Implementing Cisco Data Center Unified Fabric

Volume 1

Version 5.0

https://t.me/learningnets

DCUFI

Implementing Cisco Data Center Unified Fabric

Volume 1

Version 5.0

Student Guide

Text Part Number: 97-3211-01

https://t.me/learningnets

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

DISCLAIMER WARRANTY: THIS CONTENT IS BEING PROVIDED “AS IS.” CISCO MAKES AND YOU RECEIVE NO WARRANTIES IN CONNECTION WITH THE CONTENT PROVIDED HEREUNDER, EXPRESS, IMPLIED, STATUTORY OR IN ANY OTHER PROVISION OF THIS CONTENT OR COMMUNICATION BETWEEN CISCO AND YOU. CISCO SPECIFICALLY DISCLAIMS ALL IMPLIED WARRANTIES, INCLUDING WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT AND FITNESS FOR A PARTICULAR PURPOSE, OR ARISING FROM A COURSE OF DEALING, USAGE OR TRADE PRACTICE. This learning product may contain early release content, and while Cisco believes it to be accurate, it falls subject to the disclaimer above.

© 2012 Cisco and/or its affiliates. All rights reserved.

https://t.me/learningnets

Students, this letter describes important course evaluation access information!

Welcome to Cisco Systems Learning. Through the Cisco Learning Partner Program, Cisco Systems is committed to bringing you the highest-quality training in the industry. Cisco learning products are designed to advance your professional goals and give you the expertise you need to build and maintain strategic networks.

Cisco relies on customer feedback to guide business decisions; therefore, your valuable input will help shape future Cisco course curricula, products, and training offerings. We would appreciate a few minutes of your time to complete a brief Cisco online course evaluation of your instructor and the course materials in this student kit. On the final day of class, your instructor will provide you with a URL directing you to a short post-course evaluation. If there is no Internet access in the classroom, please complete the evaluation within the next 48 hours or as soon as you can access the web.

On behalf of Cisco, thank you for choosing Cisco Learning Partners for your Internet technology training.

Sincerely,

Cisco Systems Learning

https://t.me/learningnets

https://t.me/learningnets

Table of Contents

Volume 1

Course Introduction ***1***

Overview 1
Learner Skills and Knowledge 2
Course Goal and Objectives 3
Course Flow 4
Additional References 5
Cisco Glossary of Terms 5
Your Training Curriculum 6

Cisco Nexus Product Overview ***1-1***

Overview 1-1
Module Objectives 1-1

Describing the Cisco Data Center Network Architecture **1-3**

Overview 1-3
Objectives 1-3
Cisco Unified Fabric Fundamentals 1-4
Structured Layers: Core, Aggregation, Access 1-12
Product Placement 1-16
Positioning of Product Families in the Architecture 1-21
Summary 1-26

Identifying Cisco Nexus Products **1-27**

Overview 1-27
Objectives 1-27
Cisco Nexus Family of Products 1-28
Important Features of Cisco Nexus 7000 I/O Modules 1-47
Important Features of Cisco NX-OS 1-60
Summary 1-70
Module Summary 1-71
Module Self-Check 1-73
Module Self-Check Answer Key 1-75

Cisco Nexus Switch Feature Configuration ***2-1***

Overview 2-1
Module Objectives 2-1

Understanding High Availability and Redundancy **2-3**

Overview 2-3
Objectives 2-3
Network-Level High Availability 2-4
System-Level High Availability 2-20
Cisco IOS In-Service Software Upgrade 2-31
Summary 2-38
References 2-38

Configuring Virtual Device Contexts **2-39**

Overview 2-39
Objectives 2-39
Using VDCs in Data Centers 2-40
Virtual Device Contexts 2-44
Resource Allocation 2-48
New VDC Features in Cisco NX-OS 6.1 2-55
Configuring VDCs 2-58
Management Settings 2-66
Storage VDCs 2-71
Summary 2-76
References 2-76

https://t.me/learningnets

Configuring Layer 2 Switching Features 2-77
Overview 2-77
Objectives 2-77
Basic Interface Parameters 2-78
Cisco Nexus 7000 and Cisco Nexus 5000 Switch Feature Comparison 2-97
VLAN Configuration 2-98
STP Extensions 2-113
Summary 2-120
References 2-120
Configuring PortChannels 2-121
Overview 2-121
Objectives 2-121
Using Port Channels and vPCs 2-122
Configuring Port Channels 2-131
vPC Architecture 2-137
Configuring vPC 2-144
Configuring the FEX 2-154
Configuring Enhanced vPCs 2-164
Summary 2-170
References 2-170
Implementing Cisco FabricPath 2-171
Overview 2-171
Objectives 2-171
Implement Cisco FabricPath 2-172
Verify Cisco FabricPath 2-201
Summary 2-206
References 2-206
Configuring Layer 3 Switching Features 2-207
Overview 2-207
Objectives 2-207
Routing Protocols 2-208
First Hop Redundancy Protocols (FHRPs) 2-214
Bidirectional Forwarding Detection 2-224
Layer 3 Virtualization 2-228
Unicast RIB and FIB 2-233
Route Policy Manager 2-235
Policy-Based Routing (PBR) 2-239
IPv6 2-241
Summary 2-247
References 2-247
Configuring IP Multicast 2-249
Overview 2-249
Objectives 2-249
IP Multicast 2-250
Configuring IGMP and MLD 2-256
Configuring PIM 2-258
Configuring IGMP Snooping 2-269
Configuring MSDP 2-272
Summary 2-274
References 2-274
Module Summary 2-275
Module Self-Check 2-277
Module Self-Check Answer Key 2-286

 © 2012 Cisco Systems, Inc.

https://t.me/learningnets

Course Introduction

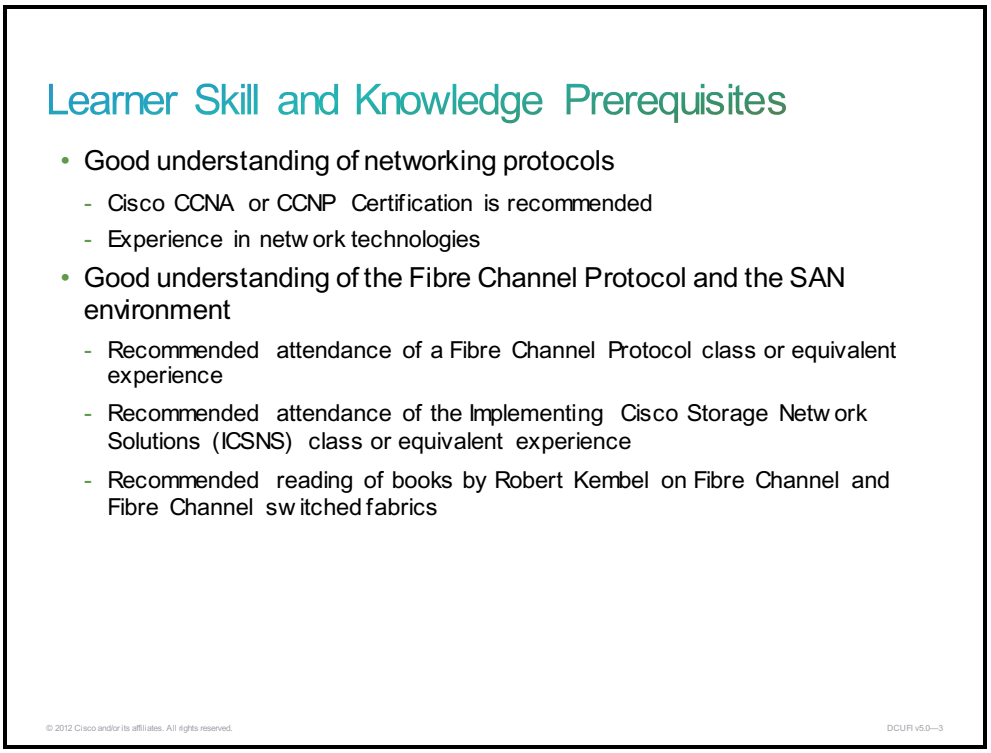
Overview

Implementing Cisco Data Center Unified Fabric (DCUFI) v5.0 is a five-day instructor-led course. The course is designed for systems and field engineers, consulting systems engineers, technical solutions architects, and Cisco integrators and partners who install and implement the Cisco Nexus 7000 and 5000 Series switches and the Cisco Nexus 2000 Fabric Extenders. The course covers the key components and procedures needed to install, configure, manage, and troubleshoot the Cisco Nexus 7000 and 5000 Series switches in the network and SAN environment.

https://t.me/learningnets

Learner Skills and Knowledge

This subtopic lists the skills and knowledge that learners must have in order to benefit fully from this course. The subtopic also includes recommended Cisco learning offerings that learners should first complete in order to benefit fully from this course.



Before attending this course, learners should be familiar with networking protocols and technologies, the SAN environment, and the Fibre Channel Protocol (FCP).

Cisco Certified Network Associate (CCNA®) or Cisco Certified Network Professional (CCNP®) level of knowledge is recommended for students attending the DCUFI course.

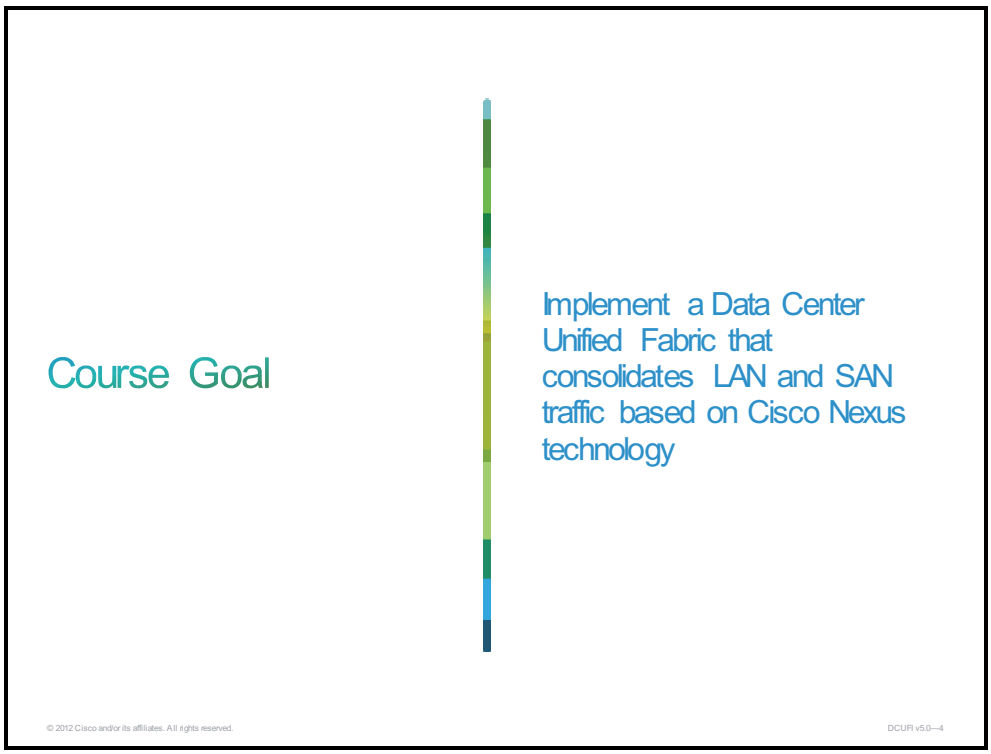
Note	The recommended courses for CCNA certification are the *Interconnecting Cisco Network Devices Part 1* (ICND1) and *Interconnecting Cisco Network Devices Part 2* (ICND2) courses.

In order to attain the appropriate level of knowledge of the Fibre Channel Protocol and SAN environment, the learner should have attended a Fibre Channel Protocol course such as the *Implementing Cisco Storage Network Solutions* (ICSNS) course. The recommended reading includes books by Robert Kembel books on Fibre Channel and Fibre Channel switched fabrics.

https://t.me/learningnets

Course Goal and Objectives

This topic describes the course goal and objectives.



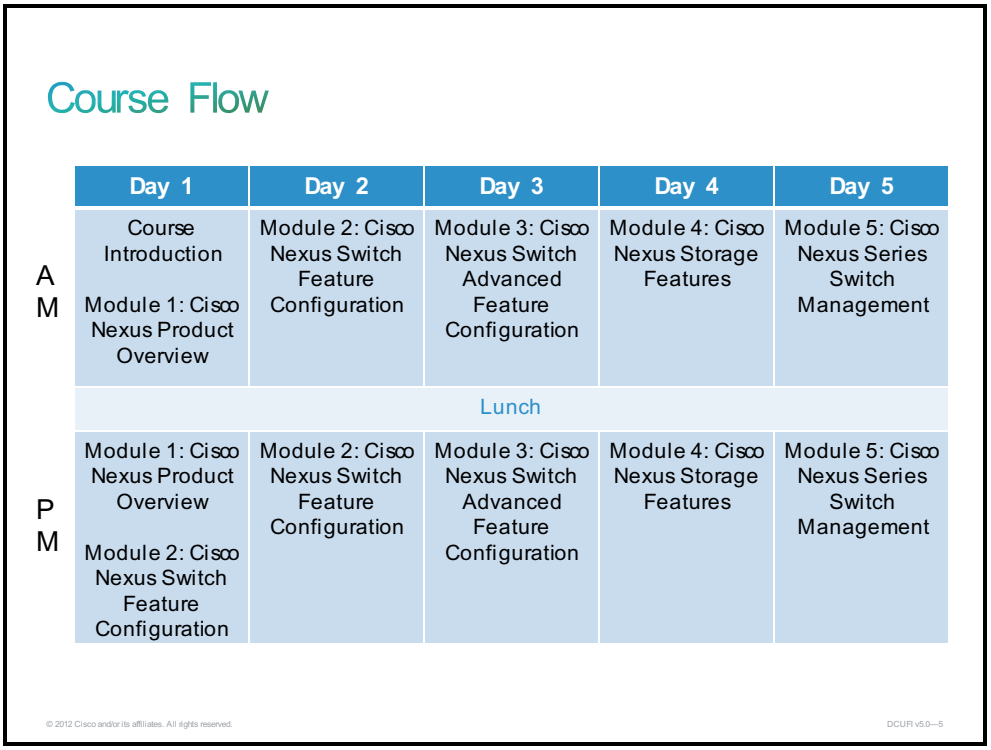
Upon completing this course, you will be able to meet these objectives:

- Identify the Cisco Nexus product family, specifically the Cisco Nexus 7000 Series switch chassis and components, the Cisco Nexus 5000 Series switch, and the Cisco Nexus 2000 Fabric Extender
- Install the Cisco Nexus products in a Cisco Data Center Business Advantage environment
- Given a requirement, identify how to plan and implement virtual device contexts into the solution
- Evaluate the security features available on the Cisco Nexus 7000 Series switch in order to identify which features should be implemented into a solution
- Evaluate and configure the Connectivity Management Processor on the Cisco Nexus 7000 Series switch and identify the management options available
- Evaluate the service-level and network-level high availability of the Cisco Nexus switches and how to use the Cisco IOS In-Service Software Upgrade feature
- Discuss the Fibre Channel Protocol, including Fibre Channel addressing, flow control, and zoning
- Translate a given design into an implementation plan for configuring Fibre Channel over Ethernet on the Cisco Nexus switch
- Understand the processes, tools, and resources for troubleshooting the data center infrastructure, interconnectivity, and operations

https://t.me/learningnets

Course Flow

This topic presents the suggested flow of the course materials.



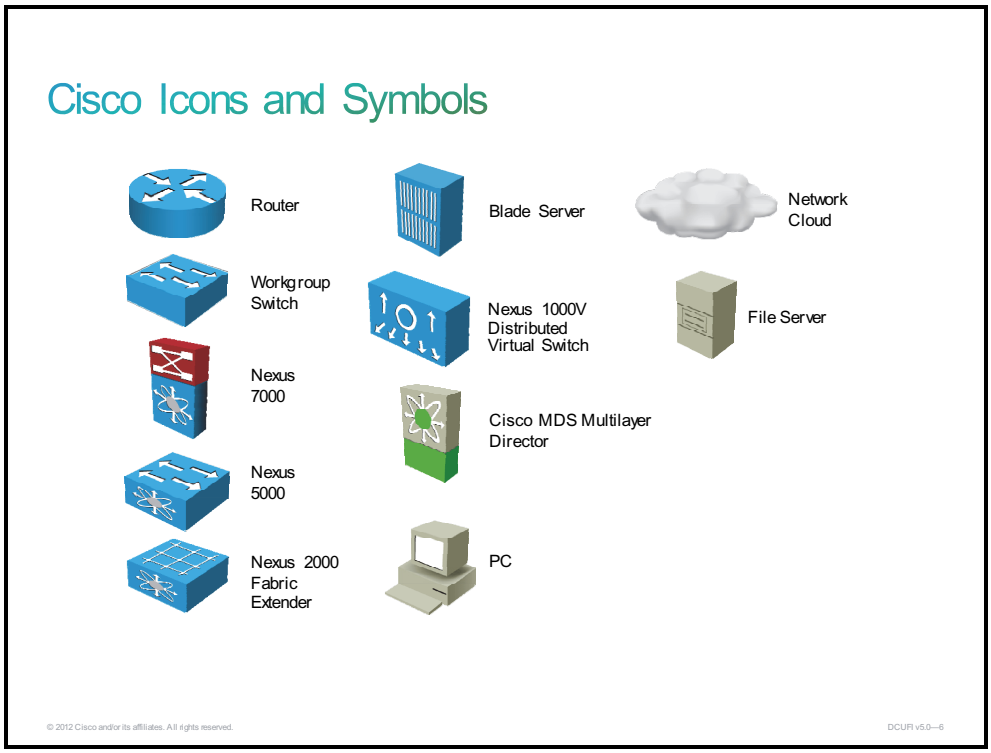
The schedule reflects the recommended structure for this course. This structure allows enough time for the instructor to present the course information and for you to work through the lab activities. The exact timing of the subject materials and labs depends on the pace of your specific class.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Additional References

This topic presents the Cisco icons and symbols that are used in this course, as well as information on where to find additional technical references.



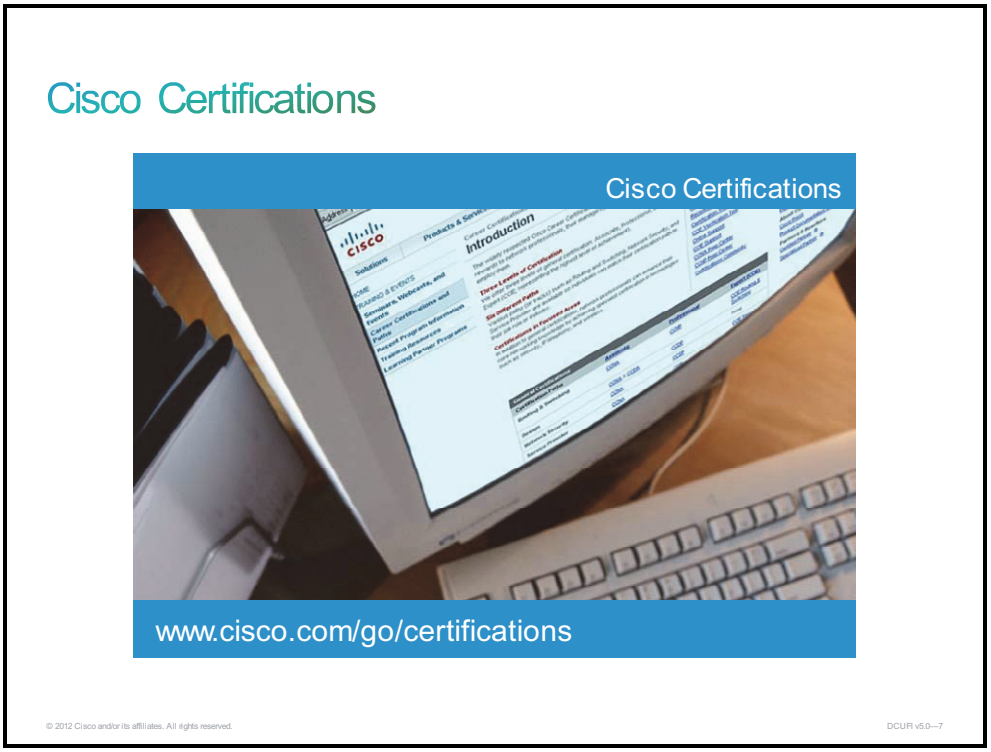
Cisco Glossary of Terms

For additional information on Cisco terminology, refer to the *Cisco Internetworking Terms and Acronyms (CIT) Guide* glossary of terms at http://docwiki.cisco.com/wiki/Internetworking_Terms_and_Acronyms_%28ITA%29_Guide.

https://t.me/learningnets

Your Training Curriculum

This topic presents the training curriculum for this course.

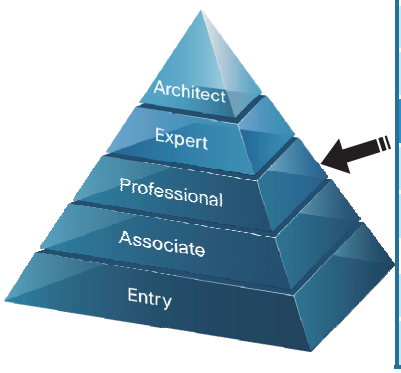


You are encouraged to join the Cisco Career Certification Community, a discussion forum open to anyone holding a valid Cisco Career Certification (such as Cisco CCIE®, CCNA®, CCDA®, CCNP®, CCDP®, CCIP®, CCVP®, or CCSP®). The community provides a gathering place for Cisco-certified professionals to share questions, suggestions, and information about Cisco Career Certification programs and other certification-related topics. For more information, visit www.cisco.com/go/certifications.

https://t.me/learningnets

Cisco Career Certifications: Cisco CCNP Data Center

Expand Your Professional Options and Advance Your Career



Cisco CCNP Data Center
Implementing Cisco Data Center Unified Fabric (DCUFI)
Implementing Cisco Data Center Unified Computing (DCUCI)
Available Exams (pick a group of 2)
Designing Cisco Data Center Unified Computing (DCUCD)
Designing Cisco Data Center Unified Fabric (DCUFD)
or
Troubleshooting Cisco Data Center Unified Fabric (DCUFT)
Troubleshooting Cisco Data Center Unified Computing (DCUCT)

www.cisco.com/go/certifications

© 2012 Cisco and/or its affiliates. All rights reserved.

https://t.me/learningnets

 © 2012 Cisco Systems, Inc.

https://t.me/learningnets

Module 1

Cisco Nexus Product Overview

Overview

In this module you will examine the Cisco Nexus Family of products, specifically the Cisco Nexus 7000 Series Switches chassis and components, the Cisco Nexus 5000 and 5500 Platform switches, Cisco Nexus 4000 and 3000 Series Switches, and the Cisco Nexus 2000 Series Fabric Extenders. You will also identify Cisco Nexus 7000 Series I/O modules and learn about the important features of the Cisco Nexus Operating System (NX-OS) Software.

Module Objectives

Upon completing this module, you will be able to describe the Cisco Unified Fabric products in the Cisco Data Center Network Architecture. This ability includes being able to meet these objectives:

- Describe the Cisco Data Center Network Architecture and its relationship to the Cisco Nexus Family of products
- Identify the Cisco Nexus Family of products and the important components of the chassis, line modules, and FEXs

https://t.me/learningnets

 © 2012 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 1

Describing the Cisco Data Center Network Architecture

Overview

The Cisco Nexus Family brings new technologies that are essential for building unified fabric and a new generation of data center. It is critical to be able to identify which device or technology is needed to solve the challenges that unified fabric poses to network design.

In this lesson, you will learn how to position the Cisco Nexus Family of products and other Cisco products in the Cisco Data Center Network Architecture.

