------------------------------------
1          up     2          --     --
2          up     1          up     3
3          up     2          --     --
-------------------------------------------
Stack Link:
-------------------------------------------------------------------------
Stack-Port       Port           Status     PeerPort       PeerStatus
-------------------------------------------------------------------------
Stack-Port1/1    10GE1/0/1      up         10GE2/0/1      up
Stack-Port1/1    10GE1/0/2      up         10GE2/0/2      up
Stack-Port2/1    10GE2/0/1      up         10GE1/0/1      up
Stack-Port2/1    10GE2/0/2      up         10GE1/0/2      up
Stack-Port2/2    10GE2/0/3      up         10GE3/0/1      up
Stack-Port2/2    10GE2/0/4      up         10GE3/0/2      up
Stack-Port3/1    10GE3/0/1      up         10GE2/0/3      up
Stack-Port3/1    10GE3/0/2      up         10GE2/0/4      up
-------------------------------------------------------------------------
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- After a stack is set up successfully, you are advised to run the save command immediately to save the configuration.

Huawei Training & Certification

Example for Configuring a Stack on CE12800/12800S



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- An enterprise builds a data center. The data center network must provide high reliability on the core layer and have a simple structure to facilitate configuration and management.
- To meet requirements of the enterprise, core switches SwitchA and SwitchB set up a stack. Ports 10GE1/0/1 to 10GE1/0/2 and 10GE2/0/1 to 10GE2/0/2 of the two switches need to be connected using stack cables and added to stack ports.
- The configuration roadmap is as follows:
  - Connect SwitchA and SwitchB using stack cables. Connect four service ports on two LPUs of each switch to improve bandwidth and reliability.
  - Configure the same stack domain ID and stack connection mode on SwitchA and SwitchB, and configure different stack member IDs and stack priorities for the two switches.
  - Create a stack port on SwitchA and SwitchB, and add the physical member ports to the stack port.
  - Save the configurations of SwitchA and SwitchB and enable the stacking function on the two switches.
  - Check whether a stack is set up successfully.

Huawei Training & Certification

Connect Stack Cables and Configure Stack Parameters

- On SwitchA, set the stack member ID to 1, stack priority to 150, stack domain ID to 10, and stack connection mode to MPU connection.

```
[~SwitchA]stack
[~SwitchA-stack] stack member 1
[~SwitchA-stack] stack priority 150
[*SwitchA-stack] stack domain 10
[*SwitchA-stack] stack link-type mainboard-direct
[*SwitchA-stack] quit
```

- On SwitchB, set the stack member ID to 2, stack priority to 100, stack domain ID to 10, and stack connection mode to MPU connection.

```
[~SwitchB]stack
[~SwitchB-stack] stack member 2
Warning: The device will use the configuration of member ID 2 after the
device resets. Continue? [Y/N]: y
[*SwitchB-stack] stack priority 100
[*SwitchB-stack] stack domain 10
[*SwitchB-stack] stack link-type mainboard-direct
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Connect SwitchA and SwitchB using stack cables. To use the MPU connection mode, connect SIP ports on the MPUs and physical member ports on the LPUs. To use the LPU connection mode, connect physical member ports on the LPUs.
- To use the LPU connection mode, run the stack link-type linecard-direct command.

Huawei Training & Certification

Configure Stack Ports

- Create a stack port on SwitchA, and add physical ports 10GE1/0/1 to 10GE1/0/2 and 10GE2/0/1 to 10GE2/0/2 to the stack port.

```
[~SwitchA]port-group group1
[*SwitchA-port-group-group1] group-member 10ge 1/0/1 to 10ge 1/0/2
[*SwitchA-port-group-group1] group-member 10ge 2/0/1 to 10ge 2/0/2
[*SwitchA-port-group-group1] shutdown
[*SwitchA-port-group-group1] quit
[*SwitchA]commit
[~SwitchA] interface stack-port 1
[*SwitchA-Stack-Port1] port member-group interface 10ge 1/0/1 to 1/0/2
[*SwitchA-Stack-Port1] port member-group interface 10ge 2/0/1 to 2/0/2
[*SwitchA-Stack-Port1] quit
[*SwitchA]commit
[~SwitchA]port-group group1
[~SwitchA-port-group-group1] undo shutdown
[*SwitchA-port-group-group1] quit
[*SwitchA]commit
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Create a stack port on SwitchB, and add physical ports 10GE1/0/1 to 10GE1/0/2 and 10GE2/0/1 to 10GE2/0/2 to the stack port. The configuration on SwitchB is similar to the configuration on SwitchA, and is not mentioned here.

Huawei Training & Certification

Check the Stack Configuration

- Run the display stack configuration command to check whether the stack configuration is the same as expected.

```
<SwitchA>display stack configuration
Oper : Operation
Conf : Configuration
*    : Offline configuration
Isolated Port: The port is in stack mode, but does not belong to any Stack-Port

Attribute Configuration:
-------------------------------------------------------------
 MemberID      Domain         Priority       Mode     Enable
Oper(Conf)   Oper(Conf)      Oper(Conf)   Oper(Conf)   Oper
-------------------------------------------------------------
1(1)          --(10)          100(150)     MB(MB)     Disable
-------------------------------------------------------------

Stack-Port Configuration:
-------------------------------------------------------------------------------
Stack-Port           Member Ports
-------------------------------------------------------------------------------
Stack-Port1          10GE1/0/1     10GE1/0/2     10GE2/0/1     10GE2/0/2
-------------------------------------------------------------------------------
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The command output on SwitchA is used as an example.

Huawei Training & Certification

Enable the Stacking Function

- Save the configuration of SwitchA and enable the stacking function.

```
<SwitchA>save
Warning: The current configuration will be written to the device.
Continue? [Y/N]: y
<SwitchA>system-view
[~SwitchA]stack
[~SwitchA-stack]stack enable
Warning: Make sure that one or more dual-active detection methods are
configured once the conversion is complete and the device enters the
stack mode.
Current configuration will be converted to the next startup saved-
configuration file of stack mode.
System will reboot. Continue? [Y/N]: y
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Save the configuration of SwitchB and enable the stacking function. Similar operation as on SwitchA.

Huawei Training & Certification
Huawei Training & Certification

Check Whether the Stack Is Setup Successfully

- View indicators on the switches. On SwitchA, the STACK indicator on one MPU is steady green. On SwitchB, the STACK indicators on both MPUs blink green. This indicates that a stack has been set up, in which SwitchA is the master switch, and SwitchB is the standby switch.
- # Log in to the stack through the console interface or management interface of any switch and run the display stack command to check whether the stack is set up successfully. To log in through the management interface, specify the IP address of the master switch.

```
<SwitchA>display stack
--------------------------------------------------------------------------
MemberID Role     MAC             Priority   DeviceType  Description
--------------------------------------------------------------------------
1        Master   006d-8835-2b00   150        CE12804
2        Standby  006d-8835-2c00   100        CE12804
--------------------------------------------------------------------------
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- After a stack is set up successfully, you are advised to run the save command immediately to save the configuration.

Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

M-LAG Feature of CE Series Switches

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Objectives

- Upon completion of this course, you will be able to:
  - Understand the principle of M-LAG.
  - Know application scenarios of M-LAG.
  - Know how to configure M-LAG on Huawei CE series switches.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. **M-LAG Background**
2. M-LAG Principle
3. M-LAG Application
4. M-LAG Configuration
5. Enhancement in the Future

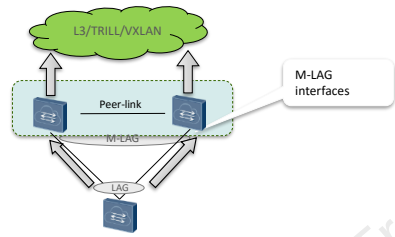
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Disadvantages of Traditional Technologies

- Multi-Chassis Link Aggregation (M-LAG) is mainly used in dual-homing scenarios where the use-side device is dual-homed to two access devices. Traditional technologies including stacking and Smart Link have the following disadvantages:
  - Stacking: Two (or more) stack member switches share the same control plane, so a single-point failure may affect the entire system, especially during an upgrade of the stack system. This technology cannot provide high reliability on the aggregation/core layer or other scenarios requiring high reliability.
  - Smart Link: Links work in active-standby mode. That is, only one link forwards data of a VLAN at a time, so the link utilization is low.
- M-LAG addresses the preceding mentioned reliability and link utilization issues.



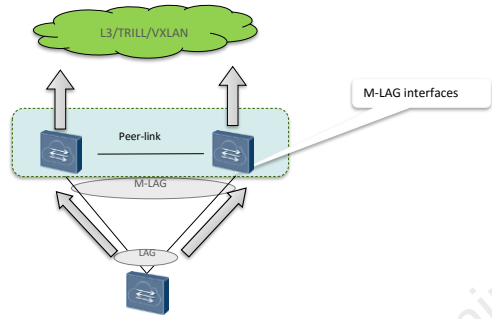
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Principle

- Multi-Chassis Link Aggregation Group (M-LAG) enables ports of different devices to be bundled into an LAG, so as to improve link bandwidth and reliability.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Forwarding plane: As an access switch connects to two aggregation/core switches through an LAG, traffic can be distributed to the two upstream switches on a per flow basis.
- Control plane: The two switches that set up an M-LAG are independent of each other and use the M-LAG protocol to communicate. A failure of one switch will not affect the other.

Huawei Training & Certification

Contents

1. M-LAG Background
2. **M-LAG Principle**
3. M-LAG Application
4. M-LAG Configuration
5. Enhancement in the Future

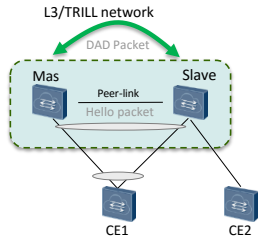
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

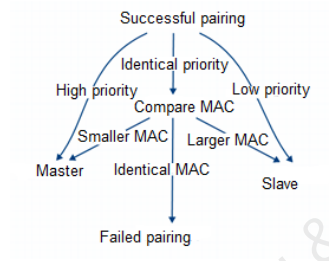
HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Principle: Device Pairing

- When the configurations on devices at both ends of the M-LAG are complete, devices at both ends perform paring.



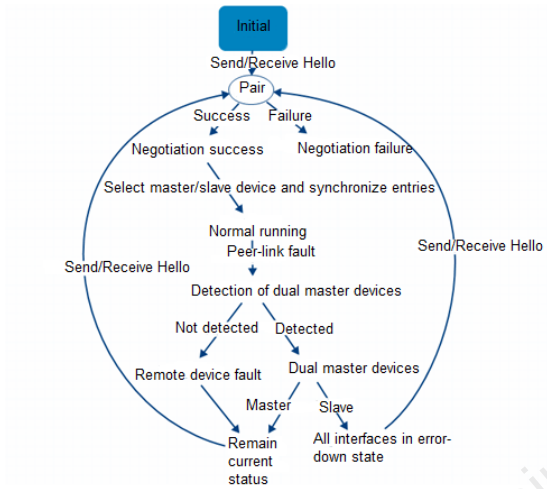


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Device at both ends of the M-LAG exchange Hello packets on the peer-link every 1s. Hello packets carry the DFS group ID, protocol version number, and system MAC address of the local end.
- After receiving Hello packets from the remote end, the local end determines whether DFS group IDs are the same. Pairing succeeds if the IDs are the same.
- The master and slave devices are selected. The device with a higher priority is the master device. If the priorities are the same, the device with a smaller MAC address is the master device. The master/backup negotiation does not affect the forwarding behavior in the M-LAG and is used when faults occur.
- Both devices synchronize information, including the system name, system MAC address, software version, M-LAG status, STP BPDU, as well as MAC, ARP, and IGMP entries.
- Devices exchange heartbeat packets over an external network (non-peer-link, L3 or TRILL network). Heartbeat packets are used for detecting dual master devices when the peer-link fails.

Huawei Training & Certification

Principle: Status Switching During Device Pairing



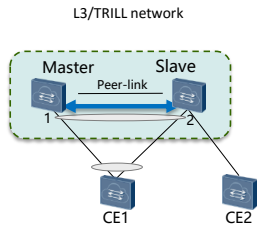
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

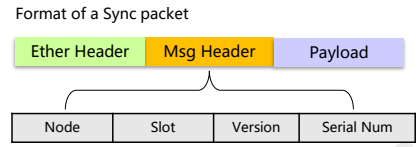
- After the device configuration is complete, they enter the initial state and send Hello packets.
- If the local device receives Hello packets from the remote device and DFS group IDs are the same, pairing is successful.
- If DFS group IDs are different, negotiation fails.
- After negotiation is successful, the local device synchronizes information including the priority and system MAC address from the remote device to determine master and slave devices.
- After master and slave devices are selected, entries and configurations are synchronized. After synchronization, M-LAG can work.
- If the peer-link fails, M-LAG immediately detects whether there are dual master devices.
- If M-LAG detects dual master devices, M-LAG considers that there are dual master devices. All interfaces except the peer-link interface and management interface are configured to enter the error-down state.
- If M-LAG does not detect dual master devices, M-LAG considers that the remote device fails and the status of the local device remains unchanged.

Huawei Training & Certification

Principle: Data Synchronization

- Entries on devices at both ends of an M-LAG must be the same, so entries must be synchronized. If entries on M-LAG member interfaces are different, traffic forwarding may fail.





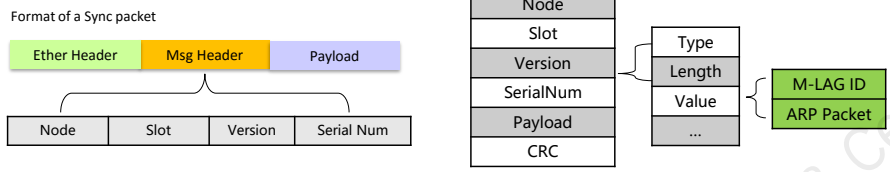
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

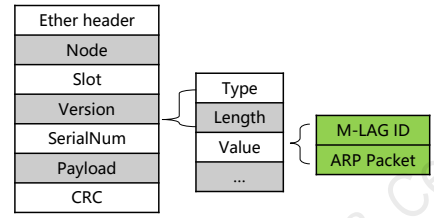
- The following configurations and entries on the two devices must be the same:
  - MAC address entries
  - ARP entries
  - L2 and L3 multicast forwarding entries
  - DHCP snooping binding entries
  - LACP system ID
  - STP configurations in the system and on interfaces
  - Other information including the global VLAN and ACL
- Entries including ARP entries, MAC address entries, and IGMP entries need to be synchronized.
- The preceding figure shows the format of a Sync packet. Fields in a Sync packet:
  - Ether header: Ethernet header
  - Msg header: self-defined message header
  - Node: node ID of the local end
  - Slot: slot ID of the receiver

Huawei Training & Certification

Principle: Example of an Entry Format

- ARP synchronization packets are used as an example. The packets use the TLV format, which can be easily extended.





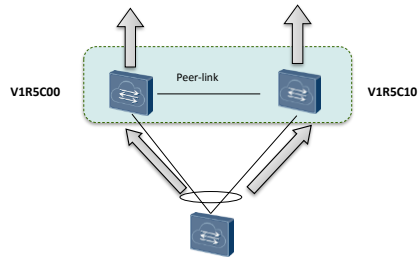
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The TLV contains the M-LAG ID and original ARP packets. Devices of any version process original ARP packets in the same manner, so version compatibility is easy to implement.
- Entries of protocols such as IGMP in addition to ARP are synchronized based on original packets, and then processed on the remote end.

Huawei Training & Certification

Principle: M-LAG Protocol Compatibility

- Generally, an M-LAG upgrade is performed on member switches one by one to minimize impact of the upgrade on the system. Therefore, the M-LAG member switches will run different protocol versions during the upgrade. Backward compatibility of the M-LAG protocol can ensure a successful upgrade and minimize packet loss.



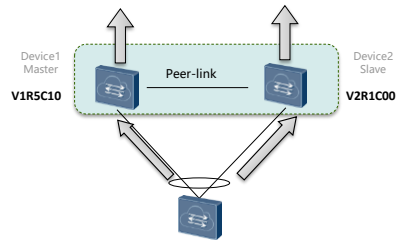
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The M-LAG protocol version remains unchanged if there are no changes in M-LAG functions. If M-LAG functions (including MAC, ARP, and IGMP entry synchronization) have changed, the M-LAG protocol version also changes.
- M-LAG member switches synchronize original information with each other. For example, to synchronize ARP and IGMP entries, a member switch sends original packets (with a little additional information) to the peer, and the peer learns the entries from these packets. This synchronization model is less dependent on the protocol version and ensures version compatibility.
- An M-LAG member device obtains the protocol version of the remote device for its software modules.
- If the software modules find that their functions have changed compared with the obtained protocol version, they still process protocol packets of the old version and send packets of the old version to the remote device.

Huawei Training & Certification

Principle: Example of M-LAG Protocol Compatibility

- In the following figure, the switch is upgraded from V1R5C10 to V2R1C00. in V2R1C00, the LACP system ID of the master device is synchronized to the slave device, and the slave device uses the synchronized LACP system ID to replace its LACP system ID.



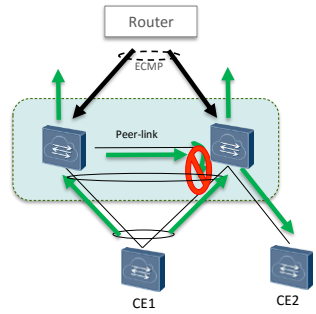
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Procedure
  - The two M-LAG switches use V1R5C10, and the same LACP system ID is manually configured on the two switches.
  - Device2 start to be upgraded. Device2 restarts after the system software package is changed. The dual-homing networking changes into a single-homing networking, and Device1 is responsible for packet forwarding.
  - After Device2 completes the upgrade, it starts with V2R1C00 and sends Hello packets for pairing. Assume that Hello packets of V2R1C00 adds one TLV. After Device1 receives packets from Device2, Device1 can only process the TLVs that can be identified by itself. After Device2 receives packets from Device1, Device2 detects that Device1 uses an earlier version. Device2 does not check the new TLV. The two devices can perform negotiation (the new TLV does not take effect).
  - After the negotiation is complete, Device2 starts to send packets for synchronizing the LACP system ID. Device1 runs an earlier version and cannot process the new type of packets. Device1 directly ignores the packets. Device2 does not receive packets for synchronizing the LACP system ID from Device1. The manually configured LACP system ID is still used and the LACP status remains unchanged.
  - Device1 starts to be upgraded in the same manner as Device2.
  - After Device1 completes the upgrade, the two devices use the same version. The manually configured LACP system ID is deleted, and the automatically negotiated LACP system ID is used.

Huawei Training & Certification

Principle: Traffic Model

- Traffic from a CE is distributed to two M-LAG member switches on a flow basis.



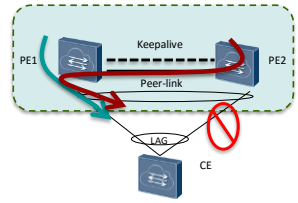
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- M-LAG member switches forward uplink and downlink traffic through local outbound interfaces preferentially.
- Broadcast traffic passes through the peer-link. The peer-link is unidirectionally isolated from the M-LAG member interfaces connected to a dual-homed CE device. (Traffic entering from the peer-link cannot be sent back to the corresponding M-LAG member interface.) The peer-link is not isolated from M-LAG member interfaces connected to single-homed CE devices or interfaces not in the M-LAG.

Huawei Training & Certification

Fault Scenario: M-LAG Member Interface Fails

- When an M-LAG member interface fails, ensure that packets on two PEs can be received by the other interface.



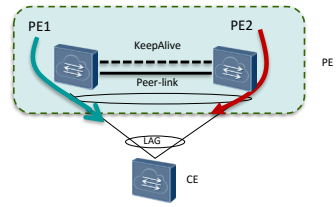
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Isolation between the peer-link and the M-LAG member interface on PE-1 needs to be disabled so that service traffic can be transmitted over the peer-link.
- PE2 deletes the MAC address entry of the CE and synchronizes the MAC address entry learned on the M-LAG member interface of PE1 to its peer-link interface.

Huawei Training & Certification

Fault Scenario: M-LAG Member Interface Recovers

- When an M-LAG member interface fails, ensure that packets on two PEs can be received by the other interface.



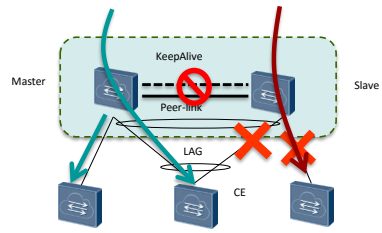
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- When the M-LAG member interface recovers, isolation between the peer-link and the M-LAG member interface on PE1 needs to be enabled again to prevent duplicate packets.
- PE2 deletes the MAC address entries on the original peer-link interface and synchronizes MAC address entries learned on the M-LAG member interface of PE1 to its M-LAG member interface.
- PE2 synchronizes the ARP entries learned on the peer-link interface to its M-LAG member interface that has recovered. This is implemented using the MAC-triggered APR entry update mechanism, which enables a device to update the outbound interface in an ARP entry after the outbound interface in the matching MAC address entry is updated.

Huawei Training & Certification

Fault Scenario: Peer-link Fails

- When the peer-link fails, M-LAG becomes ineffective. The LAG interface of the CE receives only one copy of packets.



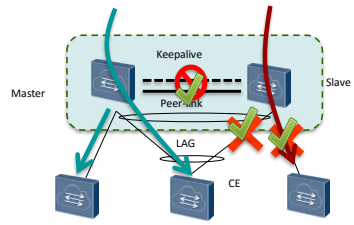
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- M-LAG member switches have determined the master and backup roles based on their priorities.
- After the peer-link fails, the slave interface sets all its physical interfaces to Down state to ensure that all traffic is forwarded by the master device.
- Note:
  - If some interfaces need to be configured in Down state, configure unpair-port to specify interfaces in Down state.

Huawei Training & Certification

Fault Scenario: Peer-link Recovers

- After the faulty peer-link recovers, M-LAG member switches need to be restored to minimize packet loss.



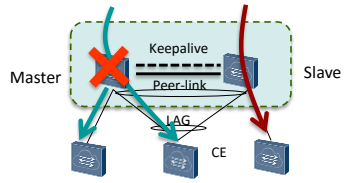
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The two switches start a negotiation again. After the negotiation is complete, the slave device sets all interfaces out of the M-LAG to Up state.
- The M-LAG member interface becomes Up 2 minutes after the negotiation succeeds.
- In an SVF system of modular and fixed switches, M-LAG member interfaces on leaf switches will become Up 5 minutes after the negotiation succeeds.

Huawei Training & Certification

Fault Scenario: Device Is Powered Off

- The master device is powered off and all interfaces go Down.
- After the slave device detects the peer-link fault, it detects whether the remote device is online through the network-side interface or heartbeat channel.
- The slave device becomes the master device and continues to forward packets.



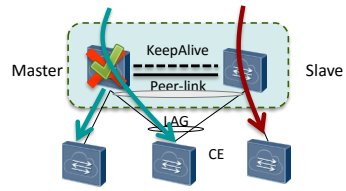
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Fault Scenario: Device Recovers

- After the master device that has been powered off recovers, the peer-link recovers. The master and slave devices perform re-negotiation. After the negotiation is complete, they re-select the master and slave devices.
- The delay for the access interface on the master device to go Up needs to be configured. After M-LAG negotiation is complete, the physical interface is still Down.
- Considering that network protocols such as routing protocols and TRILL require a period of time for negotiation and entry delivery, the interface is configured to go Up after a delay.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. M-LAG Background
2. M-LAG Principle
3. **M-LAG Application**
4. M-LAG Configuration
5. Enhancement in the Future

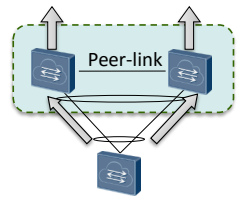
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Scenario 1: Server Access Through an M-LAG

- Uplinks of a server set up a LAG, through which the server connects to the PEs that set up an M-LAG.
- The LAG on the server can run the Link Aggregation Control Protocol (LACP).
- Some servers need to support dynamic LACP.
- Remarks: To implement load sharing, the server must have two NICs working in active-active mode.



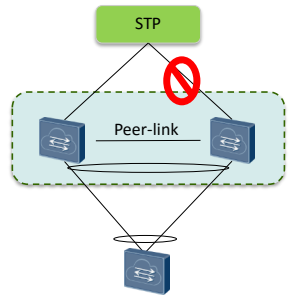
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- If LACP negotiation fails, LAG member interfaces need to be unbound. Otherwise, the server cannot obtain the configuration through DHCP. (The server has no configuration when it connects to the network.)

Huawei Training & Certification

Scenario: STP and M-LAG Combination

- The two M-LAG member switches need to connect to the STP network and STP can correctly eliminate loops.



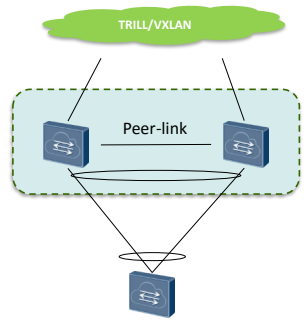
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- An M-LAG can connect to a traditional STP network.
- The two M-LAG member switches are virtualized into one STP node to eliminate loops on the STP network.
- In V100R005C00 and earlier versions, the two M-LAG member switches must be virtualized into the root bridge. This restriction does not apply to versions later than V100R005C10.
- Even if the M-LAG member switches do not connect to an STP network, STP is still enabled on the switches to prevent loops caused by misconfigurations.

Huawei Training & Certification

Scenario: TRILL/VXLAN and M-LAG Combination

- The M-LAG network-side interfaces can connect to an overlay network (such as a TRILL or VXLAN network).



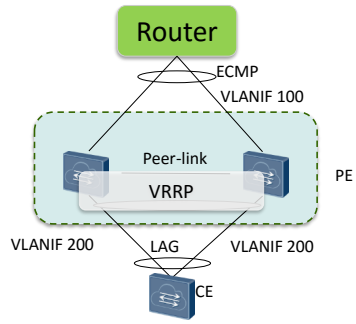
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- An M-LAG can connect to a TRILL network. In this scenario, the two M-LAG member switches are virtualized into a logical TRILL node and identified by a virtual nickname.
- An M-LAG can connect to a VXLAN network. In this scenario, the two M-LAG member switches are virtualized into a logical VXLAN node and identified by a virtual tunnel end point (VTEP) IP address.
- Regardless of whether the M-LAG connects to a TRILL or VXLAN network, the M-LAG acts like one logical device for user-side and network-side devices. The M-LAG member switches back up each other and implement load sharing.

Huawei Training & Certification

Scenario: M-LAG on the L3 Network

- For the downstream CE, the two active PEs are independent, and their L3 interfaces (VLANIF and main interfaces) have the same IP and MAC addresses.



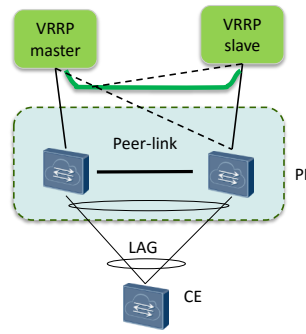
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Mode 1: Configure VRRP on VLANIF interfaces of the PEs to ensure that the VLANIF interfaces of the two PEs have the same IP and MAC addresses. Both the VRRP members are in active state.
- Mode 2: Configure the same MAC address on the VLANIF or main interfaces on the PEs to ensure that they have the same IP and MAC addresses (planned in V1R6).
- In a scenario where a single VLANIF interface is used, when the fault is rectified, ARP pre-synchronization can be used to accelerate convergence (V1R6 supports this function and the delay for the VLANIF interface to go Up needs to be set).

Huawei Training & Certification

Scenario: VRRP and M-LAG Combination

- The M-LAG provides L2 access and connects to the VRRP gateway or firewall.



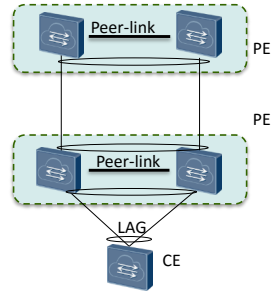
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- A VRRP gateway can connect to M-LAG member switches in single-homing or dual-homing mode.
- The M-LAG provides L2 access. Each VRRP member device is connected to a single M-LAG member switch. VRRP packets are transparently transmitted over the peer-link.
- The M-LAG provides L2 access.

Huawei Training & Certification

Scenario: M-LAG Cascading

- M-LAG cascading can be deployed to build a large L2 network in a DCN or set L2 interconnection between DCs.



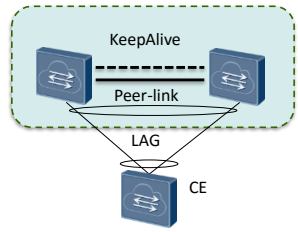
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- M-LAG cascading can be deployed to build a large L2 network in a DCN or set L2 interconnection between DCs.
- The network must be kept loop-free in 2-level M-LAG cascading configuration or various failure scenarios.

Huawei Training & Certification

Reliability Scenario: DAD

- DAD is provided to improve M-LAG reliability.

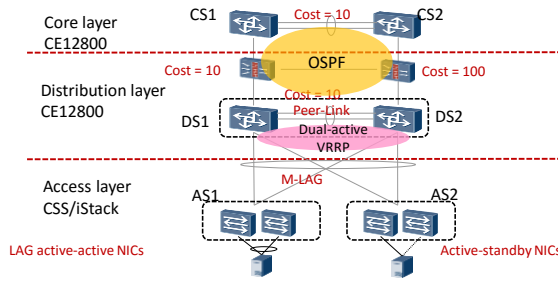


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Both switches send heartbeat packets over the heartbeat channel or external network to detect the device status.
- The switch sends DAD packets over the heartbeat channel or external network to check for a dual-active condition.
- The remote communication address can be detected automatically.
- A member switch can detect the peer communication address (IPv4 address, IPv6 address, IPV4 VPN, IPv6 VPN, or nickname) through heartbeat packets.

Huawei Training & Certification

Integrated M-LAG&CSS/iStack Scenario



- Precautions for network deployment:
  - To prevent loops caused by incorrect cable connections, MSTP or RSTP needs to be deployed in each POD, with the root bridge located on the distribution switches.
  - DAD needs to be configured on the M-LAG member switches, using the out-of-band management network or an inbound IP network as the heartbeat channel.
  - The RSTP mode needs to be configured.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Core layer
  - Double independent OSPF-enabled core switches
  - The OSPF cost configuration ensures that southbound and northbound traffic is preferentially forwarded through the path between DS1 and CS1.
  - The core and distribution switches are connected in a square topology to ensure that traffic passing through a firewall is forwarded through the same path in forward and reverse directions. This topology is cost-effective and easy to maintain.
  - The switches at the core and distribution layers (or firewalls) are interconnected through inter-card LAGs (not M-LAG).
- Distribution layer
  - The L2 transparent firewalls work in active-active mode, and their uplinks and downlinks (LAGs) are associated to ensure quick failover.
  - The distribution switches set up an M-LAG, which is connected to stack systems at the access layer to build a loop-free topology.
  - The distribution switches run VRRP and both work in active state to implement L3 gateway redundancy.
  - The M-LAG solution provides high reliability and reduces risks in an upgrade.
- Access layer
  - Access switches set up iStack or CSS systems.
  - The stack networking enables servers to connect to the access layer reliably in dual-homing mode, and also simplifies network management and maintenance.

Huawei Training & Certification

Contents

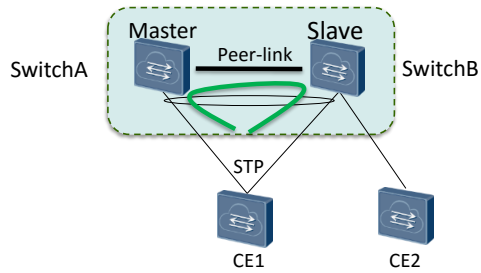
1. M-LAG Background
2. M-LAG Principle
3. M-LAG Application
4. **M-LAG Configuration**
5. Enhancement in the Future

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configuring a Virtual Root Bridge




```
[~SwitchA]stp root priority 0
[~SwitchA]stp bridge-address 1-1-1
```

```
[~SwitchB]stp root priority 0
[~SwitchB]stp bridge-address 1-1-1
```

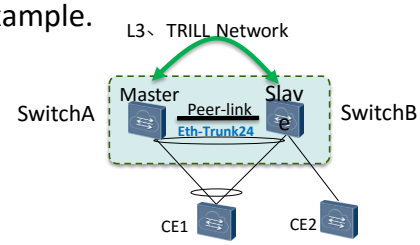
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- In V1R5C00 and earlier versions:
  - During M-LAG configuration, a temporary loop may occur.
  - M-LAG can be deployed on an STP network. M-LAG participates in STP calculation to ensure loop prevention.
  - Before performing M-LAG configuration, configure the two M-LAG switches as root bridges and configure the same priority for them. Root protection can be configured if no device has higher priority than them
  - After the configuration is completed, the two switches are virtualized into the root bridge to eliminate loops.
- In later versions of V1R5C10:
  - The switch still supports the virtual root bridge.
  - The switch provides V-STP. That is, two M-LAG switches are configured as one virtual STP node to eliminate loops. They do not need to be configured as root bridges.

Huawei Training & Certification

Creating a DFS Group

- The configurations on two M-LAG switches are similar. SwitchA is used as an example.



- Create a DFS group.

```
[~SwitchA]dfs-group 1
```

- Configure a reachable IP address or nick name in the DFS group.

```
[*SwitchA-dfs-group-1]source ip 1.1.1.1
```

- Configure a peer-link.

```
[~SwitchA-Eth-Trunk24]peer-link 1
```

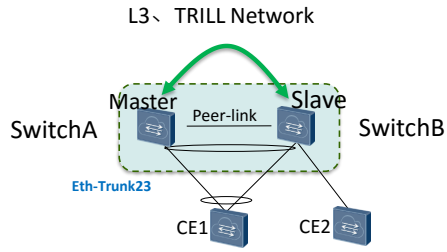
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- After the same DFS group is configured on two M-LAG switches, the two M-LAG switches send protocol packets over the peer-link for negotiation and pairing.

Huawei Training & Certification

Binding an M-LAG

- After the DFS group configuration is complete, bind the M-LAG to the LAG. The M-LAG on the two devices must be bound to the same DFS group and M-LAG ID.




```
[~SwitchA-Eth-Trunk23]dfs-group 1 m-lag 1
```

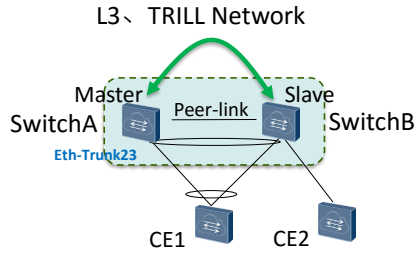
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- As long as the LAG at both ends is bound to the same DFS group and M-LAG ID, pairing can be performed on the two interfaces and an M-LAG can be established.

Huawei Training & Certification

Configuring L3 Access

- To use M-LAG for L3 access, configure VRRP on the VLANIF interface where M-LAG is configured.




```
[~SwitchA-Eth-Trunk23]dfs-group 1 m-lag 1
[~SwitchA-Eth-Trunk23]port link-type hybrid
[~SwitchA-Eth-Trunk23]port hybrid tagged vlan 100
[~SwitchA-Vlanif100]vrrp vrid 1 virtual-ip 2.2.2.2
```

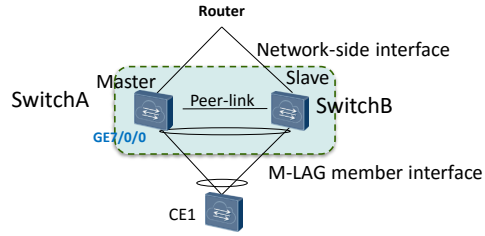
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Two M-LAG switches must be configured with the same VRID and virtual IP address. The peer-link automatically blocks VRRP packets, so VRRP works in dual-active mode. For CE1, the gateway uses the same IP address and MAC address.

Huawei Training & Certification

Configuring LACP and the Delay for an Interface to Go Up

- You are advised to configure LACP on an M-LAG to improve the M-LAG reliability and accelerate convergence.
- You are advised to configure LACP on a peer-link to enhance reliability and solve the unidirectional communication problem.
- You are advised to set the delay for the M-LAG member interface and network-side interface to go Up. For L2 traffic, the delay of about 30s is recommended.




```
[~SwitchA-GE7/0/0]set up-delay 120
```
(The command can be only used on physical interfaces, and is configured on LAG member interfaces one by one.)

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- For L3 traffic, the delay is protortional to the ARP specification. The delay increases by 1s for every 400 ARP entries. The delay on the M-LAG member interface is twice the delay on the network-side interface. For example, if there are 12K ARP entries, it is recommended the delay on the M-LAG member interface be 90s and the delay on the network-side interface be 60s:
- 12K/400 = 30s
- 30s+30s= 60s
- 30s + 30s x 2 = 90s

Huawei Training & Certification

Contents

1. M-LAG Background
2. M-LAG Principle
3. M-LAG Application
4. M-LAG Configuration
5. **Enhancement in the Future**

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Enhancement in the Future

- Two M-LAG switches must use the same configurations, so consistency check is required. If the configurations are inconsistent, measures such as alarms should be taken. The configurations includes the STP, LACP, and VRRP configurations, PV relationship, interface mode, MTU, and interface rate.
- The switch provides one-click configuration synchronization. When the configurations are performed on one device, the configurations are automatically synchronized to the other device. This function reduces the configuration workload and prevents misconfigurations.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

VS Feature of CE Series Switches

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Foreword

- Virtualized applications of servers are emerging in the cloud computing era. Virtual machines (VMs) have increased the use efficiency of physical computing resources while reducing IT system operation and maintenance (O&M) costs. In addition, VM dynamic migration enhances system reliability, flexibility, and scalability. Virtual machines can be used on many network devices that function as critical network elements.
- This course describes the necessity of virtualizing one physical device into multiple virtual devices. Specifically, this course takes the next-generation Huawei Virtual System (VS) as an example to describe the virtual machine's architecture, application scenarios, and benefits for customers.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Objectives

- Upon completion of this course, you will be able to learn about:
  - VS application advantages
  - VS technology principles
  - VS deployment modes
  - Basic VS configurations

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

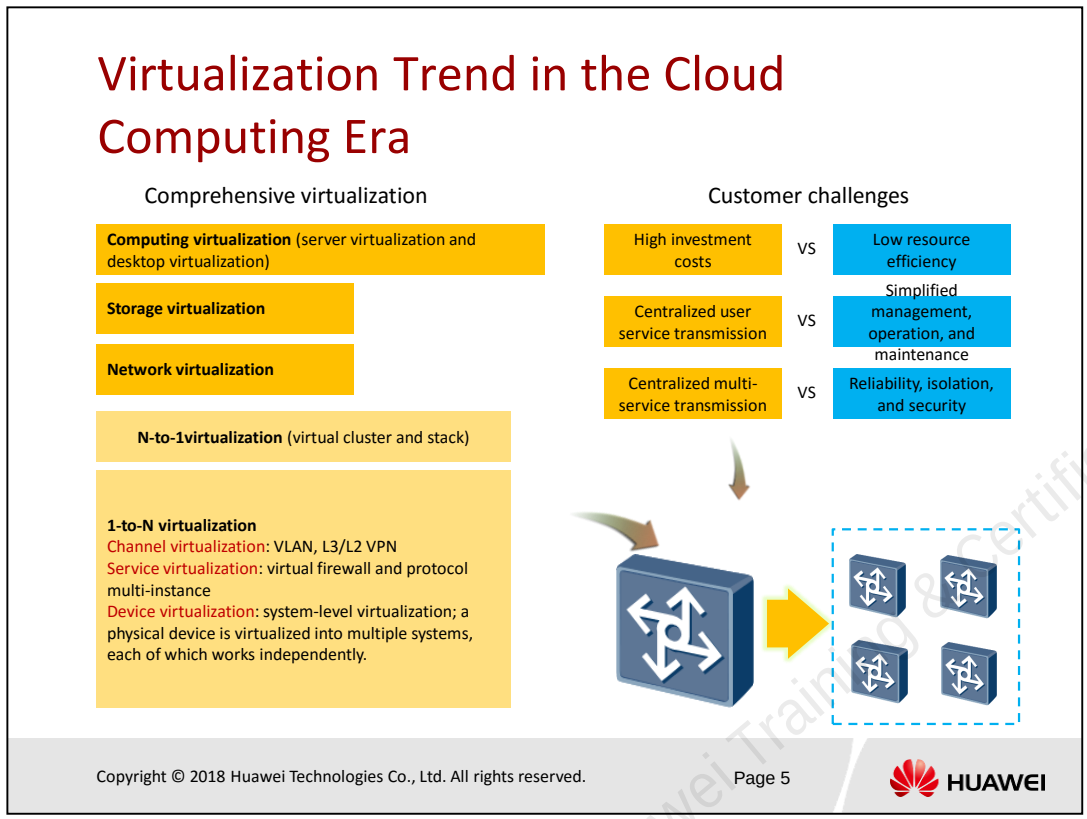
Huawei Training & Certification
Huawei Training & Certification

Contents

1. **VS Overview**
2. VS Implementation
3. VS Deployment
4. VS Typical Applications

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

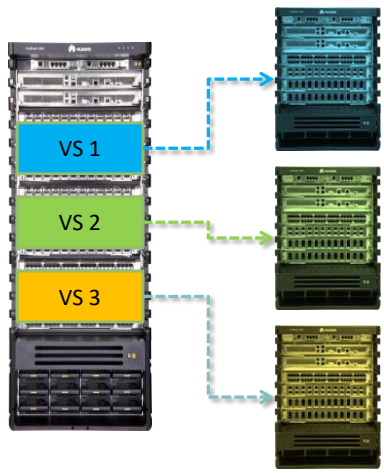
Huawei Training & Certification
Huawei Training & Certification



- Cloud computing technologies virtualize IT resources. The virtualized IT resources have become similar to other household utilities such as electricity. Users can obtain the virtualized IT resources on demand. Virtualization is one of the critical cloud computing technologies. Virtualization on different layers abstracts physical resources using the virtualization technology to share or isolate cloud resources. According to analysis from International Data Corporation (IDC), after virtualization is introduced into cloud computing, resource use efficiency is increased from 15% to 80%. Various universal hardware devices are used. In addition, IT resource O&M costs are reduced by tens of times.
- Virtualization in the cloud computing era consists of computing virtualization, storage virtualization, and network virtualization. Similar to server virtualization and desktop virtualization, network virtualization allows users to obtain network resources on demand.
- There are two types of network virtualization: N-to-1 and 1-to-N. In N-to-1 virtualization, multiple physical network resources are virtualized as a logical resource such as the stacking and cluster technology. In 1-to-N virtualization, one physical resource is virtualized into multiple logical resources.

Huawei Training & Certification

What Is VS?



VS definition

- Virtual System (VS) is 1-toN virtualization technology in switch virtualization. It virtualizes one physical switch into multiple logical switches, each of which works independently and is isolated from each other.

VS characteristics

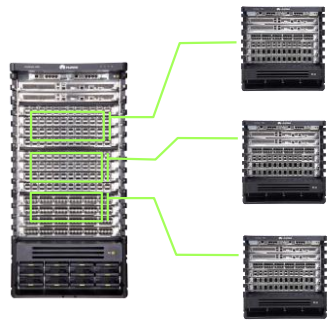
- 1-to-N virtualization: VS virtualizes one physical switch into multiple logical switches.
- VS isolation: Each VS exclusively uses its allocated resources and independently runs network resources. VSs are isolated from each other and work independently.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The virtual machine in data center switches removes barriers between physical devices, changing physical device resources into manageable logical resources. These logical resources run transparently on a physical device platform, implementing isolation and on-demand distribution of resources.
- The Huawei VS is a key feature of Huawei Cloud Fabric Data Center Solution. The Huawei VS provides the technical architecture of network device virtualization, dividing multiple logical or virtual systems on the physical devices. Each VS is a virtual machine on a network device and can be independently configured, managed, and maintained. In addition, each VS is isolated from other VSs, running and processing network services independently. Data center networks process various services and serve various user groups using the VS on physical devices, implementing the following functions:
- Enables service isolation and improves network reliability and security.
- Increases device use efficiency.
- Reduces users' investment.
- Enables isolation between user groups and manages user groups.
- Simplifies network O&M.
- Notes: Currently, the CE12800 supports VS. All products refer to CE12800 series switches unless otherwise stated.

Huawei Training & Certification

VS Setup



- VS control and management
  - VS creation and deletion
  - VS on-demand reset
  - VS suspension
- System resource allocation
  - System resource specifications are set for each VS.
  - Physical cards and interfaces are allocated to VSs.
- VS service deployment
  - (Optional) Configuring the service type on VSs
  - Delivering the service configuration to VSs

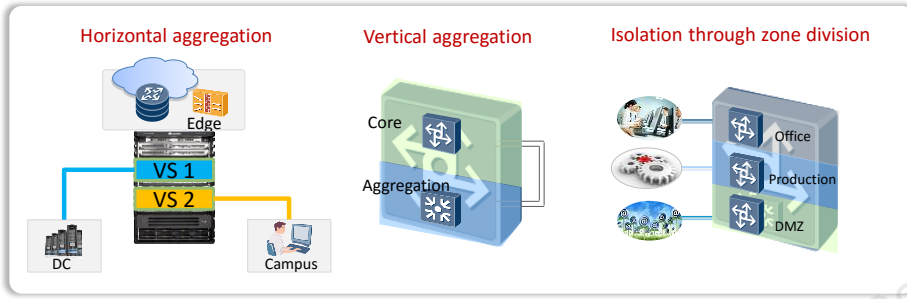
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- To put the virtualization technology into effect, devices must be abstracted, isolated, and encapsulated. The VS architecture is built into the following modes:
- Abstraction
  - The software system of physical devices is abstracted into multiple virtual machines. The virtual machine has an independent and logical control and service panel, forwarding panel, and management panel. The hardware system resources are abstracted into standardized virtual hardware to meet users' requirements. The standardized virtual hardware includes ports, boards, memory, and central processing unit (CPU) resources.
- Isolation
  - Process-level isolation is implemented between multiple virtual machines that run on the same physical device. The abstracted virtual hardware is managed as a virtual machine. Moreover, VSs do not affect each other.
- Encapsulation
  - The virtual machine is encapsulated independently from the virtual context on a specific physical device. Full-service and distributed capabilities and the fine-grained, multi-process mechanism of Huawei VRPv8 are used to build system-level dynamic migration capabilities. These system-level dynamic migration capabilities enable the flexible service deployment and improvement of virtual machine reliability as well as device use efficiency.

Huawei Training & Certification

VS Highlights



Reduce network investment costs and improve resource efficiency

Reduce the operating expenditures (OPEX)

Improve network security and reliability

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Reduce network investment costs and improve resource efficiency.
  - Inter-VS logical isolation replaces traditional physical isolation between different devices, reducing the number of required network devices and network investment costs.
  - VS implements on-demand network device resource allocation, improving switch resource efficiency.
- Reduce the OPEX.
  - Fewer physical network devices are required, reducing requirements for power consumption, refrigeration, and space and reducing the OPEX.
  - Fewer network devices need to be operated and maintained, facilitating operation and maintenance.
- Improve network security and reliability.
  - VS isolates tenants and services, reducing security risks.
  - VSs are isolated, and faults of a single VS do not affect other VSs, improving reliability.

Huawei Training & Certification

Relationship Between VS and Switch

- Relationship between VS and physical switch
  - VS virtualizes one physical switch into multiple virtual switches in a similar manner as one physical server is virtualized into multiple VMs.
  - A VS still needs to share the same MPU with a physical switch. Different VSs are identified and isolated on the MPU using different software processes without affecting each other.
  - A VS and a physical switch share the operating system and software version.
- Relationship between VSs
  - Virtual switches are isolated from each other, cannot directly communicate within the system, and cannot use the common service ports to connect to external networks.
  - VSs function as multiple independent switches.
- VS
  - VSs have their own port resources and system resources.
  - VSs have their own configuration files, administrator accounts, and management channels.
  - VSs complete service configuration and service forwarding independently.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

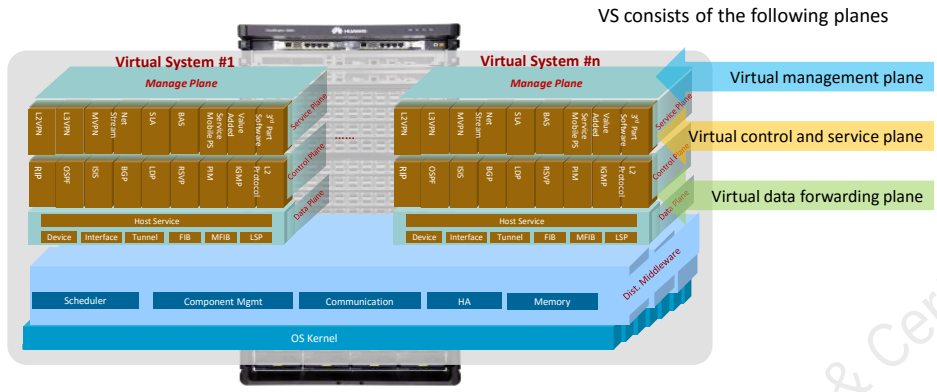
1. VS Overview
2. **VS Implementation**
3. VS Deployment
4. VS Typical Applications

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

VS Software Architecture



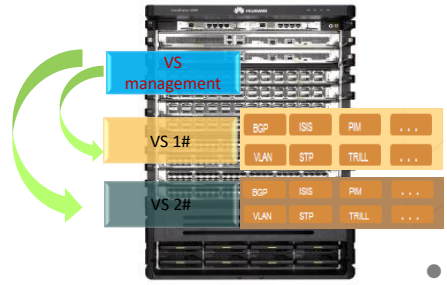
The VS uses a virtual, fine-grained, elastic, and distributed architecture

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The VS uses a virtual, fine-grained, elastic, and distributed architecture. The entire VS is constructed based on full-service and distributed middleware of Huawei VRPv8. Similar to Hypervisor in the server virtual machine, VS control components uniformly schedule and manage multiple VSs. The control components virtualize the control and service plane, data plane, and management plane so that each VS can independently deploy services, upload configuration files, and control network management. Furthermore, the control components enable the VS to provide physical device capabilities. The VS also uses the full-service and distributed capabilities to implement fine-grained and distributed deployment of services. For example, various VS service modules can be distributed on different boards, which substantially increases the hardware resource use efficiency.
- The virtual control and service plane transmits network control protocols and processes user services. Both network reliability and secure isolation are critical. The VS can run in different processes and provides fine-grained process control. The VS uses inter-process isolation and exclusive virtual memory space to prevent control protocols and services from affecting each other. Therefore, VS service reliability and secure isolation capabilities are considerably consolidated. The fine-grained process control mechanism sharply reduces the expense of each VS, and allows a physical device to virtualize 16 VSs simultaneously.

Huawei Training & Certification

VS Control and Management

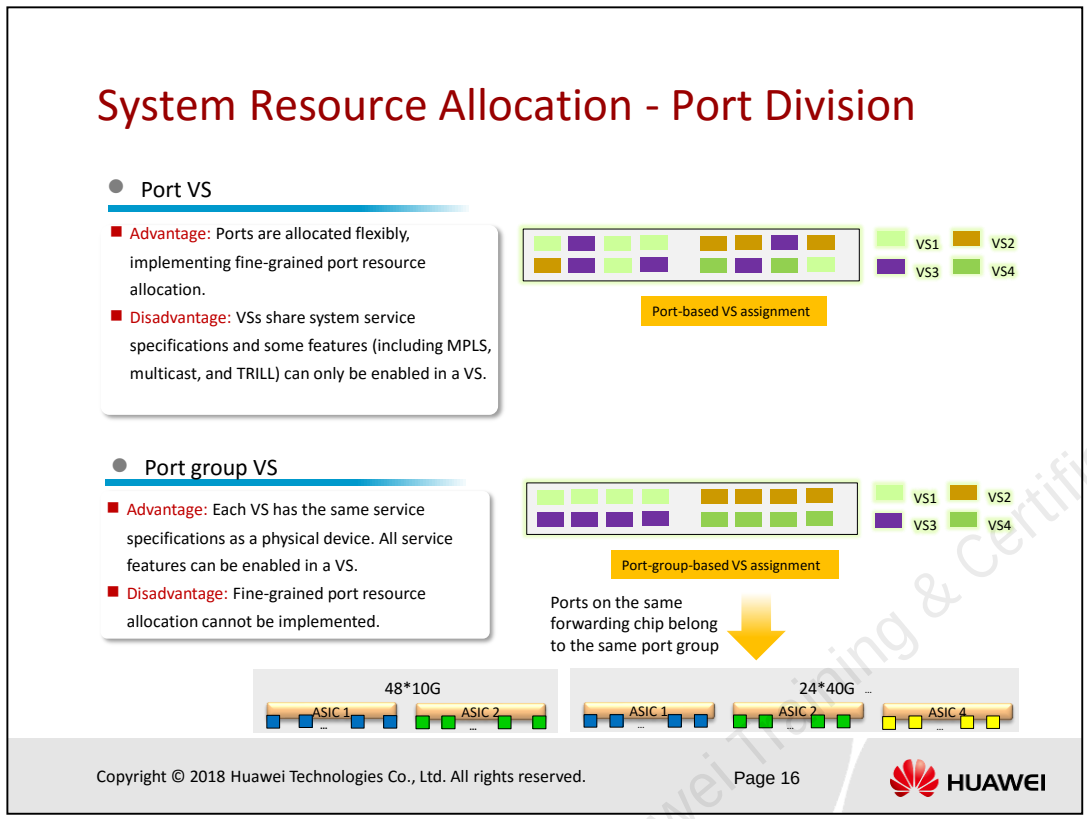


- VS management
  - Administrative VS (default VS): The physical device administrator has administration rights of the physical device and is responsible for creating, deleting, and controlling common VSs.
- VS control
  - VS on-demand reset
  - VS suspension
  - VS switching

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Administrative VS (default VS)
  - The physical device administrator has administration rights of the physical device and enters the administrative VS by default.
  - The administrator can only create, delete, or control common VSs on the administrative VS.
- VS creation and deletion
  - Common VSs are created and submitted on the administrative VS.
  - Resources such as interfaces are allocated to VSs.
  - Regardless of any phase in which a VS fails to be created, the operation can be rolled back.
- Independent VS control
  - Similar to physical devices, VSs are reset, switched, and suspended independently.
- VS on-demand reset
  - The VS stops running services and releases resources. Then the physical device loads the configuration to the VS.
- VS suspension
  - VS services are interrupted or restored.
  - Simulate physical device power-on or power-off operations.

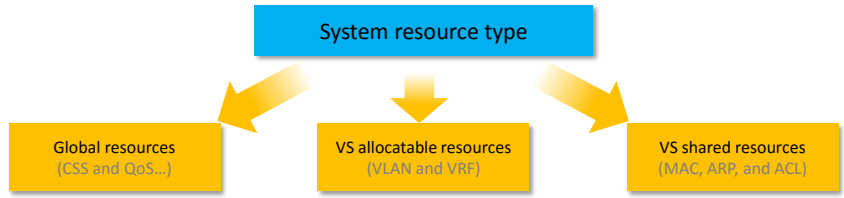
Huawei Training & Certification



- Physical ports specified for a VS are exclusively used by the VS.
- VS port division mode:
  - Port mode
    - Any port is independently allocated to a VS.
    - VSs share system service specifications and some features can only be enabled in a VS.
    - Some services such as multicast, MPLS, and TRILL can only be enabled in a VS.
  - Port group mode
    - A group of ports can be allocated to one VS.
    - The VS exclusively shares system service specifications.
    - All services can be enabled in a VS.

Huawei Training & Certification

System Resource Allocation - System Resource Specification Setting



Item	Definition	Example
Global resources	Global resources are configured only on the administrative VS and are valid for all VSs.	CSS and QoS
Allocatable resources	These resources can be configured using commands or manually configured.	VLAN, VRF, and routing table
Shared resources (valid only for port VS)	These resources cannot be manually allocated on VSs. ➢ All port VSs share system specifications of these resources and preempt the resources. ➢ Each port group VS uses system specifications without involving resource sharing.	LAG, MSTP, MAC, ARP, ECMP, and ACL

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- System resources include global resources, VS allocatable resources and VS shared resources.
  - Global resources such as QoS are configured only on the administrative VS and are valid for all VSs.
  - VS allocatable resources can be independently allocated and reclaimed on each VS, such as VLANs and VRFs.
  - VS shared resources such as MAC, ARP, and ACL are limited by the system based on VS division and are not independently configured on a VS.
- VS resource specifications are relevant to the port division mode.
  - In port mode, all VSs share resources.
    - All the VSs in port mode share 240K MAC addresses.
  - Each VS in port group mode share system specifications.
    - Each VS exclusively uses 240K MAC addresses.

Huawei Training & Certification

VS Allocatable Resources

Resource	System Specifications	Default VS Specifications	
		Group mode	Port mode
IPv4 unicast routing entries	64000000	60000	60000
IPv4 multicast routing entries	60000	1000	1000
IPv6 unicast routing entries	4096000	16000	16000
IPv6 multicast routing entries	8000	100	100
VLANs	Port mode: 4063 Group mode: 28441	4063	200
VPN instances	Port mode: 4096 Group mode: 61440	4096	256
MPLS	Supported	Supported	Not supported
TRILL	Supported	Supported	Not supported
Multicast	Supported	Supported	Not supported
Weight of the maximum CPU usage	-	5	5
Maximum memory usage	-	100	100

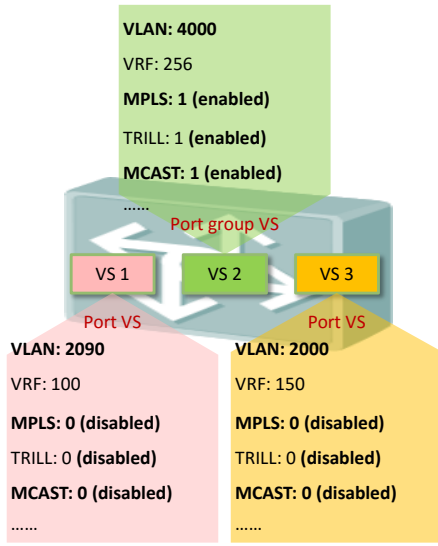
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Port group VS
  - VLAN resources
    - All port group VSs share 12K VLANs and the number of resources to be used by each port group VS is configured.
  - Other resources
    - Each port group VS exclusively uses the full specifications of resources except VLAN resources.
- Port VS
  - Resource sharing
    - All port VSs share system resources without being affecting by the mode in which port VSs are created.
    - The total number of resources used by each port VS cannot exceed the maximum system specifications.
  - Allocatable resources
    - The number of resources to be used by each port VS can be configured using commands.
  - Other resources
    - Port VSs preempt non-allocatable resources.

Huawei Training & Certification
Huawei Training & Certification

Examples of VS Allocatable Resources



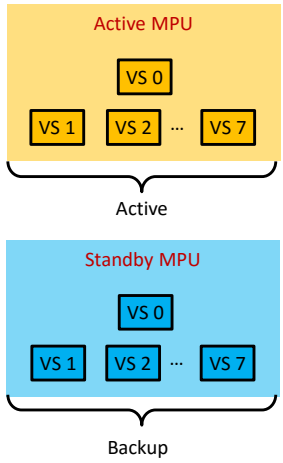
- Application example
  - Create three VSs: VS1, VS2, and VS3. VS1 and VS3 are port VSs, and VS2 is a port group VS.
  - VS2 exclusively uses various resources without competition.
  - VS1 and VS3 need to compete for various resources:
  - The total number of resources used by VS1 and VS3 cannot exceed the system specifications.
  - VS1 and VS3 can not use MPLS, TRILL, and MCAST features.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Configuration requirements on allocatable resources:
  - For port group VS
    - All port group VSs need to compete 12K VLANs, and each VS can be assigned a maximum of 4090 VLANs.
    - A port group VS can exclusively use other resources except VLANs.
  - For port VS
    - The number of routing table resources is separately configured for each port VS and limited by system memory capacity.
    - All port VSs compete for other resources except routing table resources.

Huawei Training & Certification

VS Operation



- All VSs run on the master MPU

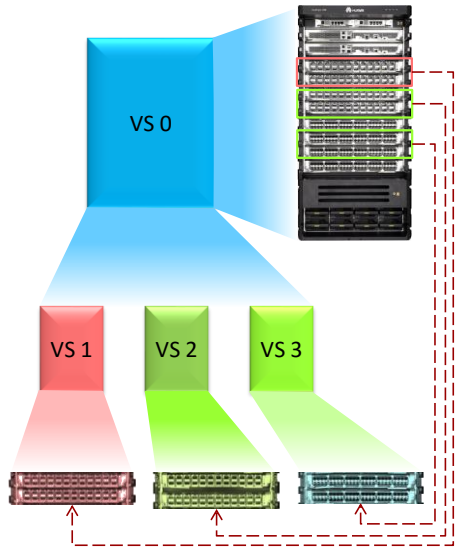
  - All VSs start and run on the active MPU. The standby MPU is a backup of the active MPU.
  - An active/standby MPU switchover will trigger active/standby MPU switchovers of all VSs.
  - VS0 functions as the administrative VS. If you restart VS0, all the VSs are restarted.
  - You can restart a VS except VS0 independently without triggering the restart of other VSs.
  - Different function modules on a VS are isolated in different modes:
    - Process isolation: The same function modules on different VSs run in different software processes, such as routing protocols.
    - Data isolation: The same function modules on different VSs run in the same process but use isolated data modules for VS isolation, such as Layer 2 functions and system management.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

VS Configuration Management Model



VS0 manages all the other VSs (like the State Council)

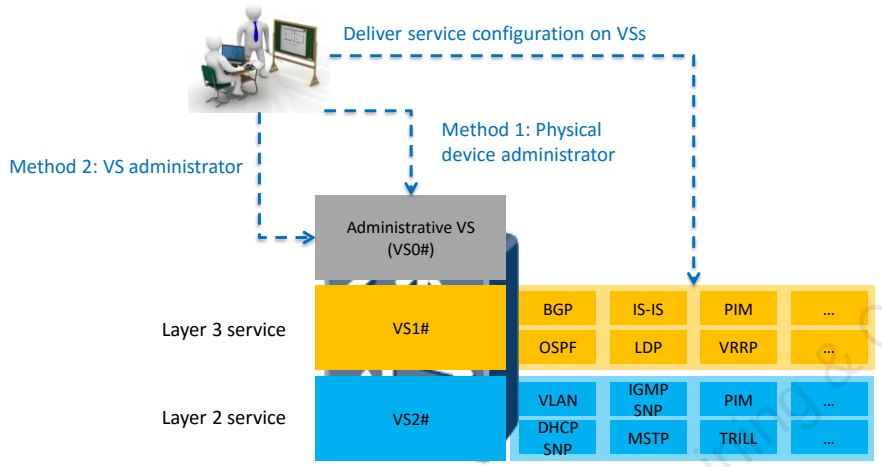
Other VSs manage themselves (like provinces, autonomous regions, and municipalities)

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- VS0 manages all the other VSs
  - Functions as the administrative VS and exists by default to manage global resources.
  - Manages resources of the entire device.
  - Performs device-level management and maintenance, such as active/standby MPU switchover, card reset, and system upgrade and restart.
  - Creates or deletes other VSs.
- Other VSs manage themselves
  - Manage and use the resources allocated to themselves.
  - Performs system-level operations, such as VS mode configuration, startup, and resource allocation/release.
  - Configures, operates, and forwards services.

Huawei Training & Certification

VS Service Deployment

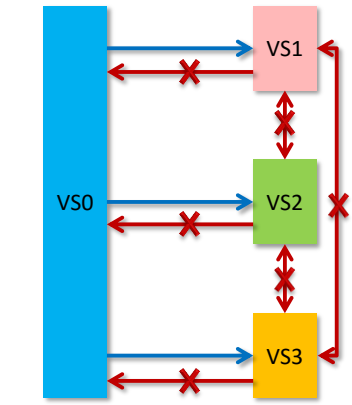


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- VS service deployment role:
  - Method 1: physical device administrator
    - Log in from the management VS to a common VS to configure services.
  - Method 2: VS administrator
    - The local VS administrator logs in to configure services.
    - The local VS administrator cannot operate other VSs.
- (Optional) Enable services on a VS.
  - Specify the service type on a VS to prevent the VS from deploying unauthorized services, and first allocate service-related resources.
  - Enable specified services (MPLS, multicast, and TRILL) on only one VS when the port mode is used.
- Deliver the service configuration on a VS.
  - All service configurations except the device management configuration can be delivered independently on a VS.
  - Service isolation among different VSs
    - For example, different VSs can have overlapping VLAN/VRF IDs.

Huawei Training & Certification

VS Configuration Management Rights Control



Allowed to perform VS configuration management

Prohibited from performing VS configuration management

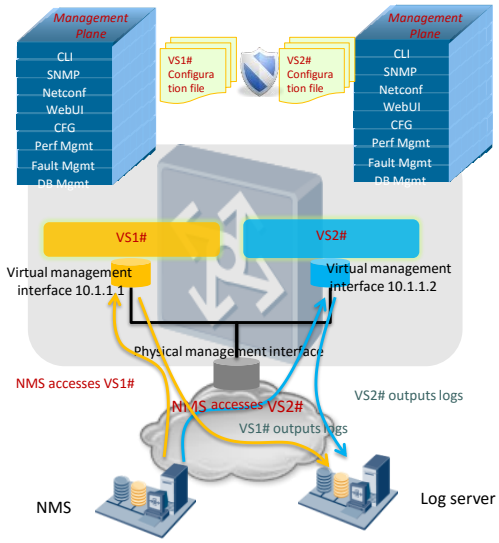
- VS management control
  - VS0 functions as the administrative VS and is allowed to control other VSs. That is, VS0 can use the system administrator account to log in to other VSs to configure and manage these VSs.
  - Other VSs can only manage themselves but not other VSs.
- Rights control when the client logs in to a VS
  - To prevent unauthorized users from logging in to the system, configure ACL rules in the system to determine whether the IP address of the remote client is valid. This ensures VS configuration management security.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- VS0 control over other VSs is saved in system operation logs for tracking.

Huawei Training & Certification

VS Operation and Maintenance



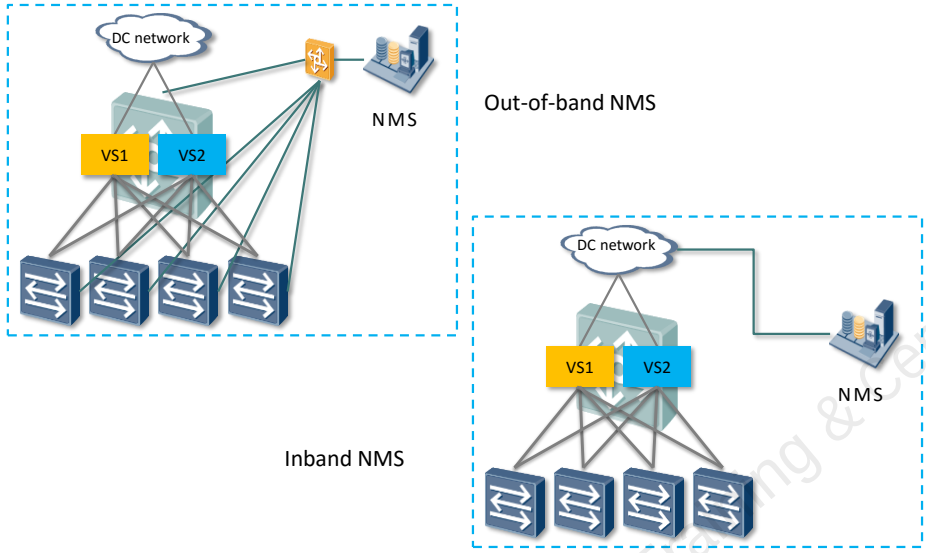
- VSs are operated and maintained as independent logical devices
  - Independent configuration operations
  - Independent configuration file
  - File system isolation
  - Isolation between the VS O&M management channel and rights
  - Accessing each VS through an independent management channel

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Independent configuration operations
- Independent configuration file
  - The local VS configuration file is saved and installed independently.
- File system isolation
  - Files (such as configuration files) of different VSs are isolated, and users in a VS cannot access the file directories in other VSs.
- Isolation between the VS O&M management channel and rights
  - Independent restart
  - Independent diagnostic commands
  - Independent logs, alarms, and NMS servers (such as log/alarm/AAA servers)
- Each VS has its own file systems, configuration files, logs, alarms, and network management servers, implementing independent O&M. Each VS has exclusive network management channels and isolation rights, meeting multiple user clusters' requirements for independent management and secure isolation. This network management mode is called independent management mode. Each VS is managed as an independent network element that has its own topology.
- To satisfy customers' various network requirements, the VS also provides the unified management mode. In this mode, each VS is uniformly managed on a physical network element and does not have its own topology. The unified management mode is applicable to service isolation. The independent management mode integrates service isolation and network isolation, while still independently managing the network.

Huawei Training & Certification

VS Operation and Maintenance Modes



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Out-of-band NMS
  - Deployment: The management network is physically independent. VSs share the management interfaces of the physical device.
  - Requirements: (1) Different VSs need to use different management IP addresses and management MAC addresses. (2) Management IP addresses of all VSs belong to the same subnet.
  - O&M: The NMS logs in to each VS to operate and maintain each VS through the management network.
- Inband NMS
  - Deployment: The management network is borne over the service network. Each VS only needs to connect to the service network without requiring deployment. Each VS needs an independent management IP address and MAC address.
  - O&M: The NMS logs in to each VS to operate and maintain each VS through the service network.

Huawei Training & Certification

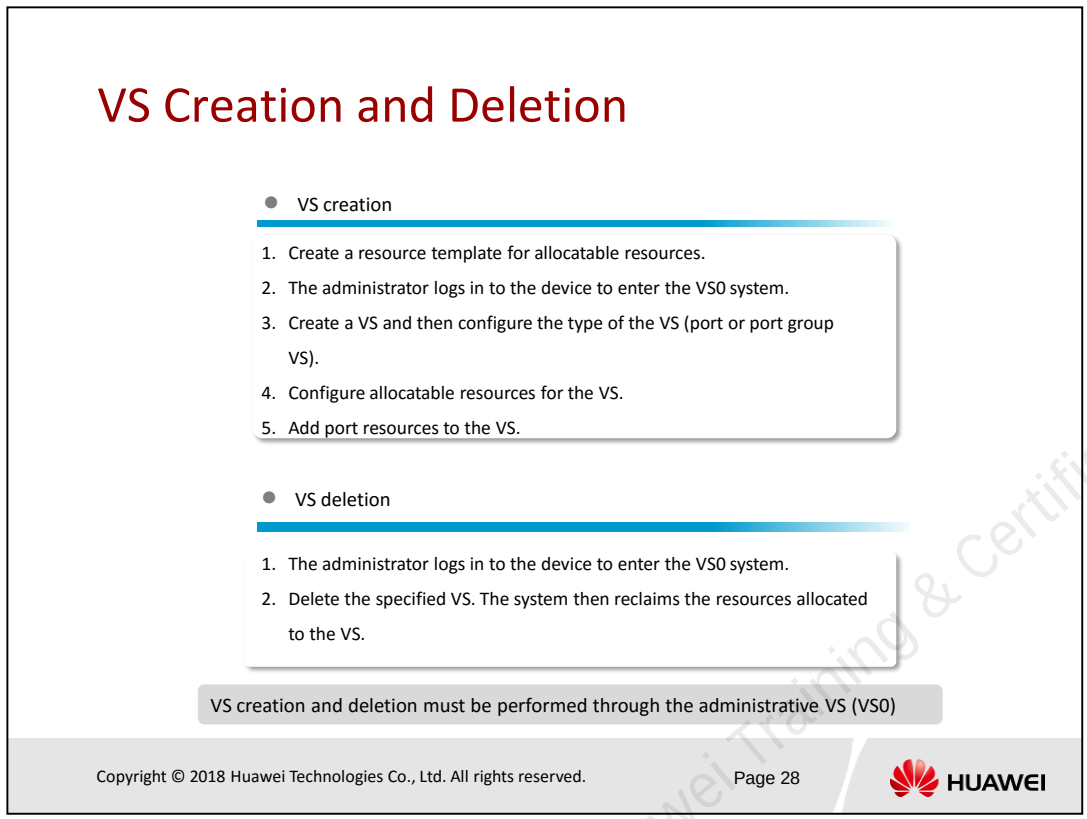
Contents

1. VS Overview
2. VS Implementation
3. **VS Deployment**
4. VS Typical Applications

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

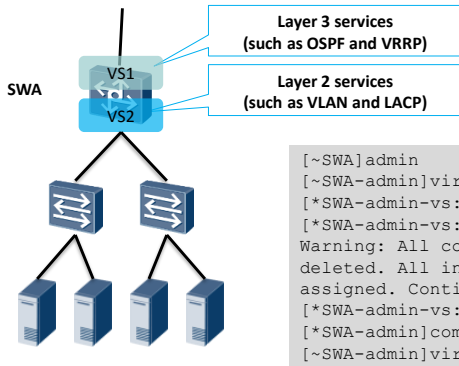
Huawei Training & Certification
Huawei Training & Certification



- When configuring VSs, pay attention to the following points:
- The VS function is controlled by a license. By default, the VS function is disabled on a device. To use the VS function, apply for and purchase a license from a Huawei agent or a Huawei local office.
- A PS supports a maximum of 16 VSs (including the admin-VS). By default, seven VSs in group mode can be created. To create more VSs in group mode, adjust the number of VLANs occupied by VSs.
- Only the PS administrator can create VSs and allocate resources to VSs.
- VSs can be created only when the PS has sufficient physical and logical resources to be allocated.
- Only physical ports on LPUs can be assigned to non-admin-VSs. However, physical member ports in a stack and member ports of an Eth-Trunk cannot be assigned to non-admin-VSs.
- After a physical port is assigned to a VS, the port cannot be configured as a physical member port.
- There is no mapping between a logical resource template and a VS. After a logical resource template is loaded on a VS, modifying the logical resource template does not affect the VS. The modified logical resource template takes effect only after it is loaded to the VS again.

Huawei Training & Certification

Example for Creating a Port Group VS



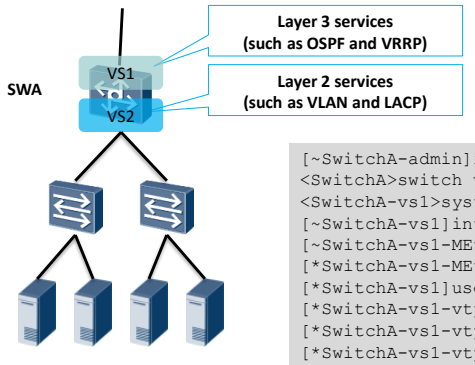

```
[~SWA]admin
[~SWA-admin]virtual-system vs1
[*SWA-admin-vs:vs1]port-mode group
[*SWA-admin-vs:vs1]assign interface 10GE 1/0/0
Warning: All configurations of the interfaces will be
deleted. All interfaces of the same group will be
assigned. Continue? [Y/N]: y
[*SWA-admin-vs:vs1]quit
[*SWA-admin]commit
[~SWA-admin]virtual-system vs2
[*SWA-admin-vs:vs2]port-mode group
[*SWA-admin-vs:vs2]assign interface 10GE 1/0/24
Warning: All configurations of the interfaces will be
deleted. All interfaces of the same group will be
assigned. Continue? [Y/N]: y
[*SWA-admin-vs:vs2]quit
[*SWA-admin]commit
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- On the network, all services need to be processed by the MBs of the PS. In such a case, when a service failure on the PS causes a failure of the PS, other services running on the PS cannot be properly forwarded. To prevent this problem, users require that Layer 2 and Layer 3 services deployed on the PS be isolated from each other to improve network security.
- Run: system-view The system view is displayed.
- Run: admin The admin view is displayed.
- Run: virtual-system vs-name A VS is created.
- Run: port-mode { group | port } [resource-template template-name] The interface assignment mode and a logical resource template are specified for the VS.
- (Optional) Run: description description Description information is configured for the VS.
- By default, no description information is configured for a VS.
- Run: assign interface interface-type interface-number Physical interfaces are assigned to the VS.
- NOTE: Before assigning physical interfaces to a VS in port group mode, run the display device port-map [slot slot-id] command in the user view or system view to view the mapping between forwarding chips and interfaces to learn which interfaces will be assigned to the VS.
- Run: commit The configuration is committed.