Objectives

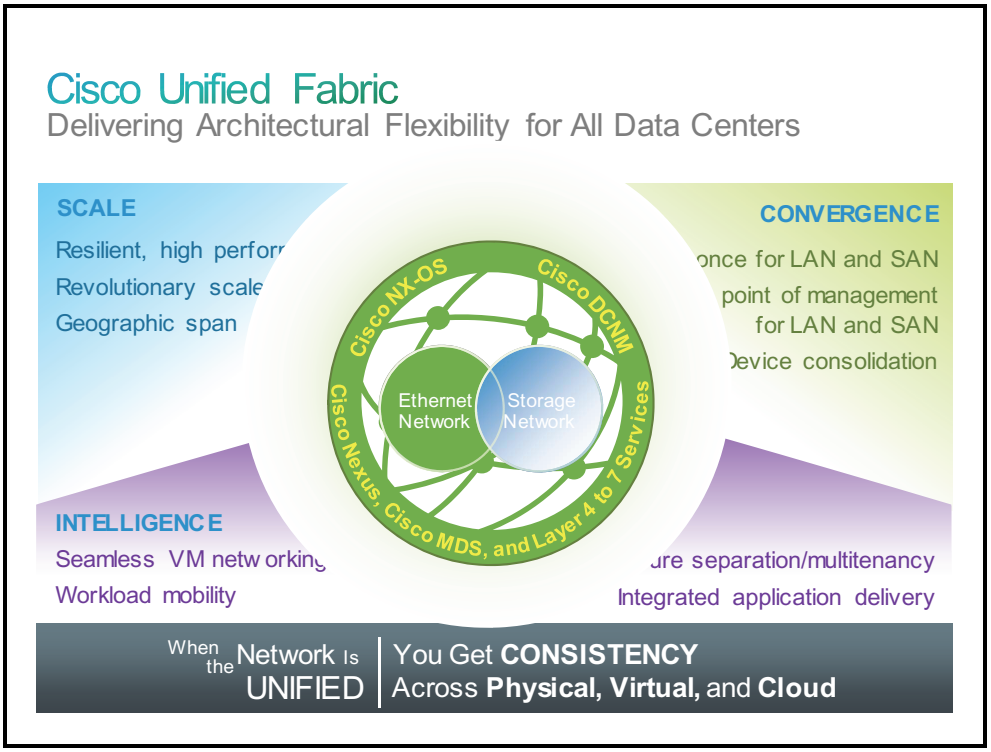
Upon completing this lesson, you will be able to describe the Cisco Data Center Network Architecture and its relationship to the Cisco Nexus Family of products. This ability includes being able to meet these objectives:

- Identify the components of the Cisco Unified Fabric solution
- Identify the structured layers of the Cisco Data Center Network Architecture
- Identify the placement of the Cisco Nexus and Cisco MDS Families of switches, Cisco UCS, Cisco Adapter FEX, and Cisco VM-FEX products in the Cisco Data Center Network Architecture
- Identify how to position different product families in the Cisco Data Center Network Architecture

https://t.me/learningnets

Cisco Unified Fabric Fundamentals

This topic identifies the components of the Cisco Unified Fabric solution.



The Cisco Unified Fabric solution provides the foundational connectivity for general-purpose, virtualized, and cloud-based data centers and unifies storage, data networking, and network services. Cisco Unified Fabric delivers architectural flexibility to address the diverse requirements of all types of data centers.

It includes the Cisco Nexus and MDS Family portfolios, the Cisco Nexus Operating System (NX-OS) and Cisco Data Center Network Manager (DCNM), along with Layer 4 to Layer 7 solutions.

Cisco Unified Fabric uniquely offers multidimensional scalability for the data center network: switch performance, system scale, and geographic span.

Business and IT agility is achieved through a flexible and highly available secure fabric that supports dynamic resource allocation, changing traffic patterns, complex workloads, and industry-leading simultaneous scalability within and across data centers.

Cisco Unified Fabric enables converged fabric. Financial efficiencies and investment protection are achieved through consolidation, multiprotocol solutions, and a single point of management for LAN and SAN. These attributes enable an evolutionary adoption without disruption to existing infrastructure and operations. Fibre Channel over Ethernet (FCoE) simplifies the data center network by converging LANs and SANs over a single lossless Ethernet network providing a “wire once, connect anything” approach. It reduces network hardware sprawl through consolidation of Ethernet and SAN switches. It also consolidates LAN and SAN cabling onto a single Ethernet cable, significantly simplifying data center management while reducing overall capital expenditures (CapEx) and operating expenses (OpEx).

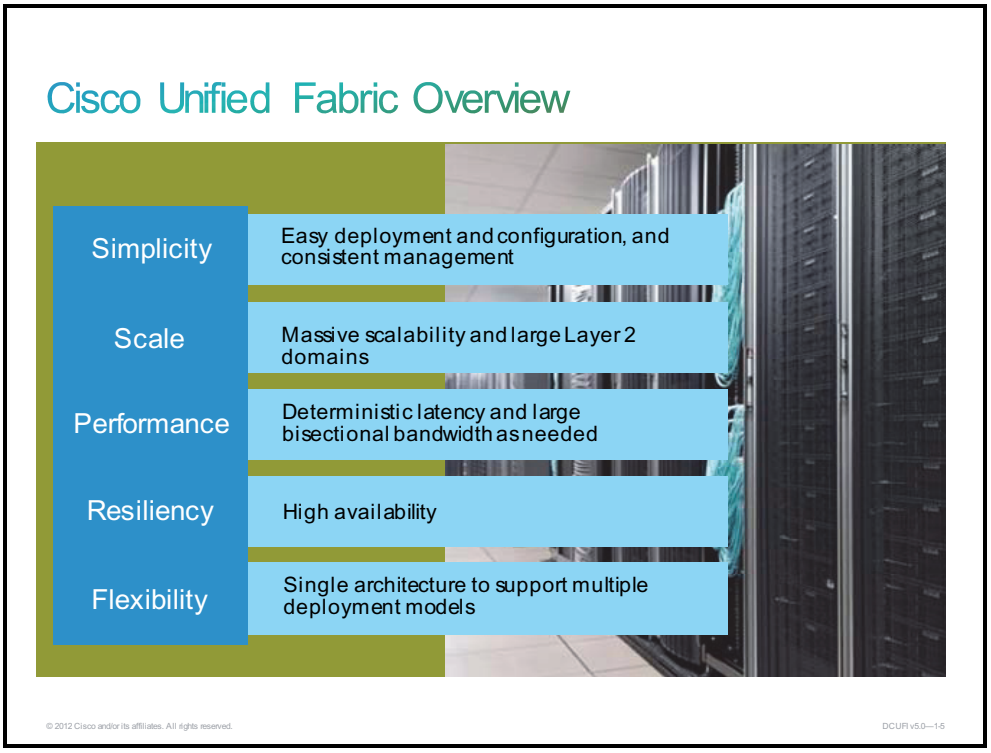
© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Cisco Unified Fabric provides intelligence. Simplified operations are achieved through embedding virtualization-aware policy-based security and intelligent, consistent services directly into the network fabric. This strategy results in application acceleration and seamless and efficient general-purpose, converged, virtualized, and cloud environments.

Cisco Unified Fabric provides consistent networking across physical, virtual, and cloud environments. This consistency enables IT as a service model for delivering agile and cost-effective network services to servers, storage, and applications. In return, the consistency helps customers reduce the percentage of budget and time that is spent on data center maintenance and instead focus on contributing to the profit line and business innovation by delivering new and improved services.

https://t.me/learningnets



The Cisco approach to the data center is to provide an open and standards-based architecture. System-level benefits such as performance, energy efficiency, and resiliency are addressed, along with workload mobility and security. Cisco offers tested, preintegrated, and validated designs, providing businesses with a faster deployment model and quicker time to market.

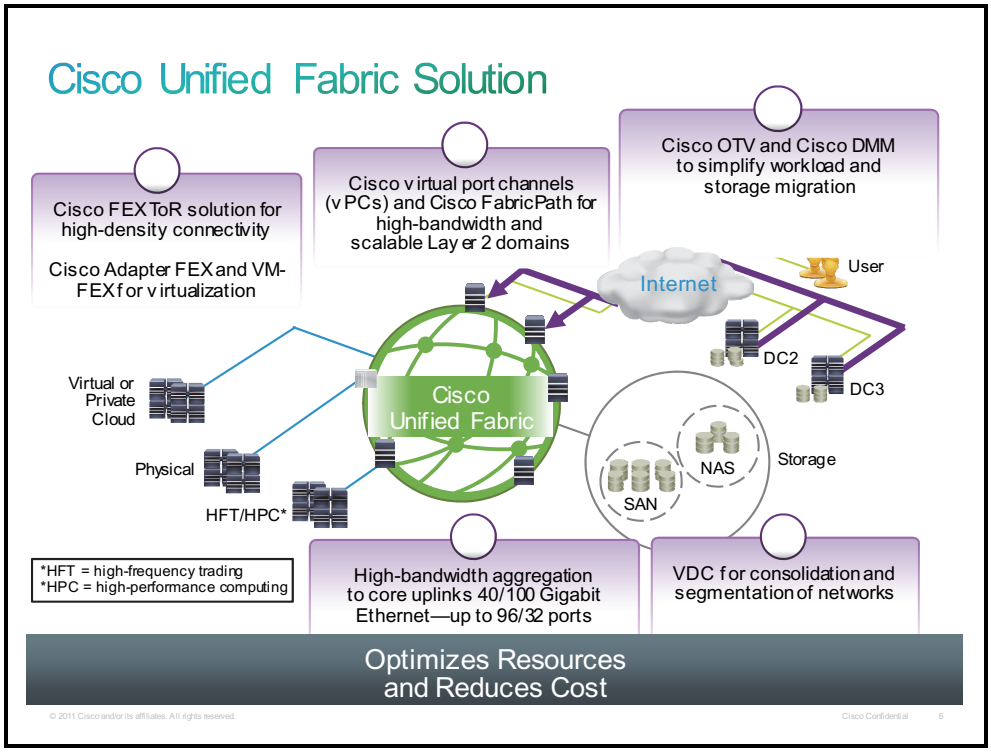
Cisco Unified Fabric delivers transparent convergence, massive three-dimensional scalability, and sophisticated intelligent services to provide the following benefits:

- Support for traditional and virtualized data centers
- Reduction in total cost of ownership (TCO)
- An increase in return on investment (ROI)

The five architectural components that affect TCO include the following:

- **Simplicity:** Businesses need the data center to be able to provide easy deployment and configuration and consistent management of existing and new services.
- **Scale:** Data centers need to be able to support large Layer 2 domains that can provide massive scalability without the loss of bandwidth and throughput.
- **Performance:** Data centers should be able to provide deterministic latency and large bisectional bandwidth to applications and services as needed.
- **Resiliency:** The data center infrastructure and implemented features need to provide high availability to the applications and services that they support.
- **Flexibility:** Businesses need a single architecture that can support multiple deployment models.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



Reducing the number of data centers to one or a few data centers requires more efficient use of space in the remaining data centers and also more network capacity to manage the increased load. Secure segmentation is also required. The Cisco Unified Fabric provides several innovations and solutions to help customers maximize space and deliver ample network capacity to accommodate small or large data center consolidation.

1. At the server access level, fabric extender (FEX) technology enables high density server deployments with easy to deploy and configure top-of-rack (ToR) Cisco Nexus 2000 Series Fabric Extenders that support Gigabit Ethernet and 10 Gigabit Ethernet connectivity. Cisco Adapter Fabric Extender (Adapter FEX) and Cisco Data Center Virtual Machine Fabric Extender (Cisco VM-FEX) provide added scalability at the server level by partitioning the server network adapters and by offloading the hypervisor, allowing for more virtual machines (VMs) to be loaded in each server.
2. To support higher density and higher VM to server ratio, 10 Gigabit Ethernet connectivity to the server is becoming commonplace. However, 10 Gigabit Ethernet connectivity can lead to bottlenecks between the aggregation and core. To avoid bottlenecks, the Nexus 7000 Series Switches offer high speed, standards-based, 40 Gigabit Ethernet and 100 Gigabit Ethernet connectivity.
3. To scale the bandwidth between the access and aggregation layer and also enable larger Layer 2 domains for virtualized pods, the Cisco Unified Fabric offers virtual port channel (vPC) and Cisco FabricPath. Unlike spanning tree, vPC and Cisco FabricPath allow all links to be active and forwarding.
4. In some situations, separate data centers may have been required to provide isolation and security. With the Cisco Unified Fabric, isolation and security can be provided with features like virtual device context (VDC) and virtual routing and forwarding (VRF). A VDC allows a single switch to be partitioned, providing complete data plane and control plane separation and fault isolation. It also provides securely delineated administrative contexts so that each VDC can be managed by a different IT staff person. VDCs allow multiple separate switches to be consolidated into one switch, for a reduced number of devices, which results in lower power usage, a reduced footprint and lower CapEx and OpEx.

https://t.me/learningnets

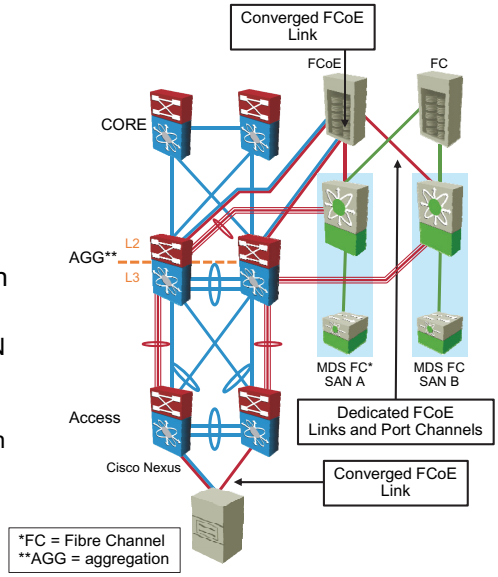
5. One of the issues of consolidating data centers is the duration of the outage during the consolidation process when data is being moved from one data center to the other. Cisco Unified Fabric offers several innovations that help alleviate the migration outage. Cisco Overlay Transport Virtualization (Cisco OTV) extends Layer 2 domains (VLANs) across any network, allowing for a seamless migration of VMs from one data center to the other. Cisco Data Mobility Manager (DMM) enables online migration of data storage across heterogeneous storage devices.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Building Blocks: Converged and Dedicated Links

- Converged links to the access switch allow:
  - Cost savings in the reduction of required equipment
  - “Cable once” for all servers to have access to both LAN and SAN networks
- Dedicated links from access to aggregation and in aggregation layer are common:
  - Separate links for SAN and LAN traffic ; both links are same I/O (10 Gigabit Ethernet)
  - Advanced Ethernet features can be applied to the LAN links
  - Maintains fabric isolation



© 2012 Cisco and/or its affiliates. All rights reserved.

Building upon the converged network adapters (CNA), the data center can use converged or dedicated links:

1. Converged links allow the enterprises to save costs through the reduction of required equipment. They enable the “cable once” approach for all servers to have access to both LAN and SAN networks. Converged links are most common as the access links to the access switch and may be used in other network layers.

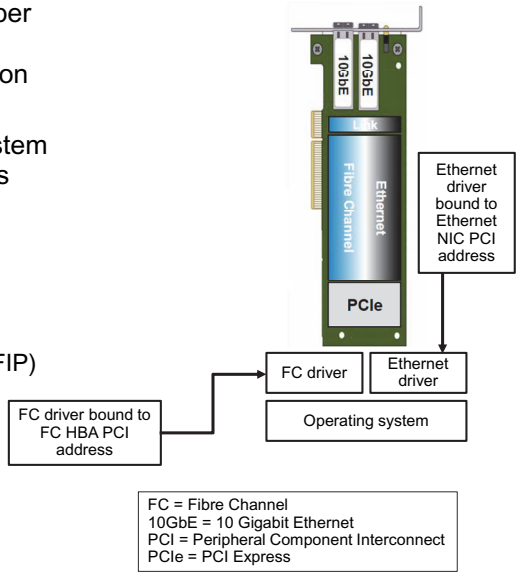
2. Dedicated links provide separation of SAN and LAN traffic. Both links can be of the same I/O type, most typically 10 Gigabit Ethernet. Advanced Ethernet features can be applied to the LAN links. The main advantage of dedicated links is the fabric isolation. This figure depicts dedicated links from access to aggregation. Dedicated links are typical in aggregation and core layers.

3. From a SAN perspective, the use of converged links does not change anything; SANs are still separated and each SAN has its own dedicated links.

https://t.me/learningnets

Building Blocks: Converged Network Adapter

- Replaces multiple adapters per server, consolidating both Ethernet and Fibre Channel on a single interface
- Appears to the operation system as individual interfaces (NICs and HBAs)
- Features:
  - Priority flow control (PFC)
  - Data Center Bridging (DCB)
  - FCoE Initialization Protocol (FIP)
  - Single chip implementation
  - Low power consumption



© 2012 Cisco and/or its affiliates. All rights reserved.

Fabric unification would not be possible without converged network adapters (CNAs). A CNA is a computer I/O device that combines the functionality of a host bus adapter (HBA) with a network interface controller (NIC). In other words it "converges" access to, respectively, a SAN and a general-purpose computer network.

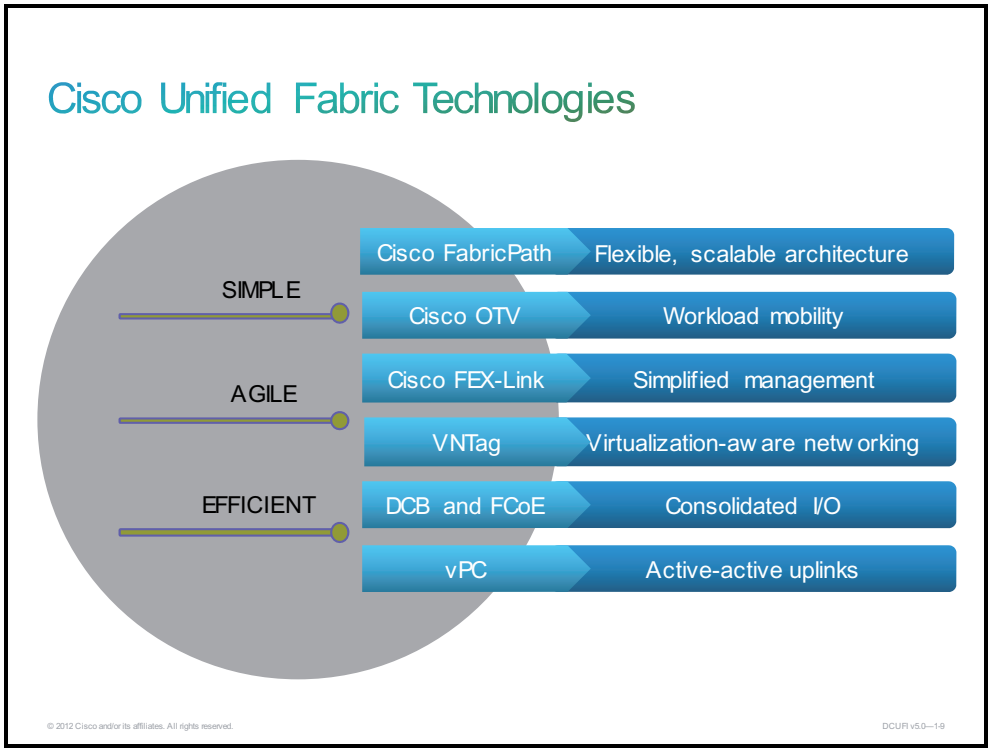
The CNA appears to the operation system as individual interfaces, that is the NIC and HBAs, respectively.

To implement unified fabric, several technologies need to be implemented on CNA:

- Priority flow control (PFC): Used for nondrop flow control on Ethernet
- Data Center Bridging (DCB): Used for feature negotiation and exchange among devices that are building unified fabric
- FCoE Initialization Protocol (FIP): Used during FCoE initialization

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



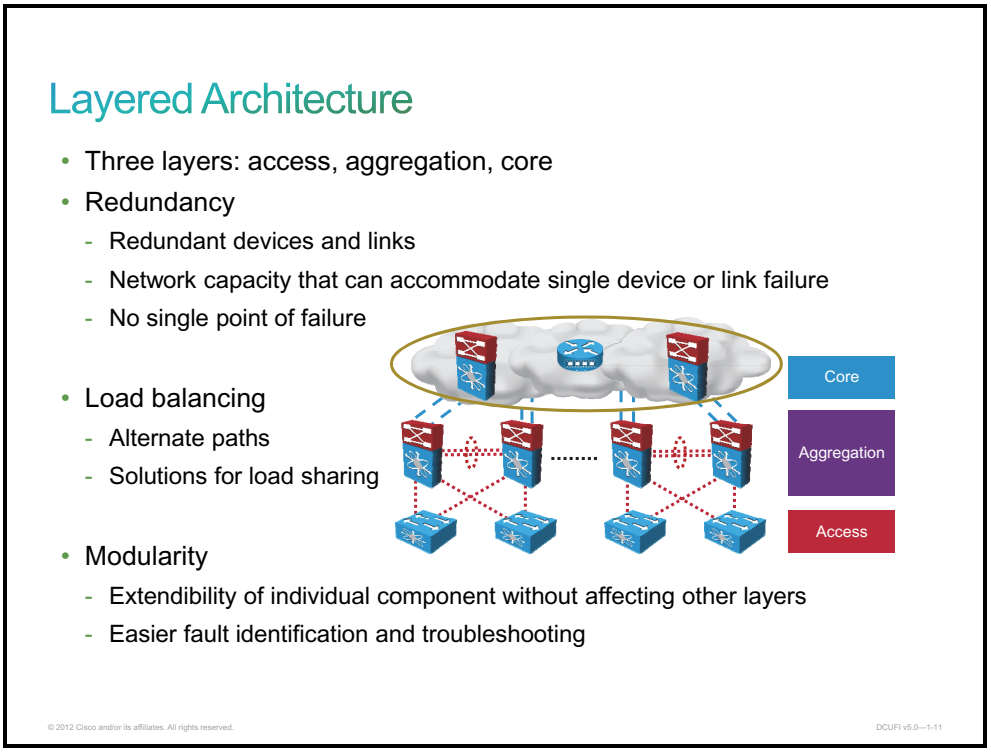
To support the five architectural attributes, the Cisco Unified Fabric evolution is continuing to provide architectural innovations.

- **Cisco FabricPath:** Cisco FabricPath is a set of capabilities within the Cisco Nexus Operating System (Cisco NX-OS) Software combining the "plug-and-play" simplicity of Ethernet with the reliability and scalability of Layer 3 routing. Cisco FabricPath enables companies to build highly scalable Layer 2 multipath networks without the Spanning Tree Protocol (STP). These networks are particularly suitable for large virtualization deployments, private clouds, and high-performance computing environments.
- **Cisco OTV:** Cisco Overlay Transport Virtualization (Cisco OTV) is an industry-first solution that significantly simplifies extending Layer 2 applications across distributed data centers. Cisco OTV allows companies to deploy virtual computing resources and clusters across geographically distributed data centers, delivering transparent workload mobility, business resiliency, and superior computing resource efficiencies.
- **Cisco FEX-Link**: Cisco Fabric Extender Link (Cisco FEX-Link) technology enables data center architects to gain new design flexibility while simplifying cabling infrastructure and management complexity. Cisco FEX-Link uses the Cisco Nexus 2000 Series Fabric Extenders to extend the capacities and benefits that are offered by upstream the Cisco Nexus Family of switches.
- **VNTag**: The virtual network tag (VNTag) provides advanced hypervisor switching as well as high-performance hardware switching. It is flexible, extensible, and service-enabled. The VNTag architecture provides virtualization-aware networking and policy control.
- **Data Center Bridging (DCB) and FCoE**: Cisco Unified Fabric provides the flexibility to run Fibre Channel, IP-based storage such as network-attached storage (NAS) and Internet Small Computer System Interface (iSCSI), or FCoE, or a combination of these technologies, on a converged network.
- **vPC**: Virtual port channel (vPC) technology enables the deployment of a link aggregation from a generic downstream network device to two individual and independent Cisco NX-OS devices (vPC peers). This multichassis link aggregation path provides both link redundancy and active-active link throughput scaling high-performance failover characteristics.

https://t.me/learningnets

Structured Layers: Core, Aggregation, Access

This topic identifies the structured layers of the Cisco Data Center Network Architecture.



The architectural components of the infrastructure are the access layer, the aggregation layer, and the core layer. The principal advantages of this model are its hierarchical structure and its modularity. A hierarchical design avoids the need for a fully meshed network in which all network nodes are interconnected. Modules in a layer can be put into service and taken out of service without affecting the rest of the network. This ability facilitates troubleshooting, problem isolation, and network management.

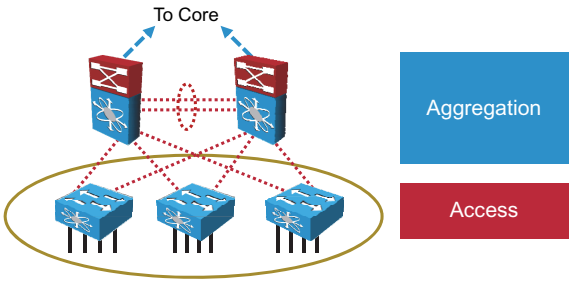
The hierarchical network model supports designing a highly available modular topology using scalable building blocks that allow the network to meet evolving business needs. The modular design makes the network easy to scale, understand, and troubleshoot by promoting deterministic traffic patterns.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Access Layer

- Provides access and aggregation for applications in an environment many features
- Provides high availability through software attributes and redundancy
- Supports convergence for voice, wireless, and data
- Provides security services to help control network access
- Offers QoS services including traffic classification and queuing
- Supports IP multicast traffic for efficient network use



© 2012 Cisco and/or its affiliates. All rights reserved.

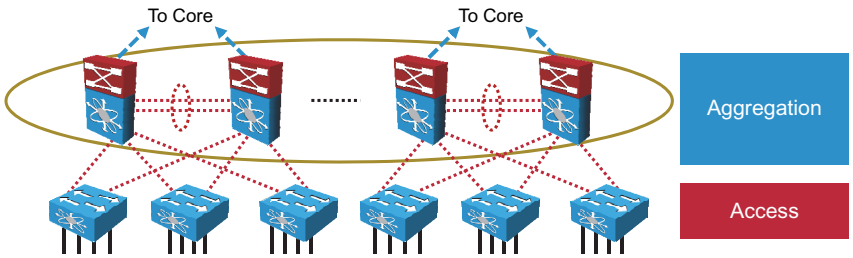
The access layer aggregates end users and provides uplinks to the aggregation layer. The access layer is generally an environment with many features including the following features:

- **High availability:** The access layer is supported by many hardware and software attributes. This layer offers system-level redundancy by using redundant supervisor engines and redundant power supplies for crucial application groups. The layer also offers default gateway redundancy by using dual connections from access switches to redundant aggregation layer switches that use a First Hop Redundancy Protocol (FHRP), such as Hot Standby Router Protocol (HSRP).
- **Convergence:** The access layer supports inline Power over Ethernet (PoE) for IP telephony and wireless access points (APs). This support allows customers to converge voice onto their data networks and provides roaming wireless LAN (WLAN) access for users.
- **Security:** The access layer provides services for additional security against unauthorized access to the network. This security is provided by using tools such as IEEE 802.1X, port security, DHCP snooping, Dynamic ARP Inspection (DAI), and IP Source Guard.
- **Quality of service (QoS):** The access layer allows prioritization of mission-critical network traffic by using traffic classification and queuing as close to the ingress of the network as possible. The layer supports the QoS trust boundary.
- **IP multicast:** The access layer supports efficient network and bandwidth management by using software features such as Internet Group Management Protocol (IGMP) snooping for IP version 4 (IPv4) multicast or Multicast Listener Discovery (MLD) for IP version 6 (IPv6) multicast.

https://t.me/learningnets

Aggregation Layer

- Aggregates access nodes and uplinks
- Provides redundant connections and devices for high availability
- Offers routing services such as summarization, redistribution, and default gateways
- Implements policies including filtering, security, and QoS mechanisms
- Segments workgroups and isolates problems



© 2012 Cisco and/or its affiliates. All rights reserved.

Availability, load balancing, QoS, and provisioning are the important considerations at the aggregation layer. High availability is typically provided through dual paths from the aggregation layer to the core and from the access layer to the aggregation layer. Layer 3 equal-cost load sharing allows both uplinks from the aggregation to the core layer to be used.

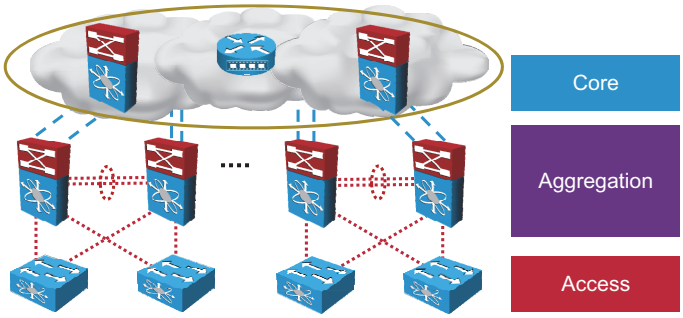
The aggregation layer is the layer in which routing and packet manipulation is performed and can be a routing boundary between the access and core layers. The aggregation layer represents a redistribution point between routing domains or the demarcation between static and dynamic routing protocols. This layer performs tasks such as controlled-routing decision making and filtering to implement policy-based connectivity and QoS. To further improve routing protocol performance, the aggregation layer summarizes routes from the access layer. For some networks, the aggregation layer offers a default route to access layer routers and runs dynamic routing protocols when communicating with core routers.

The aggregation layer uses a combination of Layer 2 and multilayer switching to segment workgroups and to isolate network problems so that they do not affect the core layer. This layer is commonly used to terminate VLANs from access layer switches. The aggregation layer also connects network services to the access layer and implements policies regarding QoS, security, traffic loading, and routing. In addition, this layer provides default gateway redundancy by using a First-Hop Resiliency Protocol (FHRP) such as Hot Standby Router Protocol (HSRP), Gateway Load Balancing Protocol (GLBP), or Virtual Router Redundancy Protocol (VRRP). Default gateway redundancy allows for the failure or removal of one of the aggregation nodes without affecting endpoint connectivity to the default gateway.

https://t.me/learningnets

Core Layer

- High-speed backbone and aggregation point for the enterprise.
- Reliability is achieved through redundancy and fast convergence.
- Aggregation layer switches are connected hierarchically.
  - Less physical cabling is required.
  - Less routing complexity is imposed.
- Separate core layer helps in scalability during future growth.



© 2012 Cisco and/or its affiliates. All rights reserved.

The core layer is the backbone for connectivity and is the aggregation point for the other layers and modules in the Cisco data center architecture. The core must provide a high level of redundancy and must adapt to changes very quickly. Core devices are most reliable when they can accommodate failures by rerouting traffic and can respond quickly to changes in the network topology. The core devices must be able to implement scalable protocols and technologies, alternate paths, and load balancing. The core layer helps in scalability during future growth.

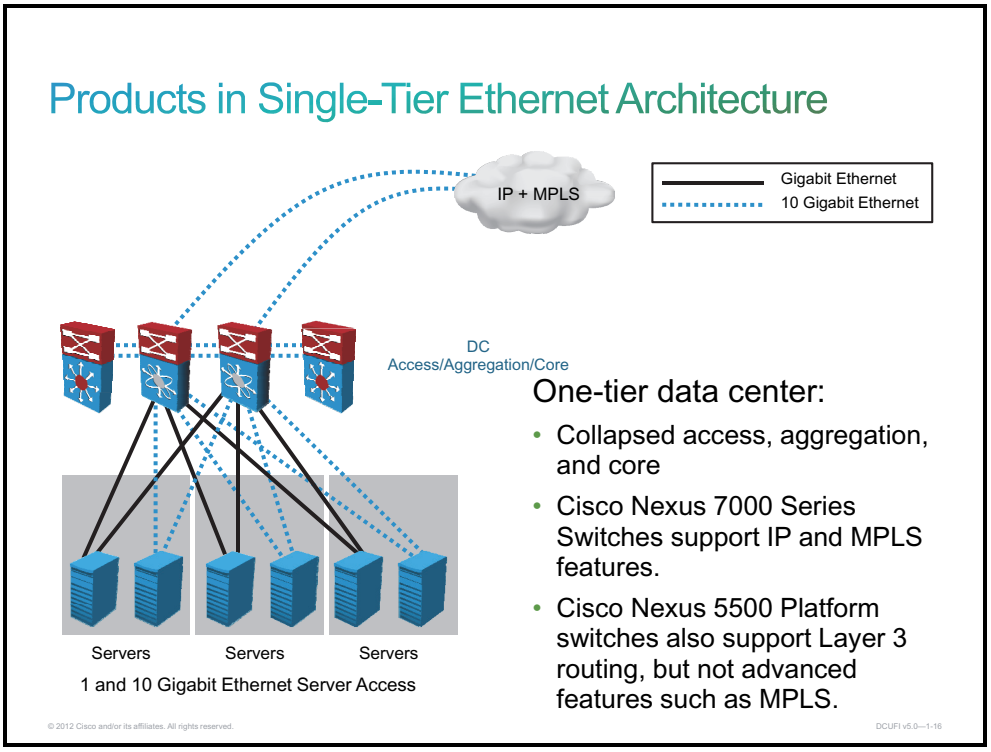
The core should be a high-speed Layer 3 switching environment that uses hardware-accelerated services. For fast convergence around a link or node failure, the core uses redundant point-to-point Layer 3 interconnections in the core. That type of design yields the fastest and most deterministic convergence results. The core layer should not perform any packet manipulation, such as checking access lists and filtering, which would slow down the switching of packets.

Without a core layer, the distribution layer switches will need to be fully meshed. The full-mesh design is difficult to scale, and increases the cabling requirements because each new building distribution switch needs full-mesh connectivity to all the distribution switches. The routing complexity of a full-mesh design increases as new neighbors are added.

https://t.me/learningnets

Product Placement

This topic identifies the placement of the Cisco Nexus and MDS Families of switches, Cisco Unified Computing System (Cisco UCS), Cisco Adapter FEX, and Cisco Data Center Virtual Machine Fabric Extender (Cisco VM-FEX) products in the Cisco Data Center Network Architecture.



The Cisco Nexus Family of products covers the access layer through to the core layer in any network infrastructure.

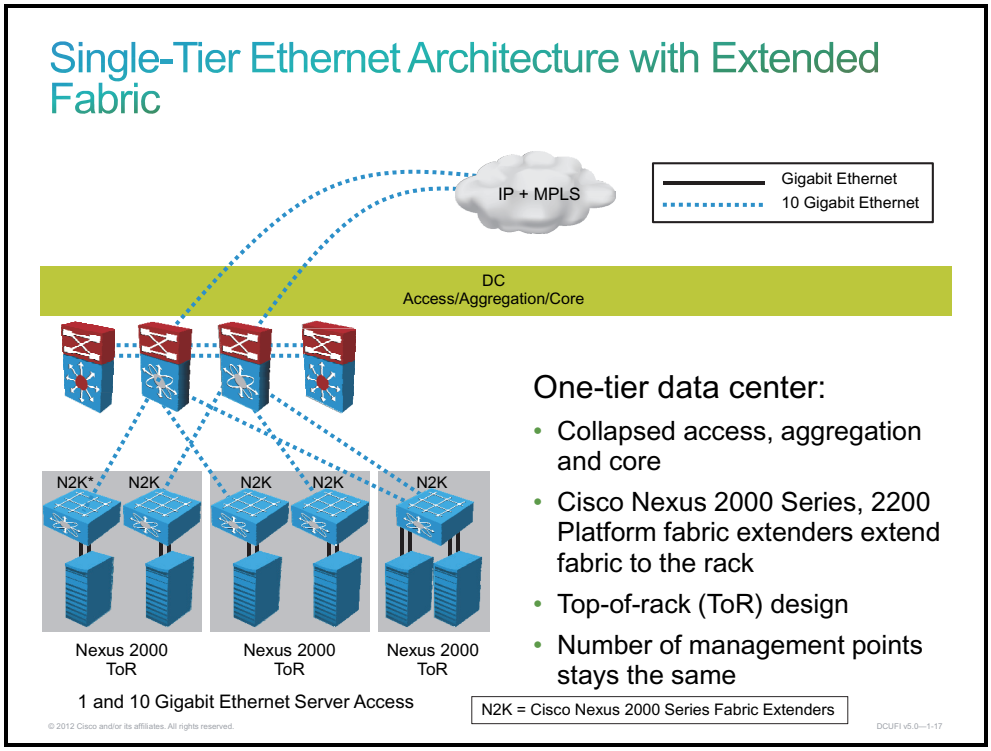
The Cisco Nexus Family of products encompasses switches that would be used at the access layer, through to switches to be used in the aggregation and core layers of the data center and network architecture. Switches in this family are not restricted to a single layer only. For example, the Cisco Nexus 7000 Series Switches could be used in the core, aggregation, or access layer where high densities of servers require 1 and 10 Gigabit Ethernet connectivity.

In the single-tier data center architecture, the Cisco Nexus 7000 Series Switches could be used for both access and core layer connectivity. The access layer connectivity for the servers would be provided by using the 48-port Gigabit Ethernet line module and, where necessary, the 32-port 10 Gigabit Ethernet line module.

Connectivity from a Cisco Nexus 7000 Series switch to the IP and Multiprotocol Label Switching (MPLS) core would be provided by using the 10 Gigabit Ethernet line modules, with a separate layer for services such as server load balancers or firewalls.

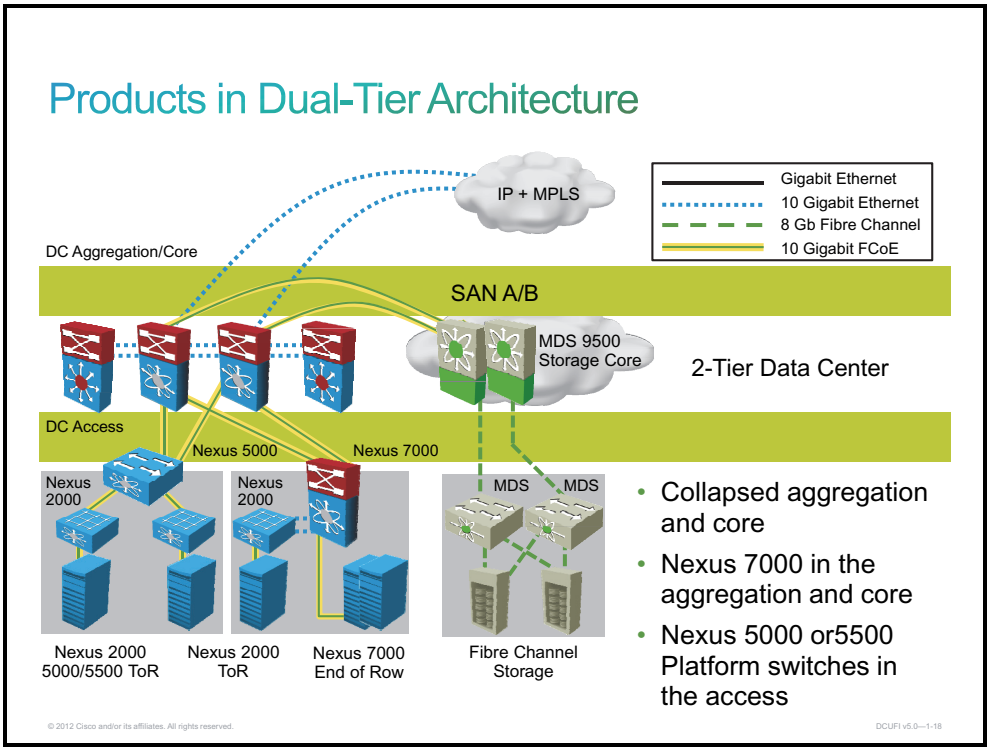
© 2012 Cisco Systems, Inc.

https://t.me/learningnets



You can expand the single-tier data center architecture by connecting a Cisco Nexus 2200 Platform fabric extender to a Cisco Nexus 7000 Series switch to provide the Gigabit Ethernet connectivity for the servers. Up to 10 Gigabit Ethernet links would connect the Cisco Nexus 2200 Platform fabric extender to the Cisco Nexus 7000 Series parent switch. This setup would provide a top-of-rack (ToR) solution for the servers with a Cisco Nexus 7000 Series switch acting as the management point, and access, aggregation, and core layers. Cisco NX-OS Software supports the Cisco Nexus 2200 Platform fabric extenders.

https://t.me/learningnets

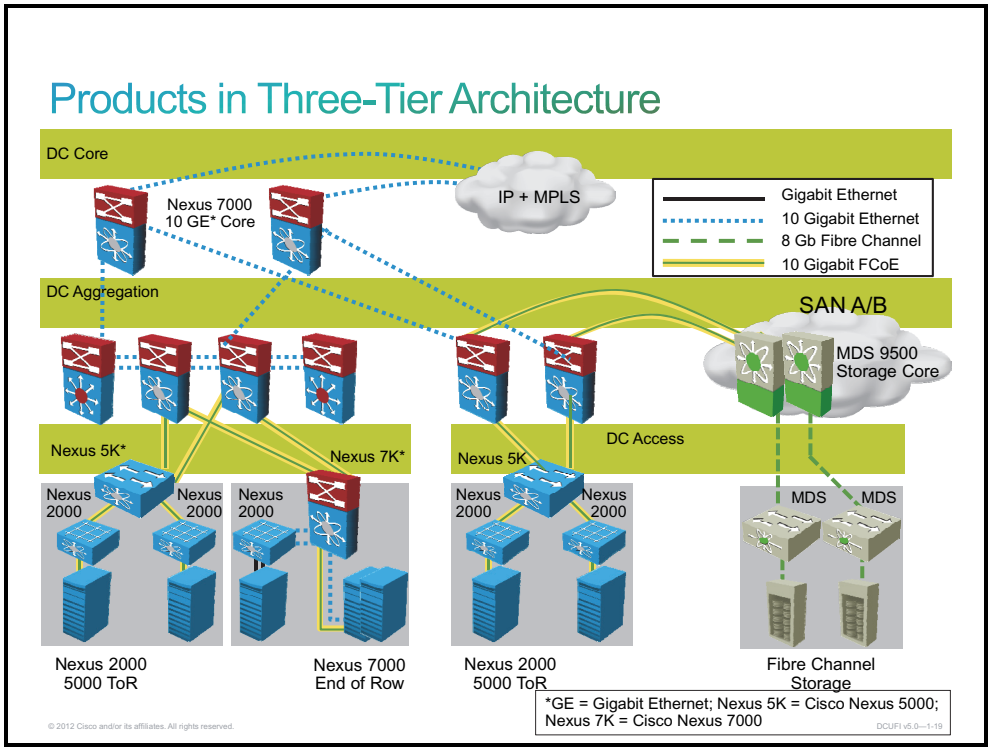


The two-tier data center option connects the Cisco Nexus 2000 Fabric Extenders to an upstream Cisco Nexus 5000 Platform or 5500 Platform switch. The Cisco Nexus 5000 or 5500 Platform switch would then connect to the Cisco Nexus 7000 Series switch. This topology provides an access layer and a collapsed core and aggregation layer. As an end-of-row (EoR) switch, the Cisco Nexus 7000 Series switch would act as a collapsed access and aggregation layer.

To support the high density of servers at the access layer, a Cisco Nexus 7000 Series switch could be deployed instead of, or in addition to, the Cisco Nexus 5000 or 5500 Platform switches.

The Cisco MDS 9000 Series Multilayer Switches provide the SAN connectivity at the access layer and the storage core layer. Optionally, an FCoE connection could be provided from the Cisco Nexus 7000 Series switch to the Cisco MDS 9000 Series core switches. This setup would support I/O consolidation at the access layer where the Cisco Nexus 5000 or 5500 Platform switches are located, using a Cisco Nexus 2200 Platform fabric extender.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



The illustration shows potential product placements within the campus, data center, and storage infrastructures.

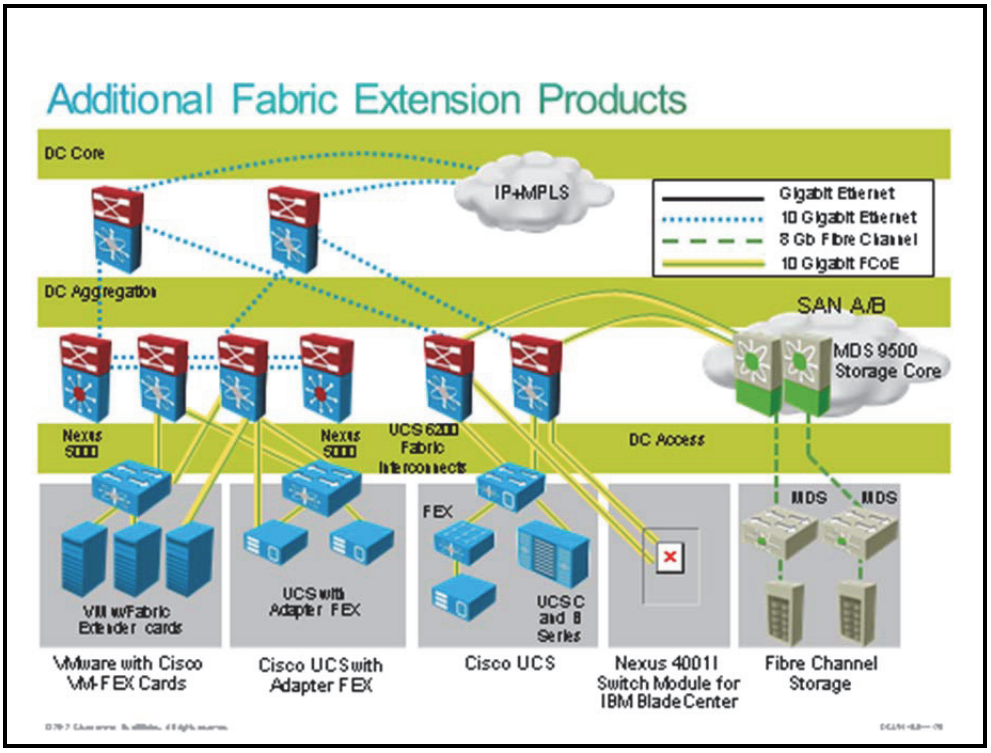
Within the data center, use of the Cisco Nexus 5000 and 5500 Platform switches, with the Cisco Nexus 2000 Series Fabric Extenders, offers the option to provide FCoE I/O consolidation at the access layer. The Cisco MDS 9000 Series Multilayer Switches would be used to support the SAN infrastructure.

Connectivity between the SAN and LAN infrastructures to support FCoE would be supported through the Cisco Nexus 7000 F1-Series line modules for the Cisco Nexus 7000 Series switch and the Cisco MDS 9500 Series core layer.

To support a services layer for services such as server load balancing and firewalling, a pair of Cisco Catalyst 6500 Series Switches would be used off the aggregation layer Cisco Nexus 7000 Series Switches.

The core layer would be provided by the Cisco Nexus 7000 Series Switches.

https://t.me/learningnets



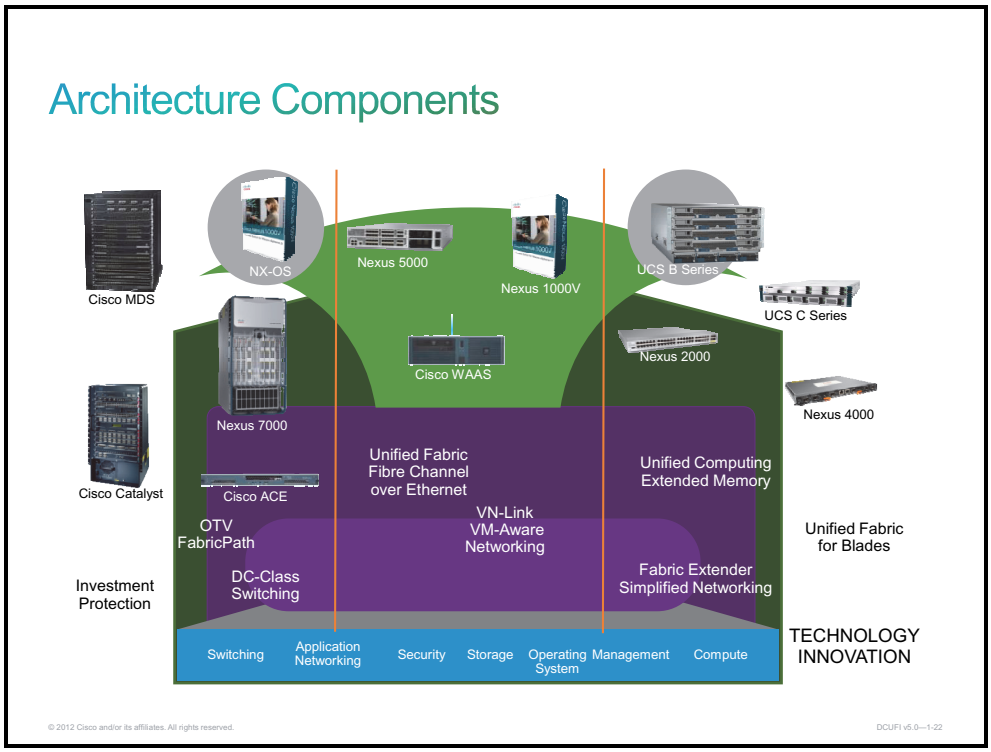
In addition to the Cisco Nexus 2000 Series Fabric Extenders, Cisco offers several other solutions to extend the fabric to the server:

- Cisco VM-FEX collapses virtual and physical networking into a single infrastructure. The Cisco VM-FEX software extends Cisco Fabric Extender Technology (FEX Technology) to the virtual machine (VM) with the following capabilities:
  - Each VM includes a dedicated interface on the parent switch.
  - All VM traffic is sent directly to the dedicated interface on the switch.
  - The software-based switch in the hypervisor is eliminated.
- Cisco UCS P81E Virtual Interface Card is a virtualization-optimized FCoE PCI Express (PCIe) 2.0 x8 10-Gb/s adapter that is designed for use with Cisco UCS C-Series Rack-Mount Servers. The virtual interface card is a dual-port 10 Gigabit Ethernet PCIe adapter that can support up to 128 PCIe standards-compliant virtual interfaces, which can be dynamically configured so that both their interface type (NIC or HBA) and identity (MAC address and world wide name [WWN]) are established using just-in-time provisioning. The Cisco UCS P81E supports network interface virtualization and Cisco VM-FEX technology.
- A combination of the Cisco UCS 6100 and 6200 Series Fabric Interconnects with the Cisco Nexus 2200 Platform fabric extenders and the Cisco UCS system.
- The Cisco Nexus 4000 Series Switches extend the benefits of the Cisco Nexus Family to blade servers. The Cisco Nexus 4000 Series provides all ports with support for both Gigabit Ethernet and 10 Gigabit Ethernet autonegotiation, for increased investment protection. It is also a Fibre Channel over Ethernet (FCoE) switch and is fully compliant with the IEEE DCB specification. The series is commonly used with, but not restricted to, the IBM BladeCenter solution.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Positioning of Product Families in the Architecture

This topic identifies how to position different product families in the Cisco Data Center Network Architecture.