Huawei Training & Certification

Configuring a Management Account for a VS




```
[~SwitchA-admin]return
<SwitchA>switch virtual-system vs1
<SwitchA-vs1>system-view
[~SwitchA-vs1]interface MEth 0/0/0
[~SwitchA-vs1-MEth0/0/0]ip address 10.1.1.10 24
[*SwitchA-vs1-MEth0/0/0]quit
[*SwitchA-vs1]user-interface vty 0 4
[*SwitchA-vs1-vty0-4]authentication-mode aaa
[*SwitchA-vs1-vty0-4]user privilege level 15
[*SwitchA-vs1-vty0-4]quit
[*SwitchA-vs1]aaa
[*SwitchA-vs1-aaa]local-user vs1_user password
irreversible-cipher VS1_password
[*SwitchA-vs1-aaa]local-user vs1_user service-type telnet
[*SwitchA-vs1-aaa]local-user vs1_user level 3
[*SwitchA-vs1-aaa]quit
[*SwitchA-vs1]commit
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- A VS functions as an individual physical device. The PS administrator must perform initialization configurations, for example, configuring an IP address or a management account for the VS to facilitate VS user login and management.
- The PS administrator can switch from the Admin-VS to a VS using a command. After entering the VS, the PS administrator has the highest privilege level and can configure an IP address or a management account for the VS in a way same as for a PS. The VS administrator can log in to the VS using the configured IP address and user name to perform service configurations.
- NOTE: Only the PS administrator can switch from the Admin-VS to a VS.
- Procedure
- Run the following command in the user view:
- switch virtual-system vs-name The PS administrator switches from the Admin-VS to the VS.
- After switching to the VS, you can run the quit command to switch back to the user view of the PS.
- Configure an IP address.
    - Run: system-view The system view is displayed.

Huawei Training & Certification

VS Operation Precautions

System upgrade

- Version upgrade and patch installation are system-level operations and do not differentiate VSs. Performing these operations will affect all VSs.

VS resource modification

VS mode

- After a VS is created, modifying the VS mode is not allowed. You can only delete and then create another VS and specify the VS

Allocatable resources

- Reduce allocatable resources: You can reduce occupied resources. If services have been configured on resources, these resources cannot be released.
- Increase allocatable resources: You can increase the number of resources provided by the system. If the specified value exceeds the number of allocatable resources, the configuration fails.

Port resources

- Delete port resources: Deleted ports are automatically reclaimed by the system.
- Add port resources: The port (port VS) or corresponding port group (port group VS) cannot be used by other created VSs.

CSS deployment

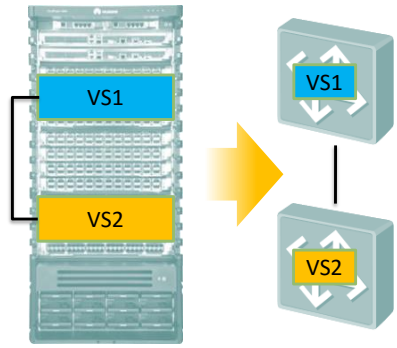
- VS can be used together with CSS. CSS must be configured before VSs are created.
- CSS does not belong to any VS. CSS management and detection are performed by the system without being controlled by VSs.
- All inter-device VSs share the CSS management/forwarding channel provided by the system for control and forwarding.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Communication Between VSs



Physical link

- VSs must connect to each other through physical links

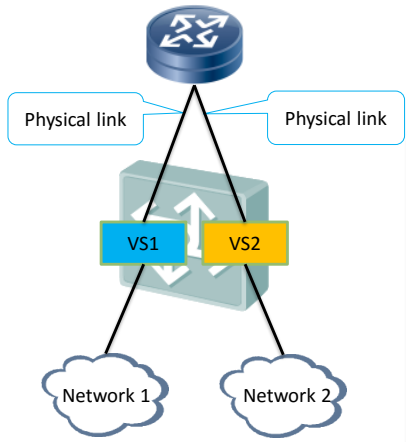
- Physical port resource isolation and logical system resource isolation are performed on multiple VSs that are virtualized from the same physical switch. Therefore, VSs cannot directly communicate.
- To enable two VSs to communicate like two physical switches, use physical cables or fibers to connect physical ports of the two VSs.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Communication Between VS and External Network Device



- VSs cannot share physical ports to communicate with external network devices

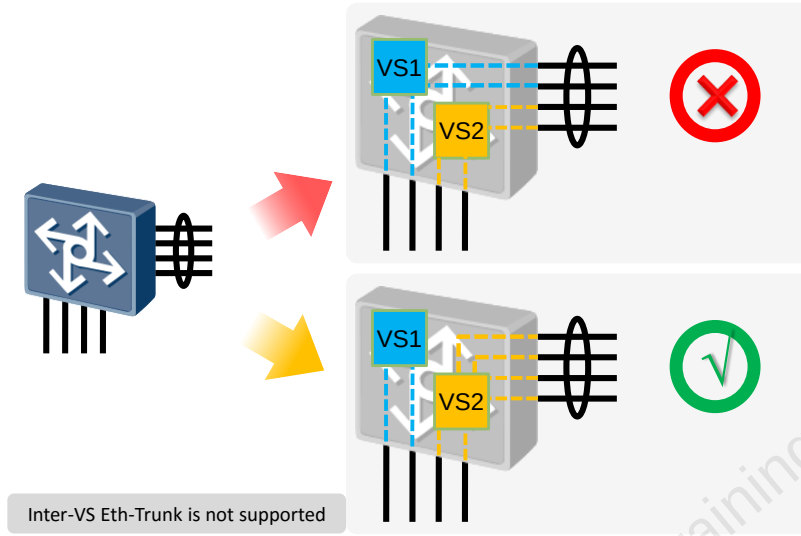
- Port resources of multiple VSs that are virtualized from the same physical switch are physically isolated. Different VSs cannot share physical ports to connect to external network devices.
- During network deployment, each VS must use its own ports to connect to external network devices.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

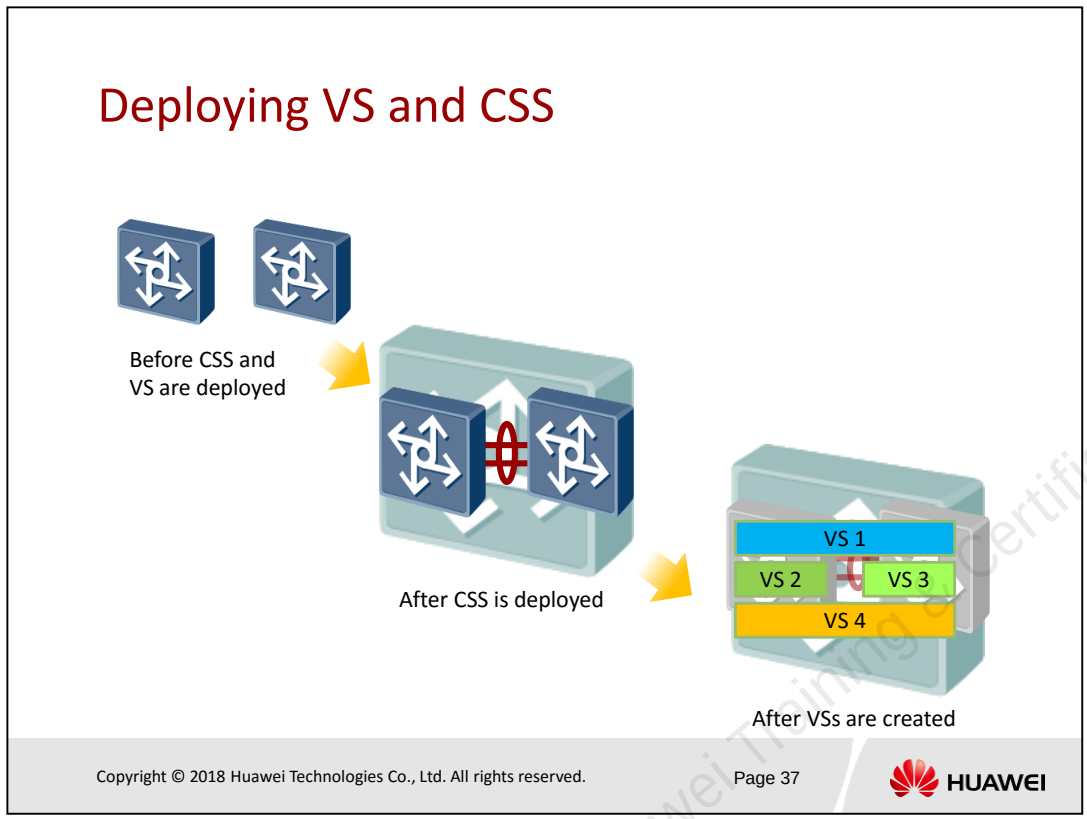
Configuring Eth-Trunk in VS



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

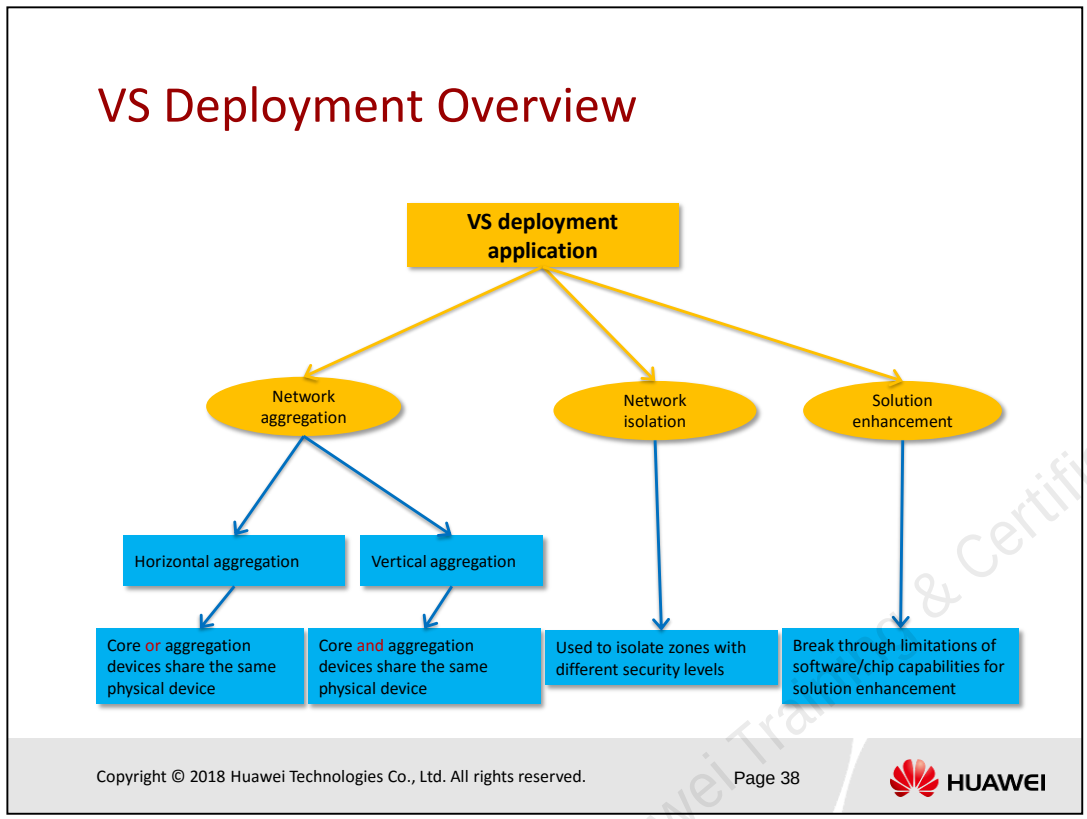
- Port resources of VSs are physically isolated. VSs cannot share ports. An Eth-Trunk port can only be configured in a VS but cannot be deployed among VSs.
- Requirements for configuring Eth-Trunk in VSs are as follows:
  - A port cannot be added to a VS if the port belongs to an Eth-Trunk.
  - An Eth-Trunk can only be configured in a specified VS. The system only allows adding the port that belongs to the VS to the Eth-Trunk.

Huawei Training & Certification



- You must configure CSS before configuring VS. When both VS and CSS need to be deployed, you must configure CSS in the system first. After CSS takes effect, you can create and configure VSs based on the CSS system.
- CSS does not belong to any VS. CSS can only be configured in VS0. CSS management channel, forwarding channel, and split detection channel do not belong to any other VS.
- Inter-device VSs share the CSS interconnection channel. Inter-CSS VSs of different physical devices, such as VS1 and VS4) use CSS management/forwarding channel for VS management and service processing.
- VSs are flexibly configured on the CSS. VSs configured in the CSS can occupy ports of one device or occupy ports of multiple devices.
- The CSS supports the same number of VSs as a physical device. The number of VSs that can be deployed on the CSS is the same as a single physical device. For example, a CE12800 running V1R1 supports a maximum of eight VSs, and the CSS containing CE12800s also supports a maximum of eight VSs.

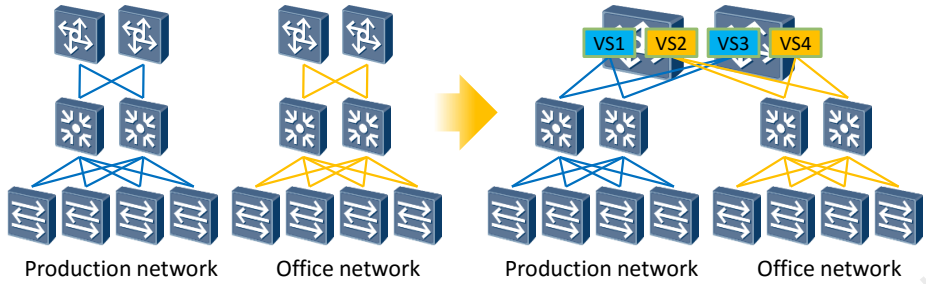
Huawei Training & Certification



- Common VS application scenarios:
  - Network aggregation
    - Horizontal network aggregation: VS can be used on the multiple networks that need to be physically isolated so that core/aggregation devices can share the same physical device. This virtualization reduces the number of required devices and saves investment costs while ensuring functions and performance.
    - Vertical network aggregation: The common networking mode is core-aggregation-access. VS can be used to virtualize independent physical devices deployed at the core and aggregation layers, retaining the original network hierarchy and service deployment.
  - Network isolation
    - Implement isolation between zones with different security levels on the same physical device.
    - Support isolation between different service functions, such as FCoE deployment.
  - Solution enhancement
    - Optimize the solution according to the gap between customer requirements and product/chip capabilities.
    - Implement differentiated networking according to the differences in network performance requirements of different services.
- VS application and deployment guide
  - Regard VS as a common physical switch: You can configure and use VS like a common physical switch.
  - When applying and deploying VS, identify the scenarios in which VS can be used to simplify network architecture and reduce costs and use VS to replace physical switches in the solution.

Huawei Training & Certification

Using VS for Horizontal Aggregation of Network Devices



Deploy core switches on VS so that two networks can share physical switches at the core layer and exclusively use different VSs for isolation

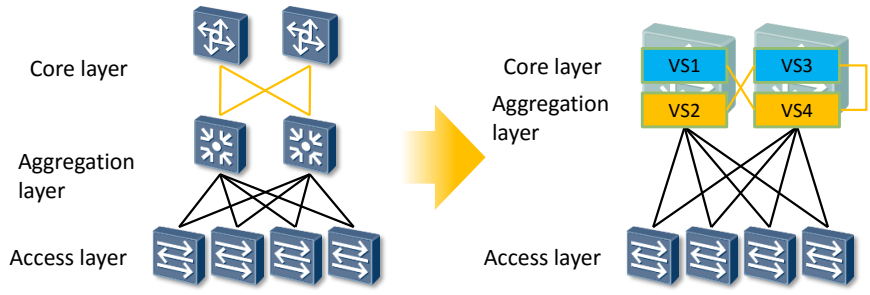
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Service status and requirements
  - Status: The customer needs to deploy two independent networks to run different services.
  - Requirements: Core network devices have super-high performance and access capabilities, which exceed current enterprise requirements of enterprises. Deploying core devices on two networks will increase the investment costs.
- Deployment solution and benefits
  - Solution: Deploy core switches on VS so that two networks can share physical switches at the core layer and exclusively use different VSs for isolation. Configure VS in port group mode and ensure that each VS has the same performance as a physical switch.
  - Benefits: Meet service requirements, make full use of device capabilities, and reduce the investment costs.

Huawei Training & Certification

Using VS for Vertical Aggregation of Network Devices



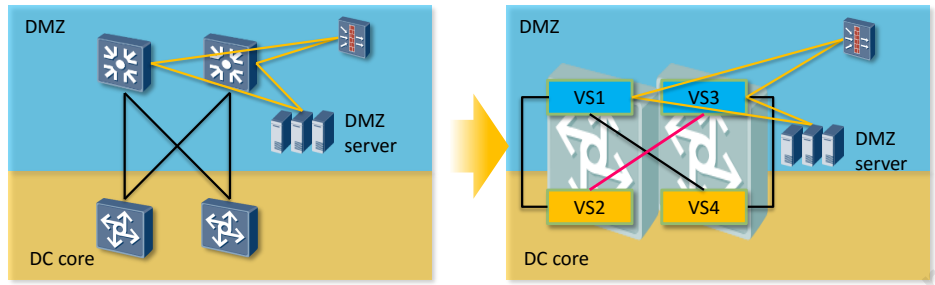
Deploy core layer devices and aggregation layer devices on VS to share the same physical switch

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Service status and requirements
  - Status: Three layers (core, aggregation, and access layers) are deployed on the customer's network.
  - Requirements: The customer wants to reduce network investment costs and simplify network device O&M without changing the existing network model.
- Deployment solution and benefits
  - Solution: Deploy core layer devices and aggregation layer devices on VS to share the same physical switch. Configure VS in port group mode and ensure that each VS has the same performance as the physical switch.
  - Benefits: Reduce network investment costs and simplify network O&M.

Huawei Training & Certification

Using VS for Isolation Between Zones with Different Security Levels



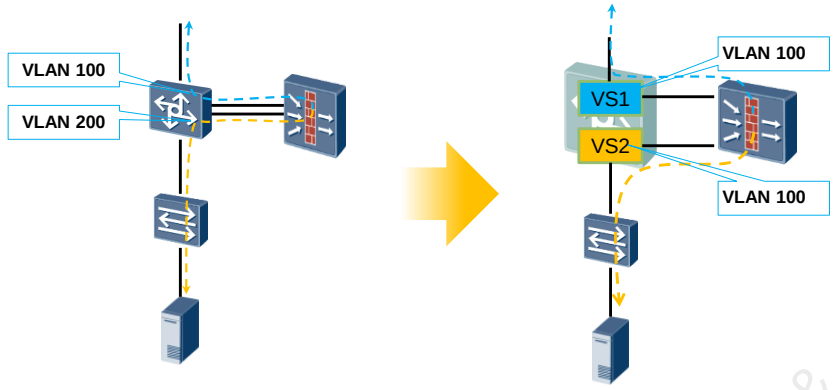
Deploy switches in different security zones on VS to share the same physical device

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Service status and requirements
  - Requirements: Different security policies need to be performed in different security zones (such as the DMZ and DC core)
  - Status: Traditionally, different security zones are designed and different devices and security policies are deployed in these security zones.
- Deployment solution and benefits
  - Solution: Deploy switches in different security zones on VS to share the same physical device, remain the original security zone partition and security model, and deploy different security policies on VSs in different zones.
  - Benefits: Keep the security mode unchanged, reduce the investment costs, and simplify network O&M.

Huawei Training & Certification

Using VS for Secure Service Deployment



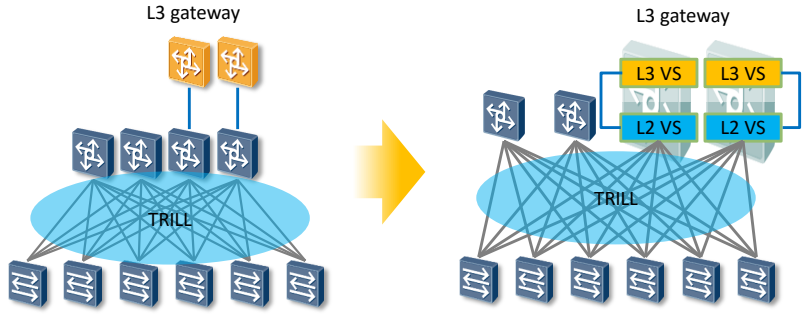
Deploy VS so that security devices physically deployed in bypass mode are logically deployed as the inline mode

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Service status and requirements
  - Status: Current secure service deployment is implemented through complex VLAN planning and configuration on switches.
  - Problem: VLAN planning is complex and the number of available VLANs is insufficient as services increase.
- Deployment solution and benefits
  - Solution: Deploy VS to allow security devices physically deployed in bypass mode to be logically deployed as the inline mode. Configurations of VSs are independent of each other.
  - Benefits: Keep the security model unchanged, simplify network deployment, and reduce O&M costs.

Huawei Training & Certification

Using VS for L3 Gateway Deployment on the TRILL Network Edge



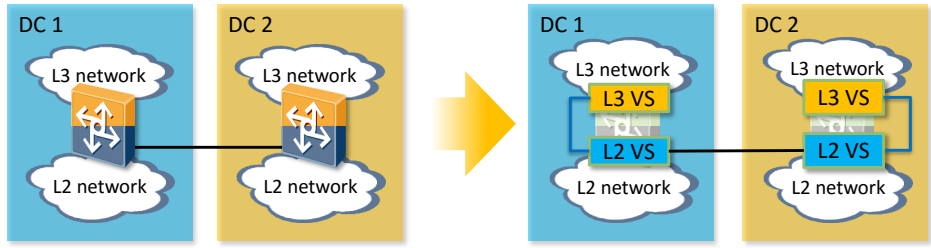
Configure two port group VSs on the TRILL edge device: one functioning as the TRILL network edge device and the other as the L3 gateway

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Service status and requirements
  - Status: A TRILL edge device cannot function as a L3 gateway because of TRILL encapsulation and forwarding modes.
  - Problem: Deploying a gateway for attached devices will increase network investment costs.
- Deployment solution and benefits
  - Solution: Configure two port group VSs on the TRILL edge device: one functioning as the TRILL network edge device and the other as the L3 gateway. Deploy VRRP on the gateway to improve reliability and ensure a loopfree network.
  - Benefits: Reduce network investment costs and simplify network deployed and O&M.

Huawei Training & Certification

Using VS for Differentiated Forwarding Resource Allocation (1)



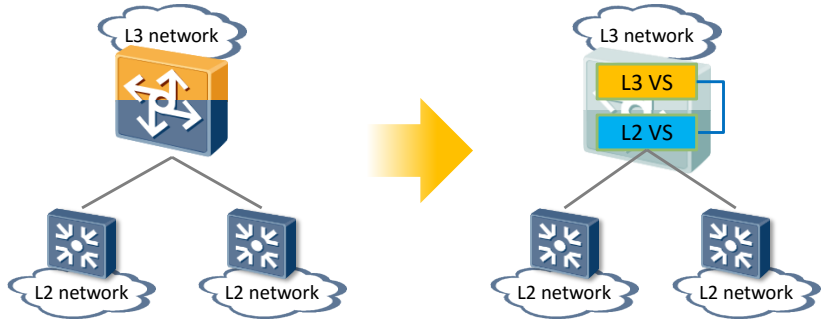
Deploy port group VSs on L2 and L3 edge devices, configure a large MAC table (200K) on the L2 VS for inter-DC L2 forwarding, and configure a large ARP table (120K) on the L3 VS for inter-subnet L3 access

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Service status and requirements
  - Status: A switch often supports both L2 and L3 forwarding, but ARP table and MAC table will compete for forwarding resources.
  - Problem: Inter-DC L2 interconnection requires more MAC entries, causing devices unable to meet L2 and L3 forwarding performance requirements.
- Deployment solution and benefits
  - Solution: Deploy port group VSs on L2 and L3 edge devices, configure a large MAC table (200K) on the L2 VS for inter-DC L2 forwarding, and configure a large ARP table (120K) on the L3 VS for inter-subnet L3 access.
  - Benefits: Improve network performance and reduce network investment costs.

Huawei Training & Certification

Using VS for Differentiated Forwarding Resource Allocation (2)



Deploy two port group VSs on the L2 and L3 edge switches, configure a large MAC table (200K) on one VS for L2 forwarding, and configure a large ARP table (120K) on the other VS for external access to internal hosts

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Service status and requirements
  - Status: A switch often supports both L2 and L3 forwarding, but ARP table and MAC table will compete for forwarding resources.
  - Problem: When switches need to connect through different ports, a large number of hosts are required for L2 interconnection. This requirement increases the number of required MAC entries, causing devices unable to meet L2 and L3 forwarding performance requirements.
- Deployment solution and benefits
  - Solution: Deploy two port group VSs on the L2 and L3 edge switches, configure a large MAC table (200K) on one VS for L2 forwarding, and configure a large ARP table (120K) on the other VS for external access to internal hosts.
  - Benefits: Improve network performance and reduce network investment costs.

Huawei Training & Certification

Contents

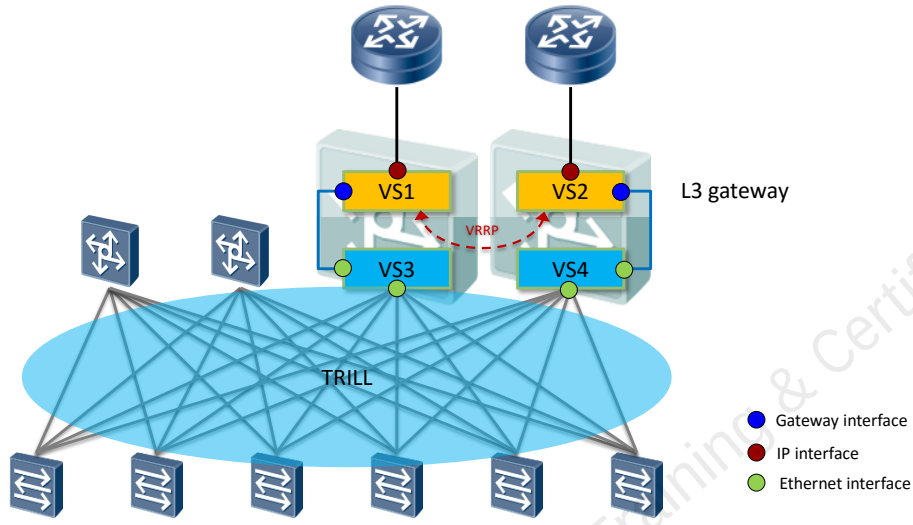
1. VS Overview
2. VS Implementation
3. VS Deployment
4. **VS Typical Applications**

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Using VS for TRILL Gateway Deployment (1)

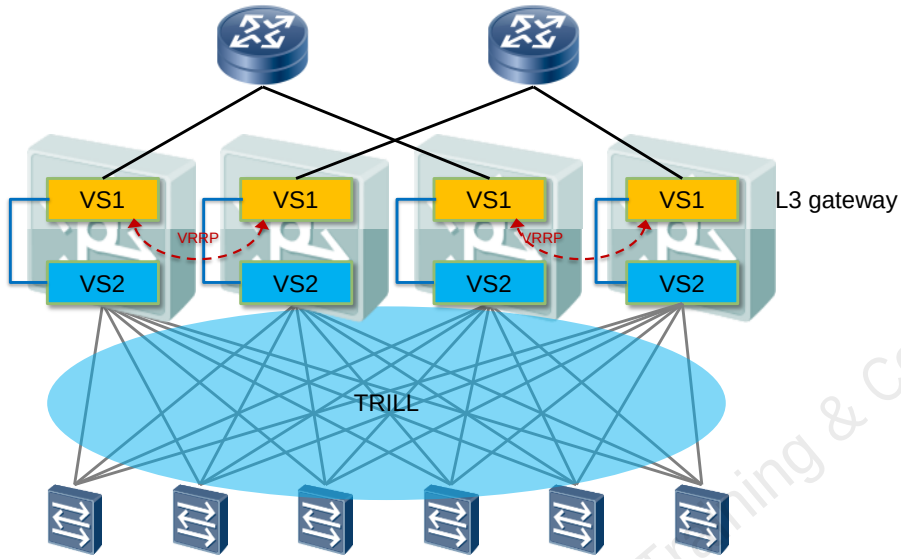


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Service deployment
  - Configure two TRILL core devices as TRILL gateways and create two port group VSs on each of the two devices.
  - Configure one VS as a TRILL core device and configure a large MAC table for the VS. Configure the other VS as an L3 gateway and configure a large ARP table for the VS.
  - Enable L2 services to be forwarded within the TRILL network and L3 services to be sent to the gateway for processing.
- Reliability deployment
  - Deploy VRRP on the two gateway VSs to improve gateway reliability.
  - Configure multiple VRRP groups and specify different master gateways to distribute services of different subnets to the two gateway VSs for load balancing.
  - For expansion, deploy gateway VSs on four TRILL core devices to improve reliability through VRRP, and load balance traffic to multiple gateway VSs through multiple VRRP groups.

Huawei Training & Certification

Using VS for TRILL Gateway Deployment (2)

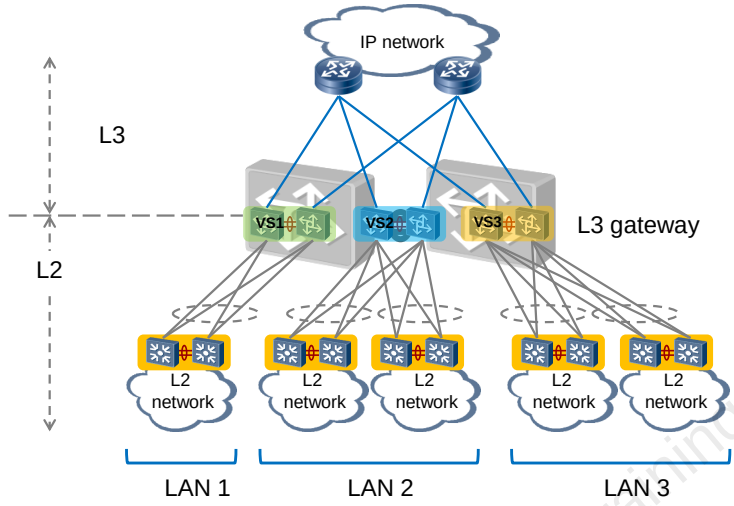


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Reliability deployment
  - Deploy VRRP on the two gateway VSs to improve gateway reliability.
  - Configure multiple VRRP groups and specify different master gateways to distribute services of different subnets to the two gateway VSs for load balancing.
  - For expansion, deploy gateway VSs on four TRILL core devices to improve reliability through VRRP, and load balance traffic to multiple gateway VSs through multiple VRRP groups.

Huawei Training & Certification

Using VS for Multi-tenant Scale Expansion on Physical Network



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Service deployment
  - Deploy CSS on core switches and then configure multiple port group VSs.
  - Connect aggregation switches in different service zones to different port group VSs.
  - Assign tenants to each service zone based on VLANs and deploy a gateway on the access VS of a service zone. A maximum of 4K tenants are supported.
  - Run routing protocols between service zones and implement L3 interconnection.
- Reliability deployment
  - Deploy VSs on different physical switches in a CSS to improve VS reliability.
  - Deploy Eth-Trunk between aggregation switches and core VSs in each service zone to improve network reliability.

Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

FCoE Feature of CE Series Switches

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Foreword

- In data centers in the cloud computing era, applications and services are rapidly and flexibly deployed. Data center networks must be highly scalable. In the original network architecture, due to requirements for high service quality and characteristics of switching device performance and protocol standards, the FC SAN scale is small. Many small-scale storage isolated sites are often formed on data center networks, causing a difficulty in expanding networks. Ethernet technologies develop continuously. At the end of 2011, server 10GE network adapters increasingly become the mainstream server configurations. Data networks step into the 40G/100G era with high density and large capacity, allowing scalability. Data networks are used to transmit storage services. The requirements for unified switching are almost certain. FCoE is used to construct the converged network.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Objectives

- Upon completion of this course, you will be able to understand:
  - FCoE background
  - FCoE implementation
  - Relevant FCoE technologies
  - Technologies used by lossless Ethernet
  - Huawei FCoE solution

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. **Data Center Network Development Trend**
2. FC Network Overview
3. FCoE Network Overview
4. FCoE Data Forwarding
5. Enhanced Ethernet Technologies
6. Huawei FCoE Network Convergence Solution
7. Basic Configurations of FCoE

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

History of Storage Technologies

1 Embedded Storage (ES)

- Deployment: Storage devices are deployed in servers.
- Characteristics: Storage medium cannot be shared, and scalability is low.
- Application: hard disk on PCs and servers



2 Direct Storage (DS)

- Deployment: Servers are directly connected to a storage array.
- Characteristics: Storage medium cannot be shared, and scalability is limited by access capabilities of storage devices.
- Application: extended local storage devices for servers



3 Network Attached Storage

- Deployment: Servers are connected to a storage array through an IP network.
- Characteristics: Shared storage medium, high scalability, and file storage
- Application: used with file servers



4 Storage Area Network (SAN)

- Deployment: Servers are connected to a storage array through an IP or FC network.
- Characteristics: Shared storage medium, high scalability, and high-efficient block storage
- Application: used with database servers

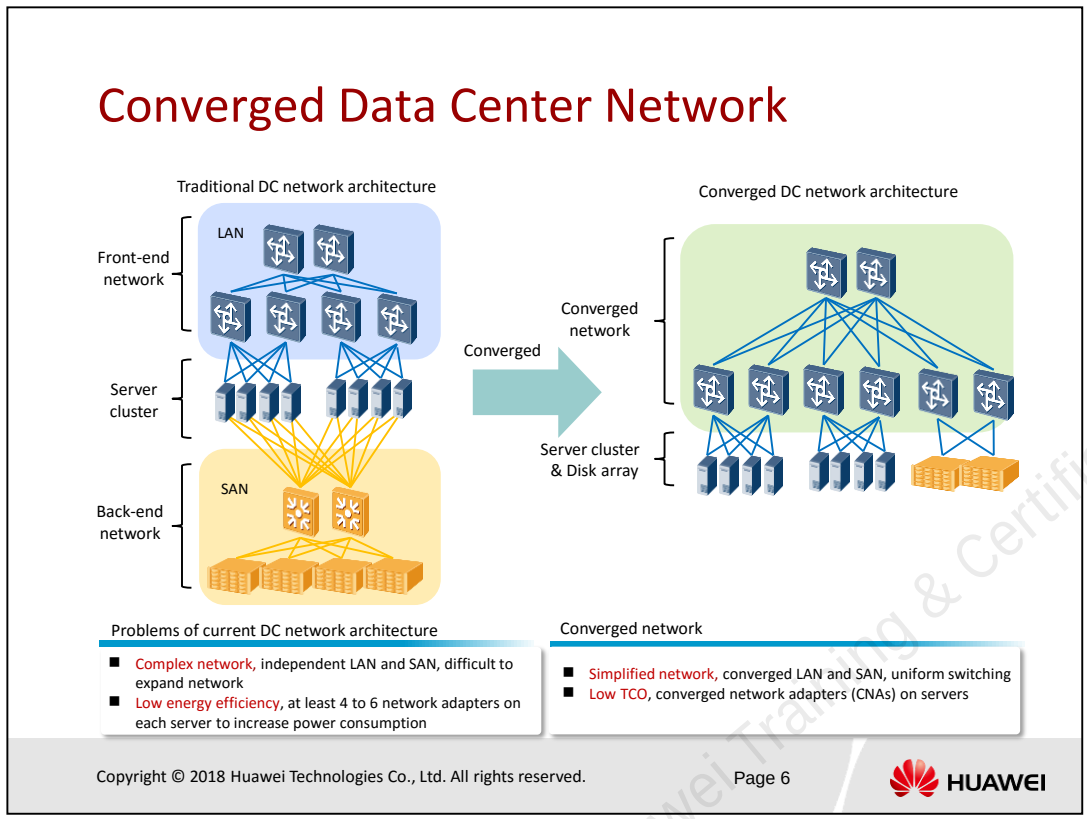


Among existing storage technologies, FC SAN can provide high-speed access to shared storage medium but is not compatible with Ethernet.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

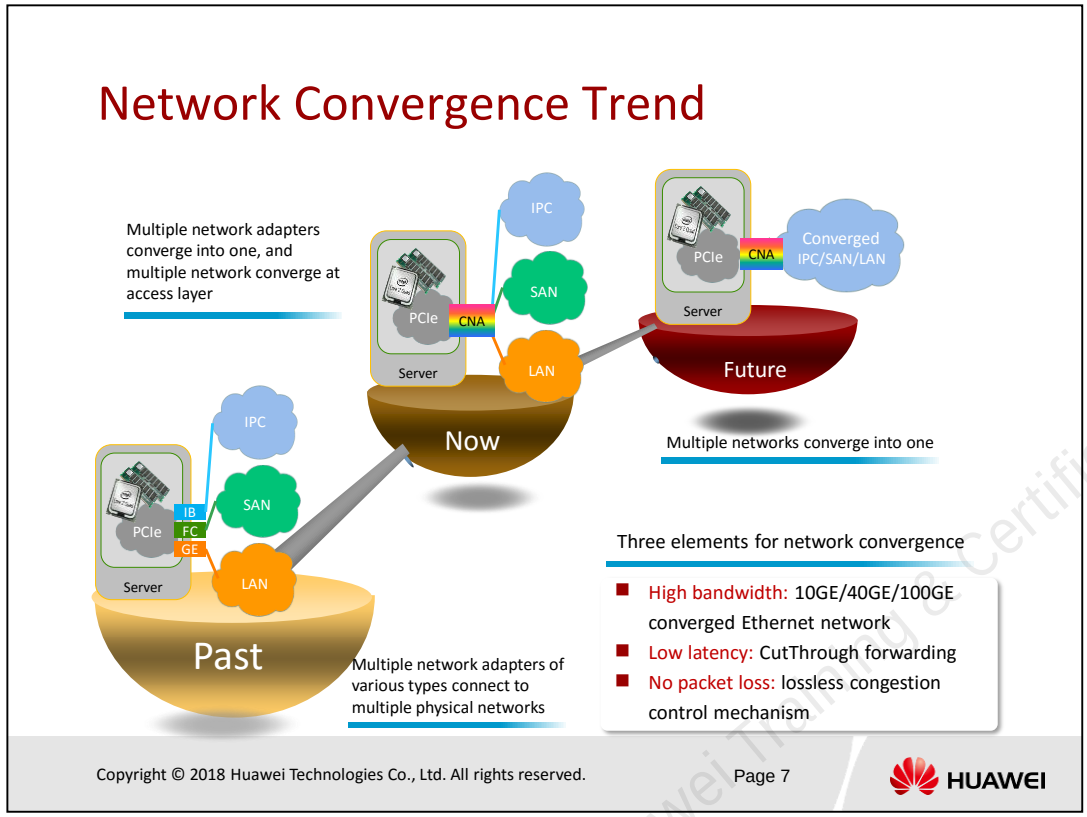
HUAWEI

Huawei Training & Certification
Huawei Training & Certification



- For traditional data center architecture, there are three independent networks: data network, SAN, and HPC network:
  - Access interfaces in data centers are often connected through Ethernet and construct a high-speed data network.
  - A data center mainly uses the NAS and FC SAN to transmit storage services.
  - Infiniband and Ethernet are often used for concurrent computing between servers.
  - A server requires at least four to six network adapters. Two network adapters connect to the FC SAN, two network adapters connect to the Ethernet, and two network adapters connect to the distributed computing network between servers.
  - Different servers use heterogeneous architecture of operating systems and upper-layer software, and different interfaces and data formats.
- For the converged data center network, only one network is required, that is, Ethernet (LAN).
  - FC switches provide access for FC storage services and forward them on LANs.
  - The HPC network can also run on a LAN.
  - A server requires only one CNA.

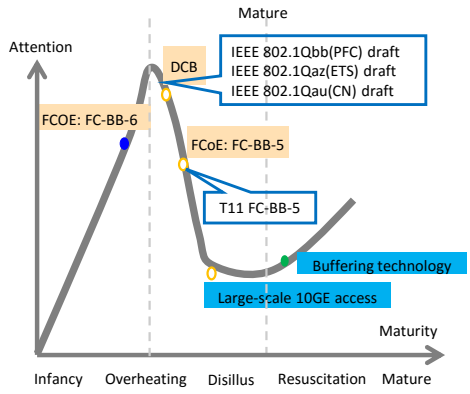
Huawei Training & Certification



- Large bandwidth: A large amount of data exists between servers and storage devices, so large bandwidth is required. Currently, Ethernet interfaces provide 10GE, 40GE, and future 100GE. In addition, there are technologies such as link aggregation and load balancing. The interfaces and technologies can meet bandwidth requirements.
- Low latency: The Cutthrough mechanism of the Ethernet ensures low latency forwarding of data packets.
- No packet loss: Traditional Ethernet technologies cannot ensure non-packet loss, even if QoS is used. To construct a lossless FCoE network, no-packet loss is the key.

Huawei Training & Certification

Technology and Standard Evolution



Large-scale 10GE access

Lossless Enhanced Ethernet

Fiber Channel over Ethernet

10GE and Lossless Enhanced Ethernet technologies have been standardized, while FCoE is still being optimized.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Large-scale 10GE access
  - 10GE CNA. The technology has been well developed and related products have been launched to market.
  - 40GE/100GE network products have emerged.
- Lossless Enhanced Ethernet
  - DCB, defines service-based back pressure congestion control mechanism based on FC BB-Credit and EE-Credit. This standard has been published.
- Fibre Channel over Ethernet
  - FC-BB-5, an FC over Ethernet standard that defines the encapsulation format, signaling control, and forwarding model of FC packets on Ethernet. This standard has been published.
  - FC-BB-6, optimizes the local forwarding mechanism and defines FDF local forwarding entities. The standard is a draft now.

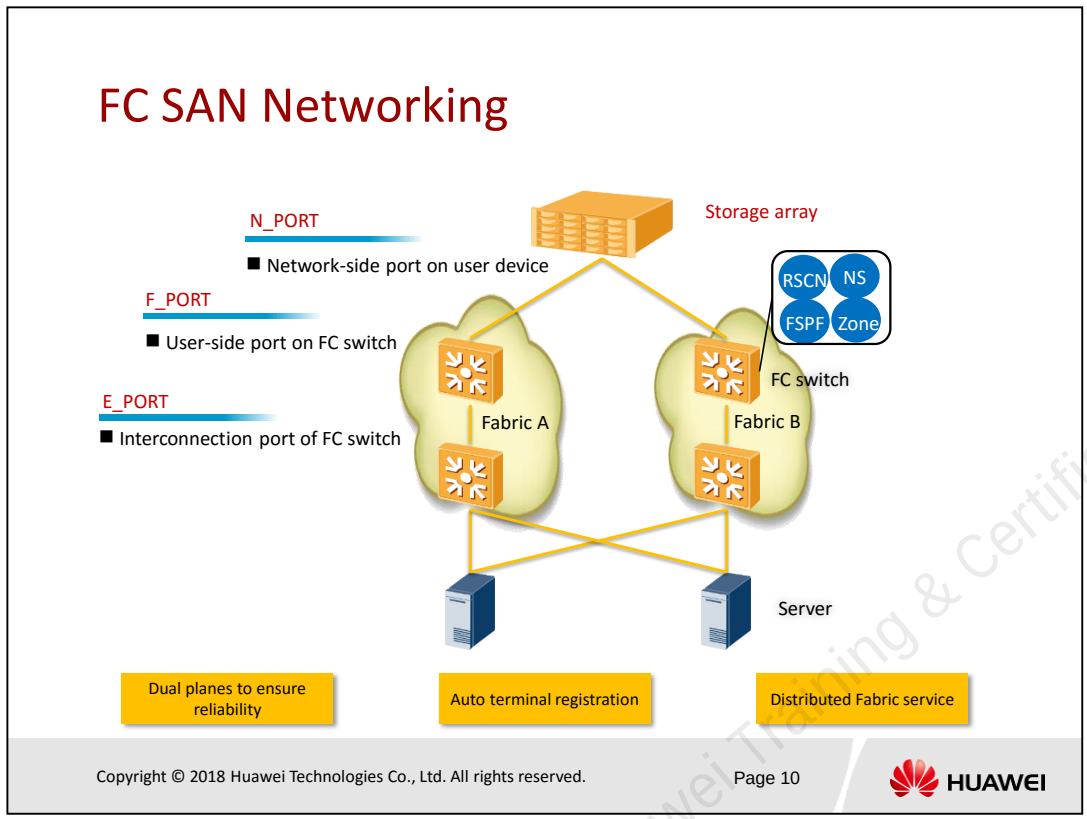
Huawei Training & Certification

Contents

1. Data Center Network Development Trend
2. **FC Network Overview**
3. FCoE Network Overview
4. FCoE Data Forwarding
5. Enhanced Ethernet Technologies
6. Huawei FCoE Network Convergence Solution
7. Basic Configurations of FCoE

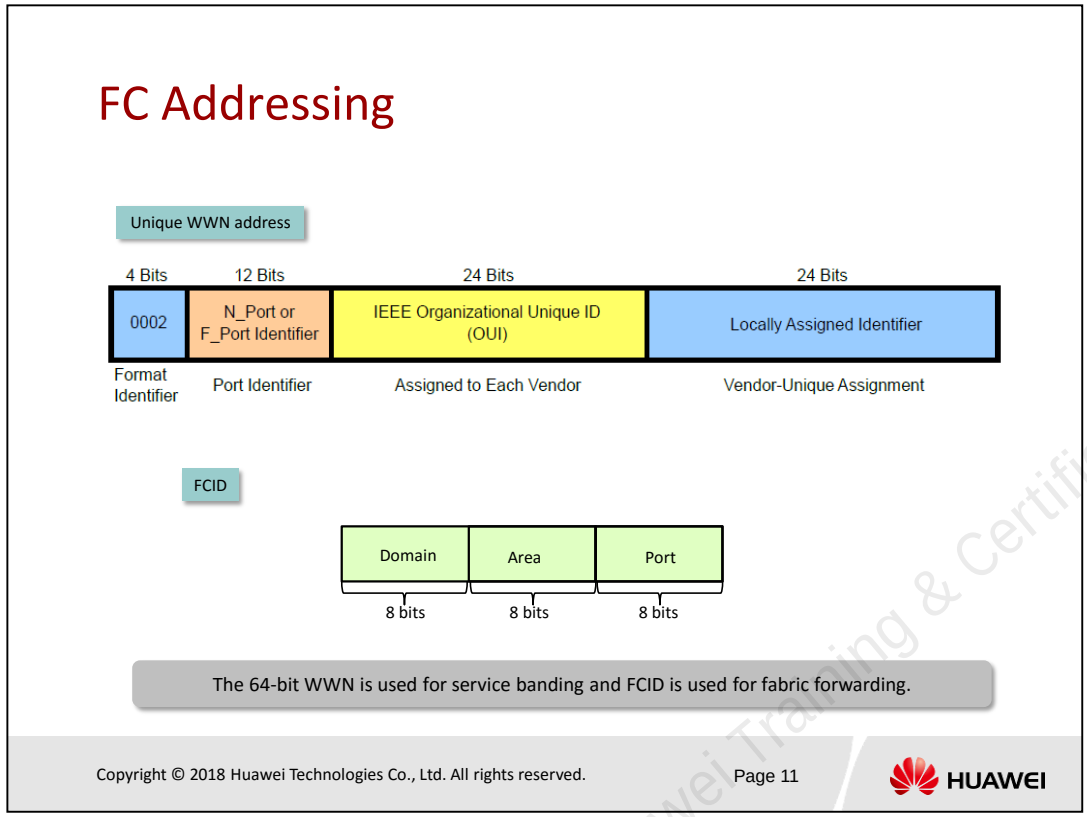
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification



- Port role:
- On the traditional FC network, FC devices are connected through FC interfaces. FC interfaces are classified into N_Ports and F_Ports:
  - N_Port (node port): FC device interface that connects to an FC switch. An FC device can be a server or storage device.
  - F_Port (fabric port): FC switch interface that connects to an FC device and provides fabric access services for the FC device.
  - Dual planes to ensure reliability: A SAN has two independent planes (Fabric networks). Each server and storage device connects to the Fabric networks through two independent links. When any node or link fails, the server software switches traffic to the other link.
  - Auto terminal registration: Switches provide services, as well as registration and state maintenance functions for users (server and storage devices).
  - Distributed Fabric service: Switches need to exchange and maintain user state (RSCN), naming service (NS), routing information (FSPF), and zone information on the local Fabric networks.

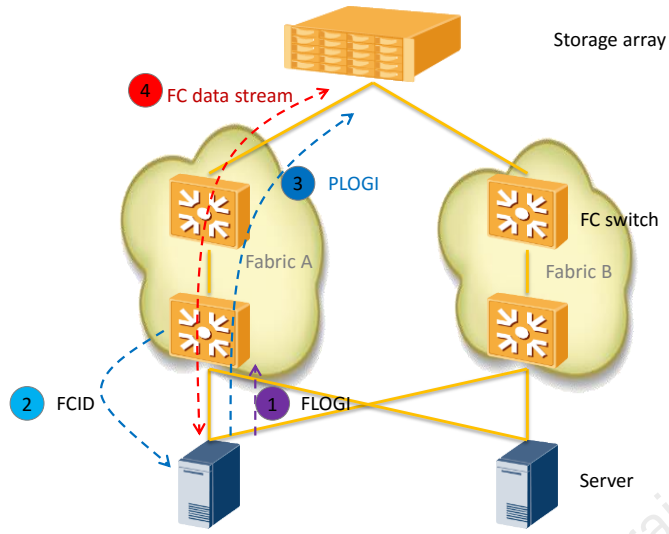
Huawei Training & Certification



- WWN, World Wide Name
  - WWN (similar to a MAC address), is a globally unique 64-bit address defined by a device vendor to identify all devices (FC NICs and switch ports) on an FC network.
- FCID, Fiber Channel Identifier
  - FCID (similar to an IP address), is a 24-bit address used for data forwarding on an FC network. Each FCID identifies a user node and is dynamically allocated by an FC switch when a user registers with the FC switch.
  - Fields of FCID: The FCID is 24 bits and consists of the Domain, Area, Port fields. The Domain field identifies an FC switch and accommodates a maximum of 236 domain IDs (similar to masks used on an IP networks).

Huawei Training & Certification
Huawei Training & Certification

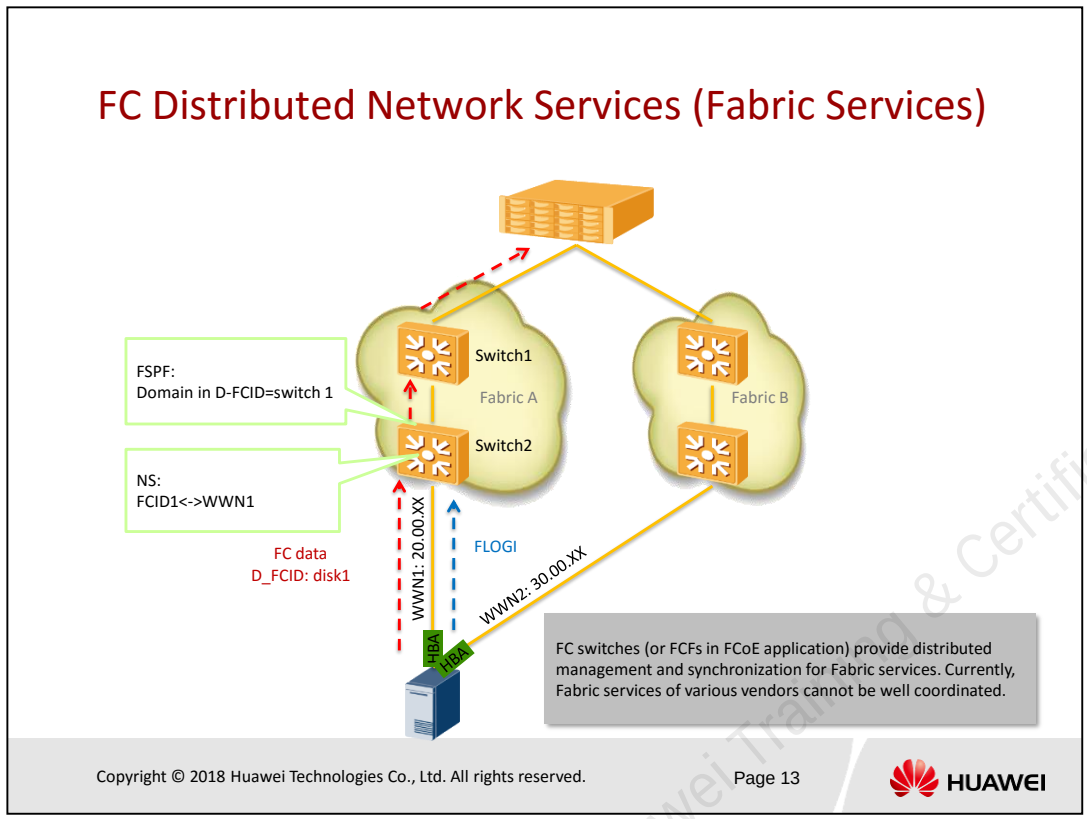
FC Terminal Registration



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Fabric Login (FLOGI)
  - Through FLOGI, a user requests service from an FC switch and obtains an FCID when the user connects to a Fabric network.
- Port Login (PLOGI)
  - Through PLOGI, a user sends a session request to the destination node. The user can access the destination node after a session is set up.
- FC Communication
  - After a server and a storage array set up a session through PLOGI, they can exchange FC data. FC switches on the Fabric network search for transmission paths for FC data based on FCIDs of the source and destination nodes.

Huawei Training & Certification



- NS
  - Name Service: a service that an FC switch provides to enable users to query information such as FCID-WWN mapping.
- FSPF
  - Fabric Shortest Path First (FSPF): similar to OSPF. It is a protocol that forwards FC data based on the domain ID in the destination FCID.
- RSCN
  - Registered State Change Notification: When a user's state changes, the FC switch connected to the user sends an RSCN message to notify the FC network of the change.
- FC switches (or FCFs in FCoE application) provide distributed management and synchronization for Fabric services. Currently, Fabric services of various vendors cannot be well coordinated.

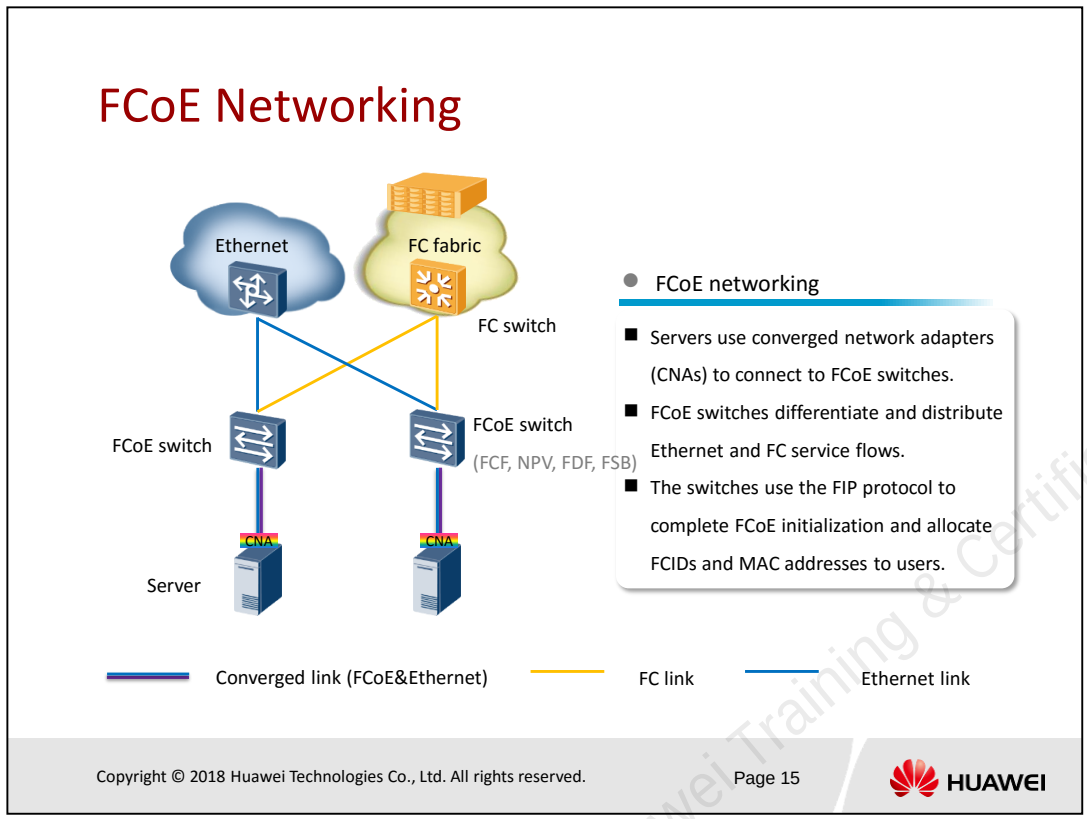
Huawei Training & Certification

Contents

1. Data Center Network Development Trend
2. FC Network Overview
3. **FCoE Network Overview**
4. FCoE Data Forwarding
5. Enhanced Ethernet Technologies
6. Huawei FCoE Network Convergence Solution
7. Basic Configurations of FCoE

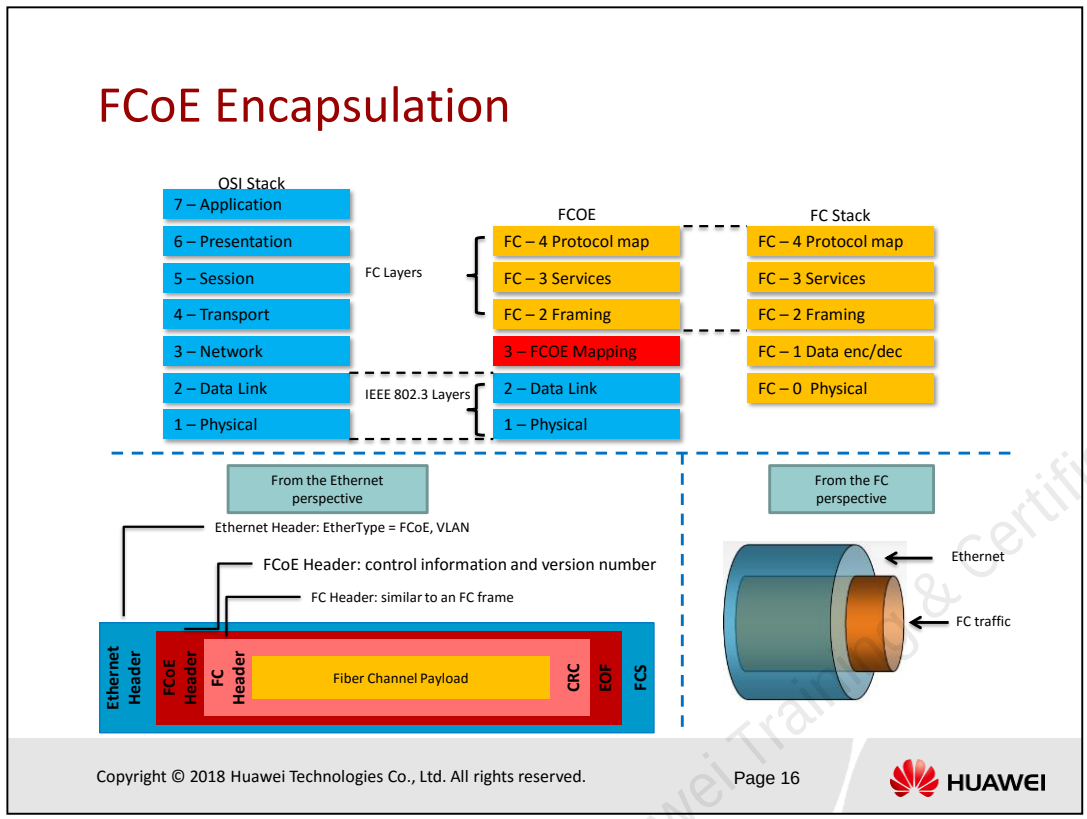
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification



- A LAN is unlimited and uses the BE model. The LAN does not ensure transmission reliability and uses the BE model in the forwarding plane, so packet loss often occurs. Hosts do not predefine the client and server. Any client can be a client or server.
- On the SAN, traffic is reliably transmitted in a deterministic direction. The FC network uses flow control based on BB_Credit to ensure lossless frame transmission. The client and server are predefined.
- FC over Ethernet (FCoE) encapsulates FC frames in Ethernet frames. After the LAN and SAN are converged, the SAN and Ethernet LAN share the same integrated network infrastructure.
- FIP: FCoE initialization protocol, allocates FCIDs, MAC addresses, and VLAN IDs to online users.
- FCoE switch classification: FCoE switches are classified into FCF, NPV, FDF, and FSB switches based on functions they provide. FCFs are core components on an FCoE network and provide key functions such as FTP and Fabric service.
- The requirements for the converged enhanced Ethernet (CEE) using FCoE are as follows:
  - 10Gpbs Ethernet
  - Lossless Ethernet
  - Ethernet Jumbo frames

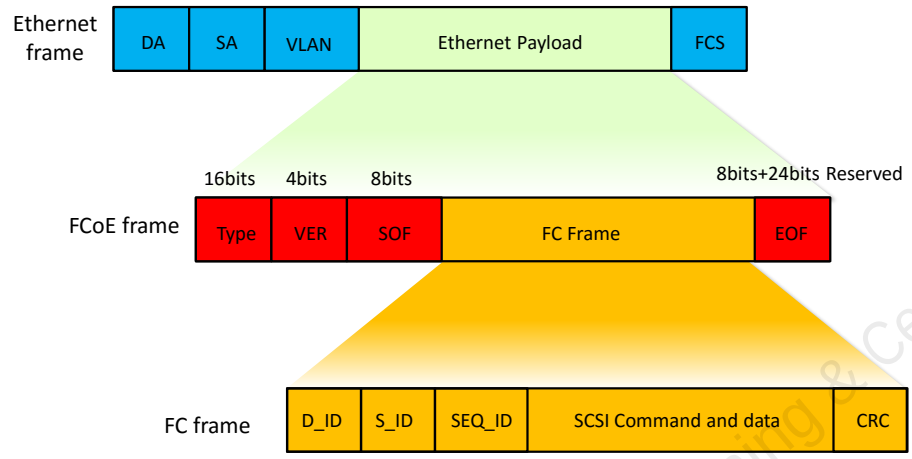
Huawei Training & Certification



- FC4: defines mapping from the upper-layer protocol to the FC protocol.
- FC3: defines common services.
- FC2: defines the frame protocol and flow control mechanism.
- FC1: defines frame encoding.
- FC0: defines the medium type.
- From the Ethernet perspective, FCoE is just another upper layer protocol to carry Ethernet frames.
- From the FC perspective, FCoE is a different way of transmitting FC traffic.

Huawei Training & Certification

FCoE Frame Format

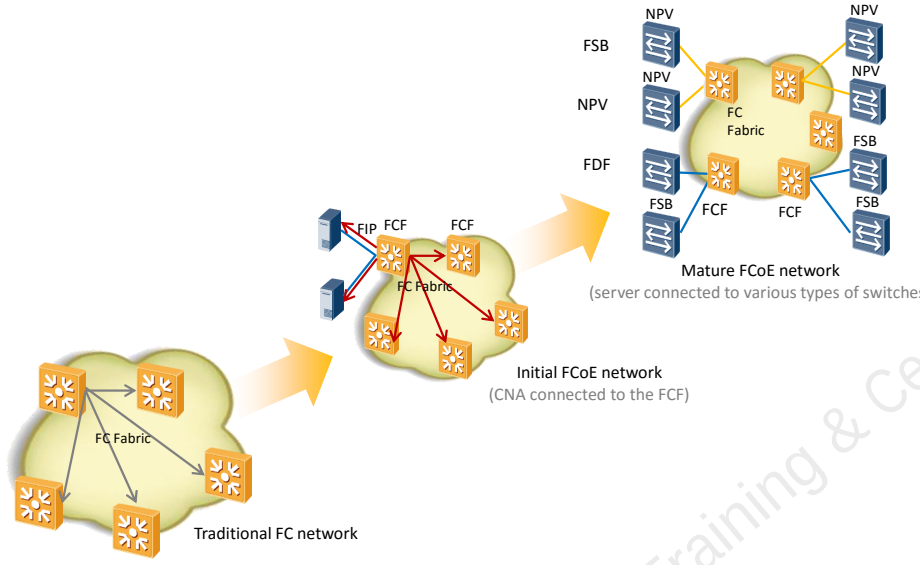


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- DA:destination MAC address, indicating the MAC address of the next hop FCF in unicast forwarding, or reserved MAC address in multicast forwarding.
- SA: source MAC address, indicating the MAC address of each hop of FCF or a terminal.
- VLAN: FIP-specified VLAN or data VLAN for FCoE packets.
- Type: Ethernet type. The value 8906h indicates FCoE packets.
- FCS: Ethernet frame check sequence.
- VER: version number
- SOF: start of frame
- EOF: end of frame
- D_ID: destination FCID
- S_ID: source FCID
- SEQ_ID: sequence ID

Huawei Training & Certification

FCoE Network Evolution

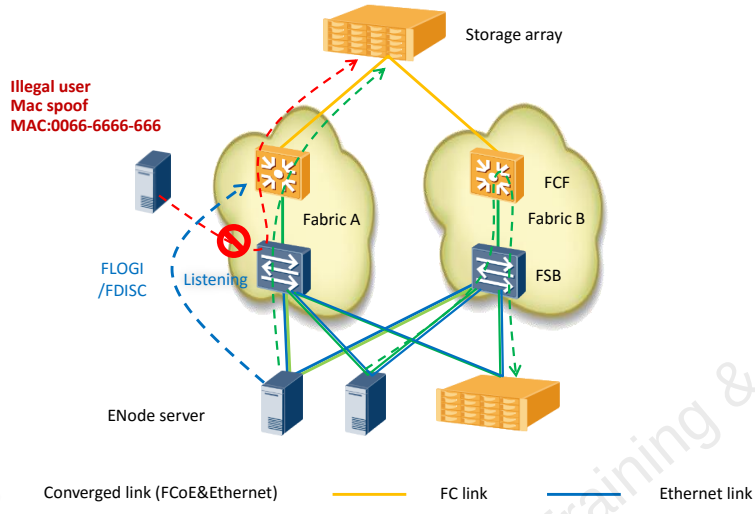


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Scale and performance limitations of a traditional FC network
  - Domain ID limitation: Each FC switch uses a domain ID. Theoretically, an FC SAN network cannot have more than 236 switching nodes.
  - Network performance degrading caused by Fabric data synchronization: FC switches need to synchronize user information, and routing information. More information needs to be transmitted when more switching nodes exist on the network. This degrades performance of the entire network.
- Characteristics of FCFs
  - Provide all FCoE functions: FCFs use the FIP protocol to complete FCoE initialization and allocate FCIDs and MAC addresses to users.
  - Inherit disadvantages of FC switches: FCFs provide Fabric service similar to that provided by FC switches and respond to registration requests from users. They can communicate with FC switches. However, the Fabric service and registration response functions occupy domain IDs, and require fabric data synchronization.
- Various FCoE access technologies, expanding network
  - Expanding network scale: various access switches are added to Fabric network perimeter.
  - NPV: provides FC access ports but does not provide Fabric service or occupy domain ID.
  - FSB: provides Ethernet access ports but does not provide Fabric service or occupy domain ID.

Huawei Training & Certification

FCoE Switch: FSB

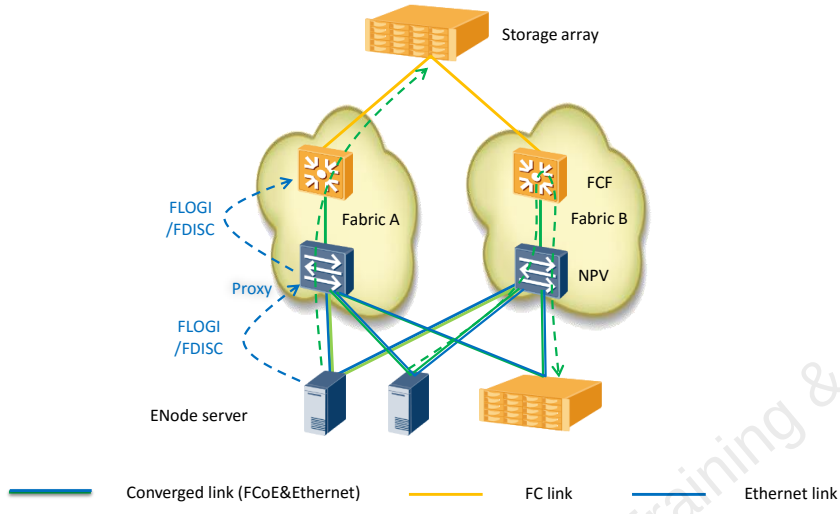


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- FSB functions:
  - FSB: FIP snooping bridge
  - Control plane: does not participate FIP protocol control, but listens on FIP messages and controls link access based on FIP messages to enhance security.
  - Forwarding plane: forwards packets based on MAC addresses.
  - Function: provides 10GE FCoE access and Ethernet transparent transmission.
- FSB advantage and disadvantage:
  - Advantage: easy to deploy and high-efficiency MAC-based forwarding
  - Disadvantage: cannot be deployed independently. An FSB supports only FCoE ports and is not compatible with FC networks. FC traffic must be transmitted by FCFs (circuitous forwarding paths).

Huawei Training & Certification

FCoE Switch: NPV

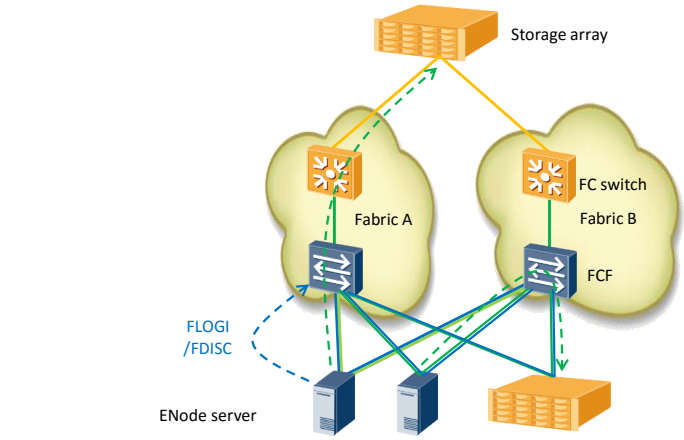


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- NPV functions:
  - NPV: N_Port Virtualization
  - Control plane: participates in part of FIP protocol control, and functions as a proxy agent for servers to send FIP requests.
  - Forwarding plane: forwards packets based on FCIDs.
  - Function: provides FCoE access and supports FC ports to connect to traditional FC networks.
- NPV advantage and disadvantage:
  - Advantage: compatible with FC networks.
  - Disadvantage: cannot be deployed independently. FC traffic must be transmitted by FCFs (circuitous forwarding paths).

Huawei Training & Certification

FCoE Switch: FCF



Converged link (FCoE&Ethernet) FC link Ethernet link

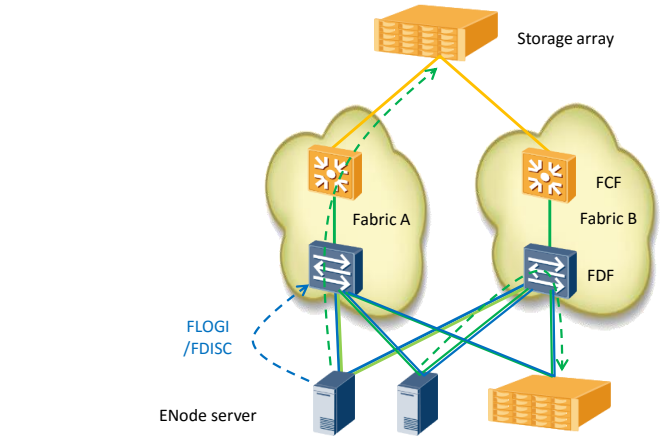
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- FCF functions:
  - FCF: FC forwarder
  - Control plane: provides FIP registration service, and provides Fabric networks with the naming service (NS), Registered State Change Notification (RSCN), and FC Shortest Path First (FSPF) services.
  - Forwarding plane: forwards packets based on FCIDs and supports local forwarding.
  - Function: provides FCoE initialization function (FIP) and Fabric service.
- FCF advantage and disadvantage:
  - Advantage: compatible with FC networks and can be deployed independently.
  - Disadvantage: The domain ID specification limits the network scale.

Huawei Training & Certification

FCoE Switch: FDF



Converged link (FCoE&Ethernet) FC link Ethernet link

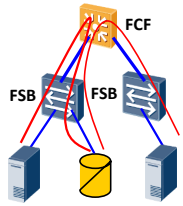
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- FDF functions:
  - FDF: FCoE data forwarder
  - Control plane: participates in part of FIP protocol control but does not provide Fabric service.
  - Forwarding plane: forwards packets based on FCIDs and supports local forwarding.
  - Function: provides FCoE access and FC ports to connect to traditional FC networks, and supports local forwarding.
- FDF advantage and disadvantage:
  - Advantage: compatible with traditional FC networks and supports local forwarding to avoid circuitous forwarding paths.
  - Disadvantage: cannot be deployed independently and must work with FCFs.
- FDF is defined in FC-BB-6 and has not been standardized.

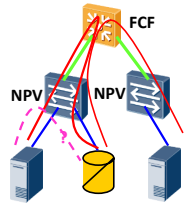
Huawei Training & Certification

FCoE Switch Models and Comparisons

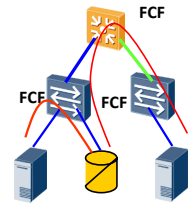
FCoE Switch Model	Whether to Support FC	Forwarding Mechanism	Optimal Path Forwarding	FC Distributed Fabric Service	Whether to Use Domain IDs	Extensibility	Complexity
FSB	No	MAC	No	No	No	Good	Low
NPV	Yes	FCID	No	No	No	Good	medium
FCF	Yes	FCID	Yes	Yes	Exclusive	Low	High
FDF	Yes	FCID	Yes	No	Shared	Good	Medium



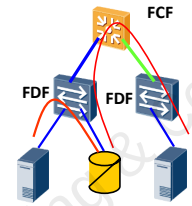
FCF+FSB networking
No path optimization



FCF+NPV networking
No path optimization



FCF networking
Path optimization



FCF+FDF networking
Path optimization

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

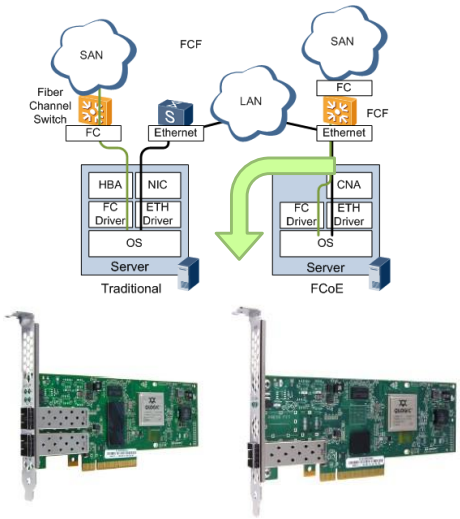
Contents

1. Data Center Network Development Trend
2. FC Network Overview
3. FCoE Network Overview
4. **FCoE Data Forwarding**
5. Enhanced Ethernet Technologies
6. Huawei FCoE Network Convergence Solution
7. Basic Configurations of FCoE

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Mandatory Entity for FCoE Data Transmission - CNA



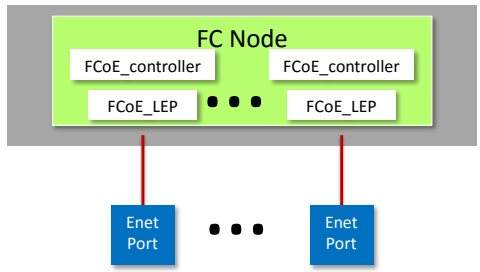
- Introduction to the CNA
  - CNA: converged network adapter
  - The CNA is a 10GE network adapter and supports the FC, Ethernet, and FCoE protocol stack. The CNA needs to support lossless network protocols such as PFC and ETS.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The CNA has one port or two ports, for example, Brocade 1010 (single port)/Brocade 1020 (dual port). It helps reduce the number of adapters on the converged network, reduce power consumption, and increase available space in data centers.

Huawei Training & Certification

Mandatory Entity for FCoE Data Transmission - ENode



ENode model

- Introduction to ENode

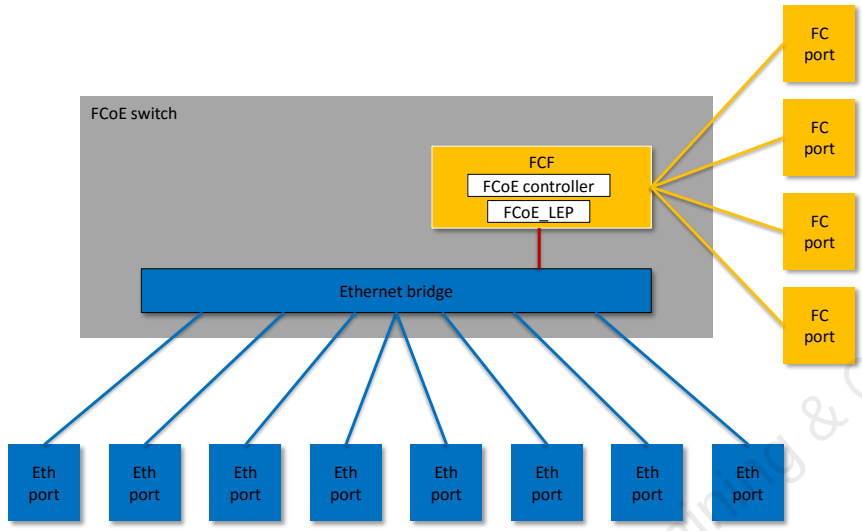
- ENode: FCoE node
- ENode: is a CNA that supports FCoE and FC.
- FCoE EndPoint (FCoE LEP): is a data forwarding component that encapsulates and decapsulates FC frames.
- FCoE controller: performs FIP functions.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

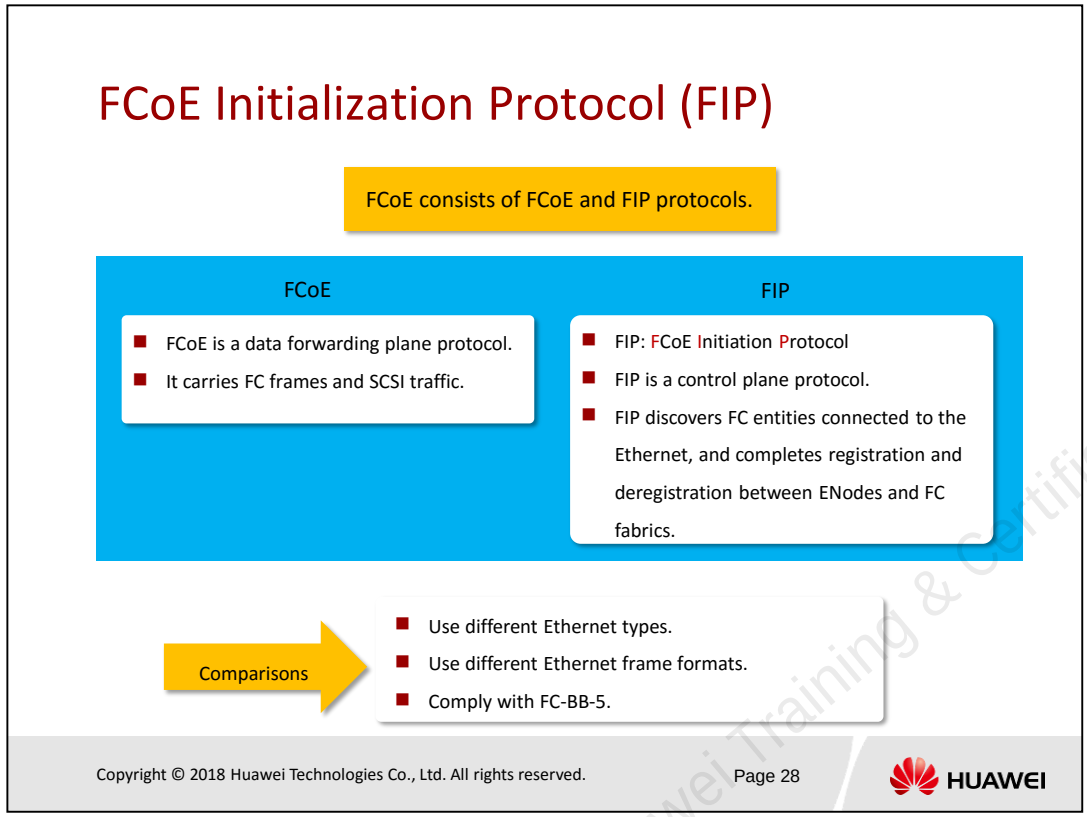
Mandatory Entity for FCoE Data Transmission - FCF



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- An FCoE forwarder (FCF) is a switch supporting both FCoE and FC, and is used to connect the SAN and LAN.
- An FCF can forward FCoE packets and encapsulate and decapsulate FCoE frames.

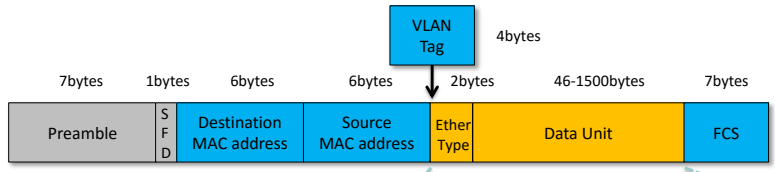
Huawei Training & Certification



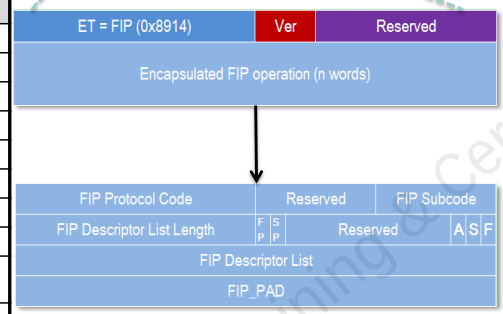
- The FCoE Initialization Protocol (FIP) is a Layer 2 protocol that discovers FC terminals on an FCoE network, implements fabric login, and establishes FCoE virtual links. An ENode can log in to the fabric using FIP to communicate with the target FC device. FIP can also maintain FCoE virtual links.

Huawei Training & Certification

FIP Frame Format

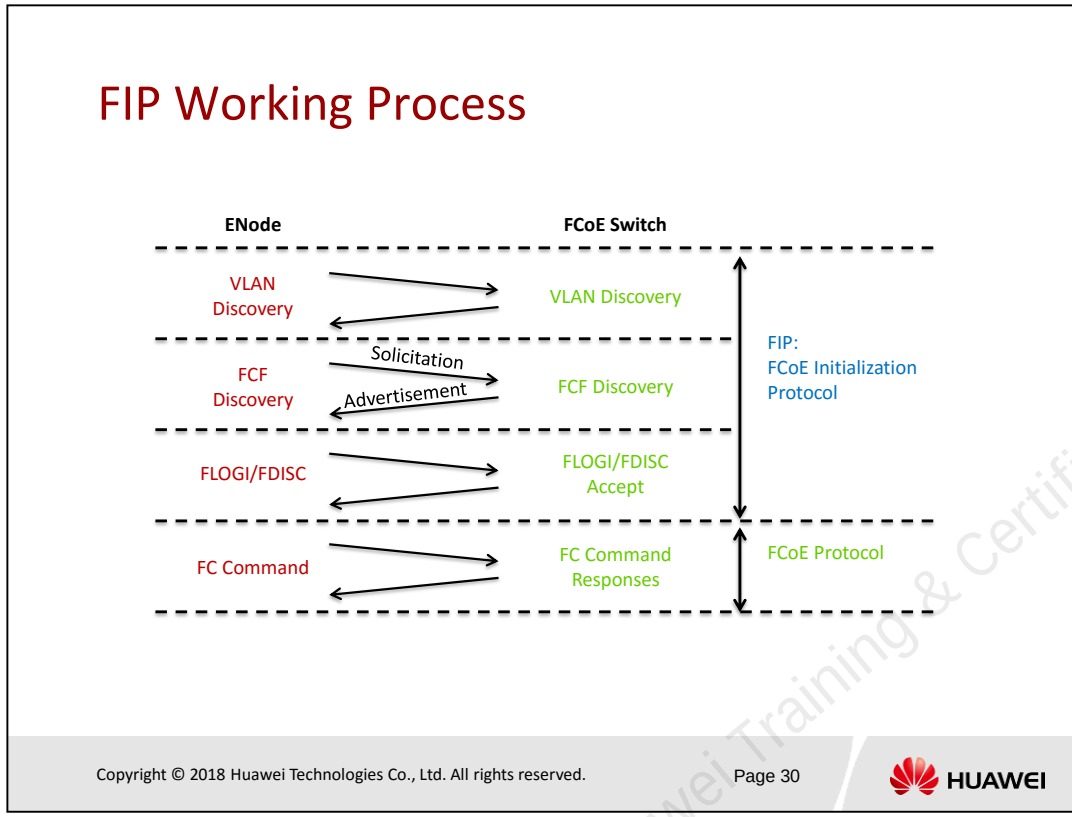


FIP Protocol Code	FIP Subcode	FIP Operation
0001h	01h	Discovery Solicitation
	02h	Discovery Advertisement
0002h	01h	Virtual Link Instantiation Request
	02h	Virtual Link Instantiation Reply
0003h	01h	FIP Keep Alive
	02h	FIP Clear Virtual Links
0004h	01h	FIP VLAN Request
	02h	FIP VLAN Notification
FFF8h to FFFEh	00h to FFh	Vendor Specific
All others	All others	Reserved



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

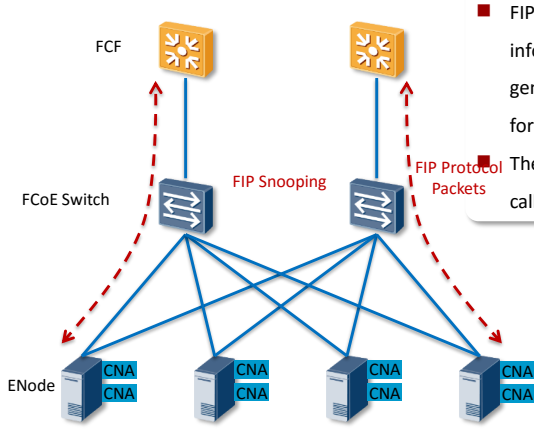
Huawei Training & Certification



- FIP is a control plane protocol and provides the following functions:
  - FIP VLAN discovery: FIP VLAN discovery discovers VLANs used by FCoE nodes such as servers or storage arrays to transmit FCoE traffic. For example, a server obtains the VLAN ID for transmitting FCoE frames through FIP, and carries the VLAN ID in the VLAN tag of FCoE frames sent by the server CAN.
  - FCF discovery: An FCoE node obtains information about all FCFs through FIP and determines to establish a virtual link with an FCF.
  - Virtual link establishment: An FCoE node sends FIP FLOGI or FIP FDISC frames to the selected FCF to establish a virtual link, and the FCF allocates an FC address to the FCoE node.
  - Virtual link maintenance: After a virtual link is established, the FCF periodically sends advertisement frames to the FCoE node, and the FCoE node periodically sends keepalive frames to the FCF to maintain the virtual link.

Huawei Training & Certification

FIP Snooping



- FIP snooping working principle
  - Applies to Enode and FCF modes.
  - FIP snooping obtains FCoE virtual link information by listening to FIP packets and generates ACL rules to control FCoE traffic forwarding.
  - The FCoE switch running FIP snooping is called an FIP snooping bridge (FSB).

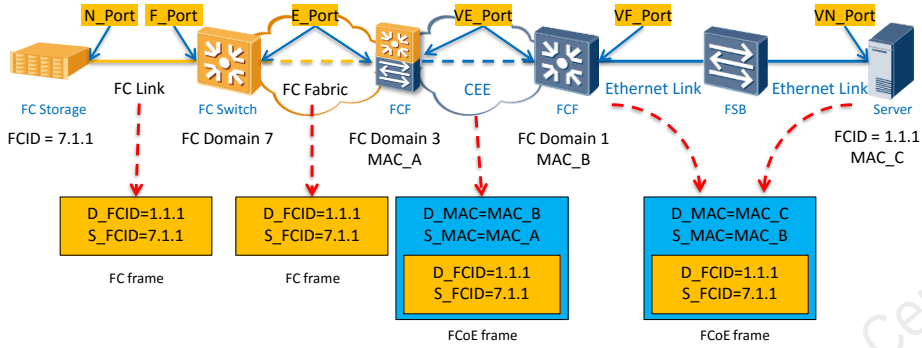
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- On an FC network, an FC switch is considered a trusted device. Other FC devices such as ENodes must log in to the FC switch before they can connect to the FC network. The FC switch then assigns addresses to the FC devices. FC links are point-to-point, and an FC switch can completely control traffic received and sent by FC devices. Therefore, an FC switch ensures that devices use the assigned addresses to exchange packets and protect FC devices against malicious attacks.
- When an FCoE switch is deployed between an ENode and an FCF, FCoE frames are forwarded on the FCoE switch based on the Ethernet protocol because the FCoE switch does not support the FC protocol. In this case, FCoE frames may not be destined for the FCF, and the point-to-point connection between the ENode and FCF is terminated.
- To achieve equivalent robustness as an FC network, the FCoE switch must forward FCoE traffic from all ENodes to the FCF. FIP snooping obtains FCoE virtual link information by listening to FIP packets and generates ACL rules to control FCoE traffic forwarding.
- The FCoE switch running FIP snooping is called an FIP snooping bridge (FSB).

Huawei Training & Certification

FCoE Data Forwarding Process



The FC address remains unchanged during end-to-end transmission, but the Ethernet MAC address in frames is replaced per hop.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- An FC port (N_Port) in the storage array sends an FC frame with destination address (D_ID) 1.1.1 and source address (S_ID) 7.1.1.
- After the FC switch with the domain ID of 7 receives the FC frame, it searches for the forwarding table and sends the frame out through the port with the shortest path to destination address 1.1.1.
- After the FC switch with the domain ID of 3 receives the FC frame, it searches for the forwarding table and detects that the frame needs to be transmitted through an Ethernet interface. Then the FC switch encapsulates the FC frame into an FCOE frame, that is, adds an FCoE header and an Ethernet header to the FC frame header. The destination MAC address in the Ethernet header of the FCoE frame is B (MAC address of the receiver FCoE switch), and the source address is A (MAC address of the sender FCoE switch).
- After the FCoE frame reaches the FCoE switch with the MAC address of B, the FCoE switch detects that the FC destination address (D_ID) in the FC frame contained in the FCoE frame is in the domain range. The FCoE switch then encapsulates the new destination MAC address C (server MAC address) and new source MAC address B in the FCoE frame and sends the FCoE frame out through an interface.
- The server (the CNA MAC address is C) receives the FCoE frame with FC source address (S_ID) 7.1.1 and FC destination address (D_ID) 1.1.1.

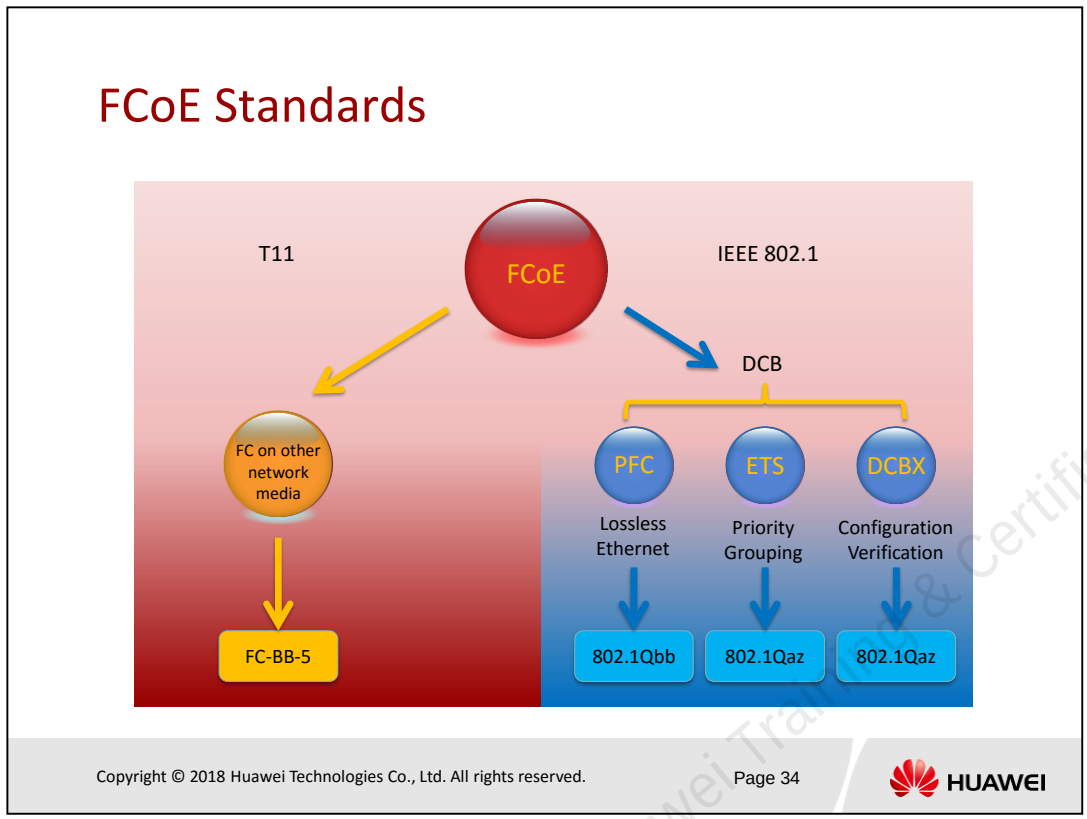
Huawei Training & Certification

Contents

1. Data Center Network Development Trend
2. FC Network Overview
3. FCoE Network Overview
4. FCoE Data Forwarding
5. **Enhanced Ethernet Technologies**
6. Huawei FCoE Network Convergence Solution
7. Basic Configurations of FCoE

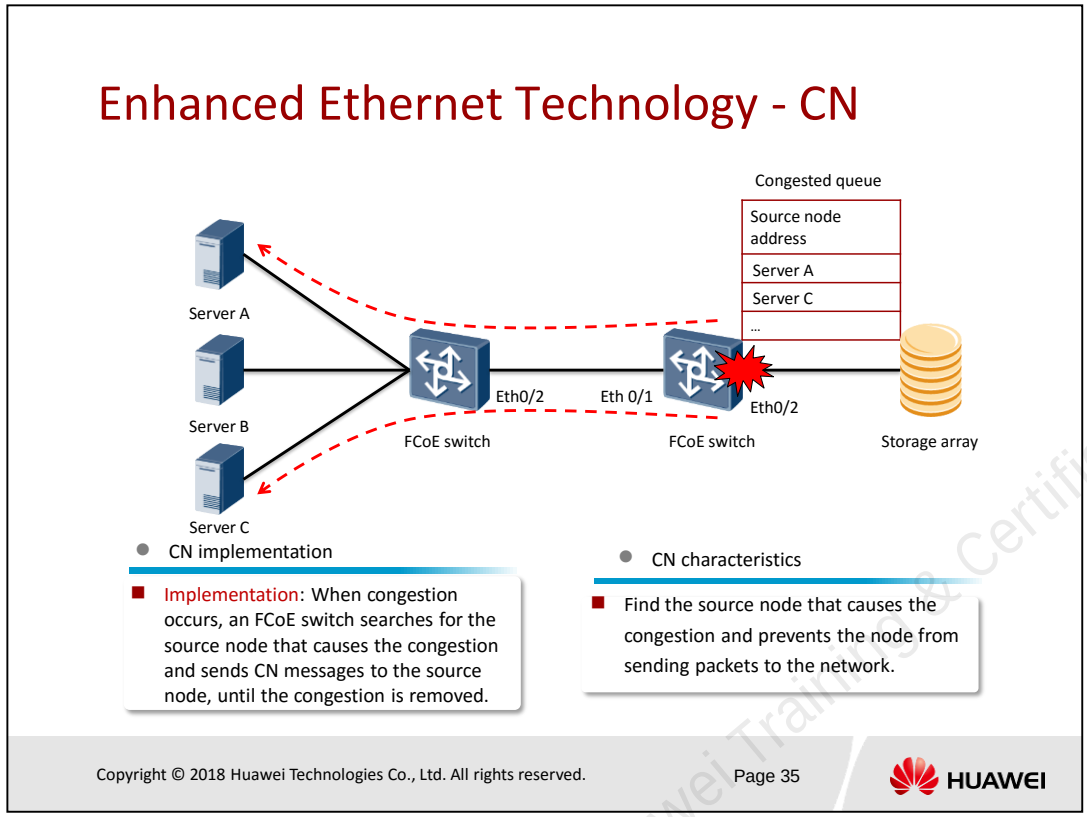
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification



- Traditional Ethernet
  - Congestion control: uses retransmission and sliding window mechanisms of TCP to perform flow control based on destinations.
  - Pause mechanism: Traditional Ethernet switches send Pause frames to request downstream devices to stop sending frames.
- Traditional FC network
  - Congestion control: uses buffer-to-buffer credit (BB_Credit) and end-to-end credit (EE_Credit) mechanisms. BB_Credit and EE_Credit are similar to the token bucket mechanism and perform flow control based on bandwidth capabilities between nodes.
- Enhanced Ethernet
  - Congestion control: combines the Ethernet Pause mechanism and FC flow control to define the multi-level congestion control mechanism based on PFC and ETS, and CN-based end-to-end congestion control mechanism.
- Storage services require lossless transmission. Traditional FC technology has a Credit mechanism to ensure lossless transmission, but Traditional Ethernet only implements "best effort" transmission. DCB technology is introduced to Ethernet to ensure QoS of FCoE service.
- Based on FC-BB-5, FCoE integrates DCB and transmits FC traffic over the enhanced Ethernet.
- Data Center Bridging (DCB) is a set of extensions to Ethernet for use in a data center environment, which is defined by the IEEE 802.1 working group. DCB is used to build lossless Ethernet, meeting QoS requirements after data center networks are converged.

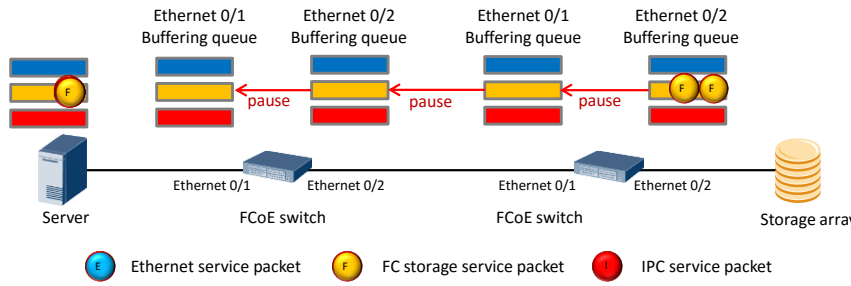
Huawei Training & Certification



- CN: Congestion Notification, 802.1Qau
- When congestion occurs on a network, the congestion point sends an instruction to the data source to limit the traffic that causes congestion and notifies the data source of canceling the limit after congestion is eliminated. This implementation limits congestion at the network edge and reduces the impact of congestion on other network devices. CN provides end-to-end management for continuously congested traffic on a Layer 2 network. For example, when a server CNA sends FCoE frames, many blocks of data flows are continuously transmitted. When congestion occurs on a switch along the transmission path, the switch sends a frame on the Layer 2 network, notifying the server CNA of limiting the traffic rate. Network congestion is therefore lessened.

Huawei Training & Certification

Enhanced Ethernet Technology - PFC

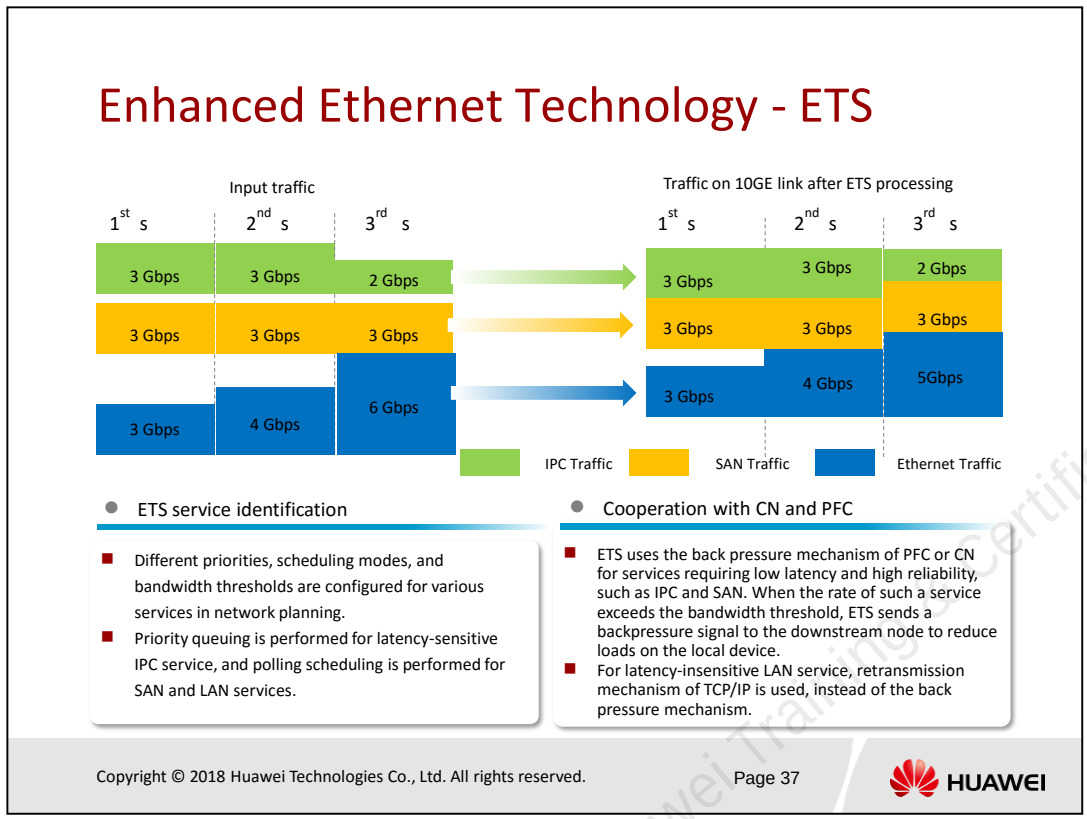


- PFC service identification
  - Different priorities (0 to 7) and queue thresholds are configured for various services in network planning.
  - Switches perform congestion control based on service priorities. When service traffic of a priority exceeds the threshold, excess packets of this service are dropped. Packets of other services are forwarded normally.
- Multi-level back pressure
  - Ethernet Pause mechanism: When congestion occurs, switches send backpressure signals to downstream devices, requesting downstream devices to stop sending service packets.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- PFC: Priority Flow Control, 802.1Qbb
- FCoE requires link sharing. On a link transmitting FCoE frames, burst traffic of one type cannot affect traffic of other types, and enqueuing of traffic of one type cannot preempt buffer resources of traffic of other types. The Ethernet Pause mechanism ensures the lossless transmission service, but can stop all traffic on a link. PFC enhances the Pause mechanism. PFC defines eight queues on an Ethernet link and allows the device to send backpressure signals to pause or restart any queue. This mechanism allows a network to create lossless service for a single queue so that traffic of the lossless service and other types of traffic coexists on the same interface. FCoE is an application of PFC. Switches supporting FCoE differentiate different types of traffic through CoS values in Ethernet frames. You can apply PFC to FCoE traffic with the specified CoS value, and use the BE mode for traffic of other types.

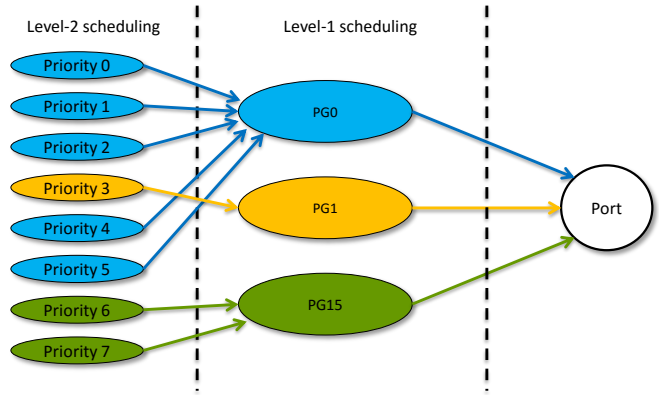
Huawei Training & Certification



- ETS: Enhanced Transmission Selection, 802.1Qaz
- A converged data center network has LAN traffic, SAN traffic, and IPC traffic. The converged network has high QoS requirements. For example, SAN traffic is sensitive to packet loss and relies on in-order deliver. IPC traffic is exchanged between servers and requires low latency. LAN traffic allows packet loss and is delivered on a best-effort (BE) basis. Hosts at both ends process packet loss and mis-sequencing, without network nodes to perform any operations. Traditional QoS cannot meet requirements of the converged network, whereas ETS uses a hierarchical flow control mechanism to implement QoS on the lossless Ethernet.
- ETS can set priorities for different types of traffic, ensure bandwidth, and allow high-priority traffic to use idle bandwidth of high-priority queues, to implement high network bandwidth use efficiency. In the preceding figure, a 10GE link carries server cluster traffic (3 Gbit/s bandwidth), FCoE storage traffic (3 Gbit/s bandwidth), and LAN traffic (BE mode). At t1 and t2, all traffic can be forwarded because the total traffic on the interface is within the interface bandwidth. At t3, the total traffic exceeds the interface bandwidth and LAN traffic exceeds given bandwidth. At this time, LAN traffic is scheduled based on ETS parameters and 1 Gbit/s LAN traffic is discarded.
- ETS also provides traffic shaping based on priority groups. This traffic shaping mechanism limits traffic bursts in a priority group to ensure that traffic in this group is sent out at an even rate.

Huawei Training & Certification

Enhanced Ethernet Technology - ETS Queue Scheduling



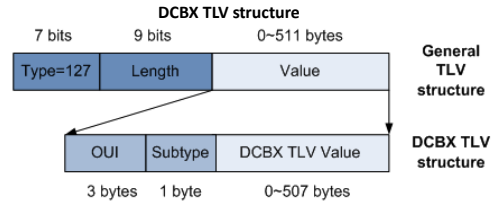
Priority Group No.	Priority Queue	Scheduling Mode	Bandwidth Usage	Traffic
PG0	0, 1, 2, 4, 5	DRR	50%	LAN
PG1	3	DRR	50%	SAN
PG15	6, 7	PQ	-	IPC

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- ETS provides two-level scheduling: scheduling based on the priority group and scheduling based on the priority.
- Compared with common QoS, ETS provides scheduling based on priority groups. ETS adds traffic of the same type to a priority group so that traffic of the same type obtains the same CoS.
- A priority group is a group of priority queues using the same scheduling mode. You can add queues with different priorities to a priority group. Scheduling based on the priority group is called level-1 scheduling.
- ETS defines three priority groups: PG0, PG1, and PG15. PG0, PG1, and PG15 process LAN traffic, SAN traffic, and IPC traffic respectively.
- As defined by ETS, PG0, PG1, and PG15 use PQ+DRR. PG15 uses priority queuing (PQ) to schedule delay-sensitive IPC traffic. PG0 and PG1 use Deficit Round Robin (DRR). Bandwidth can be allocated to priority groups based on actual networking.
- As shown in the networking diagram, the queue with priority 3 carries FCoE traffic, so this queue is added to the SAN group (PG1). Queues with priorities 0, 1, 2, 4, and 5 carry LAN traffic, so these queues are added to the LAN group (PG0). The queue with priority 7 carries IPC traffic, so this queue is added to the IPC group (PG15).

Huawei Training & Certification

Enhanced Ethernet Technology - DCBX



DCBX TLV content

Feature	Description
ETS	Priority group configuration: priority group ID and bandwidth usage of a priority group
	Priority queue configuration: priority queue ID and its priority group ID
PFC	Priority queue ID
	Whether PFC is applied to a queue
Other application protocols	Priority in FCoE traffic

Introduction to DCBX

- DCBX: Data Center Bridging Exchange Protocol
- DCBX, as a link discovery protocol, enables DCB devices at both ends of a link to discover and exchange DCB configurations.
- DCBX encapsulates DCB configurations into Link Layer Discovery Protocol (LLDP) TLVs so that devices at both ends of an FCoE virtual link can exchange DCB configurations.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- DCBX: Data Center Bridging Exchange Protocol, 802.1az
- To implement lossless Ethernet on a converged data center network, both ends of an FCoE link must have the same PFC and ETS parameter settings. DCBX, as a link discovery protocol, enables DCB devices at both ends of a link to discover and exchange DCB configurations, reducing workloads of administrators. DCBX provides the following functions:
  - Discovers the DCB configuration of the remote device.
  - Detects the DCB configuration errors of the remote device.
  - Configures the remote device if permitted.
- DCBX enables DCB devices at both ends to exchange the following DCB configurations:
  - ETS priority group
  - PFC
  - Information about other application protocols, for example, FCoE
- The DCB configuration is encapsulated into specified TLVs. The Type field has a fixed value of 127, the OUI field has a fixed value of 0x0080c2, and the Subtype field has a fixed value of 2.

Huawei Training & Certification

Contents

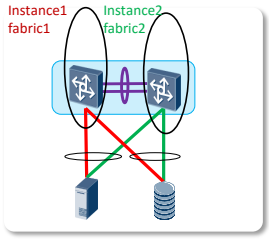
1. Data Center Network Development Trend
2. FC Network Overview
3. FCoE Network Overview
4. FCoE Data Forwarding
5. Enhanced Ethernet Technologies
6. **Huawei FCoE Network Convergence Solution**
7. Basic Configurations of FCoE

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

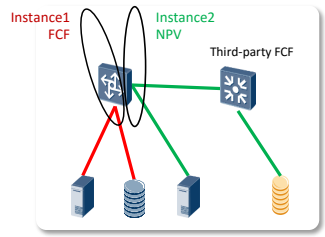
FCoE Instance

Dual planes



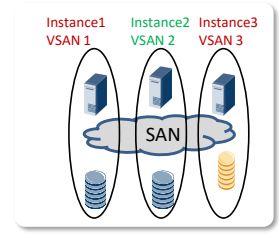
Use two instances to build dual fabrics

Multiple networks



Instances used to differentiate new networks and reconstructed NPV networks

Large-scale networking



Instances used to differentiate the VSANs (future)

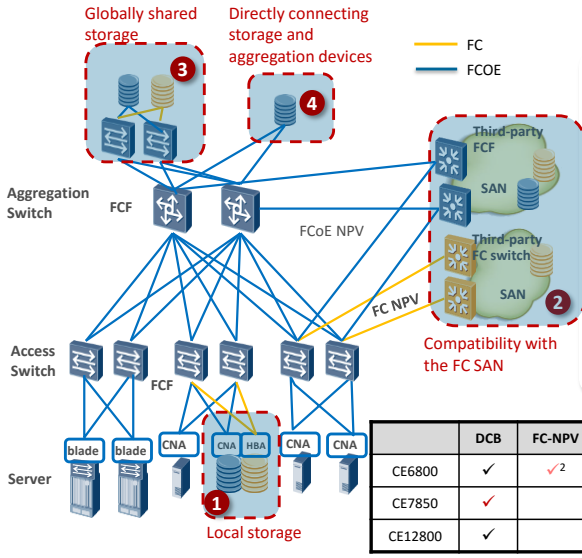
- CE switches supports three FCoE networking: FSB, FCF, and NPV. One mode is selected for each instance.
  - Cisco: The FCF and NPV cannot coexist. When the FCF and NPV are switched, the OS needs to be reset.
  - H3C: The FCF and NPV cannot coexist. When the FCF and NPV are switched, the original configuration needs to be deleted.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Each FCF instance is configured with one FCoE VLAN. Because FC interfaces are not supported, one VLAN corresponds to one VSAN. That is, one instance corresponds to one VSAN.
- The FSB, FCF, and NPV instances can coexist. In actual networking, you are advised to use FCF or NPV instances. You are not advised to use FSB instances and other instances together.

Huawei Training & Certification

Huawei FCoE Evolution



Storage Deployment Scenarios

1. Local storage
   - FC/FCoE storage deployment in a rack
2. Compatibility with the FC SAN
   - Compatibility with the FC SAN through NPV
3. Globally shared storage
   - New FC/FCoE storage devices connect to a network through CE6800s.
4. Directly connecting storage and aggregation devices
   - Directly connecting a small number of FCoE storage devices to aggregation devices, which reduces network layers and improves performance

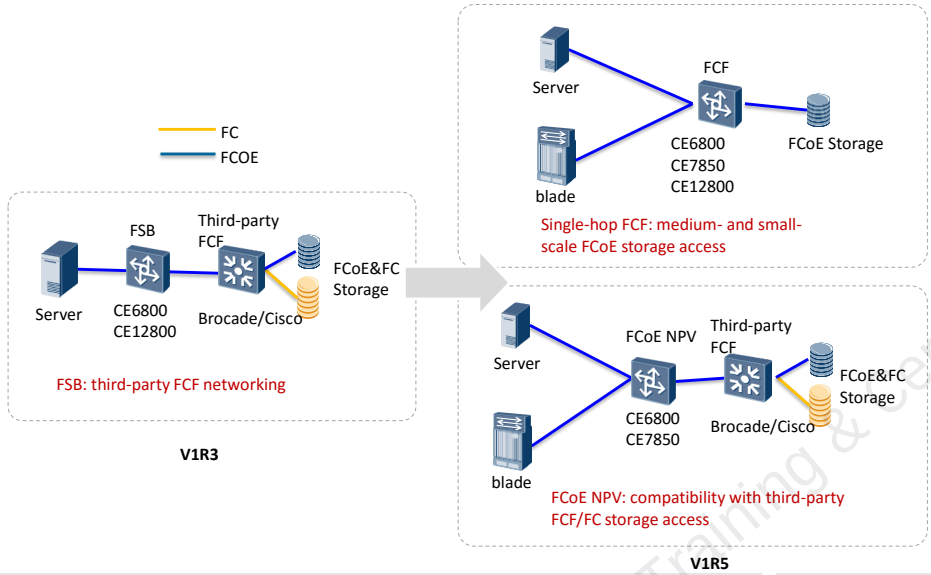
	DCB	FC-NPV	FCoE-NPV	FCF-单跳	FCF-多跳	FC端口
CE6800	✓	✓[2]	✓	✓	✓[3]	✓[2]
CE7850	✓		✓	✓	✓	
CE12800	✓		✓[1]	✓	✓	

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- CE12800 R5C00 does not support NPV.
- R5C10 supports FC interfaces and FC FCF/NPV.
- Multi-hop FCF is not supported.

Huawei Training & Certification

Huawei FCoE Solution (1)

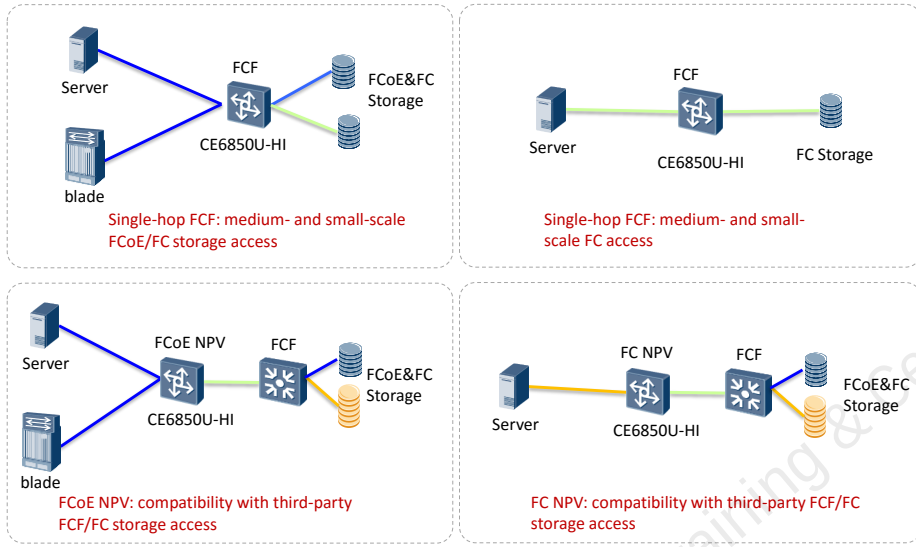


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Huawei FCoE Solution (2)



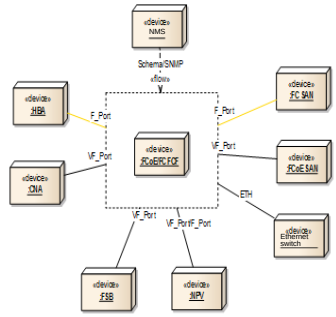
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

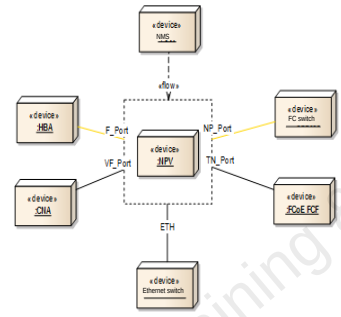
- Huawei FCoE Solution (2) (V1R5C10, 2015 Q2)

Huawei Training & Certification

CE6850U-HI

- 48*10GE + 6*40GE
  - Each 10GE interface can be configured to work in FC mode, and FC interfaces support 2/4/8 G.
  - FC interfaces need to connect to independent optical modules.
- The CE6850U-HI supports the FC FCF and FC NPV.

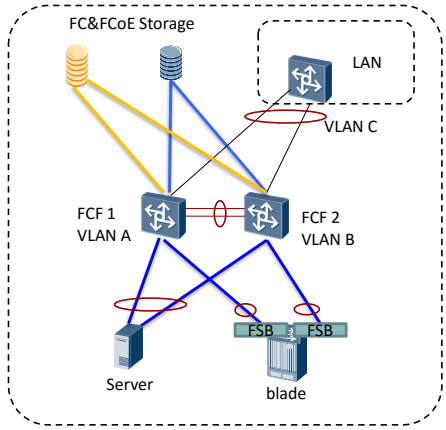




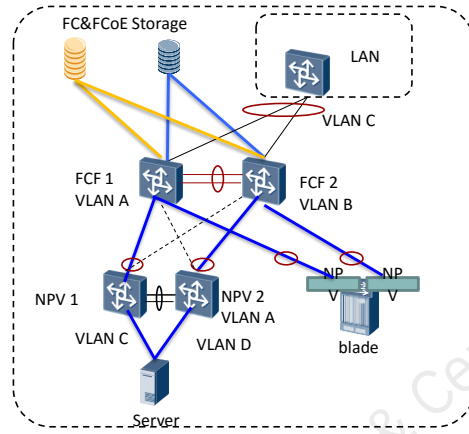
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Dual Fabric Networking



FCF multi-instance implements dual fabrics.
Each FCF instance corresponds to a VLAN, and Ethernet traffic is transmitted in other VLANs.
Each FSB can be single-homed to only one FCF.



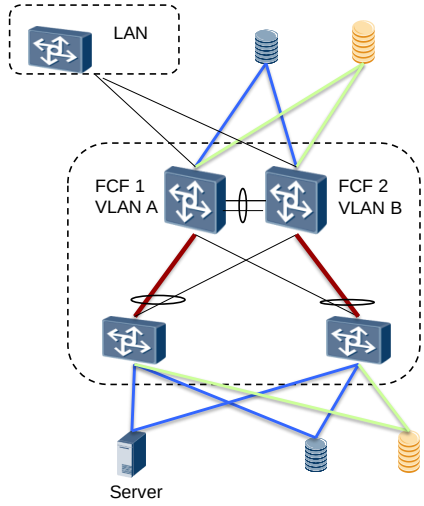
FCF/NPV multi-instance implements dual fabrics.
Each FCF instance corresponds to a VLAN, and Ethernet traffic is transmitted in other VLANs.
Each NPV switch can be single-homed to only one FCF. To implement dual fabric in dual-homing mode, configure Eth-Trunk member interfaces as VNP_Ports.
The NPV VLAN must be different from the FCF VLAN.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

ToR SVF Networking



- R5C00 supports only the FCF/NPV in ToR SVF networking.
- Servers and storage devices connect to leaf switches, or connect to leaf and spine switches respectively.
- FCoE and Ethernet traffic can be forwarded through the fabric link between leaf and spine switches.
- A fabric link can connect to the spine switch through an Eth-Trunk, and packets can be load balanced based on the source or destination FC_ID.
- When a leaf switch is dual-homed to the spine switch through a trunk, you can specify an Eth-Trunk member interface though which FCoE traffic is transmitted (Eth-Trunk load balancing may be affected).

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

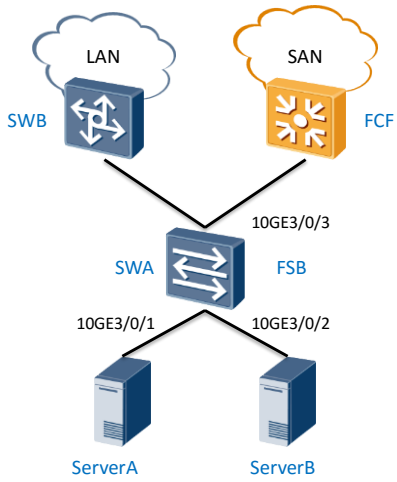
1. Data Center Network Development Trend
2. FC Network Overview
3. FCoE Network Overview
4. FCoE Data Forwarding
5. Enhanced Ethernet Technologies
6. Huawei FCoE Network Convergence Solution
7. **Basic Configurations of FCoE**

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

FSB Instance Configuration (1)



Service requirements:
The FCF is used to aggregate services only. SWA as the access switch directly connects to SWB and FCF through uplink interfaces, and connects to servers through downlink interfaces. SAN traffic needs to be correctly forwarded on SWA.

Step 1: Enable PFC on physical interfaces.

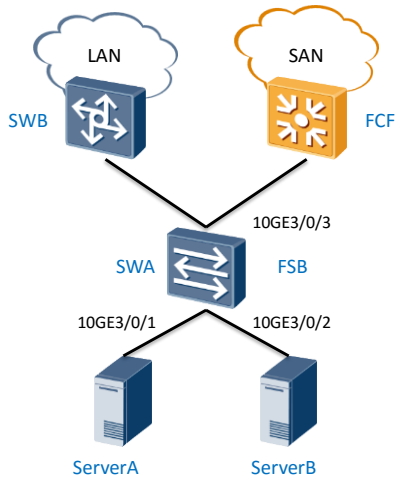
```
[SWA]dcb pfc
[~SWA-pfc]priority 3
[~SWA]interface 10GE 3/0/1
[*SWA-10GE3/0/1]dcb pfc enable mode auto
[*SWA-10GE3/0/1]quit
[*SWA]interface 10GE 3/0/2
[*SWA-10GE3/0/2]dcb pfc enable mode auto
[*SWA-10GE3/0/2]quit
[*SWA]interface 10GE 3/0/3
[*SWA-10GE3/0/3]dcb pfc enable mode auto
[*SWA-10GE3/0/3]quit
[*SWA]commit
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Configuring PFC
  - Run the system-view command to enter the system view.
  - Run the dcb pfc command to enter the PFC view.
  - Run the priority { start-priority [to end-priority] } &<1-8> command to enable PFC. By default, PFC is enabled for queues with priority 3.
  - Run the quit command to exit from the PFC view.
  - Run the interface interface-type interface-number to enter the view of the interface to be enabled with PFC.
  - Run the dcb pfc enable [mode { auto | enable }] command to enable PFC on the interface and specify the PFC working mode.
  - By default, PFC is disabled on an interface.
  - When PFC is configured to work in auto mode, configure DCBX functions.
  - Run the commit command to commit the configuration.