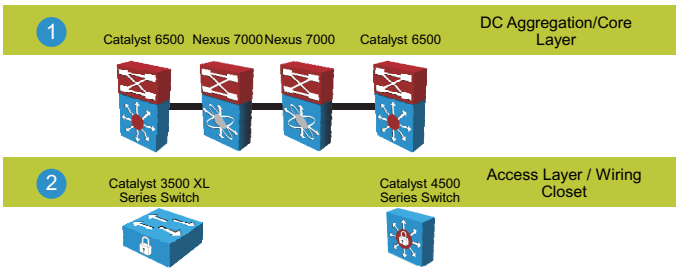


The Cisco Data Center Network Architecture encompasses a number of additional product families. This section discusses the Cisco Catalyst Family of switches, Cisco MDS Family, Cisco ASA adaptive security appliances, and Cisco Wide Area Application Services (WAAS).

https://t.me/learningnets

Cisco Catalyst Switches

1. Services modules in Cisco Catalyst 6500 Series chassis:
   - Firewall Services Module (FWSM)
   - ASA Services Module
   - Cisco ACE Application Control Engine Module
   - Intrusion Detection System (IDSM-2) Services Module
   - Network Analysis Module (NAM-3)
2. Switch fabric in the wiring closet
   - Cisco Catalyst 4900/4500X, 4500, 3750, 3560, 2960 Series Switches



© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco Catalyst switches fill two major roles in the data center environment.

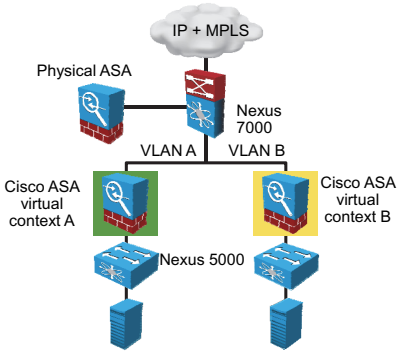
- The services edge is hosted by Cisco Catalyst 6500 Series Switches. The highly scalable Catalyst 6500 Series Switches support a range of high-performance services modules that are deployed in the data center to provide add-on services, such as firewalling, load balancing, intrusion prevention, and network analysis. Some of these services and modules are covered in detail in the later lessons.
- On the campus, the Cisco Catalyst 4900, 4500, 3750, 3560, and 2960 Series Switches could be used in the wiring closet, depending on the density of server ports that are required. The campus aggregation layer could be a pair of Cisco Catalyst 6500 Series Switches in the Virtual Switching System (VSS) mode. In that case, the Cisco Catalyst 6500 Series Switches could also provide the services layer functionality.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Cisco ASA Adaptive Security Appliances

1. Two Cisco ASA adaptive security appliances product families:
   - Standalone appliance—Cisco ASA 5500 Series Adaptive Security Appliances
   - Cisco Catalyst 6500 service blade: Cisco ASA Services Module
2. Main ASA appliance features:
   - Similar to FWSM but runs newest ASA appliance software releases (8.x)
   - Supports EtherChannel (LACP)
   - Up to 32 interfaces per virtual context



© 2012 Cisco and/or its affiliates. All rights reserved.

In addition to the Cisco Catalyst 6500 Series Firewall Services Module (FWSM), Cisco offers two product lines of the Cisco ASA appliance, the flexible and robust firewalling and VPN platform:

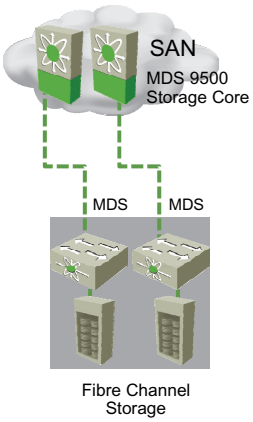
- Cisco ASA 5500 Series Adaptive Security Appliances. This family encompasses standalone appliances Cisco ASA 5505, ASA 5510, ASA 5512-X, ASA 5515-X, ASA 5520, ASA 5525-X, ASA 5540, ASA 5545-X, ASA 5550, ASA 5555-X, and ASA 5585-X Adaptive Security Appliances, that differ in throughput, supported interfaces, and computing power and are therefore targeted at small office, Internet edge, and enterprise data center deployments. Cisco ASA 5585-X is often found in the enterprise data center.
- Cisco ASA Services Module, which provides a natural migration path from the FWSM. Cisco ASA Services Module enhances the Cisco Firewall Services Module (FWSM) functionality by supporting the newest ASA 8.x software releases.

Both the 5500 series and the service blades support a range of data center features, such as Link Aggregation Control Protocol (LACP)-based EtherChannel, and virtualization with up to 32 interfaces per virtual context.

https://t.me/learningnets

Cisco SAN Product Family

- Cisco MDS 9000 Series Multilayer Switches
- Cisco MDS SAN-OS designed for storage area networks (SANs)
- Multiprotocol:
  - Fibre Channel Protocol (FCP)
  - IBM Fibre Connection (FICON)
  - Internet Small Computer System Interface (iSCSI)
  - Fibre Channel over IP (FCIP)
- Fibre Channel over Ethernet (FCoE)
- Inter-VSAN Routing
- Security:
  - Switch and Host Authentication,
  - IP Security for FCIP and iSCSI
  - RBAC
  - Zoning
  - Port Security and Fabric Binding
- QoS



© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco MDS 9500 Series Multilayer Directors are director-class SAN switches that are designed for deployment in large-scale storage networks to enable enterprise clouds and business transformation. Layering a comprehensive set of intelligent features onto a high-performance, protocol-independent switch fabric, the Cisco MDS 9500 Series addresses the requirements of virtualized data center storage environments: high availability, security, scalability, ease of management, and transparent integration of new technologies for extremely flexible data center SAN solutions. Cisco MDS 9500 Series enables seamless deployment of unified fabrics with high-performance Fibre Channel and Fibre Channel over Ethernet (FCoE) connectivity and is compatible with all generations of Cisco MDS 9000 Series Family of switches.

The multilayer architecture of the Cisco MDS 9000 Series Family enables a consistent feature that is set over a protocol-independent switch fabric. They transparently integrate Fibre Channel, FCoE, IBM Fiber Connection (FICON), Internet Small Computer Systems Interface (iSCSI), and Fibre Channel over IP (FCIP) in one system.

Virtual storage area network (VSAN) technology, access control lists (ACLs) for hardware-based intelligent frame processing, and fabric-wide quality of service (QoS) enable migration from SAN islands to enterprise-wide storage networks. Furthermore, Cisco Arbitrated Local Switching feature provides high-performance, predictable, fair switching between all hosts that are attached to the same 8-Gb/s Advanced Fibre Channel switching module and their associated storage devices.

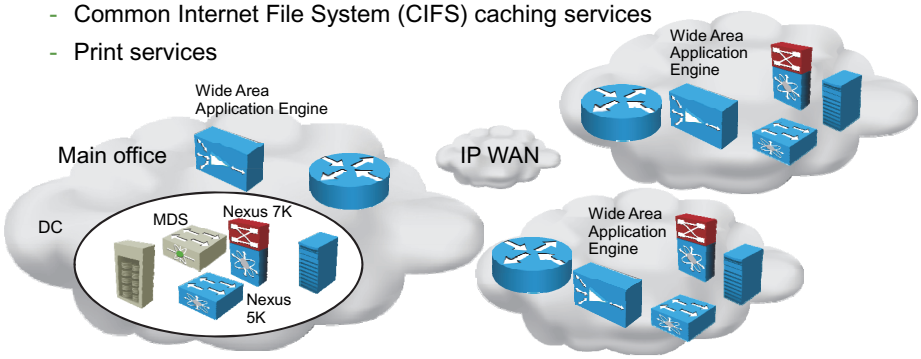
Integration of VSANs into port-level hardware allows any port in a system or fabric to be partitioned to any VSAN. Integrated hardware-based Inter-VSAN Routing (IVR) provides line-rate routing between any ports in a system or fabric without the need for external routing appliances.

In addition to support for services such as VSANs, hardware-enforced zoning, ACLs, per-VSAN role-based access control (RBAC), Cisco SME for tapes and disks, and Cisco TrustSec Fibre Channel link encryption, the Cisco MDS 9000 Series supports a comprehensive security framework consisting of RADIUS and TACACS+, Fibre Channel Security Protocol (FC-SP), Secure File Transfer Protocol (SFTP), Secure Shell (SSH) Protocol, and Simple Network Management Protocol Version 3 (SNMPv3) implementing Advanced Encryption Standard (AES).

https://t.me/learningnets

Cisco WAAS

- Cisco Wide Area Application Services
- Optimization of enterprise operations over the WAN
- Product line with these main functions:
  - Advanced compression
  - Transport file optimizations
  - Common Internet File System (CIFS) caching services
  - Print services



© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco's WAN optimization platforms scale the delivery of an optimal user experience to users, applications, and devices in data center environments, where enterprise branches are connected to the main office data center via an IP WAN network. Cisco WAAS accelerates applications, optimizes bandwidth, provides local hosting of branch IT services, and enables a smooth evolution to cloud-based services.

The Cisco WAVE Appliances: 594, 694, 7541, 7571, and 8541 are second generation WAN optimization solutions, delivering a dramatic increase in performance, with the following benefits for a data center environment:

- Comprehensive WAN optimization from data centers to branches
- Five times the performance with up to 2 Gb/s optimized WAN throughput
- Three times the scale with 150,000 TCP connections

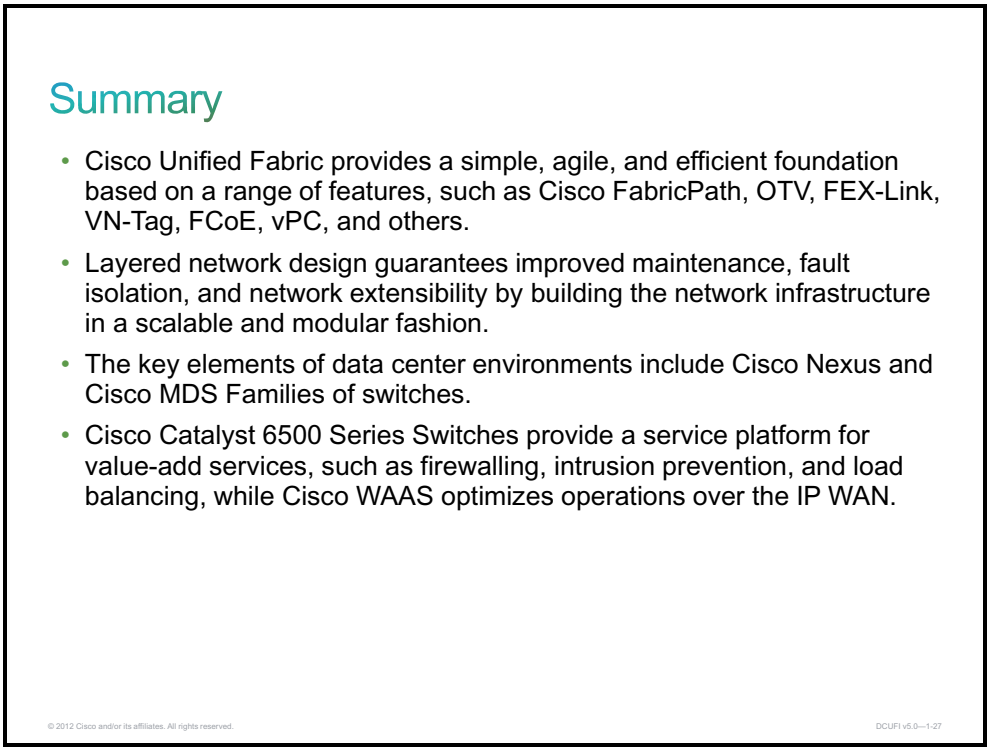
Cisco WAAS optimization is focused on these main areas:

- Advanced compression
- Transport file optimizations
- Common Internet File System (CIFS) caching services
- Print services

https://t.me/learningnets

Summary

This topic summarizes the key points that were discussed in this lesson.



© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 2

Identifying Cisco Nexus Products

Overview

In this lesson, you will learn how the Cisco Nexus Family of products can satisfy the requirements of a unified fabric that is used in the modern data center. You will also learn how to choose chassis, line modules, and fabric extenders that match the requirements of your data center.

Objectives

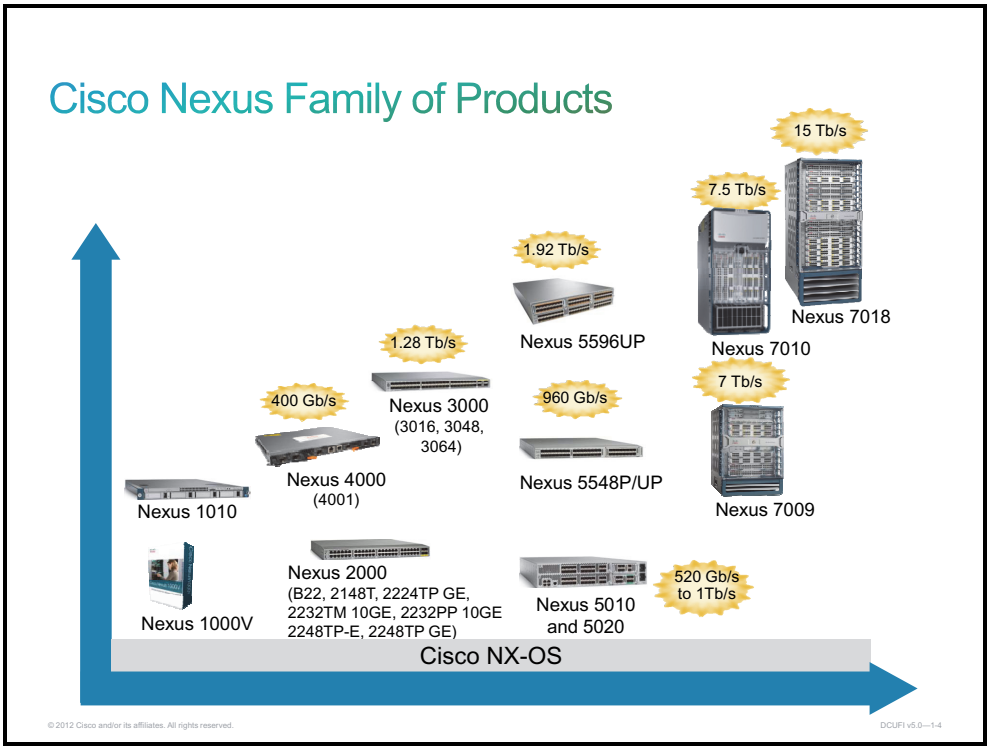
Upon completing this lesson, you will be able to identify the Cisco Nexus Family of products and the important components of the chassis, line modules, and fabric extender. This ability includes being able to meet these objectives:

- Identify the Cisco Nexus Family of products
- Identify the important features and benefits of the I/O modules of the Cisco Nexus 7000 Series Switches
- Identify the important features of Cisco NX-OS that provide high availability and scalability as well as support for Cisco Unified Fabric

https://t.me/learningnets

Cisco Nexus Family of Products

This topic identifies the components of the Cisco Nexus Family of products.



The Cisco Nexus Family of products includes the following switches:

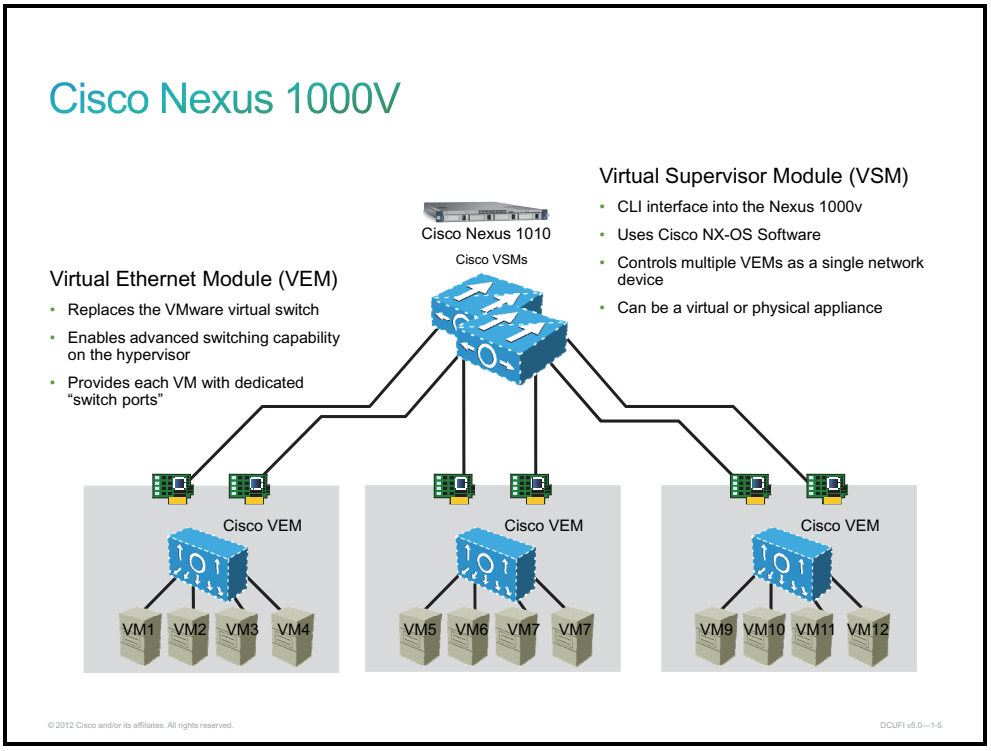
- **Cisco Nexus 1000V Series Switches:** A virtual machine (VM) access switch that is an intelligent software switch implementation for VMware vSphere environments running the Cisco Nexus Operating System (Cisco NX-OS) Software. The Cisco Nexus 1000V Series Switches operate inside the VMware ESX hypervisor and support the Cisco Virtual Network Link (Cisco VN-Link) server virtualization technology to provide the following:
  - Policy-based VM connectivity
  - Mobile VM security and network policy
  - Nondisruptive operational model for server virtualization and networking teams
- **Cisco Nexus 1010 Virtual Services Appliance:** This appliance is a member of the Cisco Nexus 1000V Series Switches and hosts the Cisco Nexus 1000V Virtual Supervisor Module (VSM). It also supports the Cisco Nexus 1000V Network Analysis Module (NAM) Virtual Service Blade (VSB) and provides a comprehensive solution for virtual access switching. The Cisco Nexus 1010 provides dedicated hardware for the Cisco Nexus 1000V VSM, making access switch deployment much easier for the network administrator.
- **Cisco Nexus 2000 Series Fabric Extenders:** A category of data center products that are designed to simplify data center access architecture and operations. The Cisco Nexus 2000 Series Fabric Extenders use the Cisco Fabric Extender Link (Cisco FEX-Link) architecture to provide a highly scalable unified server-access platform across a range of 100-Mb/s Ethernet, Gigabit Ethernet, 10 Gigabit Ethernet, unified fabric, connectivity over copper and optical links, and rack and blade server environments. The Cisco Nexus 2000 Series Fabric Extenders act as remote line cards for the Cisco Nexus 5000 Series Switches (which includes the 5000 and 5500 Platform switches) and the Cisco Nexus 7000 Series Switches.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

- **Cisco Nexus 3000 Series Switches:** The Cisco Nexus 3000 Series Switches are targeted at the high-frequency trading (HFT) market. They support up to 48 fixed, 1 and 10 Gigabit Ethernet enhanced small form-factor pluggable (SFP+) ports and up to 16 fixed quad SFP+ (QSFP+) ports, which allow a smooth transition from 10 Gigabit Ethernet to 40 Gigabit Ethernet. The product family is well suited for financial colocation deployments, delivering features such as latency of less than a microsecond, line-rate Layer 2 and 3 unicast and multicast switching, and the support for 40 Gigabit Ethernet standards technologies.
- **Cisco Nexus 4001I Switch Module for IBM BladeCenter**: The Cisco Nexus 4001I is a blade switch solution for IBM BladeCenter H and HT chassis. This switch provides the server I/O solution that is required for high-performance, scale-out, virtualized and nonvirtualized x86 computing architectures. It is a line-rate, extremely low-latency, nonblocking, Layer 2, 10 Gigabit Ethernet blade switch that is fully compliant with the International Committee for Information Technology (INCITS) Fibre Channel over Ethernet (FCoE) and IEEE 802.1 Data Center Bridging (DCB) standards. This switch is one of the Cisco Nexus 4000 Series Switches.
- **Cisco Nexus 5000 Series Switches (including the Cisco Nexus 5000 Platform and 5500 Platform switches**: A Series of line-rate, low-latency, lossless 10 Gigabit Ethernet, and FCoE switches for data center applications. The Cisco Nexus 5000 Series Switches are designed for data centers that are transitioning to 10 Gigabit Ethernet as well as data centers that are ready to deploy a unified fabric that can manage LAN, SAN, and server clusters. This capability provides networking over a single link, with dual links used for redundancy. Some of the switches included in this series are the Cisco Nexus 5000 Platform switches, 5010 and 5020, and the Cisco Nexus 5550 Platform switches, 5548UP, 5548P, and 5596UP as noted in the figure.
- **Cisco Nexus 7000 Series Switches**: A modular data center-class switch that is designed for highly scalable 10 Gigabit Ethernet networks with a fabric architecture that scales beyond 15 terabits per second (Tb/s). The switch is designed to deliver continuous system operation and virtualized services. The Cisco Nexus 7000 Series Switches incorporate significant enhancements in design, power, airflow, cooling, and cabling. The 10-slot chassis has front-to-back airflow making it a good solution for hot aisle and cold aisle deployments. The 18-slot chassis uses side-to-side airflow to deliver high density in a compact form factor. The chassis in this series include Cisco Nexus 7000 9-Slot, 10-Slot, and 18-Slot Switch chassis, also referred to as Cisco Nexus 7009, 7010, and 7018 chassis as seen in the figure.

https://t.me/learningnets



Cisco Nexus 1000V Series Switches deliver multitenant services by adding virtualization intelligence to the data center network. These softswitches are integrated with VMware vCloud Director. They are built to scale for cloud networks, with support for Virtual Extensible LAN (VXLAN). This series addresses the requirements for scalable LAN segmentation and helps to enable broader VM mobility.

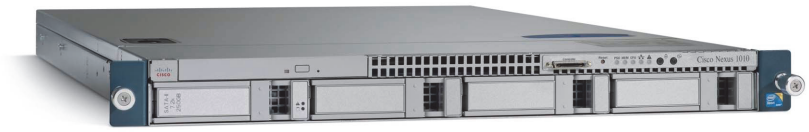
There are two components that are part of the Cisco Nexus 1000V implementation:

- **Virtual Ethernet Module (VEM)**, a software switch that is embedded in the hypervisor.
- **Virtual Supervisor Module (VSM)**, which manages networking policies and quality of service (QoS) for VMs in concert with the VEM. The VSM can control several VEMs, with the VEMs forming a switch domain that is in the same virtual data center.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Cisco Nexus 1010 Virtual Services Appliance

- Dedicated appliance hosting
  - Cisco Nexus 1000V VSM
  - Virtual service blade (VSB)
- Cisco Nexus 1000V Network Analysis Module (NAM) VSB



© 2012 Cisco and/or its affiliates. All rights reserved.

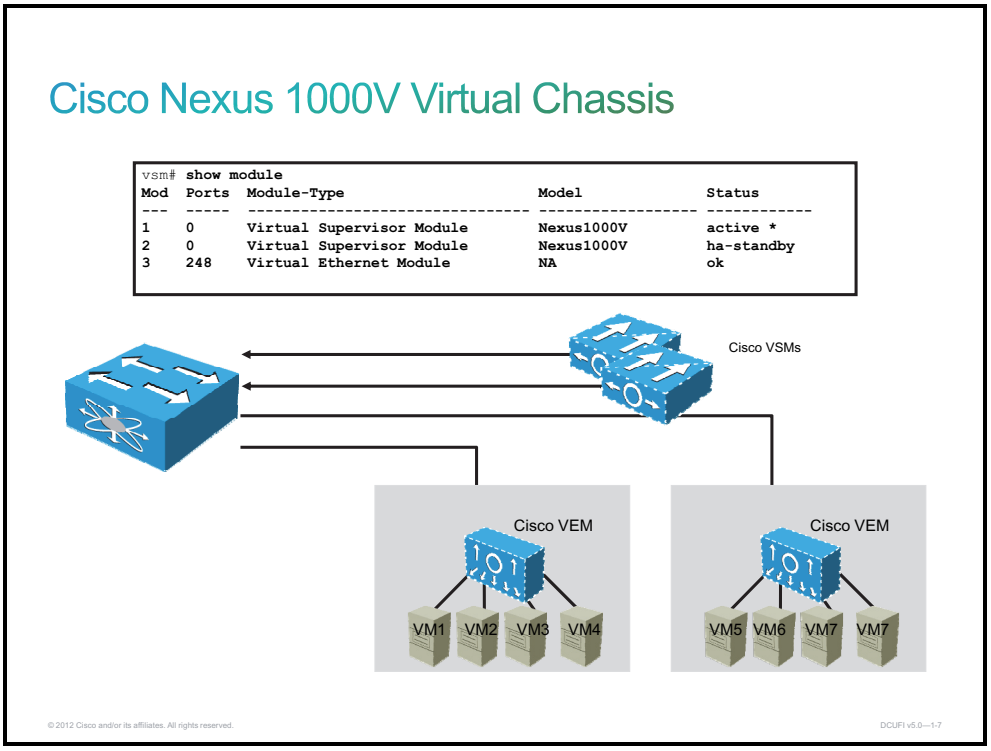
The Cisco Nexus 1010 Virtual Services Appliance server is used as an appliance to host the Cisco 1000V VSM.

It brings several benefits into the virtual switching environment:

- Offloads VSM installation and management to the network team
- Has no need for a VMware ESX license
- Installs VSM the same way as a standard Cisco switch

In addition to VSM, Cisco Nexus 1010 can be used for hosting other Cisco virtual appliances such as Cisco Virtual Security Gateway (VSG), Cisco Virtual Wide Area Application Services (vWAAS), and virtual service blades (VSBs).

https://t.me/learningnets

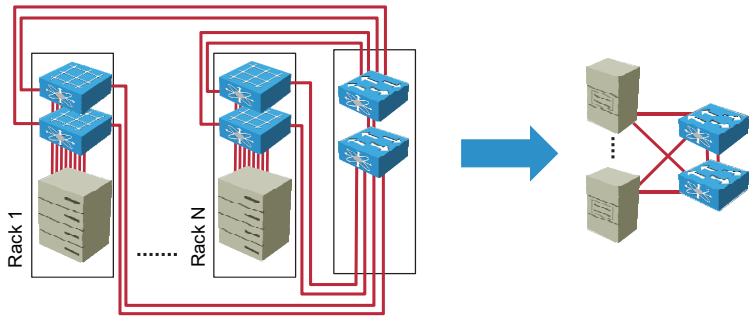


The Cisco Nexus 1000V is effectively a “virtual chassis.” It is modular, and ports can be either physical or virtual. The servers are modules on the switch, with each physical network interface virtualization (NIV) port on a module being a physical Ethernet port. Modules 1 and 2 are reserved for the VSM, with the first server or host automatically being assigned to the next available module number. The ports to which the virtual network interface card (vNIC) interfaces connect are virtual ports on the Cisco Nexus 1000V, where they are assigned a global number.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Cisco Nexus 2000 Series Fabric Extenders

- Serve as remote I/O modules of a Cisco Nexus 5000 or 5500 Platform switch or a 7000 Series switch
- Are managed and configured from parent switch
- Together, parent switches and Cisco Nexus 2000 Series Fabric Extenders combine benefits of ToR cabling with EoR management



© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 2000 Series Fabric Extenders behave as remote line cards for a parent Cisco Nexus 5000 or 5500 Platform switch or a Cisco Nexus 7000 Series switch. The fabric extenders are essentially extensions of the parent Cisco Nexus switch fabric, with the fabric extenders and the parent Cisco Nexus switch together forming a distributed modular system. Working with the Cisco Nexus Family of switches, the Cisco Nexus 2000 Series Fabric Extenders extend the capabilities and benefits that are offered by the parent Cisco Nexus switch.

This architecture enables physical topologies with the flexibility and benefits of both top-of-rack (ToR) and end-of-row (EoR) deployments.

Cisco Nexus 2000 Series Fabric Extenders connect to a parent Cisco Nexus switch through their fabric links using CX1 copper cable, short-reach or long-reach optics, and the cost-effective optical Cisco Fabric Extender Transceivers.

https://t.me/learningnets

Cisco Nexus 2000 Series Fabric Extender Models

Model	Nexus B22	Nexus 2224	Nexus 2232	Nexus 2248
Parent switches	• Nexus 5010/5020 • Nexus 5548P/UP • Nexus 5596UP	• Nexus 5010/5020 • Nexus 5548P/UP • Nexus 5596UP • Nexus 7000 (only for models 2224TP, 2248TP, 2232PP)		
Interfaces	• 10GBASE-KR internal connectors	• 24 Fixed 100 Megabit or 1 Gigabit Ethernet ports • 2 Fixed 10 Gigabit Ethernet* uplinks	• 32 1 or 10 Gigabit Ethernet or FCoE • 8 10 Gigabit Ethernet DCB or FCoE uplinks	• 48 Fixed 100 Megabit or 1 Gigabit Ethernet ports • 4 Fixed 10 Gigabit Ethernet uplinks
Description	Model B22HP dedicated to: • HP BladeSystem c3000 enclosure • HP BladeSystem c7000 enclosure		Nexus 2232PP suitable for migration from Gigabit Ethernet to 10 Gigabit Ethernet and unified fabric environments. It supports FCoE and DCB.	2248TP-E model provides enhancements for large-volume databases, distributed storage, and video editing

© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 2000 Series Fabric Extenders comprise a category of data center products that are designed to simplify data center access architecture and operations. The Cisco Nexus 2000 Series provides a scalable unified server-access platform across a range of 100 Megabit Ethernet, Gigabit Ethernet, 10 Gigabit Ethernet, unified fabric, connectivity over copper and optical links, rack, and blade server environments. The platform supports traditional Gigabit Ethernet while allowing transparent migration to 10 Gigabit Ethernet, VM-aware unified fabric technologies.

The Cisco Nexus 2000 Series offers front-to-back cooling, compatibility with data center hot-aisle and cold-aisle designs, placement of all switch ports at the rear of the unit in close proximity to server ports, and accessibility of all user-serviceable components from the front panel. The Cisco Nexus 2000 Series has redundant hot-swappable power supplies and a hot-swappable fan tray with redundant fans. The Cisco Nexus 2000 Series has two types of ports: ports for end-host attachment and uplink ports.

The family comprises these models:

- **Cisco Nexus B22HP Fabric Extender** is a blade fabric extender for HP, and offers 16 x 10GBASE-KR internal host interfaces and 8 x 10 Gigabit Ethernet fabric interfaces SFP+.
- **Cisco Nexus 2224TP, 2248TP,** and **2248TP-E** Fabric Extenders provide port density options for highly scalable 100 Megabit Ethernet and Gigabit Ethernet connectivity. The Cisco Nexus 2232PP Fabric Extender provides ease of migration from Gigabit Ethernet to 10 Gigabit Ethernet while supporting highly scalable 10 Gigabit environments.
- **Cisco Nexus 2248TP-E Fabric Extender** is a general-purpose 1 Gigabit Ethernet fabric extender with enhancements that target workloads such as large-volume databases, distributed storage, and video editing. Just like the Cisco Nexus 2248TP, the Cisco Nexus 2248TP-E supports 48 100/1000BASE-T host-facing ports and four 10 Gigabit Ethernet fabric interfaces.
- **Cisco Nexus 2232PP Fabric Extender** is the ideal platform for migration from Gigabit Ethernet to 10 Gigabit Ethernet and unified fabric environments. It supports FCoE and a set of network technologies that are known collectively as Data Center Bridging (DCB) that

https://t.me/learningnets

increase the reliability, efficiency, and scalability of Ethernet networks. These features allow the switches to support multiple traffic classes over a lossless Ethernet fabric, thus enabling consolidation of LAN, SAN, and cluster environments.

- **Cisco Nexus 2232TM Fabric Extender** supports scalable 1/10GBASE-T environments, ease of migration from 1GBASE-T to 10GBASE-T, and effective reuse of existing structured cabling. It comes with an uplink module that supports eight 10 Gigabit Ethernet fabric interfaces. The Nexus 2232TM supports DCB.

Cisco Nexus 3000 Series Switches

- Targeted at financial collocation deployments
- Ultra-low latency
- Line-rate traffic throughput (both Layer 2 and 3) on all ports
- Support for advanced unicast and multicast routing protocols

Model	Nexus 3016	Nexus 3048	Nexus 3064
Photo			
Interfaces	• 16 QSFP ports; each supports native 40 Gigabit Ethernet or 4 x 10 Gigabit Ethernet	• 48 100/1000-Mb/s ports • Four 1/10-Gb/s uplink ports	• 48 SFP ports supporting 1 and 10 Gigabit Ethernet • 4 QSFP ports; each supports native 40 Gigabit Ethernet or 4 x 10 Gigabit Ethernet
Performance	• 1.28-Tb/s switching capacity • Forwarding rate 960 mpps	• 176 Gb/s switching • capacity • 132 mpps forwarding rate	• 1.28-Tb/s switching capacity • Forwarding rate of 960 mpps

© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 3000 Series Switches include high-performance, high-density, ultralow-latency Ethernet switches. They provide line-rate Layer 2 and Layer 3 switching. The switches run the Cisco NX-OS Software, providing customers with comprehensive features and functionality. The switches are optimized for low latency and low-power consumption. They are targeted at financial colocation deployments that require support for comprehensive unicast and multicast routing protocol features at ultralow latencies.

The Cisco Nexus 3000 Series supports a wide variety of 1, 10, and 40 Gigabit Ethernet connectivity options. The 1 and 10 Gigabit Ethernet connectivity is achieved using SFP+ transceivers in the first 48 ports, and 40 Gigabit Ethernet connectivity is achieved by using QSFP+ transceivers.

QSFP+ technology allows smooth transition from 10- to 40-Gigabit Ethernet infrastructures in data centers. The Cisco Nexus 3000 Series supports connectivity over copper and fiber cables, providing excellent physical-layer flexibility. For low-cost cabling, copper-based 40-Gb/s Twinax cables can be used, and for longer cable reaches, short-reach optical transceivers are excellent.

Connectivity can be established from the QSFP ports to an upstream 10 Gigabit Ethernet switch using a splitter cable that has a QSFP transceiver on one end and four SFP+ transceivers on the other end. Similar capability can be achieved using optical transceivers by procuring third-party fiber splitters.

The Cisco Nexus 3016 Switch offers 16 QSFP+ ports, while the Cisco Nexus 3064 Switch provides four QSFP+ ports in addition to 48 SFP ports that support 1 and 10 Gigabit Ethernet.

https://t.me/learningnets

Cisco Nexus 4000 Series Switches

- Currently only one model: Cisco Nexus 4001I
- Blade switch module for IBM BladeCenter H and HT chassis
  - High-performance, scale-out, virtualized and nonvirtualized architectures
  - Line-rate, low-latency, nonblocking
- Interfaces:
  - 14 x 10 Gigabit Ethernet server-facing downlinks
    - Autosensing; can also operate in Gigabit Ethernet mode
  - 6 x 10 Gigabit Ethernet uplinks
    - Autosensing; can also operate in Gigabit Ethernet mode
  - 2 x management ports: one external 10/100/1000BASE-T port and one internal port for Advanced Management Module (AMM) connectivity

Cisco Nexus 4001I

© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 4001I Switch Module for IBM BladeCenter is a blade switch solution for IBM BladeCenter H and HT chassis, providing the server I/O solution that is required for high-performance, scale-out, virtualized and nonvirtualized x86 computing architectures. It is a line-rate, extremely low-latency, nonblocking, Layer 2, 10 Gigabit Ethernet blade switch that is fully compliant with the INCITS Fibre Channel over Ethernet (FCoE) and IEEE 802.1 DCB standards.

At the center of the Cisco Nexus 4001I is the unified switch ASIC, a new, purpose-built, high-performance, line-rate switch ASIC that delivers extremely low and consistent latency across all packet sizes independent of the configured networking features. The unified switch ASIC supports standard Ethernet as well as priority flow control (PFC), and Enhanced Transmission Selection (ETS), which is required for lossless Ethernet transmission. LAN and SAN networking protocols are delivered through Cisco NX-OS Software. Using the combination of the unified switch ASIC and Cisco NX-OS, the Cisco Nexus 4001I extends the benefits of the Cisco Nexus Family of data center switches to blade servers.

The Cisco Nexus 4001I Switch Module for IBM BladeCenter offers these features:

- Fourteen fixed 10 Gigabit Ethernet server-facing downlinks (with autosensing ports and can also operate in Gigabit Ethernet mode)
- Six fixed 10 Gigabit Ethernet uplinks (with autosensing ports and can also operate in Gigabit Ethernet mode)
- Two management ports: one external 10/100/1000BASE-T port and one internal port for Advanced Management Module (AMM) connectivity
- One RS-232 serial console port

The Cisco Nexus 4001I inserts into the high-speed slot of the IBM BladeCenter H or HT chassis. The IBM BladeCenter H and HT chassis are designed to support up to four Cisco Nexus 4001I switches per chassis.

https://t.me/learningnets

Cisco Nexus 5500 Platform Switches

- Manages 2000 Series Fabric Extenders as virtual line cards
- Unified port technology enables an interface to be configured as either:
  - 1 and 10 Gigabit Ethernet
  - Fibre Channel over Ethernet (FCoE)
  - 1-, 2-, 4-, or 8-Gigabit native Fibre Channel port
- License-based software packaging
  - Default system has Layer 2 security and management features
  - Licensed features: Layer 3 routing, multicast, and enhanced Layer 2 (FabricPath)

Model	Nexus 5548	Nexus 5596
Photo		
Interfaces	48-port switch: • 32 fixed ports, 1 and 10 Gigabit Ethernet, FCoE, or DCB • 1 expansion module slot	96-port switch: • 48 fixed ports, 1 and 10 Gigabit Ethernet, FCoE, or FC (unified ports) • 3 expansion module slots

© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 5500 Platform switches are the second generation of access switches for 10 Gigabit Ethernet connectivity. Compared with the Cisco Nexus 5000 Platform switches, the 5500 Platform introduces a license-based software packaging model. The default system software includes most Cisco Nexus 5000 Platform features, such as Layer 2 security and management features. Licensed features include: Layer 3 routing, IP multicast, and enhanced Layer 2 (Cisco FabricPath).

Cisco Nexus 5500 Platform switches offer these features:

- **Unified port technology:** The unified ports allow you to configure a physical port on a Cisco Nexus 5500 Platform switch as a 1 and 10 Gigabit Ethernet, FCoE, or 1-, 2-, 4-, or 8-Gigabit native Fibre Channel port.
- **High-density and high-availability:** The Cisco Nexus 5548P Switch provides 48 1 and 10 Gigabit Ethernet ports in 1 rack unit (1 RU), and the upcoming Cisco Nexus 5596UP Switch provides a density of ninety-six 1 and 10 Gigabit Ethernet ports in 2 RUs. The switches in the Cisco Nexus 5500 Platform are designed with redundant and hot-swappable power and fan modules that can be accessed from the front panel, where status lights offer an at-a-glance view of switch operation. To support efficient data center hot- and cold-aisle designs, front-to-back cooling is used for consistency with server designs.
- **Nonblocking line-rate performance:** All the 10 Gigabit Ethernet ports on the Cisco Nexus 5500 Platform switches can manage packet flows at wire speed. The absence of resource sharing helps ensure the best performance of each port regardless of the traffic patterns on other ports. The Cisco Nexus 5548P Switch can have 48 Ethernet ports, at 10 Gb/s, sending packets simultaneously without any effect on performance, offering true 960-Gb/s bidirectional bandwidth. The upcoming Cisco Nexus 5596UP Switch can have 96 Ethernet ports at 10 Gb/s, offering true 1.92-Tb/s bidirectional bandwidth.

https://t.me/learningnets

- **Low latency**: The cut-through switching technology that is used in the ASICs of the Cisco Nexus 5500 Platform switches enables the product to offer a low latency of 2 microsec, which remains constant regardless of the size of the packet that is being switched. This latency was measured on fully configured interfaces, with access control lists (ACLs), quality of service (QoS), and all other data path features turned on. The low latency on the Cisco Nexus 5500 Platform switches together with a dedicated buffer per port and the congestion management features make the Cisco Nexus 5500 Platform an excellent choice for latency-sensitive environments.
- **Single-stage fabric**: The crossbar fabric on the Cisco Nexus 5500 Platform switches is implemented as a single-stage fabric, thus eliminating any bottleneck within the switches. Single-stage fabric means that a single crossbar fabric scheduler has complete visibility into the entire system and can therefore make optimal scheduling decisions without building congestion within the switch. With a single-stage fabric, the congestion becomes exclusively a function of your network design; the switch does not contribute to it.
- **Congestion management**: Keeping latency low is not the only critical element for a high-performance network solution. Servers tend to generate traffic in bursts, and when too many bursts occur at the same time, a short period of congestion occurs. Depending on how the burst of congestion is smoothed out, the overall network performance can be affected. The Cisco Nexus 5500 Platform offers a complete range of congestion management features to reduce congestion. These features address congestion at different stages and offer granular control over the performance of the network.
  - **Virtual output queues**: The Cisco Nexus 5500 Platform implements virtual output queues (VOQs) on all ingress interfaces, so that a congested egress port does not affect traffic that is directed to other egress ports. Every IEEE 802.1p class of service (CoS) uses a separate VOQ in the Cisco Nexus 5500 Platform architecture, resulting in a total of eight VOQs per egress on each ingress interface, or a total of 384 VOQs per ingress interface on the Cisco Nexus 5548P Switch, and a total of 768 VOQs per ingress interface on the Cisco Nexus 5596UP Switch. The extensive use of VOQs in the system helps ensure high throughput on a per-egress, per-CoS basis. Congestion on one egress port in one CoS does not affect traffic that is destined for other classes of service or other egress interfaces. This ability avoids head-of-line (HOL) blocking, which would otherwise cause congestion to spread.
  - **Separate egress queues for unicast and multicast**: Traditionally, switches support eight egress queues per output port, each servicing one IEEE 802.1p CoS. The Cisco Nexus 5500 Platform switches increase the number of egress queues by supporting eight egress queues for unicast and 8 egress queues for multicast. This support allows separation of unicast and multicast that are contending for system resources within the same CoS and provides more fairness between unicast and multicast. Through configuration, the user can control the amount of egress port bandwidth for each of the 16 egress queues.
  - **Lossless Ethernet with priority flow control (PFC)**: By default, Ethernet is designed to drop packets when a switching node cannot sustain the pace of the incoming traffic. Packet drops make Ethernet very flexible in managing random traffic patterns that are injected into the network. However, they effectively make Ethernet unreliable and push the burden of flow control and congestion management up to a higher level in the network stack.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

PFC offers point-to-point flow control of Ethernet traffic that is based on IEEE 802.1p CoS. With a flow-control mechanism in place, congestion does not result in drops, transforming Ethernet into a reliable medium. The CoS granularity then allows some classes of service to gain a no-drop, reliable, behavior while allowing other classes to retain traditional best-effort Ethernet behavior. The no-drop benefits are significant for any protocol that assumes reliability at the media level, such as FCoE.

- **Explicit congestion notification (ECN) marking**: ECN is an extension to TCP/IP. It is defined in RFC 3168. ECN allows end-to-end notification of network congestion without dropping packets. Traditionally, TCP detects network congestion by observing dropped packets. When congestion is detected, the TCP sender takes action by controlling the flow of traffic. However, dropped packets can sometimes lead to long TCP timeouts and consequent loss of throughput. The Cisco Nexus 5500 Platform switches can set a mark in the IP header so that instead of dropping a packet, it sends a signal impending congestion. The receiver of the packet echoes the congestion indicator to the sender, which must respond as though congestion had been indicated by packet drops.

- **FCoE**: FCoE is a standards-based encapsulation of Fibre Channel frames into Ethernet frames. By implementing FCoE, the Cisco Nexus 5500 Platform switches enable storage I/O consolidation in addition to Ethernet.

- **NIV architecture**: The introduction of blade servers and server virtualization has increased the number of access-layer switches that need to be managed. In both cases, an embedded switch or softswitch requires separate management. NIV enables a central switch to create an association with the intermediate switch, whereby the intermediate switch will become the data path to the central forwarding and policy enforcement under the control of the central switch. This scheme enables both a single point of management and a uniform set of features and capabilities across all access-layer switches.

  One critical implementation of NIV in the Cisco Nexus 5000 and 5500 Platforms is the Cisco Nexus 2000 Series Fabric Extenders and their deployment in data centers. A Cisco Nexus 2000 Series Fabric Extender behaves as a virtualized remote I/O module, enabling the Cisco Nexus 5500 Platform switches to operate as a virtual modular chassis.

- **IEEE 1588 Precision Time Protocol (PTP)**: In financial environments, particularly high-frequency trading environments, transactions occur in less than a millisecond. For accurate application performance monitoring and measurement, the systems supporting electronic trading applications must be synchronized with extremely high accuracy (to less than a microsecond). IEEE 1588 is designed for local systems that require very high accuracy beyond that which is attainable using Network Time Protocol (NTP). The Cisco Nexus 5500 Platform supports IEEE 1588 boundary clock synchronization. In other words, a Cisco Nexus 5500 Platform switch will run PTP and synchronize to an attached master clock, and the boundary clock will then act as a master clock for all attached slaves. The Cisco Nexus 5500 platform also supports packet time stamping by including the IEEE 1588 time stamp in the Encapsulated Remote Switched Port Analyzer (ERSPAN) header.

- **Cisco FabricPath and TRILL**: Existing Layer 2 networks that are based on Spanning Tree Protocol (STP) have a number of challenges to overcome. These challenges include suboptimal path selection, underutilized network bandwidth, control-plane scalability, and slow convergence. Although enhancements to STP and features such as Cisco virtual port channel (vPC) technology help mitigate some of these limitations, these Layer 2 networks lack fundamentals that limit their scalability.

https://t.me/learningnets

Cisco FabricPath and Transport Interconnection of Lots of Links (TRILL) are two emerging solutions for creating scalable and highly available Layer 2 networks. Cisco Nexus 5500 Platform hardware is capable of switching packets that are based on Cisco FabricPath headers or TRILL headers. This capability enables customers to deploy scalable Layer 2 networks with native Layer 2 multipathing.