Huawei Training & Certification

FSB Instance Configuration (2)



Step 2: Enable ETS on physical interfaces.

```
[~SWA]dcb ets-profile ets1
[~SWA-ets-ets1]priority-group 0 queue 6
[*SWA-ets-ets1]priority-group 0 drr weight 60
[*SWA-ets-ets1]priority-group 1 drr weight 40
[*SWA-ets-ets1]quit
[*SWA]interface 10GE 3/0/1
[*SWA-10GE3/0/1]dcb ets enable ets1
[*SWA-10GE3/0/1]quit
[*SWA]interface 10GE 3/0/2
[*SWA-10GE3/0/2]dcb ets enable ets1
[*SWA-10GE3/0/2]quit
[*SWA]commit
```

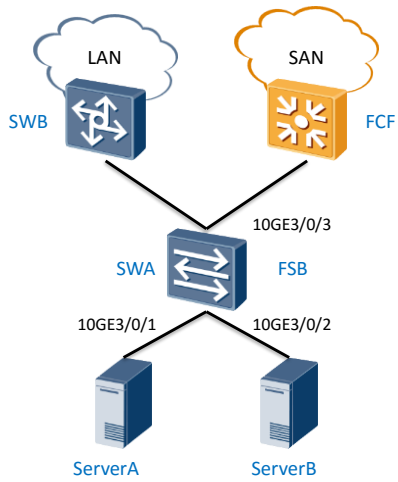
Step 3: Enable DCBX on physical interfaces.

```
[SWA]lldp enable
Info: Global LLDP is enabled successfully.
[~SWA]interface 10GE 3/0/1
[*SWA-10GE3/0/1]lldp tlv-enable dcbx
[*SWA-10GE3/0/1]quit
[*SWA]interface 10GE 3/0/2
[*SWA-10GE3/0/2]lldp tlv-enable dcbx
[*SWA-10GE3/0/2]quit
[*SWA]interface 10GE 3/0/3
[*SWA-10GE3/0/3]lldp tlv-enable dcbx
[*SWA-10GE3/0/3]quit
[*SWA]commit
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Configuring an ETS profile
- Background
- ETS provides two-level scheduling: scheduling based on the priority group and scheduling based on the priority. Priority group-based flow control includes congestion management and traffic shaping based on the priority group.
- Priority-based flow control includes queue congestion management, traffic shaping, and congestion avoidance.
- Procedure
- Run the system-view command to enter the system view.
- Run the dcb ets-profile profile-name command to create an ETS profile and enter the ETS profile view. By default, the system provides ETS profile default, which can be modified but cannot be deleted.
- (Optional) Run the priority group pg-value command to create a priority group.
- By default, an ETS profile defines three priority groups: PG0, PG1, and PG15.
- (Optional) Run the priority-group pg-value queue { start-queue [to end-queue] } &<1-8> command to add an interface queue to a priority group.
- By default, queues 0, 1, 2, 4, and 5 belong to PG0, queue 3 belongs to PG1, and queues 6 and 7 belong to PG15.

FSB Instance Configuration (3)



Step 4: Add interfaces to instances.

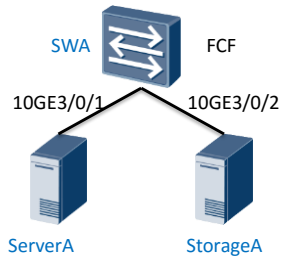
```
[SWA]fcoe fsb1 FSB
[~SWA-fcoe-fsb1-FSB]vlan 2094
[*SWA-fcoe-fsb1-FSB]fip fka-adv-period 6
[*SWA-fcoe-fsb1-FSB]quit
[*SWA]interface 10GE 3/0/3
[*SWA-10GE3/0/3]port link-type trunk
[*SWA-10GE3/0/3]port trunk allow-pass vlan 2094
[*SWA-10GE3/0/3]fcoe role vnp
[*SWA-10GE3/0/3]quit
[*SWA]interface 10GE 3/0/2
[*SWA-10GE3/0/2]port link-type trunk
[*SWA-10GE3/0/2]port trunk allow-pass vlan 2094
[*SWA-10GE3/0/2]quit
[*SWA]interface 10GE 3/0/1
[*SWA-10GE3/0/1]port link-type trunk
[*SWA-10GE3/0/1]port trunk allow-pass vlan 2094
[*SWA-10GE3/0/1]commit
[*SWA-10GE3/0/1]quit
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Configuring an interface role
- Background
- Because traditional Ethernet does not define the traffic transmission direction, roles of interfaces on the device used as the FSB must be defined during network planning. There are two interface roles:
  - VNP_Port
  - ENode_Facing
- Before configuring a role for an interface, add the interface to an FCoE VLAN.
- Procedure
- Run the system-view command to enter the system view.
- Run the interface interface-type interface-number command to enter the view of the Ethernet interface of which the role needs to be configured.
- Run the port link-type { trunk | hybrid } command to set the link type of the interface.
- By default, the link type of an interface is access.
- Run the port trunk allow-pass vlan { { vlan-id1 [to vlan-id2] } &<1-10> | all } or port hybrid tagged vlan { { vlan-id1 [to vlan-id2] }&<1-10> | all } command to add the interface to an FCoE VLAN.
- Run the fcoe role vnp command to set the interface role to VNP_Port.
- By default, an interface role is ENode_facing.
- Run the commit command to commit the configuration.

Huawei Training & Certification

FCF Instance Configuration (1)



Service requirements:
SWA functions as the FCF, and directly connects to ServerA and StorageA. Traffic between ServerA and StorageA needs to be forwarded through FCoE.

Step 1: Configure DCB.

```
The configuration is omitted.
```

Step 2: Configure an instance and bind an FCoE VLAN.VLAN

```
[SWA]fcoe fcf1 FCF
[~SWA-fcoe-fcf1-FCF]vlan 100
//The following configurations are optional,
and there are default settings.
[*SWA-fcoe-fcf1-FCF]domain 2
[*SWA-fcoe-fcf1-FCF]fabric-name
20:00:00:25:9e:fb:9f:01
[*SWA-fcoe-fcf1-FCF]fc-map 0e:fc:00
[*SWA-fcoe-fcf1-FCF]fip fka-adv-period 8
[*SWA-fcoe-fcf1-FCF]fcid persistent
[*SWA-fcoe-fcf1-FCF]commit
```

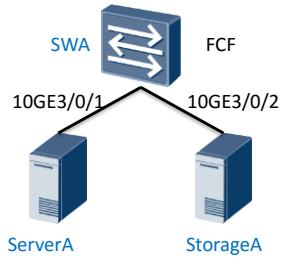
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification

- Creating an FCoE interface and an FC instance
- Background
- On a traditional FC network, all FC nodes are interconnected through a fabric. FC nodes belonging to a fabric have the same attributes. To control FCoE traffic forwarding, FCoE uses FC instances because FCoE transmits SAN traffic over the Ethernet. An FC instance defines fabric attributes such as the FCoE VLAN. A VF_Port is a virtual logical interface that is manually created on an FCF, and provides functions of a physical FC interface. The VF_Port can work properly only when the VF_Port is bound to a physical Ethernet interface.
- Procedure
- Run the system-view command to enter the system view.
- Run the fcoe fc-instance-name fcf command to create an FC instance and enter the FCoE view.
- By default, no FC instance is configured.
- Run the member interface fcoe-port fcoe-port1 [to fcoe-port2] command to add FCoE interfaces to the FC instance.
- By default, no FCoE interface is added to an FC instance.
- Run the vlan vlan-id command to add an FCoE VLAN to the FC instance.

FCF Instance Configuration (2)



Step 3: Add interfaces to instances.

```
[SWA]interface fcoe-port 1
[~SWA-10GE3/0/1]fcoe-port 1
[*SWA-10GE3/0/1]port trunk allow-pass vlan 100
[*SWA-10GE3/0/1]quit
[*SWA]interface fcoe-port 2
[*SWA-10GE3/0/2]fcoe-port 2
[*SWA-10GE3/0/2]port trunk allow-pass vlan 100
[*SWA-10GE3/0/2]quit
[*SWA]commit
```

Step 4: Configure a zone.

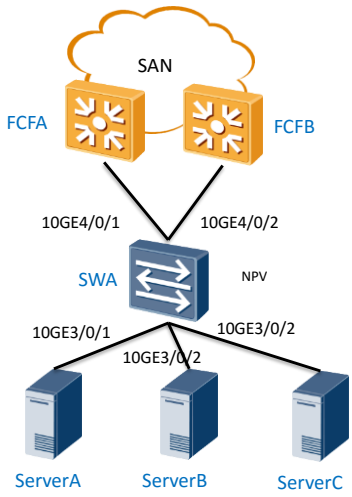
```
[SWA]zone-alias storagea
[~SWA-zone-alias-storagea]member interface
fcoe 2
[*SWA-zone-alias-storagea]quit
[*SWA]zone zone1
[*SWA-zone-zone1]member interface fcoe 1
[*SWA-zone-zone1]member zone-alias storagea
[*SWA-zone-zone1]quit
[*SWA]zoneset name zoneset1
[*SWA-zoneset1]member zone1
[*SWA-zoneset1]quit
[*SWA]active zoneset zoneset1
[*SWA]commit
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Creating an FCoE interface
- Procedure
- Run the system-view command to enter the system view.
- Run the interface fcoe-port interface-number command to create an FCoE interface and enter its view.
- Run the quit command to exit from the system view.
- Run the interface interface-type interface-number command to enter the view of the physical interface that needs to be added to the FCoE interface.
- Run the fcoe-port fcoe-port-id command to add the physical interface to the FCoE interface.
- By default, no physical interface is added to an FCoE interface.
- Run the port link-type { trunk | hybrid } command to set the link type of the interface.
- By default, the link type of an interface is access.

Huawei Training & Certification

NPV Instance Configuration (1)



Service requirements:
SWA functions as the NPV switch. SWA directly connects to ServerA, ServerB, and ServerC, and connects to the San through FCFA and FCFB. FCoE traffic between ServerA and SAN needs t be forwarded through 10GE4/0/1, and traffic of ServerB and ServerC is transmitted according to the load balancing algorithm.

Step 1: Configure DCB.

```
The configuration is omitted.
```

Step 2: Configure an instance and bind an FCoE VLAN.

```
[SWA]fcoe npv1 NPV
[~SWA-fcoe-npv1-NPV]vlan 200
[*SWA-fcoe-npv1-NPV]commit
[~SWA-fcoe-npv1-NPV]quit
```

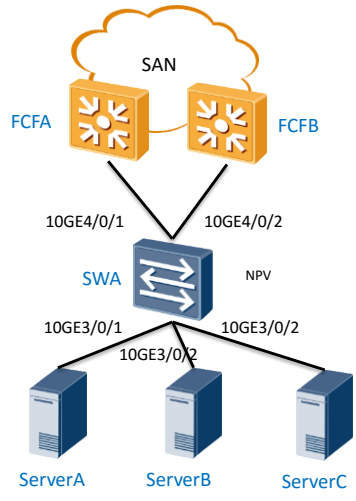
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Creating an FCoE interface and an FC instance
- Background
- On a traditional FC network, all FC nodes are interconnected through a fabric. FC nodes belonging to a fabric have the same attributes. To control FCoE traffic forwarding, FCoE uses FC instances because FCoE transmits SAN traffic over the Ethernet. An FC instance defines fabric attributes such as the FCoE VLAN. The NPV switch uses a VF_Port to connect to a VN_Port of an ENode and uses a VNP_Port to connect to a VF_Port of an FCF.
- Procedure
- Run the system-view command to enter the system view.
- Run the fcoe fc-instance-name npv command to create an FC instance and enter the FCoE view.
- By default, no FC instance is configured.
- An NPV instance corresponds to one FCF instance.
- Run the vlan vlan-id command to add an FCoE VLAN to the FC instance.
- By default, no FCoE VLAN is configured.
- Run the quit command to return to the system view.
- Run the commit command to commit the configuration.

Huawei Training & Certification

NPV Instance Configuration (2)



Step 3: Add interfaces to instances.

```
[~SWA]interface fcoe-port 1
[~SWA-10GE3/0/1]fcoe-port 1
[*SWA-10GE3/0/1]port trunk allow-pass vlan 200
[*SWA-10GE3/0/1]quit
[*SWA]interface fcoe-port 2
[*SWA-10GE3/0/2]fcoe-port 2
[*SWA-10GE3/0/2]port trunk allow-pass vlan 200
[*SWA-10GE3/0/2]quit
[*SWA]interface fcoe-port 3
[*SWA-10GE3/0/3]fcoe-port 3
[*SWA-10GE3/0/3]port trunk allow-pass vlan 200
[*SWA-10GE3/0/3]quit
[*SWA]interface fcoe-port 4
[*SWA-10GE3/0/4]fcoe-port 4
[*SWA-10GE3/0/4]port trunk allow-pass vlan 200
[*SWA-fcoe4]quit
[*SWA]interface fcoe 5
[*SWA-10GE4/0/2]fcoe-port 5
[*SWA-10GE4/0/2]port trunk allow-pass vlan 200
[*SWA-fcoe5]quit
[*SWA]fcoe npv1 npv
[*SWA-fcoe-npv-npv1]fcoe role vn_port interface fcoe-port4
[*SWA-fcoe-npv-npv1]fcoe role vn_port interface fcoe-port5
[*SWA-fcoe-npv-npv1]npv-map vf-port interface fcoe 2 to
interface fcoe 3 vn-port interface fcoe 5
[*SWA-fcoe-npv-npv1]quit
[*SWA]commit
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Creating an FCoE interface and an FC instance
- Background
- On a traditional FC network, all FC nodes are interconnected through a fabric. FC nodes belonging to a fabric have the same attributes. To control FCoE traffic forwarding, FCoE uses FC instances because FCoE transmits SAN traffic over the Ethernet. An FC instance defines fabric attributes such as the FCoE VLAN. The NPV switch uses a VF_Port to connect to a VN_Port of an ENode and uses a VNP_Port to connect to a VF_Port of an FCF.
- Procedure
- Run the system-view command to enter the system view.
- Run the interface fcoe-port interface-number command to create an FCoE interface and enter its view.
- Run the quit command to exit from the system view.
- Run the interface interface-type interface-number command to enter the view of the physical interface that needs to be added to the FCoE interface.
- Run the fcoe-port fcoe-port-id command to add the physical interface to the FCoE interface.
- By default, no physical interface is added to an FCoE interface.

Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

Cloud Data Center Underlay Network Deployment

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Foreword

- This course enables you to learn basic knowledge of cloud data center underlay network design. Your will learn NETCONF, SNMP, stack, M-LAG, BGP, and network reliability design.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Objectives

- Upon completion of this course, you will be able to:
  - Obtain general knowledge of the data center underlay network service plan, including the IP address, VLAN, IGP, and BGP plan.
  - Learn knowledge of data center reliability design which can be enhanced by using multiple technologies. In addition to the network, infrastructure, application cluster, and disaster recovery are all key points in reliability design.
  - Learn knowledge of network reliability which can be improved through reliability design including system-level reliability, device-level reliability, and link-level reliability.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. **SDN Interface Protocols and Features**
   - Introduction to the NETCONF Protocol
   - Introduction to the SNMP Protocol
2. Eth-Trunk
3. Stack
4. M-LAG
5. BGP
6. Underlay Networking Cases

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

NETCONF Concept

- The network configuration protocol (NETCONF) is a network configuration and management protocol based on Extensible Markup Language (XML).
- The NETCONF protocol provides network device installation, operation, and deletion mechanisms. It can be used for remote network device management and monitoring.
- The NETCONF protocol provides a set of standard operations and RPCs for network management.
- The NETCONF protocol adopts the SSH transmission protocol.

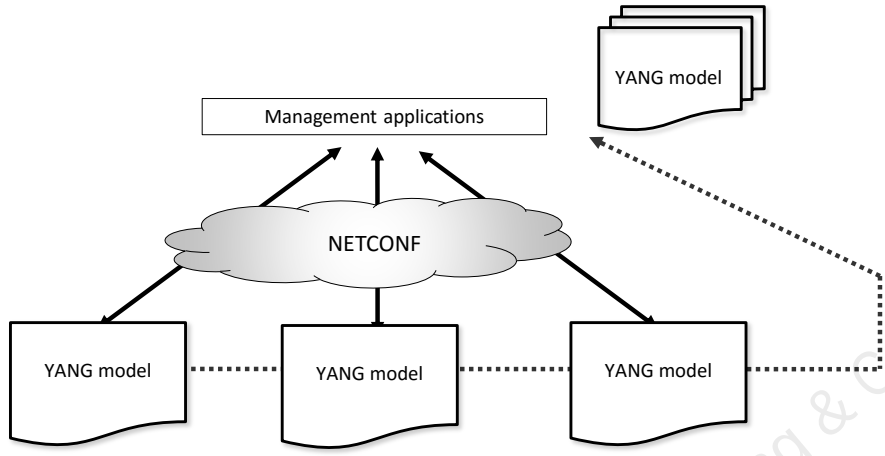
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- NETCONF is a network management protocol developed and standardized by IETF. In December 2006, it was first issued in RFC4741. In June 2011, it is revised in RFC6241.
- NETCONF is a network configuration and management protocol based on XML. NETCONF uses a simple mechanism based on remote procedure call (RPC) to implement communication between clients and servers.
- The transport layer uses the SSH protocol to carry the NETCONF protocol. The default TCP port number is 830. The NETCONF server must listen this port.
- The NETCONF protocol has the following advantages:
  - Defines messages in the XML format, and modifies configuration information based on the RPC mechanism, facilitating configuration management and implementing interoperability between devices of different vendors.
  - Reduces network faults caused by manual configuration errors.
  - Improving efficiency of system upgrading using the configuration tool.
  - Provides high extensibility, allowing different vendors to define their own protocol operations for special management functions.
  - Provides security mechanisms such as authentication and authorization, ensuring message transmission security.

Huawei Training & Certification

NETCONF Framework



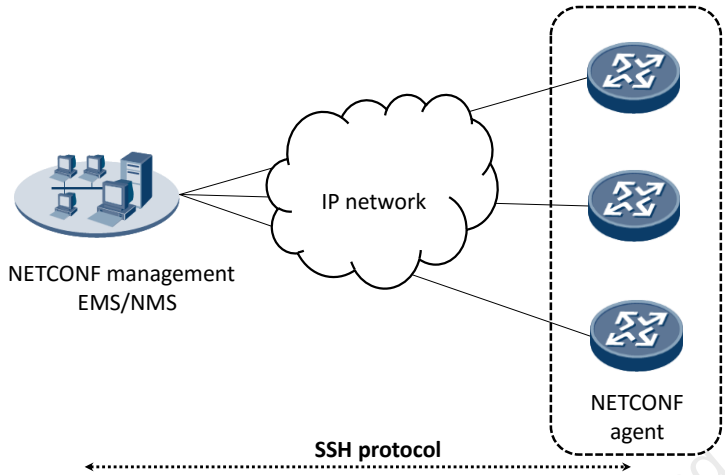
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- YANG, short for Yet Another Next Generation, was initially created as the modeling language of NETCONF. In the entire NETCONF configuration protocol system, YANG is used to describe the configuration, status, notification time, and O&M of a specific NE, and constraints between NEs.
- The YANG language is developed by the data modeling language working group (NETMOD) of IETF NETCONF, and is defined in the RFC6020.
- YANG is used to define the operation data, configuration data, notification, and modeling language of operations in the NETCONF protocol.
- YANG is a modular language using the data structure of an XML tree.
- NETCONF is mainly used together with the YANG language. The figure shows the framework. The NETCONF protocol is open to management applications and uses the Yang model to precisely define the data structure, grammar, and data semantics. Operations and modifications are performed in YANG models. The management software uses NETCONF to manage the YANG module. As a protocol, NETCONF needs to comply with specific regulations and layer dividing rules.

Huawei Training & Certification

NETCONF Network Architecture



- The NETCONF network architecture consists of three parts, including NETCONF management, NETCONF agent, and schema.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The NETCONF network architecture consists of the following parts:
- NETCONF manager
- The NETCONF manager functions as a client that uses NETCONF to manage network devices.
  - The NETCONF manager can send RPC requests to the NETCONF agent to query or modify one or more parameters.
  - The NETCONF manager can receive the alarms and events actively reported by the NETCONF agent to obtain the status of a managed device.
- NETCONF agent
- The NETCONF agent functions as a server that maintains managed network devices, responds to the RPC request, and reports management data to the NETCONF manager.
  - After receiving the RPC request, the NETCONF agent parses the request, processes the request based on the Configuration Management Plane (CMP), and sends a response to the NETCONF manager.
  - If an alarm is generated or an event occurs on a managed device, the NETCONF agent of the managed device reports the alarm or event to a NETCONF manager for the NETCONF manager to learn the status of the managed device.

Huawei Training & Certification

Contents

1. **SDN Interface Protocols and Features**
   - Introduction to the NETCONF Protocol
   - Introduction to the SNMP Protocol
2. Eth-Trunk
3. Stack
4. M-LAG
5. BGP
6. Underlay Networking Cases

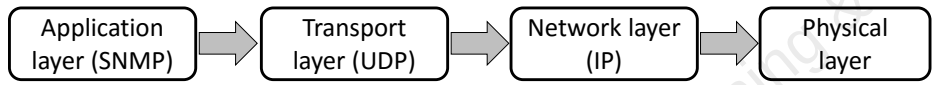
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

SNMP Overview

- The simple network management protocol (SNMP) is a standard network management protocol widely used on TCP/IP networks for device information collection, management, and modification.
- As part of the TCP/IP protocol family, SNMP messages are encapsulated into UDP packets for transmission over the Internet.

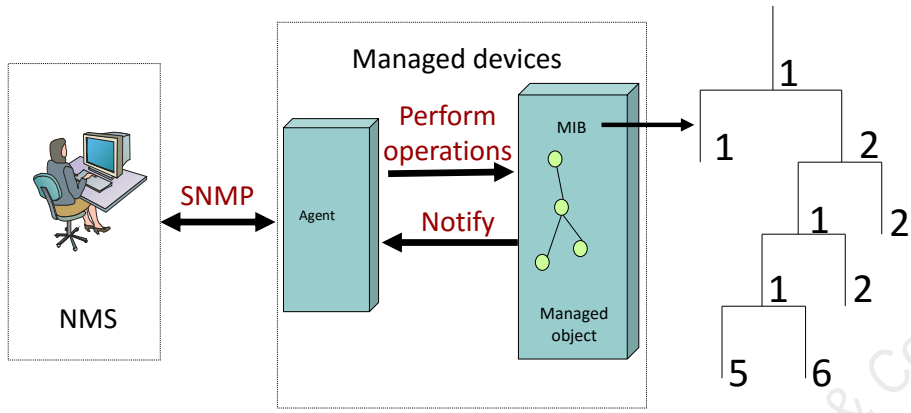


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

SNMP Architecture



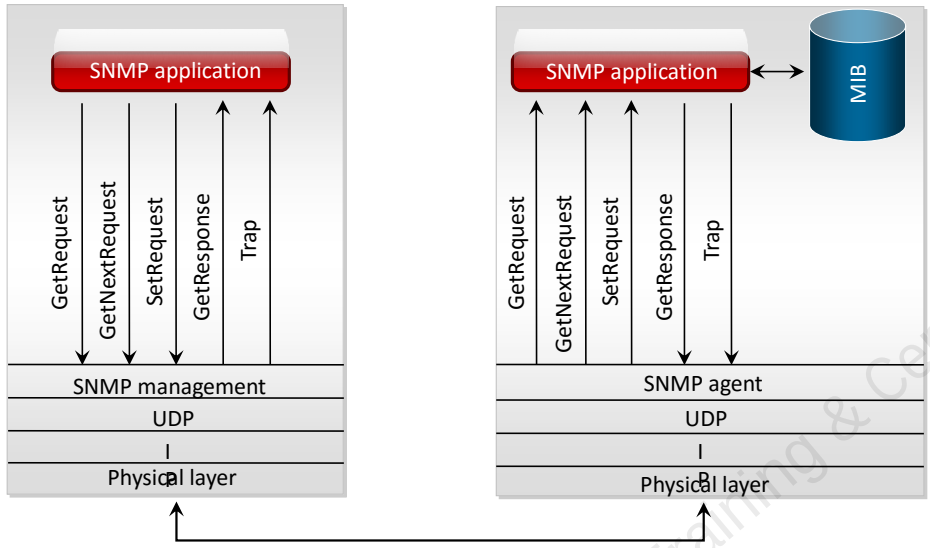
- There are three roles for SNMP-based network management: NMS, agent, and managed device.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Three roles are defined during the process of device management through SNMP:
  - NMS: Sends query packets to the managed devices, and receives alarms sent by the managed devices.
  - Agent: As a process on the managed device, the agent has following functions:
    - Receives and parses query packets sent from the NMS.
    - Reads or writes management variables based on the query type, and generates and sends response packets to the NMS.
    - Sends alarms to the NMS when particular events occur. For example, the system view is displayed or closed, or the device is restarted. Protocol modules on the device define the conditions that lead to the alarms.
  - Managed device: They are managed by the NMS, and report generated alarms to the NMS.
- To uniquely identify managed objects, SNMP organizes them in a hierarchical tree structure and identifies each one by a path starting from the tree root to the node, as shown in the figure. The nodes on the tree are the managed objects.
- An object is uniquely identified by a string of digits {1.2.1.1}. This string is its object identifier. A MIB is a set of standard variable definitions of monitored network devices. It is used to describe the hierarchical tree structure.
- Users can use a standard MIB or define a MIB based on their requirements. Using a standard MIB can reduce the costs of proxy component deployment, and thereby reduce the cost of the entire NMS.

Huawei Training & Certification

SNMP Protocol Architecture



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- SNMP defines five types of protocol data units (PDUs, that is, SNMP packets) for interactions between the NMS and the agent.
- SNMP replaces the complex command set with GET-SET operations.
- The five types of PDUs are described as follows:
  - get-request: obtains one or more parameter values from the proxy process.
  - get-next-request: obtains the next parameter value from the proxy process.
  - set-request: sets one or more parameter values of the proxy process.
  - get-response: returns one or more parameter values. This operation is performed by the agent process. It is the response to the preceding three operations.
  - trap: proactively sent by a proxy process to notify a management process of an event notice.
- The first three operations are sent by the NMS to the agent whereas the last two operations are sent from the agent to the NMS.

Huawei Training & Certification

SNMP Implementation



The agent receives NMS SNMP request packets through UDP port 161, decodes packets, checks the version number and community name authentication, reads and processes related PDUs, and returns corresponding messages to the NMS.

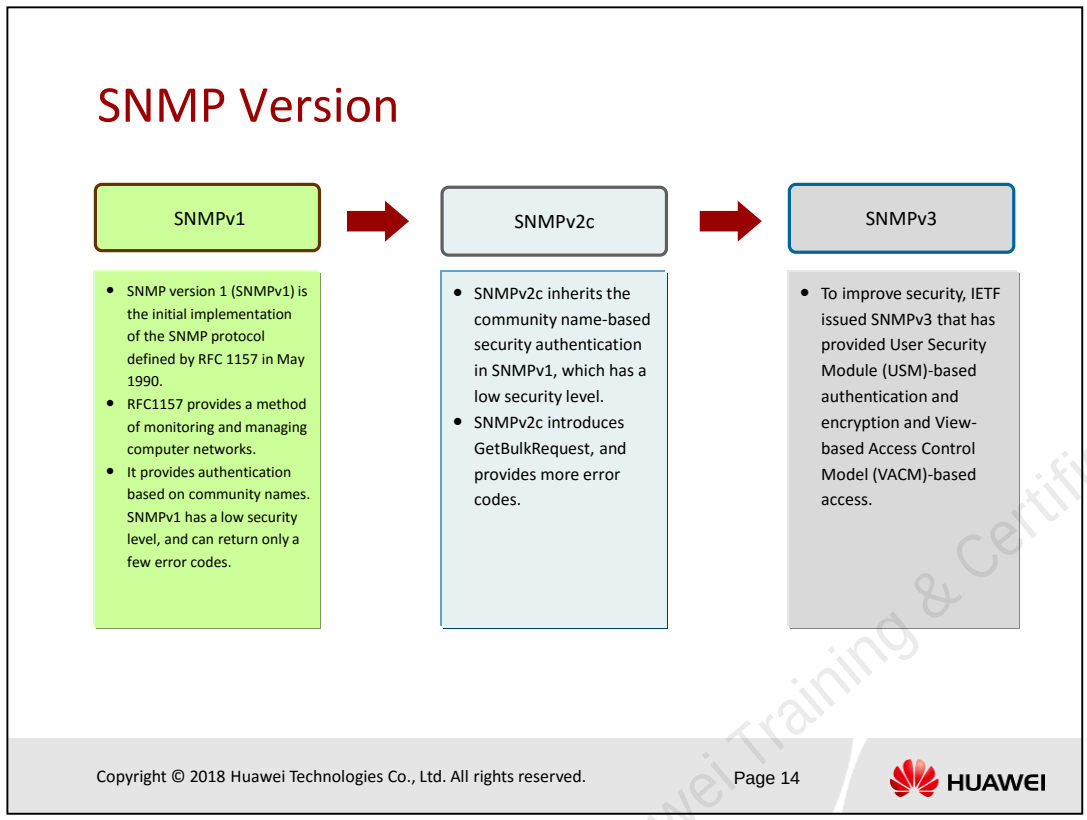
If values of some modules exceed the defined threshold, the agent will use UDP port 162 to send a trap message to the NMS.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification

- The agent receives requests from the NMS through UDP port 161.
- The agent processes a received packet in the following steps:
  - Decodes the packet. According to ASN.1 encoding rules, the agent generates a packet that is represented in the internal data structure. If an error occurs in this process and the decoding fails, the packet is discarded without further processing.
  - Checks the SNMP version. The agent compares the version of the received packet with the list of the supported versions.
  - Authenticates the community name. The agent abstracts the packet community name, which is filled in by the origin NMS.
  - Checks the community name. The agent obtains from the packet the community name, which is set by the NMS that sends the packet. SNMPv1 provides a weak security measure, which will be enhanced in SNMPv3. SNMPv1 provides a weak security measure, which will be enhanced in SNMPv3.
  - Obtains the PDU: The agent obtains the PDU from the authenticated ASN.1 object. If the process fails, the packet is discarded without further processing.
  - Processes the PDU: The agent processes the PDU according to its SNMP entity. Then the agent obtains the value of the managed from the corresponding module to generate a response packet. After that, the agent encodes the response packet and sends it back to the NMS.
  - The NMS receives the response packet and processes it in a similar way. Then the NMS displays a result.
- When alarm triggering conditions are met in the module of a managed device, the agent sends traps to the NMS. Then, network management engineers can take timely measures to cope with network faults.
- The NMS uses the 162 port to receive the trap packet and does not need to confirm the receiving of the packet.



- There are three versions of SNMP: SNMPv1, SNMPv2c, and SNMPv3.
- SNMPv1 and SNMPv2c both adopt the community name-based authentication. NMS uses the community name list to control the device access rights, whereas the agent does not check whether the community name used by the sender is authorized. Additionally, SNMP messages are not encrypted for transmission, so security in authentication and privacy is not ensured.
- SNMPv2c is an enhanced version of SNMPv1. Compared with SNMPv1, SNMPv2c supports more operations, data types, communications protocols, and error handling codes.
- SNMPv3 provides all functions of SNMPv1 and SNMPv2c, and offers a new security mechanism for authentication and encryption.
- SNMPv3 ensures data security and access control.
- SNMPv3 provides message-level security in the following ways:
  - Data security: Data modification must be authorized beforehand. Data sequence modification must be within the specific scope.
  - Data source verification: This verification makes clear which user the received data comes from. SNMPv3 ensures security on a user basis and authenticates messages by user instead of application program.
  - Data verification check: The NMS or agent checks the time a received message was generated. The message will not be accepted if the difference between the message time and the system time exceeds the specified range. Data verification check protects in-transfer messages against being tampered with, and prevents the reception and handling of malicious messages.
- NMPv3 implements access control based on protocol operations, and controls the access to managed objects.

Huawei Training & Certification

Contents

1. SDN Interface Protocols and Features
2. **Eth-Trunk**
3. Stack
4. M-LAG
5. BGP
6. Underlay Networking Cases

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Eth-Trunk

- Ethernet link aggregation, also called Eth-Trunk, bundles multiple physical links to form a logical link to increase link bandwidth.
- Multiple bundled links dynamically back up each other to improve reliability.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Multiple interfaces can be aggregated into a LAG. The member interfaces in the LAG load balance incoming and outgoing traffic to provide high reliability. Port aggregation can be divided into manual aggregation, dynamic LACP aggregation, and static LACP aggregation. Link aggregation can work in load balancing or non-load balancing mode.

Huawei Training & Certification

Advantages of Eth-Trunk

- Increase bandwidth
  - The maximum bandwidth of a link aggregation interface is the total bandwidth of member interfaces.
- Improve reliability
  - When an active link fails, traffic on the link can be switched to another available member link, ensuring high reliability of the link aggregation interface.
- Balance loads
  - In a link aggregation group, traffic is balanced among active member links.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

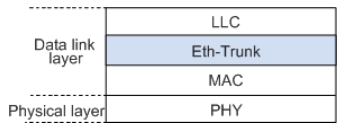
HUAWEI

- Advantages of link aggregation:
  - First, increased network bandwidth. Port aggregation binds multiple connected ports into a logical port, whose total bandwidth is the total of bandwidth of each independent port. When increasing traffic on a port limits network performance, using switches that support link aggregation can easily increase the network bandwidth (for example, two 10 Gbit/s ports can be aggregated into a 20 Gbit/s port or four 10 Gbit/s ports can be aggregated into a 40 Gbit/s port). This feature is applicable to 1 GE, 10 GE, 40 GE, and 100 GE Ethernet technologies.
  - Second, improved network reliability. When data mission rate is very high on the backbone network, its security is of great importance. Reliability of high-speed servers and backbone network connections must be ensured. Port aggregation can prevent such faults. For example, one mistakenly removed cable will not cause link disconnection. That is, once a port in the port aggregation group fails, traffic will be automatically redirected to normal connections. This process is fast, as only one access address needs to be changed. Then the switch transfers data to another port to ensure normal network connection.

Huawei Training & Certification
Huawei Training & Certification

Forwarding Principle of Eth-Trunk

- The Eth-Trunk is located in the data link layer between the MAC address layer and the LLC sub-layer.
- The Eth-Trunk can extract the source and destination MAC/IP addresses of a data frame, and implements load balancing based on the forwarding table.



Position in the OSI protocol stack

- HASH-KEY value
- The HASH-KEY value is calculated by using the hash algorithm based on the MAC or IP address of a data packet.
- Interface number
- The distribution of Eth-Trunk forwarding entries is related to the number of supported member interfaces of an Eth-Trunk. Different HASH-KEY values correspond to different outbound interfaces.

HASH-KEY	0	1	2	3	4	5	6	7	8	.....	15
PORT	1	2	3	4	1	2	3	4	1	.....	4

Eth-Trunk forwarding table

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The Eth-Trunk module forwards data frames based on the forwarding table in the following steps:
- After receiving a data frame from the MAC sub-layer, the eth-trunk module abstracts its source or destination MAC/IP address based on the load balancing mode.
- The Eth-Trunk module uses the hash algorithm to calculate the HASH-KEY value.
- The eth-trunk module finds the interface corresponding to the HASH-KEY value in the forwarding table, and sends the data frame from the interface.
- A data flow is a group of data packets with one or more identical attributes. These attributes include the source MAC/IP address, destination MAC/IP address, TCP/UDP source port number, and destination TCP/UDP port number.
- When there are multiple physical links between both devices of the Eth-Trunk, the first data frame of a data flow is transmitted on one physical link, and the second data frame is transmitted on another physical link. In this case, the second data frame may arrive the peer device earlier than the first data frame. As a result, packet mis-sequencing occurs.
- To avoid this fault, Eth-Trunk adopts the flow-based load balancing mechanism, which uses the HASH algorithm to generate HASH-KEY values for addresses in data frames. Then Eth-Trunk searches for the outbound ports in the Eth-Trunk forwarding table according to the hash-key values. Each MAC or IP address corresponds to a unique HASH-KEY value, so the outbound interfaces are different. This ensures that frames of the same data flow are forwarded on the same physical link and implements load balancing of traffic on each physical link of the LAG (that is, flow-based load balancing). Flow-based load balancing ensures sequence of load balancing but not the bandwidth usage.

Huawei Training & Certification

Eth-Trunk&LACP

- The Link Aggregation Control Protocol (LACP) is based on the RFC standard (802.3ad). It is used by switches to automatically form an aggregation link based on their configuration and start the aggregation link to receive and send data.
- LACP dynamically maintains the link status. It automatically adjusts or disables link aggregation when aggregation conditions change.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- LACP based on IEEE 802.3ad implements dynamic link aggregation and de-aggregation. LACP uses the link aggregation control protocol data unit (LACPDU) to exchange information with a peer end.
- After member ports are added to the Eth-Trunk port in LACP mode, the ports will send LACPDUs to the peer end to notify information about their system priorities, MAC addresses, port priorities, port IDs, and operation keys (used to check whether the peer end of each interface is in the same aggregation group and whether the bandwidth of each interface is consistent). After the peer receives the information, it compares the information with its stored information, and then selects interfaces that can be aggregated. Devices at both ends determine which interfaces are active interfaces, and thus active links are identified.

Huawei Training & Certification

LACP Feature

- The system exchanges protocol packets to implement auto-negotiation. The packets include the system configuration and status.
- Protocol packets are sent periodically or triggered by events:
  - Event triggering: The events such as the change of the local end status or configuration trigger the generation and transmission of new protocol packets.
  - Periodic sending: The system sends the status periodically to maintain link aggregation when the aggregation link works normally.
- Protocol packets do not carry their number, so both ends do not detect or retransmit the discarded protocol packets. Instead, both ends use a timer and periodic sending mechanism to prevent information loss.
- Slow protocol: No more than five protocol packets are sent per second.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Link Aggregation Mode

- Manual aggregation
  - The LAG number and port members are configured manually. LACP is not running on ports.
- Static aggregation
  - The LAG and port members are configured manually. LACP is running on ports.
- Dynamic aggregation
  - IEEE802.3ad-based LACP is enabled.
  - The LAG number is automatically created based on the protocol.
  - The aggregation port is automatically matched and added based on the key value.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Both manual aggregation and static LACP aggregation are manually configured, and do not allow the system to automatically add or delete manual and static aggregation ports. Each manual or static aggregation group must include at least one port. If the aggregation group has only one port, you can delete the port from an aggregation group only by deleting the aggregation group.
- The LACP protocol on a manual aggregation port is disabled, and cannot be enabled by users.
- The LACP protocol on a static aggregation port is enabled. After a static aggregation group is deleted, the member ports form one or more dynamic LACP aggregations. The LACP protocol is enabled. It cannot be disabled by users.
- Dynamic LACP aggregation is automatically created and deleted by the system. You are not allowed to add or delete member ports in a dynamic LACP aggregation group. Only ports with the same rate and duplex attributes, connected to the same device, and with the same basic configuration can be dynamically aggregated.
- The LACP protocol is enabled on a dynamic aggregation port. Dynamic link aggregation can also be created for one port, which is single-port aggregation.

Huawei Training & Certification

Contents

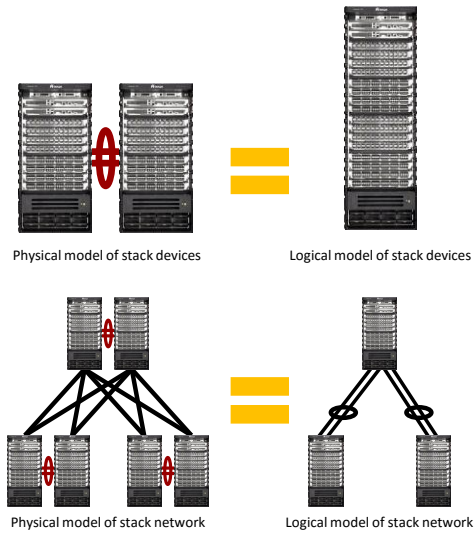
1. SDN Interface Protocols and Features
2. Eth-Trunk
3. **Stack**
4. M-LAG
5. BGP
6. Underlay Networking Cases

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

What Is Stack?



- Definition of stack

The stack technology combines multiple physical switches connected through stack cables into a logical switch system.

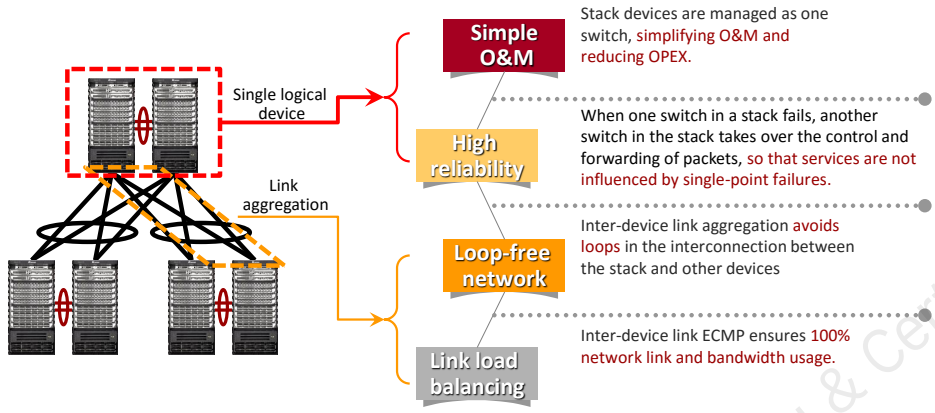
- Features of stack
  - Many-to-one virtualization: Multiple switches are virtualized into one logical switch that has a unified control plane and provides unified management.
  - Unified forwarding plane: Multiple physical switches use a unified forwarding plane that shares and synchronizes forwarding information.
  - Inter-device link aggregation: Links between physical switches in a stack are aggregated into one Eth-Trunk port for interconnection with downstream devices.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Cluster Switch System (CSS):
  - CSS is a network virtualization technology that that virtualizes clustering-capable switches into a larger logical switch.
- CSS is featured by:
  - Multiple-to-one virtualization: Multiple switches are virtualized into one logical switch that has a unified control plane and provides unified management.
  - Unified forwarding plane: Physical devices in a CSS use a unified forwarding plane that shares and synchronizes forwarding information.
  - Inter-switch link aggregation: Links between physical switches in a CSS are aggregated into one Eth-Trunk port for interconnection with downstream devices.

Huawei Training & Certification

Stack Benefits



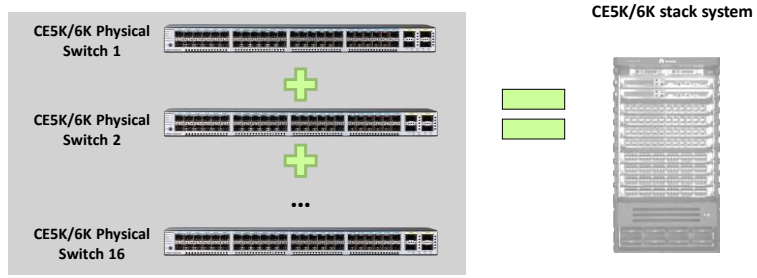
Stack is the first choice for small- and medium-sized data centers for its simple management and efficient network.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- With multiple-to-one virtualization and inter-switch link aggregation, CSS not only simplifies the network topology, but also greatly improves network performance.
  - Simple O&M: Devices in a CSS are managed as one switch, simplifying O&M and reducing OPEX.
  - High reliability: If one switch in a CSS fails, another switch in the CSS takes over the control and forwarding of packets, so that services are not influenced by single-point failures.
  - Loop-free network: CSS supports inter-switch link aggregation to prevent loops. No complex loop prevention protocol such as MSTP is required.
  - Link balancing: CSS supports equal cost multi path (ECMP) across switches, making full use of network links and bandwidths.
- CSS simplifies network architecture, improves forwarding performance, and does not lead to the loss of any network functions. CSS has all functions of the physical switches in the cluster and provides better performance. CSS gains wide customer recognition and becomes a preferred solution for simple and efficient network deployment.

Huawei Training & Certification

Stack Types and Features - iStack



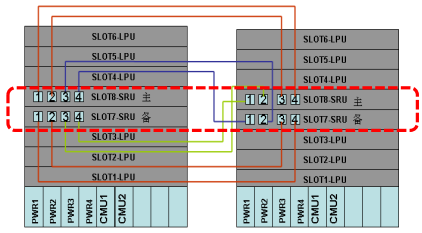
- Stack for fixed switches (iStack)
  - **Deployment mode:** Fixed switches are interconnected using high-speed copper cables or optical fibers, forming a ring or chain. The stack port must use a 10 GE or larger interface.
  - **Deployment benefits**: No dedicated stack card or stack cable is required, which increases bandwidth and avoids loop**s.**
  - **Typical product:** CE5800, CE6800, and CE7800 series switches

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

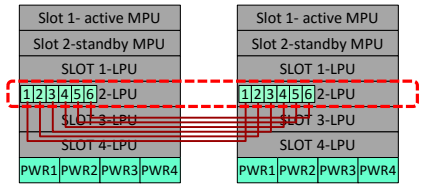
- Multiple CE5800, 6800, or 7800 switches form a stack system and logically construct a modular switch.

Huawei Training & Certification

Stack Types and Features - CSS



- Dedicated line card-based CSS
  - **Deployment mode:** Switches in a CSS are interconnected using dedicated line cards and dedicated cables.
  - **Deployment benefits:** No line card slot is required. High-speed ports are used to increase CSS interconnection bandwidth.
  - **Typical product**: S9300



- Service line card-based CSS
  - **Deployment mode:** Switches in a CSS are interconnected using service line cards and standard cables.
  - **Deployment benefit:** No extra component is required. Long-distance CSS is supported. Eth-trunk ports are used to increase CSS interconnection bandwidth.
  - **Typical product: CE12800**

The service line card-based CSS will stand out as device performance is substantially improved.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- CSS has gone through two development stages.
- At the early stage, switches set up a CSS using dedicated line cards and are interconnected by dedicated cables. Because the dedicated line cards do not occupy slots for service line cards, they do not deteriorate forwarding performance of the entire system. Meanwhile, interfaces on dedicated line cards have a higher transmission rate than interfaces on service line cards, improving bandwidth for CSS interconnection.
- At the later stage, the physical switches set up a CSS using service line cards and are interconnected by standard cables. Although several service interfaces are used for interconnection, this mode has a slight impact on the forwarding performance of the entire system because the forwarding capability of switches has greatly improved. The use of standard service interfaces and standard cables allows switches far from each other to form a cluster, making CSS deployment flexible. Moreover, the interconnected interfaces support link aggregation, which allows flexible expansion of interconnection bandwidth.

Huawei Training & Certification

Contents

1. SDN Interface Protocols and Features
2. Eth-Trunk
3. Stack
4. **M-LAG**
5. BGP
6. Underlay Networking Cases

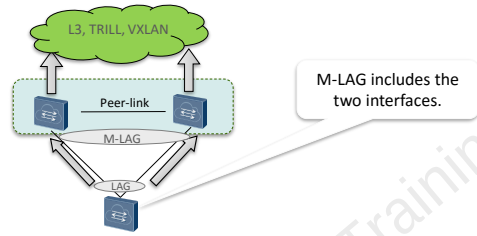
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Problems in Traditional Technologies

- The main application scenario of M-LAG is dual-homing access, that is, the user-side device is dual-homed to two access devices. Traditional technologies in this scenario include the stack and smart-link technologies, but they both have disadvantages:
  - Stack devices (two or more) share a control plane. A single-point fault can affect the entire system, especially during upgrade. Therefore, stack devices cannot meet high stability requirements of aggregation switches and core switches.
  - The smart-link technology uses the active/standby access mode. That is, for the same VLAN, there is only one link for forwarding at one time. So the link usage is low.
- To solve these problems, a new dual-homing access mode, M-LAG, is developed.



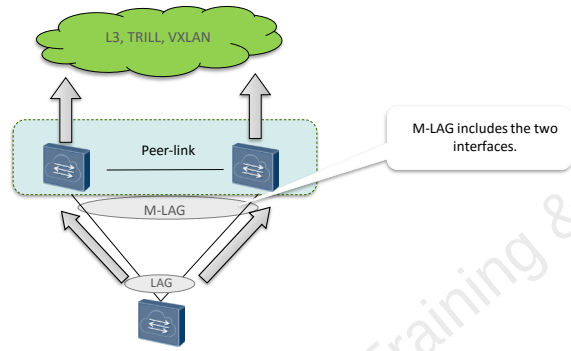
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Principles of M-LAG

- M-LAG is short for multi-chassis link aggregation. That is, different ports on different devices are integrated into an M-LAG, which can provide the functions of a common LAG.



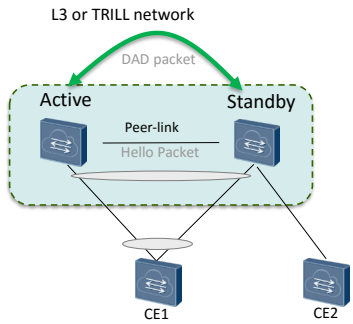
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

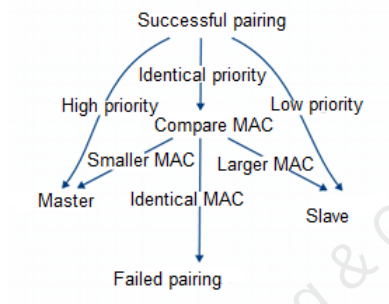
- On the forwarding plane, devices are accessed in LAG mode and traffic is load balanced to two devices based on the flow.
- On the control plane, two in dependent devices construct an M-LAG and interconnect with each other through the M-LAG protocol. One device failure affects no other device except itself.

Huawei Training & Certification

Device Pairing

- After two devices of an M-LAG are configured, they will be paired.





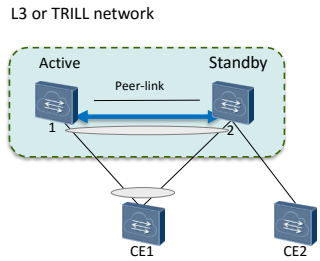
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

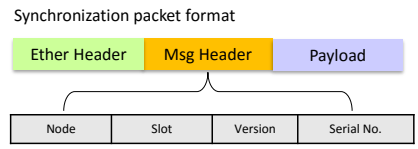
- Two devices periodically send Hello packets on peer-link (one packet per second). A Hello packet carries its own DFS group ID, protocol version number, and system MAC address.
- After receiving a Hello packet from the peer end, pairing will be performed based on the DFS group ID. If the IDs are the same, the pairing is successful.
- After the pairing succeeds, the active and standby devices are determined based on priority. The device with higher priority is the active device. If their priorities are same, the device with a smaller MAC address is the active device. Active and standby devices have no difference in forwarding traffic. It makes difference only when a fault occurs.
- After the pairing succeeds, data synchronization is performed between two devices. The information to be synchronized includes the device name, system MAC address, software version, M-LAG status, STP BPDU, MAC address, ARP, and IGMP entries.
- After the pairing succeeds, the devices send heartbeat packets through external Layer 3 or TRILL network (non-peer-link). The heartbeat packets are mainly used for dual-active detection when a peer-link fault occurs.

Huawei Training & Certification

Data Synchronization

- M-LAG is a logical LAG and entries on the two ends must be the same. In this way, entries on the two ends must be synchronized to prevent abnormal traffic.





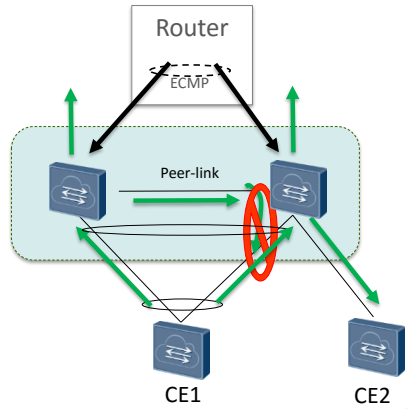
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The configuration items and entries that need to be consistent are as follows:
  - MAC address table
  - ARP table
  - Layer 2 and Layer 3 multicast forwarding table
  - DHCP snooping table
  - LACP System ID
  - Global STP configuration and port configuration
  - Global VLAN and ACL
- The entries to be synchronized include the ARP, MAC, and IGMP entries.
- The right figure shows the format of the synchronization packet. The meaning of each field is described as follows:
  - Ether header: external Ethernet header
  - Msg header: user-defined message header
  - Node: device node ID
  - Slot: slot to receive messages
  - Version: protocol version number
  - Protocol sequence number: used to enhance reliability
  - PayLoad: specific packet content, including various entries

Huawei Training & Certification

Traffic Model

- Upstream traffic on CE switches is load balanced to two M-LAG devices based on the flow.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Traffic is forwarded to M-LAG member devices. Local traffic is preferentially forwarded regardless of whether it is uplink or downlink traffic.
- Broadcast traffic passes through the peer-link. The peer-link is unidirectionally isolated from dual-homed M-LAG member interfaces (traffic entering from the peer-link cannot go out through the corresponding M-LAG member interface). The peer-link is not isolated from single-homed M-LAG member interfaces.

Huawei Training & Certification

Contents

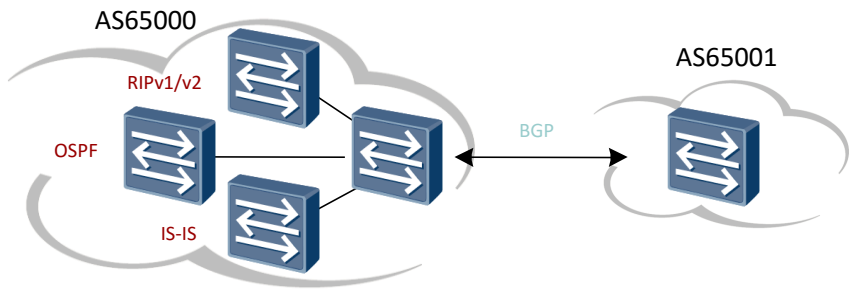
1. SDN Interface Protocols and Features
2. Eth-Trunk
3. Stack
4. M-LAG
5. **BGP**
6. Underlay Networking Cases

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Autonomous System



- Autonomous system (AS): An AS is a collection of routers that are managed by the same technology administration department and use the same routing policy.
- Internal routing protocol in an AS - interior gateway protocol (IGP)
- Routing protocol between ASs - exterior gateway protocol (EGP)

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- An autonomous system (AS) is a collection of routers that are managed by the same technology administration department and use the same internal routing policy.
- Each AS has a unique AS number, which is allocated by Internet Assigned Numbers Authority (IANA).
- Different AS numbers are used to distinguish different ASs. When the network administrator does not expect the data to pass through an AS, for example, if the AS is managed by a competitor or lacks a security mechanism, the network administrator can control the data forwarding path through the routing protocol, policy, and AS number.
- The AS number ranges from 1 to 65535. The numbers from 1 to 64511 are registered on Internet, and the numbers from 64512 to 65535 are for private networks.
- The differences between IGP and EGP are as follows:
  - IGP runs in the AS, including RIP, OSPF, and ISIS. IGP focuses on discovering and calculating routes.
  - EGP runs between ASs. It usually refers to the border gateway protocol (BGP). BGP focuses on route advertisement control and best route selecting.

Huawei Training & Certification

BGP Features

- BGP is an external routing protocol and is used to transmit routing information between ASs.
- BGP is an enhanced distance-vector routing protocol.
  - Reliable route update mechanism
  - Various metric measurement methods
  - Designed to prevent loops
- BGP adds attributes to routes.
- BGP supports classless inter-domain routing (CIDR).
- BGP provides various route filtering and routing policies.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- BGP is a dynamic routing protocol running between ASs. It is mainly used to exchange loop-free routing information between ASs. By exchanging the reachable routing information with the AS number sequence attribute, an AS topology can be constructed to eliminate routing loops and implement user-defined routing policies. Different from OSPF, RIP, and IS-IS which run inside an AS and belong to Interior Gateway Protocol (IGP), BGP belongs to Exterior Gateway Protocol (EPG). BGP is used between Internet service providers (ISP).
- BGP has been used since 1989. Its earliest three released versions are RFC1105 (BGP-1), RFC1163 (BGP-2), and RFC1267 (BGP-3). The current version is RFC4271/RFC1771 (BGP- 4), which is rapidly becoming the actual Internet border routing protocol standard.
- The following paragraph describes the features of BGP:
  - BGP is used to exchange loop-free routing information between ASs. BGP is a policy-based routing protocol and its policies are controlled through various path attributes. BGP runs on the application layer and uses TCP as a transport protocol on the transport layer. A BGP neighbor relationship is established on the basis of a TCP session. BGP has similar path transmission mode to the distance-vector routing protocol. BGP chooses paths not based on distance (most routing protocols choose paths based on bandwidth), but based on various path attributes. These attributes are carried on the route transmission paths, so BGP is also a path-vector routing protocol.Taking an AS as a router will help you understand BGP as a path-vector routing protocol. In addition, BGP also has many link state (LS) routing protocol features, for example, update mechanism in triggered and incremental mode and route advertisement with masks.

Huawei Training & Certification

Reliable Route Update Mechanism of BGP

- The transmission protocol is TCP, and the port number is 179.
- Periodic update is not required.
- Only incremental routes are sent for route update.
- KeepAlive packets are periodically sent to check the TCP connectivity.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification

- BGP enhances its own reliability by using TCP as its bearing protocol.
- BGP does not need to periodically send update messages.
- BGP transmits only updated routes in the case of a route update. This implementation reduces the bandwidth used by BGP for route distribution. Therefore, BGP applies to the Internet where a large number of routes are transmitted. During BGP initialization, all routes are sent to BGP peers, and the routing information that has been sent to the BGP peer is saved locally. When receiving a new route, the local BGP determines whether to send the route based on the saved information. If the route has not been sent, the BGP sends the route. If the route has been sent, the BGP compares the route with the sent route. If the new route is more optimal, the BGP sends the new route and updates the saved information. Otherwise, it does not send the route. When the local BGP detects an invalid route (such as the corresponding port fails), if the route has been sent, BGP will send a route withdraw message to the BGP peer. In short, BGP does not broadcast all route information every time. It only sends routes that are changed after route initialization. In this way, communication between BGP and peers can be minimized.
- In addition, BGPs send and receive KEEPALIVE messages to check TCP connectivity between each other.

BGP Packet Type

- There are four types of BGP packets:
  - Open: Used to establish neighbor relationship with peers.
  - KeepAlive: Periodically sent between BGP peers to maintain connection.
  - Update: Used to transmit route information between BGP peers.
  - Notification: Sent to peers when the BGP speaker detects an error.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Nodes running BGP are called BGP speakers. They exchange four types of packets between each other. In the packets, OPEN, KEEPALIVE, and NOTIFICATION packets are used to establish and maintain peer connections.
- OPEN: It mainly carries information such as the BGP version and AS number. Two nodes that attempt to establish a BGP peer relationship, set up a TCP session, and exchange the OPEN information to determine whether a neighbor relationship can be established.
- KEEPALIVE: It is used to maintain the BGP peer relationship and periodically exchanged to check whether the peers are reachable.
- UPDATE: It is used to exchange route information between peers. An UPDATE packet contains withdrawn routes, reachable routes, and route attributes. It is the most important BGP packet.
- NOTIFICATION: It is the error detection mechanism of BGP. Once an error is detected, the BGP speaker will send a NOTIFICATION packet. The neighbor relationship will be closed.

Huawei Training & Certification

Message Applications of BGP

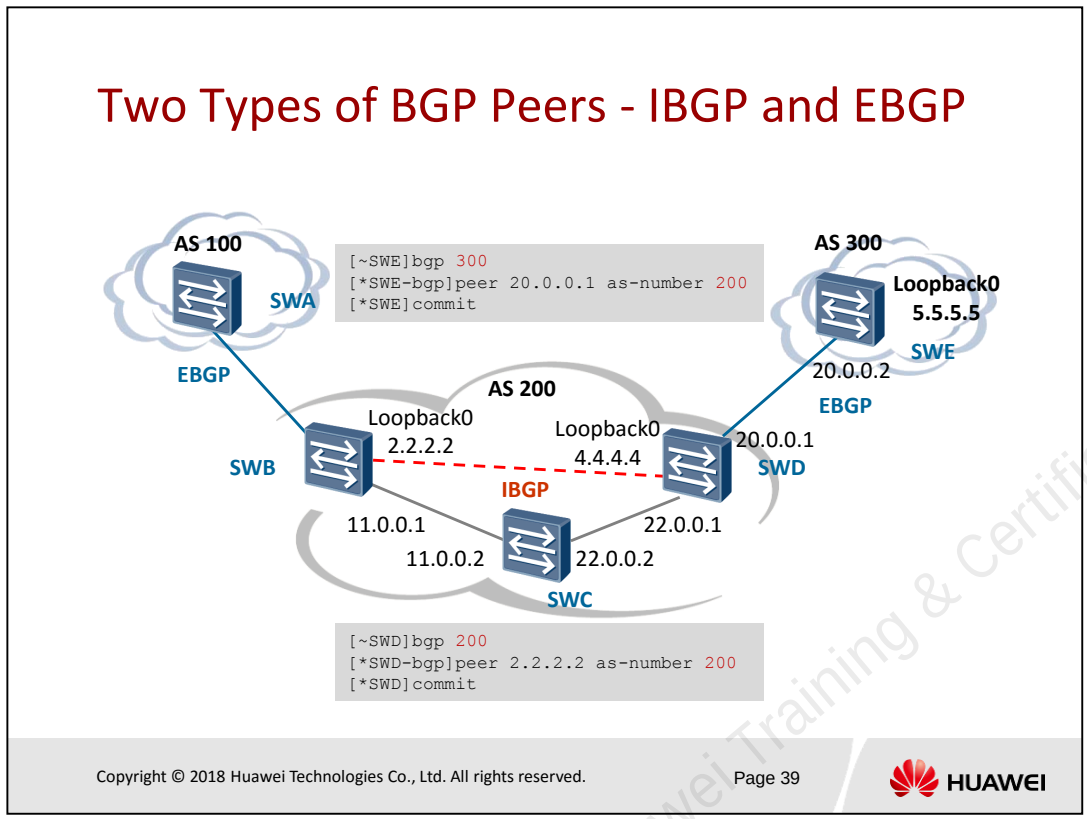
- When a BGP connection is built through TCP, an OPEN message is sent.
- After a connection is built, if there is a route changed or to be sent, an UPDATE message is sent to notify the peer of the route information.
- After the connection becomes stable, KEEPALIVE messages are periodically sent to check the BGP connection validity.
- If an error occurs when the local BGP is running, a NOTIFICATION message is sent to notify the BGP peer.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification Huawei Training & Certification

- BGP uses TCP to establish connections. The local listening port number is 179. Similar to the establishment of a TCP connection, a BGP connection also requires a series of sessions and handshakes. By means of handshake negotiation, TCP advertises its parameters, for example, port number. Handshake negotiation parameters used in BGP are BGP version, BGP connection hold time, local router ID, and authorization information. The information is carried in OPEN messages.
- After establishing a connection, BGP sends an Update message to the peer end when a route is to be sent. When advertising a route, the Update message carries the route attributes of the route to help the BGP peer to select the optimal route.
- When a local BGP route changes, an Update message is sent to notify the BGP peer of the change.
- After routing information is exchanged for a period of time, neither the local BGP router nor the BGP peer has any new route to advertise, and the BGP connection becomes stable. In this case, KEEPALIVE messages are periodically sent to check the BGP connection validity. If the local BGP router does not receive any KEEPALIVE message from the peer, the local BGP router considers the BGP connection as down, tears down the BGP connection, and deletes all the BGP routes learnt from the peer.
- On detecting an error, for example, the peer BGP version is not supported locally and the local BGP router receives an invalid Update message, the local BGP router sends a Notification message to the peer. When the local BGP router exits a BGP connection, a Notification message is also sent.



- If two peers exchanging BGP packets belong to the same AS, then they are internal BGP (IBGP) peers, such as SWB and SWD. If two peers exchanging BGP packets belong to different ASs, they are external BGP (EBGP) peers, such as SWD and SWE.
- Although BGP is running between ASs, BGP connections also need to be established between different border nodes of an AS to implement route information transmission on the entire network. For example, to enable communication between AS100 and AS300, an IBGP connection needs to be established between SWB and SWD.
- The basic BGP configuration is as follows:
  - Enable BGP (by specifying the local AS number) and access the BGP view.
  - [~SWA] bgp as-number
  - Use the bgp command to enable BGP and access the BGP view. Use the undo bgp command to disable BGP.
  - By default, BGP is disabled.
  - One node can run in only one AS, that is, only one local AS ID can be specified.
  - Specify the peer IP address and its AS number.
  - [~SWA-bgp] peer ipv4-address as-number as-number
  - Use the peer as-number command to configure the AS number of the specified peer (group) and use the undo peer as-number command to delete the AS number of the peer group.
  - By default, the peer in the peer group has no AS number.

Huawei Training & Certification

BGP Route Advertisement Principles (1)

- Once a BGP connection is established, a BGP speaker advertises all BGP routes to new peers. Among multiple paths, the BGP speaker only chooses the most optimal one.

```
[~CE5850-A]display bgp routing-table

 Total Number of Routes: 2

 BGP Local router ID is 1.1.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>i  192.168.3.0        10.1.1.2                                 0      200i
 * i                     10.2.2.2                                 0      200i
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Generally, if the BGP speaker learns multiple routes destined for the same network segment, it selects one optimal route and sends it to the routing table. However, the router will also select the optimal route. So the optimal route selected by the BGP speaker may not be used by the router. For example, a static BGP has higher priority than an optimal BGP route destined for the same network segment. So the router will add the static route rather than the BGP optimal route into the routing table.
- As shown in the slide, there are two routes to 192.168.3.0 on the CE5850-A. Their next hops are 10.1.1.2 and 10.2.2.2 respectively. BGP selects the optimal route (marked with ">") based on route selection principles (to be introduced later). Then it checks whether there is a better route on CE5850-A, such as the static route to 192.168.3.0 (the static route priority is 60, and the BGP route priority is 255. Smaller value corresponds to higher priority). If there is such a route, BGP will add the better route into the IP routing table, and otherwise choose the BGP optimal route.

Huawei Training & Certification

BGP Route Advertisement Principles (2)

- A BGP speaker only advertises its own routes to the peer.

```
[~CE5850-A]display ip routing-table
Routing Tables: Public
         Destinations : 12       Routes : 12

Destination/Mask    Proto   Pre  Cost       NextHop         Interface

        1.1.1.1/32  Direct  0    0          127.0.0.1       InLoopBack0
       10.1.1.0/30  Direct  0    0          10.1.1.1        Vlanif10
       10.2.2.0/30  Direct  0    0          10.2.2.1        Vlanif20
      127.0.0.0/8   Direct  0    0          127.0.0.1       InLoopBack0
    192.168.3.0/24  BGP     255  0          10.1.1.2        Vlanif10
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

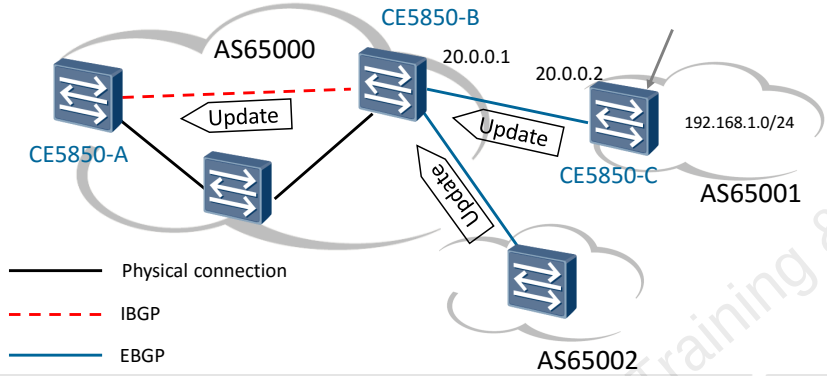
- The BGP speaker only advertises its own routes to the peer, that is, the BGP routes used in the IP routing table.
- Once the optimal BGP route (marked with ">") is sent to the IP routing table, BGP will send an Update message to advertise the route to other BGP peers (in addition to the peer receiving the BGP route).

Huawei Training & Certification

BGP Route Advertisement Principles (3)

- The BGP speaker advertises the routes obtained from EBGP to all BGP peers, including EBGP and IBGP peers.

```
[~CE5850-C]bgp 65001
[*CE5850-C-bgp]peer 20.0.0.1 as-number 65000
[*CE5850-C-bgp]network 192.168.1.0 255.255.255.0
[*CE5850-C-bgp]commit
```



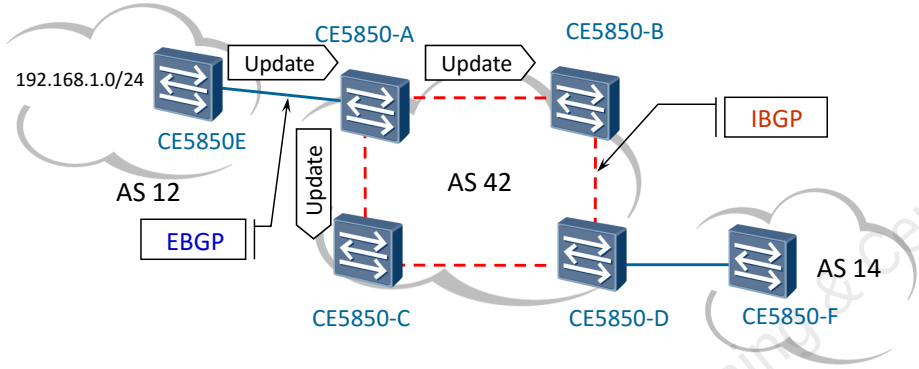
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The BGP speaker will advertise the routes obtained from EBGP to all its BGP peers, including EBGP and IBGP peers.

Huawei Training & Certification

BGP Route Advertisement Principles (4)

- The BGP speaker does not advertise the routes obtained from IBGP to its IBGP peers.

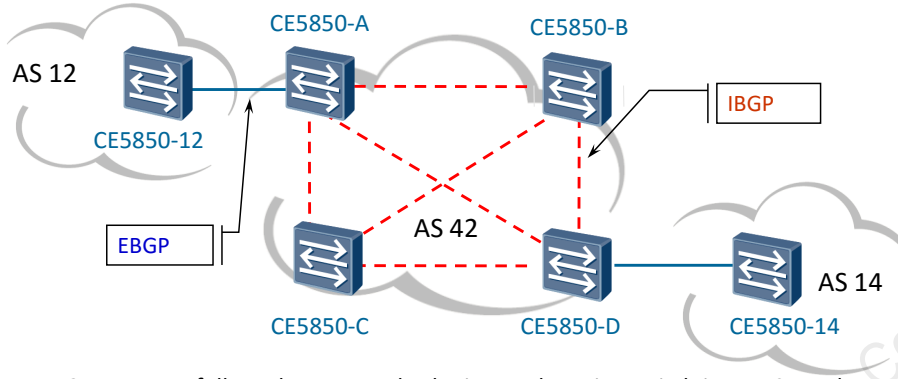


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Without this route advertisement rule, CE5850-C will advertise the route learnt from the IBGP peer CE5850-A to CE5850D. CE5850D then advertises it to CE5850B, which will advertise this route back to CE5850-A. In this way, a routing loop occurs in the AS.
- Therefore, this principle is an important measure to prevent loops in an AS. However, this principle brings another problem, that is, CE5850-D cannot receive BGP routes from AS 12. Generally, the logical full-mesh of IBGP is used to solve the problem. That is, two IBGP connections between CE5850A and CE5850D, and CE5850B and CE5850C are established.

Huawei Training & Certification

Full IBGP Interconnection



- IBGP supports full-mesh. As a result, the internal routing switch in an AS needs to maintain more IBGP sessions.
  - Route reflector
  - Confederation

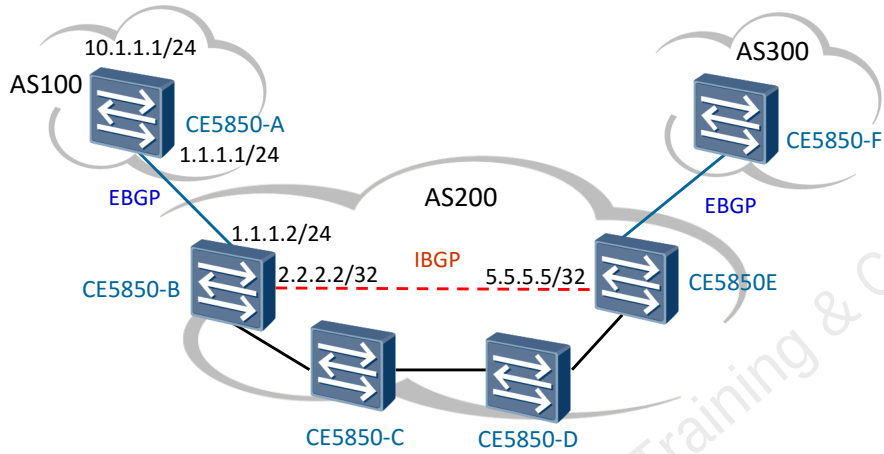
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Full-mesh IBGP connection is a way to solve the route transmission problem caused by IBGP horizontal splitting. The defect of this method is that routing switches need to spare more resources to maintain IBGP sessions in the network.
- BGP also provides the following two solutions to IBGP horizontal splitting:
  - Route reflector, proposed in RFC 2796
  - Confederation, proposed in RFC 3065

Huawei Training & Certification

BGP Route Advertisement Principles (5)

- Whether a BGP speaker advertises the routes obtained from IBGP to its EBGP peers depends on the synchronization between BGP and IGP.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Synchronization between BGP and IGP: The BGP speaker does not advertise routes obtained from IBGP peers to EBGP peers, unless the routes can also be obtained from IGP, which means the route can be transmitted smoothly in the AS.
- One of the main tasks of BGP is to advertise the network reachability information about an AS to other ASs. In the figure, CE5850-B encapsulates the route to 10.1.1.0/24 into a BGP packet, and advertises it to CE5850E through the TCP connection established by CE5850B and CE5850E. The CE5850E directly accepts this route and advertises it to RTF regardless of synchronization. Any packet from RTF or CE5850 destined for 10.1.1.0/24 must pass through CE5850-D and CE5850-C. However, as synchronization is not considered, no route to 10.1.1.0/24 exists in routing tables of CE5850-D and CE5850-C. As a result, data packets will be discarded when forwarded to CE5850-D. Therefore, data synchronization must be performed between BGP and IGP (such as RIP and OSPF). That is, before advertising a route update message received from an IBGP peer to its EBGP peers, a routing switch checks whether the destination exists in the routing table discovered by the IGP and whether non-BGP routing switches can transmit the packet to the destination. If the destination can be learnt by the IGP, the route advertises the route to its EBGP peers. Otherwise, the route is not advertised.

Huawei Training & Certification

Common BGP Route Attributes

- BGP routes mainly support the following features for CE switches:

BGP Attribute	Category
1. Origin	(well-known mandatory)
2. AS_Path	(well-known mandatory)
3. Next_Hop	(well-known mandatory)
4. Multi_Exit_Disc	(optional non-transitive)
5. Local_Pref	(well-known discretionary)

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Contents

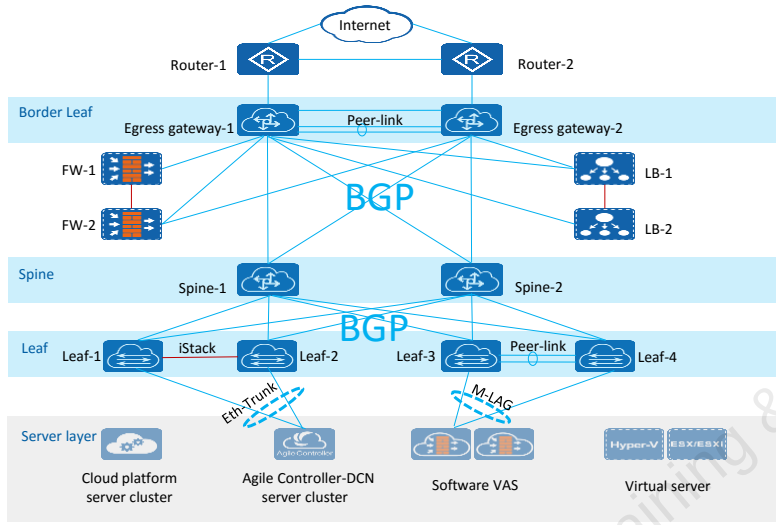
1. SDN Interface Protocols and Features
2. Eth-Trunk
3. Stack
4. M-LAG
5. BGP
6. **Underlay Networking Cases**

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Underlay Network Cases



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The underlay network uses the mature spine-leaf architecture to provide the scale-out capability and high access performance in the area. Leaf and spine nodes are fully meshed so that ECMP paths are available to ensure high availability of the network.
- Leaf nodes support multiple flexible networking solutions, such as iStack, M-LAG, and SVF, to meet access requirements of different servers.

Huawei Training & Certification

Quiz

1. What technologies are used by the data center underlay network?
2. What are the differences between stack and M-LAG technologies?

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Summary

- Limits of Traditional Network
- SDN Overview
- SDN Value
- SDN Standard Organization
- SDN Application Scenario

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

More Information

- Official website of Huawei
  - Huawei enterprise website: http://enterprise.huawei.com/cn/
  - Huawei technical support website: http://support.huawei.com/enterprise/
  - Huawei online learning website: http://learning.huawei.com/cn/
- Documentation tool
  - HedEx Lite
- Simulator
  - eNSP

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Recommendations

- Huawei e-learning website:
  - http://support.huawei.com/learning/Index!toTrainIndex
- Huawei support case library:
  - http://support.huawei.com/enterprise/servicecenter?lang=zh

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

CloudFabric VXLAN and EVPN Principles and Deployment

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Foreword

- Virtual Extensible LAN (VXLAN) is a Network Virtualization over Layer 3 (NVo3) technology that uses the MAC in User Datagram Protocol (MAC-in-UDP) mode to encapsulate packets.
- As a key cloud computing technology, server virtualization has been widely recognized and deployed because it can reduce IT and operation and maintenance (O&M) costs and enable more flexible service deployment.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Objectives

- On completion of this course, you will be able to understand:
  - Principles of VXLAN technology
  - Application scenarios of VXLAN technology
  - Configuration method of VXLAN technology
  - Principles of BGP-EVPN technology
  - Configuration method of BGP-EVPN technology

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

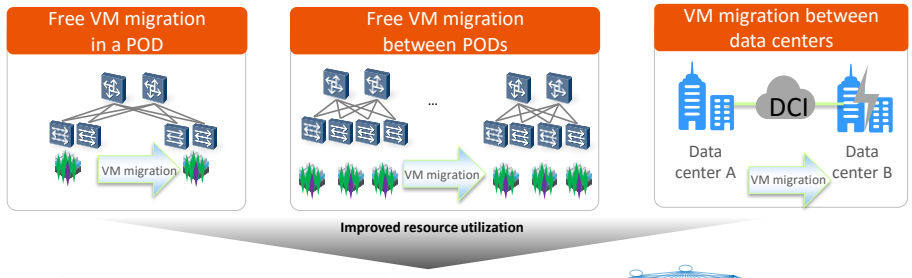
Contents

1. **Requirements and Challenges of Cloud DCs**
2. Basic Principle of VXLAN
3. VXLAN Packet Forwarding
4. Introduction to BGP-EVPN
5. VXLAN-based EVPN Deployment in a DC
6. VXLAN-based EVPN Deployment Between DCs
7. Configuration and Deployment Optimization
8. Application Cases

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

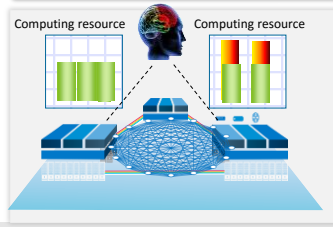
Huawei Training & Certification
Huawei Training & Certification

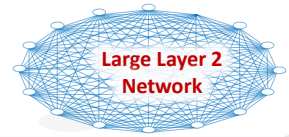
New Network Requirements of Cloud Data Center Services



Improved resource utilization

Objective: improving resource utilization





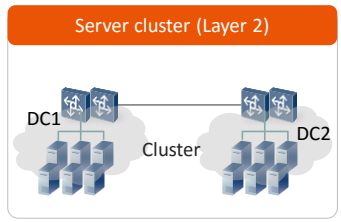
VMs can migrate freely regardless of geographical constraints, building a large Layer 2 network across geographical areas.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

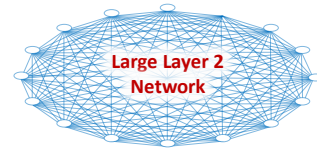
Huawei Training & Certification
Huawei Training & Certification

New Network Requirements of Cloud Data Center Services



Google is a practitioner of building large-scale server clusters and requires a non-blocking network architecture for extensive communication between servers.

- A single server cluster of Google contains about 10,000 severs.
- Internet providers in China plan to build a cluster that contains 20,000 servers.
- The number of interfaces and the network capacity are key to the cluster scale.



Building a server cluster across geographical areas to improve the system disaster recovery capability.

Large Layer 2 network:

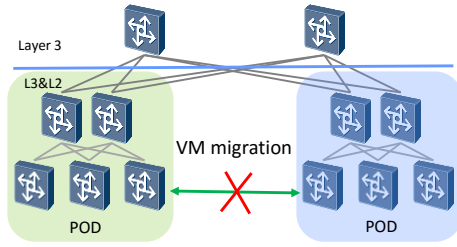
- Large-scale Layer 2 network
- Non-blocking forwarding of west-east traffic

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Why Do Traditional Networks Fail to Expand? (1)



Traditional DC network architecture

Inter-VLAN Layer 3 VM migration is not allowed. In a traditional data center network, gateways are usually deployed on aggregation switches, which communicate with each other through core switches at Layer 3. VMs can only migrate in a POD. If VMs need to migrate to another Layer 2 domain, the IP addresses of the VMs must be changed, and services will be interrupted during VM migration.

?

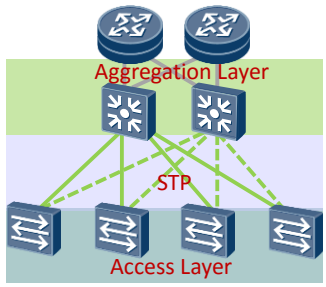
Can a single POD be expanded using existing VLAN technology?

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Why Do Traditional Networks Fail to Expand? (2)



While STP can remove network loops, it has the following disadvantages:

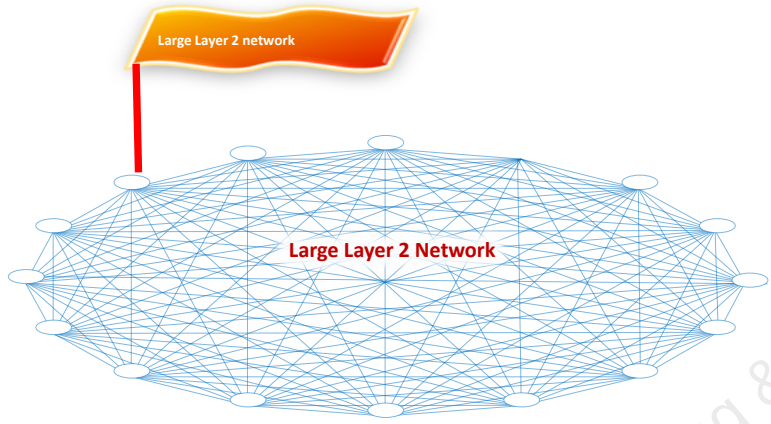
- The STP convergence time is long and an STP-enabled Layer 2 network can contain no more than 50 network nodes. Therefore, STP is unsuitable for a large-scale cloud data center network.
- When STP is used to construct a loop-free network, half of the links must be blocked, leading to low bandwidth usage.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Development Objectives of the Cloud Data Center Network

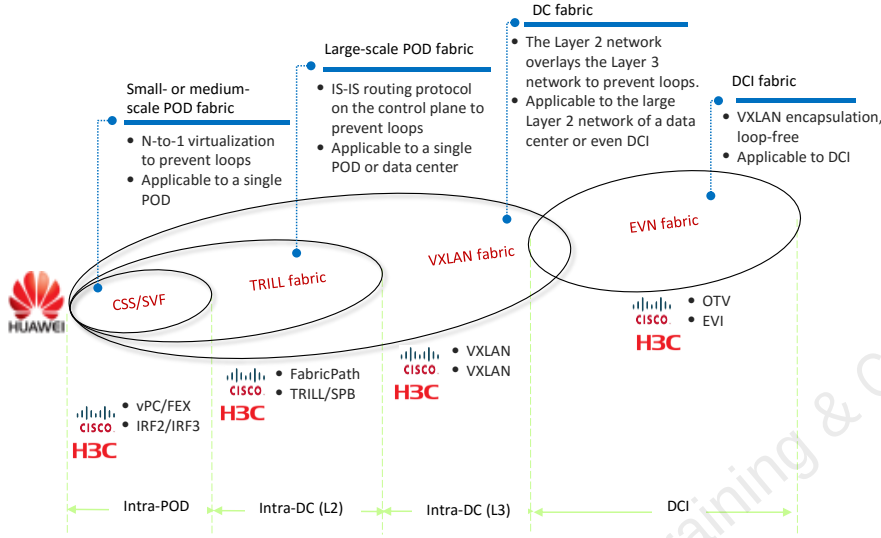


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Diversified Fabric Technologies Underlying Large Layer 2 Networks



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Contents

1. Requirements and Challenges of Cloud DCs
2. **Basic Principle of VXLAN**
3. VXLAN Packet Forwarding
4. Introduction to BGP-EVPN
5. VXLAN-based EVPN Deployment in a DC
6. VXLAN-based EVPN Deployment Between DCs
7. Configuration and Deployment Optimization
8. Application Cases

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

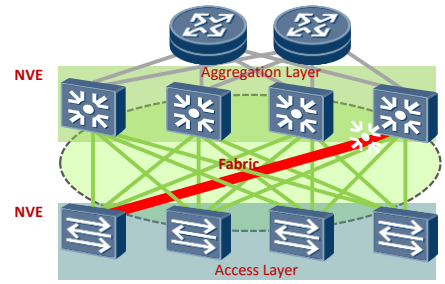
Basic Concepts of VXLAN

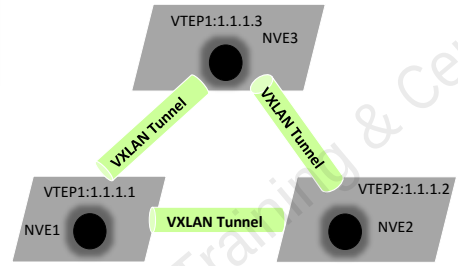
NVo3-based Layer 2 fabric

Network virtualization technology based on a Layer 3 IP overlay network is called NVo3, which is now represented by VXLAN, NVGRE, and STT.
The NVo3-enabled devices are called Network Virtualization Edges (NVEs), which are located at the edge of an overlay network and implement Layer 2 and Layer 3 virtualization.

Virtual Extensible LAN (VXLAN) is now the most influential NVo3 technology. It uses the MAC in UDP packet encapsulation mode to implement IP overlay-based virtual local area network (VLAN).

- A VXLAN Network Virtualization Edge (NVE) is identified by VXLAN tunnel endpoints (VTEPs).
- Each NVE provides at least one VTEP, which is identified by the NVE's IP address.
- A VXLAN tunnel can be established between two VTEPs.

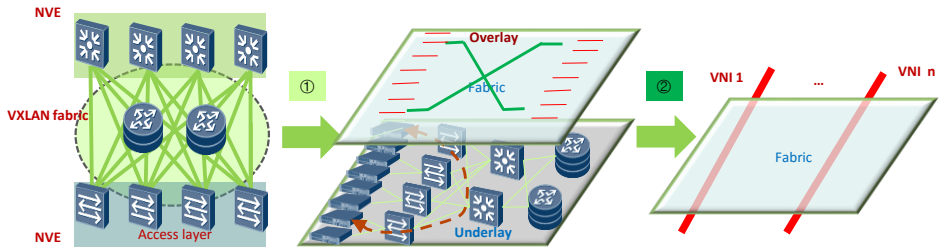




Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

VXLAN Logic Abstraction



VXLAN — Two Rounds of Virtualization

1. Virtualization for the first time: The tunneling technology is used to connect NVEs to transparently transmit Layer 2 packets. The entire network is abstracted as a super large LAN switch with an extended number of interfaces.
2. Virtualization for the second time: VXLAN Network Identifier (VNI), similar to a VLANID, virtualizes the super large LAN switch into multiple Layer 2 broadcast domains. The VNI field in the VXLAN header can be defined to extend the number of subnets from 4 K to 16 M.

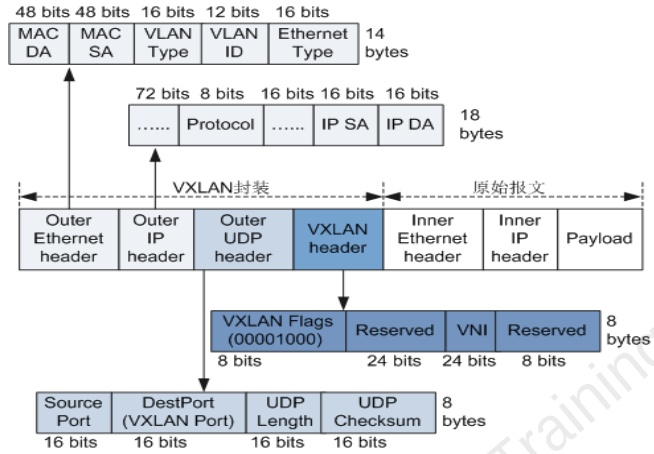
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

VXLAN Packet Format

- VXLAN is a network virtualization technology that uses the MAC-in-UDP mode to encapsulate packets. That is, VXLAN adds a UDP header and a VXLAN header to an Ethernet frame.

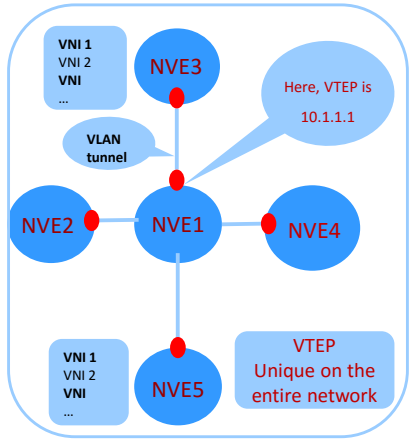


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

VTEP



Basic concept

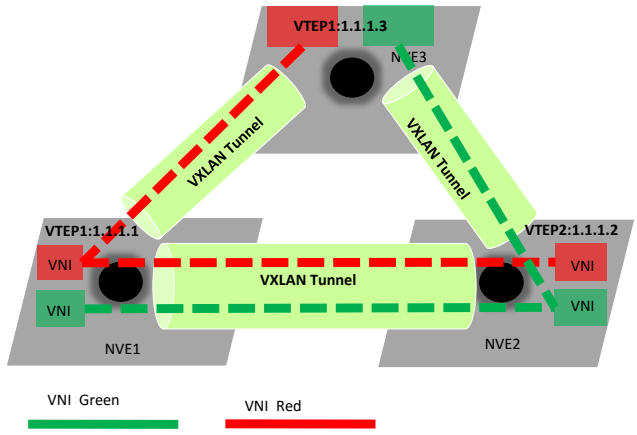
- A VXLAN NVE is identified by VXLAN tunnel endpoints (VTEPs).
- Each NVE provides at least one VTEP, which is identified by the NVE's IP address.
- A VXLAN tunnel can be established between two VTEPs and shared by all VNIs between two NVE nodes.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Relationship Between Tunnel and VNI



VNI concept

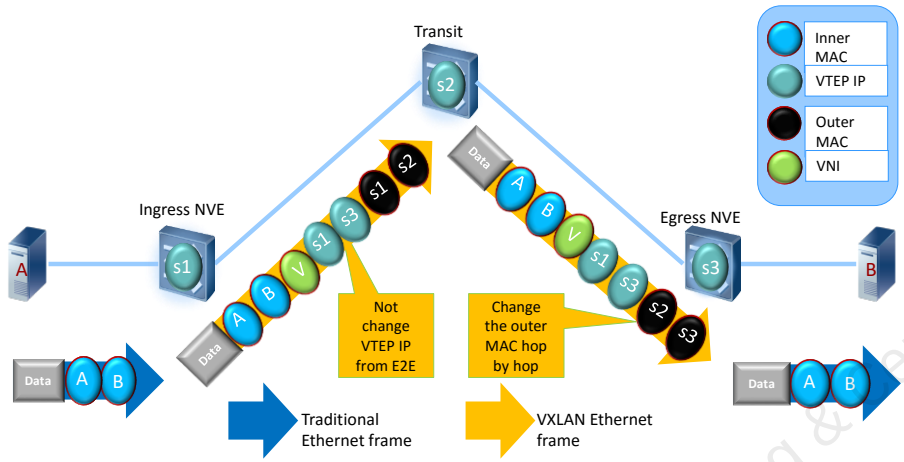
- A VNI identifies a Layer 2 domain on a VXLAN network.
- A VXLAN tunnel can be established between two VTEPs and shared by all VNIs between two NVE nodes.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

VXLAN Forwarding Data Encapsulation



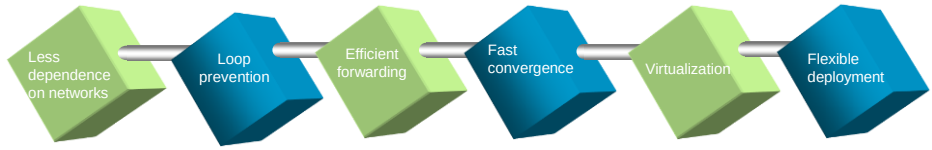
The Layer 2 packets from the source arrive at the destination over the IP network. The VXLAN network functions like a bridge fabric to hosts.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Advantages of VXLAN



It is an IP-based overlay technology requiring only IP reachability between edge devices.	Split horizon of tunnels and IP-based overlay TTL prevent loops.	Data flows are forwarded quickly using shortest path first (SPF) and equal-cost multipath (ECMP) mechanisms of IP routing.	Listen on network topology changes and complete convergence within milliseconds.	The overlay + VNI virtual network architecture supports up to16 million virtual networks.	VXLAN can be deployed on both physical devices and vSwitches.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Comparison Between VXLAN and Other Layer 2 Technologies

	Traditional Layer 2 Technology	CSS+iStack	TRILL	VXLAN
Encapsulation mode	Traditional ETH header (without TTL)	Traditional ETH header (without TTL)	TRILL (with TTL)	MAC in UDP (without TTL)
Loop prevention	MSTP	Management mode	TRILL	Loop prevention through routing protocols + split horizon of tunnels
ECMP	Not supporting ECMP	Supporting ECMP using LAG	Supporting hop-by-hop ECMP, similar to Layer 3 IP networks	Multiplexing ECMP and FRR on Layer 3 networks
Number of multicast trees	NA	NA	Few (Layer 2 RPT)	None or few (IP bidirectional multicast)
Shortest path forwarding	Not supported	Supported	Supported	Supported
Convergence time	Long, unstable convergence time	Short	Medium (hundreds of milliseconds)	Medium (hundreds of milliseconds)
Multi-tenant support	4K (isolated based on VLANs)	4K (isolated based on VLANs)	4K (isolated based on VLANs). In future, tenants can be isolated using FineLabel, and a maximum of 16M tenants will be supported.	16M (isolated based on VNIs)
Networking cost	Low	High (Inter-chassis communication occupies high bandwidth, and non-blocking forwarding is difficult to implement.)	Medium. Generally, devices on the entire network must support TRILL.	Low. Only edge nodes are required to support VXLAN, which can be extended to server vSwitches.
Network scale	Small	Medium (the number of stacked devices is limited, and non-blocking forwarding is not supported)	Large	Large
Applicable network	Applicable to hierarchical networks where the devices at each layer are aggregated to the upper layer, but not applicable to flat tree networks	Applicable to flat tree networks	Applicable to flat tree networks	Applicable to flat tree networks Applicable to SDN networks (edge node awareness, seamless extension to vSwitches)

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

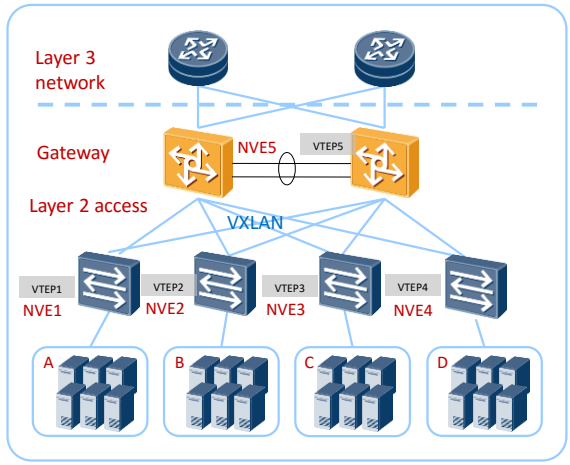
Contents

1. Requirements and Challenges of Cloud DCs
2. Basic Principle of VXLAN
3. **VXLAN Packet Forwarding**
4. Introduction to BGP-EVPN
5. VXLAN-based EVPN Deployment in a DC
6. VXLAN-based EVPN Deployment Between DCs
7. Configuration and Deployment Optimization
8. Application Cases

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Traffic Forwarding Process Within a VXLAN Subnet



Working Process

- Host A sends an ARP Request packet to Host B.
- Host B responds with an ARP Reply packet to Host A.
- Host A sends a unicast data packet to Host B.

Note: Host A, Host B, Host C, and Host D belong to the same VNI 1. ARP broadcast optimization is disabled.

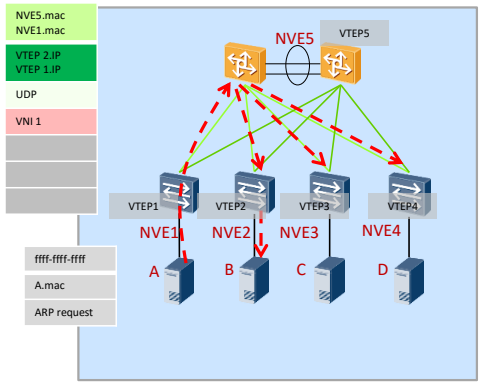
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Layer 2 Forwarding Based on MAC Addresses on the Same Network Segment (1)

ARP Request broadcast packet from A to B



1. NVE1 broadcasts the ARP Request packet in the VNI bridge domain (BD). The packet is encapsulated and forwarded over a tunnel.
2. The overlay packet is transparently transmitted on intermediate nodes.
3. NVE2/3/4/5 receive the packet, decapsulate it, and broadcast the original packet in local VNI 1.
   NVE2/3/4/5 learn the server A's MAC address.

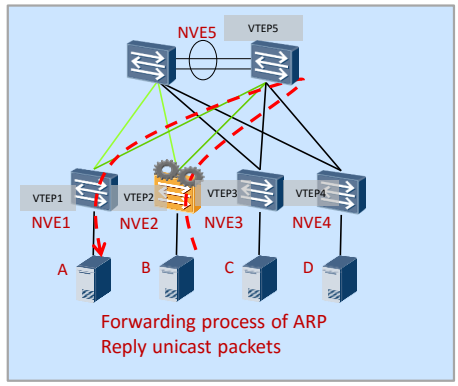
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Layer 2 Forwarding Based on MAC Addresses on the Same Network Segment (2)

ARP Reply unicast packet from B to A



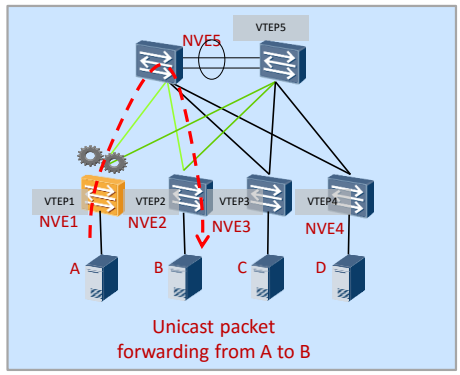
1. NVE2 searches the server A's MAC address forwarding table and finds that the matching outbound interface is a VXLAN tunnel. (from NVE2 to NVE1). The encapsulated packet is forwarded at Layer 3.
2. The overlay packet is transparently transmitted on intermediate nodes.
3. NVE1 receives and decapsulates the packet, and forwards it locally. NVE1 learns the server B's MAC address.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Layer 2 Forwarding Based on MAC Addresses on the Same Network Segment (3)

Unicast packet from A to B



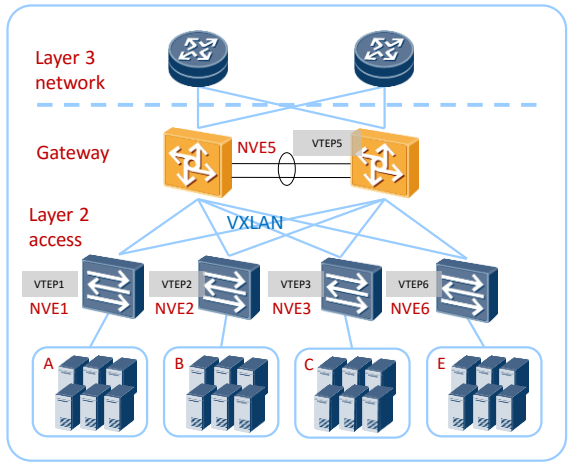
1. Both NVE1 and NVE2 have learned MAC addresses of server A and server B. Corresponding entries can be found in the MAC address table. The unicast packet forwarding process is the same as that in a VLAN.
2. The only difference lies in that VXLAN header is encapsulated to the packet. IP forwarding applies to underlay networks.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

VXLAN Data Forwarding Process (Inter-Subnet)



Working Process

Host A sends a unicast packet to Host E.

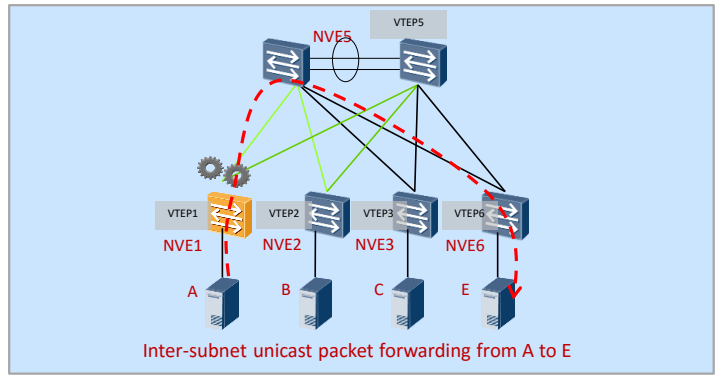
Note: NVE5 functions as a Layer 3 gateway. Host A belongs to VNI 1 and Host E belongs to VNI 2. By default, hosts and the gateway have learned ARP entries and MAC addresses of all nodes.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Host A sends a unicast packet to Host E.
- Note: NVE5 functions as a Layer 3 gateway. Host A belongs to VNI 1 and Host E belongs to VNI 2. By default, hosts and the gateway have learned ARP entries and MAC addresses of all nodes.

Huawei Training & Certification

Unicast Packet Forwarding from A to E



Inter-subnet unicast packet forwarding from A to E

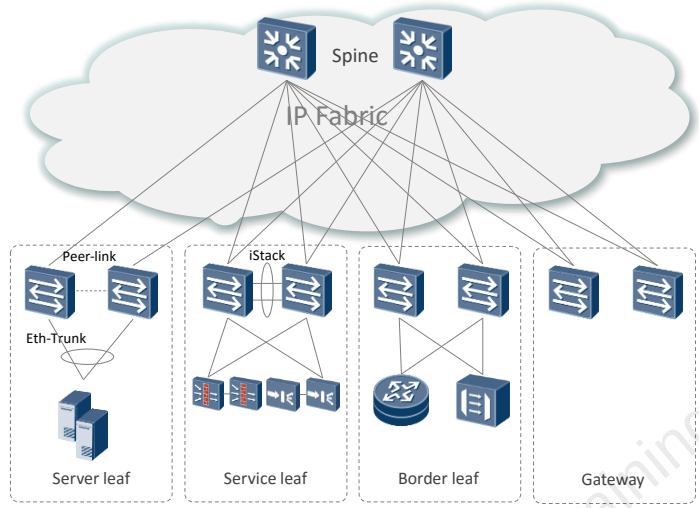
1. NVE1 searches the MAC address forwarding table of the gateway, encapsulates the packet using VNI 1, and sends the packet over the tunnel.
2. The gateway decapsulates the packet, searches for the route based on the inner IP header, and replaces the inner Ethernet header with a VXLAN header using VNI2.
3. NVE6 receives and decapsulates the packet, and forwards the inner packet according to the destination MAC address.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

VXLAN Networking Based on a Spine-Leaf Architecture

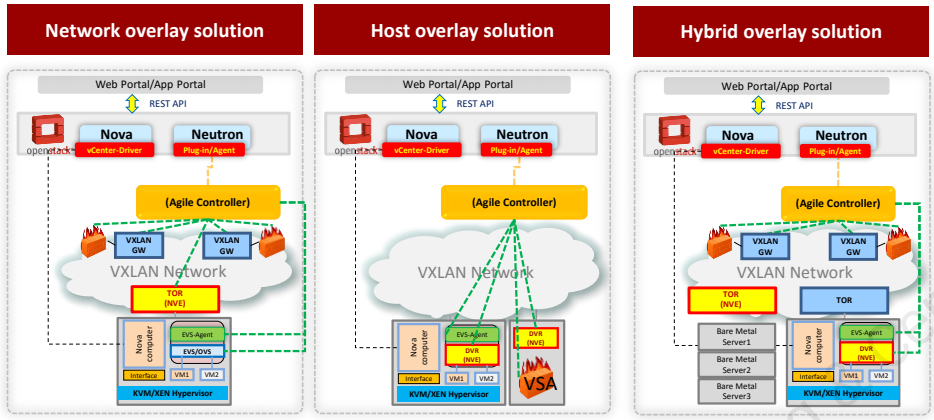


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Three VXLAN Overlay Networking Solutions



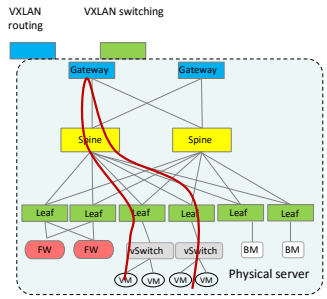
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Software solution:
  - Large Layer 2 logical networks can be automatically created without reconstructing existing physical devices, decoupling from hardware devices from other vendors, or configuring physical networks.
- Hardware solution:
  - New physical networks are required to form VXLAN overlay networks for automatic service provisioning.
- Hybrid solution:
  - SDN is used to configure or manage virtual and physical networks (switches, firewalls, and F5s) and implement automatic service provisioning.

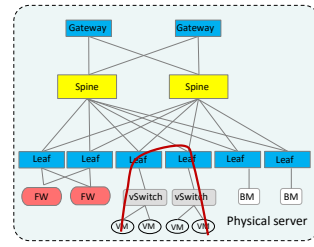
Huawei Training & Certification

Distributed and Centralized VXLAN Overlay Networks



Centralized network overlay gateway solution

- Layer 2 VXLAN VTEPs are deployed on leaf nodes.
- Layer 3 VXLAN gateways are deployed at the core layer.
- Spine: common routes



Distributed network overlay gateway solution

- Leaf nodes are both VXLAN Layer 2 VTEP gateways and Layer 3 VXLAN gateways for east-west traffic.
- The gateways for north-south traffic are deployed at the core layer.
- Spine: common routes

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

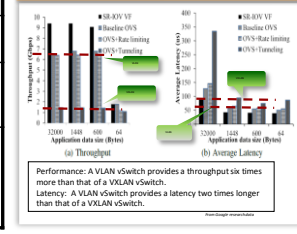
HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Network Overlay Is the Most Mature Solution

Comparison/Solution	vSwitch Functioning as a VTEP	TOR Functioning as a VTEP
Performance	The forwarding performance decreases by 30% to 40%.	No impact
Compatibility	Adapt to multiple virtualization platforms with difficulties in implementation.	Easy to adapt to multiple virtualization platforms.
Management	The O&M interfaces are separated because both hardware and software devices forward service traffic. Two interfaces need to be maintained.	Management is consistent, and the O&M interface is on TOR switches.

Better forwarding performance of TORs functioning as VTEPs



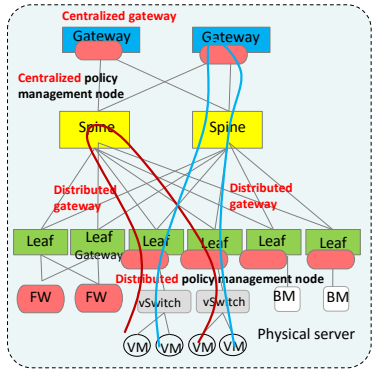
Fabric Solution	Solution Characteristics	Description
Hardware VXLAN solution	A VXLAN fabric is composed of hardware devices only.	1. Provides access to multiple types of servers (virtual servers and bare metal servers). 2. Supports Hypervisor of any types. 3. Does not occupy server CPU resources and provides high forwarding performance of hardware devices.
Software VXLAN solution	A VXLAN fabric consists of software switches.	1. Only connects to virtual servers. 2. Occupies CPU resources and implements software forwarding with poor performance. 3. A vSwitch supports limited types of Hypervisors. For example, VMware does not provide open Hypervisor vSwitch interfaces, and the Hyper-V/Docker vSwitch in the industry is immature. 4. The vSwitch is between the IT and the network, leading to separate O&M interfaces.
Hybrid VXLAN solution	A VXLAN fabric is composed of both OVSs (software switches) and physical switches.	1. Supports hybrid access of physical servers and virtual servers. 2. Same as features 2 and 3 of the software solution.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

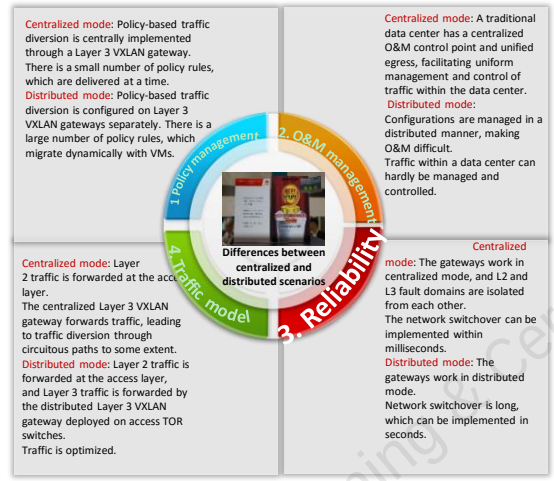
Huawei Training & Certification
Huawei Training & Certification

Hardware Centralized Networking vs. Hardware Distributed Networking



Customers select the hardware centralized VXLAN or distributed VXLAN solution

based on the application scenario (traffic model, reliability requirements, and O&M habits).

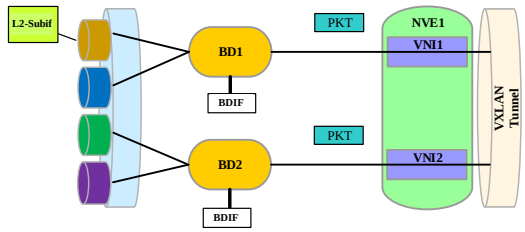


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Note: Amphenol Connector Systems (ATCS)

Huawei Training & Certification

VXLAN Gateway Service Model (1)



VXLAN gateways use the EVC service model, which contains the following components: Bridge Domain (BD), VNI, Network Virtualization Edge (NVE), Layer 2 sub-interface (L2-Subif), and VXLAN tunnel.

- L2-Subif

  It is used for user access. The sub-interface can be configured to forward singe-tagged or untagged packets.

- BD

  It identifies a Layer 2 broadcast domain with a 1:1 mapping relationship between BDs and VNIs. All broadcast domain functions are supported in a BD, such as MAC address learning, Layer 2 table lookup, and broadcast replication.

- NVE

  It is mainly used for local VTEP address management, VXLAN tunnel management, and ingress replication list management.

- VXLAN tunnel

  It is used to forward VXLAN packets and is identified by a pair of VTEP addresses (local + remote VTEPs).

- BDIF

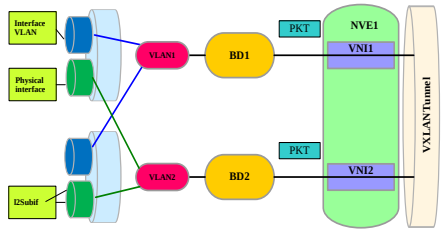
  It is a Layer 3 routing interface in the BD that forwards Layer 2 traffic at Layer 3.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

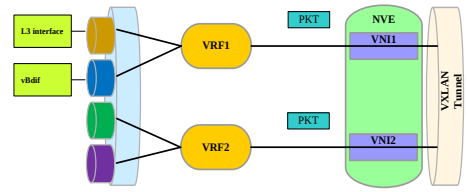
HUAWEI

Huawei Training & Certification
Huawei Training & Certification

VXLAN Gateway Service Model (2)



Global VLAN access model: It applies to L2VPN service scenarios. VLANs are bound to BDs. Traditional interfaces and VLAN interfaces can be connected to a VXLAN network. Layer 2 sub-interfaces are bound to BDs.



VRF access model: It applies to L3VPN service scenarios. VRFs are bound to VNIs. Using the BGP protocol, the outbound interface of the VXLAN tunnel iterates routing entries to provide the L3VPN service capability.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- bridge-domain 1l2 binding vlan 1

Huawei Training & Certification

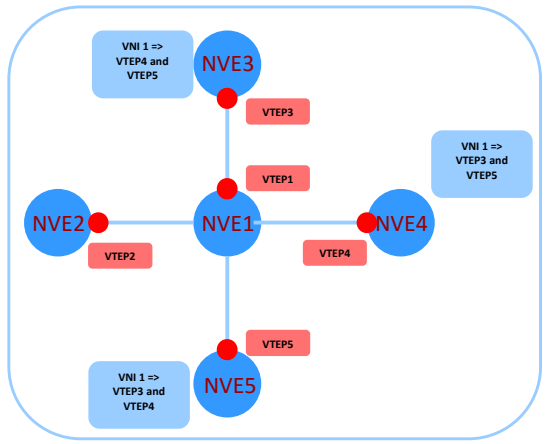
Contents

1. Requirements and Challenges of Cloud DCs
2. Basic Principle of VXLAN
3. VXLAN Packet Forwarding
4. **Introduction to BGP-EVPN**
5. VXLAN-based EVPN Deployment in a DC
6. VXLAN-based EVPN Deployment Between DCs
7. Configuration and Deployment Optimization
8. Application Cases

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Disadvantages of VXLAN (1)



Tunnel establishment

- The VXLAN standards do not contain protocols that define the method of tunnel establishment.
- The VXLAN tunnels between devices can be established through static configurations or commands on the controller.

BUM flood

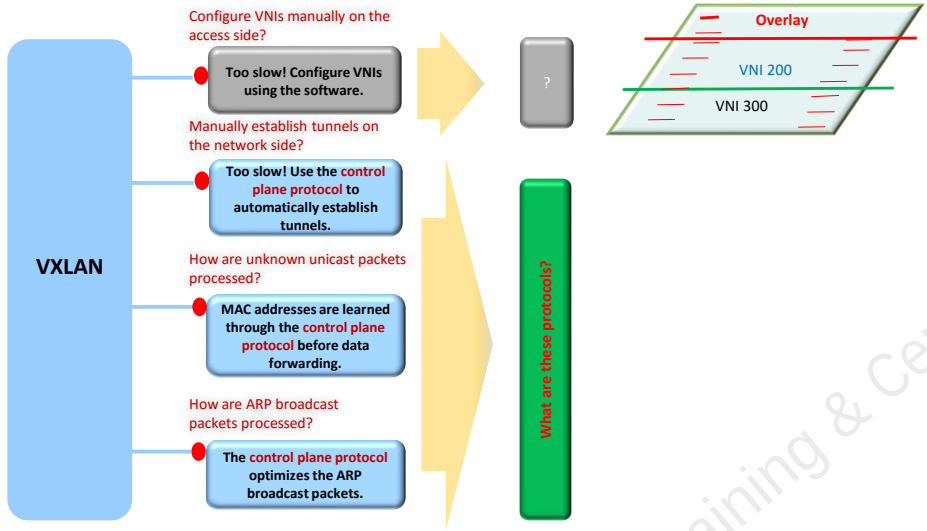
- BUM packet flood occupies bandwidth and consumes network resources.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Disadvantages of VXLAN (2)



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

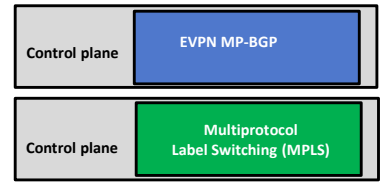
HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Introduction to BGP-EVPN

EVPN Basic Concepts

- Ethernet VPN, defined in RFC-7432, is generated to meet requirements for high bandwidth and complex QoS in the MPLS-based VPN network. The MP-BGP protocol is applied to the control plane.
- Characteristics of EVPN: The control plane and data plane are abstracted and isolated from each other. The MP-BGP control plane transmits MAC addresses and IP routes. There are several methods for encapsulation at the data plane.



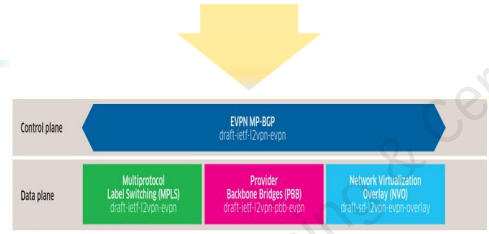
EVPN for VXLAN

- Since VXLAN has no control plane, the MP-BGP control plane of EVPN is suitable for VXLAN.
- VXLAN extends the standard EVPN.
- The following three drafts are provided:

  draft-ietf-bess-evpn-overlay

  draft-ietf-bess-evpn-inter-subnet-forwarding

  draft-ietf-bess-evpn-prefix-advertisement



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

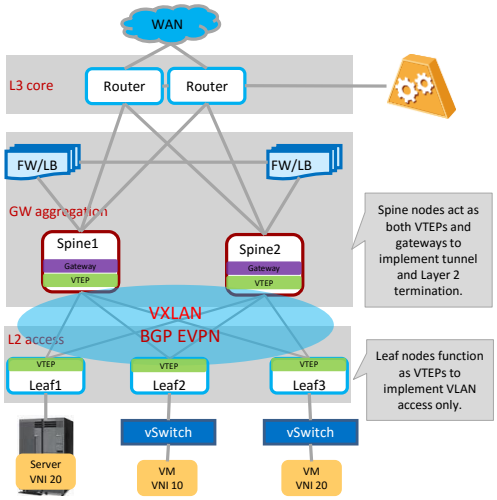
Contents

1. Requirements and Challenges of Cloud DCs
2. Basic Principle of VXLAN
3. Common Networking of VXLAN
4. Introduction to BGP-EVPN
5. **VXLAN-based EVPN Deployment in a DC**
6. VXLAN-based EVPN Deployment Between DCs
7. Configuration and Deployment Optimization
8. Application Cases

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Centralized Gateway Deployment



1. The leaf and spine nodes set up full-mesh connections. At least two spine nodes need to be deployed on the network to implement backup and load balancing. Forwarding paths are established between spine and leaf nodes through L3 routing. Normally, ECMP paths exist between the spine and leaf nodes.

2. The spine and leaf nodes all act as VTEPs.

3. A leaf node only implements L2 traffic forwarding. It maps the VLAN tag of user packets to a VNI, encapsulates the packets in VXLAN packets, and forwards them to the destination node in the local broadcast domain.

4. A spine node acts as a centralized L3 gateway to decapsulate VXLAN packets and forward them based on L3 routes, enabling communication between broadcast domains on the network and communication with external networks.

BGP EVPN acts as the VXLAN control plane to provide the following functions:

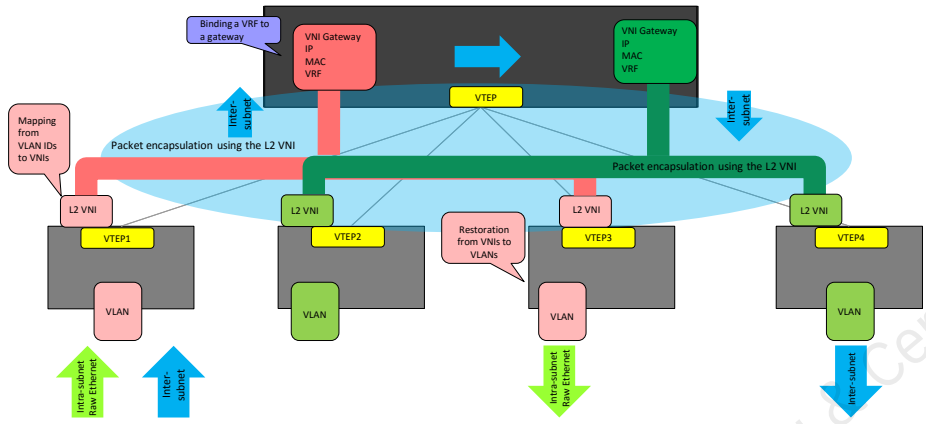
- Triggers automatic VXLAN tunnel establishment between VTEPs to avoid the need to manually configure full-mesh tunnels.
- Advertises MAC address tables to prevent flooding of unknown traffic.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The centralized Layer 3 gateway mode is similar to the traditional VLAN network, and is not the focus in this presentation.

Huawei Training & Certification

Centralized Gateway VXLAN Service Model



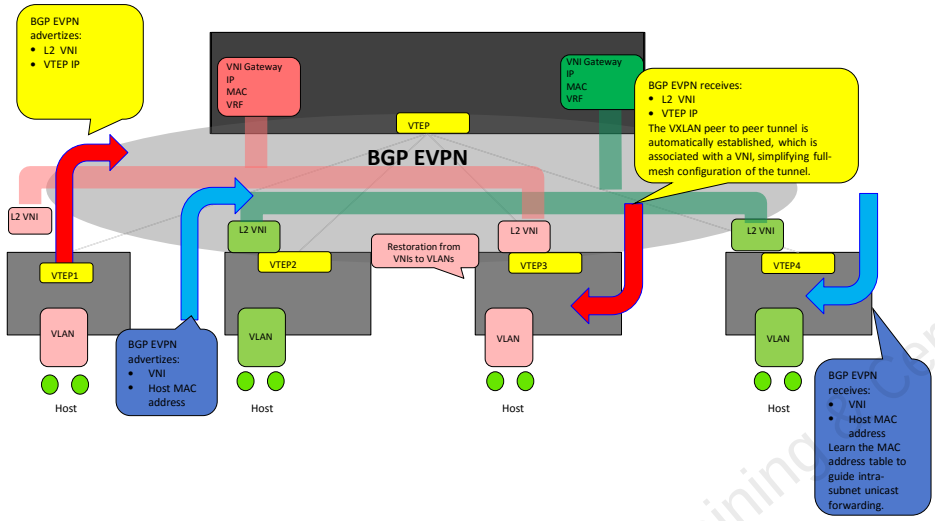
- VNIs identify Layer 2 domains and VXLAN tunnels enable intra-domain connectivity.
- Inter-subnet traffic is forwarded to the gateway based on the MAC address, and then mapped to a VRF through the gateway.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The service model complies with the following RFCs and drafts:
- RFC7348 VXLAN encapsulation
- RFC 7432 EVPN
- draft-ietf-bess-evpn-overlay

Huawei Training & Certification

BGP EVPN Protocol Packet Exchange in Centralized Gateway Deployment

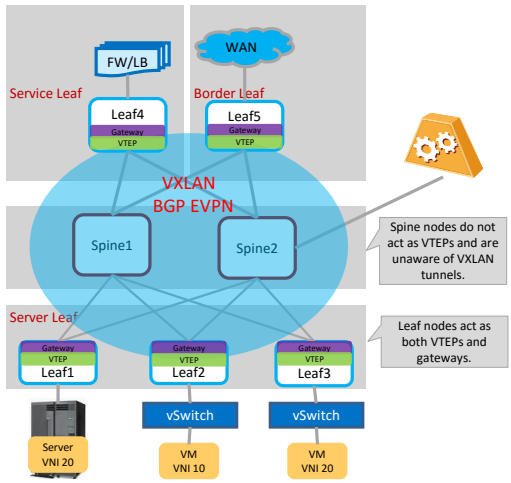


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Distributed Anycast gateway



1. The leaf and spine nodes set up full-mesh connections. At least two spine nodes need to be deployed on the network to implement backup and load balancing. Forwarding paths are established between spine and leaf nodes through IGP routing. Normally, ECMP paths exist between the spine and leaf nodes.
2. Leaf nodes act as VTEPs, whereas spine nodes are not necessarily VTEPs.
3. Each leaf node acts as an L3 gateway to terminate L3 traffic for local users.
4. Different leaf nodes provide different functions. For example, some leaf nodes connect to external networks, and some leaf nodes connect to firewalls for traffic filtering.

BGP EVPN acts as the VXLAN control plane to provide the following functions:

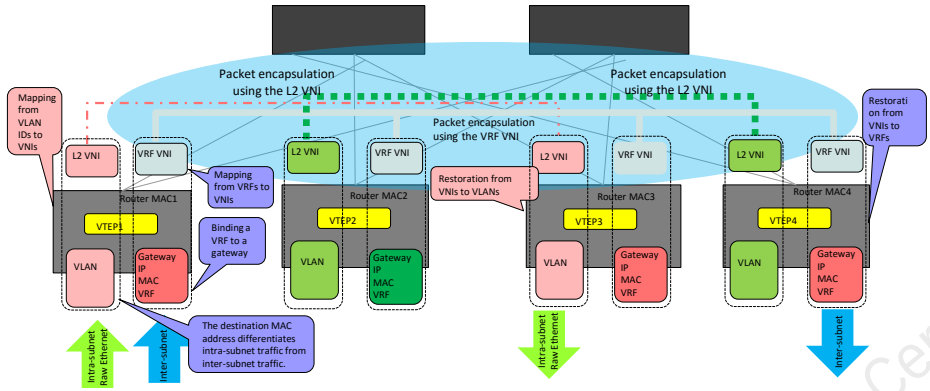
- Triggers automatic VXLAN tunnel establishment between VTEPs to avoid the need to manually configure full-mesh tunnels.
- Advertises host routes and MAC addresses to direct traffic forwarding.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Distributed VXLAN Service Model



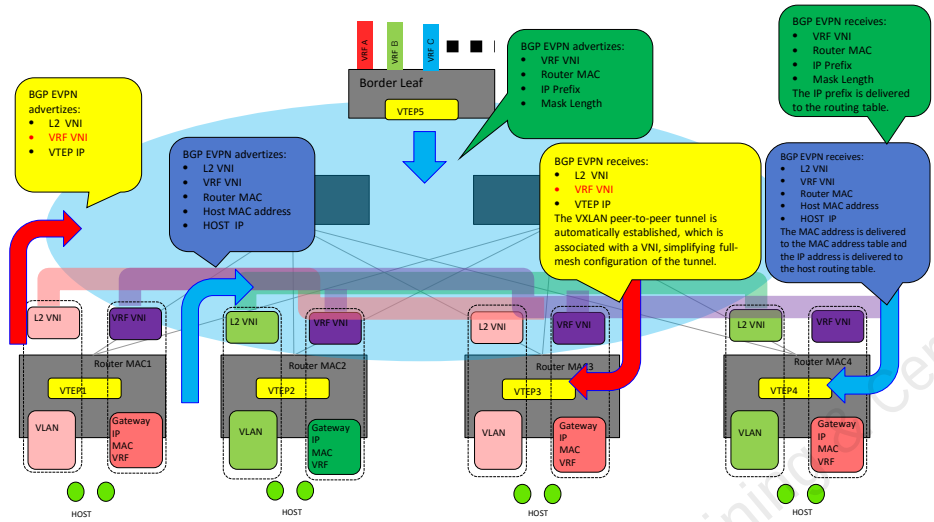
- VNIs identify Layer 2 domains and VXLAN tunnels enable intra-domain connectivity.
- VNIs identify Layer 3 VRF domains and VXLAN tunnels enable inter-VRF connectivity.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The service model complies with the following RFCs and drafts:
- RFC7348 VXLAN encapsulation
- RFC 7432 EVPN
- draft-ietf-bess-evpn-overlay
- draft-ietf-bess-evpn-inter-subnet-forwarding
- draft-ietf-bess-evpn-prefix-advertisement

Huawei Training & Certification

BGP EVPN Protocol Packet Exchange in Distributed Anycast Gateway Deployment

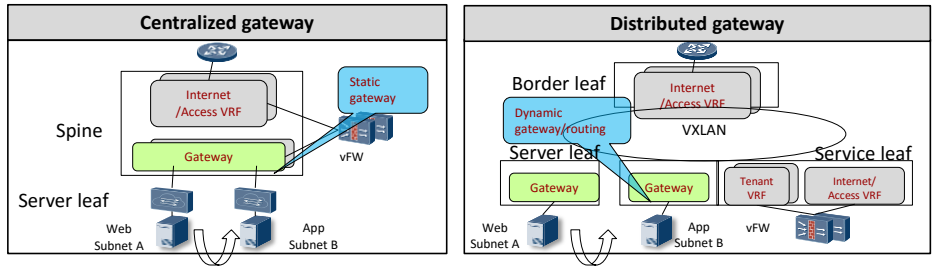


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Gateway Deployment Comparison

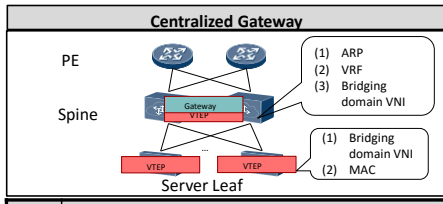


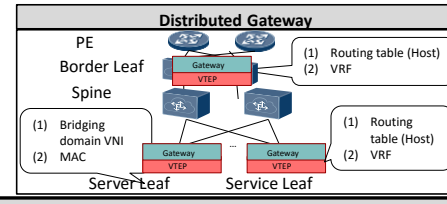
Item	Centralized Deployment	Distributed Deployment
Gateway deployment change upon VM migration	The gateways keep unchanged when VMs migrate.	The gateways migrate dynamically with VM migration. Delete gateway configurations from source access devices and create gateways on destination access devices.
Entries affected by VM migration	ARP entries are updated on the centralized gateway. MAC addresses are updated on the access devices.	Host routes are updated on all devices. MAC addresses are updated on the access devices.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Comparison of Entry Specifications



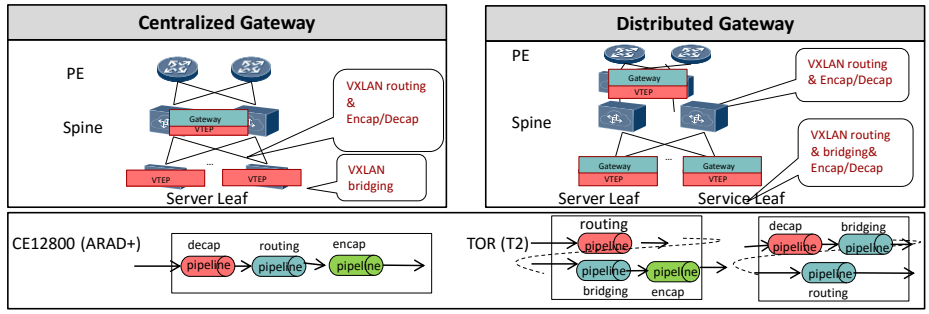


Item	Centralized Deployment	Distributed Deployment
Spine	(1) VRF: All VRFs (CE12800: a maximum of 4K VRFs) (2) Route: Subnet routes + exterior routes + Underlay routes (CE12800: a maximum of 1 M routes) (3) VNI: All VNIs (CE12800: a maximum of 4K VNIs) (4) ARP: All VMs x Number of IP addresses for VMs (CE12800: a maximum of 1M ARP entries) (5) Tunnel: Number of leaf nodes (CE12800: a maximum of 8K tunnels)	(1) VRF: All VRFs (CE12800: a maximum of 4K VRFs) (2) Route: All subnet routes + exterior routes + Underlay routes + all VM host routes (CE12800: a maximum of 1M routes) (3) VNI: Number of VNIs in a VRF (CE12800: a maximum of 4K VNIs) (4) ARP: 0 (CE12800: a maximum of 1M ARP entries) (5) Tunnel: Number of leaf nodes (CE12800: a maximum of 8K tunnels)
Leaf	(1) Route: Underlay route (CE6800: a maximum of 256K routes) (2) VNI: VNIs of connected VMs (CE6800: a maximum of 4K VNIs) (3) Tunnel: Number of leaf nodes (CE6800: a maximum of 15K tunnels) (4) ARP: 0 (CE6800: a maximum of 128K ARP entries) (5) MAC: Number of VMs in a VNI x Number of vNICs (CE6800: a maximum of 288K MAC addresses)	(1) VRF: VRFs to which connected VMs belong (CE6800: a maximum of 1K VRFs) (2) Route: VM subnets + host routes in a VRF (CE6800: a maximum of 256K routes) (3) VNI: VNIs of connected VMs (CE6800: a maximum of 4K VNIs) (4) ARP: Connected VMs x Number of IP addresses for VMs (CE6800: a maximum of 128K ARP entries) (5) Tunnel: Number of leaf nodes (CE6800: a maximum of 15K tunnels) (6) MAC: Number of VMs in a VNI x Number of vNICs (CE6800: a maximum of 288K MAC addresses)

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Forwarding Performance Comparison



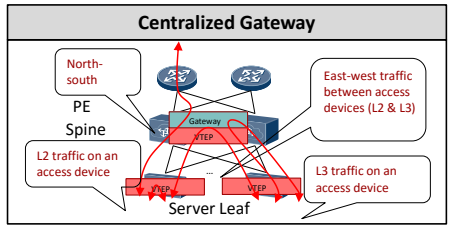
Item	Centralized Deployment	Distributed Deployment
Forwarding performance of spine nodes	Line rate ARAD + decapsulation, routing, and encapsulation through a pipeline	Line rate ARAD + decapsulation, routing, and encapsulation through a pipeline
Forwarding performance of leaf nodes	Line rate T2 completes decapsulation, Layer 2 bridging, and encapsulation through a pipeline.	The performance is halved. T2 completes decapsulation, routing, and VXLAN encapsulation through a loopback.

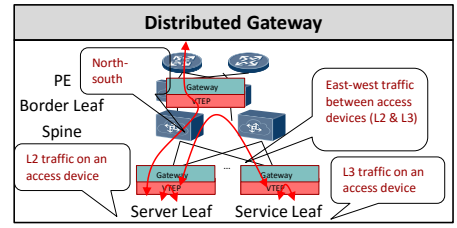
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Comparison of Forwarding Path Optimization





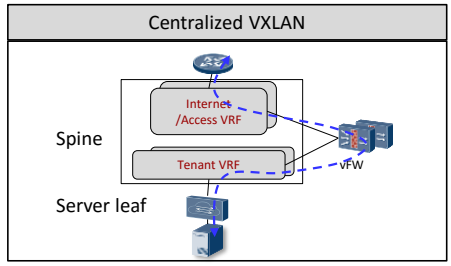
Item	Centralized Deployment	Distributed Deployment
North-south traffic (customer to servers)	Passing through the core	Passing through the core
East-west traffic between different access devices (L2 & L3)	Passing through the core	Passing through the core
East-west traffic on an access device (L2)	Local forwarding	Local forwarding
East-west traffic on an access device (L3)	Passing through the core	Local forwarding

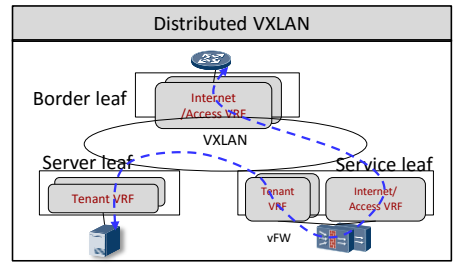
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Comparison of Firewall Traffic Filtering





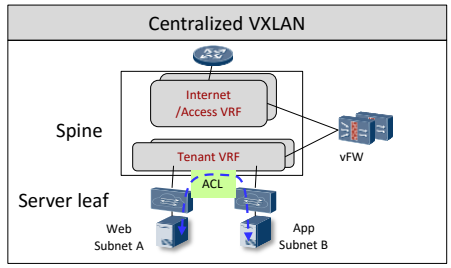
Item	Centralized Deployment	Distributed Deployment
Traffic import to firewalls	Firewalls connect to the centralized gateway in bypass mode. Policies are deployed on the centralized gateway to import traffic to firewalls.	Firewalls are not directly connected to gateways. Multiple gateways connect to firewalls through VXLAN tunnels. Policies are deployed on multiple distributed gateways.

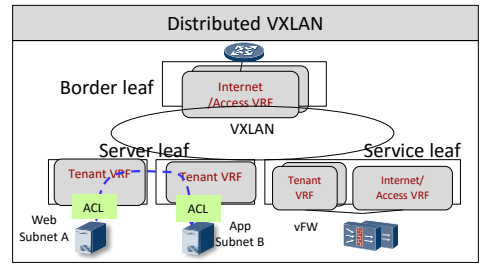
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Comparison of ACL Traffic Mapping





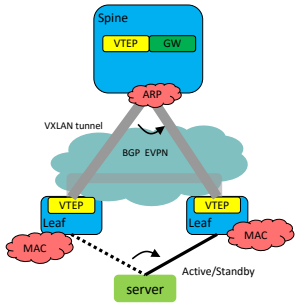
Item	Centralized Deployment	Distributed Deployment
ACL deployment location	ACL deployment on the centralized gateway.	ACL deployment on multiple distributed gateways
ACL specifications	All ACL rules are deployed on the centralized gateway (CE12800 supports a maximum of 10K ACL rules).	ACLs are deployed on access devices separately (TOR switches support a maximum of 3K ACL rules).

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

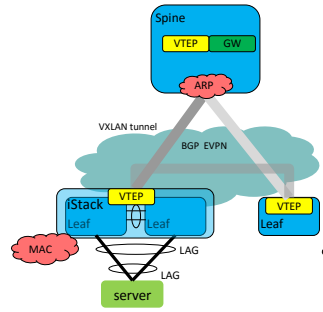
HUAWEI

Huawei Training & Certification
Huawei Training & Certification

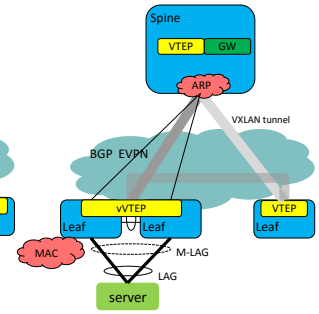
High Reliability of Access



Active and standby NICs of the server connect to switches.
The traditional high reliability mechanism adopts 1:1 protection to ensure sufficient bandwidth for services in case of the NIC fault.
Low link utilization



Two switches set up a stack and connect to bundled NICs of the server through a LAG.
In 1+1 protection mechanism, the bandwidth may be insufficient in case of link faults.
High link utilization
The number of logical NEs is small, reducing configuration workload.

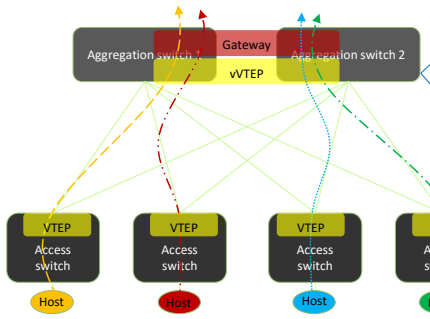


Two separate switches connect to bundled NICs of the server through an M-LAG.
In 1+1 protection mechanism, the bandwidth may be insufficient in case of link faults.
High link utilization
Independent devices, simple upgrade, and high reliability

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Highly Reliable Centralized Gateway



Multi-gateway:
Configure the same loopback address as the VTEP IP address.
Configure interfaces on gateways with the same IP address and MAC address.

1. When a gateway is faulty, it advertises low-preference routes (adjusted by the Agile Controller) to ensure that uplink and downlink traffic is sent to other gateways. After the gateway's ARP entries are re-learned, the route preference is restored and the gateway works in load balancing mode again.
2. When downlink interfaces of a gateway are faulty, the underlay routes are summarized.
3. If all uplink interfaces of a gateway are faulty, the OPS lowers the preference of the gateway VTEP address so that uplink traffic will not be sent to the faulty gateway.

MAC and IP routing tables on access devices

Destination MAC address	Remote VTEP
Gateway virtual MAC address	vVTEP

Destination VTEP route	Next-hop address
vVTEP IP	AGG1& AGG2

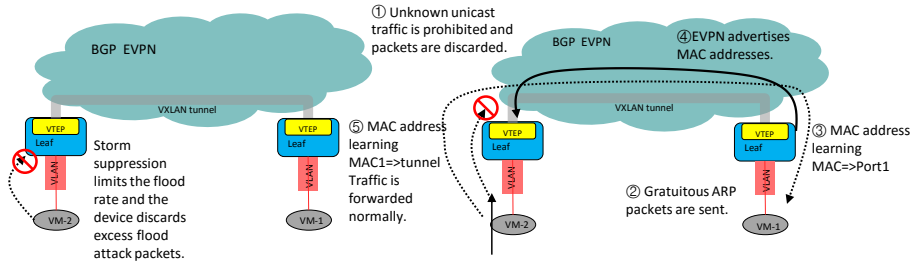
Compared with the traditional stack and VRRP gateways, the centralized gateway has the following advantages:
Gateway devices do not need to run any protocols that require subnet-based heartbeat, such as VRRP and GLBP, greatly reducing the workload on signaling packet processing.
Compared with VRRP, VXLAN supports flow-based load balancing among gateways, which can be extended.
The Agile Controller advertizes vVTEP routes to implement lossless gateway capacity expansion or upgrade.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Storm Suppression and Unknown Unicast Prohibition



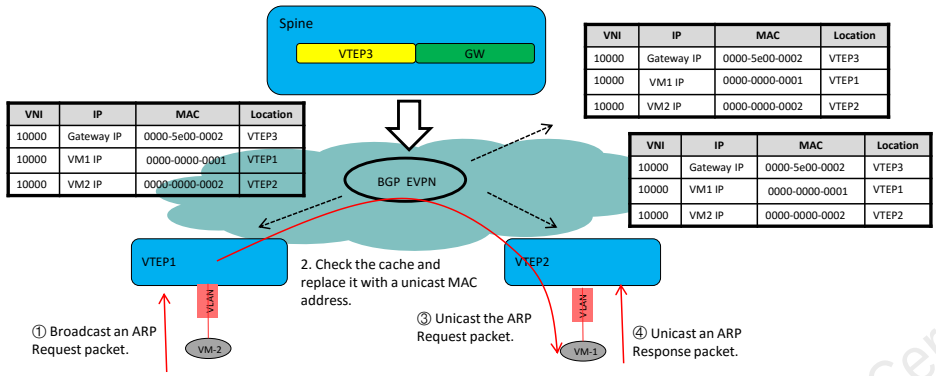
Storm suppression
When a network is running normally, the BUM flooded traffic is few (mainly ARP broadcast packets).
Storm suppression can be deployed on access interfaces of switches to reduce exceptional traffic floods on the VXLAN network and minimize the impact on host performance.

Unknown unicast prohibition
On a traditional network, MAC addresses are learned through traffic floods. After EVPN is used to advertise MAC addresses, traffic does not need to be broadcast to other switches after being learned locally. MAC addresses can be pushed to other switches using EVPN.
Flooding of unknown unicast packets can be prohibited.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

ARP Broadcast Optimization (Broadcast to Unicast)



VNI	IP	MAC	Location
10000	Gateway IP	0000-5e00-0002	VTEP3
10000	VM1 IP	0000-0000-0001	VTEP1
10000	VM2 IP	0000-0000-0002	VTEP2

VNI	IP	MAC	Location
10000	Gateway IP	0000-5e00-0002	VTEP3
10000	VM1 IP	0000-0000-0001	VTEP1
10000	VM2 IP	0000-0000-0002	VTEP2

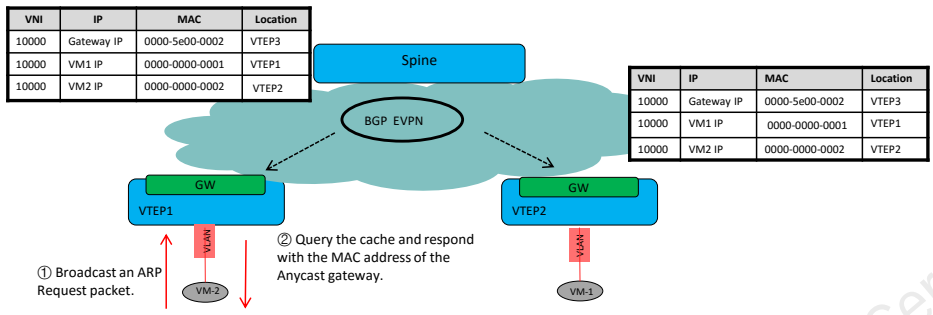
VNI	IP	MAC	Location
10000	Gateway IP	0000-5e00-0002	VTEP3
10000	VM1 IP	0000-0000-0001	VTEP1
10000	VM2 IP	0000-0000-0002	VTEP2

Compared with traditional proxy ARP, this mechanism has the following advantages:
Network devices will not terminate ARP requests and the security risk is low.
Problems may occur if ARP packets are terminated on network devices because ARP now provides more applications such as NIC Teaming detection and cluster heartbeat detection in addition to address resolution.
The destination device can receive the ARP Request packet. The ARP exchange process is complete with good protocol compatibility.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

ARP Broadcast Optimization (ARP Proxy)



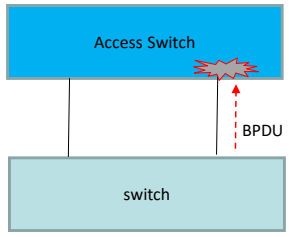
Use the MAC address of the access Anycast gateway to perform spoofing response.
IP packets on the same subnet can be forwarded through Layer 3 routing between VTEPs. This minimizes Layer 2 traffic forwarding, reducing the traffic flood risk.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

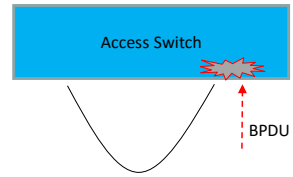
HUAWEI

Huawei Training & Certification
Huawei Training & Certification

BPDU Protection Prevents Incorrect Connection Between Switch Interfaces



After an access switch receives BPDUs, it shuts down interfaces to eliminate loops.



After an access switch receives BPDUs, it shuts down interfaces to eliminate loops.

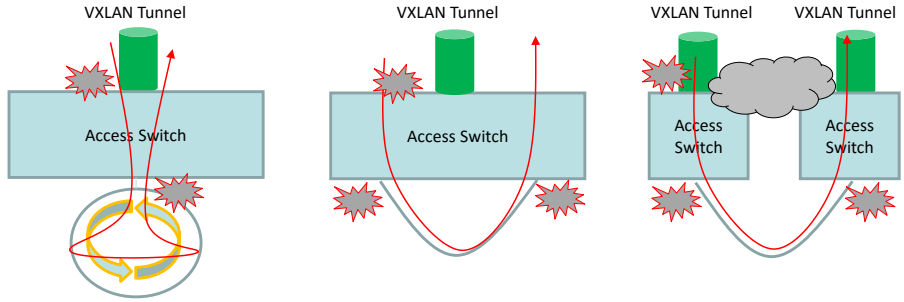
- VXLAN is a loop-free network because split horizon between tunnels prevents loops on the VXLAN network.
- If access devices are incorrectly connected, loops may occur, affecting all switches on the VXLAN network.
- By default, STP is enabled on a switch interface, from which BPDUs are periodically sent. Servers, however, do not send BPDUs. If an access switch receives BPDUs, switches must have been connected incorrectly. BPDU Protection can be deployed so that the interface shuts down once it receives a BPDU, preventing loops.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

MAC Address Flapping Detection Prevents Incorrect Linux Bridge Connection



- MAC address flapping detection can be deployed on downlink interfaces of a switch. When the switch detects loops on an interface, the interface is shut down.
- The VXLAN tunnel will not be disabled, without affecting traffic forwarding on other interfaces.
- The Agile Controller can accurately control the shutdown of interfaces by receiving MAC address flapping alarms from devices.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

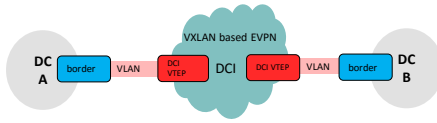
Contents

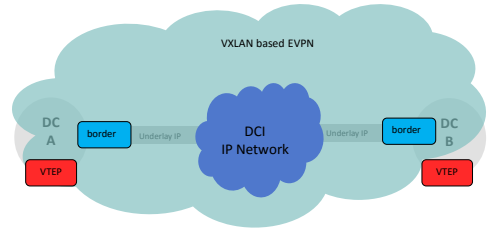
1. Requirements and Challenges of Cloud DCs
2. Basic Principle of VXLAN
3. Common Networking of VXLAN
4. Introduction to BGP-EVPN
5. VXLAN-based EVPN Deployment in a DC
6. **VXLAN-based EVPN Deployment Between DCs**
7. Configuration and Deployment Optimization
8. Application Cases

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

VXLAN-based Layer 2 DCI





Method 1:
Segmented: VTEPs in a data center and DCI VTEPs are connected through VLANs, and DCI VTEPs set up VXLAN tunnels to implement DCI.
Advantages:
- Different data centers can run different protocols.
- Data centers do not need to use the same encapsulation mode and can have different architectures.
- Data centers are connected in P2P mode logically, which facilitate bandwidth control, policy control, and broadcast storm suppression at ingress.

Disadvantages:
- DCI devices must provide high performance (all entries of data centers).
  This mode is recommended for large-scale, modular data centers.

Method 2:
Combined: VTEPs in different data centers communicate directly through the VXLAN tunnels, and the underlay network provides transparent transmission of IP packets to implement DCI. Therefore, the two data centers form one large VXLAN domain logically.
Advantages:
- Multiple data centers from one logical DC to simplify management.

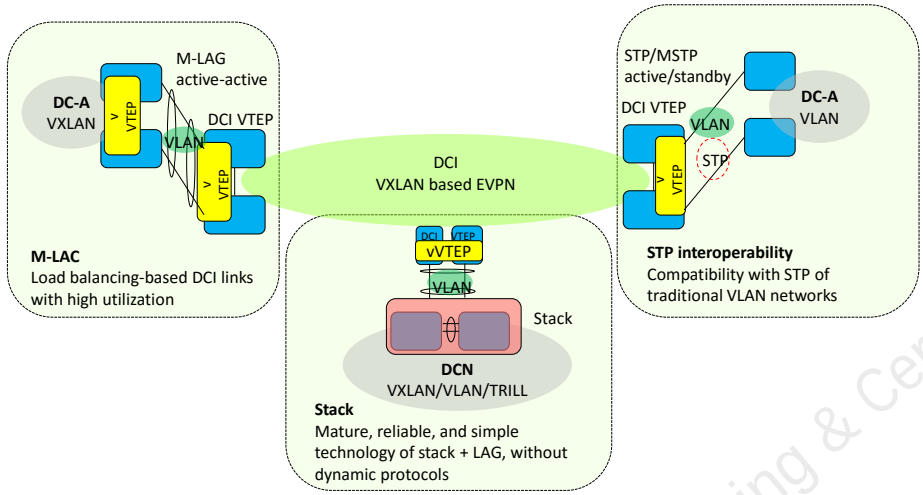
Disadvantages:
- The data centers must use the same protocols and encapsulation mode.
  This mode is recommended for small-sized data centers.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

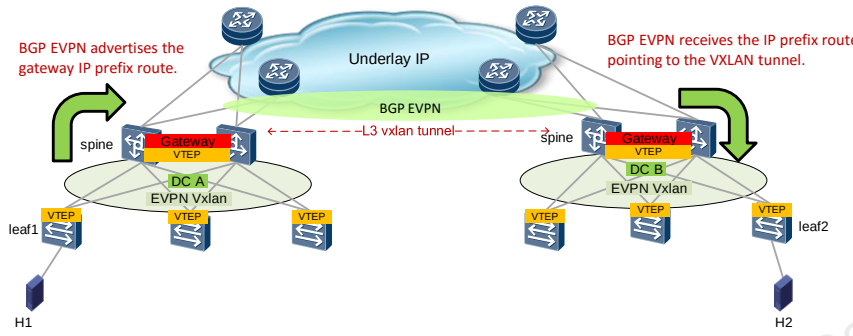
VXLAN-based Segmented Layer 2 DCI Highly Reliable Connection



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

VXLAN Implements Layer 3 Interconnection (Partially Replacing L3VPN)



EVPN allows subnet routes generated by other protocols or a local device to be advertised by BGP EVPN. Different DCs can communicate with each other through VXLAN tunnels, similar to L3VPN communication, simplifying deployment for Layer 3 interconnection between DCs.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

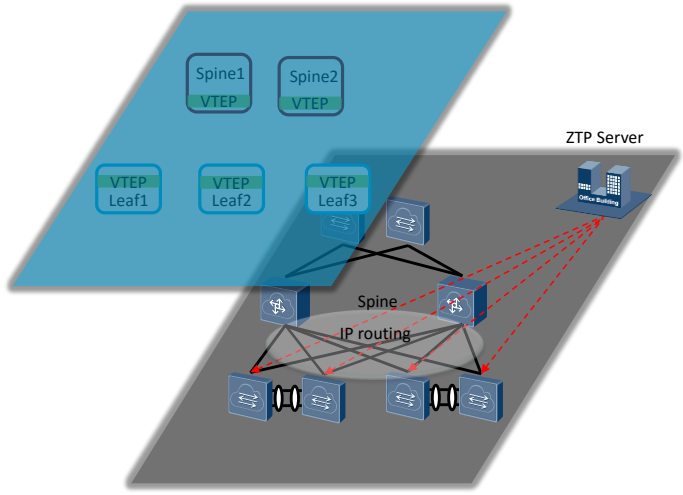
Contents

1. Requirements and Challenges of Cloud DCs
2. Basic Principle of VXLAN
3. Common Networking of VXLAN
4. Introduction to BGP-EVPN
5. VXLAN-based EVPN Deployment in a DC
6. VXLAN-based EVPN Deployment Between DCs
7. **Configuration and Deployment Optimization**
8. Application Cases

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Underlay Batch Configuration Through ZTP



Underlay deployment through ZTP

- Configuration of device management and authentication
- Configuration of device VTEP IP addresses
- Configuration of device interface connection
- Configuration of device interconnection routing protocols
- BGP EVPN session configuration

After ZTP is completed:

Devices can be managed by the Agile Controller.
IP addresses of VTEPs are reachable.
BGP EVPN peer is established.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

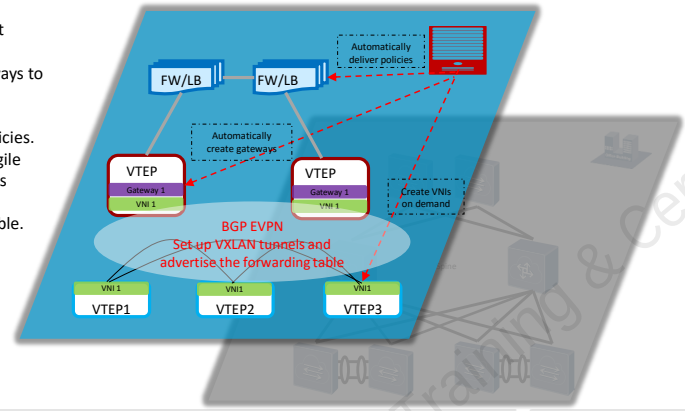
HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Simplify Overlay Service Deployment Using the Agile Controller

Simplify overlay service deployment using the Agile Controller:

- Access device
  - ✓ Create VNIs on demand and map VLAN IDs to VXLAN VNIs.
- Gateway
  - ✓ Create VNIs and L3 IP gateways automatically.
  - ✓ Advertise gateway subnet routes automatically.
  - ✓ Import traffic from gateways to firewalls automatically.
- Firewall
  - ✓ Automatically deliver policies.
- After creating VNIs on the Agile Controller, BGP EVPN triggers tunnel establishment, and advertises the forwarding table.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

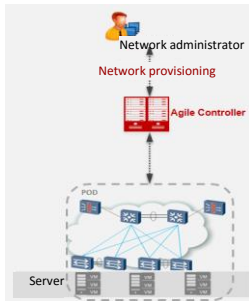
Huawei Training & Certification
Huawei Training & Certification

Optimize Overlay Service Deployment Using the Agile Controller

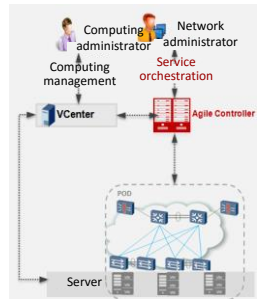
Phase 1
Network service configuration on GUIs through the Agile Controller
- Devices provide APIs for overlay service configuration.
- The Agile Controller configures overlay services for all NEs.

Phase 2
Interconnection between the Agile Controller and the VMM platform
- The Agile Controller configures overlay services for all NEs on GUIs.
- The Agile Controller obtains information about host login from the VMM to implement on-demand and flexible host service access.

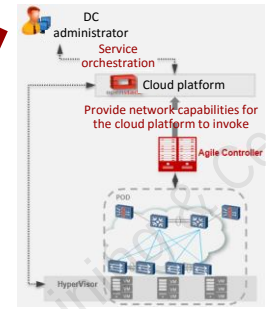
Phase 3
Interconnection between the Agile Controller and the service cloud platform
- The service cloud platform centrally allocates resources.
- The Agile Controller configures overlay services for all NEs according to instructions delivered by the cloud platform.
- The Agile Controller obtains information about host login from the cloud platform to implement on-demand and flexible host service access.



API-based network capability



API-based network capability + flexible resource allocation



On-demand network construction

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Possible Problems During Server Deployment

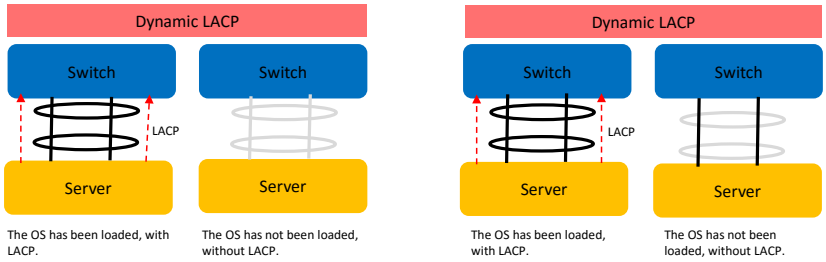
- Servers must be highly reliable. Switches connect to bundled NICs of the server using static LAGs or LACP LAGs.
- Pre-configure switches, create LAGs, and then deploy servers.
- Possible problems:
  - In the PXE phase during server deployment, only one NIC can be used to receive and send traffic. If switches work in a LAG, traffic may be hashed to a non-PXE NIC, which hinders server deployment. In this case, you need to manually delete the LAG and create it after server deployment is completed.
  - If the access VLAN and PXE VLAN (subnet) of switches are inconsistent with the service VLAN (subnet), you need to modify VLANs during server deployment.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Simplified Approach — Dynamic LACP LAG + Trunk Service VLAN



The OS has been loaded, with LACP.

The OS has not been loaded, without LACP.

The OS has been loaded, with LACP.

The OS has not been loaded, without LACP.

- Traditional LACP LAG
  - If switches fail to receive LACP packets, members in a LAG stop running.
  - In the PXE phase, the server fails to send LACP packets. As a result, the LAG cannot forward packets.
- Dynamic LACP LAG
  - If switches fail to receive LACP packets, members in a LAG inherit LAG configurations and work independently.
  - In the PXE phase, packets can be forwarded whatever NIC is used.
  - After the PXE is executed and the OS is loaded, LAGs are restored to the normal state.
- The PXE access VLAN and the service trunk VLAN are independent of each other, preventing modification of the access interface VLAN.
- The PXE VLAN can be pre-configured using ZTP. The service VLAN can be created or configured on demand by the Agile Controller.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Simplified Approach — Asymmetric VLANs for LAGs



- The trunk service VLAN covers all member interfaces in a LAG, implementing load balancing for service traffic among all members interfaces.
- Interfaces connected to the PXE NIC are added to an independent access PXE VLAN to ensure that PXE traffic is transmitted between fixed interfaces and is not hashed to other NICs.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

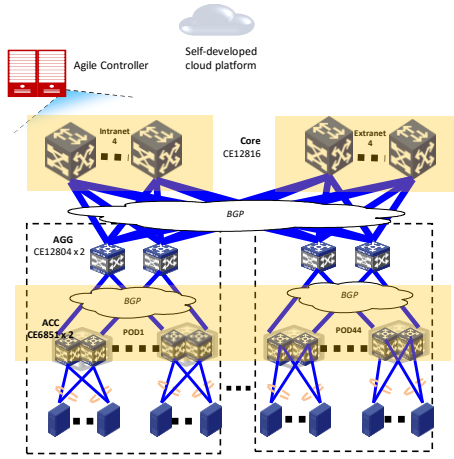
Contents

1. Requirements and Challenges of Cloud DCs
2. Basic Principle of VXLAN
3. Common Networking of VXLAN
4. Introduction to BGP-EVPN
5. VXLAN-based EVPN Deployment in a DC
6. VXLAN-based EVPN Deployment Between DCs
7. Configuration and Deployment Optimization
8. **Application Cases**

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Super Large-Scale Public Cloud for Internet Company A in China



Customer requirements

- Large scale, high performance, and low convergence ratio
- Overall DC migration
- Agile Controller-based automatic and refined O&M

Solution

- Basic networks use the Spine-Leaf architecture and overlay networks use VXLAN to set up large Layer 2 networks. VTEPs are deployed on leaf switches.
- Layer 3 VXLAN gateways are deployed in a centralized manner. Multi-active gateways are deployed to implement load balancing and improve the throughput capability.
- The network controller provides customized northbound APIs to connect to the private cloud platform and implement automatic deployment of network services.
- The Agile Controller achieves path and traffic visualization to facilitate O&M.

Customer benefits

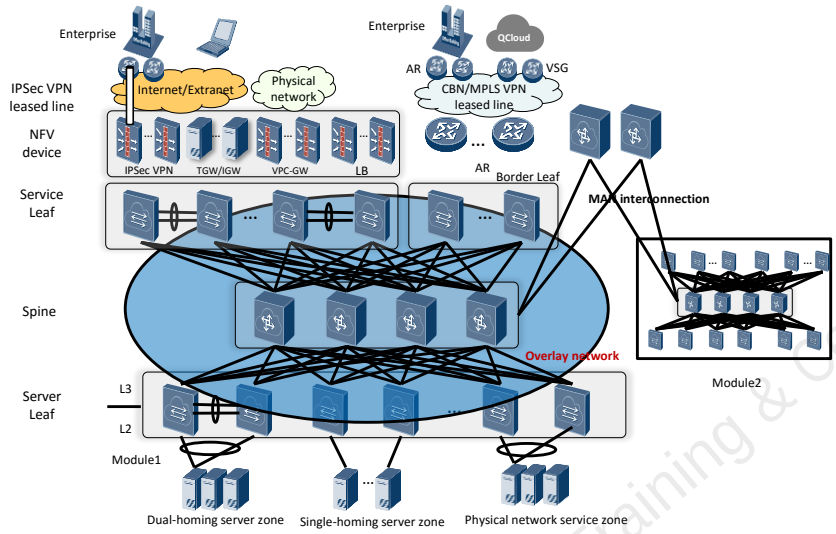
- The multi-active centralized hardware gateway ensures scale and performance, and complies with the O&M habits of management engineers in traditional data centers.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Super Large-Scale Public Cloud for Internet Company T in China

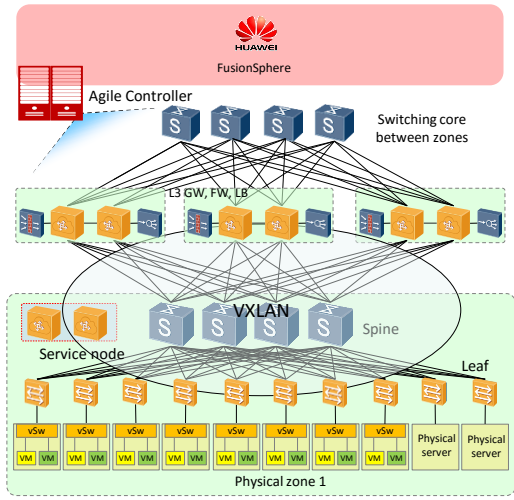


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Customer requirements
  - Large scale, high performance, and low convergence ratio
  - Overall DC migration
  - Agile Controller-based automatic and refined O&M
  - Preventing storms caused by Layer 2 loops
- Solution
  - Basic networks use the Spine-Leaf architecture and overlay networks are built using VXLAN. VTEPs and gateways are deployed on leaf switches.
  - The network controller connects to the private cloud platform and implements automatic deployment of network services.
  - The Agile Controller achieves path and traffic visualization to facilitate O&M.
  - Use ARP proxy to forward all traffic at Layer 3, eliminating loops.
- Customer benefits
  - The hardware distributed Anycast gateway does not require vSwitches to build large-scale and high-performance networks.
  - All Layer 3 traffic forwarding eliminates the potential risk of Layer 2 loops.

Huawei Training & Certification

SDN Private Cloud of XX Bank of China



Customer requirements

- The cloud platform must be open-source, independent, and controllable.
- VMs can migrate across physical zones.
- Layer 2 loops can be detected and prevented.

Solution

- Basic networks use the Spine-Leaf architecture and overlay networks use VXLAN to set up large Layer 2 networks. VTEPs are deployed on leaf switches.
- Multiple groups of Layer 3 VXLAN gateways can be deployed in a centralized mode. The gateways support service expansion.
- The network controller connects to the OpenStack cloud platform to implement automatic deployment of network services.
- Layer 2 loop detection is deployed to prevent loops on a large Layer 2 network.

Customer benefits

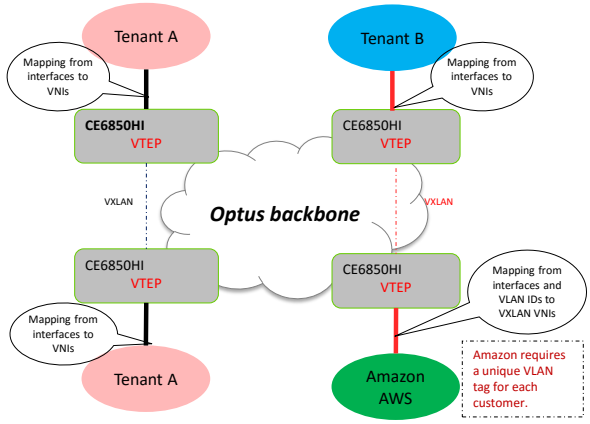
- An open cloud platform ensures long-term evolution.
- The hardware centralized gateway ensures performance and complies with the O&M habits of management engineers in traditional data centers.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Australia's Optus Uses VXLAN to Build Remote Equipment Rooms Connected with Private Lines



- Tenants rent interfaces provided by Optus to implement communication between equipment rooms.
- There are two scenarios:
  - Tenant equipment rooms communicate with each other (Optus needs to transparently transmit tenant packets with VLAN tags).
  - Tenant equipment rooms communicate with the Amazon AWS cloud platform (In addition to transparent transmission, Optus needs to add a unique VLAN tag for each user).
- Tenant equipment rooms communicate with each other through symmetric deployment of the mapping from interfaces to VNIs.
- The mapping from interfaces to VNIs is deployed on one side and the mapping from interfaces and VLAN IDs to VXLAN VNIs is deployed on the other side. In this way, each packet sent to the AWS cloud platform is added with a VLAN tag.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Summary

- Basic Principle of VXLAN
- VXLAN-based EVPN Deployment in a DC
- VXLAN-based EVPN Deployment Between DCs
- Configuration and Deployment Optimization
- Application Cases

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

More Information

- Huawei Official Websites
  - Enterprise service: http://enterprise.huawei.com/cn/
  - Technical support: http://support.huawei.com/enterprise/
  - Online learning: http://learning.huawei.com/cn/
- Documentation Tool
  - HedEx Lite
- Simulator
  - eNSP

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Recommendations

- Huawei E-Learning Website:
  - http://support.huawei.com/learning/Index!toTrainIndex
- Huawei Support Case Library:
  - http://support.huawei.com/enterprise/servicecenter?lang=zh

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

Cloud Data Center Network Security Overview

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Foreword

- An Internet Data Center (IDC) is a platform for data storage and distribution, a place where data exchange is concentrated on the Internet.
- As cloud computing advances, cloud data centers face an increasing number of new security risks and challenges. This cahpter analyzes security risks from multiple layers from the perspective of network security and provides end-to-end and comprehensive security measures for IDC networks, aiming to lay a solid foundation for IDC service development.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Objectives

- Upon completion of this course, you will be able to:
  - Describe major issues in cloud data center network security.
  - Describe design principles of cloud data center network security.
  - Deploy cloud data center network security solutions.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. **Categories of Network Security Issues**
2. Suggestions on Network Security Defense Deployment

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Cloud Data Center Network Security Requirements

- Data Center (IDC) is a system that provides a complete set of facilities and related maintenance services based on the Internet.IDCs collect, store, process, and send data in a centralized manner. IDCs are commonly built by large network server providers to host servers for small- and medium-sized enterprises or individual customers and provide virtual domain name space services.
- The network structure of an IDC has the following features:
  - Servers of multiple enterprises may be deployed in an IDC, featuring complex services.
  - The core function of the IDC is to provide network services externally. Therefore, it is vital to ensure that servers in the IDC are accessible to external networks. For this reason, border defense devices must possess high performance to guarantee network access even under attacks.
  - The IDC mainly protects intranet servers and provides various network services, and therefore are prone to be hackers' targets.
  - The IDC has huge access traffic and needs to provide remote services for a large number of tenants.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Cloud Data Center Network Security Analysis

- Based on cloud data center network security requirements, it can be concluded that cloud data network security has the following requirements:
  - Security zone-based isolation
  - Service isolation
  - Attack defense
  - Tenant access security

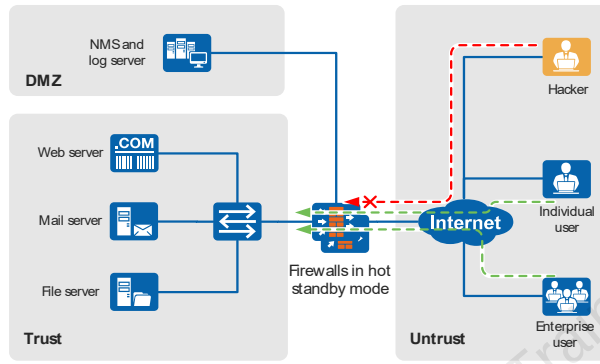
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Security Zone-based Isolation

- For cloud data centers, the major devices that require protection are servers. Generally, it is considered that the area in the data center is a trusted area, and the Internet connected to the data center is an untrusted area. To differentiate data reliability for protection, security zones need to be distinguished among Internet interconnection areas or layers.



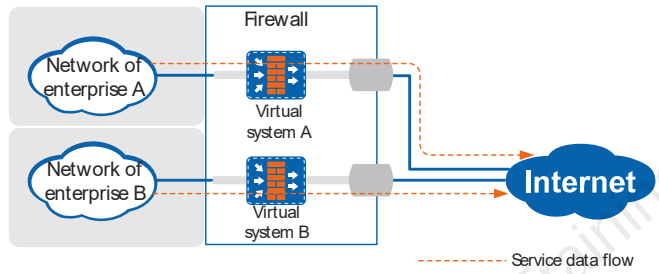
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Service Isolation

- The cloud data center stores servers of multiple enterprises and provides various services for these enterprises. The data of different enterprises with various security requirements needs to be differentiated. To meet these requirements, services need to be isolated so that different security policies can be implemented.



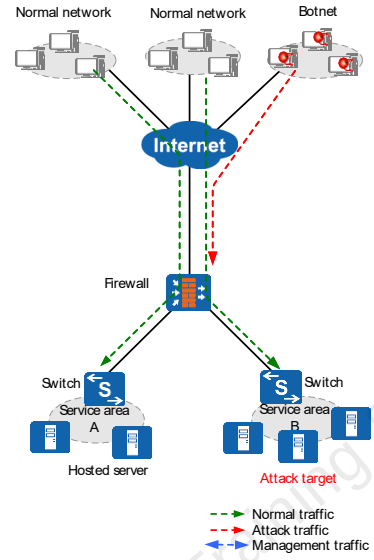
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Attack Defense

- Cloud data centers house servers of a large number of enterprises, and a large amount of data is exchanged in the data centers, making them one of the most vulnerable attack targets. Cloud data centers must guarantee server and network security and provide reliable services for tenants. Therefore, attack defense is a basic security requirement for cloud data centers.

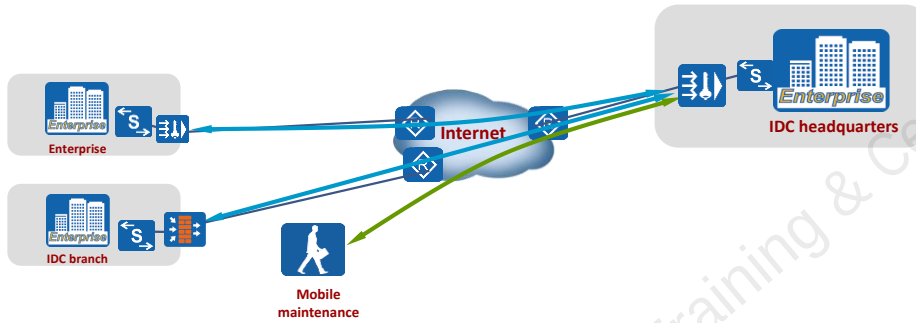


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Access Security

- IDC O&M personnel and customer maintenance personnel need to remotely access IDC and hosted devices. Therefore, secure and reliable VPN access methods shall be provided. In addition, the privacy, integrity, and authenticity of O&M management data must be guaranteed during its transmission over the Interest.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. Categories of Network Security Issues

2. **Suggestions on Network Security Defense Deployment**

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Cloud Data Center Network Security Deployment Overview

- Cloud data centers have the requirements of security zone-based isolation, service isolation, attack defense, and tenant access security. Based on these requirements, security devices such as firewalls, anti-DDoS devices, and SVNs can be deployed to improve the availability and reliability of cloud data centers.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

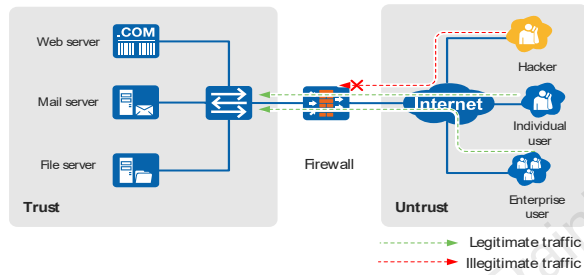
HUAWEI

- The SVN is a professional SSL VPN device.

Huawei Training & Certification
Huawei Training & Certification

Security Zone-based Isolation: Firewall Security Zone

- As network security devices, firewalls are usually deployed at network borders to filter and forward packets. Firewalls are a type of logical area filter. A security zone or zone is a security concept introduced by the device. Most security policies are implemented based on security zones. The firewall considers that data flows in the same security zone bring no security risks and therefore require no security policies. The firewall performs security checks and implements security policies only on data flows between security zones.

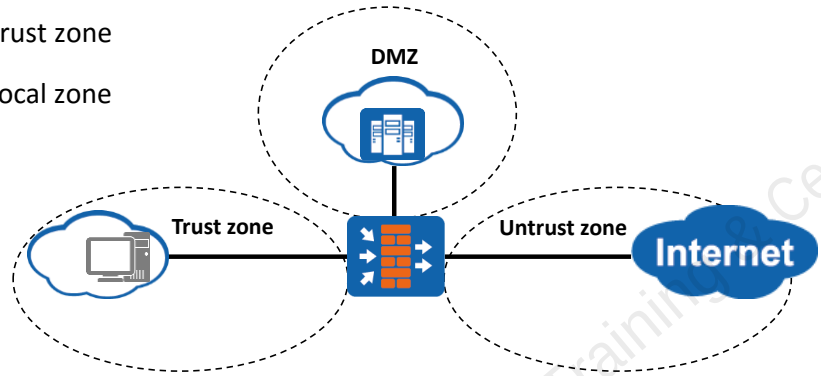


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification

Firewall Security Zones

- Default security zones
  - Untrust zone
  - DMZ
  - Trust zone
  - Local zone



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

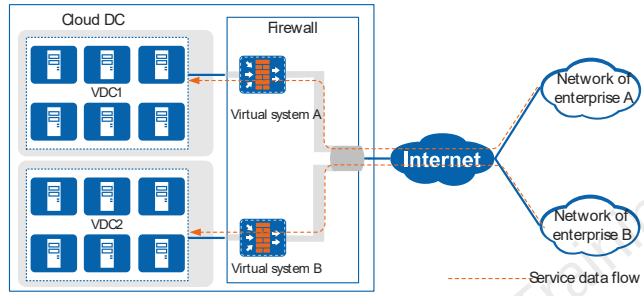
HUAWEI

- The security levels of security zones vary within a range from 1 to 100.
- In cloud data center deployment, servers are usually deployed in the Trust zone, whereas the Internet is in the Untrust zone. The DMZ is usually for hot standby.

Huawei Training & Certification

Service Isolation

- Service isolation can be implemented by dividing virtual systems.
- A virtual system is a logical device created on a physical device. Virtual systems are independent from each other.
- A firewall can be logically divided into multiple virtual systems. Each virtual system has its resources and configurations, such as interfaces, address sets, users/user groups, routing entries, and policies, and provides the same functions as a physical device. In addition, it can be configured and managed by the virtual system administrator.



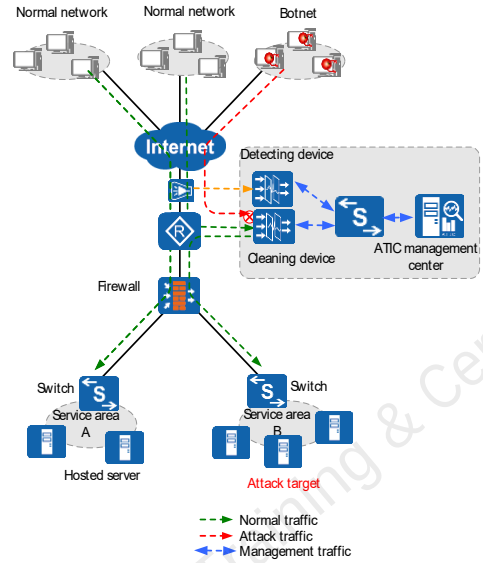
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Virtual systems have the following features:
  - Each virtual system has its own administrator and can be managed independently. With virtual systems, a large network can be divided into smaller subnets with each being served by a virtual system, simplifying the network management.
  - Each virtual system has its own configurations and routing entries so that networks connected to different virtual systems can have overlapping private addresses.
  - Each virtual system has its own resource quota so that a busy virtual system has no impact on other virtual systems.
  - The traffic of different virtual systems is separated to ensure security. However, different virtual systems can still communicate with each other if needed.
  - Virtual system technology reduces hardware investment, power consumption, and equipment footprint.

Huawei Training & Certification

Attack Defense — Anti-DDoS

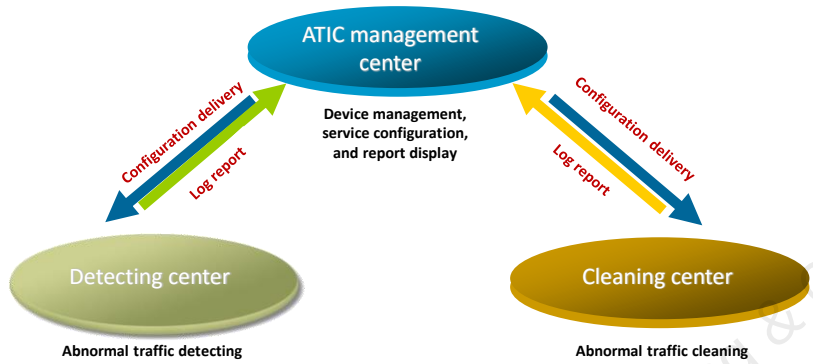
- In the past, DDoS attacks dominated Flood attacks,mostly targeting at carriers' fundamental network infrastructure, whereas existing DDoS attacks target at specific applications and services, such as enterprise applications, DNS, and email. The targets are expanding, the traffic of a single attack is small and costs low, and the attack behaviors are becoming increasingly complicated and vivid, which add difficulties to anti-DDoS.
- Huawei anti-DDoS solution not only defends against traditional traffic-type DDoS attacks but also provides application-layer defense, defense against botnets, Trojan horses, and worms, and IPv4-IPv6 dual-stack protection.
- The anti-DDoS device can be deployed at the egress to protect the entire data center network and provide value-added security services for tenants.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Anti-DDoS Solution



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

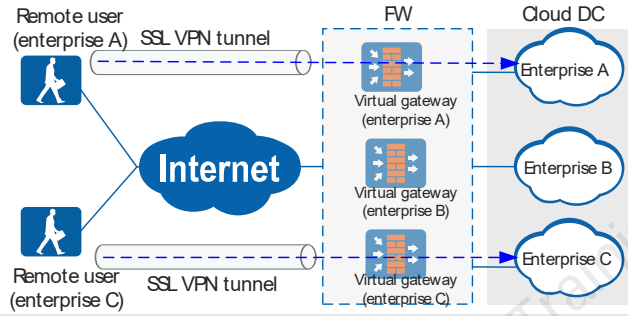
HUAWEI

- Detecting center
  - The detecting center detects traffic for attacks and reports attack events to the management center. Then the management center delivers traffic diversion policies for the cleaning device to clean the traffic. The detecting technologies are netflow-based traffic detection and application-based in-depth packet detection technologies. The former detects only large volume attack traffic, whereas the latter can further detect small volume attack traffic as well as application-layer attack traffic.
- Cleaning center
  - The cleaning center diverts and cleans traffic based on the policies delivered by the management center and then injects the cleaned traffic back. In this process, the cleaning center also logs the events and reports them to the management center. The cleaning center provides diversified DDoS traffic cleaning methods. It can accurately identify normal traffic and clean various types of abnormal traffic, including traffic-type attacks, application-layer attacks, scanning and sniffing attacks, and malformed-packet attacks. In addition, the cleaning center is capable of attack detection. If the services require low detection and cleaning performance, you can deploy only the cleaning center.
- Management center
  - The ATIC management center is in charge of the centralized management of the detecting and cleaning centers. As the center of the anti-DDoS solution, the management center provides functions, including device management, policy management, performance management, alarm management, and report management.

Huawei Training & Certification

Tenant Access Security — SSL VPN

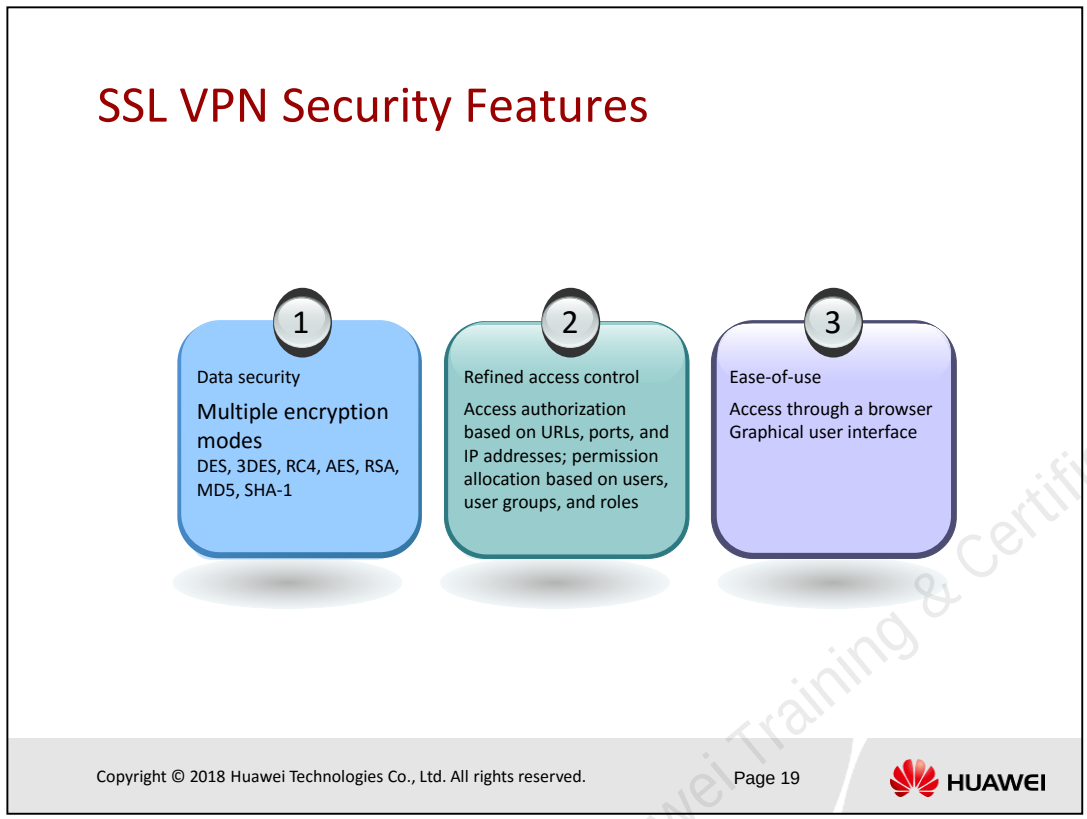
- The SSL VPN function or SVN gateway is deployed to implement secure access for remote users.
- The SVN gateway is usually deployed between the firewall at the network ingress/egress and the application server. By default, SSL port 443 is enabled on the firewall.
- A private network IP address is configured on the SVN gateway so that the SVN gateway is reachable to the intranet server.
- A public network IP address is advertised externally for user access. The IP address can be configured on the gateway, or the firewall performs NAT on the IP address.
- O&M personnel enter the gateway IP address for access by using web browsers or client software. Packets are sent to the firewall first. After the firewall identifies that the packets are destined for the SVN gateway, the firewall forwards the packets to the SVN gateway. The SVN gateway sends the packets to the IDC server after decryption.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification



- Compared with traditional VPN, SSL VPN has the following advantages in remoate access scenarios:
  - Uses the browser/server (B/S) architecture, eliminating the need for a client program.
  - Supports fine-grained permission control based on the type of accessed resources.
  - Supports multiple identity authentication modes, such as local, server, certificate-anonymous, and certificate-challenge authentication.
  - Supports host check policies to check whether the operation systems, ports, processes, and autivirus software of remote user devices meet security requirements, and provides the anti-nested remote desktop connections and anti-snapshot functions to eliminate security risks from remote user devices.
  - Supports cache clearing policies to clear the access history to enhance information security.

Huawei Training & Certification

Quiz

1. Virtual systems of firewalls can communicate with each other.

   A. T

   B. F

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Answer
  - A

Huawei Training & Certification
Huawei Training & Certification

Summary

- Cloud data center network security issues
  - Security zone-based isolation
  - Service isolation
  - Attack defense
  - Secure access
- Cloud data center network security deployment suggestions
  - Firewall security zone
  - Virtual system
  - Anti-DDoS solution
  - SSL VPN

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

More Information

- Security documentation bookshelf:
- http://support.huawei.com/enterprise/zh/index.html#popuplay=9856724

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Recommendations

- Huawei E-Learning website:
  - http://support.huawei.com/learning/Index!toTrainIndex
- Huawei support case library:
  - http://support.huawei.com/enterprise/servicecenter?lang=zh

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

Cloud Data Center Network Security Deployment

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.



Huawei Training & Certification
Huawei Training & Certification

Foreword

- Data Center (IDC) is a system that provides a complete set of facilities and related maintenance services based on the Internet for centralized data collection, storage, processing, and transmission. IDCs are commonly built by large network server providers to host servers for small- and medium-sized enterprises or individual customers and provide virtual domain name space services.
- IDCs, as the core of network data, usually are vulnerable to external attacks. Every year, there are multiple servers in IDCs being attacked, causing great loss to customers. Therefore, IDCs attach increasing importance to IDC network security.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Objectives

- Upon completion of this course, you will be able to describe:
  - Firewall networking and deployment.
  - IPS/IDS networking and deployment.
  - Anti-DDoS networking and deployment.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. Firewall Networking and Deployment
2. IPS/IDS Networking and Deployment
3. Anti-DDoS Networking and Deployment

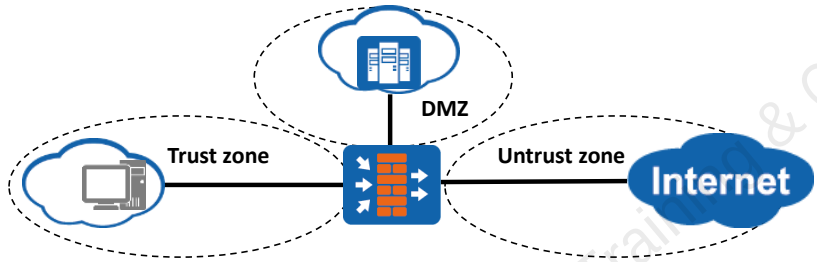
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Basic Principles

- As network security devices, traditional firewalls are usually deployed at network borders to filter and forward packets.
- Firewalls are a type of logical area filter.
- Security zone or zone for short:
  - A zone refers to a local logical security zone.
  - A zone is a network connected to one or more interfaces.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

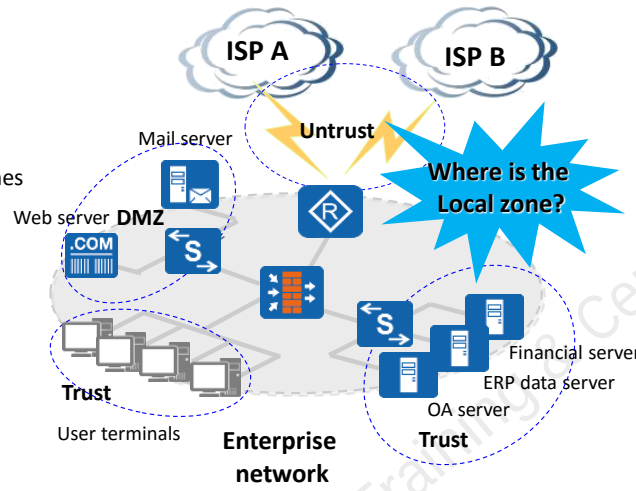
HUAWEI

- Zone functions:
  - Most security policies are implemented based on security zones.
  - Data flows in the same security zone bring no security risks and therefore require no security policies.
  - The firewall performs security checks and implements security policies only on data flows between security zones.
  - All devices on the networks connected to the same interface must reside in the same security zone, and one security zone may contain the networks connected to multiple interfaces.

Huawei Training & Certification

Security Zones

- Default security zones
  - Untrust zone
  - DMZ
  - Trust zone
  - Local zone
- User-defined security zones
  - User Zone 1
  - User Zone 2



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The firewall supports four default security zones, namely, the Untrust, DMZ, Trust, and Local zones. In addition to the default zones, the firewall also supports user-defined security zones.
- The default security zones are described as follows:

  Untrust: Indicates a security zone with a low security level of 5.

  DMZ: Indicates a security zone with a medium security level of 50.

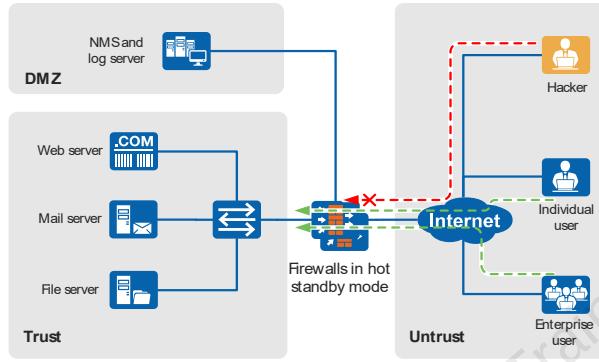
  Trust: Indicates a security zone with a relatively high security level of 85.

  Local: Indicates a security zone with the highest security level of 100.
- You do not need to create these security zones. In addition, you cannot delete them or reset their security levels.The security level ranges from 1 to 100. A larger value indicates a higher security level.
- Note that after an interface is added to a security zone, the network connected to the interface is added to the security zone, but the interface still belongs to the Local zone that is reserved by the system for the device.
- The USG supports a maximum of 32 security zones.

Huawei Training & Certification

Application Scenario - IDC Border Protection

- Data Center (IDC) is a system that provides a complete set of facilities and related maintenance services based on the Internet for centralized data collection, storage, processing, and transmission. IDCs are commonly built by large network server providers to host servers for small- and medium-sized enterprises or individual customers and provide virtual domain name space services.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The network structure of an IDC has the following features:
  - Servers in the IDC are protected. Security functions are selected based on server types.
  - Servers of multiple enterprises may be deployed in an IDC and are prone to be hackers' targets.
  - The core function of the IDC is to provide network services externally. Therefore, it is vital to ensure that servers in the IDC are accessible to external networks. For this reason, border defense devices must possess high performance and comprehensive availability mechanisms to guarantee network access even under attacks.
  - IDC traffic is complicated. If the traffic is invisible, configurations are hard to be adjusted properly.

Huawei Training & Certification

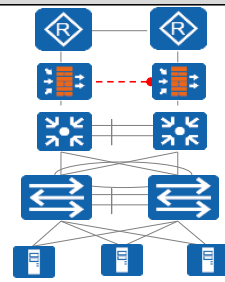
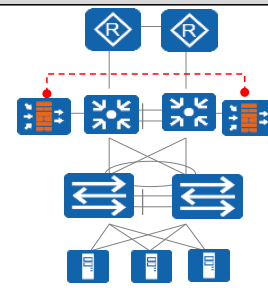
Scenario Analysis - IDC Border Protection

- Firewall application analysis:
  - Enable traffic statistics collection by IP address, user, and application to help formulate security policies.
  - Enable IP address- and application-based traffic limiting to ensure the stable operating of servers and prevent network congestion.
  - Enable intrusion prevention and antivirus to protect servers from viruses, Trojan horses, and worms.
  - Enable anti-DDoS and other attack defense functions to defend against attacks from the Internet.
  - Enable mail filtering to protect mail servers on the intranet from spams and prevent them from being blacklisted by anti-spam organizations due to unintentional spam forwarding.
  - Enable file blocking and data filtering to prevent information leaks.
  - Deploy Huawei NMS (to be purchased independently) to manage network operation logs. The logs help the administrator adjust configurations, identify risks, and audit traffic.
  - Deploy the firewalls in hot standby mode to improve availability. If a single point of failure occurs, service traffic can be smoothly switched from the active device to the standby device to ensure service continuity.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

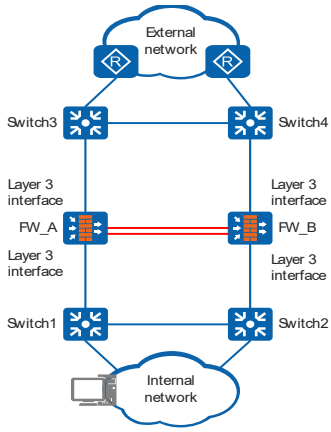
Firewall Deployment Modes

Deployment Mode	In-line Deployment (Hot Standby)	Off-line Deployment (Hot Standby)
Networking		
Strengths	Traffic deployment is clear.	Certain traffic may not pass through the firewall as specified by policies.
Issues that Need Improvement	In-line deployment has high requirements on the firewall performance.	The switch policies are complicated.

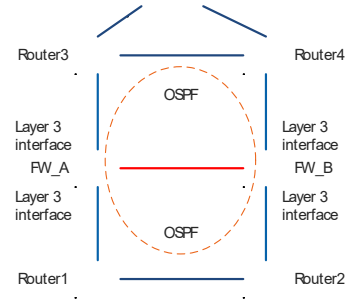
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Firewall Working Mode - Routing Mode



- Routing mode 1:
  - The firewall works in Layer 3 routing mode.
  - The upstream and downstream devices are Layer 2 devices and transparently transmit VRRP packets of the active and standby firewalls.