- **Layer 3**: The design of the access layer varies depending on whether Layer 2 or Layer 3 is used at the access layer. The access layer in the data center is typically built at Layer 2. Building at Layer 2 allows better sharing of service devices across multiple servers and allows the use of Layer 2 clustering, which requires the servers to be next to Layer 2. In some design, such as two-tier designs, the access layer may be Layer 3, although this may not imply that every port on these switches is a Layer 3 port. The Cisco Nexus 5500 Platform can operate in Layer 3 mode with the addition of a routing module.
- **Hardware-level I/O consolidation**: The Cisco Nexus 5500 Platform ASICs can transparently forward Ethernet, Fibre Channel, FCoE, Cisco FabricPath, and TRILL, providing true I/O consolidation at the hardware level. The solution that is adopted by the Cisco Nexus 5500 Platform reduces the costs of consolidation through a high level of integration in the ASICs. The result is a full-featured Ethernet switch and a full-featured Fibre Channel switch that are combined into one product.

Cisco Nexus 5000 Platform Switches

- Manages Cisco Nexus 2000 Series Fabric Extenders as virtual line cards
- Limited capabilities compared to 5500 switches
- Features:
  - Layer 2 switching (no Layer 3 routing), Layer 2 QoS
  - Fibre Channel, FCoE, and Data Center Bridging
  - High availability: ISSU, 1:1 power redundancy, N:1 fan module redundancy

Model	Nexus 5010	Nexus 5020
Photo		
Interfaces	28-port Switch • 20 fixed 10 Gigabit Ethernet and FCoE SFP+ ports • 1 expansion module slot	56-port Switch: • 40 fixed 10 Gigabit Ethernet and FCoE SFP+ ports • 2 expansion module slots

© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 5000 Platform switches use a cut-through architecture that supports line-rate 10 Gigabit Ethernet on all ports, maintaining a consistent low latency independent of packet size and service that is enabled. The switches support a set of network technologies that are known collectively as IEEE Data Center Bridging (DCB) that increases reliability, efficiency, and scalability of Ethernet networks. These features allow the switches to support multiple traffic classes over a lossless Ethernet fabric, thus enabling consolidation of LAN, SAN, and cluster environments. The ability to connect FCoE to native Fibre Channel protects existing storage system investments, which dramatically simplifies in-rack cabling.

The Cisco Nexus 5000 Platform switches integrate with multifunction adapters called converged network adapters (CNAs) to provide Cisco Unified Fabric convergence. The adapters combine the functions of Ethernet network interface controller (NICs) and Fibre

https://t.me/learningnets

Channel host bus adapters (HBAs). This functionality makes the transition to a single, unified network fabric transparent and consistent with existing practices, management software, and operating system drivers. The switch family is compatible with integrated transceivers and twinax cabling solutions that deliver cost-effective connectivity for 10 Gigabit Ethernet to servers at the rack level. This compatibility eliminates the need for expensive optical transceivers.

Cisco Nexus 5020 Switch

The Cisco Nexus 5020 Switch is a 2-rack unit (2-RU), 10 Gigabit Ethernet, FCoE, and Fibre Channel switch that is built to provide 1.04 Tb/s throughout with very low latency. It has 40 fixed 10 Gigabit Ethernet and FCoE SFP+ ports. The first 16 fixed ports support both 10 Gigabit Ethernet and Gigabit Ethernet in hardware. Two expansion module slots can be configured to support up to 12 additional 10 Gigabit Ethernet and FCoE SFP+ ports, or up to 16 Fibre Channel switch ports, or a combination of both. The switch has a serial console port and an out-of-band (OOB) 10/100/1000-Megabit Ethernet management port. The switch is powered by 1+1 redundant, hot-pluggable power supplies and 4+1 redundant, hot-pluggable fan modules to provide highly reliable front-to-back cooling.

Cisco Nexus 5010 Switch

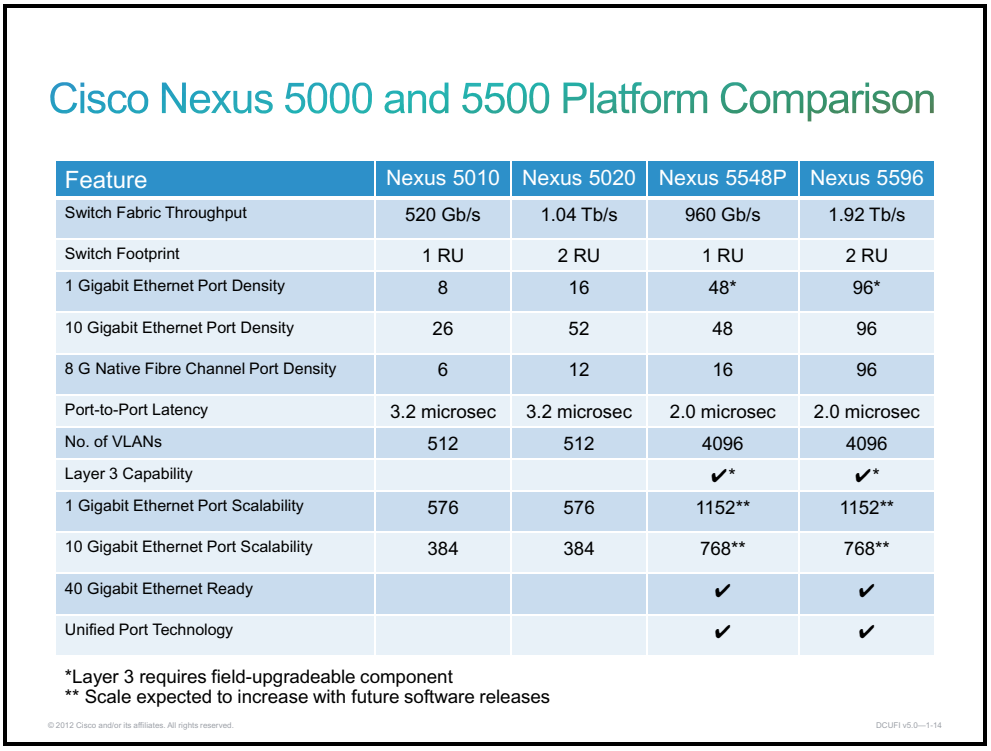
The Cisco Nexus 5010 Switch is a 1-RU, 10 Gigabit Ethernet, FCoE, and Fibre Channel switch providing more than 520-Gb/s throughput with very low latency. It has 20 fixed 10 Gigabit Ethernet and FCoE SFP+ ports. The first eight fixed ports are dual speed, supporting both 10 Gigabit Ethernet and Gigabit Ethernet in hardware. One expansion module slot can be configured to support up to six additional 10 Gigabit Ethernet and FCoE SFP+ ports, eight 4-Gb/s SFP Fibre Channel switch ports, or six 8-Gb/s SFP+ Fibre Channel switch ports, or a combination of four additional 10 Gigabit Ethernet and FCoE SFP+ ports with four additional 4-, 2-, or 1-Gb/s Fibre Channel switch ports. The switch has a serial console port and OOB 10/100/1000-Mb/s Ethernet management port. The switch is powered by 1+1 redundant, hot-pluggable power supplies and 1+1 redundant, hot-pluggable fan modules to provide highly reliable front-to-back cooling.

Expansion Modules for the Cisco Nexus 5000 Platform Switches

The Cisco Nexus 5000 Platform switches support the following expansion modules:

- Ethernet module providing six 10 Gigabit Ethernet and FCoE ports using SFP+ interfaces
- Fibre Channel plus Ethernet module providing four 10 Gigabit Ethernet and FCoE ports using SFP+ interfaces, and four ports of 4-, 2-, or 1-Gb/s native Fibre Channel connectivity using SFP interfaces
- Fibre Channel module that provides eight ports of 4-,2-, or 1-Gb/s native Fibre Channel using SFP interfaces for transparent connectivity to existing Fibre Channel networks
- Fibre Channel module that provides six ports of 8-, 4-,2-, or 1-Gb/s native Fibre Channel using SFP+ interfaces for transparent connectivity with existing Fibre Channel networks

https://t.me/learningnets



The table in the figure describes the differences between the Cisco Nexus 5000 and 5500 Platform switches. The port counts are based on 24 Cisco Nexus 2000 Series Fabric Extenders per Cisco Nexus 5500 Platform switch.

The latency improvement of a Cisco Nexus 5500 Platform switch over a Cisco Nexus 5000 Platform switch results from its architecture. The Cisco Nexus 5500 Platform is built around two custom components: a unified crossbar fabric and a unified port controller ASIC. Each Cisco Nexus 5500 Platform switch contains a single unified crossbar fabric ASIC and multiple unified port controllers to support fixed ports and expansion modules within the switch. The unified port controller provides an interface between the unified crossbar fabric ASIC and the network media adapter and makes forwarding decisions for Ethernet, Fibre Channel, and FCoE frames. The ASIC supports the overall cut-through design of the switch by transmitting packets to the unified crossbar fabric before the entire payload has been received. The unified crossbar fabric ASIC is a single-stage, nonblocking crossbar fabric that is capable of meshing all ports at wire speed. The unified crossbar fabric offers superior performance by implementing QoS-aware scheduling for unicast and multicast traffic. Moreover, the tight integration of the unified crossbar fabric with the unified port controllers helps ensure low-latency, lossless fabric for ingress interfaces requesting access to egress interfaces.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Cisco Nexus 7000 Series Switches

- 15+ Tbp/s system
- DCB and FCoE
- Modular operating system
- Device virtualization
- Cisco TrustSec solution
- Continuous operations

Model	Cisco Nexus 7009*	Cisco Nexus 7010**	Cisco Nexus 7018***
Slots	7 I/O + 2 supervisors	8 I/O + 2 supervisors	16 I/O + 2 supervisors
Height	14 RU	21 RU	25 RU
BW / Slot Fab-1	N/A	230 Gb/s per slot	230 Gb/s per slot
BW / Slot Fab-2	550 Gig / slot	550 Gb/s slot	550 Gb/s slot

*7009 = Cisco Nexus 7000 9-Slot Switch
**7010 = Cisco Nexus 7000 10-Slot Switch
***7018 = Cisco Nexus 7000 18-Slot Switch

© 2012 Cisco and/or its affiliates. All rights reserved.

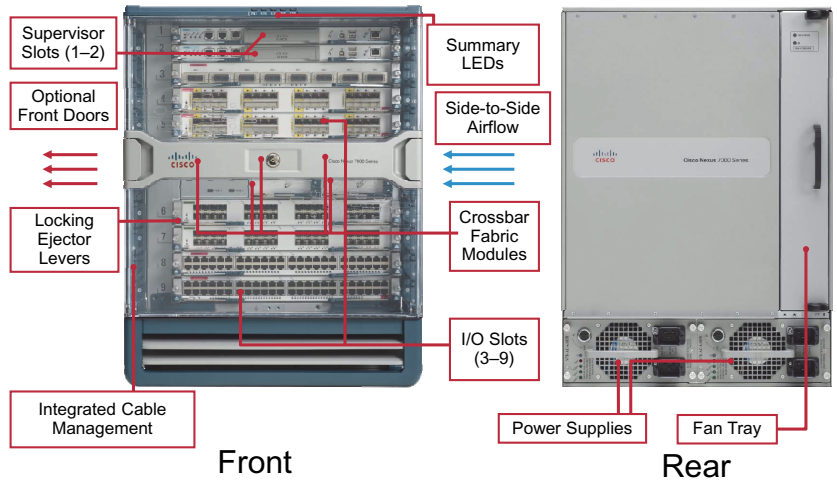
The Cisco Nexus 7000 Series Switches offer a modular data center-class product that is designed for highly scalable 10 Gigabit Ethernet networks with a fabric architecture that scales beyond 15 Tb/s. The Cisco Nexus 7000 Series provides integrated resilience that is combined with features that are optimized specifically for the data center for availability, reliability, scalability, and ease of management.

The Cisco Nexus 7000 Series Switches run the Cisco NX-OS Software to deliver an abundant set of features with nonstop operation.

- This series features front-to-back airflow with 10 front-accessed vertical module slots and an integrated cable management system that facilitates installation, operation, and cooling in both new and existing facilities.
- There are 18 front-accessed module slots with side-to-side airflow in a compact horizontal form factor with purpose-built integrated cable management easing operation and reducing complexity.
- Designed for reliability and maximum availability, all interface and supervisor modules are accessible from the front. Redundant power supplies, fan trays, and fabric modules are accessible completely from the rear to ensure that cabling is not disrupted during maintenance.
- The system uses dual dedicated supervisor modules and fully distributed fabric architecture. There are five rear-mounted fabric modules, which combined with the chassis midplane, deliver up to 230 Gb/s per slot for 4.1-Tb/s of forwarding capacity in the 10-slot form factor, and 7.8-Tb/s in the 18-slot form factor using the Cisco Nexus 7000 Series Fabric-1 Modules. Migrating to the Fabric-2 Modules increases the bandwidth per slot to 550 Gb/s. This migration increases the forwarding capacity on the 10-slot form factor to 9.9-Tb/s and on the 18-slot form factor to 18.7-Tb/s.
- The midplane design supports flexible technology upgrades as your needs change and provides ongoing investment protection.

https://t.me/learningnets

Cisco Nexus 7000 9-Slot Switch Chassis

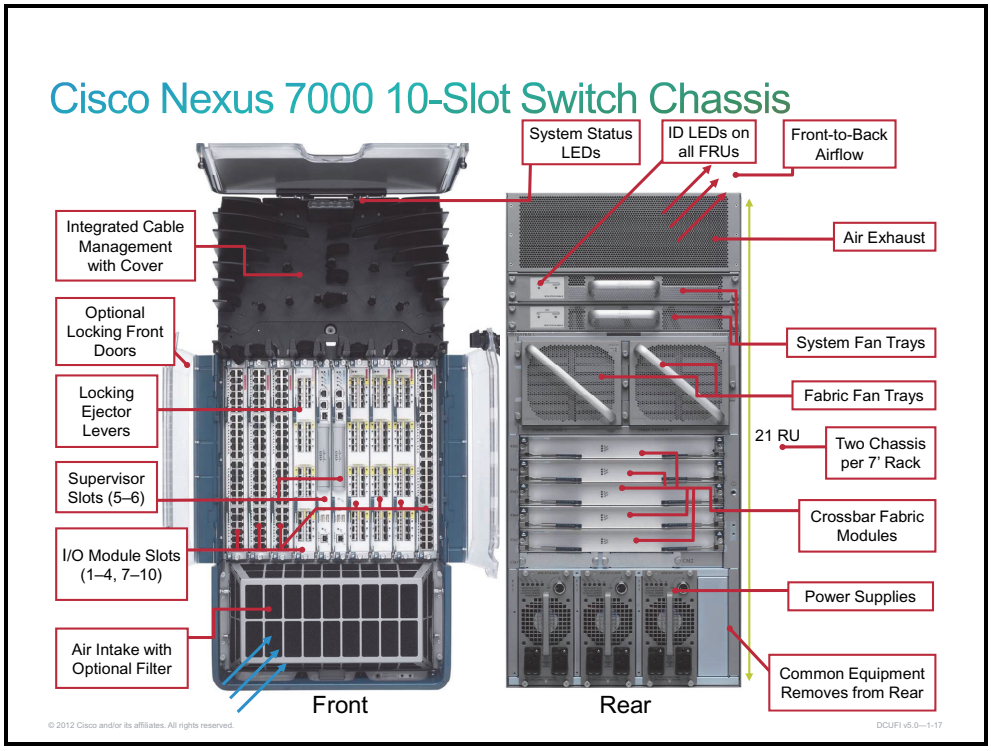


© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 7000 9-Slot Switch chassis with up to seven I/O module slots supports up to 224 10 Gigabit Ethernet or 336 Gigabit Ethernet ports. It also includes these other features:

- Airflow is side-to-side.
- The integrated cable management system is designed to support the cabling requirements of a fully configured system to either or both sides of the switch, allowing maximum flexibility. All system components can easily be removed with the cabling in place, allowing maintenance tasks to be done easily with minimal disruption.
- A series of LEDs at the top of the chassis provides a clear summary of the status of the major system components. The LEDs alert operators to the need to conduct further investigation. These LEDs report the power supply, fan, fabric, supervisor, and I/O module status.
- The purpose-built optional front module door provides protection from accidental interference with both the cabling and modules that are installed in the system. The transparent front door allows easy observation of cabling and module indicators and status lights without any need to open the doors. The door supports a dual-opening capability for flexible operation and cable installation while fitted. The door can be completely removed for both initial cabling and day-to-day management of the system.
- Independent variable-speed system and fabric fans provide efficient cooling capacity to the entire system. Fan tray redundancy features help ensure reliability of the system and support for hot swapping of fan trays.
- The crossbar fabric modules are located in the front of the chassis, with support for two supervisors.

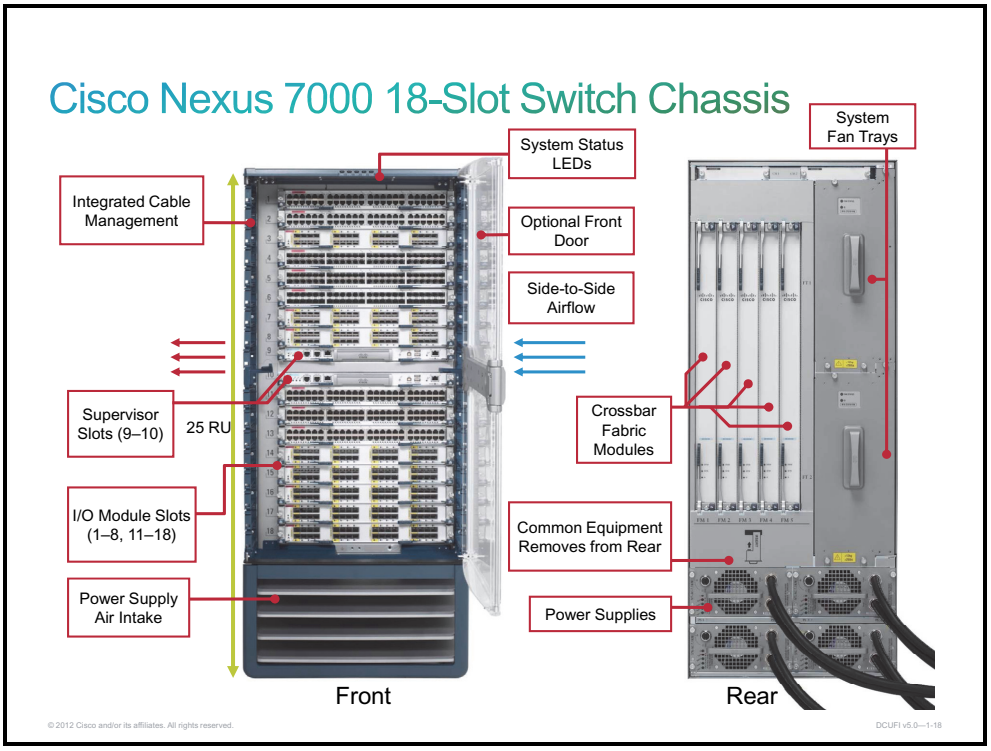
https://t.me/learningnets



The Cisco Nexus 7000 10-Slot Switch chassis with up to eight I/O module slots supports up to 256 10 Gigabit Ethernet or 384 Gigabit Ethernet ports, meeting the demands of large deployments. It also includes these other features:

- Front-to-back airflow helps ensure that use of the Cisco Nexus 7000 10-Slot Switch chassis addresses the requirement for hot-aisle and cold-aisle deployments without additional complexity.
- The system uses dual system and fabric fan trays for cooling. Each fan tray is redundant and composed of independent variable-speed fans that automatically adjust to the ambient temperature. This adjustment helps reduce power consumption in well-managed facilities while providing optimum operation of the switch. The system design increases cooling efficiency and provides redundancy capabilities, allowing hot swapping without affecting the system. If either a single fan or a complete fan tray fails, the system continues to operate without a significant degradation in cooling capacity.
- The integrated cable management system is designed for fully configured systems. The system allows cabling either to a single side or to both sides for maximum flexibility without obstructing any important components. This flexibility makes maintenance easy even when the system is fully cabled.
- The system supports an optional air filter to help ensure clean airflow through the system. The addition of the air filter satisfies Network Equipment Building System (NEBS) requirements.
- A series of LEDs at the top of the chassis provides a clear summary of the status of the system components. The LEDs alert operators to the need to conduct further investigation. These LEDs report the power supply, fan, fabric, supervisor, and I/O module status.
- The cable management cover and optional front module doors provide protection from accidental interference with the cabling and modules that are installed in the system. The transparent front door allows observation of cabling and module indicator and status lights.

https://t.me/learningnets



The Cisco Nexus 7000 18-Slot Switch chassis with up to 16 I/O module slots supports up to 512 10 Gigabit Ethernet or 768 Gigabit Ethernet ports, meeting the demands of the largest deployments. It also includes these other features:

- Side-to-side airflow increases the system density within a 25-RU footprint, optimizing the use of rack space. The optimized density provides more than 16 RU of free space in a standard 42-RU rack for cable management and patching systems.
- The integrated cable management system is designed to support the cabling requirements of a fully configured system to either or both sides of the switch, allowing maximum flexibility. All system components can easily be removed with the cabling in place, allowing maintenance tasks to be easily done with minimal disruption.
- A series of LEDs at the top of the chassis provides a clear summary of the status of the major system components. The LEDs operators to the need to conduct further investigation. These LEDs report the power supply, fan, fabric, supervisor, and I/O module status.
- The purpose-built optional front module door provides protection from accidental interference with both the cabling and modules that are installed in the system. The transparent front door allows easy observation of cabling and module indicators and status lights without any need to open the doors. The door supports a dual-opening capability for flexible operation and cable installation while fitted. The door can be completely removed for both initial cabling and day-to-day management of the system.
- Independent variable-speed system and fabric fans provide efficient cooling capacity to the entire system. Fan tray redundancy features help ensure reliability of the system and support for hot swapping of fan trays.

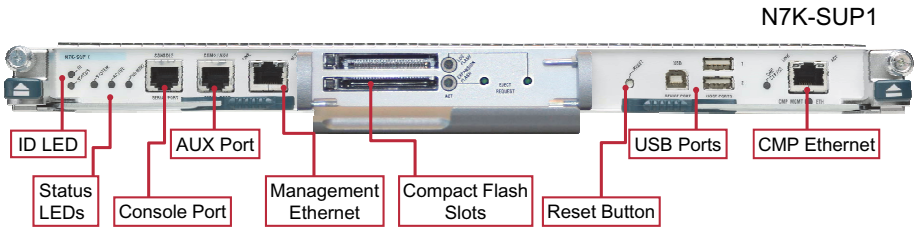
© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Important Features of Cisco Nexus 7000 I/O Modules

This topic identifies the important features and benefits of the Cisco Nexus 7000 Series I/O modules.

Cisco Nexus 7000 Supervisor Module

- Performs control plane and management functions
- Dual-core 1.66-GHz x86 processor with 4 GB DRAM
- 2 MB NVRAM, 2 GB internal bootdisk, compact flash slots
- OOB 10/100/1000 management interface
- Always-on CMP for lights-out management
- Console and auxiliary serial ports
- USB ports for file transfer



© 2012 Cisco and/or its affiliates. All rights reserved.

The first Cisco Nexus 7000 Series Supervisor Module is designed to deliver scalable control plane and management functions for the Cisco Nexus 7000 Series chassis. It is based on a dual core processor that scales the control plane by harnessing the flexibility and power of the dual cores. The supervisor controls the Layer 2 and Layer 3 services, redundancy capabilities, configuration management, status monitoring, power and environmental management, and more. It provides centralized arbitration to the system fabric for all line cards. The fully distributed forwarding architecture allows the supervisor to support transparent upgrades to higher forwarding capacity-capable I/O and fabric modules. The supervisor incorporates an innovative dedicated connectivity management processor (CMP) to support remote management and troubleshooting of the complete system. Two supervisors are required for a fully redundant system. One supervisor module runs as the active device while the other is in hot standby mode. This redundancy provides exceptional high-availability features in data center-class products.