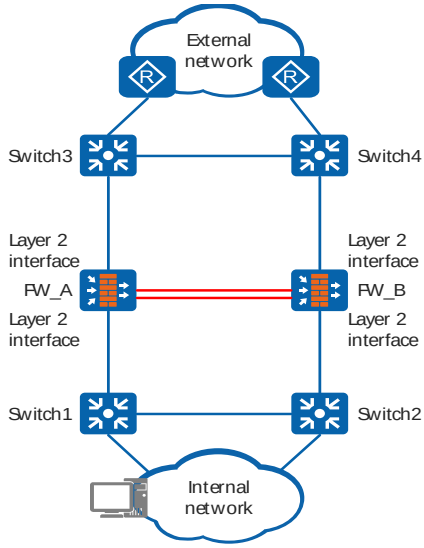
- Routing mode 2:
  - The firewall works in Layer 3 routing mode.
  - The upstream and downstream devices are Layer 3 devices and have an OSPF neighbor relationship established with the firewall.
  - The standby firewall advertises a larger cost value so that traffic is preferentially forwarded by the active firewall.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Firewall Working Mode - Transparent Mode



- Transparent mode:
  - The firewall works in Layer 2 transparent mode.
  - The upstream and downstream devices are Layer 2 devices. Usually, the standby firewall is configured not to forward packets.
  - The upstream and downstream devices are Layer 3 devices, and the two firewalls process services together.

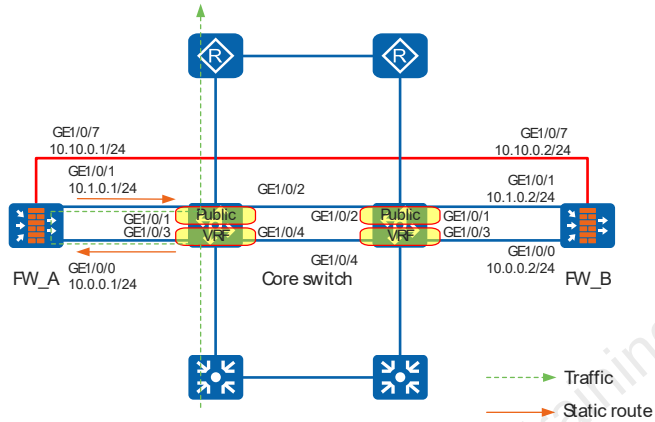
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Typical Deployment Scenarios of Cloud Data Centers

- The firewalls in active/standby deployment work in routing mode and are connected to Layer 3 devices in off-line mode.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Connecting firewalls to Layer-3 switches in off-line mode using VRRP and static routing
  - If the core switches need to use static routes to divert traffic to the firewalls, you need to configure static routes and set the next hops to the IP addresses of the firewall interfaces. However, core switches generally use OSPF to communicate with upstream routers and downstream aggregation switches. OSPF routes have higher priorities than static routes. Therefore, core switches directly forward received traffic to upstream or downstream devices using OSPF routes, but do not divert traffic to the firewalls using static routes.
  - If you intend to use static routes for traffic diversion, you must configure the virtual routing and forwarding (VRF) function on the core switches to virtualize each core switch into a public switch (Public) for connecting to the upstream switch and a virtual switch (VRF) for connecting to the downstream switch. The two virtualized switches are isolated. Therefore, traffic can be diverted to the firewalls.

Huawei Training & Certification

Contents

1. Firewall Networking and Deployment
2. **IPS/IDS Networking and Deployment**
3. Anti-DDoS Networking and Deployment

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

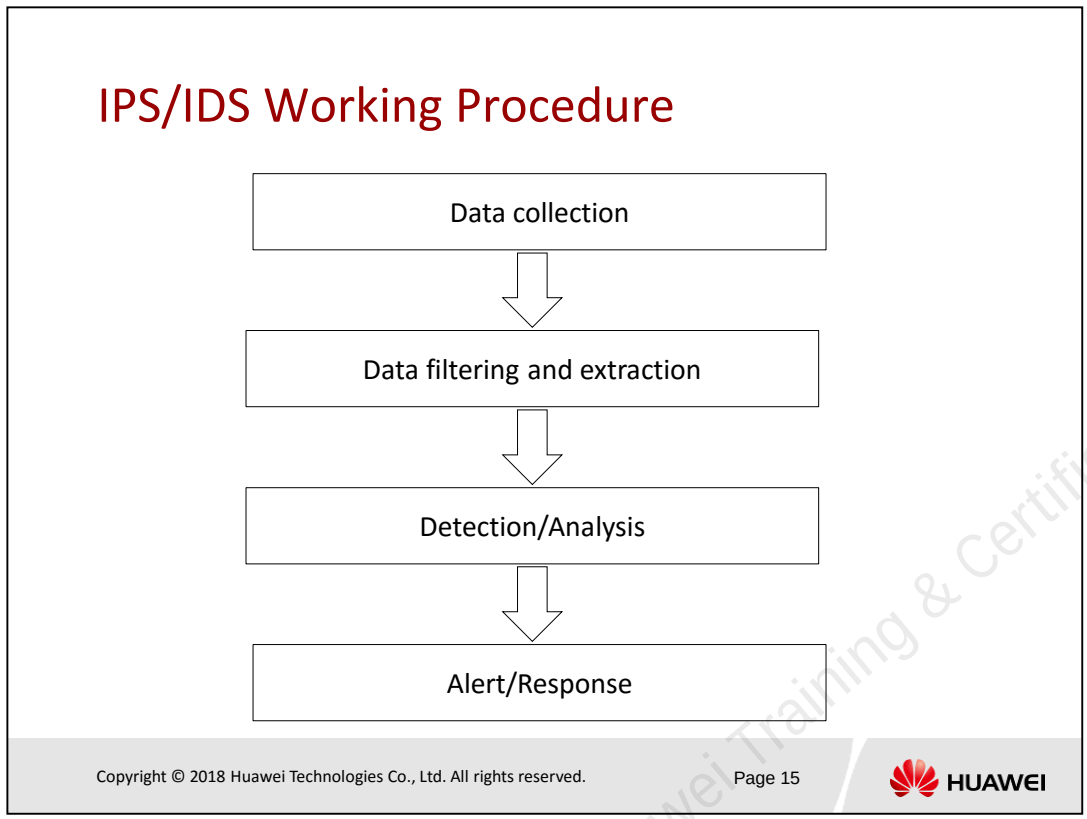
IPS/IDS Basic Concepts

- The Intrusion Detection System (IDS) is used to detect intrusion behaviors on the network, collect information from key points on the network, and analyze the information to check whether any behavior that violates security policies exists on the network.
- The Intrusion Prevention System (IPS) can not only detect intrusion behaviors but also respond to intrusions to block attacks in a timely manner.
- The IPS/IDS mainly applies to enterprises, IDCs, and campus networks to ensure application and traffic security.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

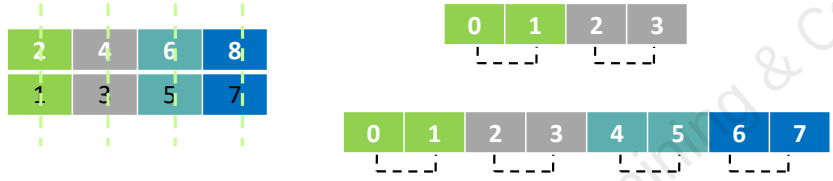


- Generally, an IDS needs to perform the following four steps to analyze and determine whether a specific behavior or event is an abnormal or attack behavior that violates security policies:
  - Data collection:
    - The IDS needs to collect necessary data for intrusion analysis.
  - Data filtering and extraction:
    - The IDS performs data filtering and extraction based on predefined settings to improve detection and analysis efficiency.
  - Detection/Analysis:
    - The IDS performs detection/analysis based on defined security policies.
  - Alert/Response:
    - Once detecting any behavior or event that violates security policies, the IDS generates an alarm and responds.

Huawei Training & Certification

IPS/IDS Working Mechanism

- By default, interfaces on IPS/IDS devices work in pairs. The interfaces on main interface boards and extension interface cards work in pairs (such as a01-b01) based on the interface sequence. The interfaces work in IPS mode by default. You can configure the interfaces to work in IDS mode as required. Sessions can be shared within an interface pair.
- Interface pairs can be combined to form an interface pair group as required. Sessions can be shared within an interface pair group.



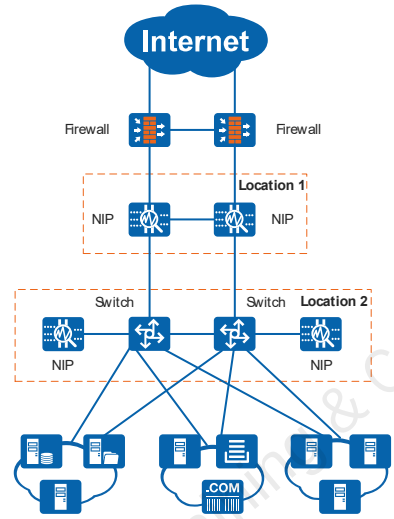
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- An interface pair is a pair of incoming and outgoing interfaces.
- After an interface pair is formed, the traffic entering the incoming interface of the interface pair is forwarded out of the outgoing interface in the interface pair, without routing table or MAC address table lookup.
- If the incoming and outgoing interfaces are the same interface, the packets entering the interface are forwarded out of the same interface after being processed.

Huawei Training & Certification

Application Scenario - Deployment In the Upstream of IDC or Enterprise Servers

- The NIP6000s are deployed in the upstream of IDC and enterprise servers to protect data security on the servers.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

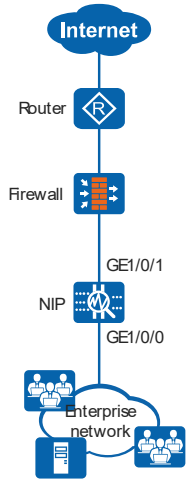
HUAWEI

- The NIP6000s can be deployed at the following positions:
  - Position 1: The NIP6000s can be deployed in-line in the upstream of servers to transparently access the network.
  - Position 2: The NIP6000s can be attached to switches or routers. In this mode, traffic exchanged between the Internet and servers is diverted to the NIP6000s for processing and then injected back afterwards.
- In this scenario, the NIP6000 mainly provides the following functions:
  - Intrusion prevention: The NIP6000 defends against worm activities targeting web, mail, and DNS servers and exploits of service and platform vulnerabilities to prevent malware from damaging, tampering with, or stealing data on the servers as well as SQL injection, scanning, guessing, and snooping attacks specific to web applications.
  - Antivirus: The NIP6000 scans the files to be uploaded to the servers for viruses to protect the servers.
  - Anti-DDoS: The NIP6000 defends against DoS and DDoS attacks targeting servers.

Huawei Training & Certification

IPS/IDS In-Line Deployment

- IPS devices are deployed in-line and connected to the network in Layer 2 mode.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

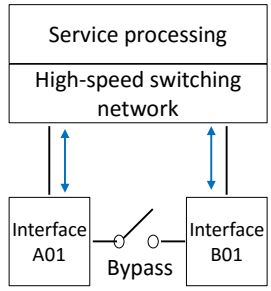
HUAWEI

- In-line deployment applies only to IPS devices so that they can block attacks in a timely manner once detecting them.
- The NIP is deployed in the downstream of an egress firewall or a router and transparently connects to the network.
- You only need to connect the interfaces in an interface pair respectively to the upstream and downstream devices.For example, as shown in the figure, GigabitEthernet 1/0/0 and GigabitEthernet 1/0/1 form an interface pair. Traffic enters the device from GigabitEthernet 1/0/1 and is forwarded out from GigabitEthernet 1/0/0.

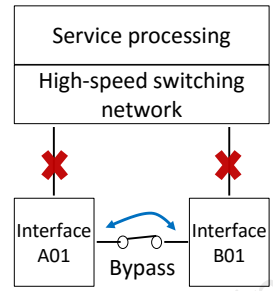
Huawei Training & Certification

IPS/IDS Availability Design

- When the NIP is deployed in in-line mode, an external optical bypass device can be deployed to prevent service interruption caused by link faults and improve availability.



Properly working



Bypass triggered by a system anomaly

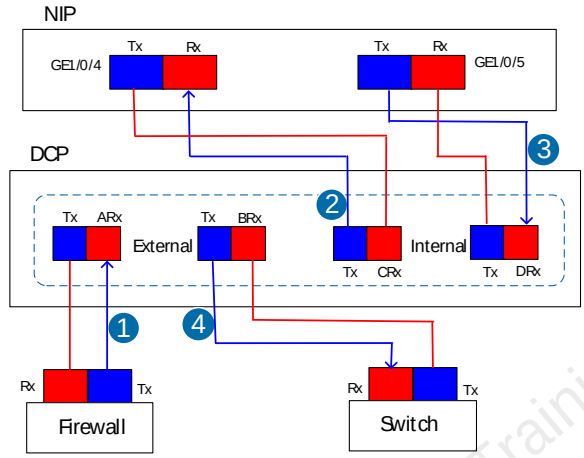
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The external optical bypass device prevents optical link interruption caused by device faults in in-line deployment scenarios. If the device is faulty or is being upgraded, the interface cannot transmit optical signals. In this case, the service traffic can be switched to the bypass device by setting the optical switch of the bypass device. The optical links become straight-through, preventing network breakdown caused by the failure to transmit optical signals and improving network reliability.

Huawei Training & Certification

IPS In-line Deployment and External Optical Bypass

- The following figure shows the cabling and traffic between the NIP and external optical bypass device.



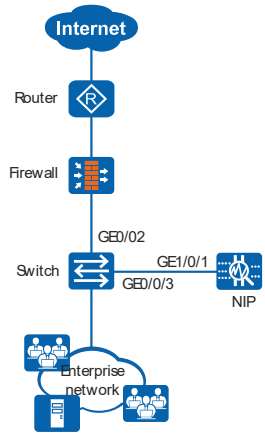
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Dual-channel optical line protection board (DCP)
- Take the downstream traffic of a link as an example. The traffic is forwarded as follows: When the external optical bypass device is in the working channel, the traffic from the firewall enters the DCP through port ARX and is forwarded to port CTX and then to the NIP for processing. After being processed by the NIP, the traffic is forwarded back to the DCP through port DRX and eventually enters the switch through port BTX. The traffic path is 1 -> 2 -> 3 -> 4.
- When the external optical bypass device is in the protection channel, the traffic from the firewall enters the DCP through port ARX and then directly enters the switch through port BTX, without passing through the internal interface of the bypass device and the NIP. The traffic path is 1 -> 4.

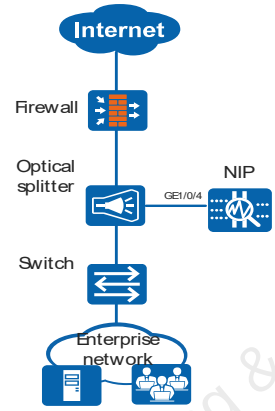
Huawei Training & Certification

IPS/IDS Off-Line Deployment

- Off-line deployment is the most typical deployment mode of the IPS/IDS.



- The IPS/IDS is associated with the switch to detect threats on the network by receiving mirrored traffic from the switch.



- The IPS/IDS is associated with the optical splitter, which is introduced to copy traffic to the NIP to detect threats on the network.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- IDS devices are usually deployed in off-line mode. IDS devices are used mainly for recording attack events and web application traffic conditions to provide evidence for network security event audit and user behavior analysis. In off-line deployment, IDS devices usually do not take actions for attacks. To meet special requirements, you can set the global working mode to protection mode for the IDS devices to take defense actions. The devices can send TCP RST packets to block connections, but the capability is limited.
- The IPS/IDS provides the following services and functions when it is deployed in off-line mode: Intrusion prevention/Antivirus: If you set the global working mode to protection mode, the NIP can block connections with limited capability by sending TCP RST packets.
  - Attack defense: The NIP detects attacks but does not take actions.
  - URL filtering: not supported.
  - Bandwidth management: not supported.

Huawei Training & Certification

Contents

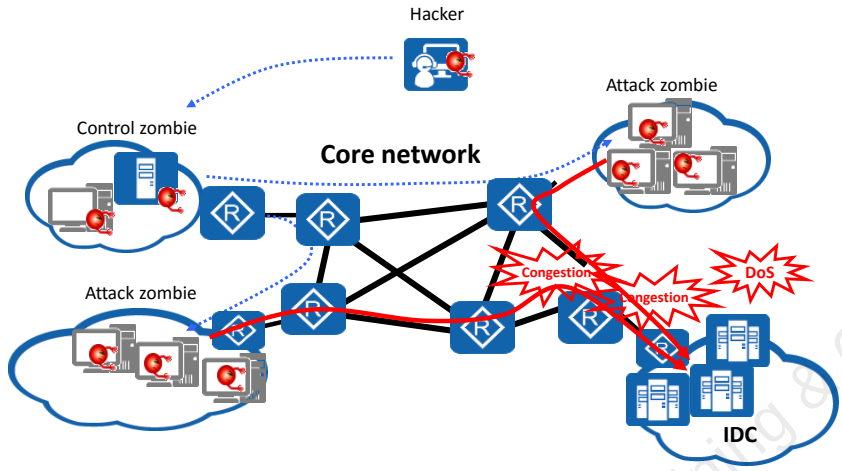
1. Firewall Networking and Deployment
2. IPS/IDS Networking and Deployment
3. **Anti-DDoS Networking and Deployment**

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

DDoS Attacks



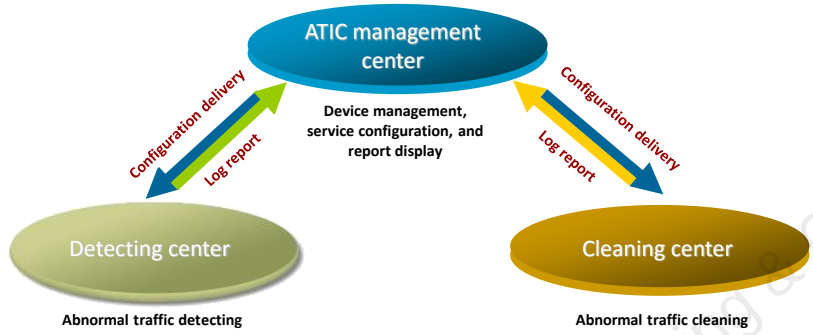
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification

- The DDoS attacks on the IP network are mainly TCP flooding, ICMP flooding, UDP flooding, and fragment-based flooding attacks. Previously, attacks with a relatively small volume of traffic mainly affect user hosts and networks. Currently, heavy-traffic attacks have affected the backbone layer of the IP network, causing abnormal forwarding on device boards, disconnection of switches, and network relay congestion. In addition, among all types of attacked users, the situations are especially severe for VIP customers (such as Internet cafes, business users, and enterprise users) in the MAN. To conclude the network faces the following DDoS threats:
  - The attack traffic volume is evolving from small to large, which directly threatens the IDC and congests the egress bandwidth of VIP customers.
  - The attack type changes from the original protocol layer to the application layer, which can disable the attacked server from providing connections or services just within a short period of time.

Major Components of the Anti-DDoS Solution



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

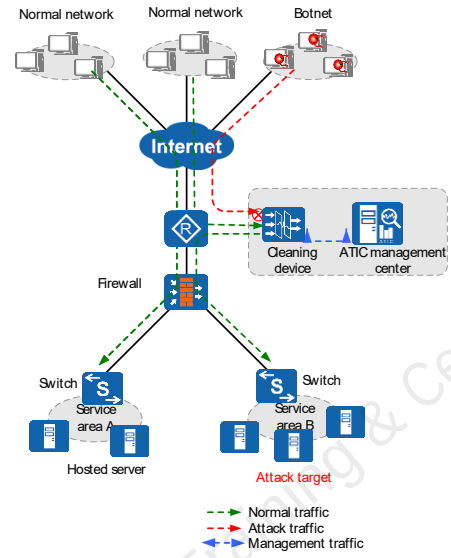
HUAWEI

- This slide describes the basic architecture and three components of Huawei anti-DDoS solution and how these components relate to each other in the solution:
  - Detecting center
  - Cleaning center
  - ATIC management center
- The detecting center detects DDoS attacks and reports attack events and traffic logs to the ATIC.
- The cleaning center diverts and cleans attack traffic, injects cleaned traffic, and reports cleaning logs to the ATIC.
- The ATIC provides a UI for users to manage anti-DDoS. On the UI, users can deliver configurations, receive detecting or cleaning logs, view attack traffic cleaning reports, and view device status.

Huawei Training & Certification

Application Scenario - IDC Protection

- With large volumes of egress traffic, diversified service types, and high availability requirements, IDCs are vulnerable to flood attacks and new application-layer attacks.



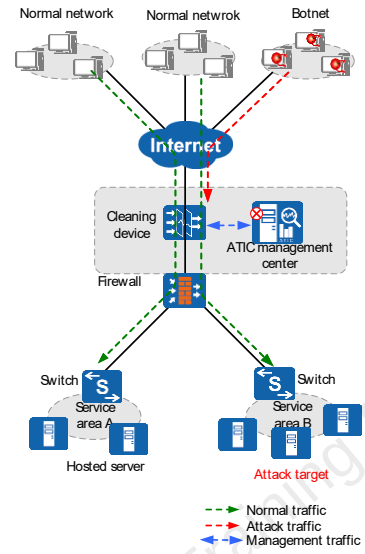
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Anti-DDoS can be deployed to filter DDoS, zombie, Trojan horse, and worm attacks, greatly improving the IDC's defense capability. Refined defense can be provided for web servers in the IDC to prevent web page tampering and Trojan horse infection.
- As shown in the figure, the cleaning device is deployed in off-line mode at the IDC egress to perform static diversion on traffic entering the IDC. When an attack occurs, the cleaning device performs real-time defense, cleans abnormal traffic, and forwards normal traffic. The cleaning device can protect multiple types of services. For different IDC customers, carriers can provide various defense functions, such as DNS authorization service protection, professional web protection, network gaming protection, and cloud service protection, to create a secure network environment.

Huawei Training & Certification

Anti-DDoS In-line Deployment

- The networking of in-line deployment is simple and does not require additional interfaces. In addition, the protection device can monitor bidirectional traffic in real time, delivering better capability than off-line deployment in defending against certain attacks. However, since all traffic will pass through the anti-DDoS device, the anti-DDoS device must be sufficiently available to prevent single points of failures and reduce false positives at the algorithm level



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

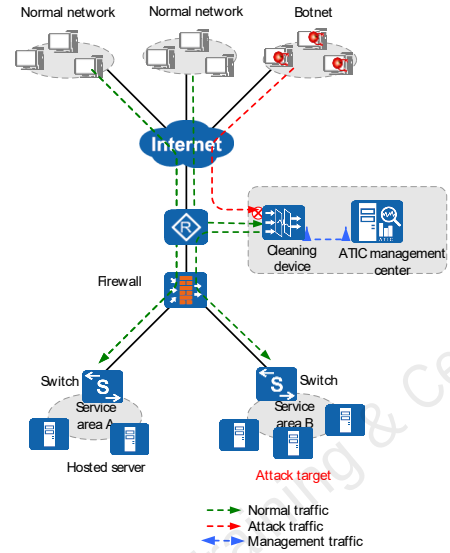
HUAWEI

- In in-line deployment, only the cleaning device needs to be deployed. The cleaning device shall be deployed in the downstream of the gateway but in the upstream of the firewall. In addition, the cleaning device needs to be connected to a bypass card or in hot standby mode to improve availability and ensure that service traffic can pass through when the cleaning device is faulty

Huawei Training & Certification

Anti-DDoS Off-line Deployment (1)

- In heavy-traffic scenarios, if anti-DDoS devices need to process all traffic, a large amount of forwarding performance will be consumed, increasing security investment. In addition, in certain carrier scenarios with complex networking, it is difficult to implement in-line deployment. And certain network service providers hope to prevent short link interruption caused by in-line deployment. Therefore, off-line deployment is introduced. Off-line deployment falls into two modes: independent deployment of the cleaning device and associated deployment of the detecting and cleaning devices.



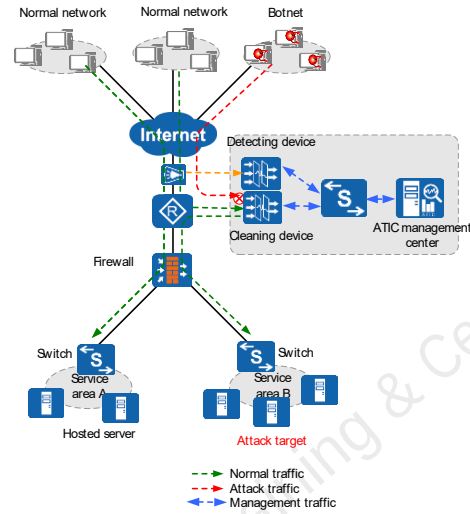
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Compared with in-line deployment, off-line deployment prevents the original networking from being damaged. In addition, the traffic direction change technology is introduced to control the traffic direction by means of traffic diversion and injection. In this way, abnormal traffic can be processed.

Huawei Training & Certification

Anti-DDoS Off-line Deployment (2)

- In associated deployment, the detecting and cleaning devices are associated through the ATIC. The detecting device detects abnormal traffic and reports the traffic to the ATIC. The ATIC delivers a traffic diversion task to the cleaning device, and the cleaning device performs traffic diversion, cleaning, and injection.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Anti-DDoS device association falls into:
  - Association between the detecting and cleaning devices
    - The anti-DDoS detecting device uses the application-based in-depth packet inspection technology to accurately identify traffic, clean up various abnormal traffic, including traffic attacks, application-layer attacks, scanning and snooping attacks, and malformed packet attacks, with a short detection delay. This can be widely applied in carrier, enterprise network, and IDC scenarios.
  - Association between the Netflow device and the anti-DDoS cleaning device
    - The Netflow traffic detection technology is applicable to heavy traffic detection due to the relatively large sampling ratio and Netflow protocol restrictions. Most backbone networks and MANs have Netflow devices deployed. Therefore, association with Netflow devices is relatively low in the cost and applicable to flood attack detection on MANs or backbone networks.

Huawei Training & Certification

Quiz

1.Which of the following are default security zones of firewalls? ( )

A. Local

B. ISP

C. DMZ

D. Default

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Answer:
  - AC

Huawei Training & Certification
Huawei Training & Certification

Summary

- Firewall deployment modes in a cloud data network
- IPS/IDS deployment modes in a cloud data network
- Anti-DDoS deployment modes in a cloud data network

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

Principles and Deployment of Converged Storage in Cloud DCs

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Foreword

- This course introduces principles and deployment of converged storage in cloud DCs and makes trainees know the definition, features, and basic configuration and deployment of converged storage.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Objectives

- On completion of this course, you will be able to know:
  - Basic principles of SAN, NAS, and RAID
  - Features of converged storage
  - Cabling and deployment of converged storage

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

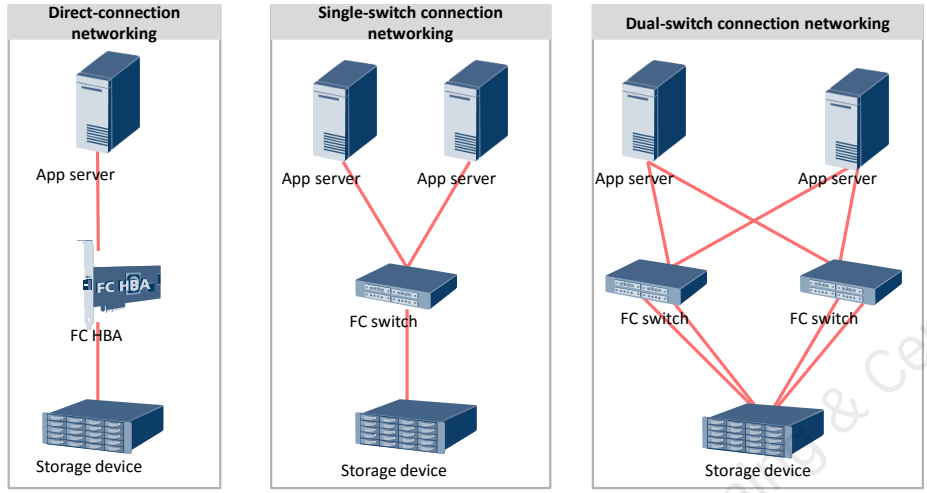
Contents

1. **SAN Principles**
   - FC SAN
   - IP SAN
2. NAS Principles
3. RAID Principles
4. Basic Features of Converged Storage
5. Deployment and Management of Converged Storage

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Definition and Networking of SAN

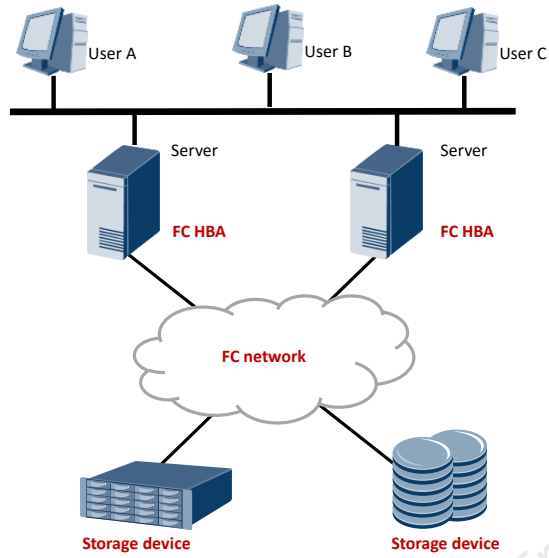


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Direct connection: A host connects to a storage device through an FC HBA. This networking mode is simple and cost-effective, but storage resources are hardly shared by other hosts.
- Single-switch: A host connects to a storage device through an FC switch. This networking mode features high scalability and allows multiple hosts to share storage resources offered by the same storage device. However, once the switch fails, the storage resources become unavailable.
- Dual-switch: A host connects to a storage device through more than one path. This networking mode features high scalability and prevents a single point of failure of the switch.

Huawei Training & Certification

FC SAN System Components



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. **SAN Principles**
   - FC SAN
   - IP SAN
2. NAS Principles
3. RAID Principles
4. Basic Features of Converged Storage
5. Deployment and Management of Converged Storage

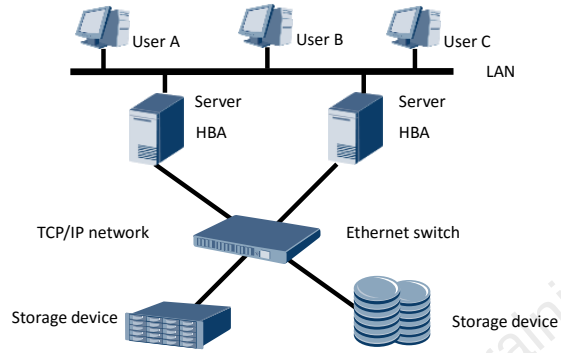
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

IP SAN Basics

- What is IP SAN?
- IP SAN is a storage area network architecture built on Ethernet and using TCP/IP as the underlying transport protocol.
- The typical protocol that implements IP SAN is iSCSI, which defines the encapsulation mode of SCSI instruction sets in IP transmission.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- IP SAN is a combination of TCP/IP protocols and SCSI instruction sets. Boasting standard protocols, low cost, and easy maintenance, IP SAN plays an important role in the network storage field.
- IP SAN transmits data blocks over the IP network. Major differences between IP SAN and FC SAN lie in transmission protocols and media. Mainstream IP SAN protocols include iSCSI, FCIP, and iFCP. iSCSI develops the fastest. To some extent, IP SAN refers to iSCSI-based SAN.
- IP SAN encapsulates SCSI instruction sets over TCP/IP. It is like we choose an express company to deliver goods to the destination. No matter which express company we choose, the goods will finally arrive at the destination. iSCSI delivers better openness and scalability because it is a standard protocol based on TCP/IP protocols and SCSI instruction sets. That is why iSCSI is so widely used.

Huawei Training & Certification

IP SAN Basics: iSCSI Connection Mode

- Depending on the connection mode between application servers and storage devices, IP SAN is classified as follows:



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- iSCSI devices use IP ports as their host ports, through which iSCSI devices are connected to Ethernet switches to form a TCP/IP-based SAN. Depending on the connection mode adopted by hosts, hosts and storage devices are connected in three modes:
- NIC + initiator software mode: The host uses standard NICs to connect to the network. The functions of the iSCSI and TCP/IP protocols are provided by the host CPU to run software. This mode requires the lowest cost because it uses the NICs universally integrated in hosts. However, this mode needs CPU resources to process iSCSI and TCP/IP protocols, which decreases host performance.
- TOE + initiator software mode: The host incorporates a TOE NIC. The functions of the iSCSI protocol are provided by the host CPU, but those of the TCP protocol are provided by the TOE NIC, relieving the workload from the host CPU.
- iSCSI HBA mode: The functions of the iSCSI and TCP/IP protocols are provided by the iSCSI HBA installed on the host. The host CPU requires the least overhead.

Huawei Training & Certification

Contents

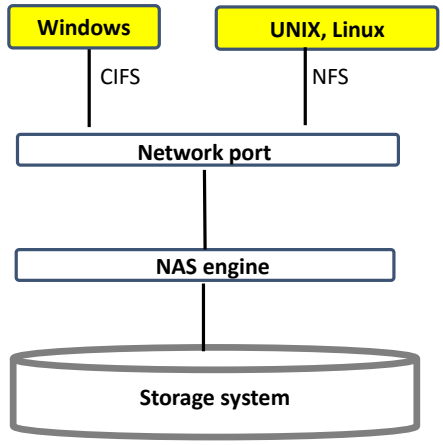
1. SAN Principles
2. **NAS Principles**
3. RAID Principles
4. Basic Features of Converged Storage
5. Deployment and Management of Converged Storage

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

NAS System Components



- NAS engine
- Provides a file system and carries the file system and operating systems of various front-end protocols.
- Network port
- Provides network protocols for user interaction, such as NFS and CIFS.
- Storage system
- The mainly used technologies are RAID, SCSI, SAS, and FC.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Introduction to CIFS

- CIFS
  - Common Internet File System (CIFS) is a protocol that enables application programs to access files and services of a remote Internet computer. CIFS can be regarded as fulfillment of an application protocol, such as File Transfer Protocol (FTP), and Hypertext Transfer Protocol (HTTP).
- Architecture
  - Client/Server mode
- Application
  - File sharing in Windows
- Transmission protocol
  - TCP/IP

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- CIFS works in client/server mode and is dedicated to file sharing in Windows. A client sends a request to a remote server for services, and then the server responds to the request. In Windows, CIFS is used to share NAS system resources.

Huawei Training & Certification

Introduction to NFS

- NFS
  - NFS is a technology for sharing files among UNIX systems. It allows data to be stored on servers and easily accessed from clients over a network. NFS allows directories and files to be shared on the network. With NFS, users and programs can access files in the remote system as easily as they access files at their local hosts.
- Architecture
  - Client/Server mode
- Application
  - UNIX and Linux
- Transmission protocol
  - TCP or UDP

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- NFS is the mainstream file sharing system used among heterogeneous platforms in the UNIX environment. It was originally developed by Sun Microsystems in 1984, allowing directories and files to be shared among heterogeneous systems, such as FreeBSD, SCO, and Solaris, over the network. Through NFS, users and programs can access files in a remote system just like accessing local files. NFS enables each computer node to utilize network resources as conveniently as it uses local resources. In the environment where various computers, operating systems, network architectures, and transmission protocols are operating, NFS helps achieve remote file access and sharing.
- NFS uses the client/server mode and involves a client program and a server program. The server program allows other computers to access file systems, and the process is called "output". The client program accesses the shared file system, and the process is called "input". Files are transmitted by blocks (a block = 8 KB). During operations, a block may be sliced into a fragment of a smaller size. NFS enables file access and sharing between servers and clients, and allows clients to access data saved in remote storage devices.

Huawei Training & Certification

Contents

1. SAN Principles
2. NAS Principles
3. **RAID Principles**
4. Basic Features of Converged Storage
5. Deployment and Management of Converged Storage

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Basic Concepts and Principles of RAID (1)

- RAID: short for redundant array of independent disks



- Implementation modes:
- Hardware RAID
- Software RAID

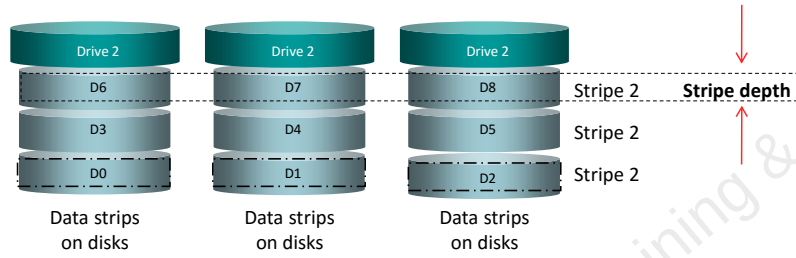
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Functions of the RAID technology are as follows:
  - Disk data is stripped so that data can be accessed in block. This reduces mechanical disk seek time and quickens data access speed.
  - Data of several disks in an array can be accessed simultaneously. This reduces mechanical disk seek time and quickens data access speed.
  - Data redundancy protection is implemented by images or by storing parity information.
- Currently, RAID can be implemented in two modes: hardware RAID and software RAID.
  - Hardware RAID uses a RAID adapter (or RAID card) that integrates a processor. The RAID adapter has its own control processor, I/O processing chip, and memory, saving CPU resources and improving data transfer speed. The RAID controller manages routes, buffer, and data flows between hosts and disk arrays.
  - Software RAID does not have its own processor or I/O chip and fully depends on the host CPU. Therefore, low-speed CPUs can hardly meet the software RAID requirements. Software RAID has two types: driver-based and OS-based. Software RAID prolongs the CPU processing period and relies on the operating system. Enterprise-class storage systems generally use software RAID.

Huawei Training & Certification

Basic Concepts and Principles of RAID (2)

- Strip: one or multiple continuous sectors on a disk. It is the element of a stripe.
- Stripe: strips in the same positions (or with the same numbers) on multiple disk drives in a disk array.

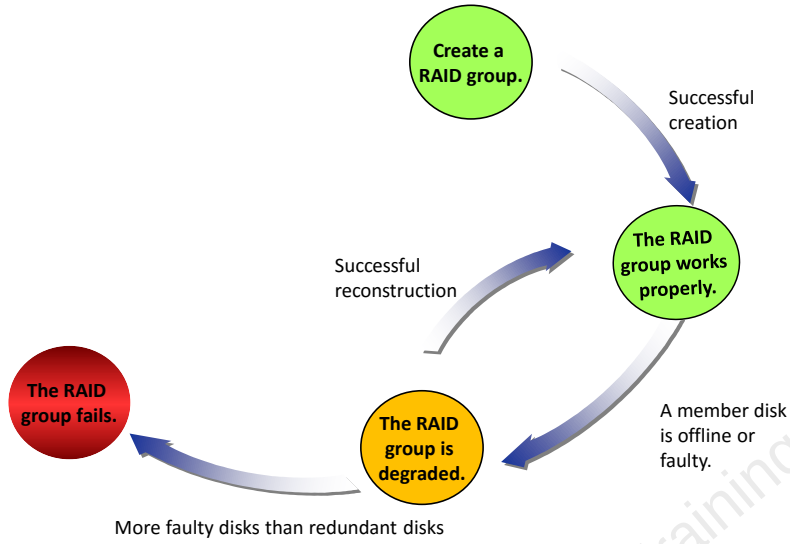


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- It is recommended that member disks of a RAID group be of the same type (specifically, the same disk capacity, disk interface type, and disk speed). If disk types are different, the RAID group performance is determined by the disk type with the lowest performance.
- Stripe width
  - Number of member disks in a stripe
- Stripe depth
  - Capacity of a strip

Huawei Training & Certification

Basic Concepts and Principles of RAID (3)



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- When a degraded RAID group is being reconstructed, if another member disk fails and the number of failed member disks is greater than that of the redundancy disks of the RAID group, the RAID group will fail. In that case, data in the RAID group will become inaccessible.

Huawei Training & Certification

RAID Level Comparison

RAID Level	RAID 0	RAID 1	RAID 3	RAID 5	RAID 6	RAID 10	RAID 50
Fault tolerance	No	Yes	Yes	Yes	Yes	Yes	Yes
Redundancy type	No	Copy	Parity check	Parity check	Parity check	Copy	Parity check
Hot spare disk	No	Yes	Yes	Yes	Yes	Yes	Yes
Read performance	High	Low	High	High	High	Minor	High
Random write performance	High	Low	Lowest	Low	Low	Minor	Low
Sequential write performance	High	Low	Low	Low	Low	Minor	Low
Min. number of disks	2	2	3	3	4	4	6
Available capacity	N x Capacity of a single disk	(1/N) x Capacity of a single disk	(N-1) x Capacity of a single disk	(N-1) x Capacity of a single disk	(N-2) x Capacity of a single disk	(N/2) x Capacity of a single disk	(N-2) x Capacity of a single disk

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Do not configure too many member disks for a RAID group. For example, a RAID group with more than 20 member disks delivers a lower performance and higher failure rate than a RAID group with eight or nine member disks (the recommended number of member disks for RAID 5).

Huawei Training & Certification

RAID 2.0 Principles (1)

- RAID 2.0
- RAID 2.0 is an enhanced RAID technology. As disk capacity become larger, it will take longer time to reconstruct a disk and thus the reconstruction window of a traditional RAID group becomes larger, which increases the probability of permanently losing data when another disk fails during reconstruction. RAID 2.0 is developed to minimize the data loss probability.
- RAID 2.0+
- RAID 2.0+ provides a finer resource granularity (tens of KB) than RAID 2.0 to serve as a unit of standard allocation and reclamation of storage resources, similar to VMs in computing virtualization. This technology is also called virtual block technology.
- Huawei RAID 2.0+
- Huawei RAID 2.0+ is a brand-new RAID technology developed by Huawei to overcome the disadvantages of traditional RAID and keep in line with the storage architecture virtualization trend. RAID 2.0+ adopts two-layer virtualized management instead of the traditional fixed management. Based on the underlying disk management that employs block virtualization (Virtual for Disk), Huawei RAID 2.0+ uses Smart-series efficiency improvement software to deliver efficient resource management that features upper-layer virtualization (Virtual for Pool).

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The core of RAID 2.0 is block-based virtualization. RAID 2.0 divides disks into multiple contiguous storage spaces of a fixed size. Each storage space is called a chunk (CK).
- Based on a RAID policy, multiple CKs form a RAID group, that is, a chunk group (CKG). A CKG is divided into multiple small data blocks (extents), and extents from different CKGs form LUNs. Hot spare CKs are spread on different disks.
- An extent is a minimum space unit of a LUN in the OceanStor 18000 series. An extent is from multiple CKGs, and chunks in a CKG are from tens, hundreds, or even thousands of disks, but not from some fixed disks like in a traditional RAID group. Therefore, the LUN space can be scattered onto an increased number of disks. RAID 2.0 divides physical space and data space into scattered blocks. This maximizes system read/write performance, facilitates system expansion, enables on-demand space allocation, and achieves data relocation based on data access frequency. Therefore, RAID 2.0 is the basis of all Smart features.
- Hot spare space is scattered onto multiple disks as chunks, and several CKGs can be reconstructed almost at the same time. Therefore, writing data into a hot spare disk is no longer a performance bottleneck, and the reconstruction time is greatly reduced.

Huawei Training & Certification

RAID 2.0 Principles (2)

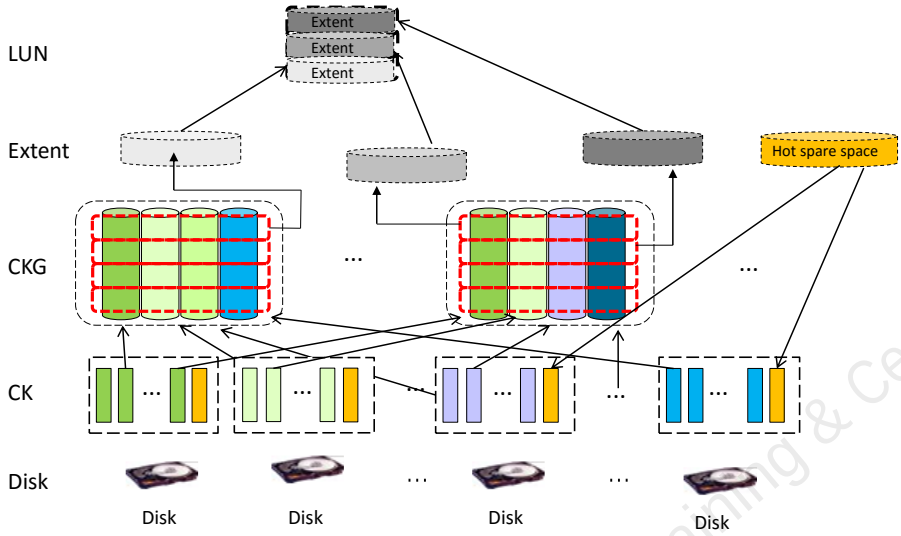
- RAID 2.0 divides disks into multiple contiguous storage spaces of a fixed size. Each storage space is called a chunk (CK).
- Based on a RAID policy, multiple CKs form a RAID group, that is, a chunk group (CKG).
- A CKG can be divided into multiple extents. A LUN is created by extents from different CKGs.
- Hot spare CKs are spread on different disks.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

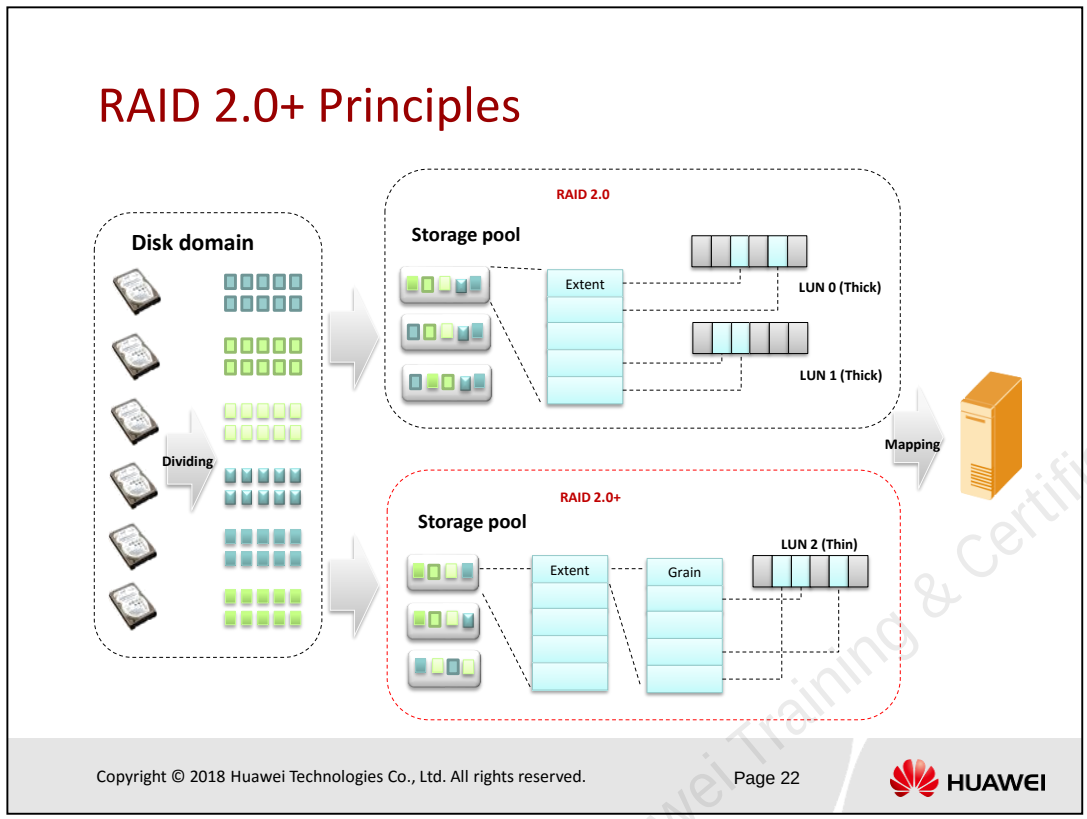
RAID 2.0 Principles (3)



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Thick LUNs are mapped into LUNs in a unit of extent.
- Extents are further divided into grains that have a finer granularity.
- Thin LUNs are mapped to LUNs in a unit of grain.

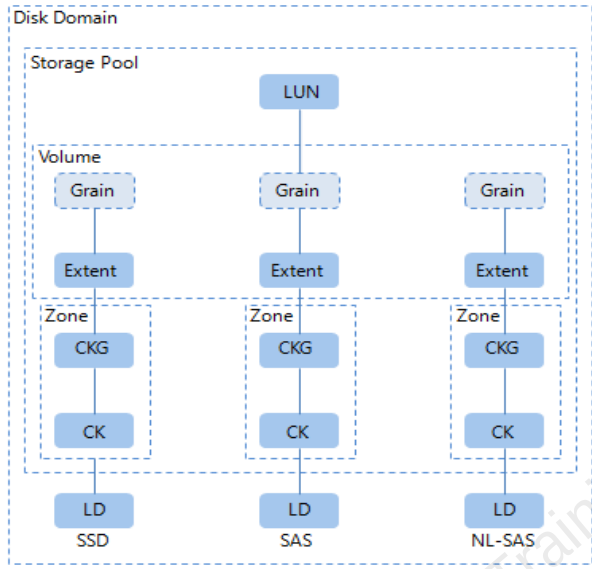
Huawei Training & Certification



- RAID 2.0+ divides each disk in the disk domain into 64 MB chunks (CKs). Disks of the same type in the disk domain form a disk group. Multiple disks are selected from a disk group at random and CKs are selected from these disks to form a chunk group (CKG) based on a specific RAID algorithm. A CKG is divided into extents of a fixed size. Based on extents, a thick LUN is mapped to a LUN. Extents are divided into grains. Based on grains, a thin LUN is mapped to a LUN.

Huawei Training & Certification

RAID 2.0+ Principles: Logical Software Objects



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

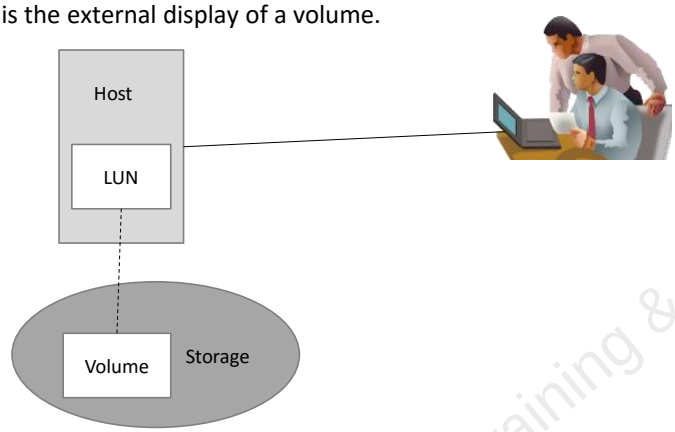
HUAWEI

Huawei Training & Certification

- Disk Domain: A disk domain can be used to create multiple storage pools. The disks that compose a disk domain can be SSD, SAS, and/or NL-SAS disks. Disk domains are isolated from each other, including their performance, storage resources, and faults.
  - Storage pool and tier: Created based on a specific disk domain, a storage pool dynamically allocates CK resources of the disk domain and enables these CKs to compose CKGs based on the RAID policy defined for each storage tier. Then, the CKGs provide applications with storage resources that have RAID protection.
  - A disk group (DG) is a collection of disks of the same type in the same disk domain. The disk types can be SSD, SAS, and NL-SAS.
  - Logical drives (LDs) are managed by a storage system. Each LD corresponds to a physical disk.
  - Each chunk (CK) is a 64 MB physical space from disk space in a storage pool. CKs are the basic components of a RAID group.
  - A chunk group (CKG) is a logical storage unit that consists of CKs from different disks in the same DG based on a RAID algorithm. A storage pool allocates resources from a disk domain by taking CKG as the smallest unit.
  - An extent is a logical storage space with a fixed size divided from a CKG. The size is adjustable. Extent is the smallest unit (granularity) for data migration and hotspot data statistics collection. It is also the smallest unit for space application and release in a storage pool.
  - In thin LUN mode, extents are divided into grains that have a finer granularity. Thin LUNs allocate resources in a unit of grain. LBA in grains is continuous. Thin LUNs are mapped to LUNs in a unit of grain. A thick LUN does not involve grains.

RAID 2.0+ Principles: Volume/LUN

- A volume is an internal management object in a storage system.
- A LUN is a storage unit that can be directly mapped to a host for data read and write. A LUN is the external display of a volume.

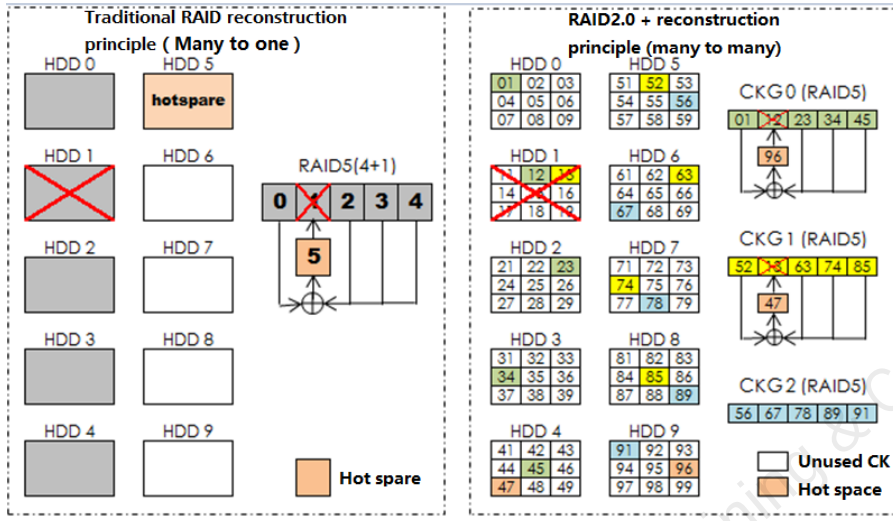


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- A volume organizes all extents and grains of a LUN. You can dynamically apply for or release extents to increase and decrease the actual space occupied by the volume.

Huawei Training & Certification

Advantages of RAID 2.0+: Fast Thin Reconstruction



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- In the schematic diagram of traditional RAID, HDDs 0 to 4 compose a RAID 5 group, and HDD 5 serves as a hot spare disk. If HDD 1 fails, an XOR algorithm is used on HDDs 0, 2, 3, and 4 to reconstruct data and write the reconstructed data onto HDD 5.
- In the schematic diagram of RAID 2.0+, if HDD 1 fails, its data is reconstructed at a CK granularity, where only the allocated CKs are reconstructed. All disks in the storage pool participate in the reconstruction. The reconstructed data is distributed on multiple disks (HDDs 4 and 9 in the figure above).
- Efficient fault handling of RAID 2.0+ accelerate reconstruction. In addition to the bad sector repair and disk failure reconstruction, RAID 2.0+ provides bad block repair that only reconstructs used space based at a CK granularity. By efficiently identifying used space, RAID 2.0+ implements thin reconstruction upon a disk failure to further shorten reconstruction time, mitigating data loss risks.

Huawei Training & Certification

Contents

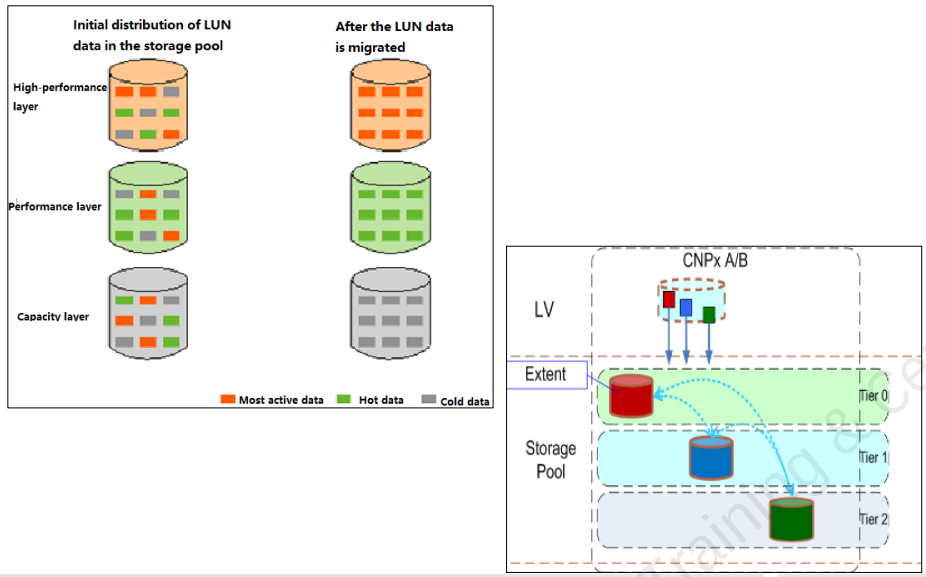
1. SAN Principles
2. NAS Principles
3. RAID Principles
4. **Basic Features of Converged Storage**
5. Deployment and Management of Converged Storage

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

SmartTier

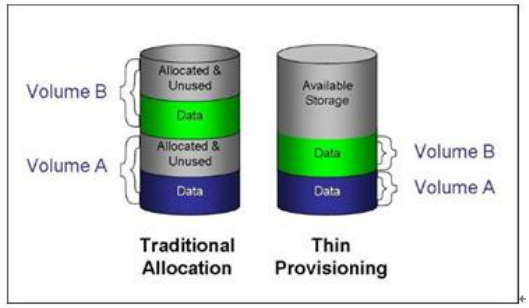


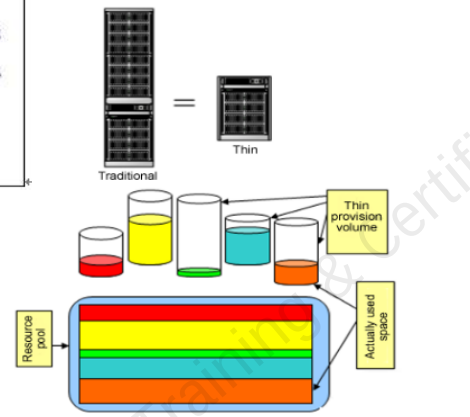
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The SmartTier dynamic storage tiering feature (SmartTier for short) helps select an appropriate storage tier for each data block based on its activity level. This feature improves the storage system performance and lowers the total cost of ownership (TCO).
- Data that is more active will be promoted to higher-performance storage media (such as SSDs), whereas data that is less active will be demoted to more cost-effective storage media with larger capacities (such as NL-SAS disks).
- SmartTier manages data placement intelligently at the LUN level. It analyzes and collects statistics about the activity level of data and determines an appropriate tier for data.
- SmartTier helps improve the overall storage system performance and reduces customers' costs. By the cost we mean the cost of purchase of software and hardware, footprint, energy, and administration.
- SmartTier does not interrupt ongoing services or affect data access. The statistical collection, analysis, and relocation executed by SmartTier are based on the SmartTier policy and performance requirements of applications. SmartTier does not affect service continuity and data availability during the mentioned actions.

Huawei Training & Certification

SmartThin





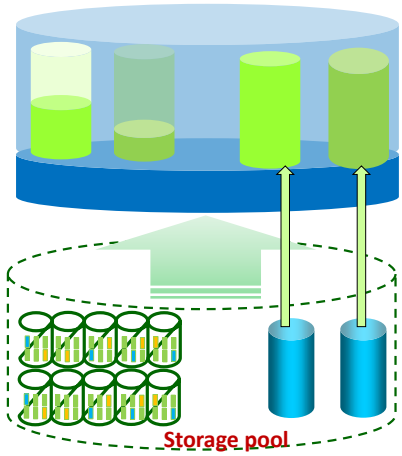
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- SmartThin is a technology that manages and allocates storage resources on demand. Based on virtualization, SmartThin reduces physical storage resources to be deployed, maximizing storage space utilization.
  - Provides a storage management approach that enables on-demand storage resource allocation.
  - Provides thin LUNs and allocates physical storage space as users demand it.
  - SmartTier, SmartThin, and SmartQoS can be used at the same time.

Huawei Training & Certification

SmartVirtualization



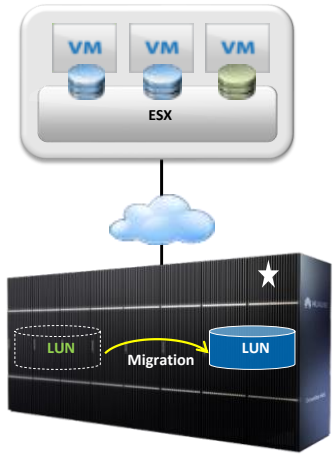
- SmartVirtualization takes over heterogeneous LUNs without using RAID 2.0+ block-level virtualization.
- Heterogeneous LUNs support HyperSnap and HyperReplication, but do not support SmartThin, SmartTier, or HyperClone.
- When snapshot or replication is configured for heterogeneous LUNs, they are automatically allocated space by their resident storage pool for snapshot COW.
- The V3 series defaults to using Write Back for heterogeneous LUNs.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

SmartMigration



Application scenarios

- Migrating data from peer vendors' products using SmartVirtualization
- Exchange between thin LUNs and thick LUNs
- RAID level migration

Purposes and benefits

- Online LUN migration, ensuring service continuity
- Complete data replication from the source LUN to the target LUN, ensuring data consistency
- Support for data migration between heterogeneous storage systems, allowing taking over peers' data

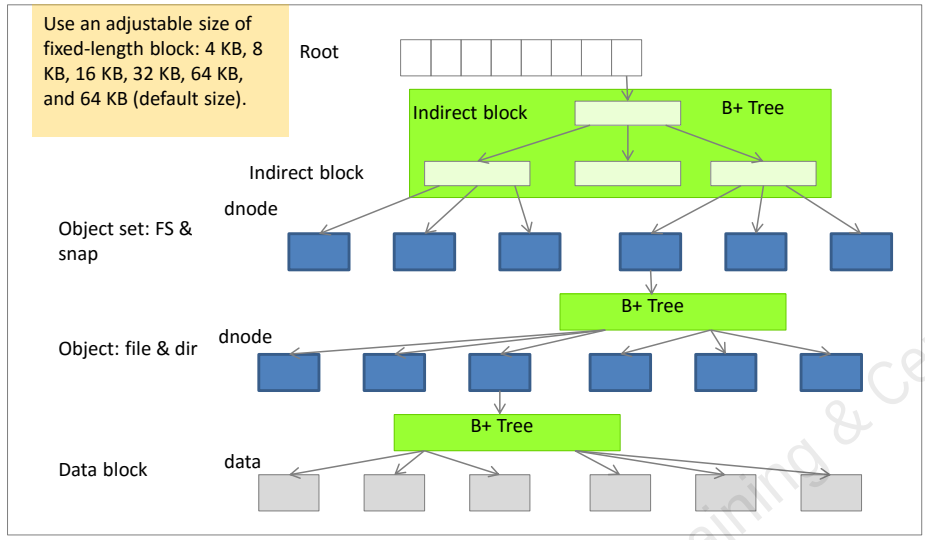
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

File System (1)

Use an adjustable size of fixed-length block: 4 KB, 8 KB, 16 KB, 32 KB, 64 KB, and 64 KB (default size).

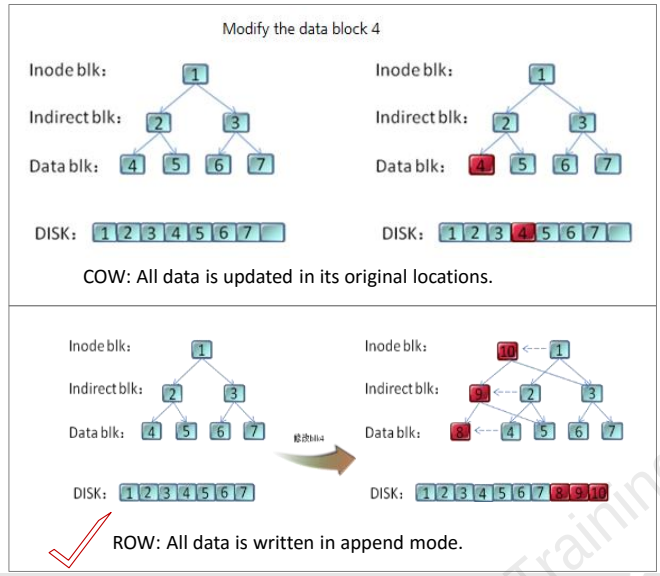


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

File System (2)

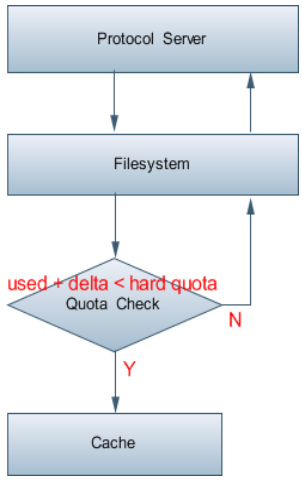


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

SmartQuota



- SmartQuota is a resource (capacity) management technology that manages directory quotas (including quotas of capacity and file quantity).
- SmartQuota affects system performance because it checks in each I/O operation whether the capacity or file quantity quota is violated.
- If a hard quota is exceeded, the I/O operation is denied and space shortfalls are reported.
- If a hard quota is not reached, the newly added capacity or files will be added to the existing ones.
- If a soft quota is exceeded, an alarm is raised.

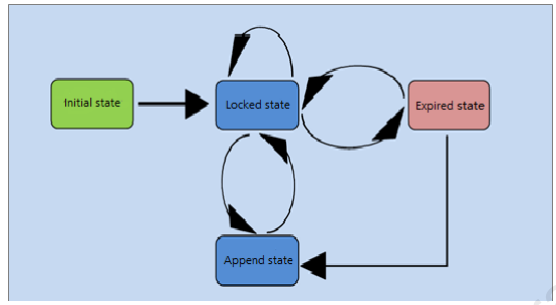
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

SmartLock

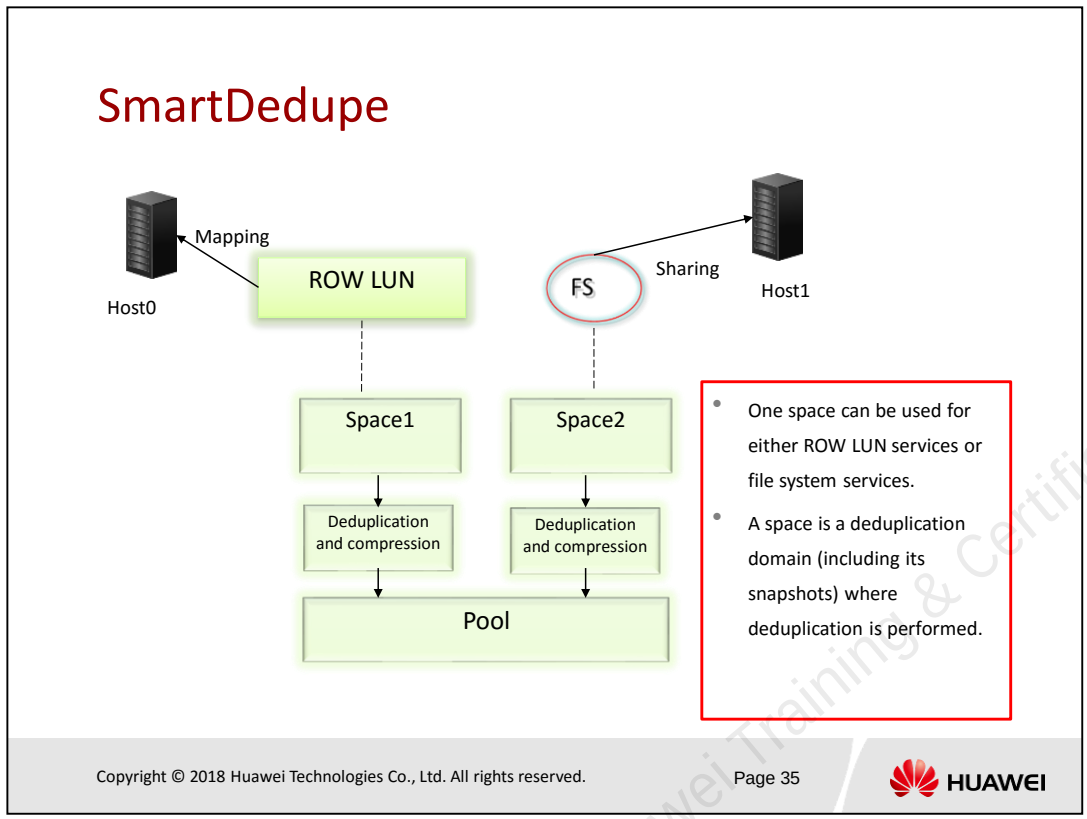
- With the Write Once Read Many (WORM) technology, a file enters the read-only state immediately after it has been written. In read-only state, the file can be read only but cannot be deleted, modified, or renamed. The major application scenario is file archiving.
- WORM has two modes: enterprise WORM and compliance WORM. The former allows administrators to flexibly manage files, and is mainly used in internal control of enterprises. The latter implements data protection in compliance with laws and regulations, allowing enterprises to avoid legal risks in confidential file archiving.
- Currently, the V3 series supports compliance WORM only.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

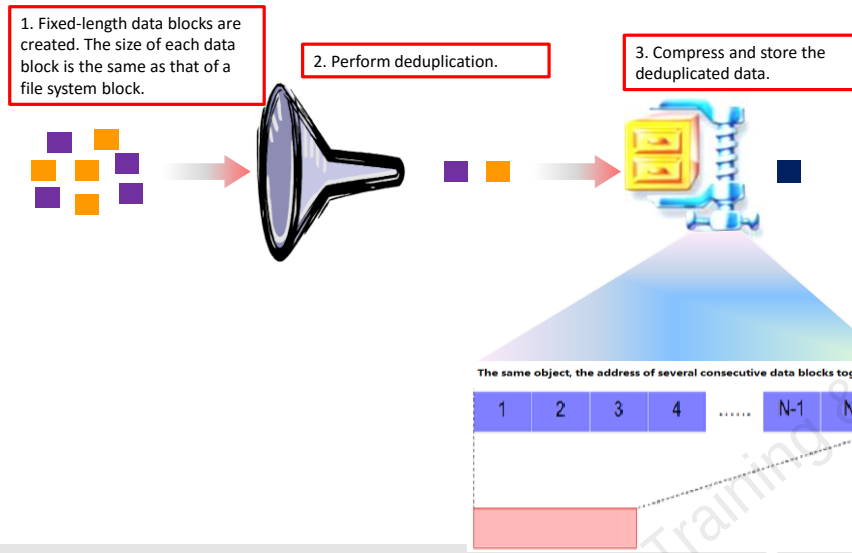
Huawei Training & Certification
Huawei Training & Certification



- SmartDedupe and SmartCompression reduce the amount of data to be stored, improving storage efficiency of file systems. In V3, ROW LUNs and file systems are derived from spaces. Data is deduplicated before being written into pools from spaces.

Huawei Training & Certification

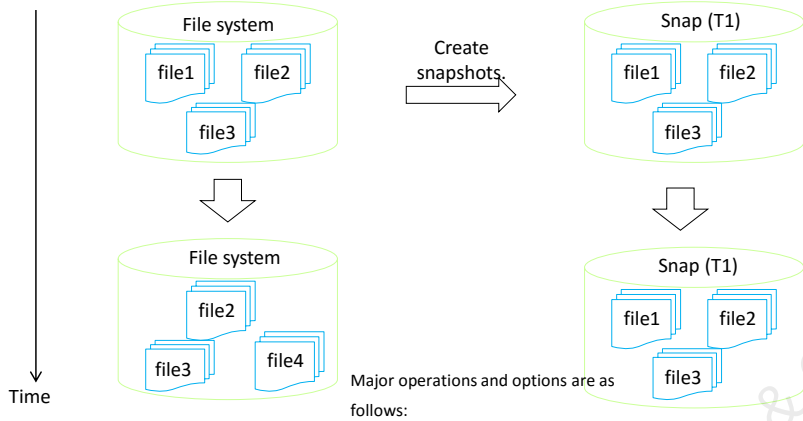
SmartCompression



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

HyperSnap



Major operations and options are as follows:

Manual snapshot creation

Periodic snapshot creation

Snapshot rollback

Snapshot deletion

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification

- By generating a consistent image of the source file system at a certain point in time, the file system snapshot function enables users to quickly obtain a data duplicate identical to the source file system. This process does not interrupt the work of the source file system. A snapshot is available immediately after being generated. Data read and write on the snapshot do not affect data in the source file system. Likewise, data updates in the file system do not affect the snapshot. This feature is applicable to online backup, data restoration, data analysis, and application testing scenarios.

Huawei Training & Certification

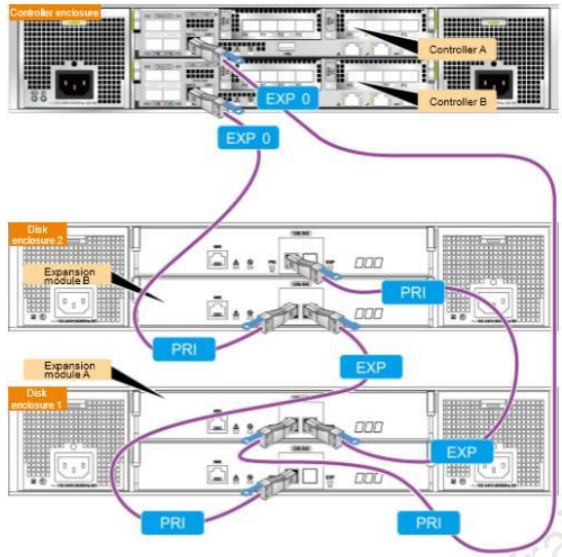
Contents

1. SAN Principles
2. NAS Principles
3. RAID Principles
4. Basic Features of Converged Storage
5. **Deployment and Management of Converged Storage**
   - OceanStor V3 Network Cabling
   - OceanStor V3 Service Configuration and Management

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Connecting Disk Enclosures (2200/2600/5300/5500 V3)



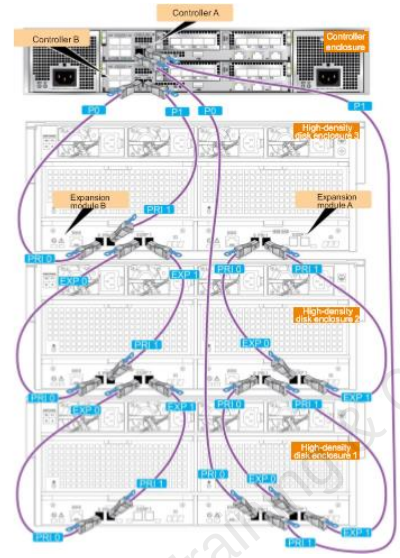
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- All EXP expansion ports on a storage system connect only to PRI expansion ports. Otherwise, services are interrupted.
- The PRI port on expansion module A or B of the disk enclosure with coffer disks must be connected to port P0 of controller A or B.
- If a storage system is paired with more than two disk enclosures, form multiple loops and ensure that the disk enclosures are evenly distributed to each loop.
- Disk enclosures except the ones with coffer disks in the same expansion loop must form two independent links that are redundant for each other to archive optimal networking reliability.

Huawei Training & Certification

Connecting Disk Enclosures (High-Density Disk Enclosures)

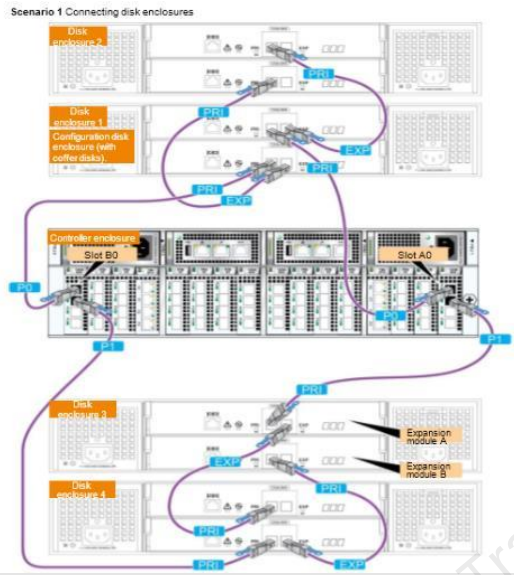
- High-density disk enclosures are connected through mini SAS HD cables. The cables must be naturally bent. At least 97 mm clearance is reserved in front of the disk enclosure for cabling.
- Non-high-density and high-density disk enclosures cannot be simultaneously connected to a SAS interface module.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Connecting Disk Enclosures (5600/5800 V3)



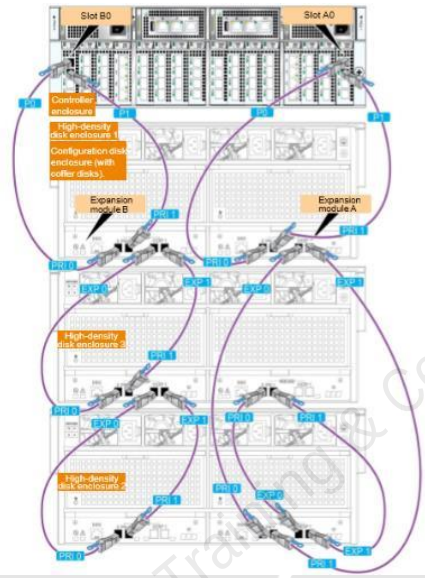
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- All EXP expansion ports on a storage system connect only to PRI expansion ports. Otherwise, services are interrupted.
- The PRI port on expansion module A or B of the disk enclosure with coffer disks must be connected to port P0 of controller A or B.
- If a storage system is paired with more than two disk enclosures, form multiple loops and ensure that the disk enclosures are evenly distributed to each loop.
- Disk enclosures except the ones with coffer disks in the same expansion loop must form two independent links that are redundant for each other to archive optimal networking reliability.

Huawei Training & Certification

Connecting Disk Enclosures (High-Density Disk Enclosures)

- High-density disk enclosures are connected through mini SAS HD cables. The cables must be naturally bent. At least 97 mm clearance is reserved in front of the disk enclosure for cabling.
- Non-high-density and high-density disk enclosures cannot be simultaneously connected to a SAS interface module.
- Only ports 0 and 1 or ports 2 and 3 are paired to connect to the same high-density enclosure.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Contents

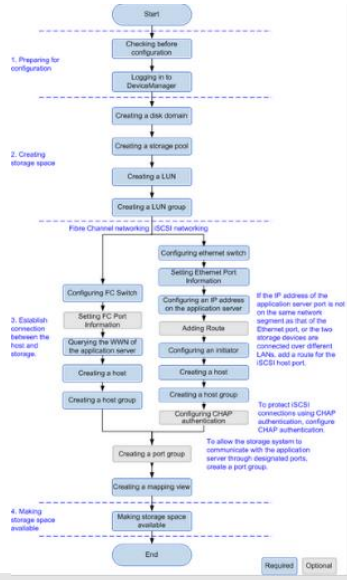
1. SAN Principles
2. NAS Principles
3. RAID Principles
4. Basic Features of Converged Storage
5. **Deployment and Management of Converged Storage**
   - OceanStor V3 Network Cabling
   - OceanStor V3 Service Configuration and Management

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Process for Configuring Storage Space: SAN

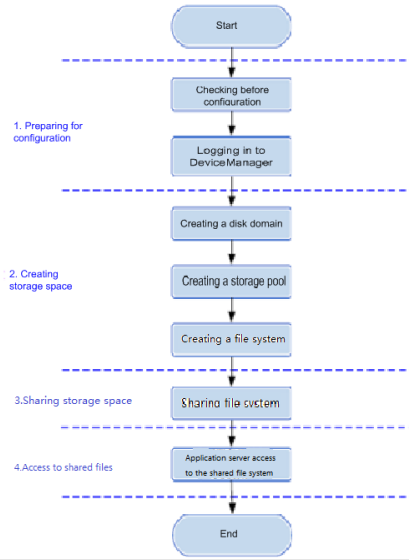


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Process for Configuring Storage Space: NAS



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Preparations

- Preparations include:
  - Pre-configuration check
  - DeviceManager check
- Check that software installation and initial configuration meet requirements for storage space configuration.

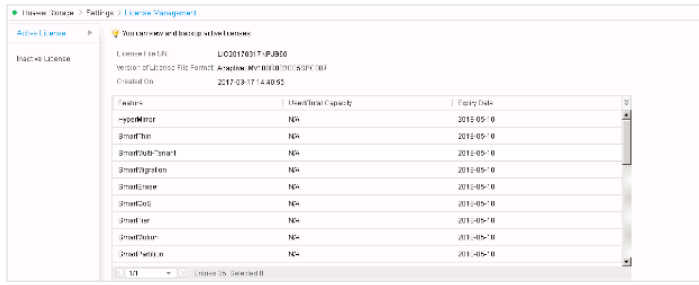
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Checking the License File Before Configuration

- Log in to DeviceManager. Check whether the information about basic features or value-added features in the license file is the same as purchased.
- If they are different, contact Huawei technical support engineers.



- To obtain the guide for applying for license, go to the following website:
- http://support.huawei.com/enterprise/docinforeader.action?contentId=DOC1000064535

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Checking DeviceManager

- DeviceManager is a piece of integrated storage management software developed by Huawei. It has been loaded to the storage system before delivery.
- You can log in to DeviceManager on any maintenance terminal connected to the storage system by entering the management network port IP address of the storage system and the local or domain user name in a browser.

OS	Fully Compatible Browser	Partially Compatible Browser
Windows 7+ Professional (32/64)	•Internet Explorer 9 to 11 •Firefox 25 to 35 •Chrome 27 to 40	Internet Explorer 8
Windows Server 2012	•Internet Explorer 10, 11 •Firefox 25 to 35 •Chrome 27 to 40	-
Windows 8	•Internet Explorer 10, 11 •Firefox 25 to 35 •Chrome 27 to 40	-
Windows 8.1	•Internet Explorer 11 •Firefox 25 to 35 •Chrome 27 to 40	-
Mac OS X 10.5 or later	Safari 5.5 to 8.0	-
Ubuntu 11	Firefox 25 to 35	-

- Note: DeviceManager has Demo. Download address: SUPPORT > IT > Storage > Cloud Storage > Tools and Platform > ISM-U

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Procedure for Logging in to DeviceManager

- Step 1 Start Internet Explorer on the maintenance terminal.
- Step 2 Enter https://XXX.XXX.XXX.XXX:8088 in the address box and press Enter.
- Step 3 (Optional) Set a login mode and language.
- Step 4 Enter the user name, password, and verification code.
- Step 5 Click log in .

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

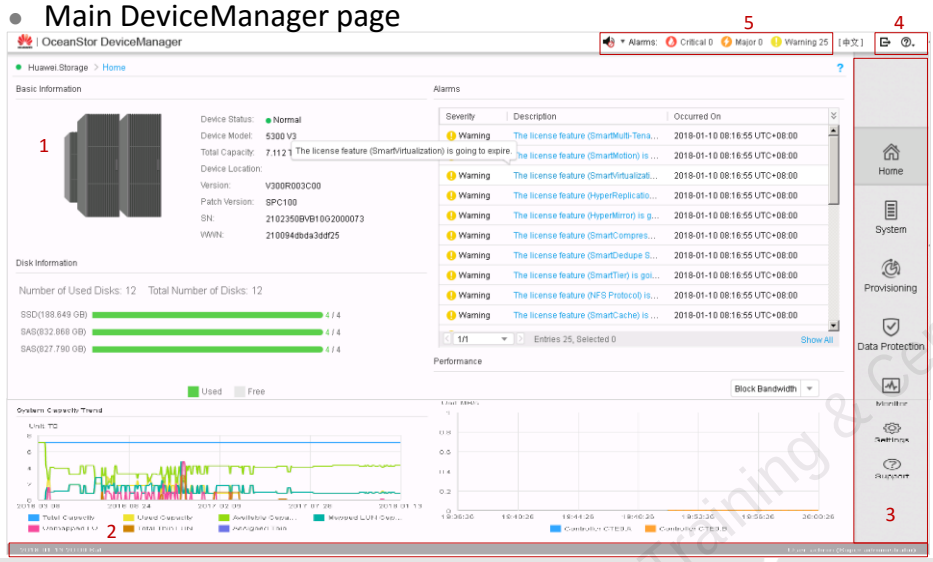
HUAWEI

- In step 2, XXX.XXX.XXX.XXX indicates the management network port IP address of a storage system.
- Step 3 (Optional) Set a login mode and language.
  - 3.1 Click 高级.
  - 3.2 Select a login mode in 认证模式.
  - 3.3 Choose a language from the 语言 list.

Huawei Training & Certification
Huawei Training & Certification

Introduction to the DeviceManager Operation Page

- Main DeviceManager page

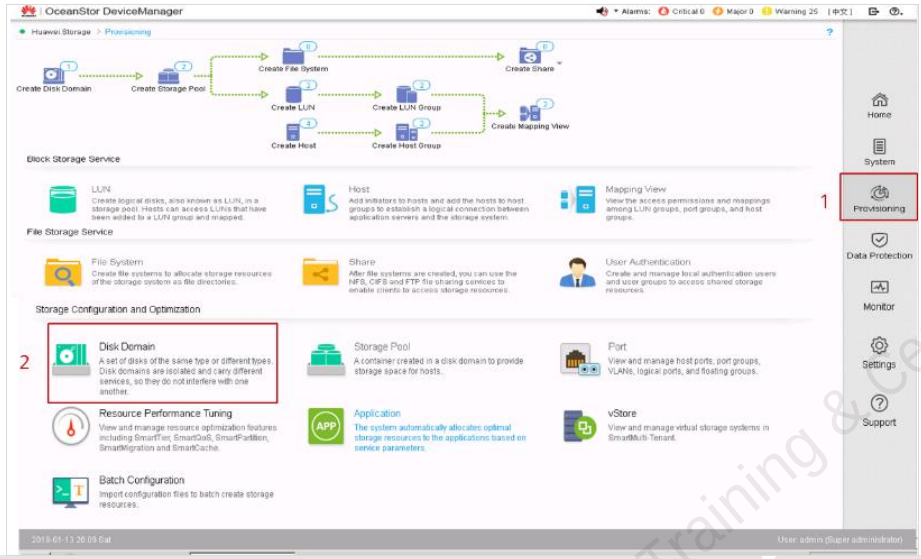


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- 1. Information display area: Displays available functions related to the current operation.
- 2. 2. Status bar: Displays the user name and login time of the current user.
- 3. Navigation bar: Lists each functional module of the entire storage system and logically divides the storage system into several modules.
- 4. Logout and Help area: Lists the logout and help buttons.
- 5. Fault statistics area: displays the number of faults by severity and helps users learn about the running status of the system.

Huawei Training & Certification

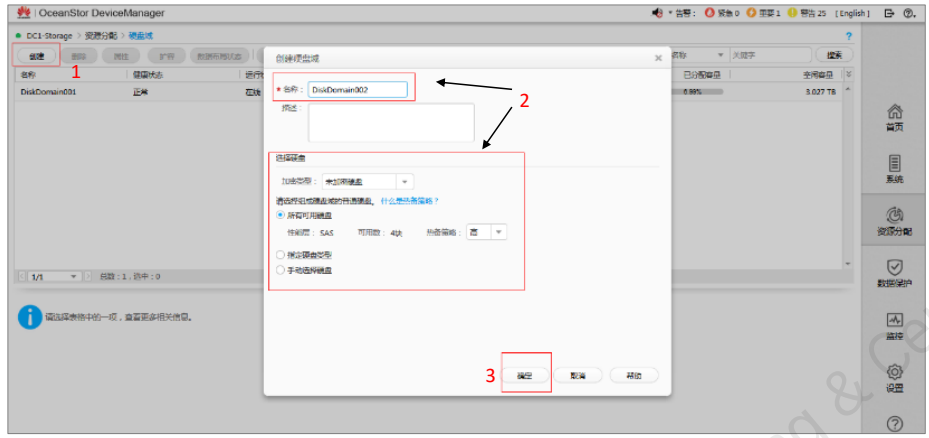
Configurations: Creating a Disk Domain (1)



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configurations: Creating a Disk Domain (2)

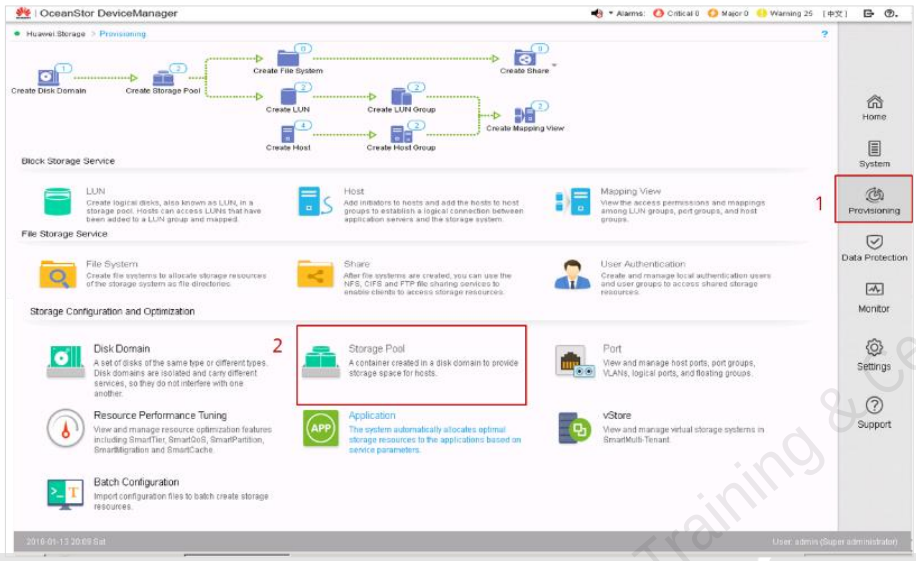


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- 1. Click Create.
- 2. Set parameters, including the name, description, and disk type.
- 3. Click OK.

Huawei Training & Certification

Configurations: Creating a Storage Pool (1)

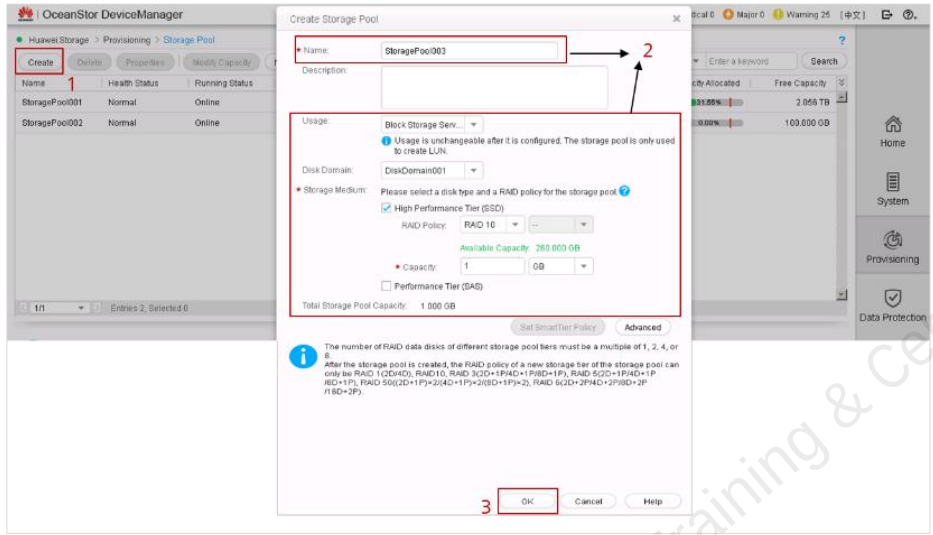


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Navigation path: 资源分配 > 存储池

Huawei Training & Certification

Configurations: Creating a Storage Pool (2)

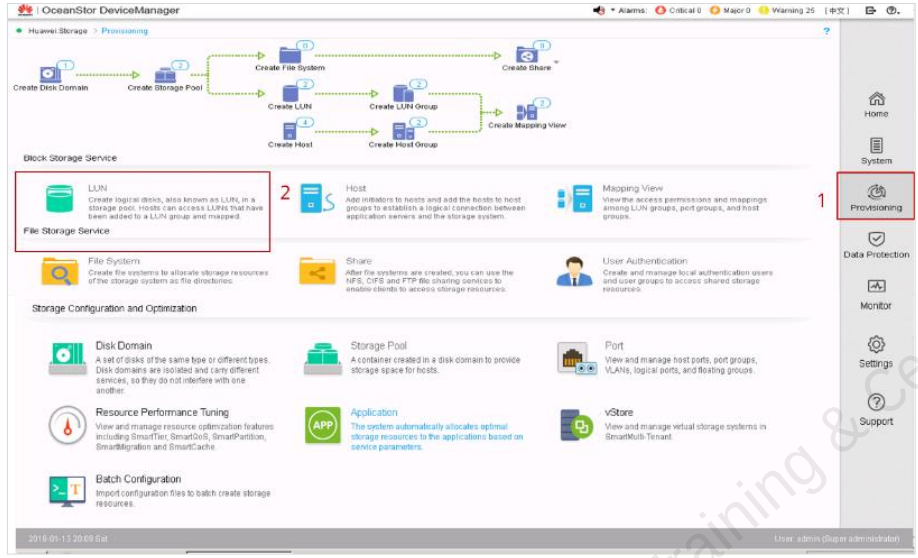


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- 1. Click Create.
- 2. Set parameters, including the name, description, disk domain, storage medium, and advanced attribute.
- 3. Click OK.

Huawei Training & Certification

Configurations: Creating a LUN (1)

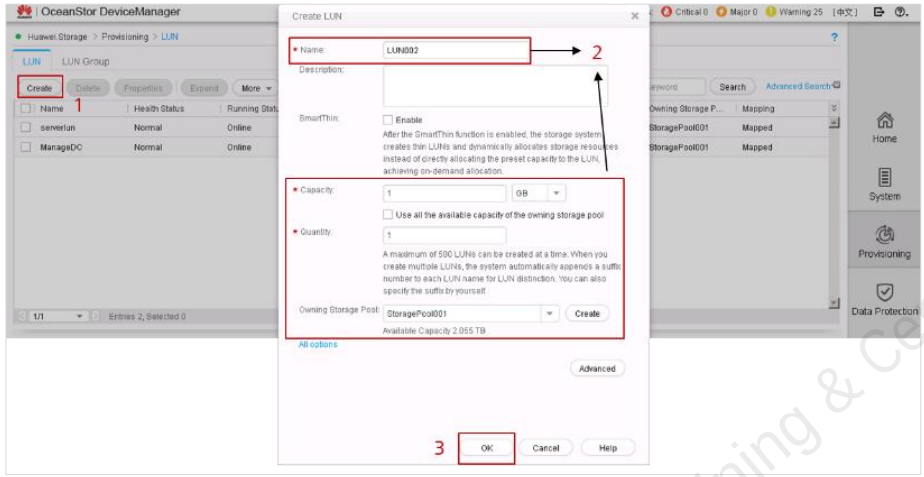


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Navigation path: 资源分配 > LUN

Huawei Training & Certification

Configurations: Creating a LUN (2)

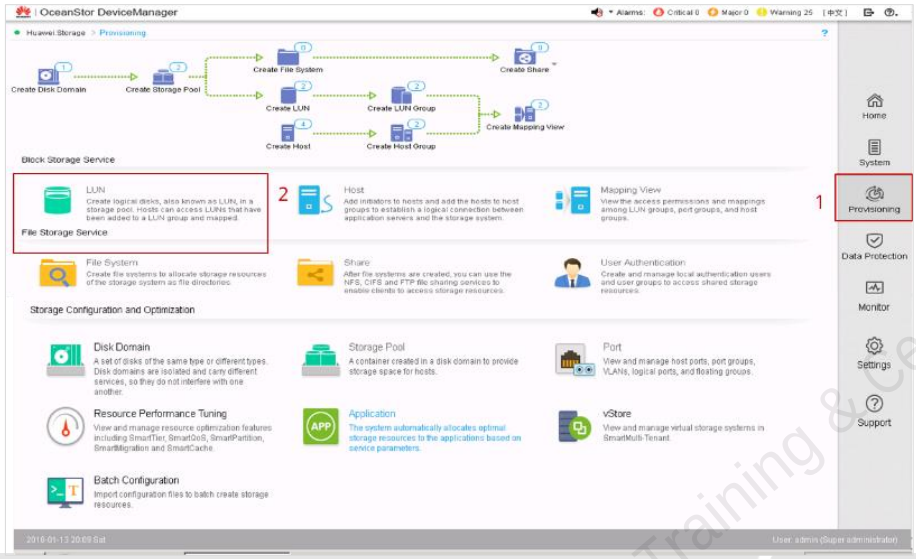


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- 1. Click Create.
- 2. Set parameters, including the name, description, SmartThin, capacity, quantity, and owning storage pool, and advanced attributes.
- 3. Click OK.

Huawei Training & Certification

Configurations: Creating a LUN Group (1)

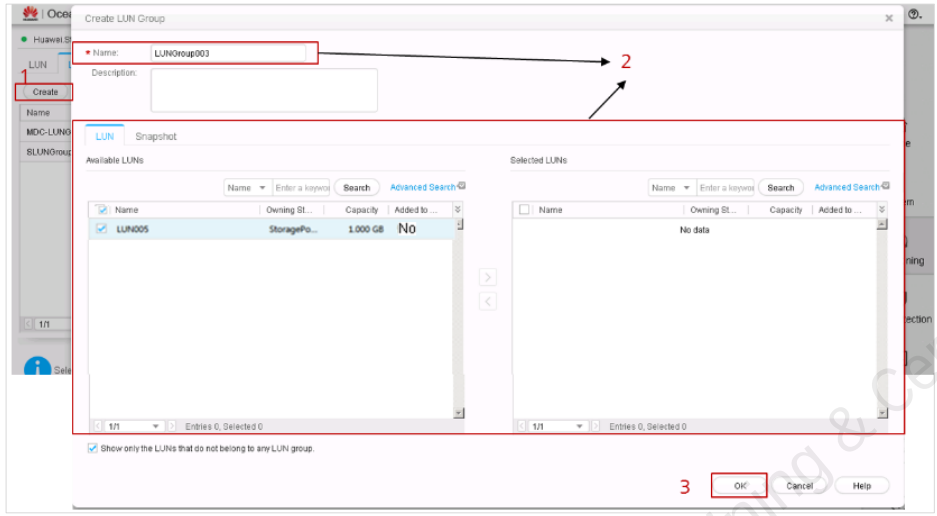


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Navigation path: 资源分配 > LUN

Huawei Training & Certification

Configurations: Creating a LUN Group (2)

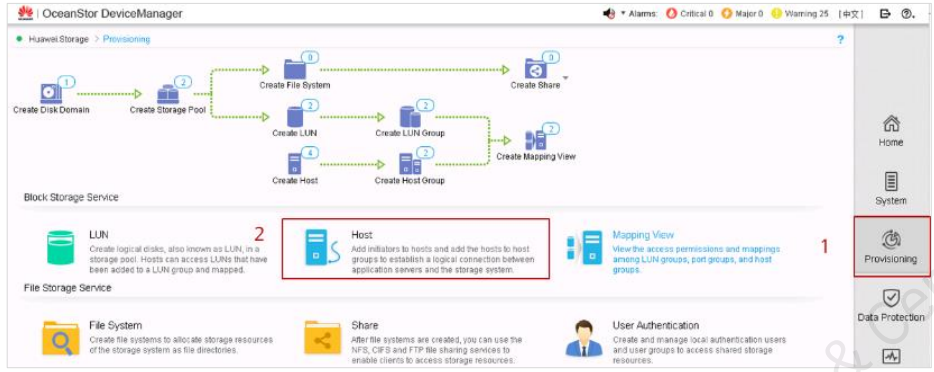


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- 1. Click a LUN group.
- 2. Click Create.
- 3. Set parameters, including the name, description, and optional LUN and snapshot.
- 3. Click OK.
- 3. Click Close.