Initial shipment of supervisor modules (Supervisor1 [Sup1]) was with 4 GB of DRAM. An upgrade to 8 GB of DRAM is needed for newer features such as Multiprotocol Label Switching (MPLS) or storage virtual device context (VDC).

https://t.me/learningnets

To deliver a comprehensive set of features, the Cisco Nexus 7000 Series Supervisor Module offers the following:

- Continuous system operation
  - Active and standby supervisor
  - Segmented and redundant OOB provisioning and management paths
  - Virtualization of the management plane
  - Integrated diagnostics and protocol decoding with an embedded control plane packet analyzer
- Upgradable architecture
  - Fully decoupled control plane and data plane with no hardware forwarding on the module
  - Distributed forwarding architecture, allowing independent upgrades of the supervisor and fabric
  - Cisco Unified Fabric-ready
  - Transparent upgrade capacity and capability, which are designed to support 40 and 100 Gigabit Ethernet
- Superior operational efficiency
  - System locator and beacon LEDs for simplified operations
  - Dedicated OOB management processor for “lights out” management

Cisco Nexus 7000 CMP

- OOB management and monitoring capability for:
  - Supervisor module control processor
  - All modules
  - Entire Cisco Nexus 7000 Series system
- Capability to reboot all components, including power supplies
- Dedicated processor, memory, bootflash memory and Ethernet management port

© 2012 Cisco and/or its affiliates. All rights reserved.

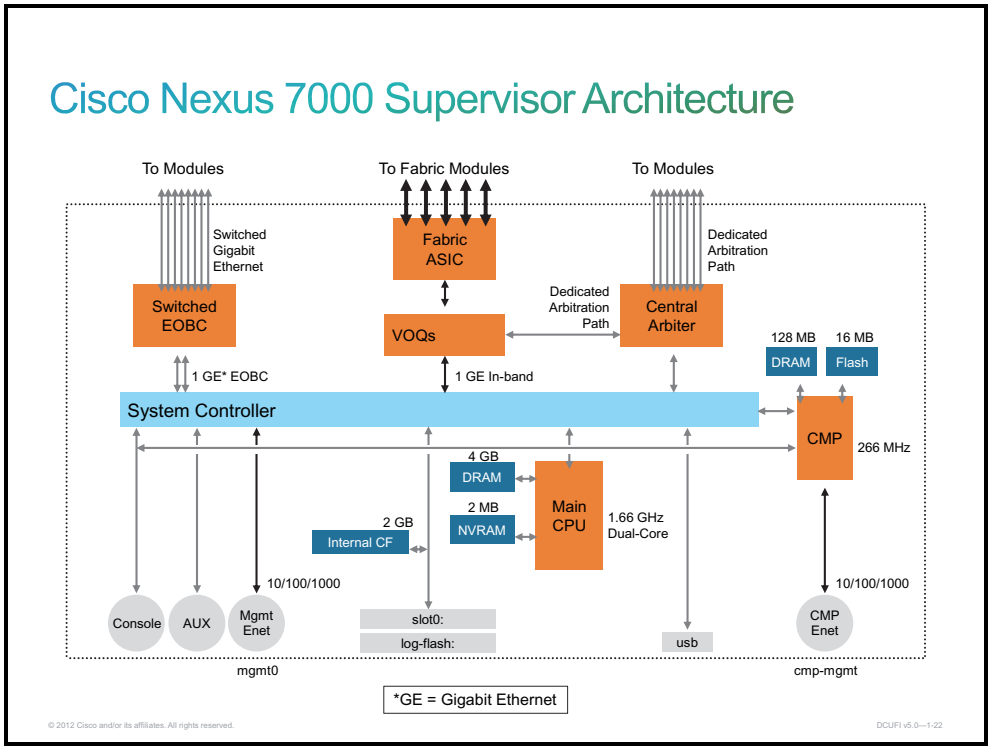
The CMP provides a complete OOB management and monitoring capability independent from the primary operating system. The CMP enables “lights out” remote monitoring and management of the supervisor module, all modules, and the Cisco Nexus 7000 Series system without the need for separate terminal servers with the associated additional complexity and

https://t.me/learningnets

cost. For example, the CMP can monitor the supervisor module control processor on the active supervisor module and can reboot the control processor or Cisco NX-OS device.

The CMP delivers the remote control through its own dedicated processor, memory, and bootflash memory and a separate Ethernet management port. The CMP can reset all system components, including power supplies. It can also reset the host supervisor module to which it is attached, allowing a complete system restart.

Management of large-scale data center networks requires proactive management tools to verify connectivity and mechanisms for capturing and analyzing traffic. The power-on self-test (POST) and Cisco Generic Online Diagnostics (Cisco GOLD) provide proactive health monitoring both at startup and during system operation. The supervisor module uniquely provides a built-in packet capture and protocol decoding tool that allows analysis of control plane traffic.

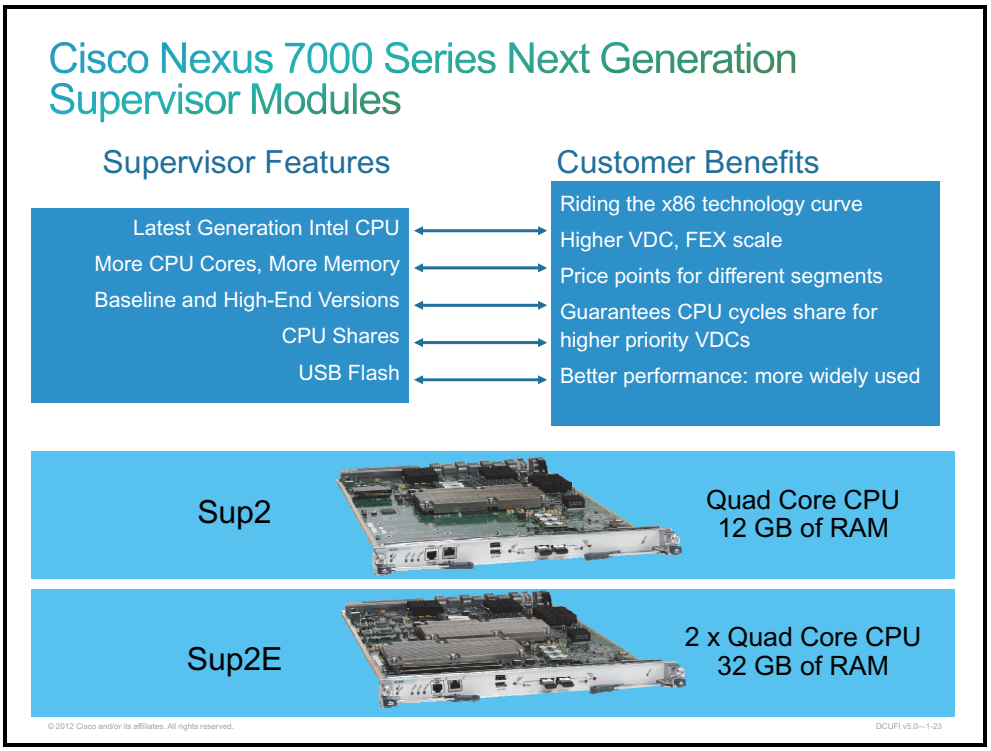


The figure shows the architecture of the Cisco Nexus 7000 Series Supervisor Module. The Ethernet out-of-band channel (EOBC) provides the communication between the supervisor and the line modules. There is a redundant pair of EOBC links. Traffic going into the fabric is controlled using central arbitration.

https://t.me/learningnets

Internal EOBC

Cisco Nexus 7000 uses a switched EOBC for management and control traffic between the supervisors and I/O modules and between the supervisors and fabric modules. On the supervisor modules, Ethernet connectivity is provided using an onboard 24-port Ethernet switch on a chip, with a 1-Gb/s Ethernet link from each supervisor to each I/O module, each supervisor to each switch fabric module (up to five), and between the two supervisors. Two additional redundant 1-Gb/s Ethernet links are used on each supervisor to connect to the local CPU within the supervisor. This design provides a highly redundant switched-Ethernet-based fabric for control and management traffic between the supervisors and all other processors and modules within the system.



The second generation of Cisco Nexus 7000 Series Supervisor Modules comes in two versions: Supervisor2 (Sup2) and Supervisor2 Enhanced (Sup2E).

Sup2 has a quad-core CPU and 12 GB of memory, compared to Cisco Nexus 7000 Series Sup1 Module, which has single-core CPU and 8 GB of memory.

Sup2E is the enhanced version of Sup2 with two quad-core CPUs and 32 GB of memory.

Sup2 and Sup2E have much higher CPU and memory and provide better user experience. Sup2E is recommended for all customers that need higher scale such as VDC and fabric extender (FEX) scale.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Cisco Nexus 7000 Series Supervisor Module Comparison

	Sup1	Sup2	Sup2E
CPU	Dual-Core Xeon	Quad-Core Xeon	2 x Quad-Core Xeon
Speed	1.66 GHz	2.13 GHz	2.13 GHz
Memory	8 GB	12 GB	32 GB
Flash Memory	Compact Flash	USB	USB
CMP	Supported	Not Supported	Not Supported
Cisco NX-OS Release	4.0 or later	6.1 or later	6.1 or later
VDCs	4	4+1	8+1
FEX	32 FEX/1536 Ports	32 FEX/1536 Ports	48 FEX/1536 Ports
CPU	Dual-Core Xeon	Quad-Core Xeon	2 x Quad-Core Xeon

© 2012 Cisco and/or its affiliates. All rights reserved.

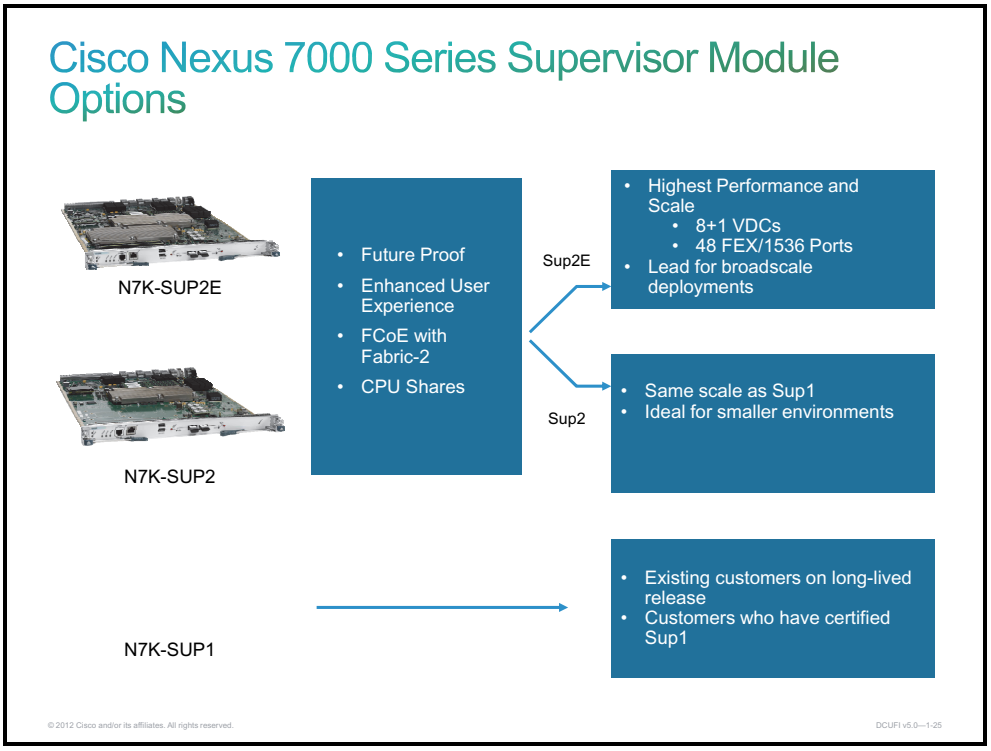
As opposed to Sup1, Sup2 introduces Admin VDC, a special-purpose VDC that is used as a single administration point for a chassis without any user traffic flowing over it.

Sup2 is similar in performance to Sup1 but for FCoE on Cisco Nexus 7000 F2-Series Ethernet modules, Sup2 is required.

In large-scale deployments, Sup2E should be used, which provides the possibility to scale VDCs to 8 + 1 VDC (an eight-user and one admin VDC). Also, Sup2E can scale up to 48 FEXs with 1536 server ports on it.

Both Sup2 and Sup2E support the same Cisco NX-OS features. The main difference between the two supervisors is in the performance and scale.

https://t.me/learningnets



Sup2 and Sup2E have more powerful CPUs and more memory as well as next generation ASICs, which together result in improved performance such as enhanced user experience, faster boot up and switchover times, and higher control plane scale, such as higher VDC and FEX scale. Both Sup2 and Sup2E support FCoE with Cisco Nexus 7000 F2-Series and also have a new feature called "CPU Shares" which allows you to allocate specific amounts of the CPU for higher-priority VDCs.

All existing Cisco Nexus 7000 Series I/O modules and fabric modules are supported with Sup2 and Sup2E

You cannot mix Sup1 and Sup2 in the same chassis. Please note that mixing these two would be a disruptive migration requiring removal of Sup1 modules.

Sup2 and Sup2E can be mixed for migration only. Please note that this is a nondisruptive migration.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Cisco Nexus 7000 Series Fabric Modules

- Separate fabric module
  - Dedicated model for each chassis type
- Parallel fabric channels to each I/O and supervisor module slot
- Central switching element for distributed forwarding on I/O modules

Model	Cisco Nexus 7000 Fabric-1 Module	Cisco Nexus 7000 Fabric-2 Module
Software requirements	• 9-slot fabric module: N/A • 10-slot fabric module: Cisco NX-OS Release 4.0 or later • 18-slot fabric module: Cisco NX-OS Release 4.1(2) or later	• 9-slot fabric module: Cisco NX-OS Release 5.2 or later • 10-slot fabric module: Cisco NX-OS Release 6.0 or later • 18-slot fabric module: Cisco NX-OS Release 6.0 or later
Performance	46 Gb/s per slot per fabric	110 Gb/s per slot per fabric
Max performance per slot	Up to five active fabric modules deliver up to 230 Gb/s per slot	Up to five active fabric modules deliver up to 550 Gb/s per slot

© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 7000 Series fabric modules for the Cisco Nexus 7000 Series chassis are separate fabric modules that provide parallel fabric channels to each I/O and supervisor module slot. Up to five simultaneously active fabric modules work together, delivering up to 550 Gb/s per slot in the case of the Cisco Nexus 7000 Fabric-2 Module. Through the parallel forwarding architecture, a system capacity of more than 15 Tb/s is achieved with the five fabric modules. The fabric module provides the central switching element for fully distributed forwarding on the I/O modules.

All fabric modules are connected to all module slots. The addition of each fabric module increases the bandwidth to all module slots up to the system limit of five modules. The architecture supports lossless fabric failover, with the remaining fabric modules load-balancing the bandwidth to all the I/O module slots, helping ensure graceful degradation.

The combination of the Cisco Nexus 7000 Series fabric module and the supervisor and I/O modules supports virtual output queuing (VOQ) and credit-based arbitration to the crossbar switch to increase performance of the distributed forwarding system. VOQ and credit-based arbitration facilitate fair sharing of resources when a speed mismatch or contention for an uplink interface exists. The fabric architecture also enables future support for lossless Ethernet and unified I/O capabilities.

The table summarizes the software requirements and performance figures for the Cisco Nexus 7000 Fabric-2 Module and the older Cisco Nexus 7000 Fabric-1 Module.

https://t.me/learningnets

Innovative New I/O Modules

Model	6-Port 40 Gigabit Ethernet Module	2-Port 100 Gigabit Ethernet Module	48-Port 1/10 Gigabit Module
Series	M-2 Series with XL Option	M-2 Series with XL Option	F2-Series
Module			
Positioning	Core, aggregation	Core	Access, aggregation
Software requirements	Cisco NX-OS Software Release 6.1 or later	Cisco NX-OS Software Release 6.1 or later	Cisco NX-OS Software Release 6.0 or later
Compatibility	Supported in all Cisco Nexus 7000 Series chassis and with Fabric-1 or Fabric-2 fabric modules	Supported in all Cisco Nexus 7000 Series chassis and with Fabric-1 or Fabric-2 fabric modules	Supported in all Cisco Nexus 7000 Series chassis
Other	Jumbo frames (9216 bytes), Precision Time Protocol (PTP) based on IEEE 1588.		Jumbo frames (9216 bytes)

More details provided in next slides

© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco Nexus 7000 Series Switches offer many I/O modules. This table introduces three new modules that represent the latest technological advancements. They belong to two module series:

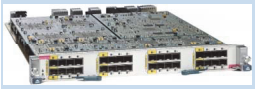
- **Cisco Nexus 7000 M-2 Series**, consisting currently of a six-port 40 Gigabit Ethernet module and a two-port 100 Gigabit Ethernet module, is targeted at core and aggregation layers. This product line requires Cisco NX-OS Software Release 6.1 or later. It is supported in all Cisco Nexus 7000 Series chassis and with Fabric-1 or Fabric-2 fabric modules.
- **Cisco Nexus 7000 F2-Series**, including the 48-port 1 and 10 Gigabit module, is targeted at access and aggregation layers. This product line requires Cisco NX-OS Software Release 6.0 or later. It is supported in all Cisco Nexus 7000 Series chassis.

These innovative modules are covered in detail in the following pages.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Other I/O Examples

Model	48-Port Gigabit Ethernet Module with XL Option	32-Port 10 Gigabit Ethernet Module with XL Option	32-Port 10 Gigabit Ethernet Module
Module			
Features	XL Option (license-based extension of routing and MAC address table size)	• M1-XL forwarding engine • Larger Forwarding Information Base (FIB)	• M1 Series • 8 ports at line rate, or • Up to 32 ports sharing 80 Gb/s of bandwidth
I/O	48 ports in two versions: • Fiber Gigabit Ethernet SFP optics • Copper 10/100/1000 with RJ-45 connectors	32 ports of 10 Gigabit Ethernet (SFP+ pluggable optic module)	• 32 ports of 10 Gigabit Ethernet (SFP+ pluggable optic module) • Ports are organized into eight groups of 4 ports
Positioning	Access	Access, aggregation	Access, aggregation
Software requirements	• NX-OS 5.0 or later (fiber) • NX-OS 5.1 or later (copper)	Cisco NX-OS Software Release 5.1 or later	Cisco NX-OS Software Release 4.0 or later

These modules are not covered in detail

© 2012 Cisco and/or its affiliates. All rights reserved.

This table lists three out of many available Cisco Nexus 7000 Series I/O modules, including a brief overview of the features, ports, positioning, and software requirements.

https://t.me/learningnets

M-2 Series 6-Port 40 Gigabit Ethernet Module

- M-2 Series 6-Port 40 Gigabit Ethernet Module with XL Option
- Large network cores, service provider and Internet peering environments
- Optional Scalable Feature License, which enables enhanced XL mode

Item	Non-XL Mode	XL Mode (with Scalable Feature License)
MAC entries	128K	128K
IPv4 routes	128K	Up to 1M*
IPv6 routes	64K	Up to 350K*
NetFlow entries	512K	512K
ACLs	64K	128K

* Actual limit depends on prefix distribution.

© 2012 Cisco and/or its affiliates. All rights reserved.

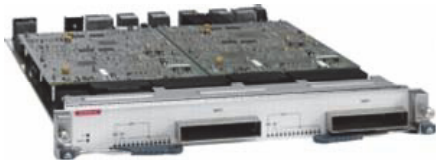
The Cisco Nexus 7000 M-2 Series 6-Port 40 Gigabit Ethernet Module with XL Option is a highly scalable module offering full-featured, nonblocking 40 Gigabit Ethernet performance on each port. It facilitates the deployment of high-density, high-bandwidth architectures, especially in large network cores and in service provider and Internet peering environments.

The Cisco Nexus 7000 M2-Series Module has a number of features that support the highest performance and comprehensive features. With an optional Scalable Feature License, the module can operate in enhanced XL mode, which enables use of the full forwarding table, essential for large-scale deployments such as Internet peering environments. This larger forwarding table can support multiple copies of the full Internet route table for use in Internet-facing deployments with virtual routing and forwarding (VRF) and VDC support. The ability to operate in either non-XL or XL mode makes the module versatile and flexible, without requiring a hardware module change or upgrade. The table lists the performance specifications for the Cisco Nexus 7000 M-2 Series Module operating in non-XL and XL modes.

https://t.me/learningnets

M-2 Series 2-Port 100 Gigabit Ethernet Module

- Cisco Nexus 7000 M-2 Series 2-Port 100 Gigabit Ethernet Module
- License-based enhanced XL Option
- Other M-2 series features:
  - Comprehensive Layer 2 and Layer 3 functionality
  - MPLS forwarding in hardware
  - Online insertion and removal (OIR)
  - Cisco TrustSec
    - SGT-based ACLs
    - MAC security (IEEE 802.1AE) using AES cipher



© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 7000 M-2 Series 2-Port 100 Gigabit Ethernet Module with XL Option provides the XL Option enhancements of the M-2 Series, and is particularly well-suited for high-bandwidth core environments.

Among additional features of the M2-Series, these deserve particular attention:

- Comprehensive Layer 2 and Layer 3 functionality that includes Layer 3 routing protocols
- Multiprotocol Label Switching (MPLS) forwarding in hardware
- Online insertion and removal (OIR)
- Cisco TrustSec solution, in particular Security Group Access Control List (SGACL), and MAC security (IEEE 802.1AE) using Advanced Encryption Standard (AES) cipher.

https://t.me/learningnets

F2-Series 48-Port 1 and 10 Gigabit Ethernet

- Cisco Nexus 7000 F2-Series 48-Port 1 and 10 Gigabit Ethernet Module
- Switch-on-a-chip (SoC) architecture
  - Single ASIC implements all the module functions: from ingress buffering, forwarding lookups, ACLs, QoS, and virtual output queuing (VOQ).
  - Each SoC manages four front-panel interfaces
- Layer 2 switching, Layer 3 forwarding, FabricPath, vPC+
- Support for Cisco Nexus 2000 Fabric Extender
- Integrated FCoE:
  - Host and target support
  - Virtual expansion port (VE-port)
  - FCoE Inter-Switch Links (ISLs)

© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 7000 F2-Series 48-Port 1 and 10 Gigabit Ethernet Module offers outstanding flexibility and wire-rate performance on each port.

It is built with switch-on-a-chip (SoC) architecture, in which a single ASIC implements all the module functions: from ingress buffering, to forwarding lookups and ACLs and QoS tables, to fabric interfaces and virtual output queuing (VOQ). Each SoC manages four front-panel interfaces. This type of design increases performance while lowering the power and cooling requirements of the module.

The comprehensive feature set includes classic Layer 2 and Layer 3 forwarding. In addition, the module delivers Cisco FabricPath technology based on IETF TRILL. Cisco FabricPath consists of a set of multipath Ethernet technologies, combining the reliability and scalability benefits of Layer 3 routing with the flexibility and "plug-and-play" aspects of Layer 2 Ethernet networks.

Note Cisco FabricPath is explained in detail later in the course.

The Cisco Nexus 7000 F2-Series Module can be used with the Cisco Nexus 2000 Series Fabric Extenders.