Huawei Training & Certification

Configurations: Creating a Host

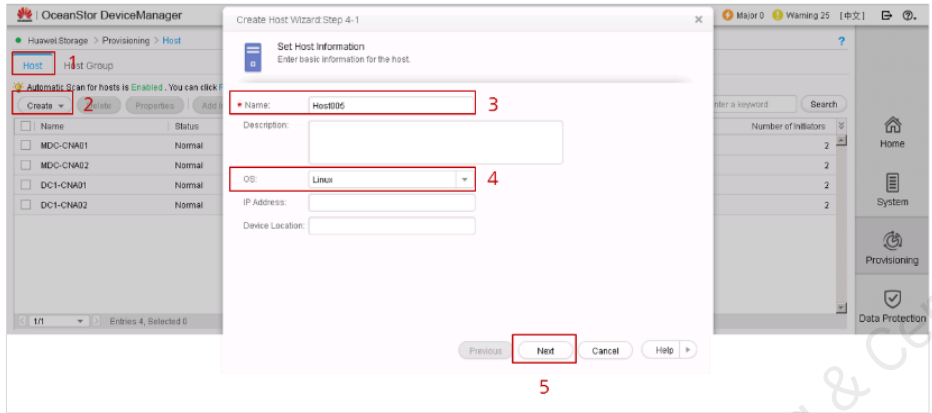


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Navigation path: 主机>主机

Huawei Training & Certification

Configurations: Manually Creating a Host (1)

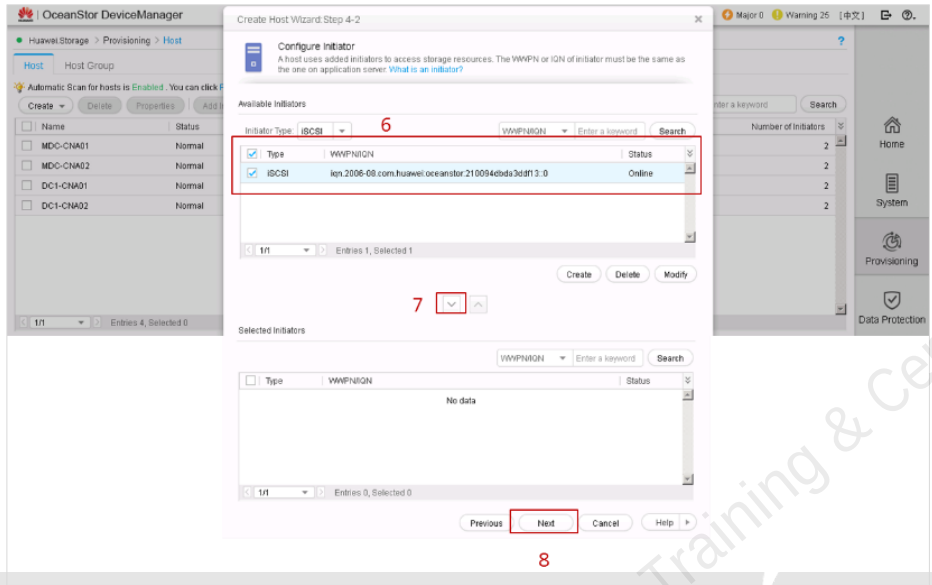


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- 1. Click Create > 手动创建.
- 2. Set parameters, including the name, description, OS, IP address, and device location.
- 3. Click Next.
- 4: Select an available initiator.
- 3. Click Next.

Huawei Training & Certification

Configurations: Manually Creating a Host (2)

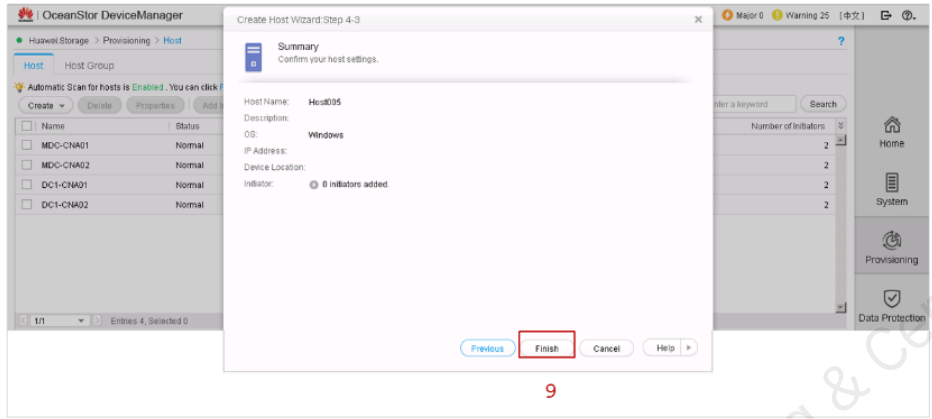


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- 1. Click 完成.
- 2. Click OK.
- 3. Click Close.

Huawei Training & Certification

Configurations: Manually Creating a Host (3)



9

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Configurations: Creating a Host Group (1)



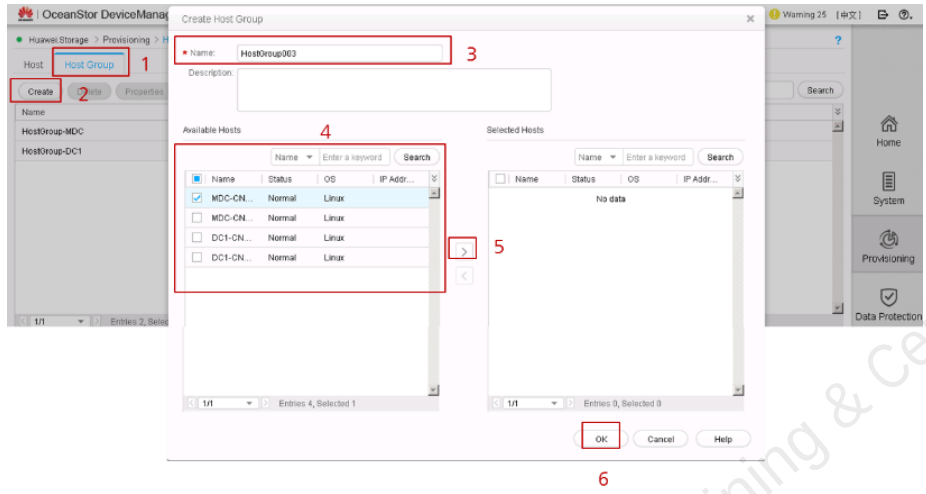
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Navigation path: 主机 > 主机

Huawei Training & Certification

Configurations: Creating a Host Group (2)

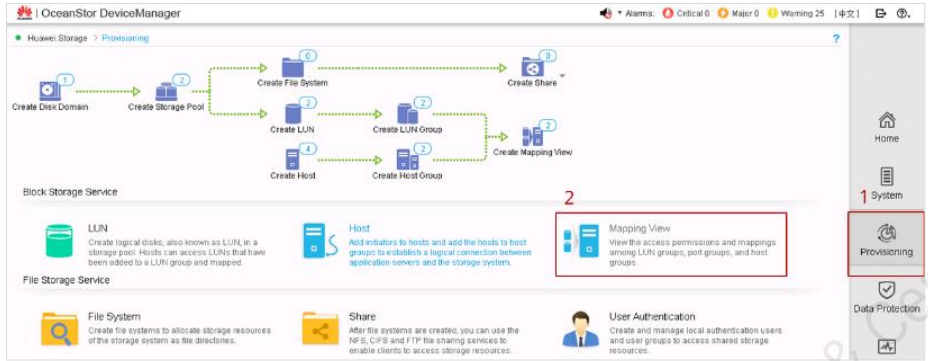


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- 1. Click a host group.
- 2. Click Create.
- 3. Set the name, description, and available host.
- 4. Click OK.
- 5. Click Close.

Huawei Training & Certification

Configurations: Creating a Mapping View (1)

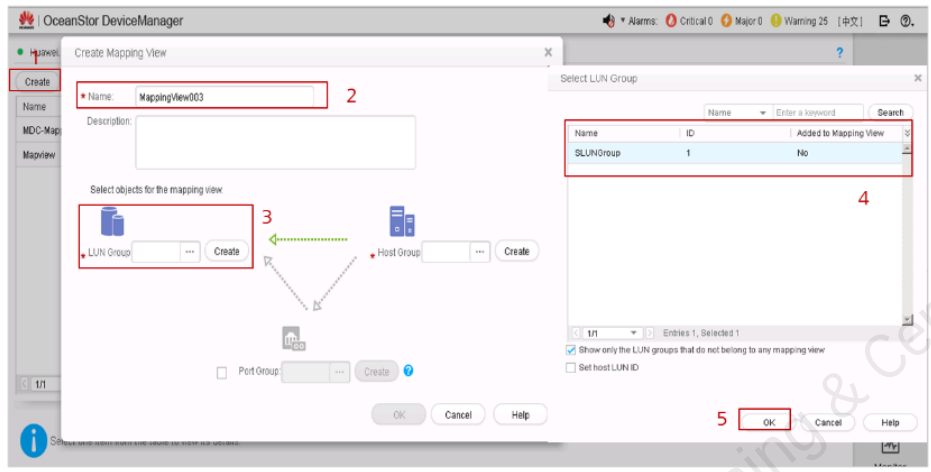


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Navigation path: 主机 > 映射视图

Huawei Training & Certification

Configurations: Creating a Mapping View (2)

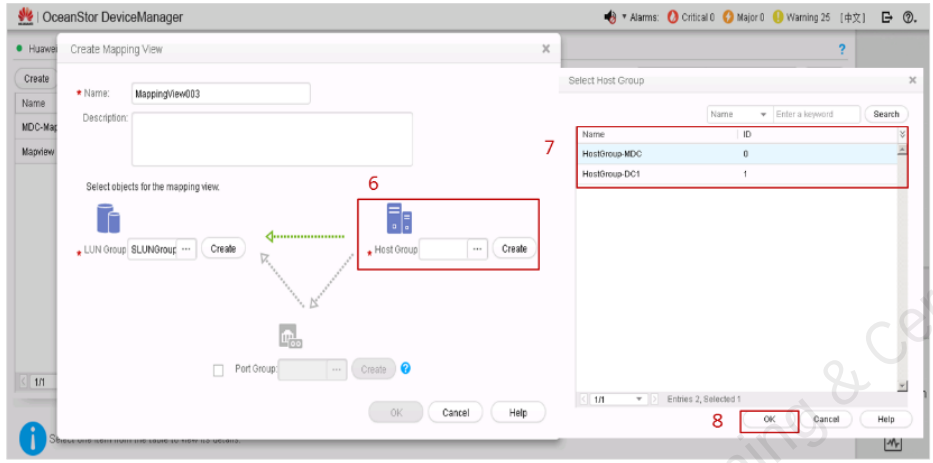


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- 1. Click Create.
- 2. Set the name and description.
- 3. Select a LUN group.
- 4. Select a host group.

Huawei Training & Certification

Configurations: Creating a Mapping View (3)



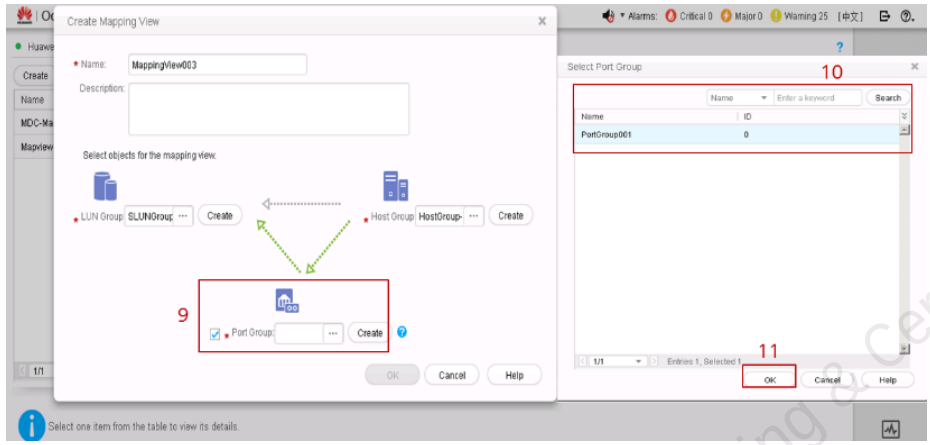
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- 1. Select, create, or ignore a port group based on site requirements.
- 2. Click OK.
- 3. Click Close.

Huawei Training & Certification
Huawei Training & Certification

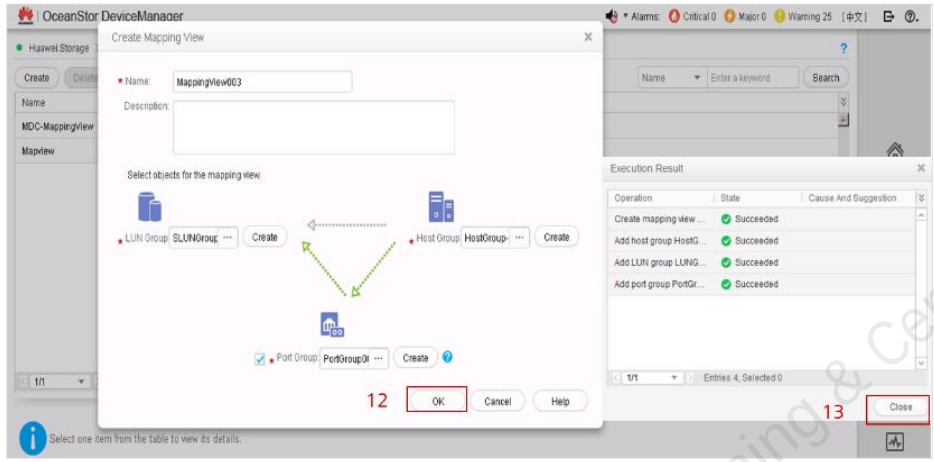
Configurations: Creating a Mapping View (4)



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

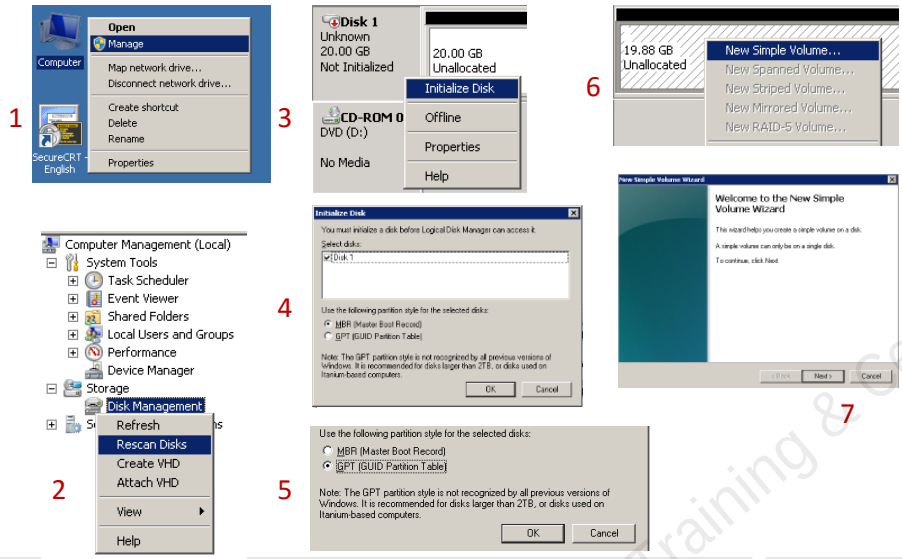
Configurations: Creating a Mapping View (5)



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configurations: Scanning for LUNs on a Windows Host (1)

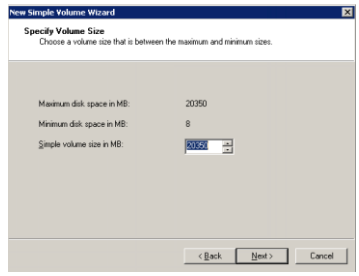


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

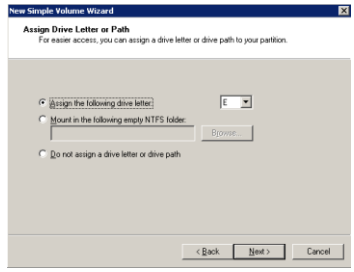
HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configurations: Scanning for LUNs on a Windows Host (2)

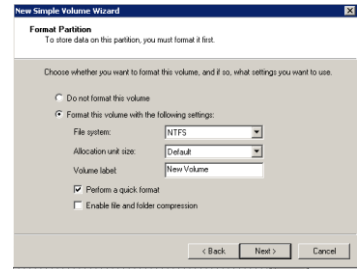
8



9



10



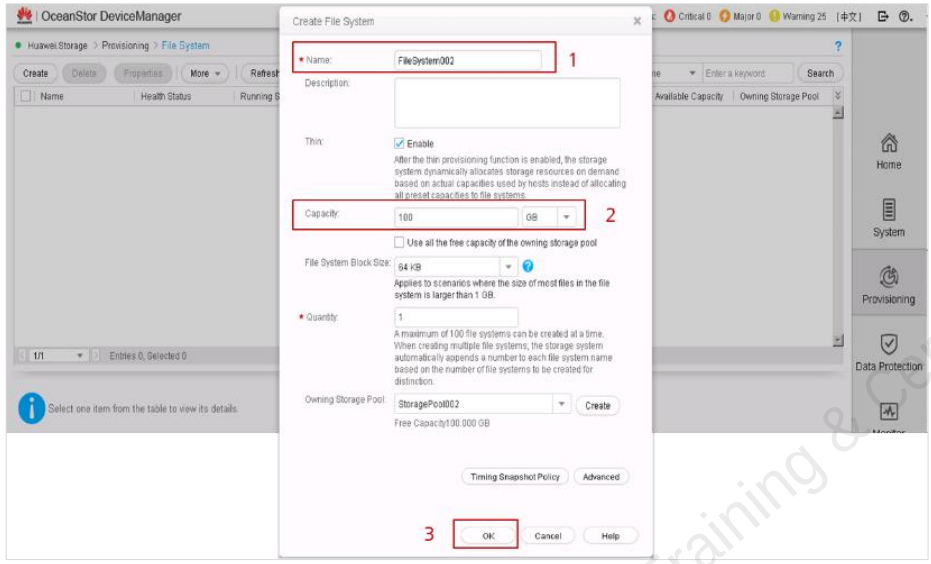
11



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Configurations: Creating a File System

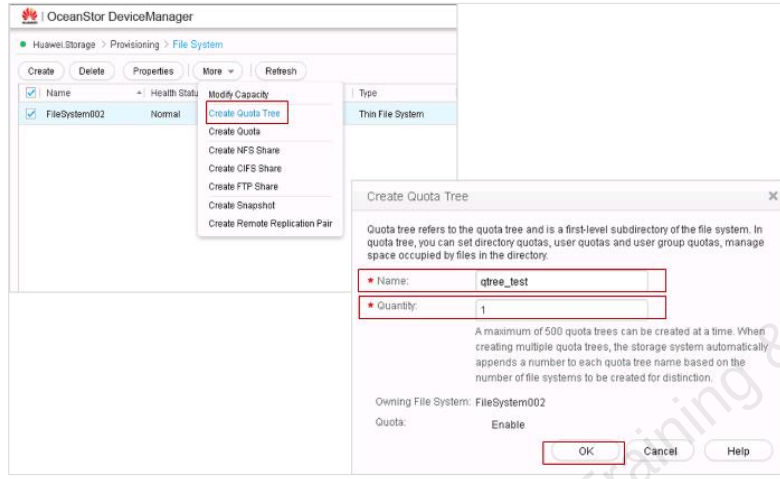


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configurations: Quota Tree

- A quota tree is a special directory of a file system. You can set a directory quota on the quota tree to manage the space used by all files under the directory.



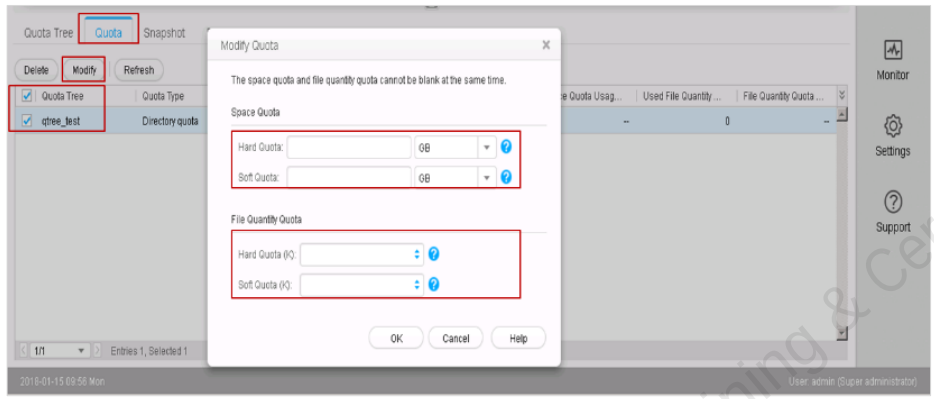
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configurations: Creating a Quota

- Quotas can be defined only for quota trees of a file system based on customer requirements.

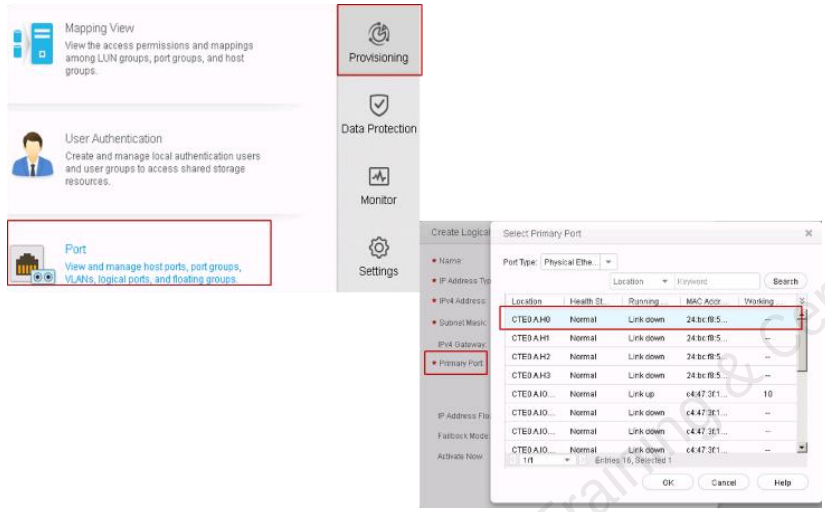


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configurations: Creating a Logical Port

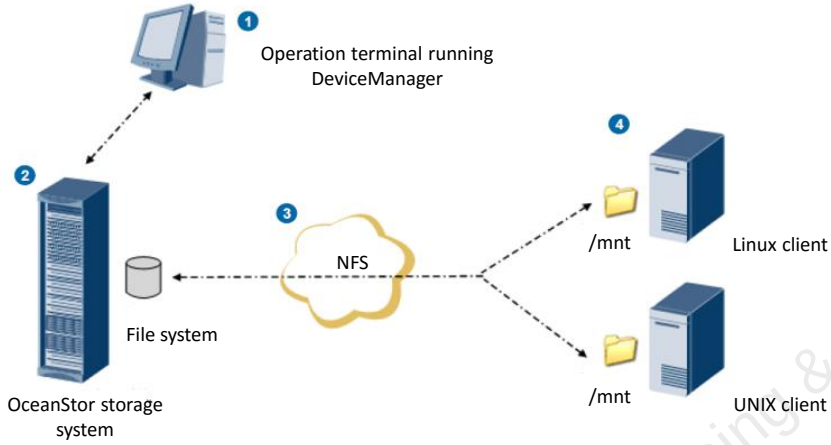
- An Ethernet port can be used to create multiple logical ports to provide services for the client.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Configurations: NFS Sharing Principles



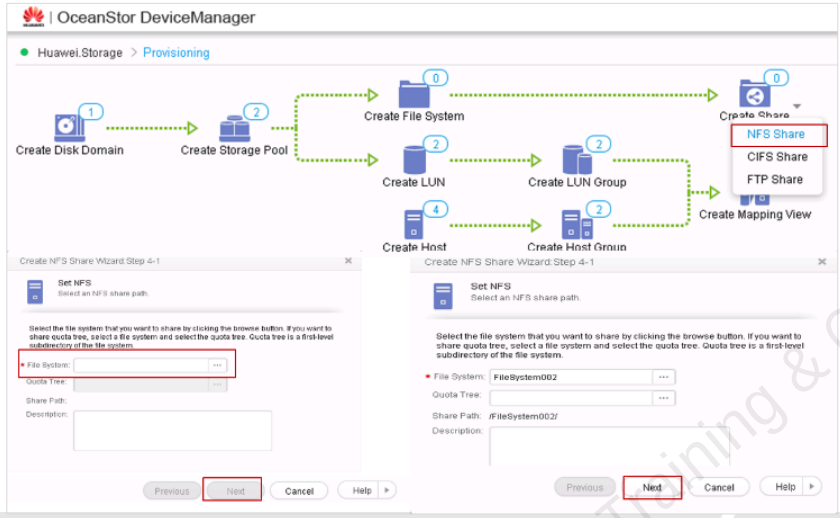
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configurations: Setting NFS Sharing

- NFS sharing allows files to be shared between Linux/UNIX clients.

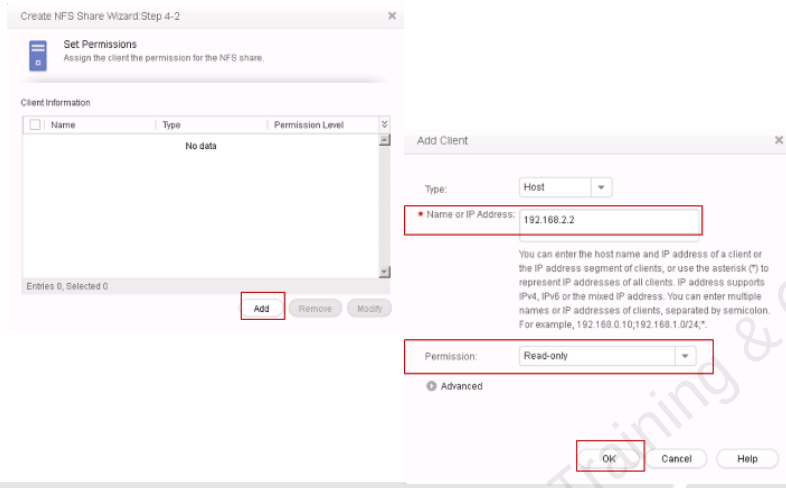


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Configurations: Setting an NFS Share Permission

- An NFS share permission can be set to read only or r/w based on the IP address, IP network segment, or network group.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configurations: Mounting an NFS Share

- Log in to the Linux/UNIX client and run the showmount command to check the NFS sharing list.
- Log in to the Linux/UNIX client and run the mount command to mount the NFS share.

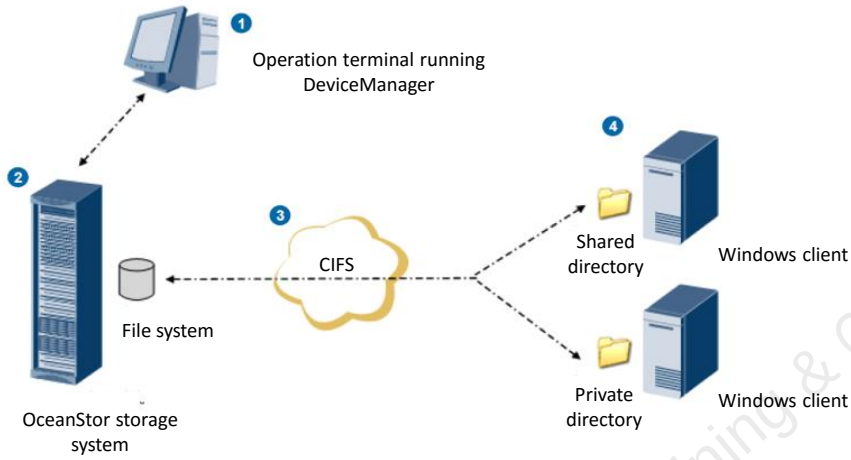
```
[root@localhost ~]# showmount -e 172.16.16.75
Export list for 172.16.16.75:
/test1 172.16.16.72
[root@localhost ~]# mount -t nfs 172.16.16.75:/test1 /mnt/nfs
[root@localhost ~]# mount
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configurations: CIFS Sharing Principles



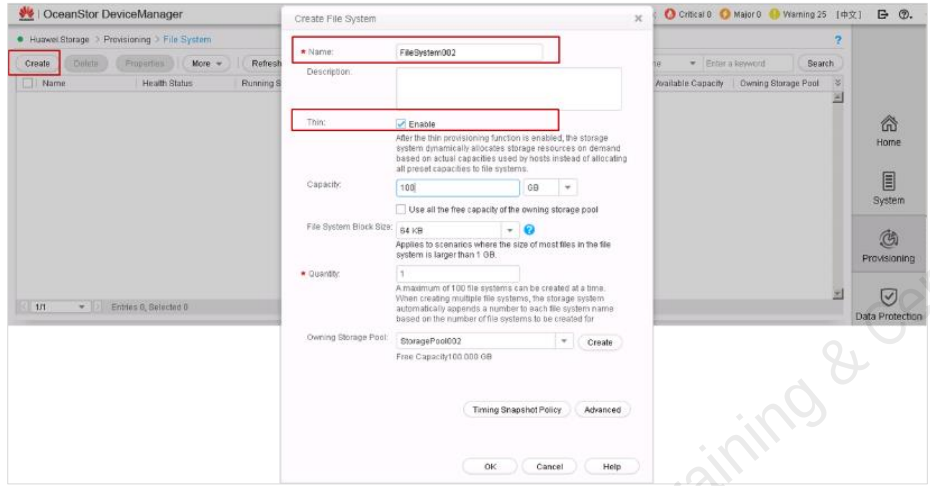
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configurations: Setting CIFS Sharing

- CIFS sharing allows files to be shared between Windows/MAC clients.

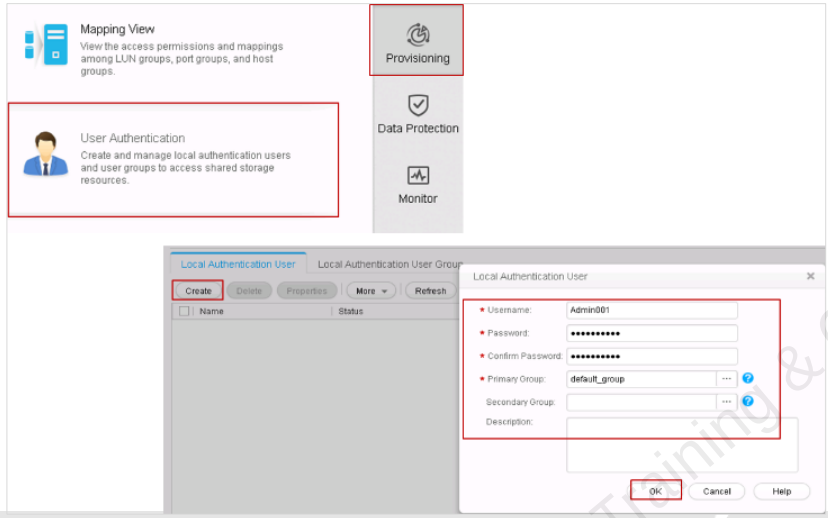


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Configuration: Creating a Local User

- A local user refers to the user who is created on the storage array.

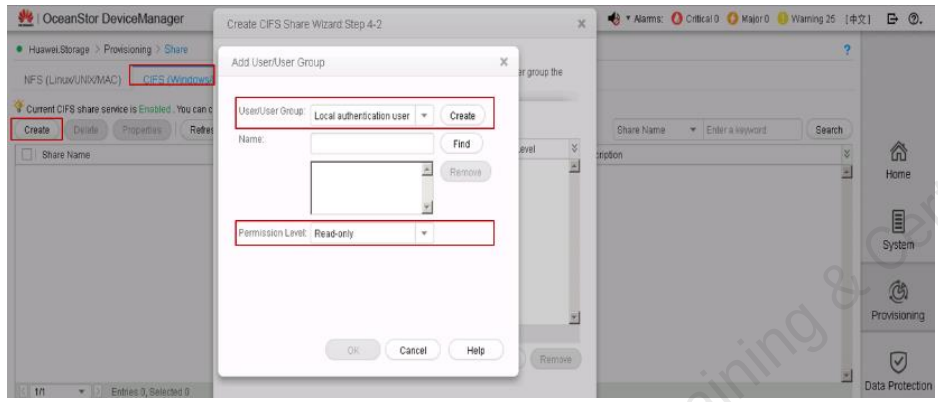


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configurations: Setting a CIFS Share Permission

- A CIFS sharing permission can be set based on the local user or domain user.

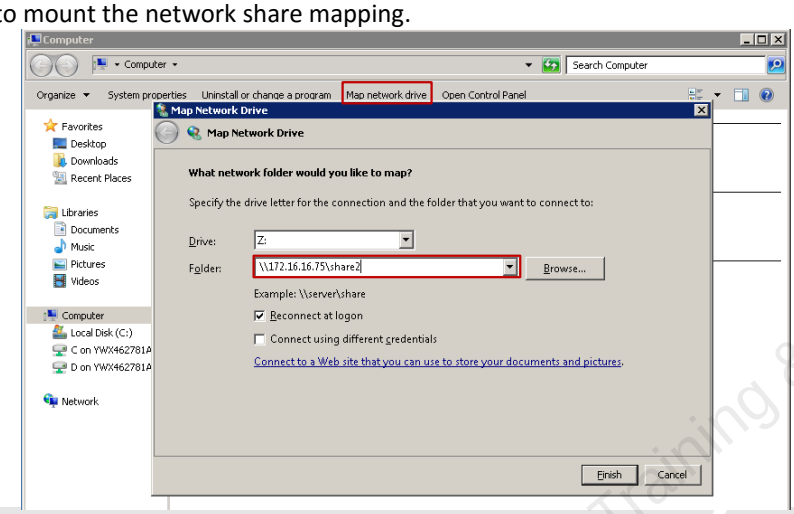


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configurations: Mounting a CIFS Share

- Log in to the domain Windows client and use the Windows resource manager to mount the network share mapping.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Quiz

1. Single Answer:
2. Which of the following is the advantage of the block virtualization architecture?
3. A. Faster reconstruction B. Improved disk reliability C. Tiered storage easily achieved D. Sounds awesome
4. Multiple-Choice Answer:
5. Which of the following are optional steps for storage space configuration?
6. A. Add a route. B. Configure CHAP authentication. C. Create a port group. D. Create a disk domain.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Summary

- SAN principles
- NAS principles
- RAID principles
- Basic features of converged storage
- Deployment and management of converged storage

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

More Information

- Huawei Enterprise Business Website
  - http://enterprise.huawei.com/cn/products/itapp/server/index.html
  - Product documents, compatibility lists, product software, and operation guides are available at the website.
- Huawei Service Hotline
  - Enterprise Service Support
  - http://support.huawei.com/enterprise/zh/index.html
  - Warranty service is available at the website.
- Huawei Server Information Self-Service Platform
  - http://enterprise.huawei.com/topic/Self_service_server/knowledge.html
  - Information about Huawei server installation, surveys, and O&M and the required tools are available at the website.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Recommendations

- Huawei Learning Website
  - http://support.huawei.com/learning/en/newindex.html
- Huawei Support Case Library
  - http://support.huawei.com/enterprise/servicecenter?lang=en

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

Huawei General Purpose Servers Deployment and O&M

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.



Huawei Training & Certification
Huawei Training & Certification

Foreword

- This course guides you through the deployment and O&M of Huawei general purpose servers. It helps you learn about the classification, basic configuration, and iBMC operations of Huawei general purpose servers.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Objectives

- On completion of this course, you will be able to command:
  - Classification of Huawei general purpose servers
  - Configuration and deployment of Huawei general purpose servers
  - Routine maintenance of Huawei general purpose servers
  - Use of Huawei server maintenance tools

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

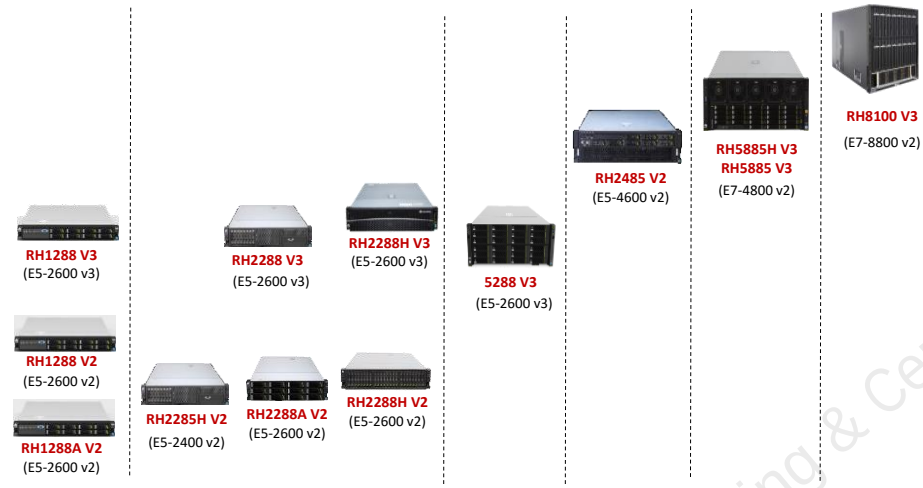
Contents

1. **Introduction to Huawei General Purpose Servers**
2. Server Configuration and Deployment
3. Routine Maintenance
4. Huawei Server Tools

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Huawei Rack Server Series

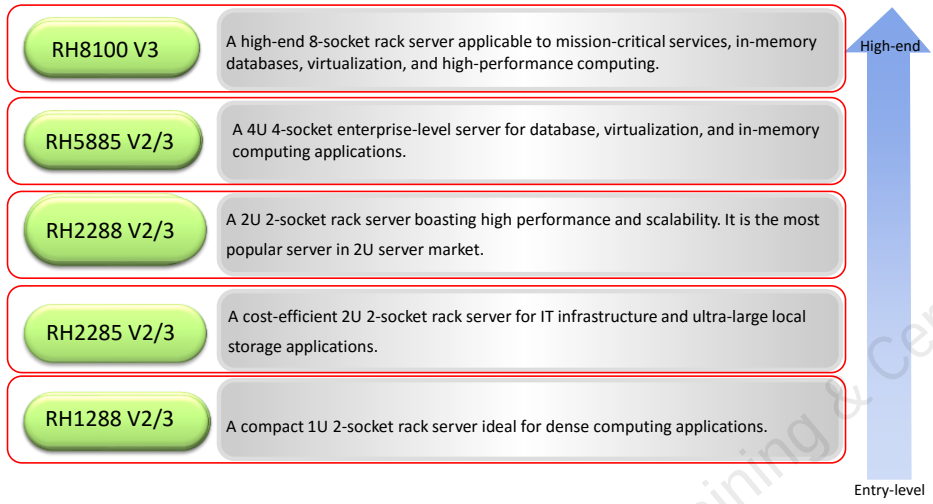


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Positioning of Huawei Rack Server



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. Introduction to Huawei General Purpose Servers
2. **Server Configuration and Deployment**
   - iBMC Configuration
     - BIOS settings
     - RAID Planning and Configuration
     - OS Installation
3. Routine Maintenance
4. Huawei Server Tools

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Logging in to the iBMC WebUI (Contd.)

- The client used to access the Intelligent Baseboard Management Controller (iBMC) WebUI must meet the following requirements:

OS	Browser	JRE
Windows 7 32-bit Windows 7 64-bit	Internet Explorer 8.0/10.0	JRE 1.6.0 U25 JRE 1.7.0 U40
	Mozilla Firefox 9.0/23.0	
	Chrome 13.0/31.0	
Windows 8 32-bit Windows 8 64-bit	Internet Explorer 10.0/11.0	JRE 1.6.0 U25 JRE 1.7.0 U40
	Mozilla Firefox 9.0/23.0	
	Chrome 13.0/31.0	
Windows Server 2008 R2 64-bit	Internet Explorer 8.0/10.0/11.0	JRE 1.6.0 U25 JRE 1.7.0 U40
	Mozilla Firefox 9.0/23.0	
	Chrome 13.0/31.0	

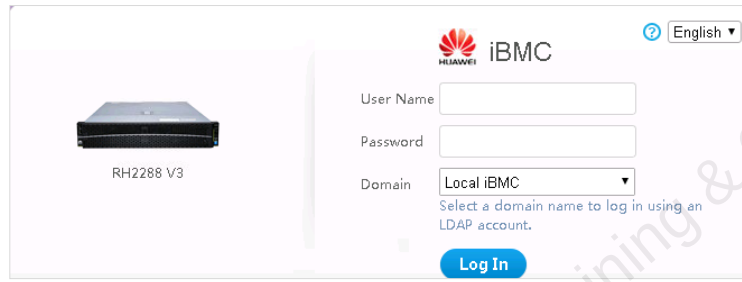
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Logging in to the iBMC WebUI (Contd.)

- Open the browser, enter https://iBMC IP address in the address box, and press Enter.
- In the security alert page displayed, click Continue to this website.
- In the login dialog box, enter the user name and password, and click Log In.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Login Procedure

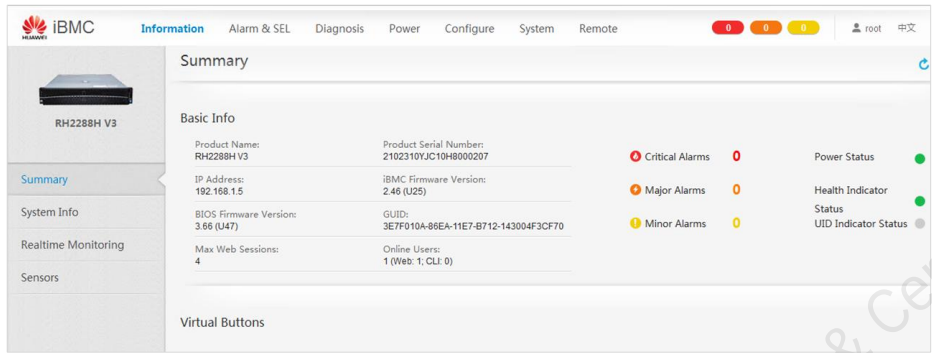
- Select the language to be used.
- Enter the user name. The default user name is root.
- Enter the password. The default password is Huawei12#$.
- Choose Local iBMC or a domain to log in.
- Click Log In.

- The BMC WebUI window is displayed. You can click Reset to set the parameters again.
- Default IP address of the iBMC management network port on rack V3 servers: 191.168.2.100
- Subnet mask: 255.255.255.0
- User name: Root
- Password: Huawei12#$

Huawei Training & Certification

Logging in to the iBMC WebUI

- iBMC WebUI

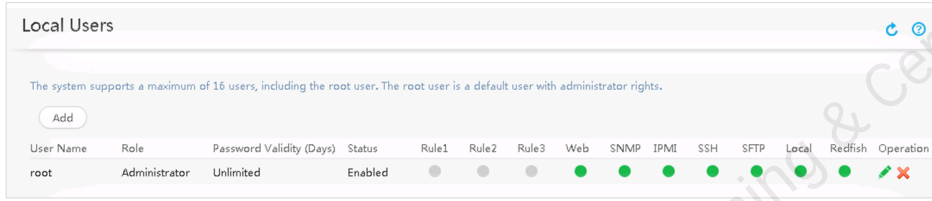


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configuration - User Configuration

- On the Configuration page, you can:
  - View and manage the users who have logged in to the iBMC.
  - View and configure security settings.
  - View and configure LDAP user information.



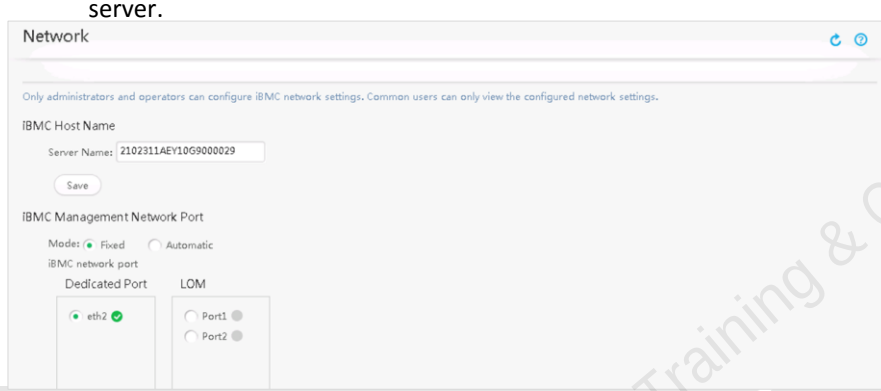
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The iBMC provides only the access for LDAP users. Using the user domain, group domain, LDAP user name, and password of the domain controller to log in to the iBMC system improves system security.
- LDAP users can log in to the iBMC only through the iBMC WebUI.

Huawei Training & Certification

Configuration - Network Configuration

- On the Network page, you can:
  - Set the server host name.
  - Set the mode and IP address of the management network port on the server.

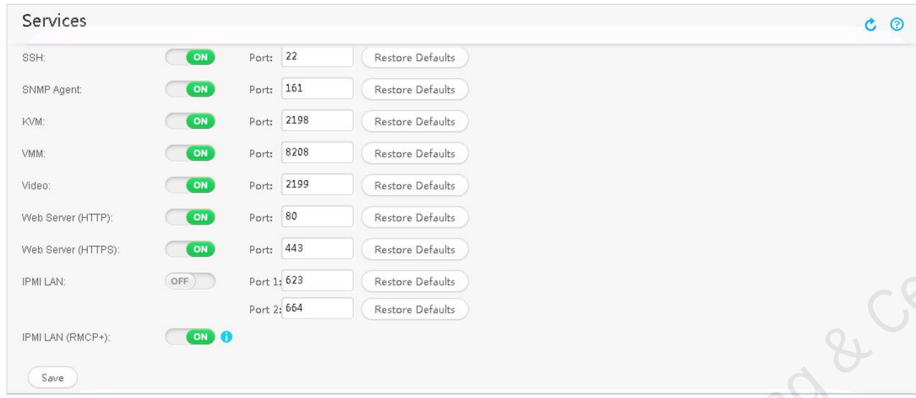


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configuration - Port Configuration

- On the Services page, you can view and set system service information.



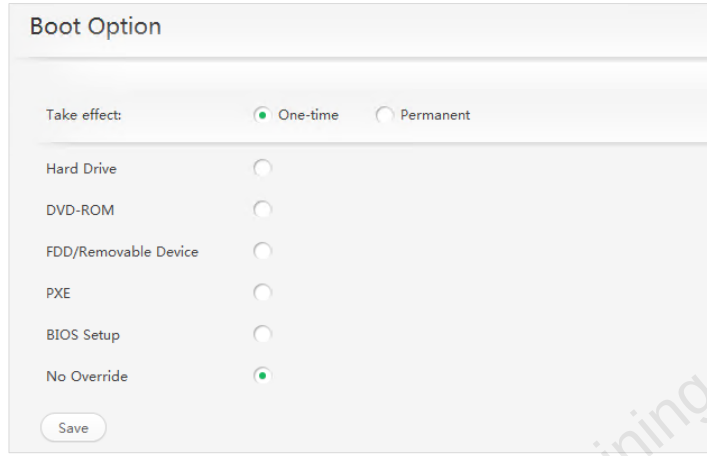
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configuration - System Startup

- On the Boot Option page, you can set the first boot device for the OS.



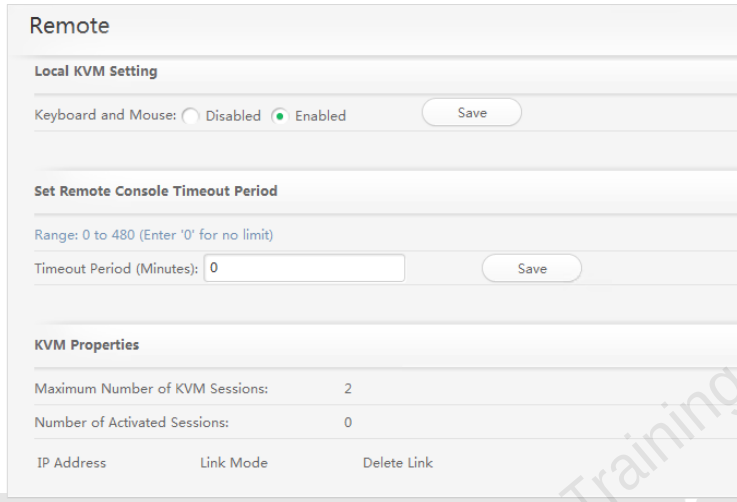
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Remote Control

- On the Remote Control page, you can view the maximum number of sessions and activated sessions of the KVM and virtual media. You can also use the remote virtual console to access the server OS and manage the server.

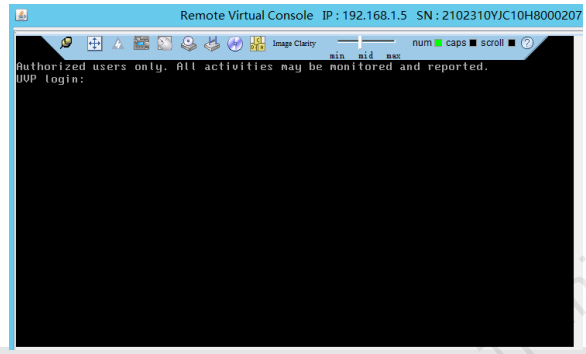


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Remote Virtual Console

- You can use the remote virtual console to remotely access and manage the server, install and repair the OS, and install device drivers.
- With the remote virtual console, you can:
  - Remotely access the server and manage the server on a real-time basis using the keyboard and mouse of the local PC.
  - Mount files and folder on the local PC to the server using the virtual floppy disk drive or CD-ROM drive.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- On the server side, the use of the virtual floppy disk drive or the CD-ROM drive is the same as the use of a USB device.

- The KVM window consists of three areas.
- 1: Toolbar: displays all the buttons you can use to manage the server remotely.

- 2: Real-time desktop. The desktop of the server is displayed on a real-time basis. You can use the mouse to perform operations or run commands on the desktop.

- 3: Status bar: displays the prompt information of the server desktop and the communication data between the server and the local PC.

Huawei Training & Certification

Contents

1. Introduction to Huawei General Purpose Servers
2. **Server Configuration and Deployment**
   - iBMC Configuration
   - BIOS Settings
   - RAID Planning and Configuration
   - OS Installation
3. Routine Maintenance
4. Huawei Server Tools

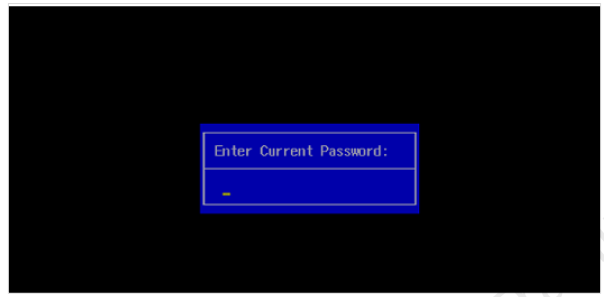
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Entering the BIOS

- Operation Procedure
  - Connect the multi-port cable, keyboard, mouse, and monitor to the server, or open the Remote Control page of the iBMC WebUI.
  - Power on the server. For details, see the user guide of the server you use.
  - When the information for pressing the shortcut keys is displayed, press DEL within 3 seconds until the BIOS Setup Utility is displayed.
  - If the Enter Current Password dialog box is displayed during the startup process, enter the password and press Enter.



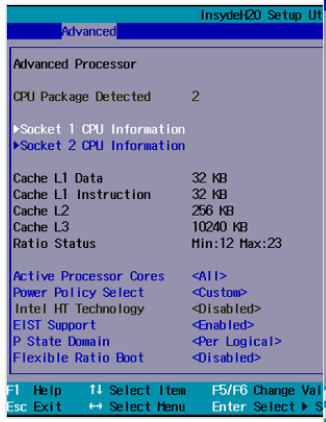
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

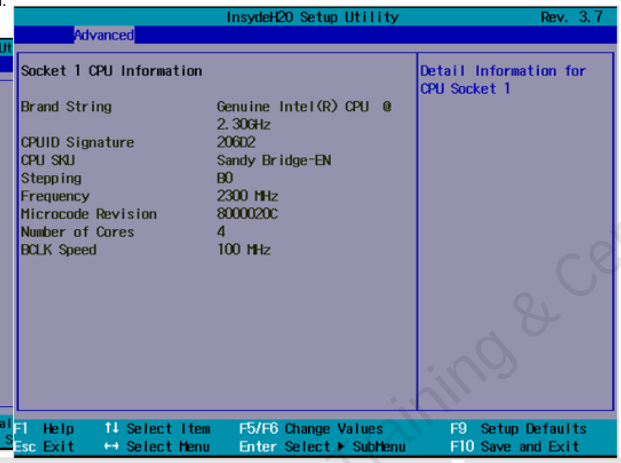
HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Querying CPU Information

- Choose Advanced > Advanced Processor and press Enter.
  - The Advanced Processor screen is displayed.
- Select Socket x CPU Information and press Enter.
  - The CPU Information is displayed.



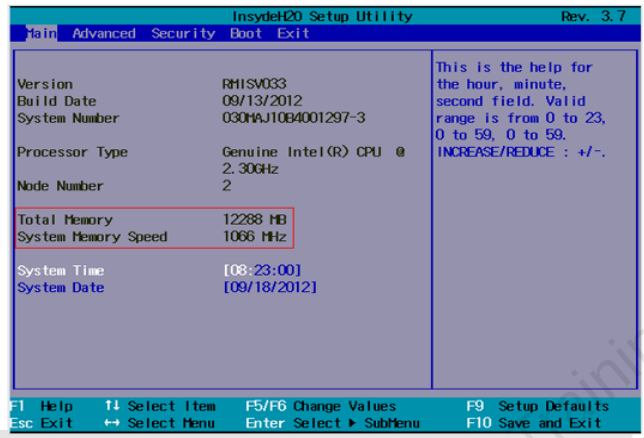


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Querying Memory Information

- Click the Main tab and press Enter.
- Select Advanced > Memory Configuration and press Enter.
  - The memory configuration information is displayed.

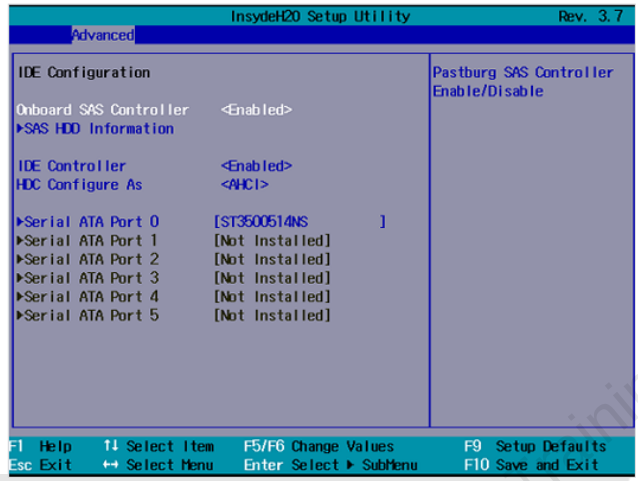


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Querying Hard Disk Information

- Choose Advanced > IDE Configuration and press Enter.
- The IDE Configuration screen is displayed.



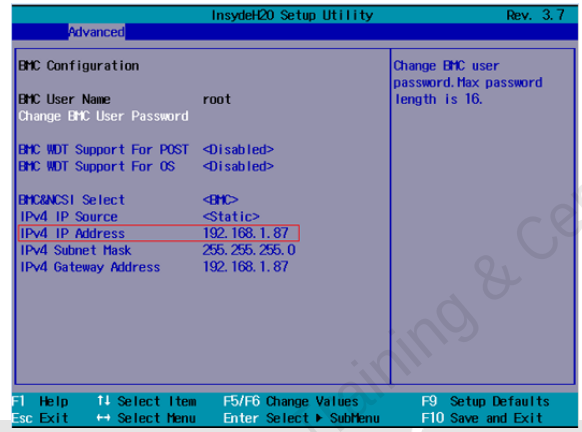
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Querying the BMC IP Address

- Choose Advanced > IPMI BMC Configuration.
- On the IPMI BMC Configuration screen, choose BMC Configuration.
  - The BMC Configuration screen is displayed.



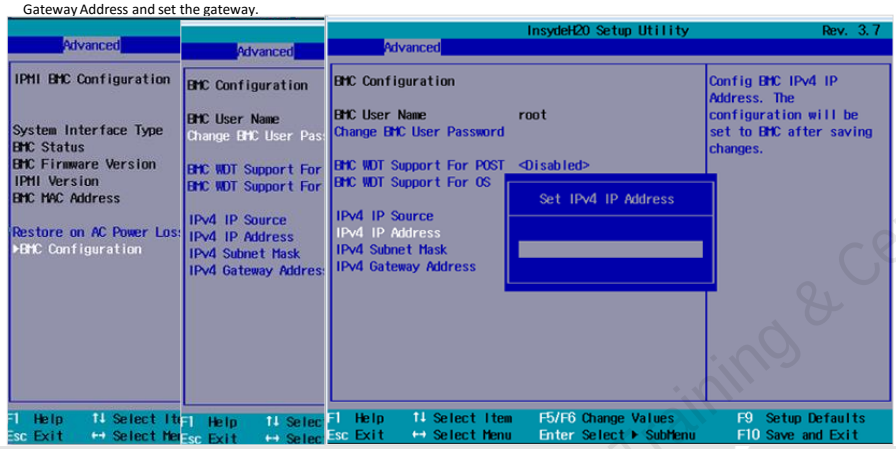
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Setting the BMC IP Address

- Choose Advanced > IPMI BMC Configuration.
- On the IPMI BMC Configuration screen, choose BMC Configuration.
  - The BMC Configuration screen is displayed.
- Select Static for IPv4 IP Source, select IPv4 IP Address and set the IP address, select IPv4 Subnet Mask and set the subnet mask, and select IPv4 Gateway Address and set the gateway.

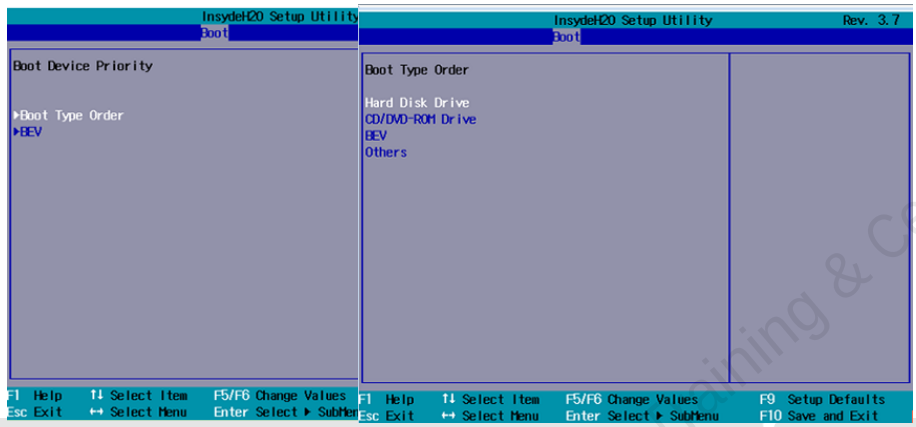


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Setting the Server Boot Mode

- Choose Boot > Legacy.
  - The Boot Device Priority screen is displayed.
- Select Boot Type Order.
- On the Boot Type Order screen, set the boot device priority as required.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. Introduction to Huawei General Purpose Servers
2. **Server Configuration and Deployment**
   - iBMC Configuration
   - BIOS Settings
   - RAID Planning and Configuration
   - OS Installation
3. Routine Maintenance
4. Huawei Server Tools

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Common RAID Levels and Application Scenarios

RAID Level	Application Scenarios
RAID 0	Applications demanding high read/write speed but considerably low security, such as graphics workstations.
RAID 1	Servers, databases, and storage, which demand random data write and high security.
RAID 3	Applications demanding data transmission and high security, such as video editing and large database storage.
RAID 5/6	Applications demanding random data transmission and high security, such as financial applications and database storage.
RAID 10	Applications demanding random read and write and high security, such as bank and financial applications.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configuring the LSI SAS3108 Card (Contd.)

- When "Press <Ctrl><R> to Run MegaRAID Configuration Utility" is displayed during the server startup process, press Ctrl+R.



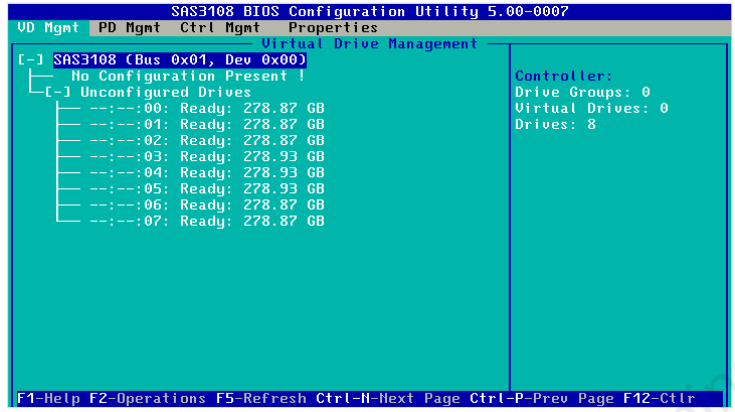
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configuring the LSI SAS3108 Card (Contd.)

- The SAS3108 BIOS Configuration Utility screen is displayed.

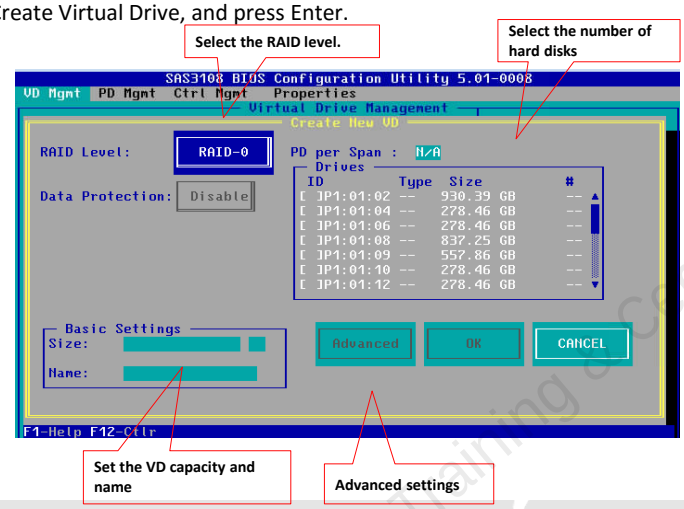


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configuring the LSI SAS3108 Card (Contd.)

- Use ↑ and ↓ to select SAS3108 (Bus 0x01, Dev 0x00).
  - Press F2, select Create Virtual Drive, and press Enter.

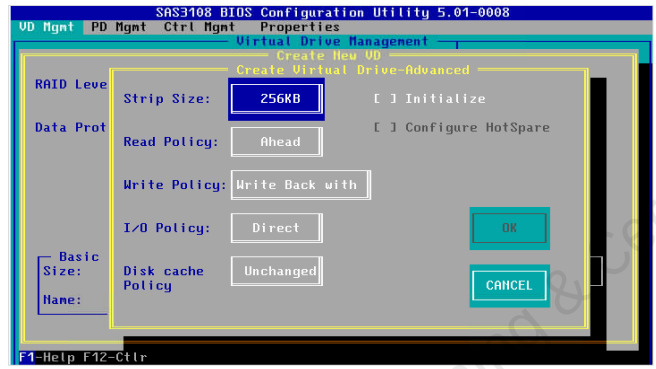


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configuring the LSI SAS3108 Card (Contd.)

- Set the following parameters:
- Strip Size
  - Set the strip size of the hard disk.
  - Read Policy
  - Default
  - Ahead
- Write Policy
  - Write back
  - Transparent writing
  - BBU write back
- I/O Policy
  - Direct
  - Cache

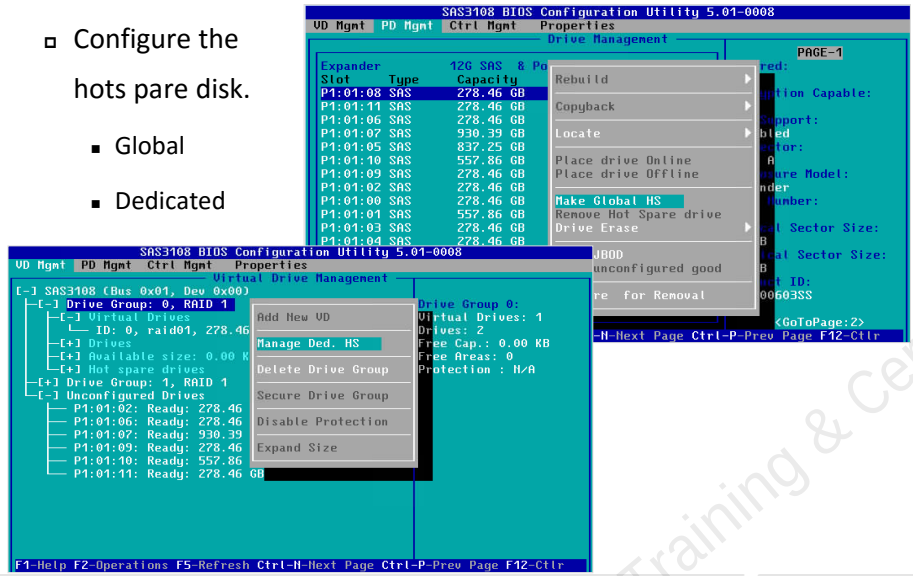


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configuring the LSI SAS3108 Card (Contd.)

- Configure the hots pare disk.
  - Global
  - Dedicated



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

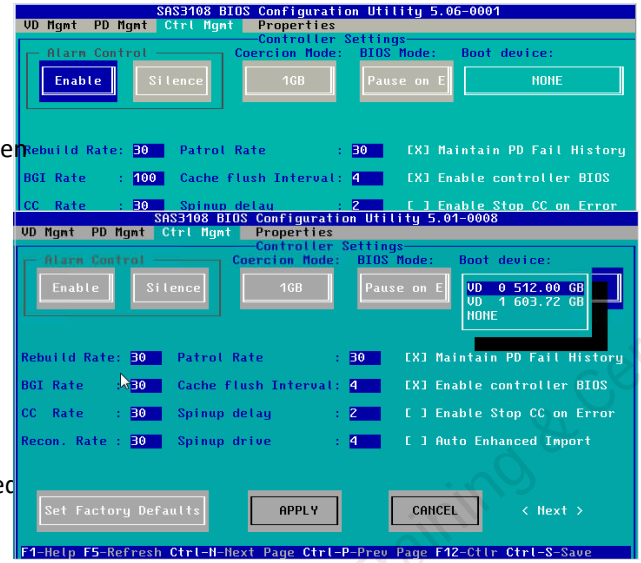
- Configure a global hot spare disk.
- Press Ctrl+N.
  - PD Mgmt. is displayed.
- Select the hard disk to be set.
- Press F2 and select Make Global HS.

- Configure a dedicated hot spare disk.
- On the VD Mgmt page, select the RAID to be configured.
- Press F2 and select Make Ded. HS.

Huawei Training & Certification

Configuring the LSI SAS3108 Card

- Set the boot device.
  - Press Ctrl+N.
  - The Ctrl Mgmt screen is displayed.
  - Select the boot device from Boot Device, and press Enter.
  - The selected boot virtual device is used to install the OS.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

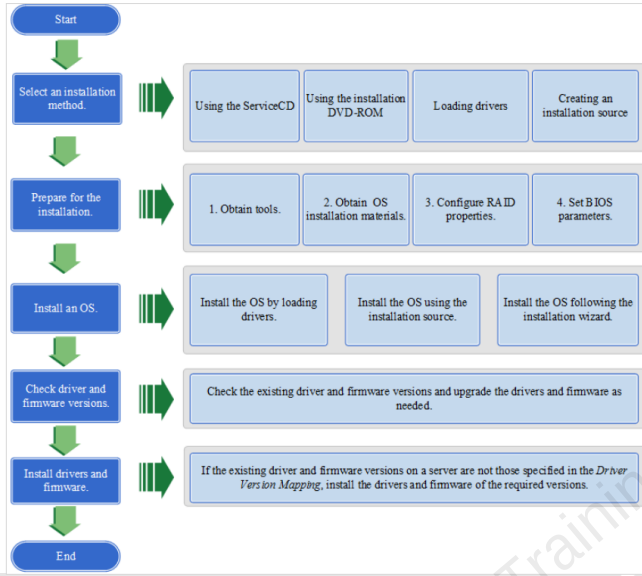
Contents

1. Introduction to Huawei General Purpose Servers
2. **Server Configuration and Deployment**
   - iBMC Configuration on Huawei Servers
   - BIOS Settings
   - RAID Planning and Configuration
   - OS Installation
3. Routine Maintenance
4. Huawei Server Tools

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

OS Installation Process

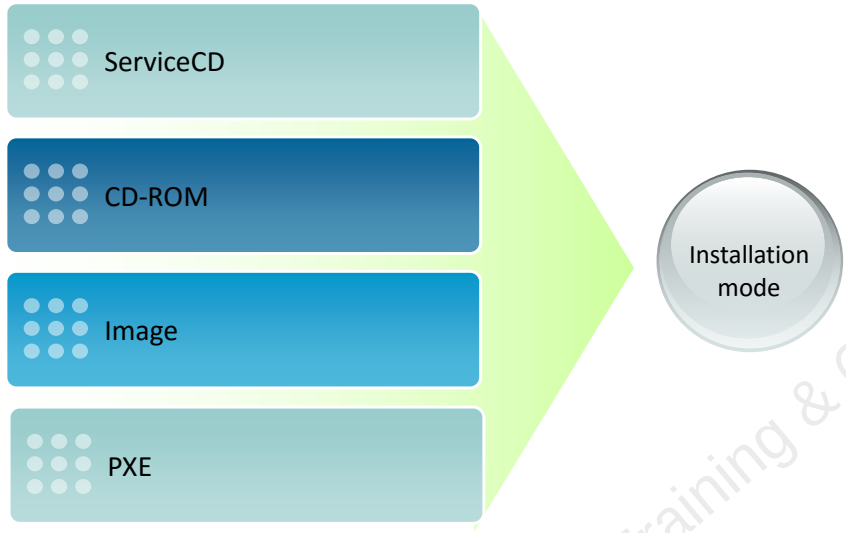


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

OS Installation Modes



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Installing OS Using ServiceCD

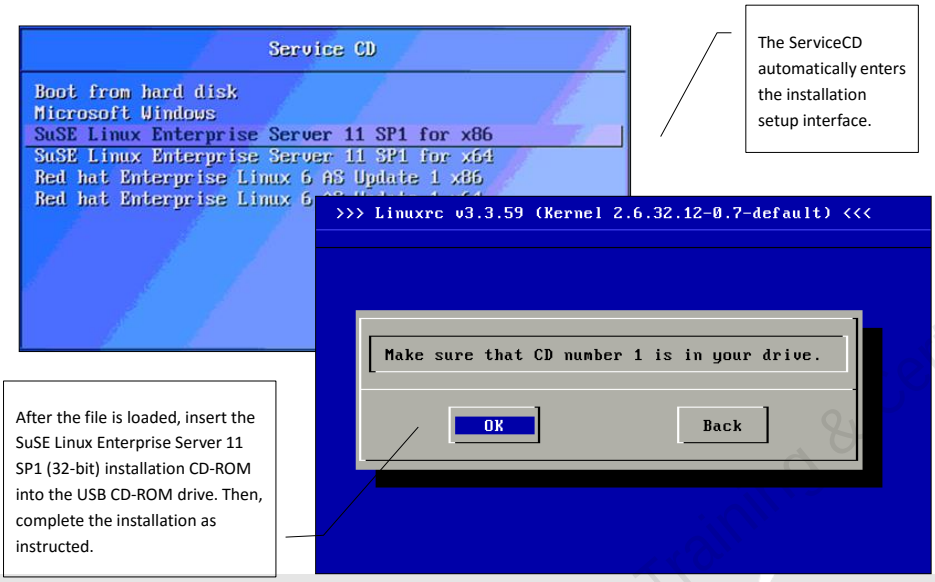
- ServiceCD function
  - The ServiceCD minimizes man-machine interaction during the installation process by automatically directing the installation of the OSs, drivers, and patches. This shortens the time for installing the OS.
- ServiceCD features
  - Helps users install mainstream OSs and provides the latest device drivers based on the hardware detected.
  - Minimizes the number of restart times during the hardware configuration and OS installation process.
  - Guides users to complete the installation through a simplified user interface.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Installation Procedure



The ServiceCD automatically enters the installation setup interface.

After the file is loaded, insert the SuSE Linux Enterprise Server 11 SP1 (32-bit) installation CD-ROM into the USB CD-ROM drive. Then, complete the installation as instructed.

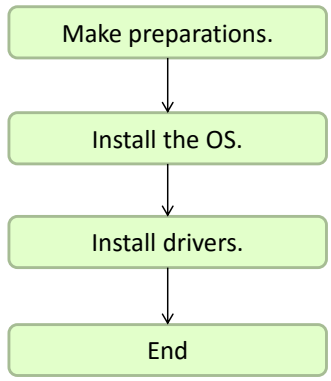
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification

Installing OS Using CD-ROM

- You need to install drivers, such as the RAID card and network adapter drivers, during the OS installation using the CD-ROM.



Get ready the OS installation CD-ROM, patch program, drivers, and USB flash drive.

Install the OS.

Install the drivers matching the OS.

Complete the installation.

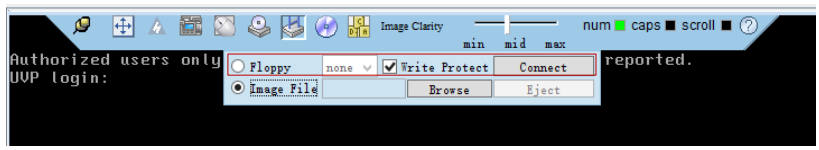
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Installing SLES Using the CD-ROM

- Enter the BIOS and set the first boot device to CD-ROM drive.
- When the startup window is displayed, press F6. In the displayed menu, select Yes to mount the virtual drive image file.
- On the Remote Control page of the iBMC WebUI, use the virtual media function to mount the RAID card drive image file.



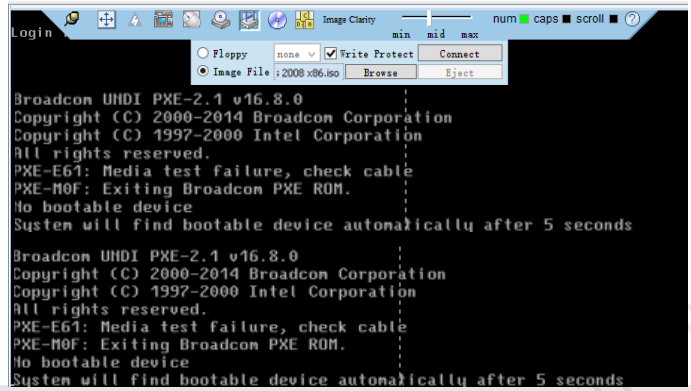
- On the driver loading interface, select the corresponding virtual drive or virtual floppy drive, select OK, and press Enter to read the driver file.
  - After the driver is installed correctly, the current disk model is displayed.
- Select Back, and press Enter. Then, reload the OS CD-ROM to install the OS.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Installing OS Using Image

- Image
  - An image is made of a series of files in a certain format. It can be used only after being decompressed by using the virtual CD-ROM drive or virtual floppy drive tool.
  - The virtual console provided by the iBMC WebUI allows you to mount an image and install the OS on the OS.
  - The virtual console provides the virtual CD-ROM drive and virtual floppy drive.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Installing OS Using Image - Preparations

- Get ready a PC or laptop running Windows for accessing the server.
- Configure network data to enable communication between the server and the PC or laptop.
- Configure RAID for the server.
- Obtain the OS image.
- Download the driver image.

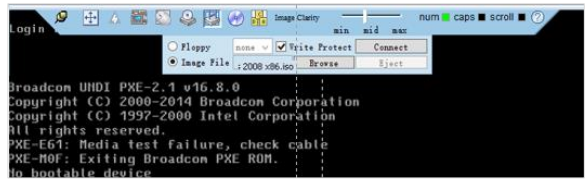
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

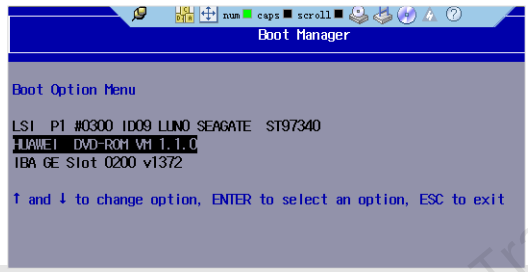
Huawei Training & Certification
Huawei Training & Certification

Installing SLES (Contd.)

- On the virtual remote console interface, select Image File, click Browse to select the ISO file, and click Connect.



- Restart the server.
- During the server startup process, press F11 and select HUAWEI DVD-ROM VM 1.1.0.

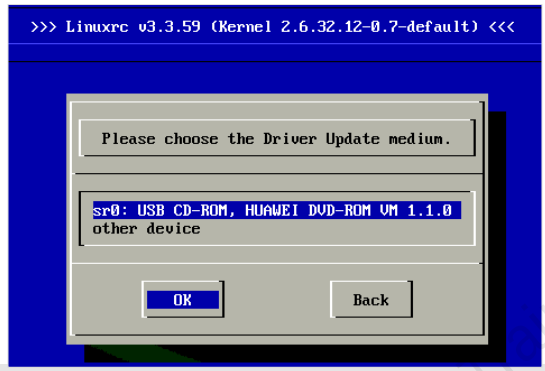


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Installing SLES (Contd.)

- When the startup window is displayed, press F6.
- In the displayed menu, select Yes to mount the virtual drive image of the RAID card driver.
- On the screen displayed, select OK and press Enter.

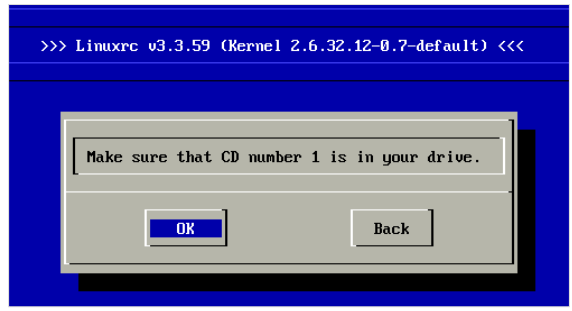


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Installing SLES (Contd.)

- After the controller driver is automatically installed, the 2308 controller image file is displayed. The system loads the system installation image file.
- Click Back.
- On the screen displayed as follows, click OK and press Enter, and install the OS as instructed.



- After the OS is installed, install the drivers, such as the NIC driver, in the same way.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Installing the OS Using PXE (Contd.)

- Preboot eXecute Environment (PXE)
  - PXE is a remote computer boot technology that enables a computer to start from the network.
  - It is a client-server environment. The PXE client obtains an IP address from the DHCP server and downloads the remote OS using FTFP to the local computer.
  - To install the OS using PXE, you only need to install the OS on one or more servers and set the other servers to boot from PXE.
  - Installation procedure:
  - Connect one end of the network cable to the NIC1 port and the other end to the server to ensure communication between them.



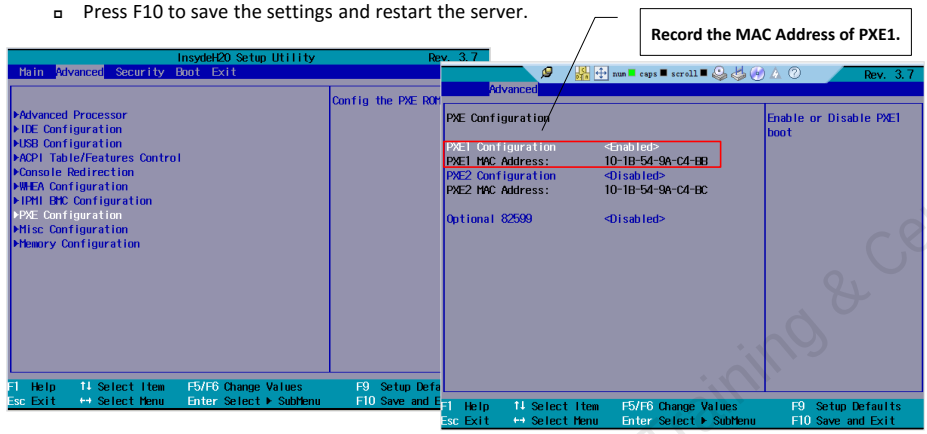
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Installing the OS Using PXE (Contd.)

- On the BIOS interface, click the Advanced tab and select PXE Configuration.
- On the PXE Configuration screen, set PXE1 Configuration to Enabled, and set other parameters to Disabled.
- Press F10 to save the settings and restart the server.

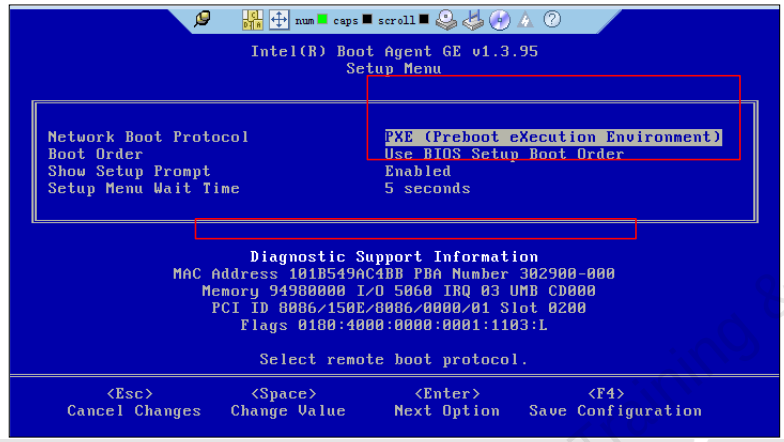


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

PXE Configuration

- During the server startup process, press Delete to access the BIOS and configure the NIC.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification

Contents

1. Introduction to Huawei General Purpose Servers
2. Server Configuration and Deployment
3. **Routine Maintenance**
4. Huawei Server Tools

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Fault Information Collection

- When a server is faulty, collect the following information for fault diagnosis:
  - Basic fault information, including basic information about the customer, server model and configuration, and fault symptom.
  - Server hardware log for identifying server system fault.
    - Collect server hardware information through the iBMC.
  - Service layer logs, including the logs related to the OS and service software.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Collecting Fault Information Using the iBMC (Contd.)

- Collect the following information on the iBMC WebUI:
  - iBMC and BIOS version information
  - Hardware (CPU, memory, and RAID card) configuration information (model and quantity)
  - Server health status, including the CPU temperature, fan speed, power consumption, and hard disk status.
  - Server historical logs (used to analyze the server status before and after the fault occurrence time)
  - Last screen (information displayed on the screen before the server is down)

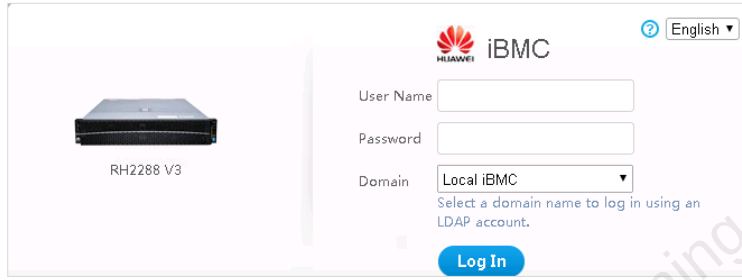
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Collecting Fault Information Using the iBMC (Contd.)

- Open the browser, enter https://iBMC IP address in the address box, and press Enter.
- In the security alert page displayed, click Continue to this website.
- In the login dialog box, enter the user name and password, and click Log In.



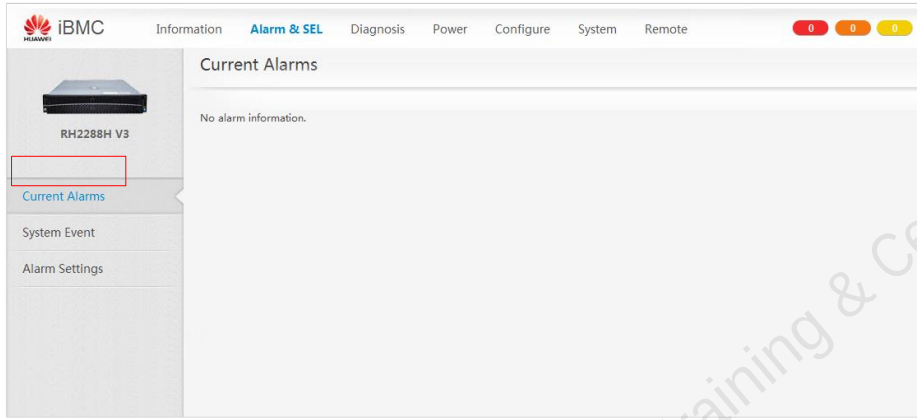
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Collecting Fault Information Using the iBMC (Contd.)

- Choose Alarm & SEL from the main menu, and select Current Alarms or System Events.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

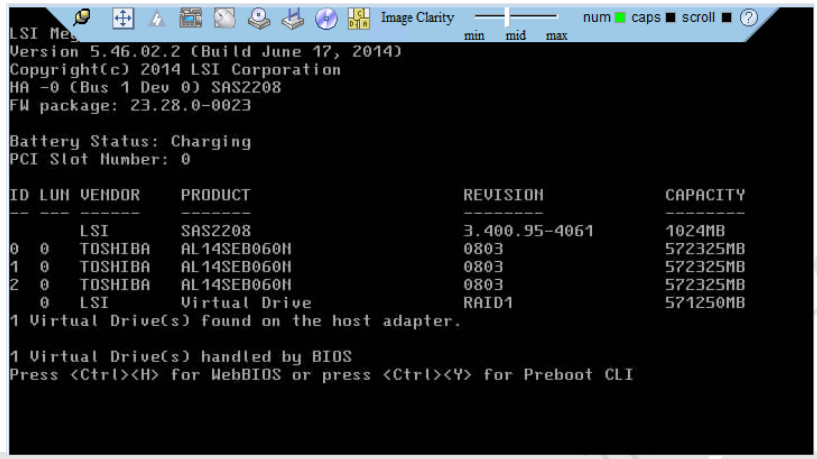
HUAWEI

- Red: Indicates a critical alarm.
  - A critical alarm may power off the server, and even interrupt system services. You must take corrective action immediately.
- Orange: Indicates a major alarm.
  - A major alarm has a major impact on the system. It affects the normal operating of the system or may cause service interruption.
- Yellow: Indicates a minor alarm.
  - A minor alarm has a minor impact on the system, but you need to take corrective action as soon as possible to prevent a more severe alarm.

Huawei Training & Certification
Huawei Training & Certification

Collecting Fault Information Using the iBMC (Contd.)

- Choose Remote Console from the main menu, and click the remote virtual console.



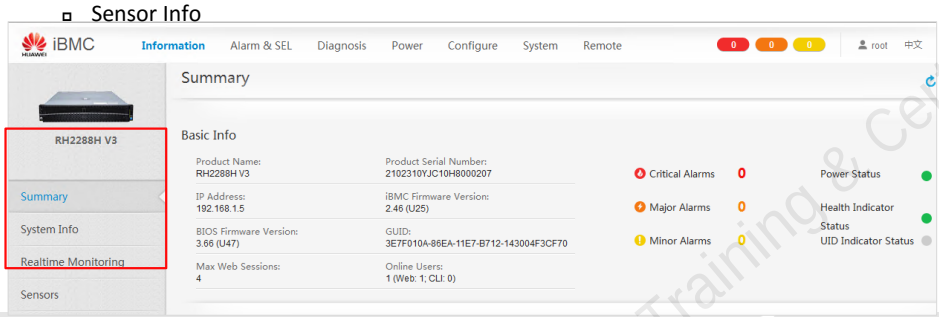
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification

Collecting Fault Information Using the iBMC (Contd.)

- Choose Information from the main menu, and select the following from the navigation tree to obtain server information:
  - Overview
  - System Info
  - Real-Time Monitoring
  - Sensor Info



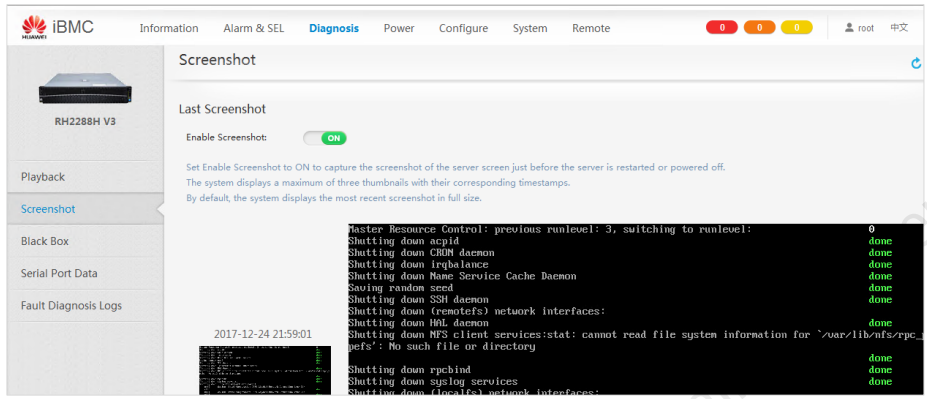
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Collecting Fault Information Using the iBMC (Contd.)

- Choose Diagnostics from the main menu, and select Screenshot.
- You can see the last screen under Capture Screenshot area.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification

Contents

1. Introduction to Huawei General Purpose Servers
2. Server Configuration and Deployment
3. Routine Maintenance
4. **Huawei Server Tools**

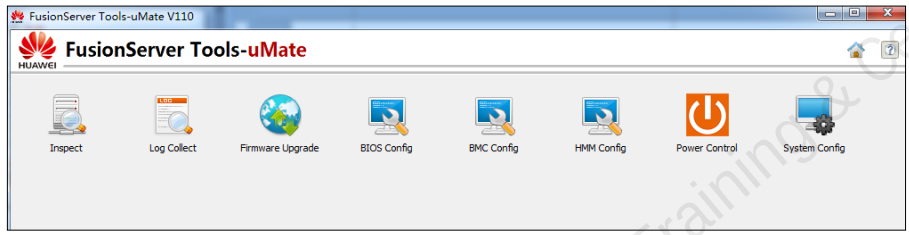
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

uMate

- uMate is a tool used for routine maintenance of Huawei servers. uMate provides the following functions:
  - Batch server inspection
  - BMC information collection
  - Batch firmware upgrade
  - BIOS configuration, BMC configuration, MM configuration
  - Power control
  - System configuration
- For details, see the uMate User Guide.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Page57

- Batch server inspection: uMate can automatically discover multiple servers and obtain server health status.
- Log collection: Users can use uMate to collect log information, which helps locate faults.
- Batch firmware upgrade: uMate implements firmware upgrade of servers in batches, which improves O&M efficiency.
- Power control: Users can use uMate to perform power control on multiple servers at a time.

Huawei Training & Certification

Before Using uMate

- Console requirements
  - Use Windows or SUSE11 sp1 or later if you want to use uMate GUI.
  - User Windows or Linux if you want to use uMate CLI.
- Prerequisites
  - Obtain the BMC IP address, user name, and password of the rack server or IP address, user name, and password of the active management module of the blade server.
  - Before performing batch operations, obtain the IP network segment and user root password of each device.
  - The servers and uMate are in the same LAN.

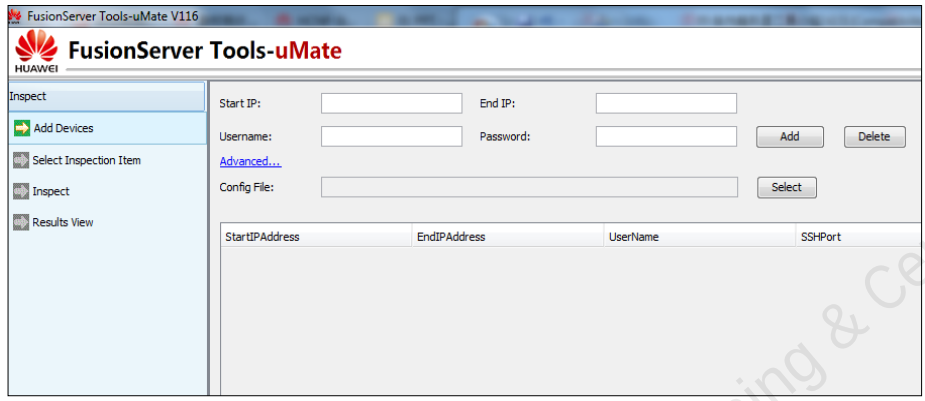
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Using uMate to Inspect Servers (Contd.)

- In the uMate directory, double-click uMate.exe and choose Inspect.

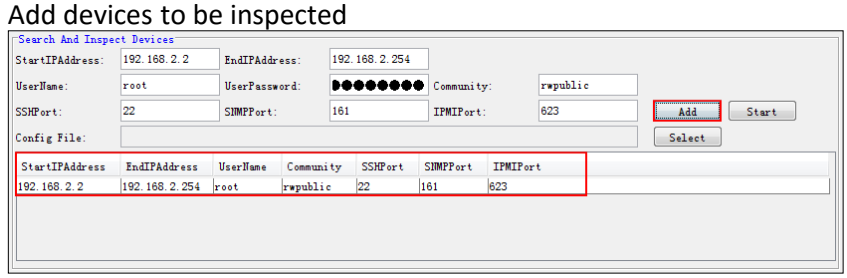


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

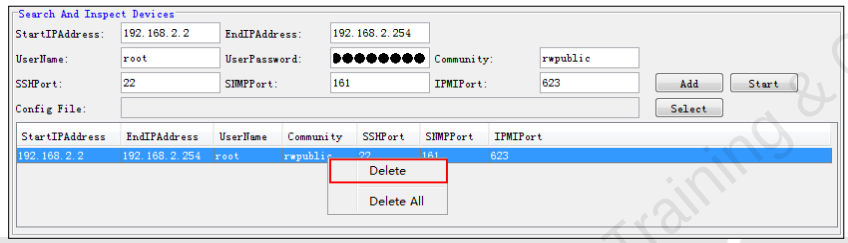
HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Using uMate to Inspect Servers (Contd.)

- Add devices to be inspected



- Delete devices



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Using uMate to Inspect Servers (Contd.)

- Set parameters

Parameter	Description	Remarks
StartIPAddress	IP address of the start server.	
EndIPAddress	IP address of the end server.	
UserNmae	BMC user name.	The default value is root.
UserPassword	BMC user password.	
Community	Read/Write community name.	Set this parameter when SNMPv1 or v2c is used.
Use the default settings for the following parameters.		
IPMIPort	IPMI port number.	The default value is 623.
SNMPPort	SNMP port number.	The default value is 161.
SSHPort	SSH port number.	The default value is 22.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

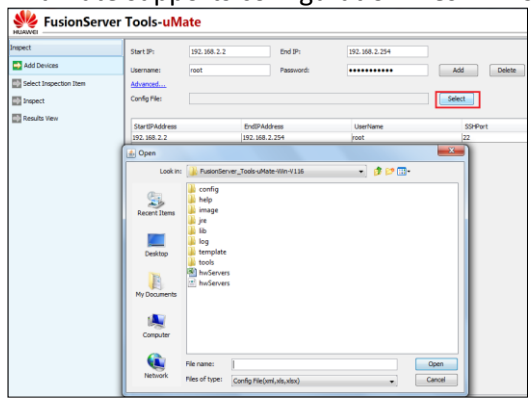
Page61

- You can query the IPMI port, SNMP port, and SSH port information through the iBMC.

Huawei Training & Certification
Huawei Training & Certification

Using uMate to Inspect Servers (Contd.)

- If the BMC passwords of the servers are different, use the configuration file.
- uMate supports configuration files in .xls or .xml format.



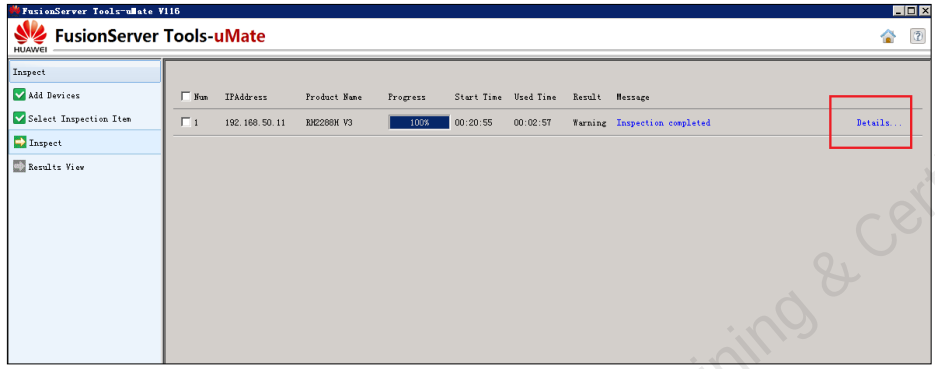
StartIPAddress	EndIPAddress	UserName	UserPassword	Com
192.168.2.98	192.168.2.98	root	Huawei12#$	rvpu
191.191.2.80	191.191.2.80	root	huaweiosta	rvpu
191.191.2.59	191.191.2.59	root	Huawei12#$	rvpu
191.191.2.65	191.191.2.65	root	Huawei12#$	rvpu
191.191.2.76	191.191.2.76	root	Huawei12#$	rvpu
191.191.2.53	191.191.2.53	root	root	rvpu
191.191.2.60	191.191.2.60	root	root	rvpu
191.191.2.76	191.191.2.76	root	Huawei12#$	rvpu
192.168.2.98	192.168.2.98	root	Huawei12#$	rvpu
191.191.2.179	191.191.2.179	root	Huawei12#$	rvpu
192.168.44.100	192.168.44.100	root	Huawei12#$	rvpu
192.168.44.117	192.168.44.117	root	Huawei12#$	rvpu
192.168.2.200	192.168.2.200	root	Huawei12#$	rvpu
192.168.2.106	192.168.2.106	root	root	rvpu
192.168.2.244	192.168.2.244	root	Huawei12#$	rvpu
192.168.1.94	192.168.1.94	root	Huawei12#$	rvpu
192.168.2.110	192.168.2.110	root	Huawei12#$	rvpu

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Using uMate to Inspect Servers (Contd.)

- It takes about 3 to 5 minutes to collect logs of a server.
- After log collection is complete, click Details to view detailed information about a server.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Using uMate to Inspect Servers (Contd.)

- After log collection is complete, the log collection report is displayed.

1.1 Device Health Report:

IP Address	Product Name	Result	Qty of Items	Qty of Error	Qty of Warning	Qty of Success
192.168.50.11	RH2288H V3	Warning	12	0	1	11

User Name	Start Time	Use Time	Error Message

4. Inspection Details:

Item	Result	Information
Shelf Check		
Power Status	OK	Power On
System Health	OK	
CPU Health	OK	CPU1: OK CPU2: OK
Memory Health	OK	DIMM000: OK DIMM001: OK DIMM010: OK DIMM011: OK DIMM020: OK DIMM021: OK DIMM030: OK DIMM031: OK DIMM100: OK DIMM101: OK

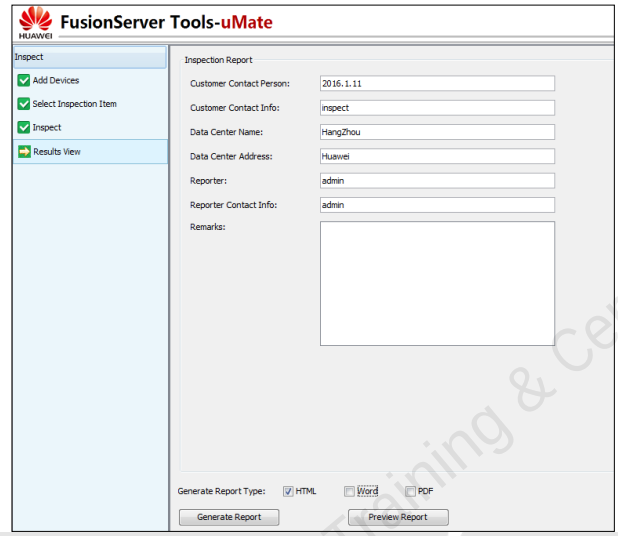
The health status of components, such as CPUs and memory, is displayed.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Using uMate to Inspect Servers

- After the inspection is complete, select the format of the report to be generated and generate a report.
- In the report, if the result contains Major or Minor, collect logs of the server for further analysis.

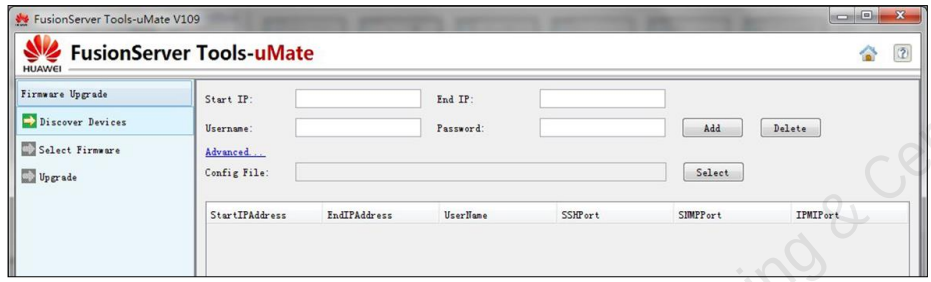


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Using uMate to Upgrade Firmware (Contd.)

- Before the upgrade, check whether the device to be upgraded is in the supported device list.
- In the main window of the uMate, click Upgrade Management in the left and click Servers Upgrade.



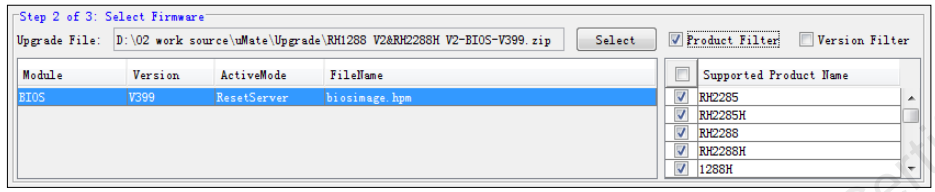
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- SMM upgrade is not supported.

Huawei Training & Certification
Huawei Training & Certification

Using uMate to Upgrade Firmware (Contd.)

- The upgrade file is in .zip format.
- After the upgrade file is selected, the tool parses the version.xml file and displays the file content and devices to be upgraded.

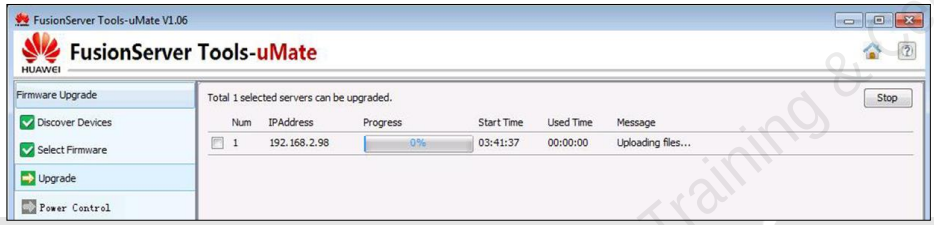


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Page67
Huawei Training & Certification
Huawei Training & Certification

Using uMate to Upgrade Firmware (Contd.)

- Select the Product Filter check box above the product name. The system filters the device discovery list based on the selected product name.
- If you select Version Filter above the product name, the servers of the same version will be filtered out. Otherwise, the servers of the same version are displayed.
- Select the servers to be upgraded and click Upgrade. If the BIOS is upgraded, the original configuration will not be retained.

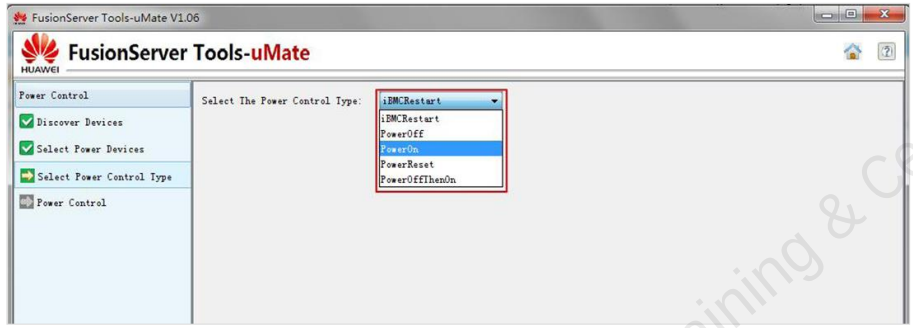


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Using uMate to Upgrade Firmware

- Perform power control on multiple servers, for example, resetting the iBMC, powering on/off, powering cycle, and resetting servers.
- The power control function is integrated into the Servers Upgrade function.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification

Quiz

1. True or False
   A. The iBMC supports both GUI and CLI (T or F).
   B. All rack V3 servers support memory hot swap (T or F).
2. Multiple Choices
   A. Which of the following methods can be used to set the management network port IP address of the server iBMC? ( )
   B. A. iBMC Web UI B. iBMC CLI C. BIOS D. LCD

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- True or False
- 1.T; 2.F

- Multiple Choices
- 1.ABC

Huawei Training & Certification
Huawei Training & Certification

Summary

- Introduction to Huawei General Purpose Servers
- Server Configuration and Deployment
- Routine Maintenance
- Huawei Server Tools

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

More Information

- Huawei enterprise technical support
  - http://support.huawei.com/enterprise/en/index.html
  - You can obtain product documentation, compatibility information, product software, and maintenance services.
- Huawei server information self-service platform
  - http://enterprise.huawei.com/topic/Self_service_server_en/knowledge.html
  - You can obtain Huawei server installation, commissioning, and O&M information and tools.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Recommendations

- Huawei learning website
  - http://support.huawei.com/learning/Index!toTrainIndex
- Huawei support case library
  - http://support.huawei.com/enterprise/servicecenter?lang=en

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

Huawei KunLun Server Application Scenarios and Deployment

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.



Huawei Training & Certification
Huawei Training & Certification

Foreword

- This section describes the application scenarios and deployment of Huawei KunLun servers. On completion of this course, you will be able to understand the application scenarios, hardware architecture, and typical deployment of Huawei KunLun servers.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Objectives

- On completion of this course, you will be able to understand:
  - Positioning and features of KunLun mission-critical servers
  - Typical application scenarios of KunLun mission-critical servers
  - Hardware architecture of KunLun mission-critical servers
  - How to perform initial configuration on KunLun series servers

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. **Positioning and Application of Huawei KunLun Servers**
   - Product Positioning and Benefits
   - Typical Application Scenarios
2. Hardware Architecture and Features of Huawei KunLun Servers
3. Initial Configuration of Huawei KunLun Servers

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Enterprise Transformation Requires an Open Platform for Critical Business

Open platform: outstanding performance and reliability



"Current Linux applications require higher performance and reliability than what are provided by a standard and open platform.""

Better cost-benefits: limited IT investment budget, lower TCO expected



"With tight IT budgets, the high OPEX of UNIX servers takes up much of the investment that should have been used for service innovation."

Unlock technology potential: compatibility with new applications



"The underlying enclosed architecture of UNIX servers is not compatible with emerging applications such as cloud computing and Big Data, preventing the new technologies from coming into full play."

Mitigate supply chain risks: open standards to prevent technology binding



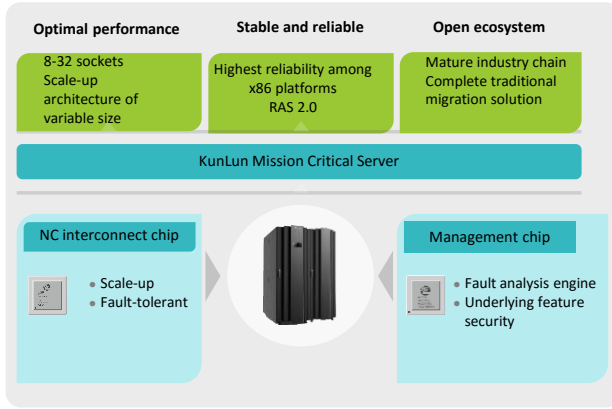
"Susceptible to vendor lock-in and supply risks due to over reliance on vendors for core technologies."

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

World's First 32-Socket Mission Critical Server



- KunLun products use Huawei proprietary BIOS developed based on American Megatrends (AMI)'s code base. The BIOS is customizable and scalable and supports diverse in-band and out-of-band configurations.

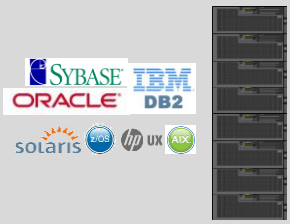
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Replacing Traditional UNIX Servers to Reduce TCO

Traditional RISC/EPIC Server



- High TCO
- Complex O&M
- Lack of open source support; slow in innovation

Transformation

KunLun Server for Mission-Critical Services



- 30% less TCO
- Standard and easy to maintain
- Support by lots of commercial and open-source software; fast in innovation

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

KunLun Offers Much More Computing Resources Per Node Than Common x86 Servers

SAP HANA 12 TB/24 TB Per Node



- Customer requirements: 12 TB or more per HANA node
- Typical customer: CEPSA

HPC Fat Node



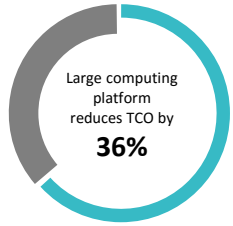
- Definition: An HPC fat node is a single computer system that uses many processors.
- Customer requirements: higher computing density and larger memory capacity per system
- Typical customers: meteorological institute, research center

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

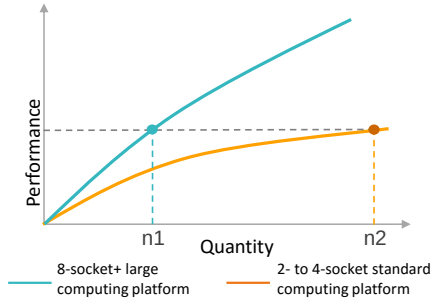
Huawei Training & Certification
Huawei Training & Certification

Effectively Saving TCO



Comparison of IT expenditures for 100 users in OLTP service scenarios

Expenditure Type	2-4P	8P+	Expenditure Saved
Server O&M	38924	20015	48.6%↓
Server and network	23541	14797	37.2%↓
Software License	25647	20249	21%↓
DC infrastructure	14307	11818	17%↓



IDC Analyze the Future

Single nodes are capable of offering more and more computing resources, which help drive down the O&M costs for clusters. More and more customers are considering using platforms with more computing resources to carry their mission-critical workloads.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

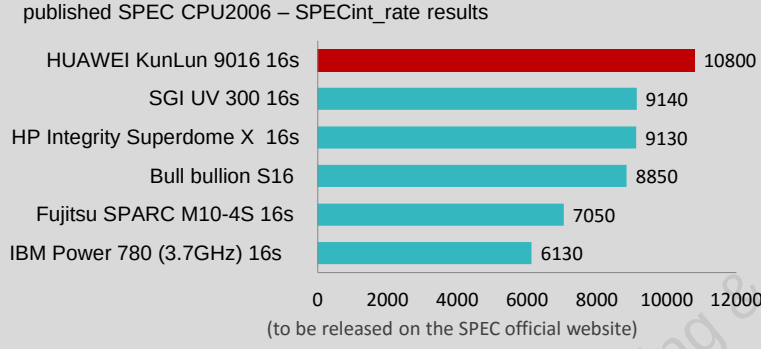
HUAWEI

Huawei Training & Certification
Huawei Training & Certification

KunLun Product Highlights (1)

Higher Performance Than UNIX Servers

Huawei-developed Scale-Up Architecture



(to be released on the SPEC official website)

- Scale-up coefficient 1.97x, most efficient in the industry
- Industry-leading SPECint_rate performance

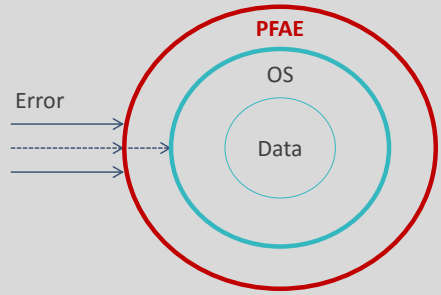
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

KunLun Product Highlights (2)

Ultimate Reliability, Comparable to High-End x86 Servers

RAS 2.0: Proactive Fault Management



- PFAE, maximizing service uptime
- Physical and logical partitioning, improving resource utilization
- Pioneering hot-swappable CPUs and memory modules

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

KunLun Product Highlights (3)

Open ecosystem: 30% lower TCO

Compatibility with Mainstream x86 Ecosystem



- World-leading IT compatibility verification and solution incubation center
- Fully compatible with mainstream OSs, databases, and virtualization environments
- Ecosystem-based cooperation to build joint solutions

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Contents

1. **Positioning and Application of Huawei KunLun Servers**
   - Product Positioning and Benefits
   - Typical Application Scenarios
2. Hardware Architecture and Features of Huawei KunLun Servers
3. Initial Configuration of Huawei KunLun Servers

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Three Typical Application Scenarios of KunLun: Database

Database

Single DB/Consolidation/Middleware + DB hybrid deployment

- CPU high-speed interconnect for faster transaction responses
- Meet the requirements of OLTP for low delay and high concurrency
- Lower TCO than scale-out

- UNIX servers
  - Closed architecture, high TCO
  - Complex O&M
  - Lack of open source support; slow in innovation
- KunLun
  - High performance, scalability, and reliability; compatible with mainstream open ecosystems; ideal replacement for UNIX servers
  - Accelerating innovation in applications such as Big Data and cloud computing
  - Easy maintenance, high security and controllability

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Three Typical Application Scenarios of KunLun: In-Memory Computing

In-memory computing

SAP HANA 12 TB/24 TB single-node

- Industry's largest-scale in-memory computing resources
- SAP HANA 12 TB or above scenarios

- UNIX servers
- No support for SAP HANA in-memory databases
- In the Big Data and cloud computing era, open source application innovations are supported by open platforms but not by closed architecture platforms.
- KunLun
- 576 cores and 768 DIMMs, industry-leading performance and largest scale of computing resources per node
- Best RAS features among open platforms
- Open platform that supports for various analysis and mining application innovations

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Three Typical Application Scenarios of KunLun: HPC Fat Nodes

HPC fat node

- High-bandwidth, low-delay intra-board communication
- Open ecosystem, diversified applications

- UNIX servers or 4-/8-socket x86 fat nodes
- Support for MPI and other parallel applications through inter-server network connection
- HPC systems require high computing cost effectiveness. Using UNIX servers as fat nodes are the most cost-effective mode for HPC systems.
- Unsatisfying 8-socket x86 single-node computing performance
- KunLun fat node
- Inner-server high-bandwidth, low-latency interconnect, delivering higher efficiency than inter-server interconnect
- Applicable to the Open-MP and other applications that require more computing resources per node and also applicable to the MPI parallel programming mode

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Database/OLTP Scenario

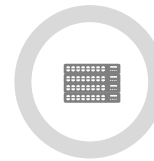
- **Single databases**

For single-database deployment (cannot be split), the overall performance can be improved only by expanding single-node processing capabilities, not by adding servers. For example, **online banking and BOSS core database.**



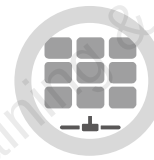
- **Hybrid deployment (databases + applications)**

Application-driven architecture is used in traditional UNIX server application scenarios. Due to the architecture restrictions, to bind applications with the database, various services must be deployed on a single server.



- **Database resource consolidation**

For datacenter consolidation and migration, IT server equipment must realize resource sharing, on-demand use, and simplified maintenance.

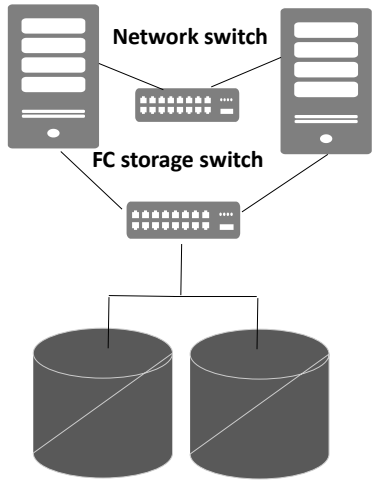


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Database Scenario - Single OLTP Database



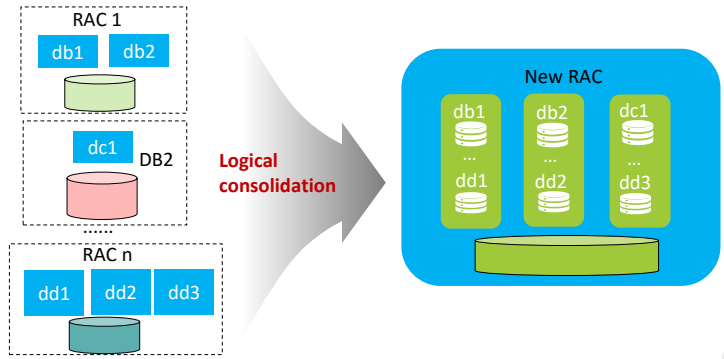
- Application Scope:
  - For single-database deployment (cannot be split), the overall performance can be improved only by expanding single-node processing capabilities, not by adding servers. For example, online banking and BOSS core database.
- Competitiveness:
  - Local HA databases for the production system: most essential measure to ensure service continuity of production databases
- Emergency database for mission-critical services: quickly takes over mission-critical services from the production database if certain modules or the entire system of the production database fails.
- Performance assurance: KunLun's powerful single-node processing capability and two-node cluster solution can meet the requirements of ultra-large single-node database systems, preventing performance from becoming performance bottlenecks.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Database Scenario - Database Consolidation



Oracle 11g database physical consolidation:

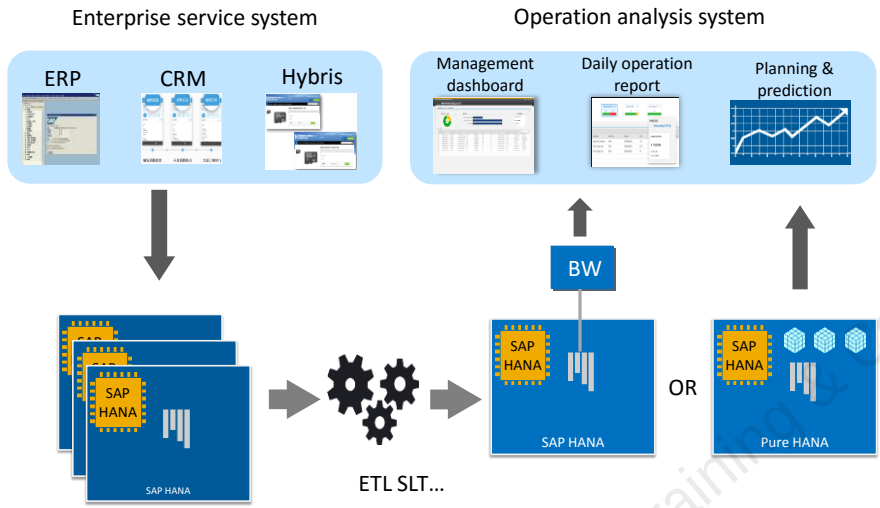
- Quick deployment and management
- Lower O&M complexity and costs
- KunLun 9032 perfectly consolidates multiple existing databases thanks to its powerful performance and partitioning capability.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

SAP HANA Scenario (1)



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

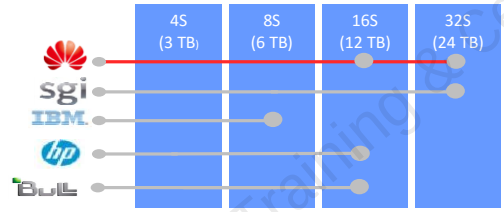
HUAWEI

Huawei Training & Certification
Huawei Training & Certification

SAP HANA Scenario (2)



- **Scenarios**
  - DDR4 DIMMs with over 6 TB memory capacity, high performance, high specifications.
  - Large SAP mission-critical service systems: such as the enterprise resource planning system, supply chain management system, and customer relationship management system.
- **Competitive Strength**
  - Support for scale-up to 24 TB memory capacity per node, the world's largest single-node specifications.
  - Huawei-developed SSD, optimized for the HANA data flow model.
  - Requiring no external disk enclosures; comprehensive services and one-stop hardware service.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

HPC Fat Node





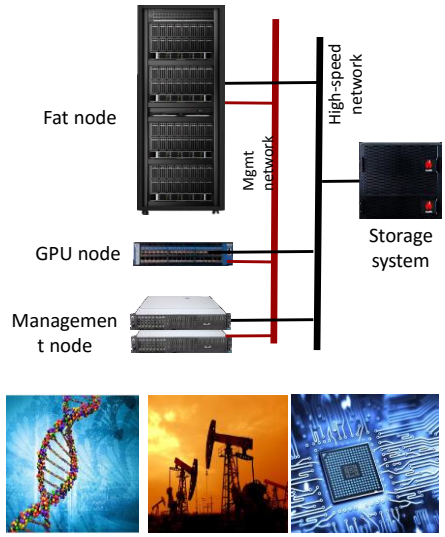
- **Business Scenarios**
  - CAE simulation, life science, oil exploration, and meteorology
- **Application Scope**
  - Applicable to applications that require high computing density and large memory capacity (usually over 1 TB)

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

HPC Fat Node Solution Features



- Service scenario
  - CAE simulation, life science, oil exploration, and meteorology
- Application scope
  - Applicable to applications that require high computing density and large memory capacity (usually over 1 TB)
- KunLun fat node solution
  - Offers higher bandwidth in a single node, delivering better acceleration ratios.
  - Uses multiple communication methods, including Open-MP and MPI in a single device, making it applicable to a vast range of applications.
  - Serves as the fat node of the largest specifications on the Intel Xeon architecture.
  - Supports GPU computing acceleration.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. Positioning and Application of Huawei KunLun Servers
2. **Hardware Architecture and Features of Huawei KunLun Servers**
3. Initial Configuration of Huawei KunLun Servers

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

KunLun 9032/9016/9008 Front View



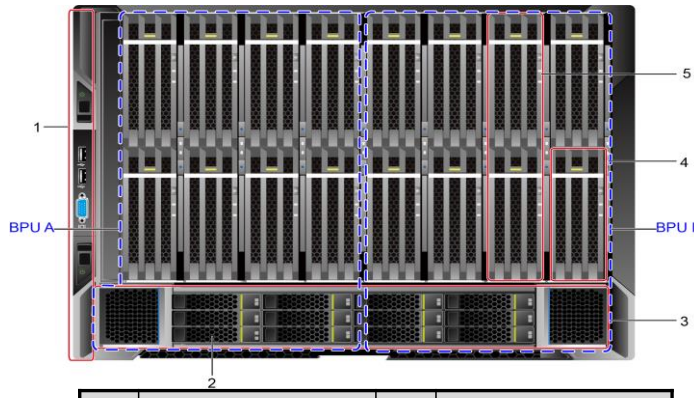
No.	Component
1	APD
2	SCE 4
3	SCE 3
4	CME
5	SCE 2
6	SCE 1

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

SCE Front View

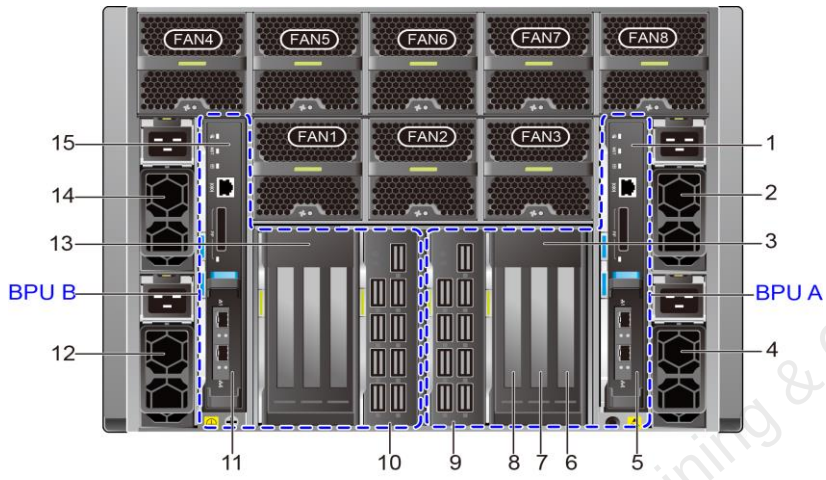


No.	Component	No.	Component
1	Left mounting ear	2	Hard disk
3	FIO	4	MBM
5	SCM	-	-

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

SCE Rear View



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

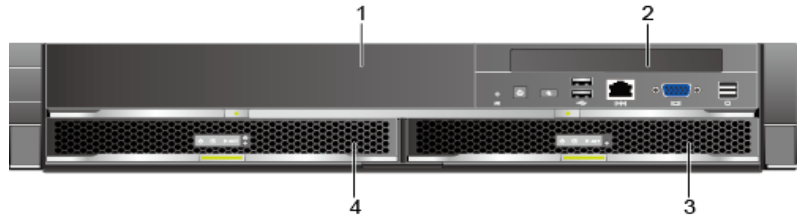
SCE Rear Panel Components

No.	Component	No.	Component
1	LPM 2	2	PSU 4
3	BIO 2	4	PSU 3
5	LOM	6	PCIe card 3 (non-hot-swappable)
7	PCIe card 2 (non-hot-swappable)	8	PCIe card 1 (non-hot-swappable)
9	NCM 2	10	NCM 1
11	LOM	12	PSU 1
13	BIO 1	14	PSU 2
15	LPM 1	-	-

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

CME Front View



No.	Component	No.	Component
1	CIM	2	DVD
3	CMC 2	4	CMC 1

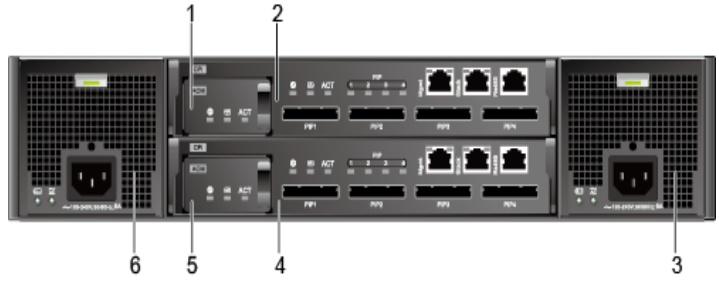
- CMC: central management console
- CIM: central interface module

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

CME Rear View



No.	Component	No.	Component
1	ACM 2	2	CPI 2
3	PFM 2	4	CPI 1
5	ACM 1	6	PFM 1

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

CME Rear Panel Components

No.	Component	Description
3/6	PFM	Supplies power to and dissipate heat from the CME.
1/5	ACM	Provides a single-source 100 MHz reference clock, which is used by the QuickPath Interconnect (QPI), node interconnect (NI), and PCIe ports.
2/4	CPI	Provides partition control functions, including synchronous power control, reset, error isolation, and status and TSC clock verification, CPU overheat protection, and CPU hot swap control.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

CME Front Panel Ports



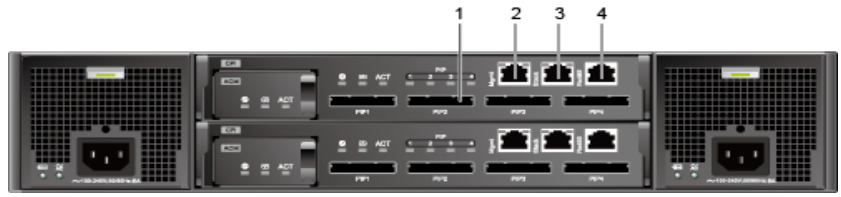
1. USB port 2. Serial port
3. VGA port 4. LCD connector port

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

CME Rear Panel Ports



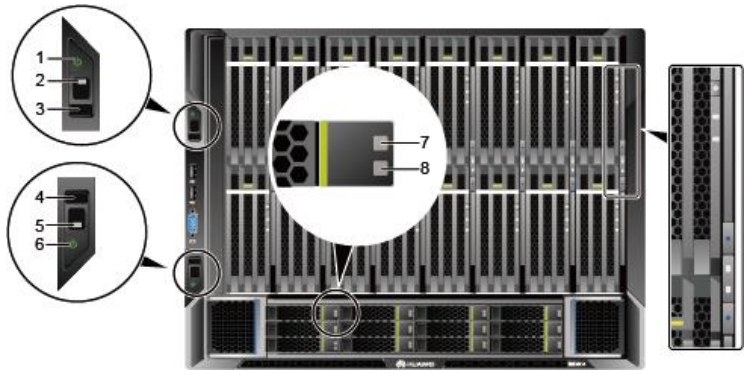
1. Partition interconnect port (PIP)
2. Management network port (MGMT)
3. Stacking port (STACK)
4. RS485 serial port

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

SCE Front Panel Indicators and Buttons



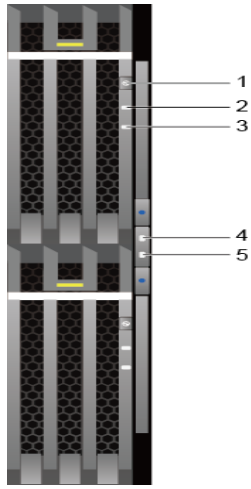
1. BPU A power button/indicator 2. BPU A health indicator 3. BPU A UID button/indicator
4. BPU B UID button/indicator 5. BPU B health indicator 6. BPU B power button/indicator
7. Hard disk fault indicator 8. Hard disk activity indicator

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

SCM Indicators and Buttons



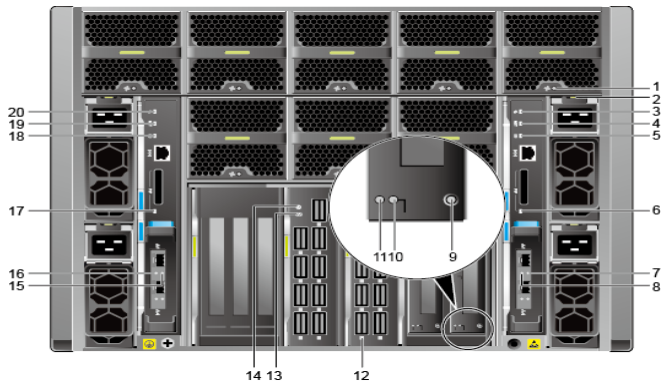
- 1. MBM power button/indicator
- 2. MBM ATTN indicator
- 3. MBM backup indicator
- 4. CPU board module status indicator
- 5. CPU board module ATTN indicator

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

SCE Rear Panel Indicators and Buttons



1. Fan module status indicator
2. PSU status indicator
3. BPU A health indicator
4. BPU A UID indicator
5/18. Primary BPU indicator

6. LPM 2 PIC installation status/health indicator

7/16. Network port data transmission status indicator

8/15. Network port connection status indicator

9. PCIe card hot swap button

17. LPM 1 PIC installation status/health indicator

19. BPU B UID indicator
20. BPU B health indicator

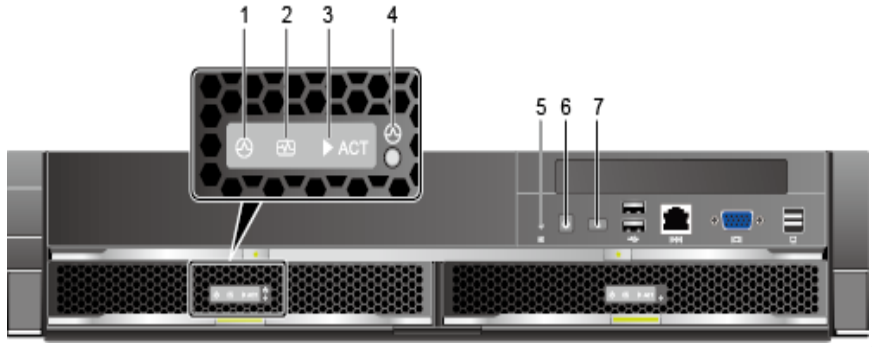
10. PCIe card power indicator
11. PCIe card status indicator
12. NI port status indicator
13. NCM health indicator
14. NCM startup indicator

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

CME Indicators and Buttons

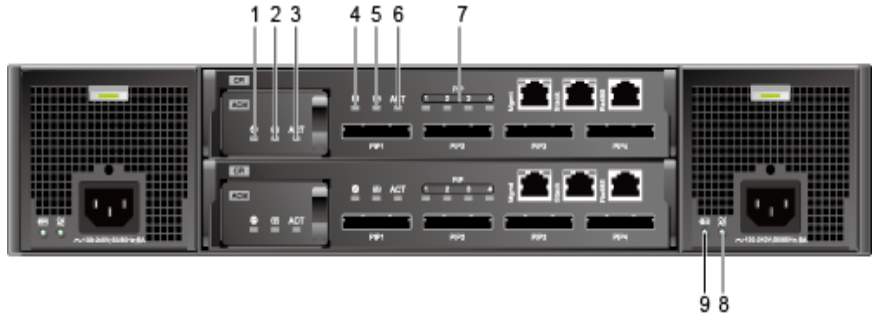


1. CMC power-on indicator
2. CMC health indicator
3. CMC active/standby indicator
4. CMC reset button
5. System health indicator
6. System power button/indicator
7. CIM UID indicator

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

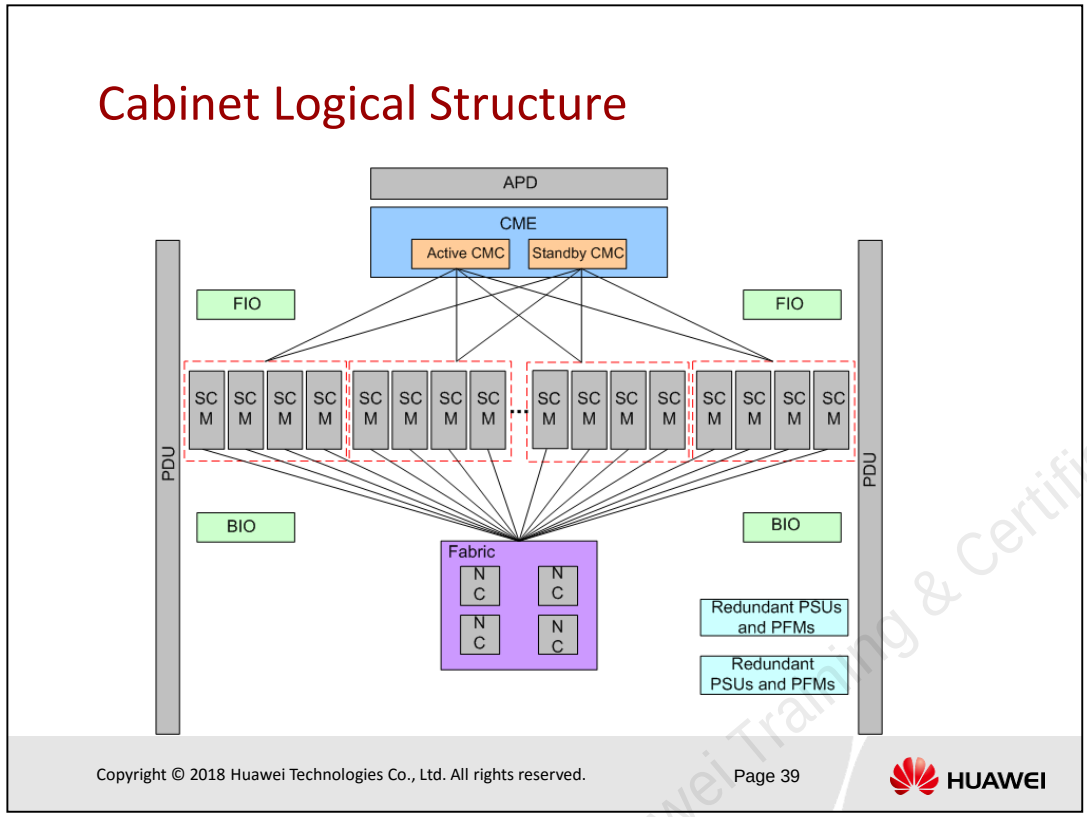
CME Rear Panel Indicators



1. ACM startup indicator
3. ACM health indicator
3. ACM active/standby indicator
4. CPI startup indicator
5. CPI health indicator
6. CPI active/standby indicator
7. PIC installation status/health indicator
8. PFM status indicator
9. BBU status indicator

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification
Huawei Training & Certification



- The CME is connected to the SCEs in all physical partitions through the I/O backplane. The four physical partitions and their SCMs are interconnected through the NC backplane and high-speed CXP optical cables.

Huawei Training & Certification

RAS Definition

- RAS stands for Reliability, Availability and Serviceability.
- Reliability: refers to the capability of a product to sustain specific functions in a given time under given conditions. Generally speaking, it refers to the capability of a server to keep operating properly, free from faults.
- Availability: refers to the capability of a product to function properly for as long time as possible. Generally speaking, it refers to the capability of a server to function properly despite faults.
- Serviceability: refers to the capability of a server to locate and recover from faults. Generally speaking, it refers to the capability of a server to quickly recover from faults.
- The core idea behind the RAS design is to maximize customer service availability or minimize the breakdown possibility. A highly available standalone system must have highly reliable underlying hardware and software design, high fault tolerance performance, fast recovery capability, and rapid service capability.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

RAS Importance

- For mission-critical servers, the core requirement is that the system can provide uninterrupted services. This is because the loss caused by server breakdown varies with applications and more mission-critical services mean more loss in case of breakdown.
- Enterprises are becoming increasingly dependent on IT systems, and the breakdown cost is increasing. According to statistics from AppDynamics, the annual average growth rate of unplanned breakdown costs is 65% from 2010 to 2014. Unexpected downtime may not only cause financial loss but only have other negative impact: damage to the corporate image due to extensive media reporting, increase in the customer churn rate, and employees' work schedule foul-ups.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Basic RAS Features

- Huawei KunLun series servers are open-architecture mission-critical servers based on Intel® Xeon® E7 processors. They deliver the same reliability as traditional midrange computers. The design of each key component and each link introduces RAS features for protection. KunLun servers support the following basic RAS features:
  - Memory reliability
  - CPU reliability
  - IIO reliability

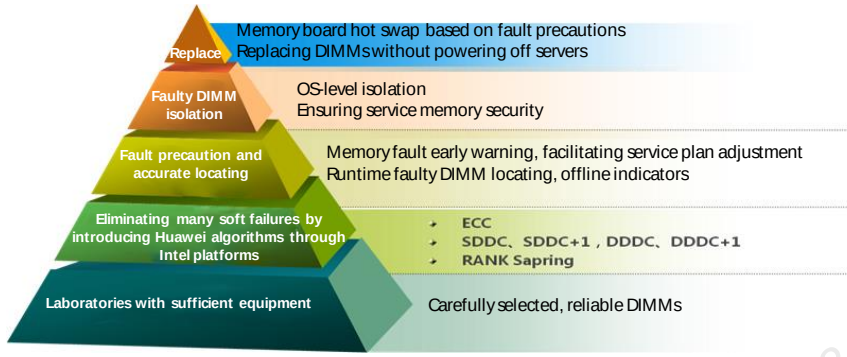
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Memory RAS

End-to-end memory protection mechanism

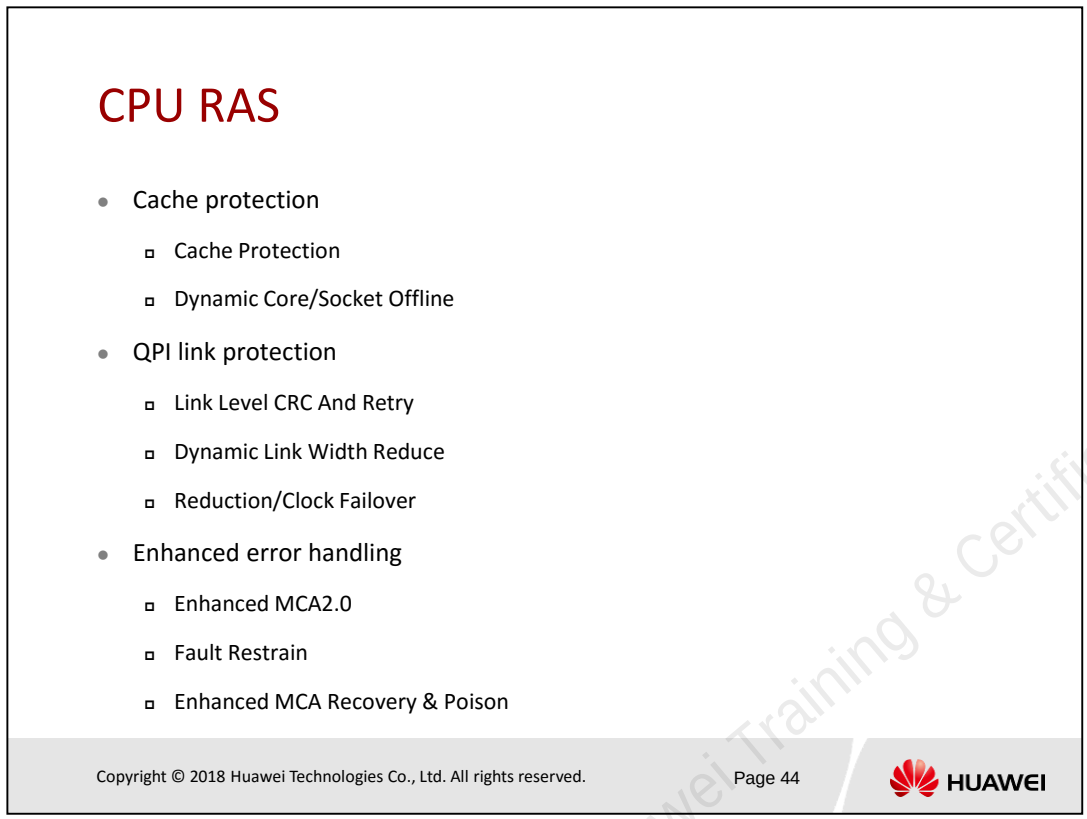


- Strict DIMM selection measures
- Memory fault prevention
- Memory error detection and correction
- Precise locating of faulty DIMMs
- DIMM fault warning
- Dynamic isolation of faulty DIMMs

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- DIMMs are important components of a server and serve as the bridge for communication with CPUs. All server programs run in memory. During system running, CPUs switch data to be computed to the memory for computing, and send the result after the computing is complete. Therefore, memory stability determines server stability. As the memory technology develops rapidly, the chip manufacturing process keeps improving, the operating voltage is decreasing, and the memory capacity is increasing. But more and more memory reliability problems are being reported.
- Currently, RAID is being used to reduce service interruptions caused by hard disk is faults, but no such mechanisms are available for memory. Memory faults usually lead to system breakdown and service interruptions. Because of the inherent features of memory chips, the number of transistors included in the memory linearly increases with the capacity. Therefore, the risk of memory failures increases linearly with the memory capacity.

Huawei Training & Certification - Huawei Training & Certification



- Cache protection: The latest Intel® Xeon® E7 series CPUs provide excellent cache protection mechanisms. During CPU initialization, the system checks the L1 and L2 cache of each core and L3 cache of each CPU and records the check results in registers. The BIOS isolates faulty cores based on the check results. Secondly, errors of the L1 instruction cache can be detected by a parity check. When an error is detected, the error can be corrected by retrying. The ECC feature is configured for L2 and L3 cache to detect multi-bit errors and correct single-bit errors.
- Dynamic Core/Socket Offline: Huawei KunLun servers support real-time cache error monitoring and alarming. When the number of correctable errors exceeds the threshold, an alarm is reported. The OS automatically gets the core or socket offline to isolate the core or socket where errors occur.
- QPI link protection: A QPI link is a bidirectional full-duplex high-speed link consisting of a pair of clock lanes and 20 pairs of data lanes, which is used to connect CPUs.
- Link Level CRC And Retry: QPI links support 8-bit CRC and retransmission capabilities. The CRC mechanism is used to check errors and ensure correct data transmission over QPI links.
- Dynamic Link Width Reduction/Clock Failover: If hardware faults occur on QPI links, for example, errors occur on some data lanes of QPI links, errors cannot be rectified by retransmission because physical links are faulty. To address this problem, QPI links provide the Dynamic Link Width Reduction feature to dynamically decrease the QPI bandwidth and isolate the faulty data lanes.

Huawei Training & Certification

IIO RAS

- **IIO error detection and repot**
  - Error Detect&Report
  - Dynamic Core/Socket Offline
- **PCIe fault tolerance**
  - PCIe Link Retraining and Recovery
  - Link CRC Error Check and Retry
  - PCIe Corrupt Data Containment
  - PCIe Live Error Recovery
  - PCIe hot-Swap (Add/Delete/Repair)

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Enhanced RAS Features

- In addition to the basic RAS features, KunLun servers also support the following enhanced RAS features:
- KPAR/LPAR partitioning technology
- NI link protection (for KunLun 9016 and KunLun 9032 only)
- Online capacity expansion
- Critical memory area mirroring
- MBM and CPU board module hot swap

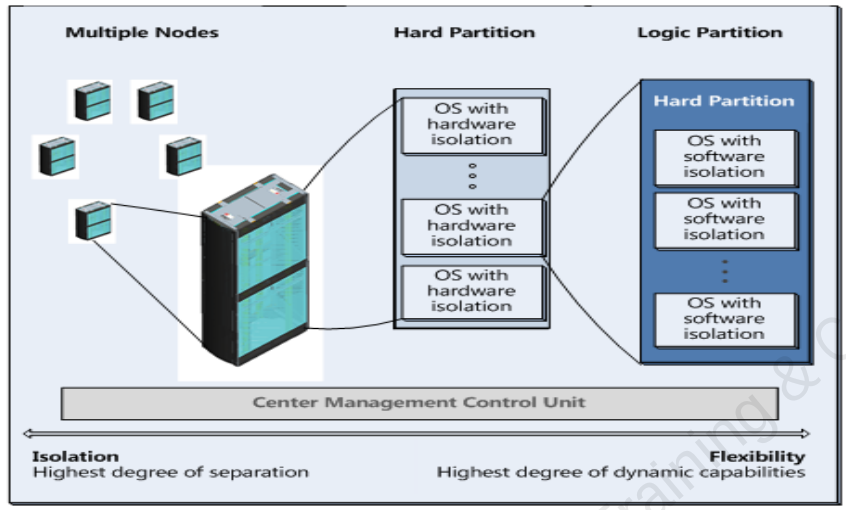
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

KPAR/LPAR Partitioning

KPAR/LPAR technology



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

KPAR

- KPAR is a physical partitioning feature provided by Huawei KunLun servers. It can be used to thoroughly isolate hardware and electrical signals between partitions, ensuring that applications running in a partition are not affected by events of other partitions.
- The KunLun 9008 supports two physical partitions, the KunLun 9016 supports four, and the KunLun 9032 supports eight. Each physical partition has its own OS, applications, peripherals, and networks.
- The CMC in the system uses dedicated programmable hardware to define the boundaries between physical partitions and forcibly isolate a physical partition from the operations in other physical partitions. Each physical partition can be independently powered on and off without affecting other partitions. Users can use a management GUI to configure and monitor physical partitions, and remotely manage them, such as powering on or off, adding, and deleting physical partitions.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- KPAR is a physical partitioning feature provided by Huawei KunLun servers. It can be used to thoroughly isolate hardware and electrical signals between partitions, ensuring that applications running in a partition are not affected by events of other partitions.
- LPAR is a logical partitioning technology. It supports smaller partitioning granularities than KPAR. The minimal partitioning granularity supported by LPAR is CPU core. LPAR greatly improves the hardware resource usage of servers and supports software fault isolation. Applications or OS faults in a logical partition do not affect other logical partitions.

Huawei Training & Certification

KPAR Features

- Easy switching: The CMC allows one-click switching between operating modes of the system, without altering hardware components or upgrading any software.
- Unified management UI: The CMC provides a unified WebUI for management no matter whether the server operates in single-system or physical partition mode.
- Equal allocation of service resources: When the server operates in physical partition mode, service resources, including CPUs, DIMMs, and local storage resources, are equally allocated to each partition.
- High reliability: The physical partitions are electrically isolated from each other to implement fault isolation, improving system reliability over fivefold compared with virtualization software-based partitioning.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

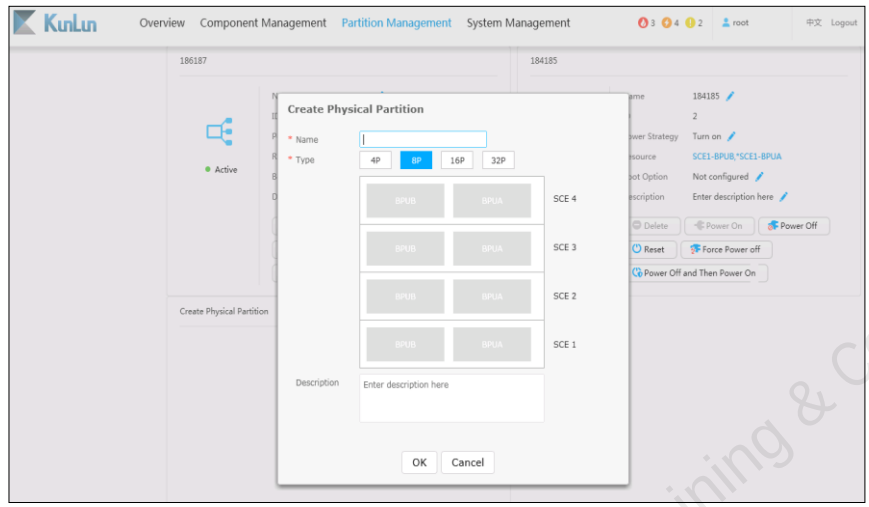
Physical Partition Specifications

Physical Partition Type	9032	9016	9008
32S	Supports the single-system mode	Not supported	Not supported
24S	Not supported	Not supported	Not supported
16S	Supports 2 physical partitions	Supports the single-system mode	Not supported
8S	Supports 4 physical partitions	Supports 2 physical partitions	Supports the single-system mode
4S	Supports 8 physical partitions	Supports 4 physical partitions	Supports 2 physical partitions

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

KPAR Management and Configuration



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

LPAR (Requiring OS Support)

- LPAR is a logical partitioning technology. It supports smaller partitioning granularities than KPAR. The minimal partitioning granularity supported by LPAR is CPU core. LPAR greatly improves the hardware resource usage of servers and supports software fault isolation. Application or OS faults in a logical partition do not affect other logical partitions.
- The OS of each logical partition runs independently. Different logical partitions can run OSs of different versions and different applications. Users can configure and manage logical partitions and dynamically add or delete logical partitions on the GUI. Operations on a logical partition do not affect other partitions, providing high flexibility.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Faulty Partition Isolation and Flexibility

Physical partitions	• Partitions are electrically isolated, providing high performance and reliability. • Hardware faults in a physical partition do not affect other partitions. • The minimal partitioning granularity is four CPUs. • Hypervisor is not required.
Logical partitions	• Logical partitions are implemented based on physical partitions. • The minimal partitioning granularity is CPU cores. A maximum of 1000 logical partitions are supported, providing high flexibility. • Partitions are isolated at the software level. Software faults in a partition do not affect other partitions. • A configuration GUI is provided for users to dynamically add or delete logical partitions.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

NI Link Protection (1)

- Link Level CRC and Retry & Dynamic Link Width Reduction

  An NI port is configured with four links that can independently transmit and receive packets. Each link consists of eight lanes. The NI link supports CRC and retransmission for data transmitted LAN, the NI link automatically isolates the faulty lane and decreases its bandwidth by 50%, dramatically improving system reliability.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

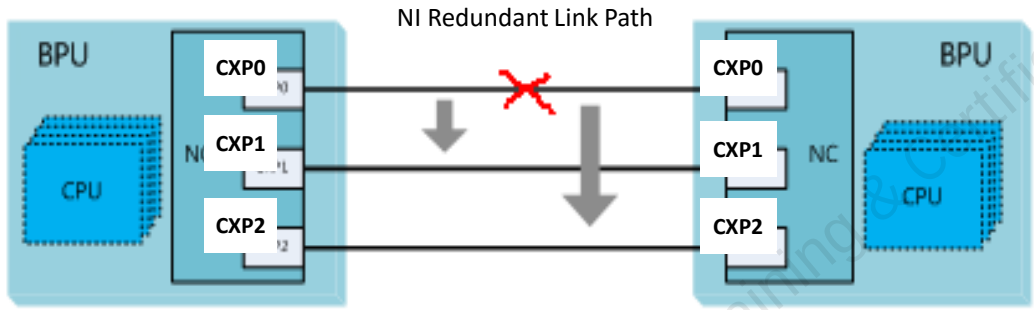
- Protection mechanisms also apply to the QPI links. NC chips are interconnected through node interconnect (NI) links. The NI link also provides the following RAS features:

Huawei Training & Certification

NI Link Protection (2)

- Fully Redundant Link Path & NI Link Hotplug

  Four links on an NI port back up each other. When a link is faulty, the port can still work as long as one link on the port runs properly. Four NI links are provided by three CXP cables. NI links support redundancy, so CXP cables support hot swap. When a CXP cable is faulty, data originally transmitted over the cable is automatically migrated to the functional CXP cables. The faulty CXP cable can be replaced when the system is running, facilitating maintenance.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

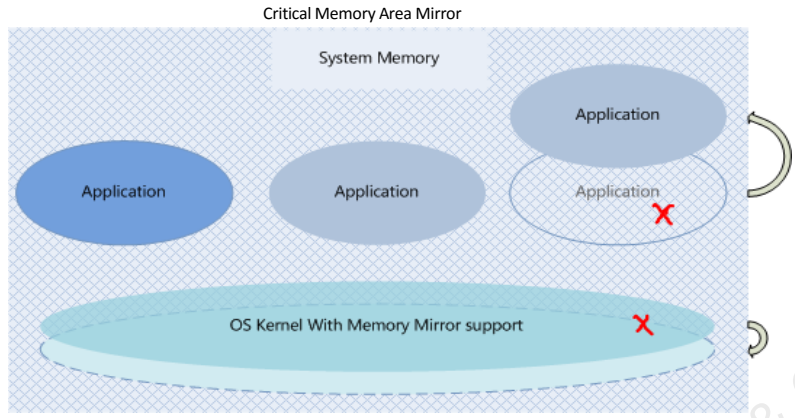
Online Expansion and High Scalability

- If customers want to expand the capacity of IBM Power servers, they need to purchase all hardware devices at the beginning and expand resources by purchasing licenses later. Huawei servers can help customers reduce initial investment. Customers can only purchase required devices at the beginning. As services develop, customers can purchase required devices to expand the system capacity, such as adding more CPUs, DIMMs, or PCIe devices. System expansion can be performed without powering off the system or interrupting services.
- Customers can adopt new technologies (such as new-generation CPUs and DIMMs) to improve system performance. KunLun servers support CPU evolution of the next three generations, meeting increasing service demands of customers and maximizing return on investment (ROI).

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Critical Memory Area Mirroring (1)



- Memory mirroring is an effective method for improving memory reliability. Similar to RAID 1, this feature mirrors a memory channel for another memory channel. When data is written into the memory, the data is written into the two channels but is read from only one channel. If an error occurs and data cannot be restored by using ECC or SDDC, the data is read from the other channel.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

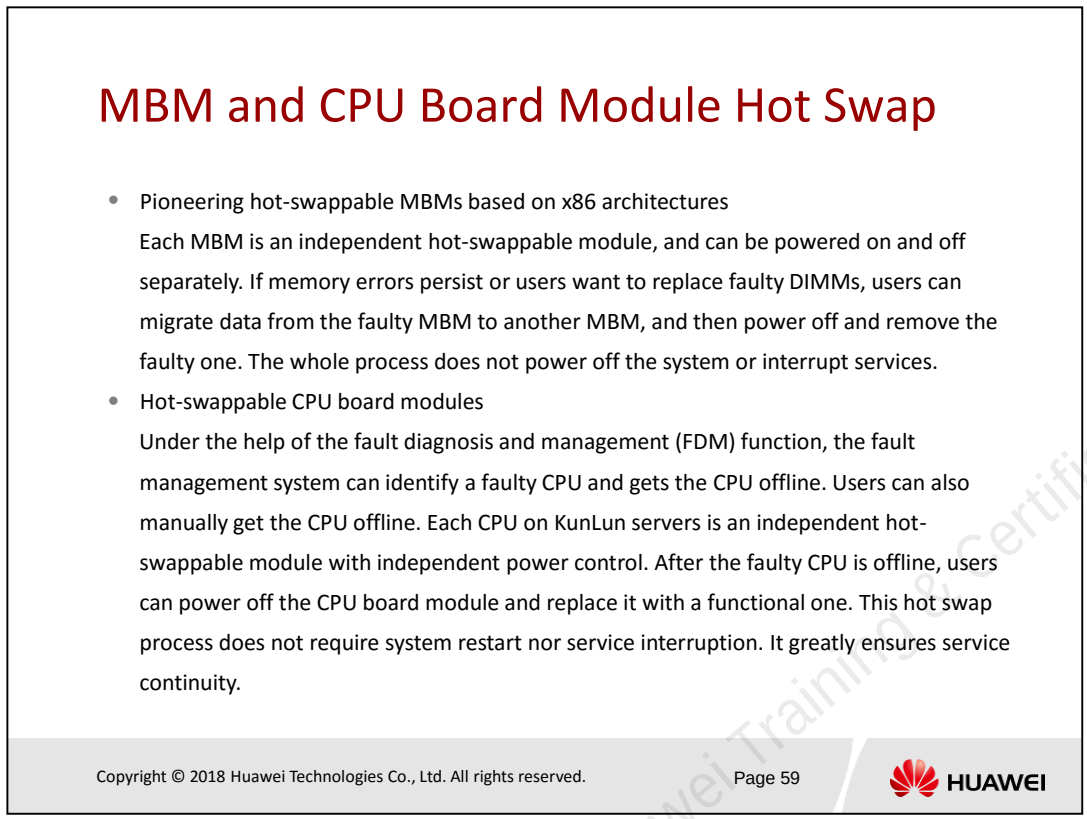
Critical Memory Area Mirroring (2)

- Almost all types of memory errors can be rectified by using memory mirroring. However, the traditional memory mirroring function consumes half of memory capacity. To improve reliability with less cost, KunLun servers provide the Critical Memory Area Mirroring feature. This feature can be configured for certain memory areas. For example, the critical memory areas used by the OS kernel or Hypervisor are mirrored with priority. Interfaces are provided to upper layer applications, which enable users to allocate mirroring memory to critical applications.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- If an uncorrectable error occurs on an upper layer application, the OS restarts the application to rectify the error without affecting other applications.
- If an uncorrectable error occurs on the OS kernel, the OS usually restarts automatically and therefore services are interrupted. After the Critical Memory Area Mirroring feature is enabled, mirror memory is allocated to the OS kernel, and the error can be rectified by switching to the standby mirror memory.
- The Critical Memory Area Mirroring feature provides a cost-effective method for protecting the system. The memory where important data is stored is mirrored, and the memory where unimportant data is stored or idle memory are not mirrored. In this manner, system reliability is ensured while memory costs are reduced.

Huawei Training & Certification



- MBM hot swap
- KunLun 9016 and KunLun 9032 support hot-swappable memory board modules (MBMs) based on x86 architecture. Each MBM is designed as an independent hot-swappable module with independent power control. When a memory fault occurs, if errors persist after granules or ranks are isolated or users want to replace the faulty DIMM to ensure reliability because isolation operations may degrade the memory fault tolerance capability, users can migrate data from the faulty MBM to another MBM, power off the faulty MBM, and remove the faulty MBM. Then, users can locate the faulty DIMM on the MBM based on the fault indicator, and replace the faulty DIMM with a functional one. The whole process does not power off the system or interrupt services. This feature enables users to replace faulty components during system running, ensuring system reliability.
- CPU board module hot swap
- KunLun 9016 and KunLun 9032 support hot-swappable CPU board modules. Under the help of the fault diagnosis and management (FDM) function, the FMS can identify the faulty CPU and gets the CPU offline. Users can also manually get the CPU offline. Each CPU on KunLun 9016 and KunLun 9032 is an independent hot-swappable module with independent power control. After the faulty CPU is offline, users can power off the CPU board module and replace it with a functional one. This hot swap process does not require system restart nor service interruption. It greatly ensures service continuity.

Huawei Training & Certification

Contents

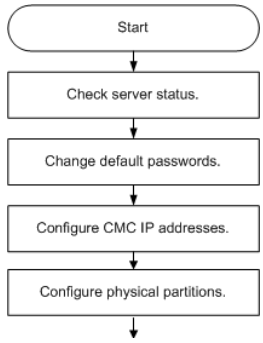
1. Positioning and Application of Huawei KunLun Servers
2. Hardware Architecture and Features of Huawei KunLun Servers
3. **Initial Configuration of Huawei KunLun Servers**
   - Configuring the CMC
   - Configuring Physical Partitions
   - Configuring the BIOS and RAID
   - Installing an OS

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configuration Process (1)

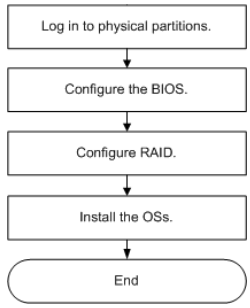


Task	Description
Check server status.	• Log in to the CMC WebUI. • Check software versions. • Check server alarms.
Change initial passwords.	• Change the initial password of the default CMC user. • Change the initial password for the CMC U-Boot.
Configure CMC IP addresses.	Configure the following data: • Static IP addresses for the active and standby CMCs • Floating IP address and default gateway
Configure physical partitions.	• A KunLun server can be divided into multiple physical partitions, with each partition running its own OS. The number of processors in each partition can be 4, 8, 16, or 32. By default, a server works as a single system. • A KunLun server must be partitioned equally.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configuration Process (2)



Task	Description
Log in to physical partitions through the remote virtual console.	Log in to each physical partition of the server through the remote virtual console of the CMC WebUI.
Configure the BIOS.	Configure the server BIOS, including setting the server boot priority and BIOS password, selecting a network driver, and enabling the Preboot Execution Environment (PXE) function for a network interface card (NIC).
Configure RAID.	The KunLun series servers support LSI SAS3108 controller cards. Create RAID arrays for each physical partition based on site requirements.
Install the OSs.	Install an OS for each physical partition.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Default CMC Configurations

Name	Description
Default user name and password	• User name: root • Default password: Huawei12#$
Static IP addresses and subnet mask of the active and standby CMCs	A static IP address is the fixed IP address of a CMC. The value is the IP address of port eth0, which can be obtained by running the ifconfig command. You can use the static IP address to log in to the CMC CLI and WebUI from a PC. • CMC 1: 192.168.2.1 • CMC 2: 192.168.2.2 • Subnet mask: 255.255.255.0
Floating IP address and subnet mask of the CMCs	A floating IP address is used by the CMC for communicating with an external network, which is bound to the active CMC. The value is the IP address of port eth0:0, which can be obtained by running the ifconfig command. The floating IP address is still bound to the active CMC after the active and standby CMCs are switched over. • Floating IP address: 192.168.2.3 • Subnet mask: 255.255.255.0

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Logging in to the CMC WebUI

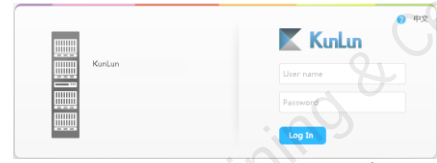
1. Set an IP address for the PC, and ensure that the IP address is on the same network segment as the IP address of the CMC management network port.
2. Connect the PC to the CMC management network port by using a network cable.
3. Open Internet Explorer, enter **https://*IP address*** in the address box, and press **Enter**. ***IP address*** is the IP address of the CMC management network port. The default IP addresses are as follows:

   Static IP addresses: 192.168.2.1 (CMC 1) and 192.168.2.2 (CMC 2)

   Floating IP address: 192.168.2.3

   You are advised to use the floating IP address to log in to the CMC WebUI.
4. In the displayed security alert dialog box, click **Continue to this website (not recommended)**.

   The login page is displayed, as shown in the figure on the right.



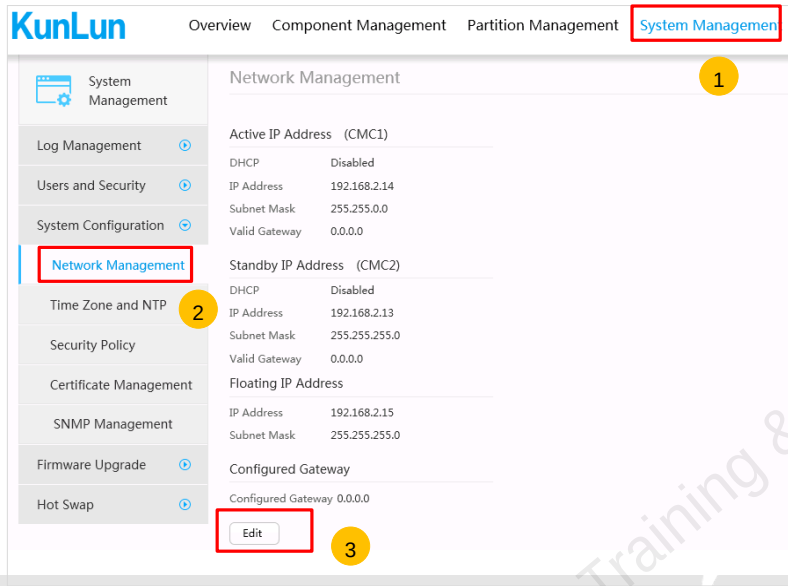
5. Enter your user name and password. The default user is **root**, who belongs to the administrator group. The default password is **Huawei12#$**.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Changing the Default IP Addresses of the CMCs on the WebUI

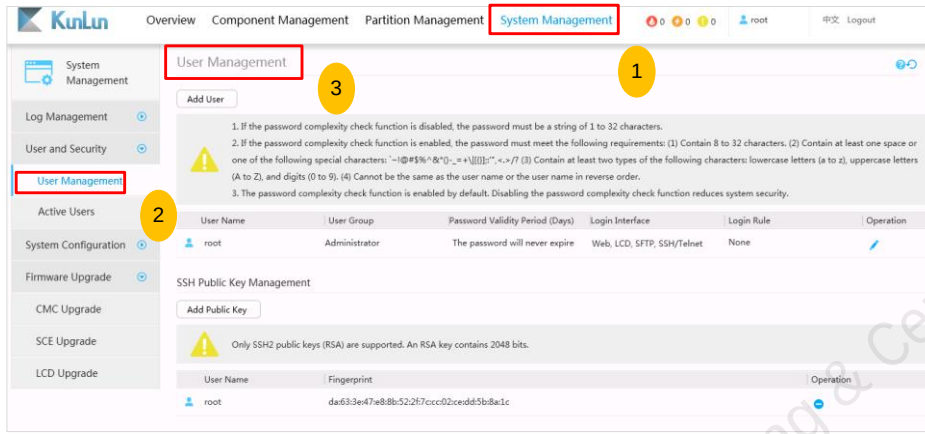


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Adding a User on the WebUI



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Permission description
- Administrator: has permission to perform all query and setting operations.
- Operator: has permission to perform query and setting operations but no permission to upgrade firmware, manage users, or enable and disable password complexity check. To be specific, they cannot run the following commands: firmwareupdate, useradd, userdel, userpasswd, userpropertychg, and userpasswdcomplex.
- Common user: has only the query permission. They can only run the smmget command.

Huawei Training & Certification
Huawei Training & Certification

Contents

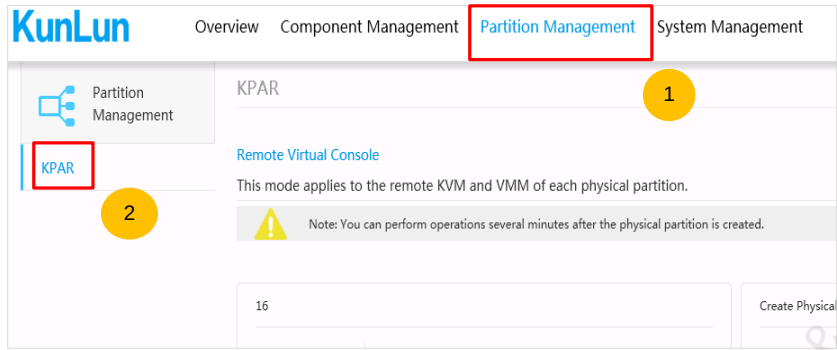
1. Positioning and Application of Huawei KunLun Servers
2. Hardware Architecture and Features of Huawei KunLun Servers
3. **Initial Configuration of Huawei KunLun Servers**
   - Configuring the CMC
   - Configuring Physical Partitions
   - Configuring the BIOS and RAID
   - Installing an OS

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configuring Physical Partitions (1)



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

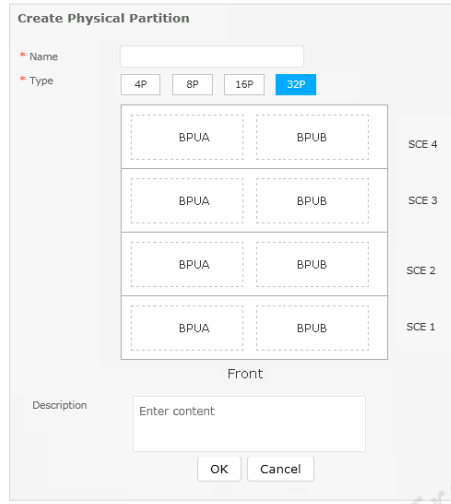
HUAWEI

- Configure a resource pool for unique resources such as MAC addresses, WWNNs, and WWPNs. WWPNs or WWNNs can share a resource pool with MAC addresses. A WWPN or WWNN is a MAC address with the prefix 1000 or 2000 respectively.
- MAC address assignment policies are as follows:
  - You can configure consecutive MAC address segments.
  - The MAC address segment 0x78D752CB0000 to 0x78D752CBFFFF is available for use.
  - Select a unique MAC address segment for each chassis to prevent MAC address conflicts.

Huawei Training & Certification

Configuring Physical Partitions (2)

- Click + in the **Create Physical Partition** area.

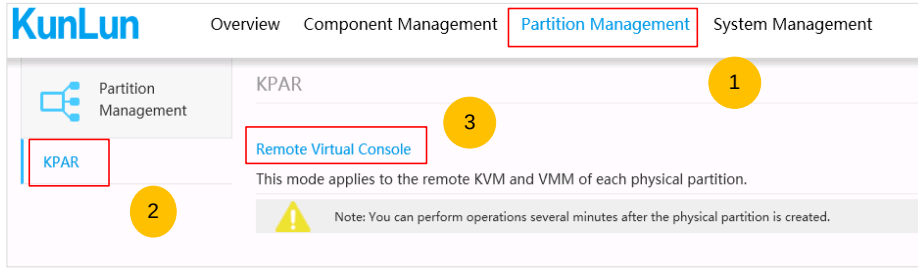


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Logging in to the Remote Virtual Console (1)

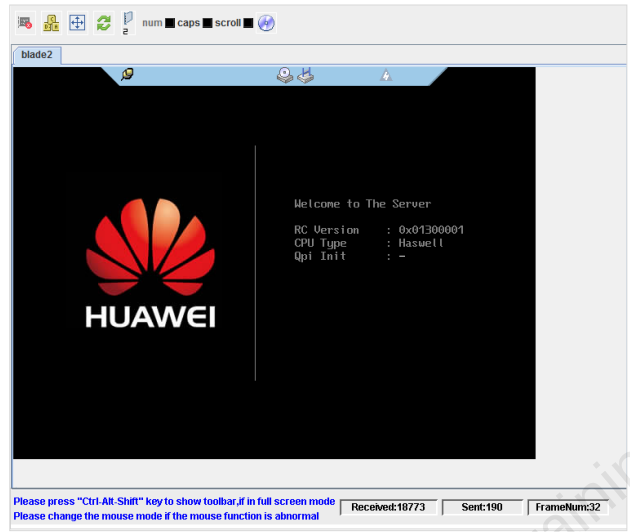


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Logging in to the Remote Virtual Console (2)



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Logging in to the Remote Virtual Console (3)

The following table lists the OSs, browsers, and Java runtime environment (JRE) required for using the Remote Virtual Console.

OS	Browser	JRE
Windows 7 32-bit Windows 7 64-bit	• Internet Explorer 9.0 • Internet Explorer 11.0	• JRE 1.6.0 U25 • JRE1.8.0 U45
	• Mozilla Firefox 34.0	
	• Google Chrome 42.0 • Google Chrome 43.0	
Windows 8 32-bit Windows 8 64-bit	• Internet Explorer 9.0 Internet Explorer 11.0	• JRE 1.6.0 U25 • JRE1.8.0 U45
	• Mozilla Firefox 34.0	
	• Google Chrome 42.0 • Google Chrome 43.0	
Windows Server 2012 64-bit	• Internet Explorer 9.0 • Internet Explorer11.0	• JRE 1.6.0 U25 • JRE1.8.0 U45
	• Mozilla Firefox 34.0	
	• Google Chrome 42.0 • Google Chrome 43.0	

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. Positioning and Application of Huawei KunLun Servers
2. Hardware Architecture and Features of Huawei KunLun Servers
3. **Initial Configuration of Huawei KunLun Servers**
   - Configuring the CMC
   - Configuring Physical Partitions
   - Configuring the BIOS and RAID
   - Installing an OS

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configuring the BIOS (1)

Procedure:

1. Restart a physical partition, log in to the remote virtual console of the partition, and press **Delete** repeatedly to go to the **Main** BIOS screen.
2. Enter the password and go to the BIOS setup utility.
3. Set the server boot priority (for example, hard disk/network/DVD-ROM drive).
4. Change the BIOS password.

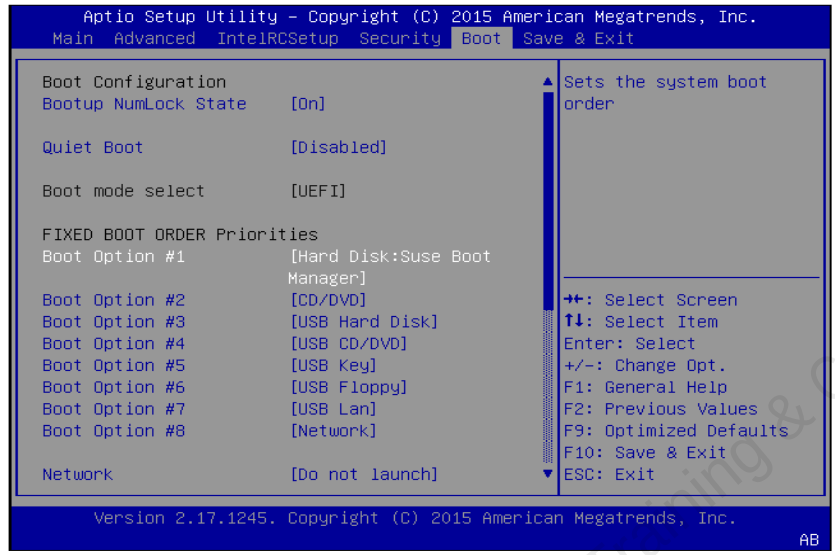
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The current password of the system administrator is required before you change the password. The system will be locked if an incorrect password is entered three consecutive times. You can unlock the system by restarting it.
- Set the administrator password:
- The password must be a string of 8 to 16 characters and contain special characters (including spaces) and at least two types of the following characters: uppercase letters, lowercase letters, and digits.
- The new password cannot be the same as any one of the previous five passwords.
- After the password is set, the following parameters are displayed:
- Clear Supervisor Password
- Power on Password
- Press F10 to save the settings.

Huawei Training & Certification

Configuring the BIOS (2)



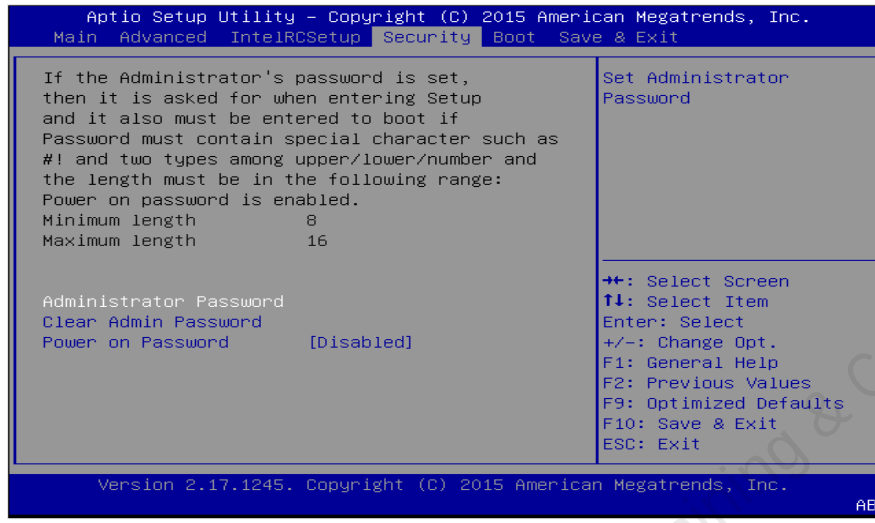
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The current password of the system administrator is required before you change the password. The system will be locked if an incorrect password is entered three consecutive times. You can unlock the system by restarting it.
- Set the administrator password:
- The password must be a string of 8 to 16 characters and contain special characters (including spaces) and at least two types of the following characters: uppercase letters, lowercase letters, and digits.
- The new password cannot be the same as any one of the previous five passwords.
- After the password is set, the following parameters are displayed:
- Clear Supervisor Password
- Power on Password
- Press F10 to save the settings.

Huawei Training & Certification

Configuring the BIOS (3)



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The current password of the system administrator is required before you change the password. The system will be locked if an incorrect password is entered three consecutive times. You can unlock the system by restarting it.
- Set the administrator password:
- The password must be a string of 8 to 16 characters and contain special characters (including spaces) and at least two types of the following characters: uppercase letters, lowercase letters, and digits.
- The new password cannot be the same as any one of the previous five passwords.
- After the password is set, the following parameters are displayed:
- Clear Supervisor Password
- Power on Password
- Press F10 to save the settings.

Huawei Training & Certification

Configuring RAID (1)

Go to the RAID configuration screen and create RAID 0 or RAID 1 for the system disk before OS startup.

Procedure:

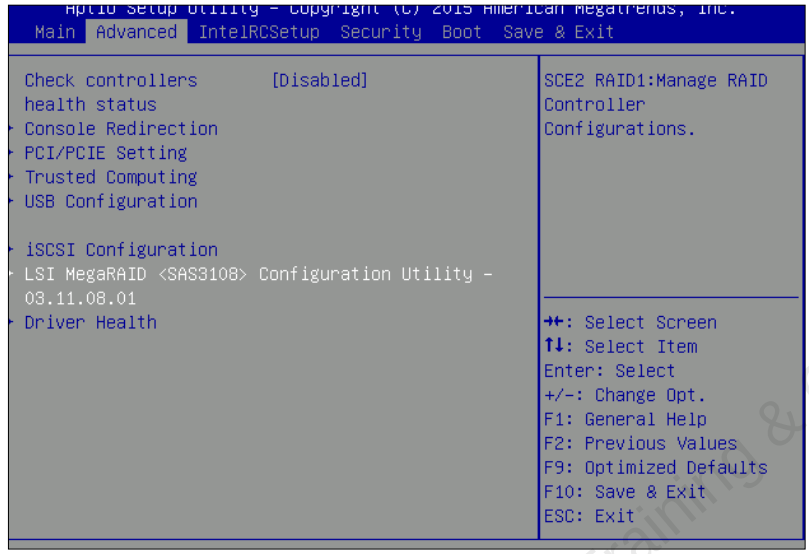
1. Start the server.
2. During startup, press **Delete** as prompted.
3. Select the LSI SAS3108 RAID controller card. Choose **Main Menu** > **Controller Management > Create Virtual Drive** > **Select RAID level**, and press **Enter**. Select a RAID level and press **Enter**.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configuring RAID (2)

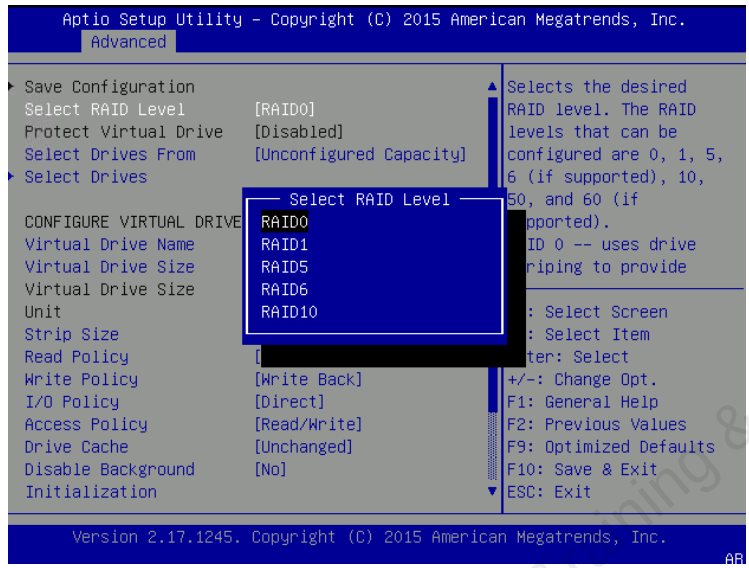


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configuring RAID (3)



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. Positioning and Application of Huawei KunLun Servers
2. Hardware Architecture and Features of Huawei KunLun Servers
3. **Initial Configuration of Huawei KunLun Servers**
   - Configuring the CMC
   - Configuring Physical Partitions
   - Configuring the BIOS and RAID
   - Installing an OS

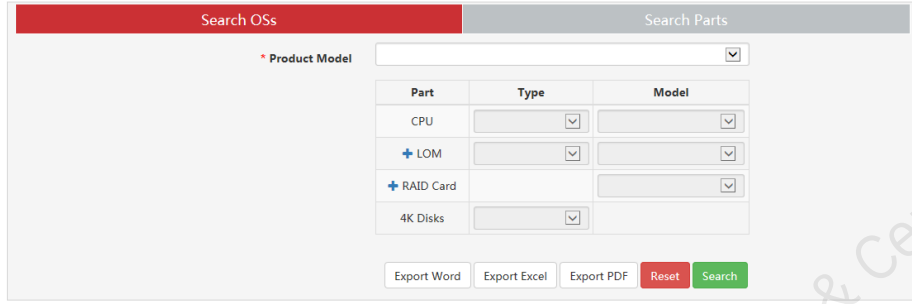
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Preparing for OS Installation (1)

- Determine the OS to be installed based on your application requirements using the server compatibility checker.

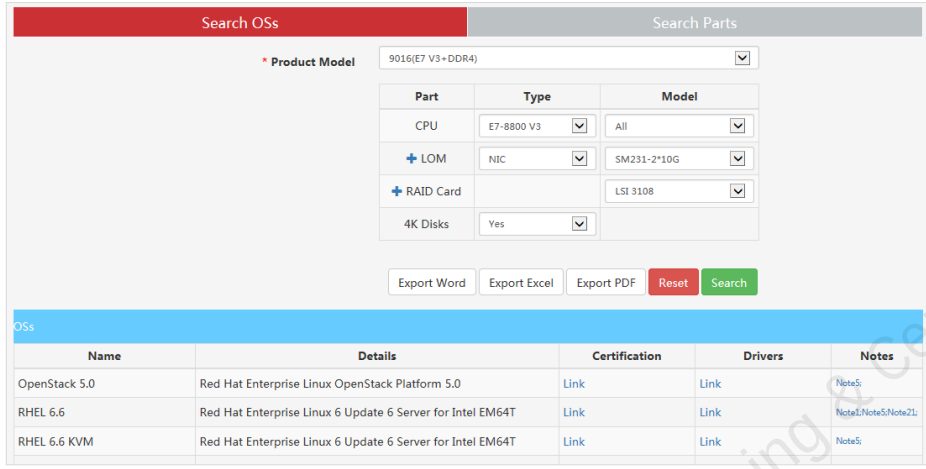


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Visit http://support.huawei.com/onlinetoolsweb/ftca/en.

Huawei Training & Certification
Huawei Training & Certification

Preparing for OS Installation (2)



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

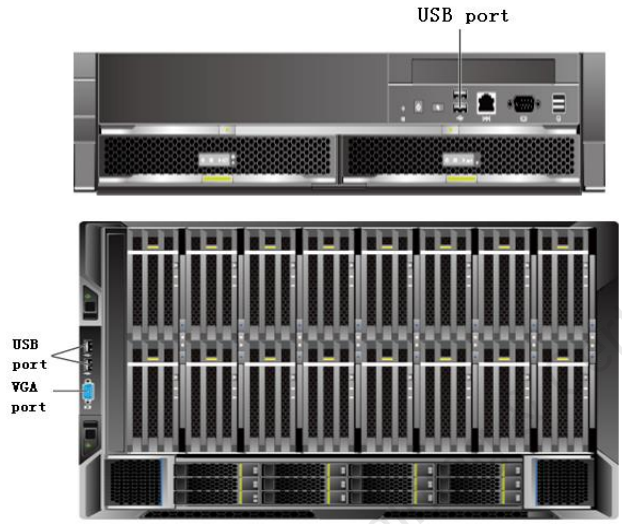
- Select the product model, and component type and model in the compatibility checker to obtain compatibility information shown on this slide.

Huawei Training & Certification

Installation Media - Local DVD-ROM

Install an OS using a local DVD-ROM:

1. Connect the mouse and keyboard to the USB ports on the CME.
2. Connect the USB DVD-ROM drive to the USB port on an SCE.
3. Connect a monitor to the VGA port on the SCE.
4. Enter the BIOS and set the DVD-ROM drive as the boot device.
5. Install the OS in the OS disk.



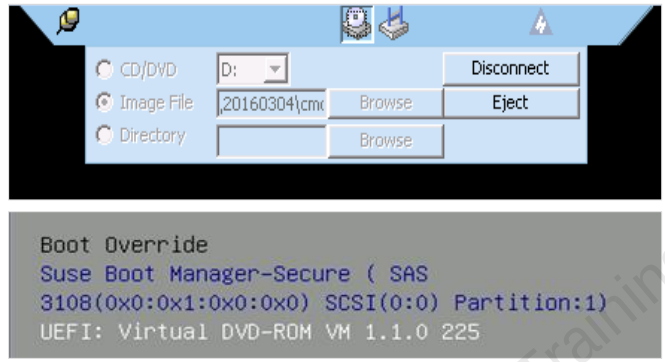
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Installation Media - Image

Install an OS by using an OS image:

1. Log in to the WebUI of the local CMC.
2. Select the remote KVM and load the OS image (.iso) by using the virtual DVD-ROM drive.
3. Restart the compute node and select **UEFI: Virtual DVD-ROM VM 1.1.0 255** during the startup.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Quiz

1. (True or False) The KunLun 9016 supports a maximum of eight physical partitions.
2. (Multiple-Answer Question) What are the application scenarios of KunLun servers?
3. A. Database
4. B. In-memory computing
5. C. HPC fat node
6. D. DR and backup

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Summary

- Positioning and Application of Huawei KunLun Servers
- Hardware Architecture and Features of Huawei KunLun Servers
- Initial Configuration of Huawei KunLun Servers

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

More Information

- Huawei enterprise technical support
  - http://support.huawei.com/enterprise/en/index.html
  - You can obtain product documentation, compatibility information, product software, and maintenance services.
- Huawei server information self-service platform
  - http://enterprise.huawei.com/topic/Self_service_server_en/knowledge.html
  - You can obtain Huawei server installation, commissioning, and O&M information and tools.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Recommendations

- Huawei learning website
  - http://support.huawei.com/learning/Index!toTrainIndex
- Huawei support case library
- http://support.huawei.com/enterprise/servicecenter?lang=en

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

Huawei Blade Server Application Scenarios and Deployment

www.huawei.com

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.



Huawei Training & Certification
Huawei Training & Certification

Foreword

- The course describes what Huawei blade servers can be used for and how to deploy this kind of servers. You will know basic components and how to deploy and configure each component of a blade server. You will also understand the relationships among compute nodes, management modules, and switch modules.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Objectives

- On completion of this course, you will be able to understand:
  - Product positioning, highlights, and application scenarios of E9000 servers
  - Hardware structure of E9000 servers
  - Hardware components of E9000 servers
  - How to configure E9000 servers

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. **E9000 Positioning and Application Scenarios**
2. E9000 Overview
3. E9000 Deployment and Management

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

E9000 Positioning

- The E9000 is a high-end computing platform.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- The E9000 is a high-end computing platform.
- The E9000 is a high-performance, high-end enterprise server designed for elastic computing and telecom computing. The server converges computing, storage, and network resources to support operator- and enterprise-level core applications. The E9000 leads the industry in reliability, availability, and serviceability (RAS), computing density, energy saving, midplane bandwidth, intelligent management and service, elastic configuration and flexible expansion for computing and storage resources, network latency, and network acceleration.

✓ Enhances competitiveness of operators' and enterprises' software services by providing equivalent quality and capability like midrange computers.

✓ Ensures the same competitiveness for operators' universal services and Internet services as cost-effective servers.

✓ Provides high switching bandwidth and short switching latency for services that require convergence of computing and media.

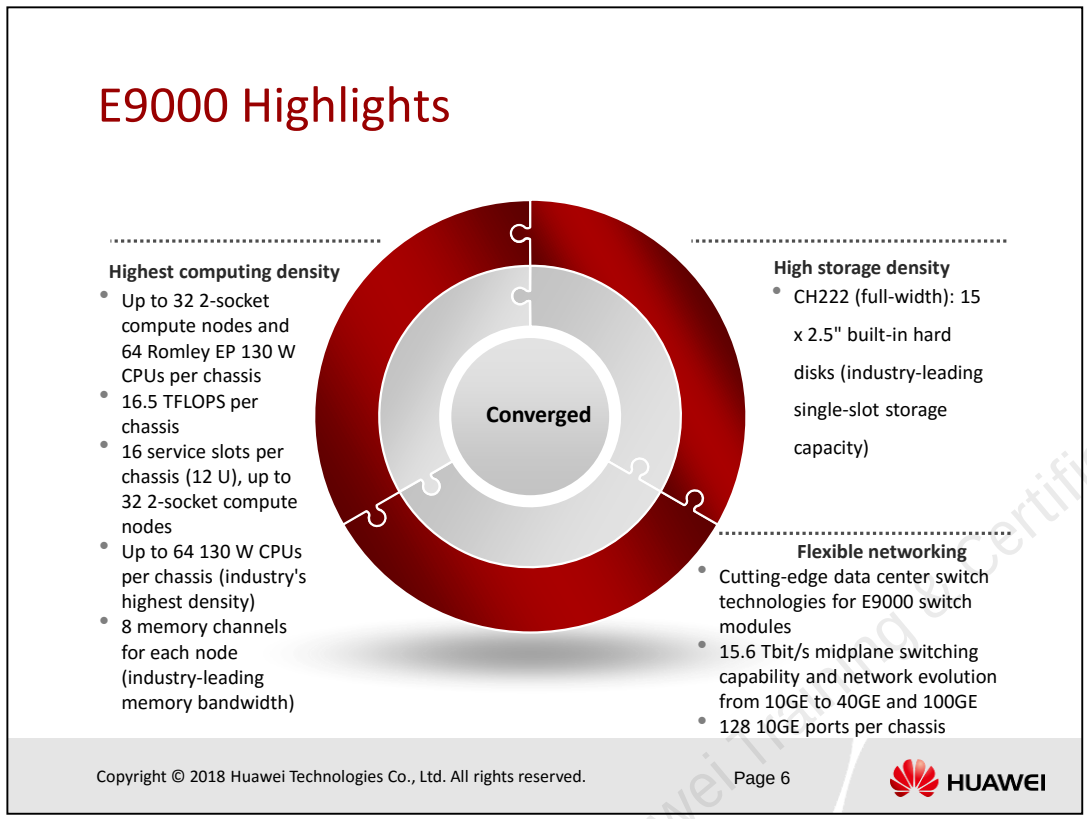
- The E9000 meets various workload requirements in the following typical scenarios:

✓ Cloud computing: The E9000 contains compute nodes with high-performance CPUs and large-capacity memory, which facilitate flexible virtual machine (VM) deployment.

✓ Traditional IT applications: Service processes including web server, application server (middleware), and database processes can be deployed on VMs running on E9000 compute

nodes. External storage devices can be connected to the compute nodes. In this way, the E9000 can run traditional IT applications including office automation (OA), enterprise resource planning (ERP), and business intelligence (BI).

Huawei Training & Certification
Huawei Training & Certification



- Hadoop applications: The E9000 converges computing and storage resources and provides ultra-high-bandwidth switch modules as the optimal platform for big data and high-performance computing applications.
- High-performance computing: The E9000 applies to HPC scenarios and provides InfiniBand switching capability (very high throughput and very short latency).
- The , as a universal service processing platform, provides an industry-leading hardware structure by integrating computing, storage, and network resources in a 12 U chassis.
- High-performance compute nodes
  - The E9000 supports Intel x86 CPUs and is compatible with future generations of CPUs.
  - Various specifications of compute nodes: 2-socket or 4-socket, half-width or full-width
  - Mezzanine cards provide various network interfaces and are compatible with accelerator cards.
- Large capacity of built-in storage
  - The E9000 provides half-width storage expansion modules, which can be integrated with half-width compute nodes to form full-width compute nodes. The full-width nodes converges storage and computing resources.
  - Each compute node supports up to 15 2.5'' SAS/SATA HDDs or SSDs.
  - Disks are installed in removable drive trays and are hot swappable.
  - Power failure protection is provided for the write cache.

Huawei Training & Certification

E9000 Product Family

Chassis



E9000 Chassis

- Modular design: compute nodes, storage nodes, switching modules, power supply modules, and cooling modules can be contained in a single chassis.
- 12U high, providing 8 full-width or 16 half-width slots.
- Compatible with the next three generations of Intel processors.
- Supports next-decade network technology evolution.

Compute nodes

CH121 CH121 V3	CH140	CH220 CH220 V3	CH221	CH222 CH222 V3	CH240	CH242 CH242 V3
						
Half-width 2-socket compute node High density Large memory	Half-width 2-socket twin compute node Ultra-high density Ultra-high computing performance	Full-width I/O expansion node Large memory Large expansion	Full-width I/O expansion node Large memory Large expansion	Full-width storage expansion node Large memory Ultra-large storage	Full-width 4-socket compute node Ultra-high computing performance (Intel E5-4600) Ultra-large memory	Full-width 4-socket compute node Ultra-high computing performance (Intel E7-4800) Ultra-large memory

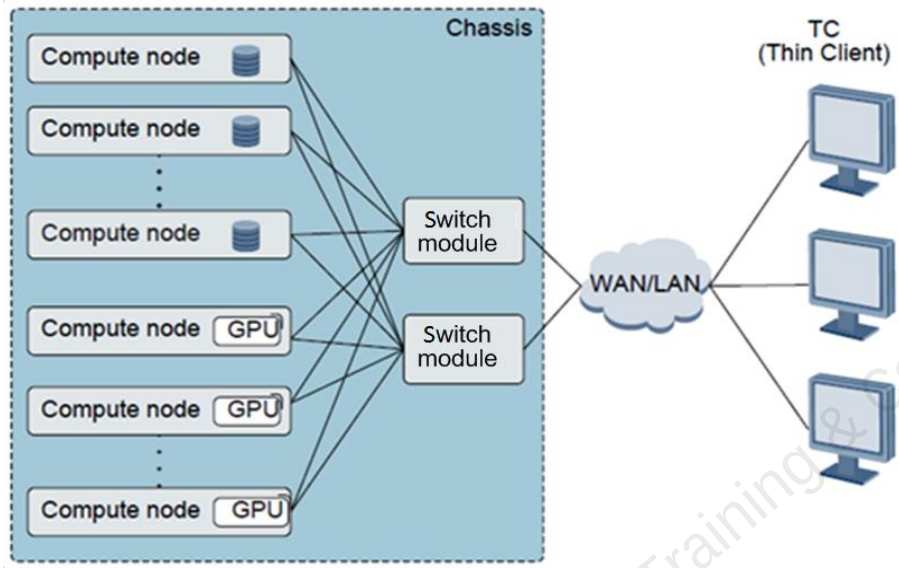
Switch modules

CX110/CX111	CX116	CX310/CX312	CX311	CX317	CX610	CX911 CX912/CX915	CX210
							
GE switch module	GE pass-through module	10GE converged switch module	10GE/FCoE converged switch module	10GE pass-through module	QDR/FDR switch module	10GE/FC multi-plane switch module	8G FC switch module

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Using the E9000 for Desktop Cloud



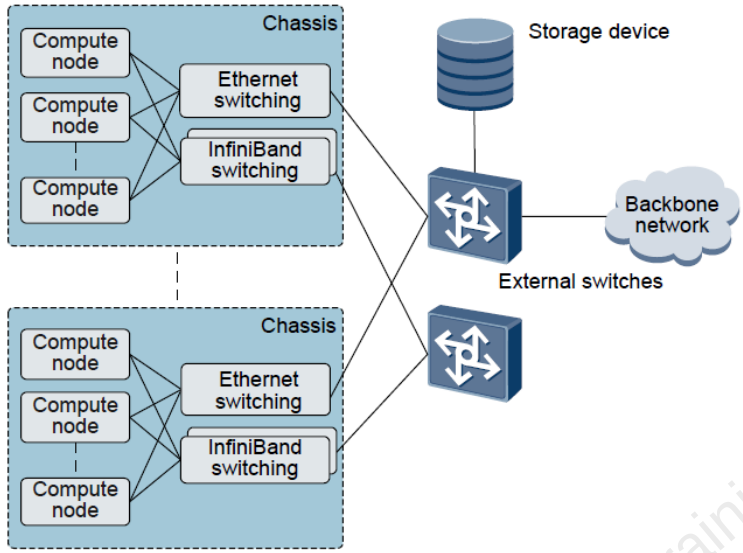
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification

- Desktop Cloud is a Virtual Desktop Infrastructure (VDI) solution provided by Huawei. It provides virtualized desktop systems based on infrastructure as a service (IaaS).
- The E9000 chassis converges computing, storage, and network resources. All the VDI applications can be deployed in a single chassis to implement the Desktop Cloud solution.
- In this example, the provides full-width compute nodes that can be configured with hard disk groups or graphics processing unit (GPU) cards. A compute nodes configured with a hard disk group supports more than 40 VMs. One chassis supports hundreds of VMs, which meets requirements of most enterprises. The details are as follows:
  - Compute nodes configured with hard disk groups provide dozens of TB, high-reliability distributed storage. Compute nodes configured with nonvolatile random access memory (NVRAM) provide high-speed caching and high throughput.
  - Compute nodes configured with GPU cards apply to VM applications of professional graphics processing workstations.
  - The built-in switch modules provide multiple service switching planes. GE switch modules are used for the Desktop Cloud scenario. The switch modules among multiple chassis can be connected for extension.

Using the E9000 for High-Performance Computing



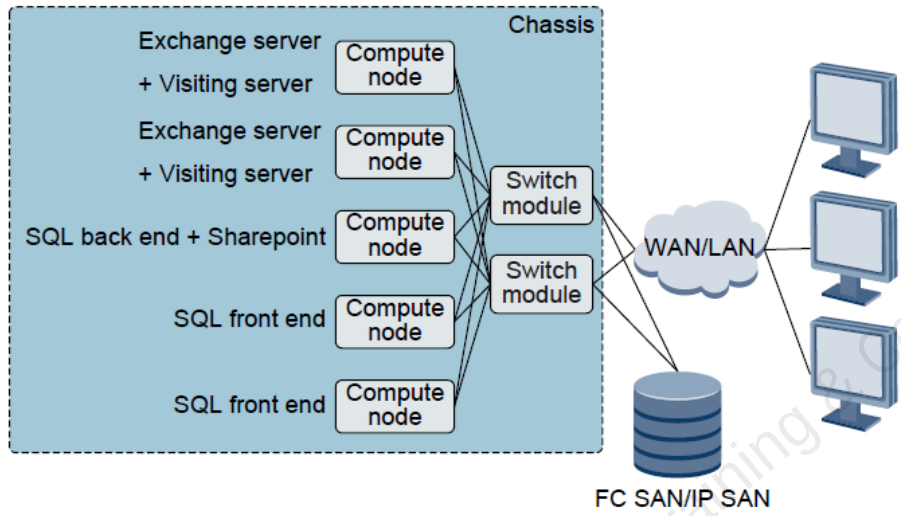
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The E9000 incorporates InfiniBand switching to provide high-performance computing.
- High-performance computing can solve difficult computing problems, such as hydrodynamics, collision simulation, seismic processing, graphics comparison, and price modeling, that require a lot of computing resources. Multiple E9000 chassis are networked with InfiniBand. The compute nodes in these chassis constitute computing small or medium clusters. The nodes can process data concurrently from a massive data computing task.
- This figure shows the networking of the for high-performance computing. Two switch modules in the chassis form an Ethernet switching network, and the other two switch modules form an InfiniBand switching network.
- The Ethernet switching network connects to an external backbone network to receive data computing tasks and return the computing results.
- The InfiniBand switching network serves as the internal computing cluster network to provide high-bandwidth and short-latency parallel computing network resources.

Huawei Training & Certification

Using the E9000 for Traditional IT Applications



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The E9000supports traditional IT applications such as OA, ERP, and BI. The IT applications can run on the E9000 by creating VMs on the compute nodes and deploying web servers, application servers (middleware), and database servers on the VMs.
- This figure shows the networking of the E9000 in Huawei's OA FusionCube solution. In this solution, Microsoft's team collaboration software including Exchange and SharePoint and backend software including SQL Server are deployed on the same equipment.
- In the OA FusionCube solution:
  - All data is stored on a fiber channel (FC) storage area network (SAN) or IP SAN.
  - A large-scale team collaboration environment can be established on the E9000 by creating VMs on the half-width compute nodes and deploying MS SharePoint, Exchange, and SQL Server on the VMs.

Huawei Training & Certification

Contents

1. E9000 Positioning and Application Scenarios

2. E9000 Overview

- Hardware Architecture
  - Components

3. E9000 Deployment and Management

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

E9000 Chassis Appearance

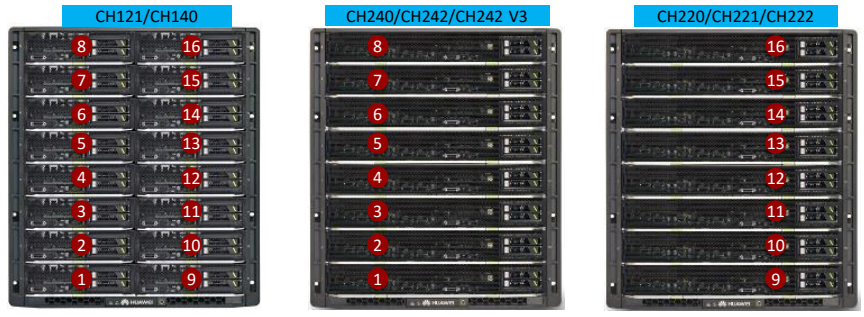


Item	Specifications
Dimensions (H x W x D)	533 mm x 442 mm x 840 mm
Color	Black
Weight	Empty chassis: 70 kg (154.32 lb) Fully configured: 220 kg (485.02 lb)
Rated input voltage	220 V AC 48 V DC
Rated input current	16 A
Rated power consumption	9000 W to 15000 W

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Front View of the E9000 Chassis



Half-width slot numbers | Full-width slot numbers (1) | Full-width slot numbers (2)

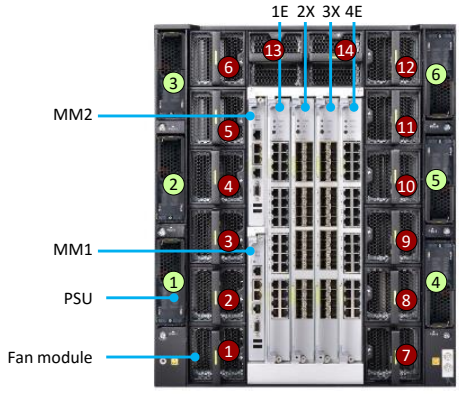
- Eight full-width or 16 half-width compute nodes are supported in the front slots.
- Compute node types vary, and slots of full-width compute nodes are numbered as follows:
- When full-width compute nodes such as CH240, CH242, and CH242 V3 are installed, slots are numbered 1–8 from bottom to top in sequence.
- When full-width compute nodes such as CH220, CH221, and CH222 are installed, slots are numbered 9–16 from bottom to top in sequence. Full-width compute nodes such as CH220, CH221, and CH222 are expanded from the half-width compute node CH121.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Rear View of the E9000 Chassis



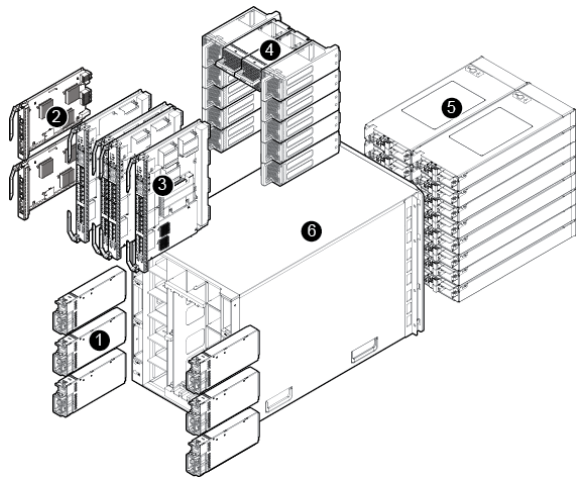
Item	Description
1MM and 2MM	Two slots are provided for installing management modules. The working mode is active/standby.
1E, 2X, 3X, and 4E	Slots for installing two pairs of switch modules.
Fan module	14 fan modules are supported.
PSU	Six PSUs are in N+N redundancy.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Fully loaded E9000 has a total of 6 power supply units (PSUs), supporting N+N and N+1 redundancy. System power varies under different E9000 PSU configurations. According to system power and the power supply modes in the equipment room, you can configure PSUs correspondingly: 3000 W AC PSUs, 2000 W AC PSUs, or 2500 W DC PSUs. Hot swap and redundancy of PSUs are supported.
- The E9000 heat dissipation system contains 14 fan modules that are divided into three zones. Each fan module contains two fan units. Every zone is configured in N+1 redundancy, which ensures that the heat dissipation effect is not affected even if a fan or a fan module fails. The E9000 uses the forced air-cooling technology and front-to-rear ventilation channels. That is, air is drawn into the chassis from the front and dissipated from the rear of the chassis. The fan modules can automatically adjust their speed based on the operating temperature of compute nodes in the chassis or controlled by the management module (MM).
- With four switch module slots, the E9000 provides external network ports and supports their inter-subrack compute node switching capabilities. In the above E9000 rear view, the switch module slots are numbered 1E, 2X, 3X, and 4E from left to right.

Huawei Training & Certification

Components of the E9000 Chassis



No.	Module	Quantity
1	Power supply unit	6
2	Management module	2
3	Switch or I/O module	4
4	Fan module	14
5	Compute node (including the service processing module)	16
6	Chassis	1

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Contents

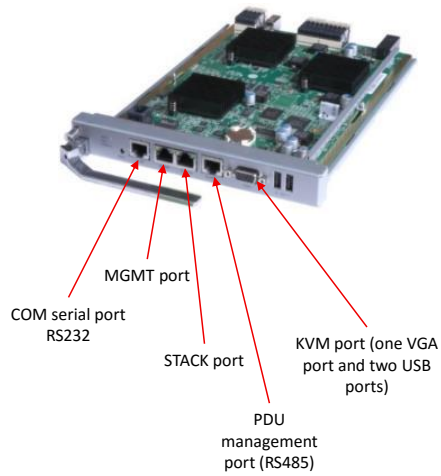
1. E9000 Product Positioning and Application Scenarios
2. **E9000 Product Overview**
   - Hardware Structure
   - Components
3. E9000 Deployment and Management

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

MM910 Management Module



- Installation slot:
- M1 and M2
- Two MM910 management modules work in active/standby mode.
- Feature:
- Manages chassis devices, sensors, and events.
- Intelligently adjusts fan speed.
- Provides IPMI and SNMP interfaces for secondary development.
- Supports KVM over IP and centralized KVM display.
- Supports management over the CLI or WebUI.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Hardware management ports.

接口名称	接口类型	遵循的规范与协议	说明	出线方式
管理维护网口	RJ45	1000Base-T	每个MM910提供1个。	后出线
机箱级联网口	RJ45	1000Base-T	每个MM910提供1个。	后出线
管理维护串口	RJ45	RS232	每个MM910提供1个。 每个交换模块提供2个。	后出线
			每个计算节点通过高密线缆提供1个。	前出线
配电管理口	RJ45	RS485	每个MM910提供1个。	后出线
本地KVM接口	USB VGA	USB 1.0/1.1/2.0 VGA	每个MM910的面板提供1个VGA和2个USB接口。	后出线
			每个计算节点通过高密线缆提供1个VGA和3个USB接口。	前出线
USB接口	USB	USB 1.0/1.1/2.0	每个计算节点在PCB（Printed Circuit Board）上提供用于连接U盘和Flash扣卡的USB接口。	-

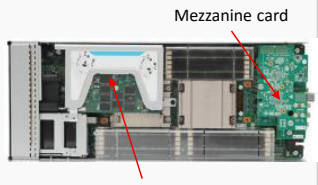
Huawei Training & Certification

CH121 V3 Half-Width Compute Node

Front view



Axial view



Specifications	Half-Width Two-Socket Blade Server
Number of processors	1 or 2
Processor model	Intel Xeon E5-2600 v3 series: 4-core, 6-core, and 8-core
DIMM slots	24 DDR4 DIMMs up to 1.5 TB memory capacity
Number of hard disks	2 x 2.5" SSDs or SAS/SATA HDDs
Built-in storage	2 SATA DOMs 2 microSD cards (RAID 1). Hot swap without opening the chassis is supported. 1 USB 3.0
RAID support	RAID 0 and RAID 1 (LSISAS2308 RAID controller card)
PCIe expansion	2 PCIe x16 mezzanine modules 1 standard full-height half-length PCIe x16 card

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Intel ® Xeon ® E5-2600 V3 (Haswell-EP) series processors ensure high system performance by providing up to 18 cores, a maximum frequency of 3.5 GHz, an L3 cache of 45 MB, and a maximum of two 9.6 GT/s Quick Path Interconnect (QPI) links. Each CH121 V3 supports two processors with 36 cores and 72 threads to maximize concurrent execution of multithreaded applications.
- Each CH121 V3 supports up to 24 load-reduced DIMMs (LRDIMMs), which provides a maximum memory capacity of 768 GB as well as high speed and availability.
- Intel ® Xeon ® E5-2600 V3 (Haswell-EP) series processors provide a maximum theoretical memory bandwidth of 68.3 GB/s (64 bit/8 x 2133 MHz x 4 channels).
- The I/O performance of a system configured with only SSDs is much higher than that of a system configured with both SSDs and HDDs or only HDDs. An SSD supports approximately 100 times more I/O operations per second (IOPS) than a typical HDD.
- Each CH121 V3 provides two integrated Gigabit Ethernet (GE) ports for connecting to a switch module.
- CH121 V3 supports 80-lane PCIe 3.0 I/O. Compared with PCIe 2.0, PCIe 3.0 improves the maximum I/O bandwidth by 60% (8 GT/s per lane).
- Intel® integrated I/O technology enables the PCIe 3.0 controller to be integrated into the Intel® Xeon E5-2600 V3 (Haswell-EP) series processors. This shortens I/O latency and enhances overall system performance.
- CH121 V3 supports flexible configuration of mezzanine modules with diverse network ports.
- CH121 V3 supports standard PCIe SSD cards.

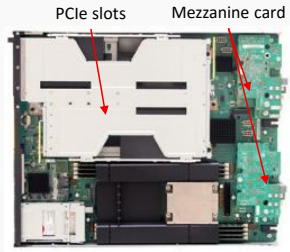
Huawei Learning & Certification - Huawei Training & Certification

CH220 V3 Full-Width I/O Expansion Compute Node

Front view



Top view



Specifications	Full-Width Two-Socket Blade Server
Number of processors	1 or 2
Processor model	Intel ® Xeon ® E5-2600 v3 series
DIMM slots	16 DDR4 DIMMs up to 1 TB memory capacity
Number of hard disks	2 x 2.5'' SSDs or SAS/SATA HDDs. PCIe SSDs are supported.
Built-in storage	2 SATA DOMs 2 microSD cards (RAID 1). Hot swap without opening the chassis is supported. 1 USB 3.0
RAID support	RAID 0 and RAID 1
PCIe expansion	2 PCIe x16 + 2 PCIe x8 mezzanine modules 6 standard PCIe cards of 3 combination modes: • a. 6 full-height half-length single-slot cards • b. 1 full-height full-length dual-slot card + 4 full-height half-length single-slot cards • c. 2 full-height full-length dual-slot cards

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Intel® Xeon® E5-2600 V3 (Haswell-EP) series processors ensure high system performance by providing up to 18 cores, a maximum frequency of 3.5 GHz, an L3 cache of 45 MB, and a maximum of two 9.6 GT/s QPI links. Each CH220 V3 supports two processors with 36 cores and 72 threads to maximize concurrent execution of multithreaded applications.
- Each CH220 V3 supports up to 16 LRDIMMs, which provides a maximum memory capacity of 256 GB as well as high speed and availability.
- Intel® Xeon® E5-2600 V3 (Haswell-EP) series processors provide a maximum theoretical memory bandwidth of 68.3 GB/s (64 bit/8 x 2133 MHz x 4 channels).
- The I/O performance of a system configured with only SSDs is much higher than that of a system configured with both SSDs and HDDs or only HDDs. An SSD supports approximately 100 times more IOPS than a typical HDD.
- Each CH220 V3 provides two integrated GE ports for connecting to a switch module.
- Each CH220 V3 supports 120-lane PCIe 3.0 I/O. Compared with PCIes 2.0, PCIe 3.0 improves the maximum I/O bandwidth by 60% (8 GT/s per lane).
- Intel® integrated I/O technology enables the PCIe 3.0 controller to be integrated into the Intel® Xeon E5-2600 V3 (Haswell-EP) series processors. This shortens I/O latency and enhances overall system performance.
- CH220 V3 supports flexible configuration of mezzanine modules with diverse network ports.
- CH220 V3 supports various PCIe cards to ensure non-convergence bandwidth. You can configure the PCIe cards in GPU/GPGPU, SSD, or hybrid mode as required.

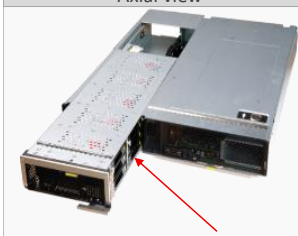
Huawei Training & Certification

CH222 V3 Full-Width Storage Expansion Compute Node

Front view



Axial view



Specifications	Full-Width Two-Socket Blade Server
Number of processors	1 or 2
Processor model	Intel ® Xeon ® E5-2600 v3 series
DIMM slots	24 DDR4 DIMMs up to 1 TB memory capacity
Number of hard disks	15 x 2.5'' SSDs or SAS/SATA HDDs
Built-in storage	2 SATA DOMs 2 microSD cards (RAID 1). Hot swap without opening the chassis is supported. 1 USB 3.0
RAID support	RAID 0, 1, 10, 5, 50, 6 and 60. The RAID cache is 512 MB or 1 GB.
PCIe expansion	2 PCIe x16 mezzanine modules 1 standard full-height half-length PCIe x16 card

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Intel® Xeon® E5-2600 V3 (Haswell-EP) series processors ensure high system performance by providing up to 18 cores, a maximum frequency of 3.5 GHz, an L3 cache of 45 MB, and a maximum of two 9.6 GT/s QPI links. Each CH222 V3 supports two processors with 36 cores and 72 threads to maximize concurrent execution of multithreaded applications.
- Each CH222 V3 supports up to 24 LRDIMMs, which provides a maximum memory capacity of 768 GB as well as high speed and availability.
- Intel® Xeon® E5-2600 V3 (Haswell-EP) series processors provide a maximum theoretical memory bandwidth of 68.3 GB/s (64 bit/8 x 2133 MHz x 4 channels).
- The I/O performance of a system configured with only SSDs is much higher than that of a system configured with both SSDs and HDDs or only HDDs. An SSD supports approximately 100 times more IOPS than a typical HDD.
- Each CH222 V3 provides two GE ports for connecting to a switch module.
- CH222 V3 supports 80-lane PCIe 3.0 I/O. Compared with PCIes 2.0, PCIe 3.0 improves the maximum I/O bandwidth by 60% (8 GT/s per lane).
- Intel® integrated I/O technology enables the PCIe 3.0 controller to be integrated into the Intel® Xeon E5-2600 V3 (Haswell-EP) series processors. This shortens I/O latency and enhances overall system performance.
- CH222 V3 supports flexible configuration of mezzanine modules with diverse network ports.
- CH222 V3 supports flexible hard disk configurations and provides elastic and expandable storage capacities up to 15 2.5-inch hard disks to satisfy users' needs.

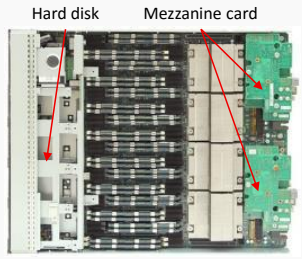
Huawei Training & Certification

CH242 V3 Full-Width Compute Node

Front view



Axial view



Specifications	Full-Width Four-Socket Blade Server
Number of processors	2 or 4
Processor model	Intel ® Xeon ® E7-4800 v2 series: 8-core, 10-core, 12-core, and 15-core
DIMM slots	32 DDR3 DIMMs. The maximum memory bandwidth is supported and the working frequency is 1600 Hz.
Number of hard disks	8 or 4 x 2.5'' SSDs or SAS/SATA HDDs
RAID support	RAID 0, 1, 10, 5, 50, 6, and 60 with a RAID cache of 1 GB.
PCIe expansion	4 PCIe x16 mezzanine modules 2 standard full-height half-length PCIe x16 cards (CH242 V3-4HDD supports one more standard full-height 3/4-length PCIe card.)

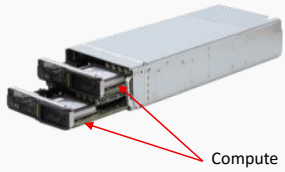
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

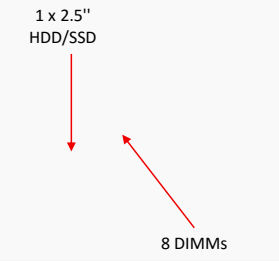
- Intel® Xeon® E7-4800 V2 series processors ensure high system performance by providing up to 15 cores, a maximum frequency of 2.8 GHz, an L3 cache of 37.5 MB, and a maximum of three 8 GT/s QPI links. Each CH242 V3 supports four processors with 60 cores and 120 threads to maximize concurrent execution of multithreaded applications.
- Each CH242 V3 supports up to 32 LRDIMMs, which provides a maximum memory capacity of 1024 GB as well as high speed and availability.
- Intel® Xeon® E7-4800 series processors provide a maximum memory bandwidth of 409.6 GB/s (64 bit/8 x 1600 MHz x 32 channels).
- The I/O performance of a system configured with only SSDs is much higher than that of a system configured with both SSDs and HDDs or only HDDs. An SSD supports approximately 100 times more IOPS than a typical HDD.
- Each CH242 V3 provides two integrated GE ports for connecting to a switch module.
- CH242 V3 supports 128-lane PCIe 3.0 I/O. Compared with PCIe 2.0, PCIe 3.0 I/O improves the maximum bandwidth by 60% (8 GT/s per lane).
- Intel® integrated I/O technology enables the PCIe 3.0 controller to be integrated into the Intel® Xeon E7 series processors. This shortens I/O latency and enhances overall system performance.
- CH242 V3 supports various flexible hard disk configurations and provides elastic and expandable storage capacities to satisfy users' needs. Each CH242 V3-8HHD supports a maximum of eight 2.5-inch hard disks, and each CH242 V3-4HHD supports a maximum of four 2.5-inch hard disks.

CH140 V3 Half-Width Twin Compute Nodes

Front view



Top view



Specifications	CH140 V3 Half-Width Twin Compute Nodes
Number of processors	2 or 4 (1 or 2 for each compute node)
Processor model	Intel Xeon E5-2600 V3 series
DIMM slots	16 DDR4 DIMMs (8 for each compute node)
Number of hard disks	1 x 2.5'' SSD/SATA for each compute node
PCIe expansion	2 mezzanine modules for each compute node
OS support	Microsoft Windows Server 2008/2012 Red Hat Enterprise Linux SUSE Linux Enterprise Server Oracle Linux CentOS Citrix XenServer VMware
Operating temperature	5° C to 35° C (41° F to 95° F)
Dimensions (D x W x H)	Depth: 537.2 mm (21.15 in.) Width: 210 mm (8.27 in.) Height: 60.46 mm (2.38 in.)

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Overview of E9000 I/O Module

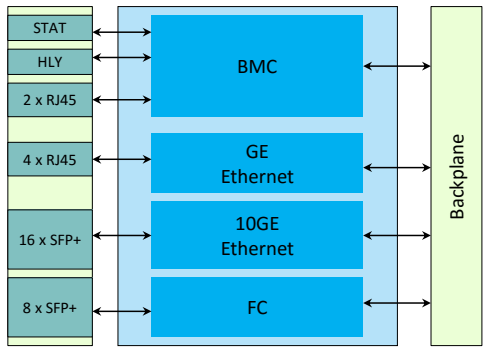
Switch Module Type	Switch Module Model
GE	 CX110  CX111  CX116
FC	 CX210-8G  CX220
Integration of 10 GE and FCoE	 CX310  CX311  CX312  CX317
InfiniBand QDR and FDR	 CX610  CX611
Multiple-plane	 CX911  CX912  CX915

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Interconnection Module: CX911





Uplink ports

- 4 x GE + 16 x 10GE + 8 x 8G FC
- Downlink ports
- 16 x GE + 32 x 10GE + 16 x 8G FC/FCOE
- Backplane directly connected links
- 2 x 40GE (can be configured as stacking or interconnection channels)
- 10GE switching plane
- L2/L3/Stack/FSB/TRILL/VRRP/IPV6
- FC switching plane
- MX510: ODM Qlogic
- Supports connection to FC switches and storage in NPV and full-switch mode.
- Matching mezzanine modules: MZ910 (slots 1 to 12 for FCoE and slots 13 to 16 for FC)
- GE switching plane is not available.

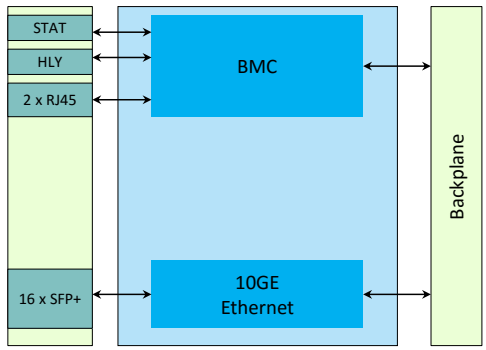
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Interconnection Module: CX310





Uplink ports

- 16 x 10GE
- Downlink ports
- 32 x 10GE
- Backplane directly connected links
- 1 x 40GE (can be configured as a stacking or interconnection channel)
- 10GE switching plane
- L2/L3/Stack/FSB/TRILL/VRRP/IPV6
- The four 10GE ports on the front panel can be configured as one 40GE port, for example, 1/2/3/4 and 5/6/7/8.
- Matching mezzanine modules: MZ510/MZ512/MZ311

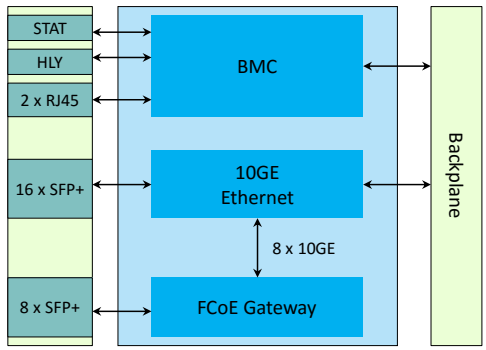
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Interconnection Module: CX311





Uplink ports

- 16 x 10GE + 8 x 8G FC
- Downlink ports
- 32 x 10GE
- Backplane directly connected links
- 1 x 40GE (can be configured as a stacking or interconnection channel)
- 10GE switching plane
- L2/L3/Stack/FSB/TRILL/VRRP/IPV6
- The four 10GE ports on the front panel can be configured as one 40GE port, for example, 1/2/3/4 and 5/6/7/8.
- FC switching plane
- MX510: ODM Qlogic
- Supports connection to FC switches and storage in TR (NPV) and full-switch (Switch) mode.
- Matching mezzanine modules: MZ510/MZ512

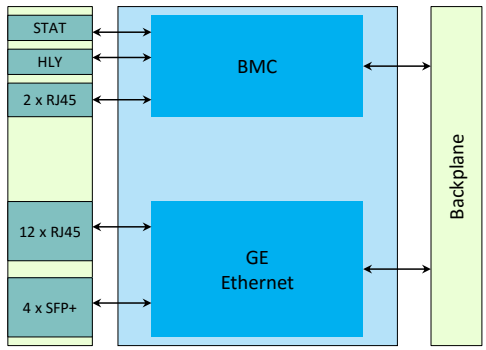
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Interconnection Module: CX111





- Uplink ports
  - 12 x GE + 4 x 10GE
  - Downlink ports
  - 32 x GE
  - Backplane directly connected links
  - None (Only 10GE ports on the front panel are available for stacking.)
  - GE switching plane
  - L2/L3/Stack/VRRP/IPV6
  - Matching mezzanine modules: MZ110/MZ111

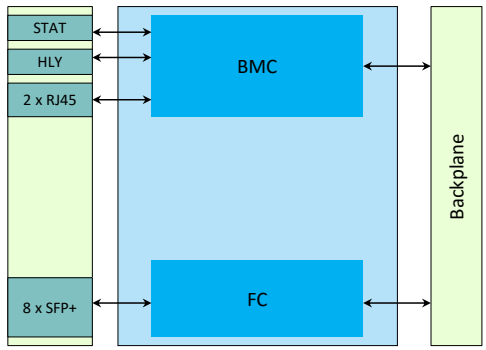
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Interconnection Module: CX210





Uplink ports:

- 8 x 8G FC
- Downlink ports
- 16 x 8G FC
- Backplane directly connected links
- None
- FC switching plane
- MX210: ODM Brocade
- Supports connection to FC switches and storage in AG (NPV) and Native (Switch) mode.
- Matching mezzanine modules: MZ910. All ports that connect to blades are FC ports.
- MZ910 10GE ports are not available.

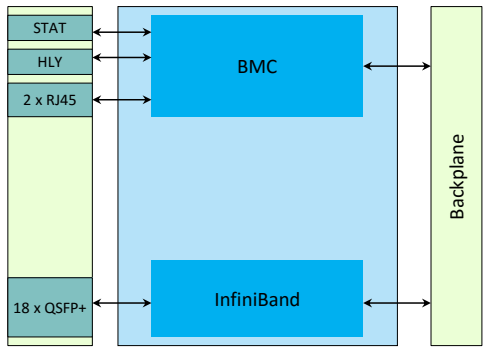
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Interconnection Module: CX611





- Uplink ports
  - 18 x 40/46G IB
  - Downlink ports
  - 16 x 40/56G IB
  - Backplane directly connected links
  - None
  - Applicable to low-latency and high-bandwidth applications, such as HPC clusters and high-performance distributed file systems.
  - Matching mezzanine modules: MZ611

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

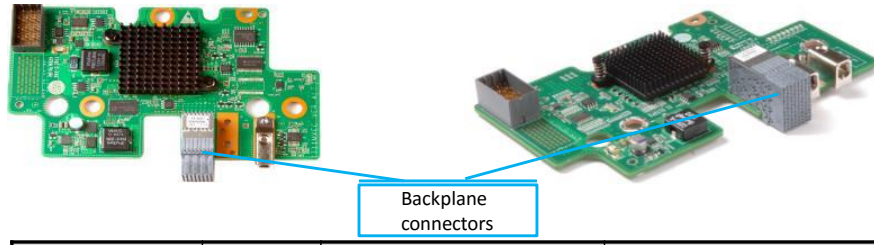
Switch Modules

	Module Type	Uplink	Backplane Directly Connected Link	FC Module Type
CX911	Multi-plane	Ethernet: 16 x 10GE FC: 8 x 8G FC	2 x 40GE	MX510: Qlogic
CX912	Multi-plane	Ethernet: 16 x 10GE FC: 8 x 8G FC	2 x 40GE	MX210: Brocade
CX915	Multi-plane	Ethernet: 12 x GE + 4 x 10GE FC: 8 x 8G FC	None	MX510: Qlogic
CX310	Converged	Ethernet: 16 x 10GE	1 x 40GE	None
CX311	Converged	Ethernet: 16 x 10GE FC: 8 x 8G FC	1 x 40GE	MX510: Qlogic
CX210	FC (8G)	FC: 8 x 8G FC	None	MX210: Brocade
CX110	Ethernet	Ethernet: 12 x GE + 4 x 10GE	2 x 40GE	None
CX111	Ethernet	Ethernet: 12 x GE + 4 x 10GE	None	None
CX611	InfiniBand	InfiniBand: 18 x 40G/56G	None	None

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Mezzanine Modules



Category	Model	Port	Description
GE mezzanine module	MZ110	2 x 2 x GE, PCIe 2.0 x4	
	MZ111	2 x 2 x GE, PCIe 2.0 x2	
CNA mezzanine module	MZ510	2 x 10GE, PCIe 2.0 x8	Support FCoE/iSCSI and multiple PFs, PF bandwidth and VLAN configurations.
	MZ512	2 x 2 x 10GE, 2 x PCIe 2.0 x8	
Multi-plane mezzanine module	MZ910	2 x 10GE + 2 x 8G/10 FCoE, PCIe 3.0 x8	
RoCE NIC	MZ311	2 x 2 x 10GE, 2 x PCIe 3.0 x8	Supports the RoCE feature.
IB mezzanine module	MZ610	2 x QDR 40G, PCIe 3.0 x8	
	MZ611	2 x FDR 56G, PCIe 3.0 x8	

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Mapping Between Mezzanine Modules and Switch Modules

Interconnection Module Category	Interconnection Module	Mezzanine Module	
GE switch module	CX110	MZ110	4 x GE
	CX111	MZ110	4 x GE
Converged switch module	CX310	MZ510	2 x 10GE CNA
		MZ512	2 x (2 x 10GE) CNA
	CX311	MZ510	2 x 10GE CNA
		MZ512	2 x (2 x 10GE) CNA
Multi-plane switch module	CX911	MZ910	2 x 10GE + 2 x 8G FC/10G FCoE
	CX912	MZ910	2 x 10GE + 2 x 8G FC/10G FCoE
	CX915	MZ910	2 x 10GE + 2 x 8G FC/10G FCoE
FC switch module	CX210	MZ910	2 x 10GE + 2 x 8G FC/10G FCoE
InfiniBand switch module	CX611	MZ610	2 x 40G QDR
		MZ611	2 x 56G FDR
Pass-through module	CX116	MZ110	4 x GE
	CX317	MZ510	2 x 10GE CNA
		MZ512	2 x (2 x 10GE) CNA

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

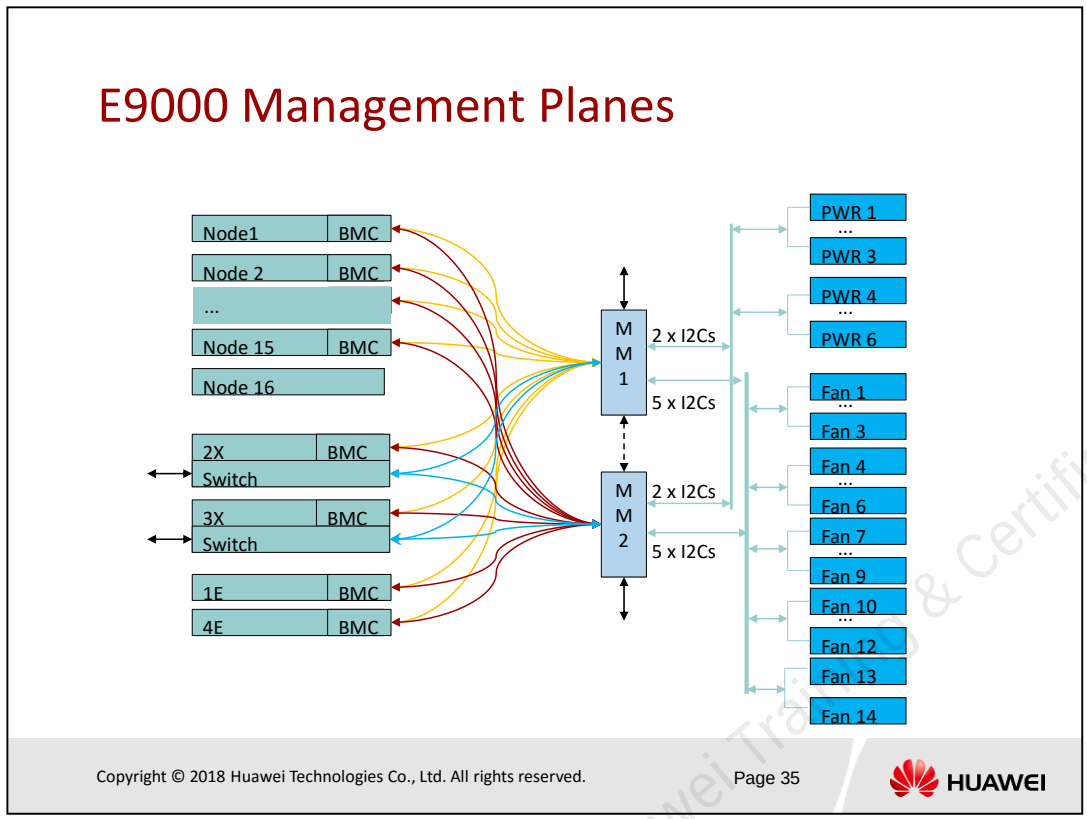
Contents

1. E9000 Positioning and Application Scenarios
2. E9000 Overview
3. **E9000 Deployment and Management**
   - Overall Architecture Management
   - Management Module Deployment and Management
   - Switch Module Deployment and Management
   - Compute Node Deployment and Management

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

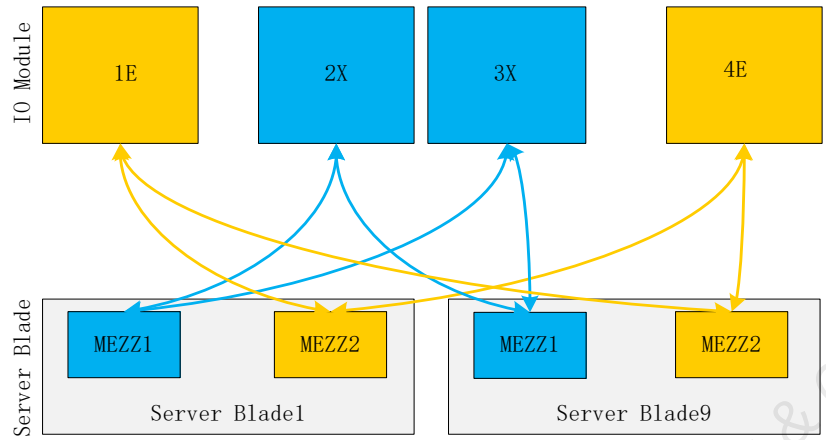
Huawei Training & Certification
Huawei Training & Certification



- Management modules provide centralized, out-of-band device management.
  - Two management modules work in active/standby mode. There are heartbeat message and data synchronization channels between the management modules.
  - Each management module provides I2C buses for fan and PSU management.
  - Each management module provides Ethernet switching for BMC out-of-band management of the front slots and switch module slots.
  - The system provides external management ports on management modules or switch modules.

Huawei Training & Certification

Connections Between Mezzanine Cards (NICs) and Interconnection Modules



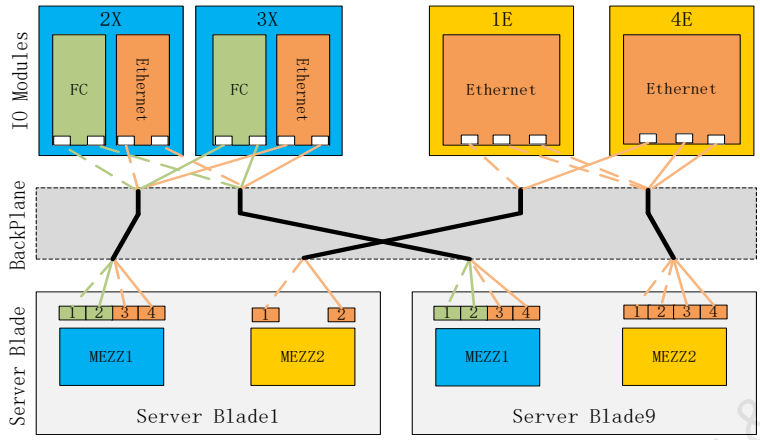
- For a half-width blade, MEZZ1 connects to I/O modules in slots 2X and 3X, and MEZZ2 connects to interconnection modules in slots 1E and 4E.
- For a full-width blade, MEZZ1 and MEZZ3 (optional) connect to interconnection modules in slots 2X and 3X, and MEZZ2 and MEZZ4 connect to interconnection modules in slots 1E and 4E.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Ports Between NICs and Interconnection Modules



- MEZZ1 of Blades 1 to 9 is a FC/Ethernet NIC that connects to both the FC and Ethernet switching planes.
- MEZZ2 of Blade 1 is a two-port NIC. MEZZ2 of Blade9 is a four-port NIC, two ports connecting to the switch module in slot 1E and the other two ports connecting to the switch module in slot 4E.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Mezzanine cards can be classified as:
- (1) FC+Ethernet, Ethernet, and IB mezzanine cards based on the port type
- (2) two-port and four-port mezzanine cards based on the port quantity
- (3) GE/10GE and 40G/56G IB mezzanine cards based on the port rate. For each mezzanine card, a maximum of two ports can be used for connecting to the interconnection modules.

Huawei Training & Certification

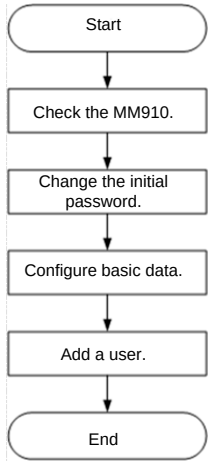
Contents

1. E9000 Positioning and Application Scenarios
2. E9000 Overview
3. **E9000 Deployment and Management**
   - Overall Architecture Management
   - Management Module Deployment and Management
   - Switch Module Deployment and Management
   - Compute Node Deployment and Management

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

MM910 Configuration Process



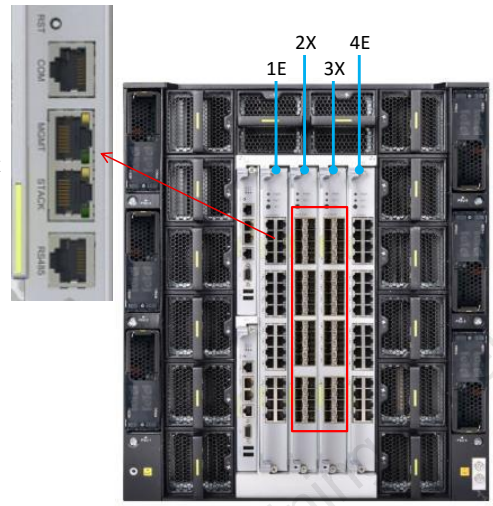
Step	Description
Check the MM910s.	View the indicator on the panel and ensure that the MM910s are working properly. View the MM910 versions and ensure that the versions meet the office requirements. Verify the active/standby switchover function of the MM910s.
Change the initial password.	Change the initial password of the default user root on the active MM910 for improved O&M security.
Configure basic data.	Configure the static IP addresses of the active and standby MM910s. Configure the time zone, Network Time Protocol (NTP), floating IP address, and Simple Network Management Protocol (SNMP) trap receiving address in the active MM910.
Add users.	Add users in the active MM910. The new user accounts can log in to the OS and WebUI of the MM910 for routine maintenance.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

MM910 Configuration Methods

- Ports:
  - Serial port
  - Management network port (MGMT) on the MM910
  - Ports on switch modules in slots 2X and 3X
- Management interfaces:
  - CLI
  - WebUI
- Configuring the initial IP address:
  - Use the serial port.
  - Obtain the IP address automatically (the DHCP server must have been configured).
  - Use the chassis intelligent display.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Default MM910 Settings

Item	Description
Default user name and password	• User name: root • Initial password: Huawei12#$
Static IP addresses and subnet mask of the active and standby MM910s	Static IP address is the permanent IP address of an MM910, mapping to the IP address of eth0 in the ifconfig command output. You can use a PC to log in to the MM910 CLI and WebUI over the static IP address. Default static IP addresses and subnet mask of the two MM910s: • MM910 1: 192.168.2.1 • MM910 2: 192.168.2.2 • Subnet mask: 255.255.255.0
Floating IP address and subnet mask of the MM910s	The IP address used by the active MM910 for communicating with an external network. The IP address maps to the IP address of eth0:0 in the ifconfig command output. The floating IP address can be used to log in to the active MM910 after an active/standby switchover. Default floating IP address and subnet mask of the MM910s: • Floating IP address: 192.168.2.3 (192.1.2.92 for the MM910 earlier than (U54) 2.20) • Subnet mask: 255.255.255.0

- Note: The static and floating IP addresses can be changed on the chassis intelligent display.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Using the MM910 CLI

- Log in to the CLI by using the HyperTerminal and set the following parameters:
  - Bits per second: 115200
  - Data bits: 8
  - Parity: None
  - Stop bits: 1
  - Flow control: None
- Basic commands:
  - smmget (for information query)
  - Syntax: smmget [-h] [-l location] [-t target] -d dataitem
  - smmset (for settings)
  - Syntax: smmset [-h] [-l location] [-t target] -d dataitem -v value

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- For more commands, see the Tecal E9000 MM910 Management Module Command Reference and the Tecal E9000 iMana Command Reference.
- Note:
- A maximum of four users can log in to the MM910 CLI over SSH or Telnet concurrently.
- To log in to the MM910 CLI over the serial port, you must connect the serial ports of the terminal and the MM910 by using a serial cable.
- To log in to the MM910 CLI over the network port, you must connect the network ports of the terminal and the MM910 by using a network cable, and ensure that the IP addresses of the two network ports are in the same network segment.
- Parameters in the commands are described as follows:
  - -h: obtains help information about a command. This parameter cannot be followed by other parameters.
  - -l location: specifies the target component that the command is executed for. If the parameter is not specified, the command is executed for the connected MM910.
  - -t target: specifies the target part on the target component, for example, sensor, indicator, or network port.
  - -d dataitem: specifies the attributes to be queried or set.

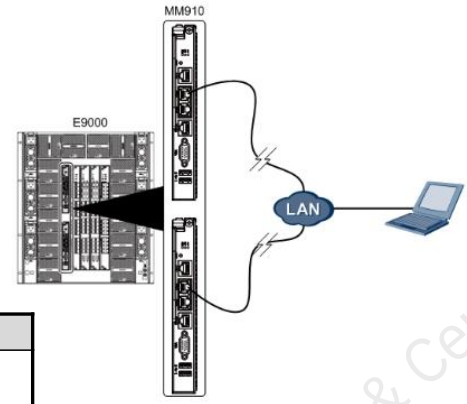
Huawei Training & Certification

- -v value: sets a new value.

Huawei Training & Certification
Huawei Training & Certification

Using the MM910 WebUI

- Log in to the active MM910 WebUI and enter the floating IP address or the static IP address of the active MM910 in the browser. To log in to the standby MM910 WebUI, enter the static IP address of the standby MM910.



Software		Version
OS		Windows XP 32-bit Windows 7 32-bit
Browser	Internet Explorer	IE8.0
	Mozilla Firefox	Mozilla Firefox3.0
JRE		JRE1.6 or later

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- By default, the ports on the switch modules in slots 2X and 3X are used as the management network ports for connecting to an external network. In this situation, do not connect the MGMT port on the MM910 to the same external network at the same time. Otherwise, a network storm will occur, which will result in network interruption. You can run the outportmode command on the CLI to change the management network port to the MGMT port on the MM910.

Huawei Training & Certification

WebUI Login Page (1)

- For MM910 (U54) 2.20 or later



- For MM910 earlier than (U54) 2.20

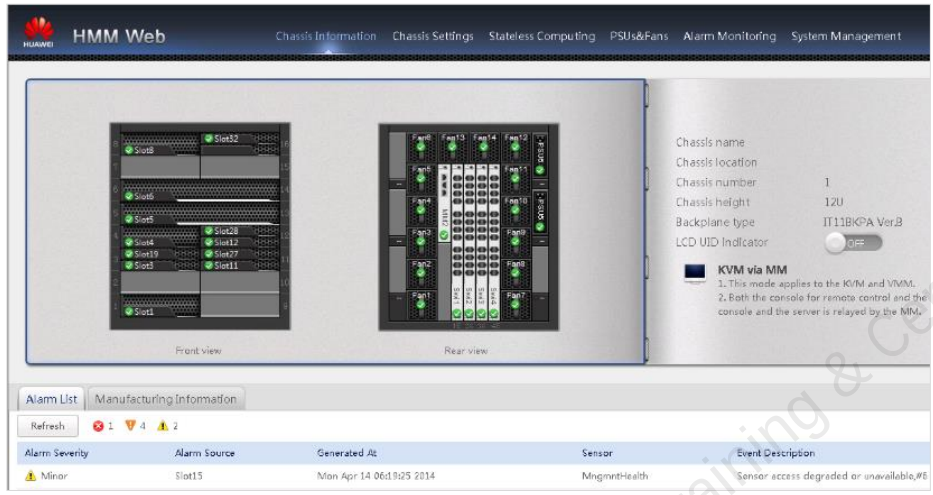


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

WebUI Login Page (2)

- For MM910 (U54) 2.20 or later

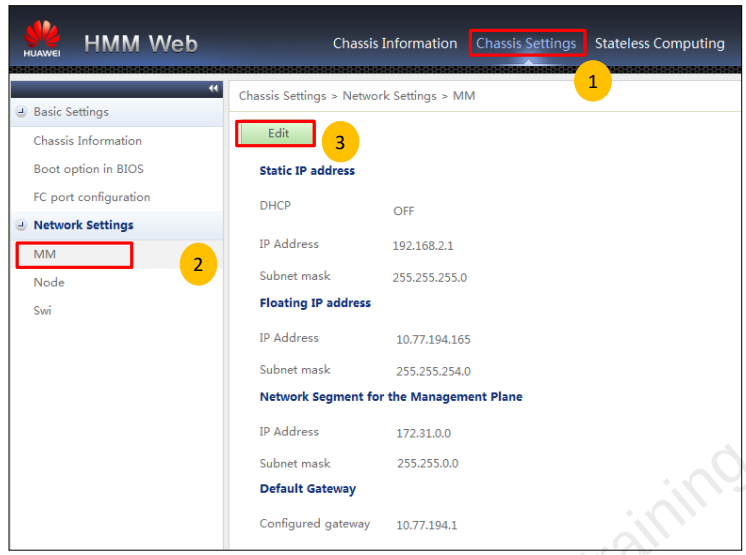


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configuring Basic Data in the WebUI

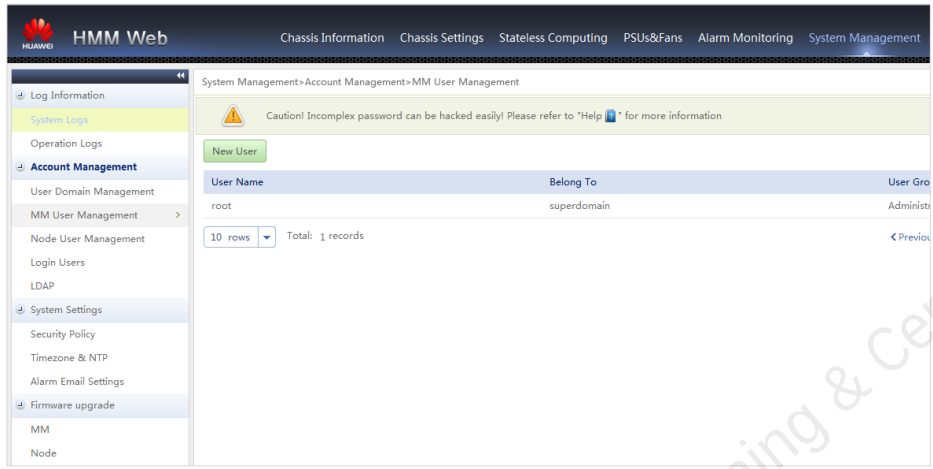


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Adding a User in the WebUI



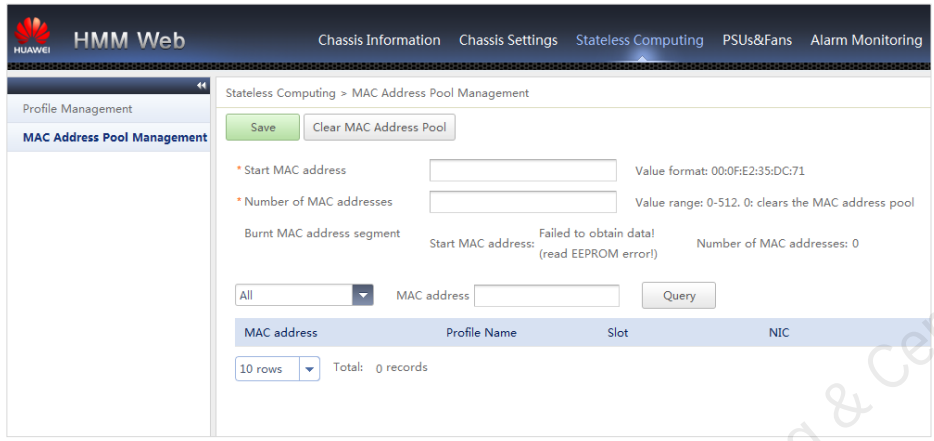
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Roles and rights:
- Administrators: have rights to query and set all information.
- Operators: have rights to query and set information except for performing upgrades (firmwareupdate), managing users (useradd, userdel, userpasswd, userpropertychg), and enabling/disabling password complexity checking (userpasswdcomplex).
- Common users: only have the rights to query information (smmget).

Huawei Training & Certification

Configuring a MAC Address Pool



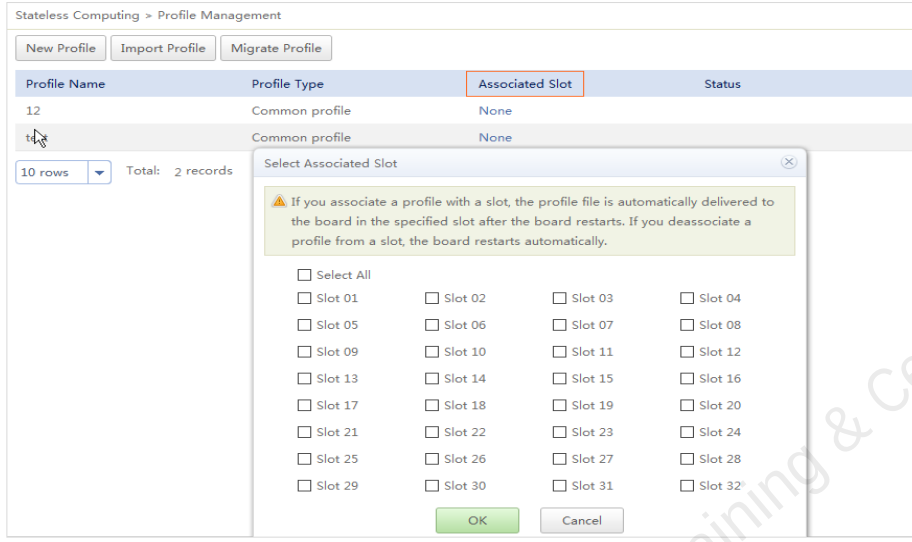
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Configure a resource pool for the unique resources such as MAC addresses, WWNNs, and WWPNs. WWPNs or WWNNs can share a resource pool with MAC addresses. A WWPN or WWNN is a MAC address with the prefix 1000 or 2000 respectively.
- MAC address assignment policies:
  - You can configure consecutive MAC address segments.
  - The MAC address segment 0x78D752CB0000 to 0x78D752CBFFFF is available for use.
  - Select different MAC address segments for different chassis to prevent MAC address conflicts.

Huawei Training & Certification

Configuring Binding



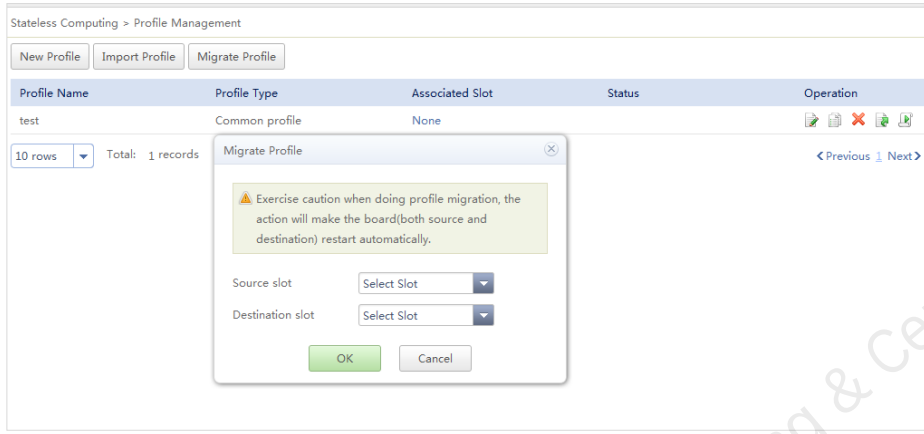
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- After binding profiles to slots, the profiles will be automatically delivered to the inserted compute nodes. The system will generate an alarm if the profile does not match the hardware of the inserted compute node. For example, the profile does not match the NIC or the configured mezz module slot for a NIC does not exist.

Huawei Training & Certification

Configuring Migration

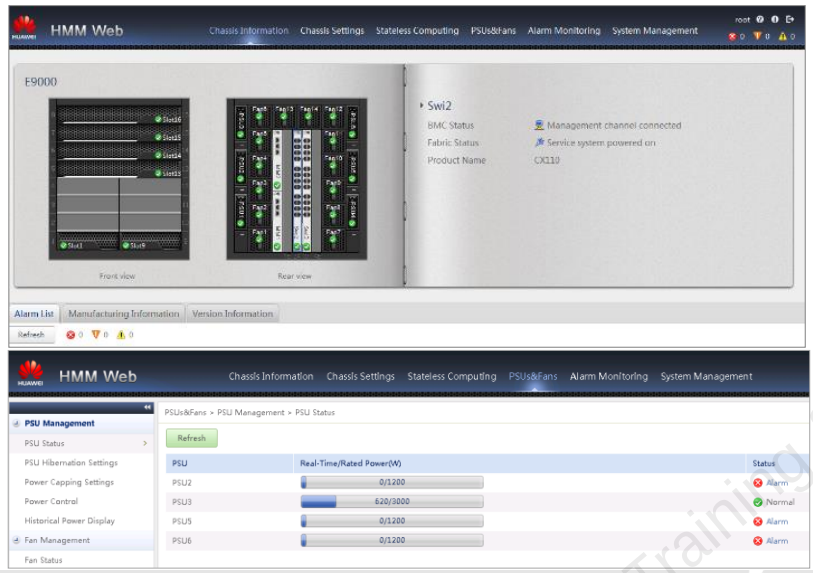


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Migrate configurations among slots. After the migration, the configurations are deleted from the source slot and take effect in the destination slot.

Huawei Training & Certification

PSU Management



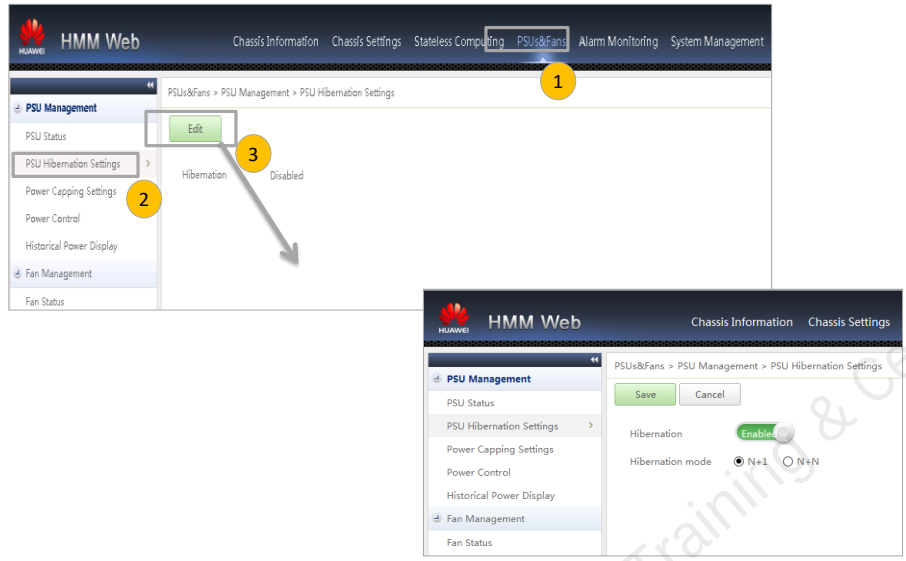
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- 3000 W AC, 2000 W AC, and 2500 W AC PSUs can monitor their own input and output and report faults to the management module.
- On HMM Web, you can view the status (health, alarms, and hybernation), rated power, and current power of the six PSUs on the PSU Status page.
- The management module checks the power requirements of the chassis every five seconds and calculates the total power of all PSUs in the chassis. When a power exception occurs and PSUs cannot supply enough power for the chassis, the chassis power exception mode is triggered, and compute nodes enter the low power consumption mode. After the power exception is solved, the system recovers automatically.

Huawei Training & Certification

PSU Hibernation



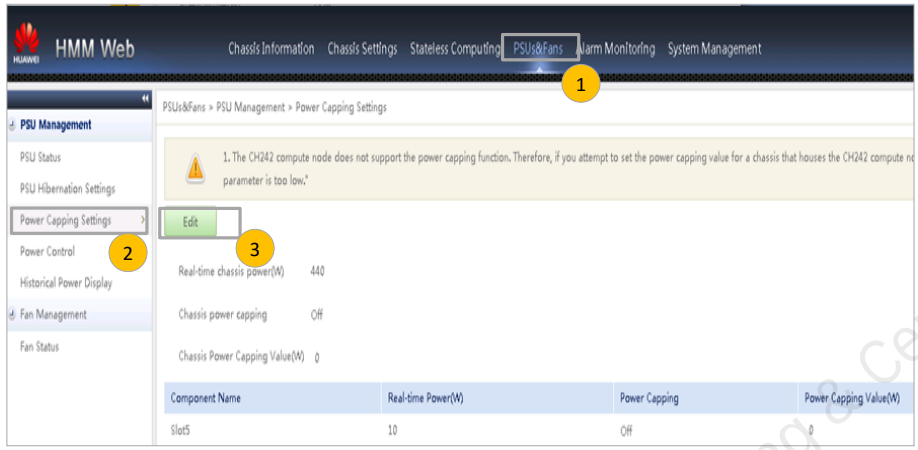
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- When the chassis power is low, to prevent multiple redundant PSUs from wasting power, the BMC can hibernate the extra PSUs to save energy based on the PSU configuration and current chassis power. 2000 W, 2500 W, and 3000 W PSUs support the hybernation feature under an appropriate voltage.
- The feature automatically sets redundant PSUs to the hibernation mode. When the chassis power increases due to a newly-installed component or a PSU exception occurs, the management module terminates the hibernation and wakes up all hibernated PSUs. When the chassis power increases or decreases, the management module compares the sum of rated power of all operating PSUs with the total power consumption of the entire chassis and determines the number of PSUs to hibernate based on the hibernation mode.

Huawei Training & Certification

Power Capping Settings



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification

- The power capping feature limits the total power for the entire chassis and improves the compute node density in the equipment room. The management module can set the maximum power for the entire chassis or for a single compute node.
- Maximum power of a compute node = (Maximum power of the entire chassis – Operating power of rear modules – Operating power of fan modules and PSUs – Maximum power preset for some compute nodes)/(Total number of compute nodes in the chassis – Number of compute nodes whose maximum power has been preset)
- The maximum power of a compute node is updated every five seconds. If the power changes, a new value for the maximum power per compute node is set. The capping value may take effect three seconds after setting. Therefore, the latency can be eight seconds. The previous power capping value is considered in power adjustment. If no compute node is added, the power deviation is about 10%.
- Power capping: The Intel management engine (ME) adjusts the power to approach the maximum value by frequency modulation and voltage regulation (during the adjustment, the iMana on a compute node periodically sends the current power information of the compute node to the Intel ME).
- The management module can set the maximum power for the entire chassis or for a single compute node.
- Example:

Contents

1. E9000 Positioning and Application Scenarios
2. E9000 Overview
3. **E9000 Deployment and Management**
   - Overall Architecture Management
   - Management Module Deployment and Management
   - Switch Module Deployment and Management
   - Compute Node Deployment and Management

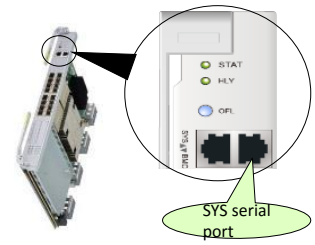
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

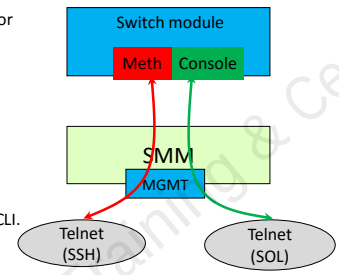
HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Switch Module Login Methods

- You can log in to a switch module in any of the following ways:
- 1. Serial port (speed: 115200 bit/s)
- The switch panel provides two serial ports.
- SYS serial port: used to connect to the Ethernet/FC switching plane.
- BMC serial port: used to log in to the BMC of the switch module.
- You can use the BMC to select the connected switching plane (Ethernet or FC) for the SYS serial port.
- 2. Management network port
- Set the IP address of the management network port through the MM910 or the serial port of the switch module.
- Log in to the management network port over SSH or Telnet on the MGMT port of the MM910.
- 3. SOL
- On the MGMT port of the MM910, run telnet MM910 floating IP address 1101 in the MM910 CLI and select a serial port to log in to the serial port CLI.





Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- You can log in to a switch module in any of the following ways:

1. Serial port (speed: 115200 bit/s)

- The switch panel provides two serial ports.
- SYS serial port: used to connect to the Ethernet/FC switching plane.
- BMC serial port: used to log in to the BMC of the switch module.
- You can use the BMC to select the connected switching plane (Ethernet or FC) for the SYS serial port.

2. Management network port

- Set the IP address of the management network port through the MM910 or the serial port of the switch module.
- Log in to the management network port over SSH or Telnet on the MGMT port of the MM910.

3. SOL

- On the MGMT port of the MM910, run **telnet *MM910 floating IP address* 1101** in the MM910 CLI and select a serial port to log in to the serial port CLI.

Huawei Training & Certification

Enabling SOL for the MM910

- **Step 1**: Log in to the MM910 CLI, and query whether SOL is enabled for the MM910 (the SOL function is disabled by default).
  - **:smmget -l smm -d solenable**
  - If "SOL is enable." is displayed, you can log in over SOL directly.
  - If "SOL is disable." is displayed, perform step 2.
- **Step 2**: Enable SOL for the MM910.
  - **smmset -l smm -d solenable -v enable**
  - Information similar to the following is displayed:
  - Success
- **Step 3**: Restart the MM910 after the SOL function is changed from disabled to enabled. (If the SOL function is already enabled, the restart is not required.)
  - **reboot**
  - The following information is displayed:
  - This command will reboot SMM board. Continue? [Y/N]:
- **Step 4**: Type **Y** and press **Enter**.
  - The active MM910 restarts, and an active/standby failover occurs. It takes about 2 minutes for the original active MM910 to restart.
- **Step 5**: Log in to the CLI of the original standby MM910.
- **Step 6**: Perform steps 3 to 6 to restart the original standby MM910.
- Note: After using the SOL function, run the **smmset -l smm -d solenable -v disable** command to disable SOL, and restart the MM910 to ensure system security.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Using the SOL Function of the MM910 (1)

- Step 1 Log in to the MM910 over SOL.
- In the MM910 CLI, run the following command:
- telnet MM910 floating IP address 1101
- The following screen is displayed:

```
==================================================================
            Welcome to SMM SOL Server
     Please log in with SMM account and password.
==================================================================
user name:
```

- Step 2 Enter the user name (root) and password (default: Huawei12#$) to log in to the MM910.
- The screen for selecting a slot is displayed.

```
Log in Success!
================================================================================
   please input the SOL Blade1~Blade16(1 ~ 16), Swi1~Swi4(33 ~ 36) and COM#(n)
   press Ctrl+R to return
================================================================================
Blade1~Blade16(1 ~ 16)
Swi1~Swi4(33 ~ 36)
Please input your choice:
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Using the SOL Function of the MM910 (2)

- **Step 3** Type the slot number of the compute node or switch module to be connected, for example, 34, and press Enter.
- All serial ports on the switch module to be connected are displayed as follows:

- **1 BMCcom** # BMC serial port
- **2 fabriccom** # Serial port on the 10GE switching plane of the CX9XX/CX3XX switch module or on the GE switching plane of the CX11X switch module
- **3 baseccom** # Serial port on the GE switching plane of the CX9XX switch module
- **4 Fccom** # Serial port on the FC switching plane of the CX911/CX311 switch module
- **COM#(1 ~ 3)**

- Type a number corresponding to the required serial port, and press **Ctrl+R** to return to the screen for selecting the chassis slot.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Default User Names and Passwords for a Switch Module

User Name	Login Method	Default Password	User Type	User Status
NA	No user name is required for logging in to the BMC of the switch module.	Huawei12#$	Administrative User	Enabled
NA	Log in to the 10GE switching plane over the SYS serial port or over SOL.	Huawei12#$	Administrative User	Enabled
root	Log in to the 10GE switching plane over SSH.	Huawei12#$	Administrative User	Enabled
admin	Log in to the FC switching plane of the CX311/CX911 switch module over the SYS serial port, SOL, or SSH.	Huawei12#$	Administrative User	Enabled

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Configuring the Meth IP Address of a Switch Module Through the CLI (1)

- Log in to the MM910 CLI, and set the IP address of the management network port on a switching plane:
- **smmset -l swiN:fruM -d swipcontrol -v ipaddress maskaddress**
- Parameters description:
- N indicates the slot number of the switch module. It ranges from 1 to 4, mapping to slots 1E, 2X, 3X, and 4E on the panel respectively.
- M indicates the ID of a switching plane. The value 1 indicates the onboard GE switching plane, 2 indicates the 10GE switching plane, and 3 indicates the FC switching plane.
- For example, to set the IP address of the management network port on the 10GE switching plane of the switch module in slot 2X to 192.168.102.105, and the subnet mask to 255.255.255.0, run the following command:
- **root@SMM:/#smmset -l swi2:fru2 -d swipcontrol -v 192.168.102.105 255.255.255.0**
- If the following information is displayed, the setting is successful:
- Success

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configuring the Meth IP Address of a Switch Module Through the CLI (2)

- Query the IP address of the management network port to verify the setting:
- **smmget -l swiN:fruM -d swipcontrol**
- Parameters description:
- N indicates the slot number of the switch module. It ranges from 1 to 4, mapping to slots 1E, 2X, 3X, and 4E on the panel respectively.
- M indicates the ID of a switching plane. The value 1 indicates the onboard GE switching plane, 2 indicates the 10GE switching plane, and 3 indicates the FC switching plane.
- For example, to query the IP address of the management network port on the 10GE switching plane of the switch module in slot 2X, run the following command:
- **root@SMM:/#smmget -l swi2:fru2 -d swipcontrol**
- The command output is as follows:
- Ip address :192.168.102.105
- Mask :255.255.255.0

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Configuring the Meth IP Address of a Switch Module Through the WebUI



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

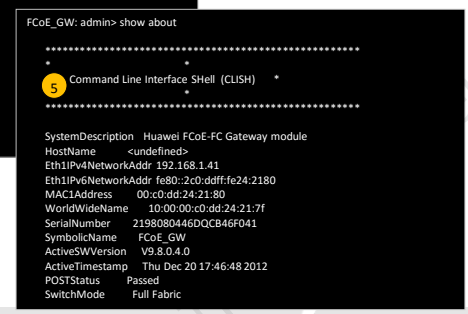
Huawei Training & Certification
Huawei Training & Certification

Logging In to a Switch Module over SOL



1. Log in to the MM910 CLI and run the **telnet 0.0.0.0 1101** command. Alternatively, run the **telnet MM910 floating IP address 1101** command directly in a telnet tool.
2. Enter the user name and password of the MM910.

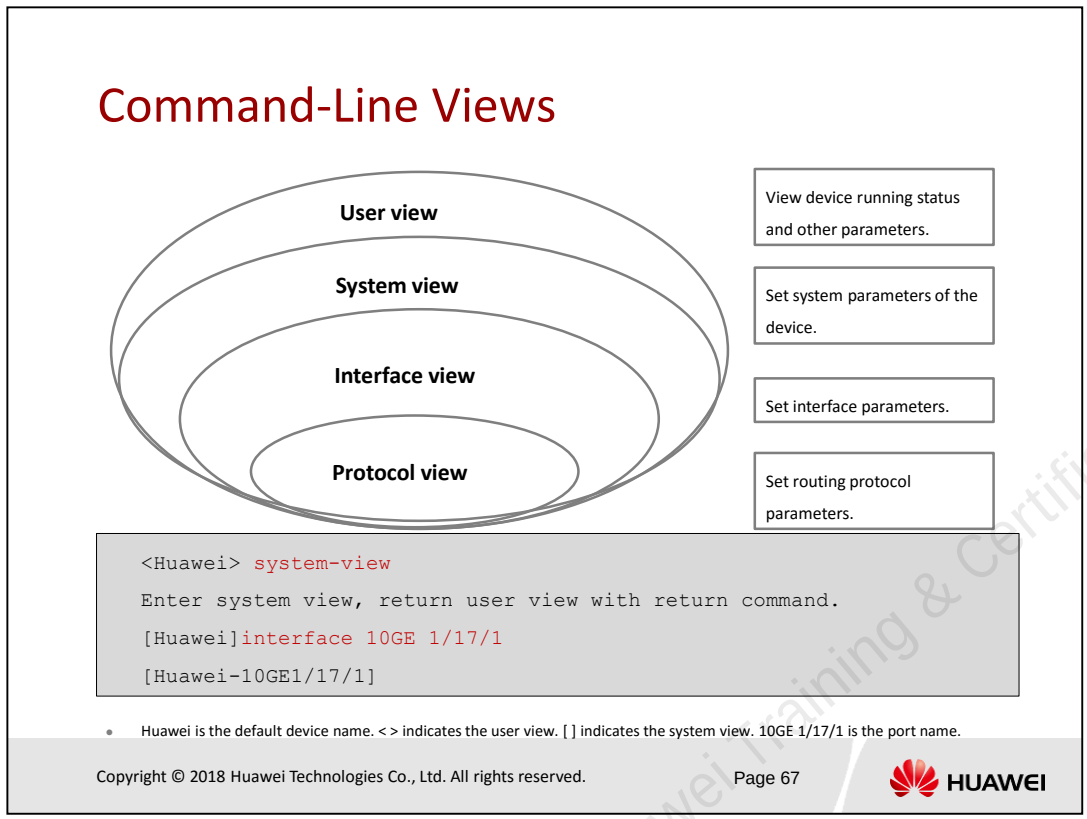
3. Select the slot number of the switch module to be connected to. (33-36 map to slots 1E, 2X, 3X, and 4E respectively.)
4. Select the serial port for connecting to the switch module.
5. Enter the password to open the serial port CLI of the switch module.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

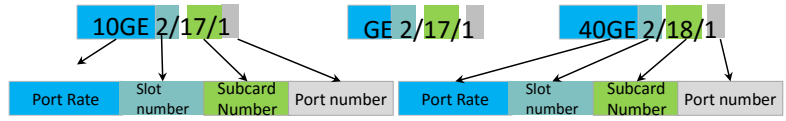
Huawei Training & Certification



- VRP's structured commands-line views define the specific view for executing each command line. This example shows the commonly used views. Each command belongs to one or more command-line views and can be executed only in the designated views. The user view shows first after you enter the VRP configuration interface. In the user view, you can view running status and statistics of the device.
- To modify system parameters, you need to enter the system view. You can enter other configuration views, such as the port view and the protocol view, from the system view.
- The prompt < > indicates the user view and [ ] indicates the system view.

Huawei Training & Certification

Ethernet Port Numbering



Function	Subcard Number	Switch Model	Port Rate	Port Type	Port Number	Example
Ethernet ports used for connecting to the mezzanine cards of compute nodes	1-8	CX9XX	10GE	NA	1-3	10GE 2/1/1
	9-16		10GE		1-2	
	1-6	CX3XX	10GE		1-2	
		CX110	GE		1-2	GE 2/1/1
Panel ports used for cascading chassis or connecting to uplink switches	17	CX9XX	10GE	SFP+	1-16	10GE 2/17/1
		CX311				
		CX110	GE	RJ45	1-12	GE 2/17/12
Ports for connecting switch modules, used as a stack port generally	18	CX9XX	40GE	NA	1-2	40GE 2/18/2
		CX311			1	
		CX110	40GE		1-2	
Ports used for connecting to the management modules	19	CX9XX	NA	NA	NA	NA
		CX3XX	GE	NA	1-2	GE 2/19/1
		CX110	GE	NA	1-2	
Extension ports of switch modules	20	CX9XX	NA	NA	NA	NA
		CX3XX	NA	NA	NA	
		CX110	10GE	SFP+	1~4	10GE 2/20/1

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Port rate: GE/10GE/40GE. Slot number: stack ID. By default, it is the slot number of the switch module: 1, 2, 3, and 4. You can run the stack command to change the slot number.
- The CX9XX uses the GE switching plane (rather than 10GE switching plane) to connect to the management modules. The GE switching plane does not connect to compute nodes.

Huawei Training & Certification

CLI Keyboard Shortcuts

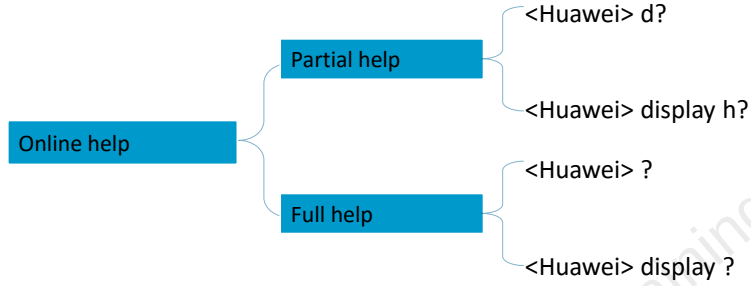
Common keys	Inserts a character at the current location of the cursor if the editing buffer is not full, and moves the cursor to the right.
Backspace	Deletes the character before the cursor and moves the cursor to the left. If there is no character before the cursor, the system just does not respond.
↑ or <Ctrl P>	Obtains the previous history command, if any.
↓ or <Ctrl N>	Obtains the next history command, if any. Otherwise, clear the command.
Tab	Completes a partial command line.
<Ctrl C>	Stops displaying information and running commands.
Space	Displays the next additional screen.
Enter or Return	Displays the next line or executes the latest command.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Obtaining the Command Help





Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Immediate Execution or Execution After Commit

```
<Huawei> system-view immediately
[Huawei] interface 10GE 2/17/1
[Huawei-10GE2/17/1] port link-type trunk
[Huawei] port trunk allow-pass vlan 100
```

Commands take effect immediately after being entered

```
<Huawei> system-view
[~Huawei] interface 10GE 2/17/1
[*Huawei] port link-type trunk
[*Huawei] port trunk allow-pass vlan 100
[*Huawei] commit
```

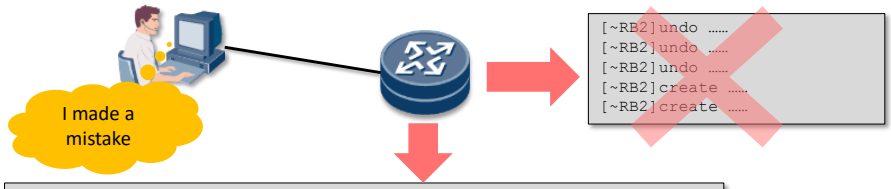
Commands take effect after the commit command is run.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Rolling Back Configuration




```
[~RB2]undo ……
[~RB2]undo ……
[~RB2]undo ……
[~RB2]create ……
[~RB2]create ……
```

```
<RB2>display configuration commit list
------------------------------------------------------------------------
No.  CommitId    Label        User            TimeStamp
------------------------------------------------------------------------
1    1000002222  -              huawei        2012-10-16 16:10:37
2    1000002221  -              huawei        2012-10-16 15:57:58
……
50   1000002173  -             anonymous  2012-10-08 15:39:56
------------------------------------------------------------------------
```

```
<RB2>rollback configuration to commit-id 1000002220
Warning: This operation will revert configuration
changes to the previous status. Continue? [Y/N]:y
Loading rollback changes
Committing
Check rollback result
Configuration rollback succeeded.
```

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Saving Configuration

```
<HUAWEI>save
Warning: The current configuration will be written to the device.
Continue? [Y/N]:Y
Now saving the current configuration to the slot 4
Info: Save the configuration successfully.
slot 4: generate local config file successfully.
<HUAWEI>
```

- Run the save command to save all configuration of a switch module to the MM910. (After the switch module is restarted, it obtains the configuration from the MM910.)
- The configuration is also saved in the cfg_local file in the switch module.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI
Huawei Training & Certification
Huawei Training & Certification

Setting the IP Address of the Management Network Port

```
<Huawei> system-view
[~Huawei] interface Meth 0/0/0
[~Huawei] ip address 192.168.100.101 255.255.255.0
[*Huawei] commit
```

- In the HMM WebUI (Chassis Settings > Network Settings > Swi), set the IP address of the management port.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Contents

1. E9000 Positioning and Application Scenarios
2. E9000 Overview
3. **E9000 Deployment and Management**
   - Overall Architecture Management
   - Management Module Deployment and Management
   - Switch Module Deployment and Management
   - Compute Node Deployment and Management

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Compute Node Configuration

- Compute nodes are installed in slots 1 to 16 on the front panel of the E9000 chassis.
- Compute node configuration operations include:
  - Configuring the BMC
  - Configuring the RAID
  - Configuring the BIOS
  - Installing the OS









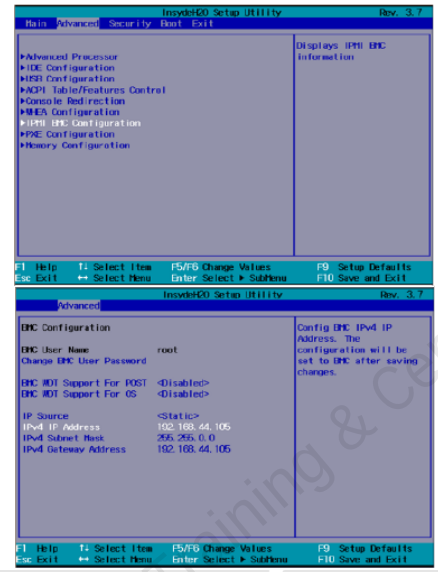
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Huawei Training & Certification

Configuring the BMC

- Using the MM910 CLI
- Run the following command:
- smmset -l bladeN -t BMC -d bmcip -v ip mask
- N: indicates the compute node slot number.
- Using the BMC serial port
- 1. Log in to the MM910 CLI and switch the serial
- port of a compute node to the BMC.
- For example, run the following command for the
- compute node in slot 16:
- smmset -l blade16 -d switchcom -v 4
- 2. Set the IP address of the management network
- port:
- ipmcset -t Network port number -d ipaddr -v BMC IP Subnet mask



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- A compute node is managed by the local BMC.
- You can log in to the BMC over the network port or serial port of the compute node.
- Before logging in to the BMC over the network port, set the IP address of the port in any of the following ways:
  - Use the MM910 CLI.
  - Use the local KVM and BIOS.
  - Use the BMC logged in to over serial port.

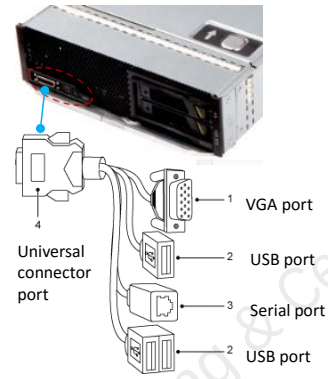
Huawei Training & Certification

Configuring the RAID and BIOS

- Configure the RAID and BIOS in any of the following ways:
  - Local KVM (through the universal connector port)
  - Virtual KVMCentralized KVM (MM910 WebUI)
  - Remote KVM of the MM910



Remote KVM of the MM910



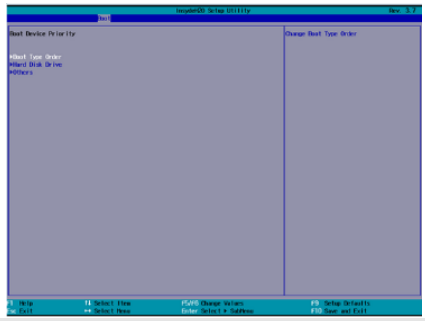
Local KVM (through the universal connector port)

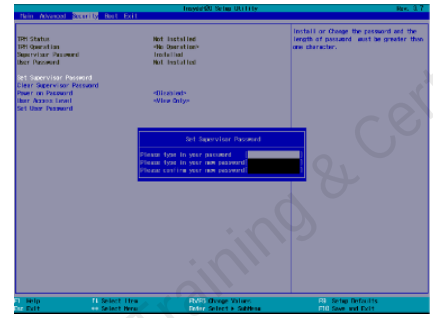
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

BIOS Configuration Procedure

- Procedure:
- Restart the compute node.
- Press Delete on the startup screen (on the remote KVM or local KVM).
- Enter the user name and password to enter the BIOS screen.
- Set the server boot mode: hard disk drive/PXE/CD-DVD ROM drive
- Change the BIOS password.



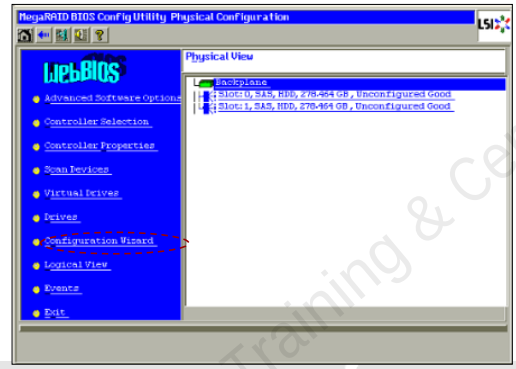


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

RAID Configuration Procedure

- Task: Configure RAID 0 or RAID 1 for the hard disks before the OS startup.
- Procedure:
- Restart the compute node.
- As soon as the following message is displayed, press Ctrl+H on the remote or local KVM.
- "Press <Ctrl><H> for WebBIOS or press <Ctrl><Y> for Preboot CLI"
- Configure the RAID as prompted.
- Initialize the virtual disks.



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

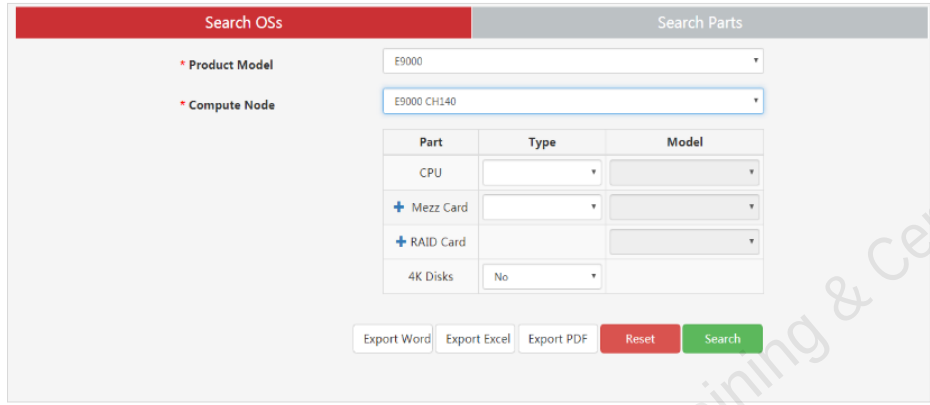
HUAWEI

- Note: The procedure on this slide applies to the LSI SAS2208 controller card. For the LSI SAS2308 controller card, press Ctrl+C.

Huawei Training & Certification

Preparing for OS Installation (1)

- Determine the OS to be installed based on your application requirements using the server compatibility checker.

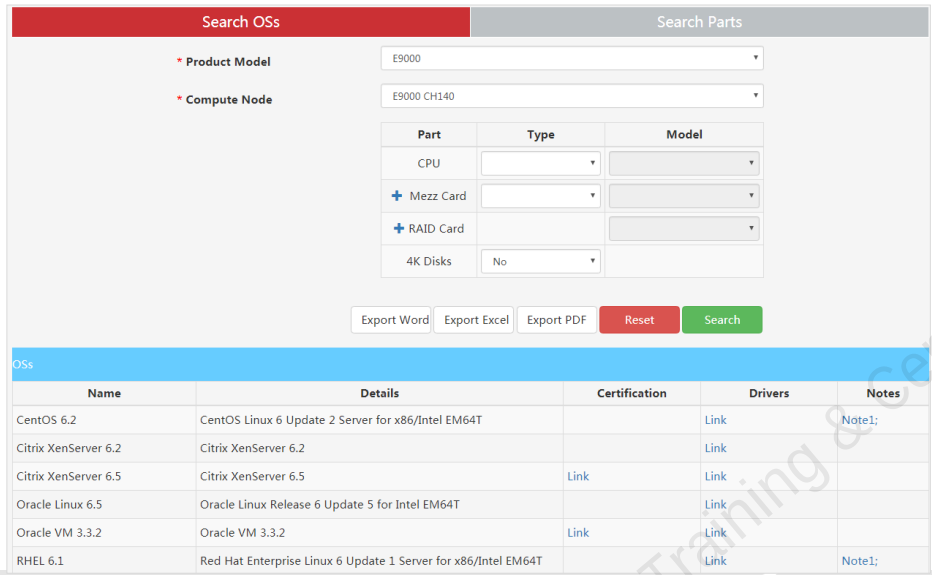


Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

- Visit http://support.huawei.com/onlinetoolsweb/ftca/en.

Huawei Training & Certification

Preparing for OS Installation (2)



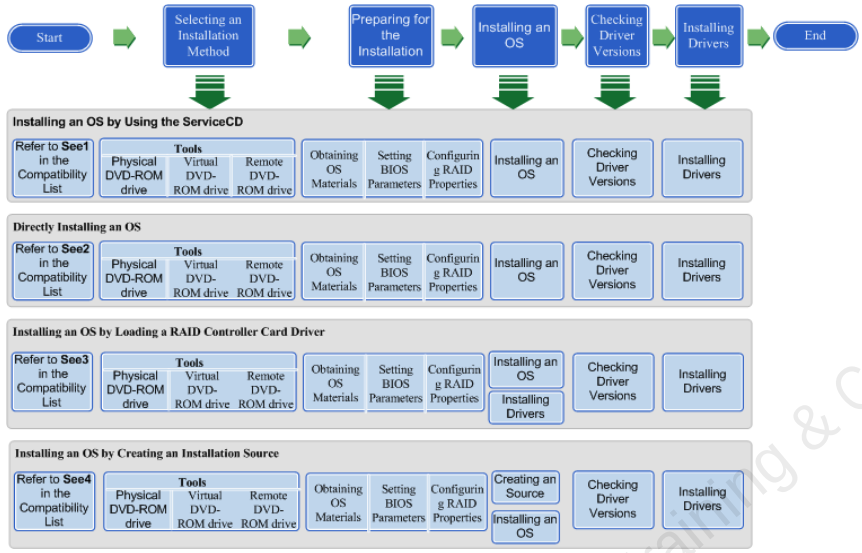
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Select the product model, and component type and model in the compatibility checker to obtain compatibility information shown on this slide.

Huawei Training & Certification

OS Installation Methods



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- You can install an OS on a Tecal server by using one of the following methods:
  - Installing an OS by using the ServiceCD

As an OS installation wizard tool, the ServiceCD simplifies the installation of OSs and drivers on a local computer. If you choose this method, you can use a physical CD/DVD-ROM drive to read content from the ServiceCD or use a virtual CD/DVD-ROM drive to load the ServiceCD image file.

  - Directly installing an OS by using an OS DVD or ISO file

It is a common method to install an OS by using the OS installation DVD or image file. If you choose this method, you can use a physical CD/DVD-ROM drive to read content from the installation DVD or use a virtual CD/DVD-ROM drive to load the image file.

  - Installing an OS by loading a RAID controller card driver

If you want to install an OS on a Tecal server by loading a RAID controller card driver, you need to obtain the OS installation DVD or image file, and use the virtual CD/DVD-ROM drive or FDD in iMana delivered with the Tecal server to load the OS installation DVD or image file.

  - Installing an OS by creating an installation source

You need to compile an installation file for a special OS to incorporate the required driver into the ISO file. If you choose this method, you need to use a virtual CD/DVD-ROM drive to load the image file.

Huawei Training & Certification

Methods for Obtaining OS Installation Resources

Installation Method	Software	How to Obtain
Installing an OS by using the ServiceCD	ServiceCD DVD or ISO file	The ServiceCD DVD is delivered with the product. Obtain the Service CD ISO file as follows: Visit http://support.huawei.com/enterprise/en and choose Product Support > Software Download > Server > FusionServer Tools. Select the latest version and download the software whose name contains ServiceCD.
	OS installation DVD or ISO file	The OS installation DVD or ISO file is provided by Huawei or the customer, depending on the contract.
Directly installing an OS	OS installation DVD or ISO file	The OS installation DVD or ISO file is provided by Huawei or the customer, depending on the contract.
	Server driver package	Visit http://support.huawei.com/enterprise/en and choose Product Support > Software Download > Server. Select the target server and download the software whose name contains Driver.
Installing an OS by loading a RAID controller card driver	OS installation DVD or ISO file	The OS installation DVD or ISO file is provided by Huawei or the customer, depending on the contract.
	Server driver package	Visit http://support.huawei.com/enterprise/en and choose Product Support > Software Download > Server. Select the target server and download the software whose name contains Driver.
Installing an OS by creating an installation source	OS installation DVD or ISO file	The OS installation DVD or ISO file is provided by Huawei or the customer, depending on the contract.
	Server driver package	Visit http://support.huawei.com/enterprise/en and choose Product Support > Software Download > Server. Select the target server and download the software whose name contains Driver.

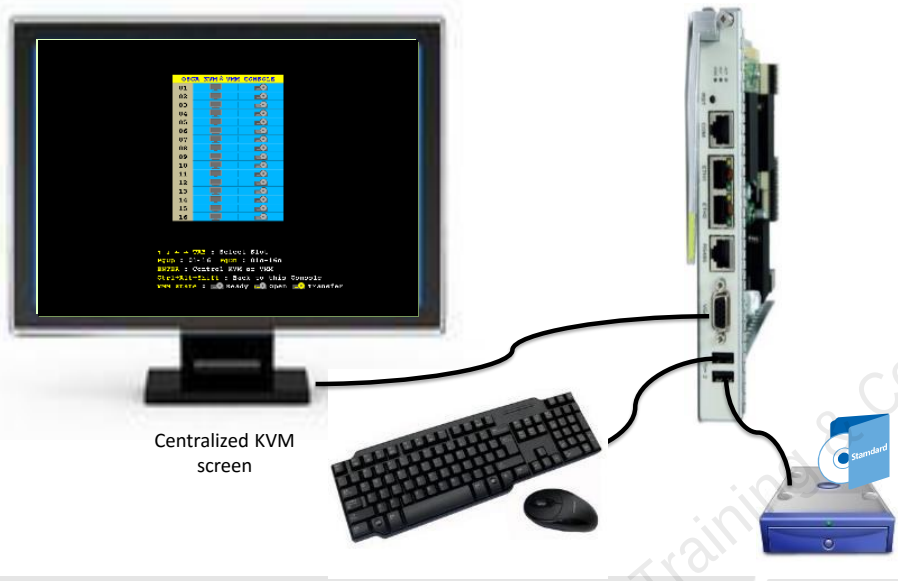
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- You can log in to a server by using a PC or KVM.
- If you log in to the Virtual Console on a server by using a PC, you can use the physical CD/DVD-ROM drive and virtual media (including the virtual CD/DVD-ROM drive and virtual floppy disk drive) on the Virtual Console. This applies to the four installation methods.
- If you log in to a server by using a KVM, you can use only the physical CD/DVD-ROM drive.

Huawei Training & Certification

Installation Media - Integrated KVM/Virtual Media for OS Installation



Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

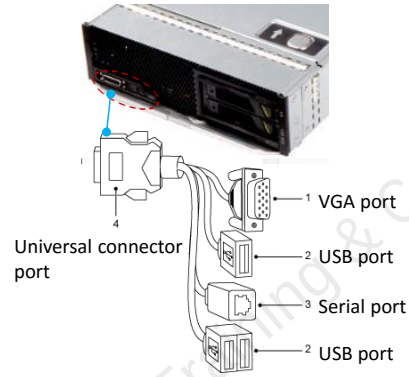
- Compute nodes can be managed by the integrated KVM of the MM910.
  - Connect a monitor to the VGA port on the MM910.
  - Connect a CD/DVD-ROM drive, mouse, and keyboard to the USB ports.
  - Select a compute node using the integrated KVM.
  - Select the compute node connected to the virtual media.
  - Reset the compute node, and select DVD (created by the virtual media) for booting in the BIOS.

Huawei Training & Certification

Installation Media - Local CD/DVD for OS Installation

Use a local CD/DVD to install an OS:

- Connect the universal connector port to the port on a compute node.
- Connect a monitor to the VGA port.
- Connect a mouse, keyboard, USB drive to the USB ports.
- Enter the BIOS and set the boot mode to DVD.
- Use the OS disk or ServiceCD for installation.



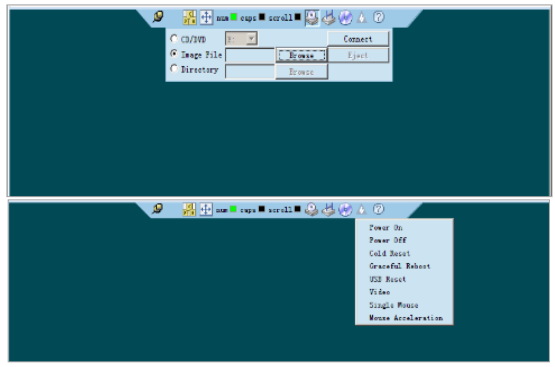
Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

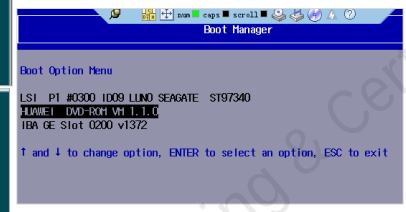
Huawei Training & Certification

Installation Media - Image Installation

Install an OS by using an image file:

- Log in to the local BMC or MM910 over web.
- Select the remote KVM and load the OS image file (.iso) by using the virtual CD/DVD-ROM drive on the UI.
- Restart the compute node, and select **HUAWEI DVD-ROM VM 1.1.0** during startup.





Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Training & Certification

Quiz

1. (Single-Answer) How many half-width service boards can be installed in the front slots of an E9000 chassis?
2. A. 8 B. 16 C. 24 D. 32
3. (Multiple-Answer) How can an OS be installed on a compute node?
4. A. Using a local CD/DVD
5. B. Using an image file
6. C. Using the PXE
7. D. Using integrated virtual media

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Summary

- E9000 Positioning and Application Scenarios
- E9000 Overview
- E9000 Deployment and Management

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

More Information

- Huawei enterprise technical support
  - http://support.huawei.com/enterprise/en/index.html
  - You can obtain product documentation, compatibility information, product software, and maintenance services.
- Huawei server information self-service platform
  - http://enterprise.huawei.com/topic/Self_service_server_en/knowledge.html
  - You can obtain Huawei server installation, commissioning, and O&M information and tools.

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Recommendations

- Huawei Learning Website
  - http://support.huawei.com/learning/Index!toTrainIndex
- Huawei Support Knowledge Base
  - http://support.huawei.com/enterprise/servicecenter?lang=en

Copyright © 2018 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Huawei Training & Certification
Huawei Training & Certification

Thank You

www.huawei.com

Huawei Training & Certification
Huawei Training & Certification

Recommendations

- Huawei Learning Website
  - http://learning.huawei.com/en
- Huawei e-Learning
  - https://ilearningx.huawei.com/portal/#/portal/ebg/51
- Huawei Certification
  - http://support.huawei.com/learning/NavigationAction!createNavi?navId=_31&lang=en
- Find Training
  - http://support.huawei.com/learning/NavigationAction!createNavi?navId=_trainingsearch&lang=en

More Information

- Huawei learning APP

Huawei Training & Certification

版权所有© 2018 华为技术有限公司