It also delivers integrated FCoE, greatly simplifying the network infrastructure and reducing costs by enabling the deployment of unified data center fabrics to consolidate data center traffic onto a single network. In addition to FCoE host and target support, the module provides virtual E Port (VE Port) support, allowing creation of FCoE Inter-Switch Links (ISLs) and enabling scalable, multihop FCoE topologies.

https://t.me/learningnets

Nexus 7000 I/O Modules Comparison: M and F Series

	M-1Series	M-2 Series	F1-Series	F2-Series
Line Rate 10 Gigabit Ports (9/10/18-slot Chassis)	56 / 64 / 128	168 / 192 / 384	161 / 184 / 368	336 / 384 / 768
Latency	~ 20 microsec	~ 15 microsec	~ 5 microsec	~ 6 microsec
Power / Line Rate 10 Gigabit Ethernet Port	~ 80 watts / port	~ 32 watts / port	~ 10 watts / port	~ 10 watts / port
VoQ Buffer Per Card	32 MB	128 MB	48 MB	72 MB
VLANs per VDC	4K	4K	4K	4K
L2 MAC Table	128K	128K	16K / SoC	16K / SoC
L3 IPv4 Routes	128K - 1M	Up to 1M	NA	32K
L3 IPv6 Routes	64K - 350K	Up to 350K	NA	16K
Adjacency Table	1M	1M	NA	16K
ECMP	1 … 16	1 … 16	1 … 16	1 … 16
NetFlow	Full / Sampled	Full / Sampled	NA	Sampled [1]
ACL Entries	64K - 128K	128K	2K / SoC	16K / SoC
SPAN and ERSPAN Sessions	2	2 bidirectional + 11 Tx/Rx unidirectional [1]	2 bidirectional + 11 Tx/Rx unidirectional [1]	2 bidirectional + 11 Tx/Rx unidirectional [1]
Sampled SPAN and ERSPAN	No	No	Yes	Yes
IEEE 1588 PTP	No	Yes [1]	Yes	Yes
MPLS	Yes	Yes [1]	No	No
Cisco OTV	Yes	Yes [1]	No	No
FabricPath	No	No	Yes	Yes
FCoE	No	No	Yes	Yes [1,2]

[1] Feature will be enabled in future NX-OS version
[2] Requires SUP2

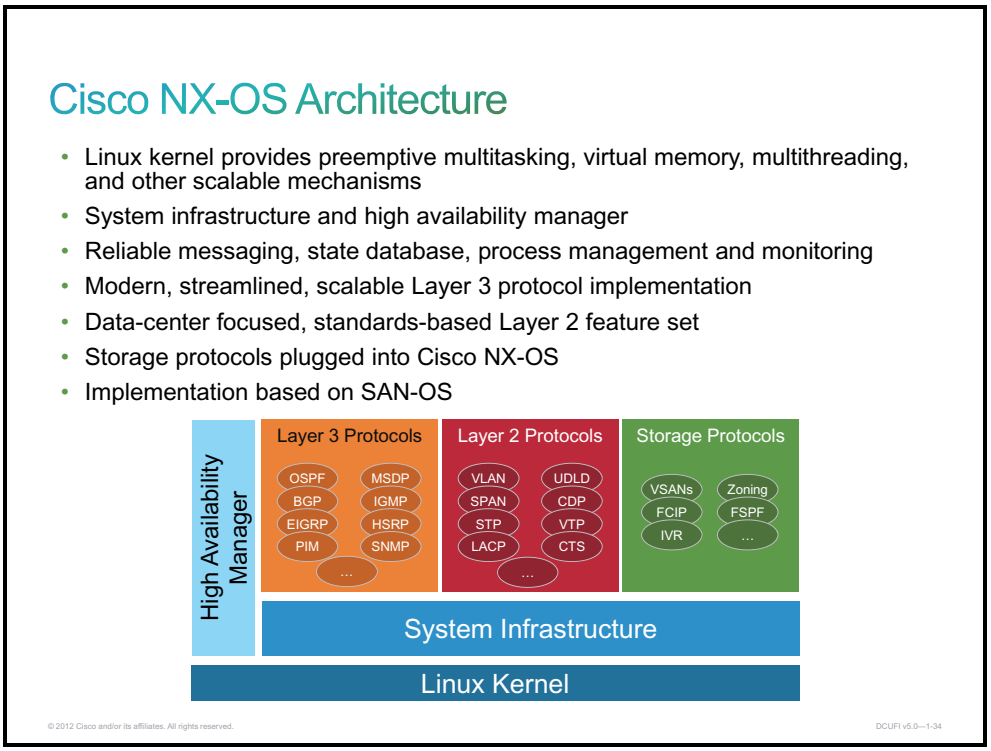
© 2012 Cisco and/or its affiliates. All rights reserved.

This table shows a comparison of the Cisco Nexus 7000 Series I/O modules: M-1 and M-2 Series and F1 and F2-Series modules.

https://t.me/learningnets

Important Features of Cisco NX-OS

This topic identifies the important features of Cisco NX-OS that provide high availability and scalability as well as support for the Cisco Unified Fabric.



The Cisco NX-OS Software is a data center-class operating system that is built with modularity, resiliency, and serviceability as its foundation. The Cisco NX-OS helps ensure continuous availability and sets the standard for mission-critical data center environments. The Cisco Nexus Family of switches and Cisco MDS 9000 Series Multilayer Switches share this common operating system that focuses on data center features and protocols, availability, and operational considerations.

The services within Cisco NX-OS are designed as nonkernel space (user space) processes that perform a function or set of functions for a subsystem or feature set.

Each service (essentially a feature) and each service instance is run as a separate independent protected process. This approach provides a highly fault-tolerant software infrastructure and fault isolation between services. In short, a failure in a service instance (such as IEEE 802.1Q or Multiple Spanning Tree [MST]) will not affect any other services running at that time (such as Link Aggregation Control Protocol [LACP]). Additionally, each instance of a service can run as an independent process. This implies that two instances of a routing protocol (for example, two instances of Open Shortest Path First [OSPF]) run as separate processes, thereby providing fault isolation even between those two instances of the same service.

This resilient highly modular architecture allows the system to provide the highest level of protection and fault tolerance for all services running on the platform. This approach also facilitates rapid fault isolation, resolution, and modular software upgrades to address specific issues, while minimizing the impact to other critical services and the overall system.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Cisco NX-OS Features

Category	Features
Resiliency	Modular software design, ISSU, process survivability, reliable interprocess communication, stateful supervisor failover
Extensibility	Common software throughout the DC, DC-class Ethernet switching, Virtual Port Channel (vPC), Cisco Overlay Transport Virtualization (OTV), FabricPath, Locator/ID Separation Protocol (LISP), IP routing, IP multicast, Data Center Bridging (DCB)
Security	802.1AE link-layer cryptography, security group access control lists (SGACLs) based on security group tags, switch and host authentication, port security, fabric binding
Efficiency	Cisco Fabric Extender Link (FEX-Link), priority flow control (PFC), Cisco Fabric Analyzer, Etheranalyzer, N-port ID virtualization (NPIV)
Virtualization	Virtual device contexts (VDCs), MPLS VRFs, virtual storage area networks (VSANs), Cisco Adapter FEX

© 2012 Cisco and/or its affiliates. All rights reserved.

The main features and benefits include the following:

- Flexibility and scalability
  - Software compatibility
  - Common software throughout the data center
  - Modular software design
  - VDCs
  - Support for Cisco Nexus 2248TP GE Fabric Extender
- Availability
  - Continuous system operation
  - Cisco In-Service Software Upgrade (Cisco ISSU)
  - Quick development of enhancements and problem fixes
  - Process survivability
  - Stateful supervisor failover
  - Reliable interprocess communication
  - Redundant switched EOBCs
  - Network-based availability
- Serviceability
  - Troubleshooting and diagnostics
  - Switched Port Analyzer (SPAN)
  - Ethanalyzer
  - Smart Call Home

https://t.me/learningnets

- Cisco GOLD
- Cisco IOS Embedded Event Manager (Cisco IOS EEM)
- Cisco IOS NetFlow

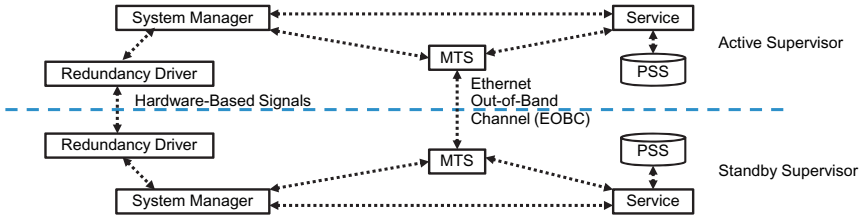
- Manageability
  - Programmatic XML interface
  - Simple Network Management Protocol (SNMP)
  - Configuration verification and rollback
  - Port profiles
  - Role-based access control (RBAC)
  - Cisco Data Center Network Manager (Cisco DCNM)
  - CMP support
- Traffic routing, forwarding, and management
  - Ethernet switching
  - Cisco Overlay Transport Virtualization (Cisco OTV)
  - Ethernet enhancement
  - Cisco FabricPath
  - IP routing
  - IP multicast
  - QoS
  - Traffic redirection
- Network Security
  - Cisco TrustSec solution
  - Data path intrusion detection system (IDS) for protocol conformance checks
  - Control-Plane Policing (CoPP)
  - Dynamic ARP Inspection (DAI)
  - DHCP snooping
  - IP Source Guard
  - Authentication, authorization, and accounting (AAA) and TACACS+
  - Secure Shell version 2 (SSHv2)
  - Simple Network Management Protocol version 3 (SNMPv3)
  - Port security
  - IEEE 802.1X authentication and RADIUS support
  - Layer 2 Cisco Network Admission Control (NAC) LAN port IP

Note The important features are covered in detail in this course.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Cisco NX-OS System Redundancy

- Three critical core infrastructure services:
  - System manager
  - Persistent storage service (PSS)
  - Message and transaction service (MTS)
- Redundant setup with dual supervisors:
  - Mirrored services run on each supervisor module
  - Configuration and operating state that is synchronized between them
  - One supervisor active, the other standby
- This method is applicable only to stateful processes.



© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco NX-OS uses three critical core infrastructure services to provide overall high availability functionality:

- System manager
- Persistent storage service (PSS)
- Message and transaction service

In a redundant configuration, such as when dual supervisor modules are in operation, mirrored services run on each supervisor module, with configuration and operating state synchronized between them. One of those supervisors will operate as the active supervisor while the other operates in a standby facility until it is activated in a switchover.

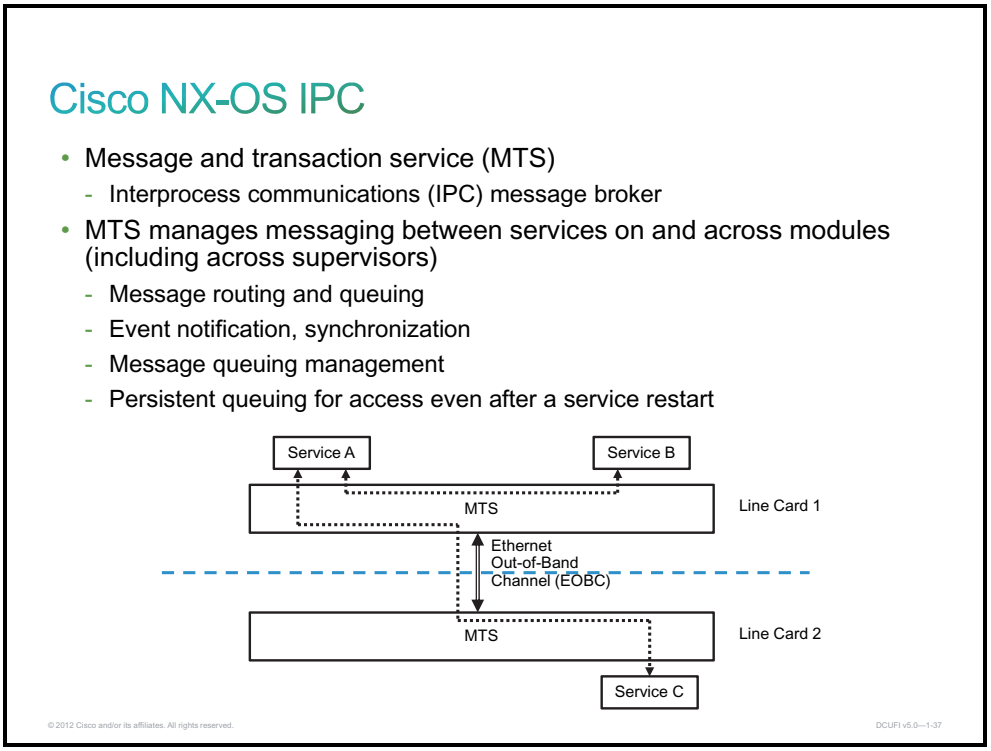
The system manager orchestrates overall system function, service management, and system health monitoring. It is also responsible for maintaining overall high availability states, enforcing high availability policies, and managing system configuration. It is the system manager that is responsible for launching, stopping, monitoring, and restarting services. The system manager is also used for initiating and managing the syncing of service states and intersupervisor states for Stateful Switchover (SSO). It is also the system manager that will initiate a supervisor switchover if it determines that the current supervisor has undergone an unrecoverable failure or if critical core services are undergoing errors and cannot be restarted reliably. In addition, being the overall control and monitoring process, the system manager is responsible for "punching" or triggering the keepalive indicator for the hardware-based watchdog timer on the supervisor. The lack of this periodic heartbeat from the system manager within the keepalive timeout period of the watchdog timer will indicate a nonresponsive system manager, which will trigger a hardware-based supervisor reset (single supervisor) or switchover (dual supervisors). The health of the system manager is also monitored by a kernel-level module receiving periodic heartbeats that are sent by the system manager process. This process allows the system to take corrective action in response to an unresponsive system manager that has exceeded the heartbeat timeout period.

The PSS is the base infrastructure component that is responsible for storing and managing the operational run-time information and configuration of the other platform services. It is used by

https://t.me/learningnets

other system services to recover state in the event of a service restart. PSS provides a managed, structured API to read and write data to and from the storage system, essentially functioning as a database of state and run-time information. Services wishing to use the PSS infrastructure are able to checkpoint their state information periodically, as needed. This ability allows services to later recover to the last known operating state preceding a failure, thereby allowing for a stateful restart. This state recovery capability is available to Cisco NX-OS services in both single- and dual-supervisor configurations, and helps enable a service to transparently return to operation without service disruption to data-plane traffic, neighboring devices, or other internal services.

For example, even in a single supervisor configuration, the PSS enables the stateful restart of a service such as spanning tree without impacting the overall spanning-tree topology or stability.

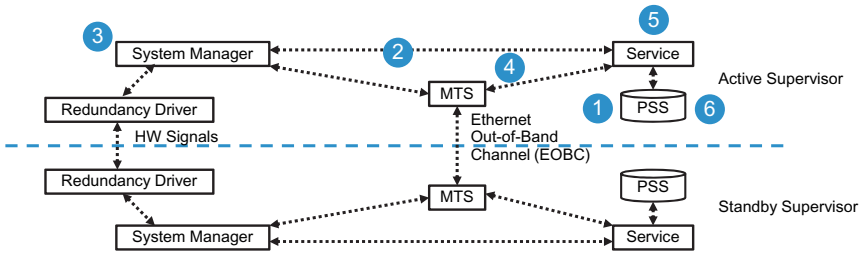


The Cisco NX-OS MTS is a high performance interprocess communications (IPC) message broker that specializes in high availability semantics. MTS manages message routing and queuing between services on and across modules (including across supervisors). MTS facilitates the exchange of messages such as event notification, synchronization, and message persistency between system services and system components. MTS also facilitates and manages message queuing between services and, as a result, can maintain persistent messages and logged messages in queues for access even after a service restart.

https://t.me/learningnets

Cisco NX-OS Service Restart

1. Service checkpoints the state to the PSS.
2. System manager detects a problem with the service.
3. System manager checks the high-availability policy for the service.
4. System manager issues a stateful restart of the service.
5. The service restarts.
6. The service retrieves the runtime state from the PSS.



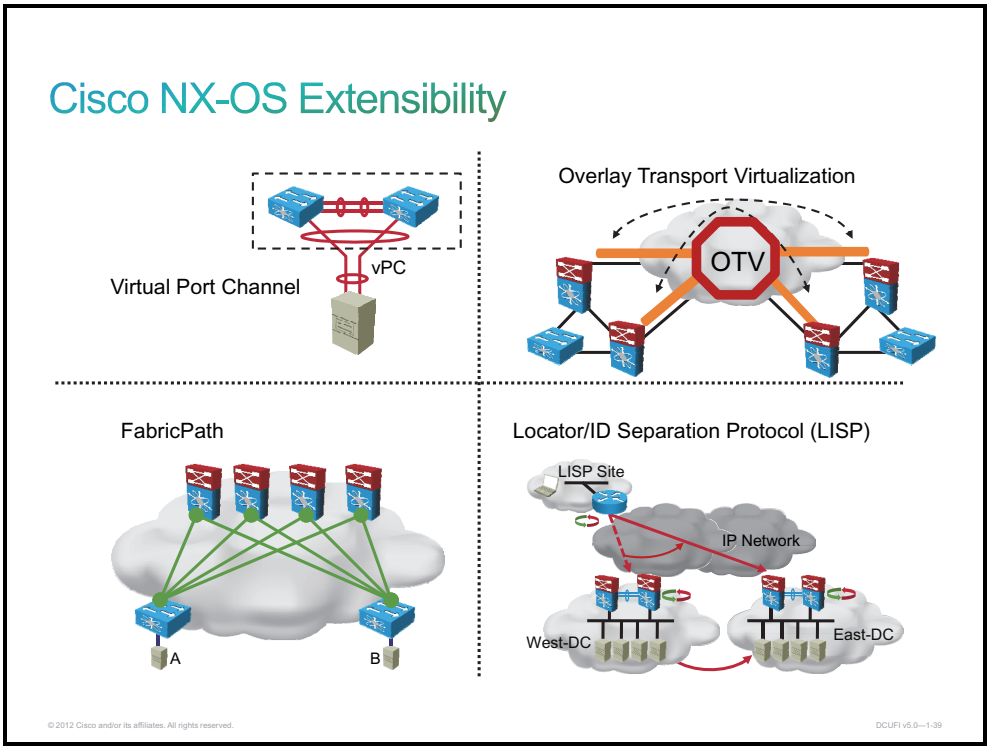
© 2012 Cisco and/or its affiliates. All rights reserved.

The services in Cisco NX-OS are capable of undergoing rapid restart. A service restart may be initiated automatically by the system manager in the event of critical fault detection. When a restart is initiated, a service process is sent a signal to stop and is then shut down and rapidly restarted, typically within milliseconds. This restart allows an error condition to be cleared and a service to be reset if necessary.

A service can undergo different types of restarts, stateful or stateless. A service may be restarted by the system manager depending on current errors, failure circumstances, and configured high availability policy for the service. If the service is issued a stateful restart, the new service process will retrieve the previous run-time state data from PSS and resume operations from the last checkpointed service state. Most of the services in Cisco NX-OS are designed to support stateful restart capabilities by using the high availability infrastructure services of PSS and MTS. If the service is issued a stateless restart, it will initialize and run as if it had just been started with no prior state.

Not all services are designed for stateful restart. Layer 3 routing protocols (Intermediate System-to-Intermediate System [IS-IS], OSPF, Border Gateway Protocol [BGP], and Request in Progress [RIP]) for IP version 4 (IPv4), IP version 6 (IPv6), and IP multicast are not designed to leverage the state persistence of PSS within Cisco NX-OS. Instead, these protocols are currently designed to rebuild their operational state using information that is obtained from neighbors.

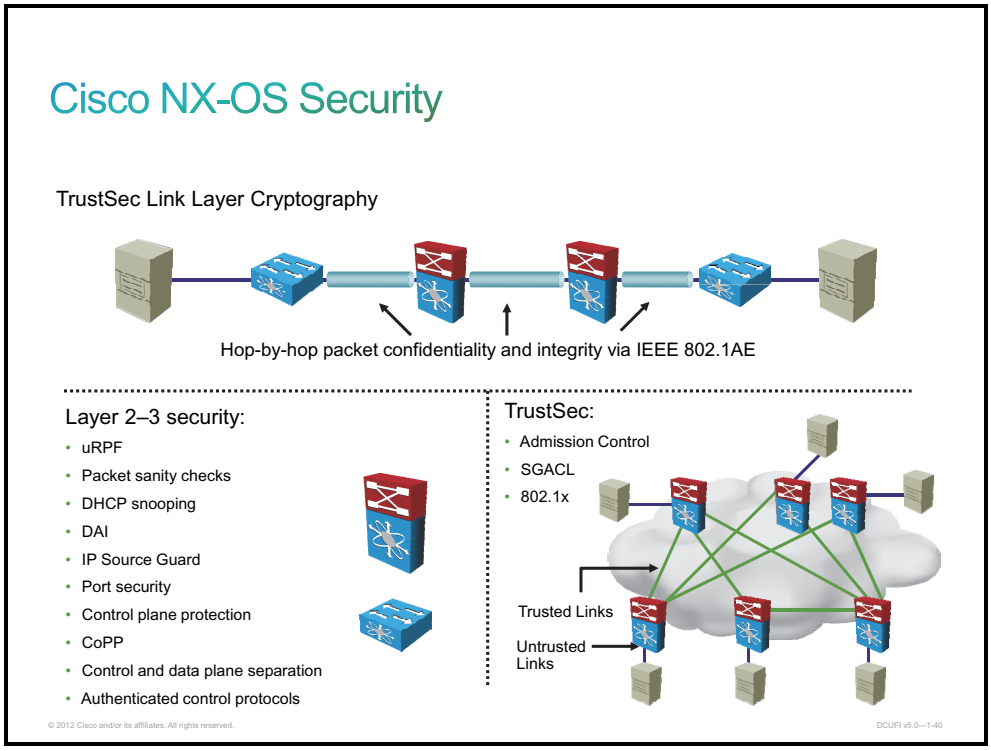
https://t.me/learningnets



Cisco NX-OS features can be grouped into several categories. One of them is the extensibility. Cisco NX-OS offers multiple extensibility solutions:

- The vPC allows redundant uplink access while the redundant upstream devices appear as a single logical device to the access device or host. This solution supports load balancing in addition to redundancy.
- Cisco OTV allows a MAC-transparent LAN extension to other enterprise sites. This solution is based on frame tunneling inside overlay IP packets. It uses IP multicast for optimal forwarding.
- Cisco FabricPath is a solution that implements IS-IS routing principles in a Layer 2 domain, thus removing the need for Spanning Tree Protocol (STP), and offering benefits, such as load balancing over parallel paths.
- LISP is a technology that provides mobile access based on mapping a local address to globally reachable addresses.

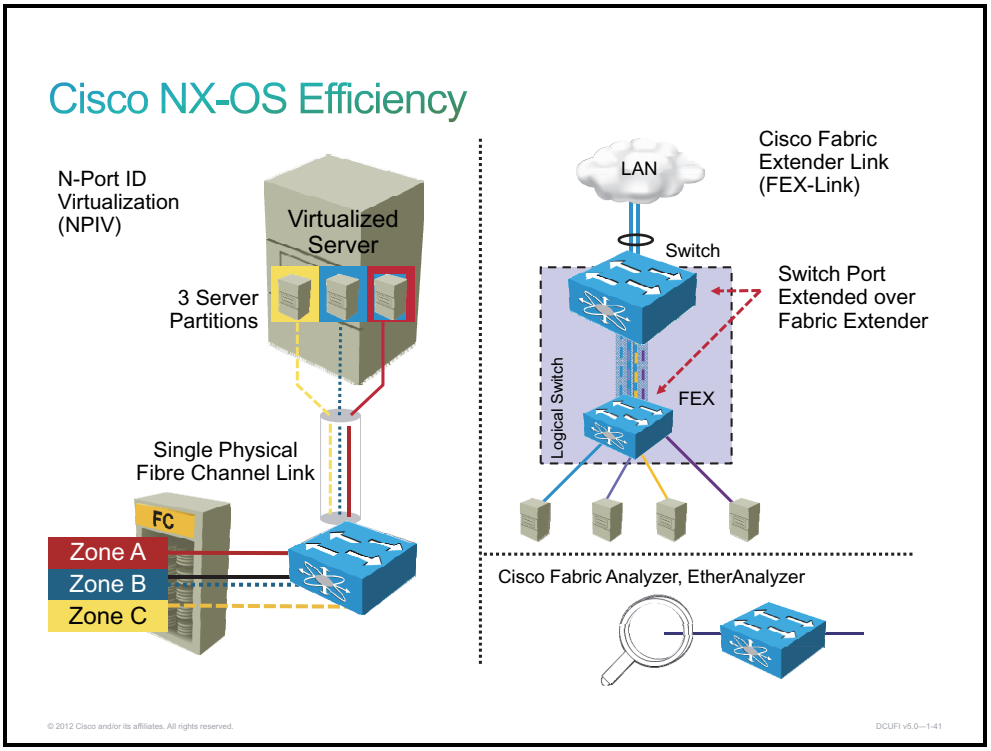
© 2012 Cisco Systems, Inc.
https://t.me/learningnets



Cisco NX-OS includes these security features:

- Cisco TrustSec data link layer cryptography, which provides protection for packets by encrypting packets on egress and decrypting packets on ingress at the device. Within the device itself, packets are in plaintext format, allowing the network to continue performing all packet inspection functions.
- Classic Layer 2 and Layer 3 security features, such as Unicast Reverse Path Forwarding (uRPF), DHCP snooping, DAI, IP Source Guard, port security, control plane protection, CoPP, and authenticated control protocols.
- Cisco TrustSec security architecture that builds secure networks by establishing clouds of trusted network devices. Each device in the cloud is authenticated by its neighbors. Communication on the links between devices in the cloud is secured with a combination of encryption, message integrity checks, and data-path replay protection mechanisms.

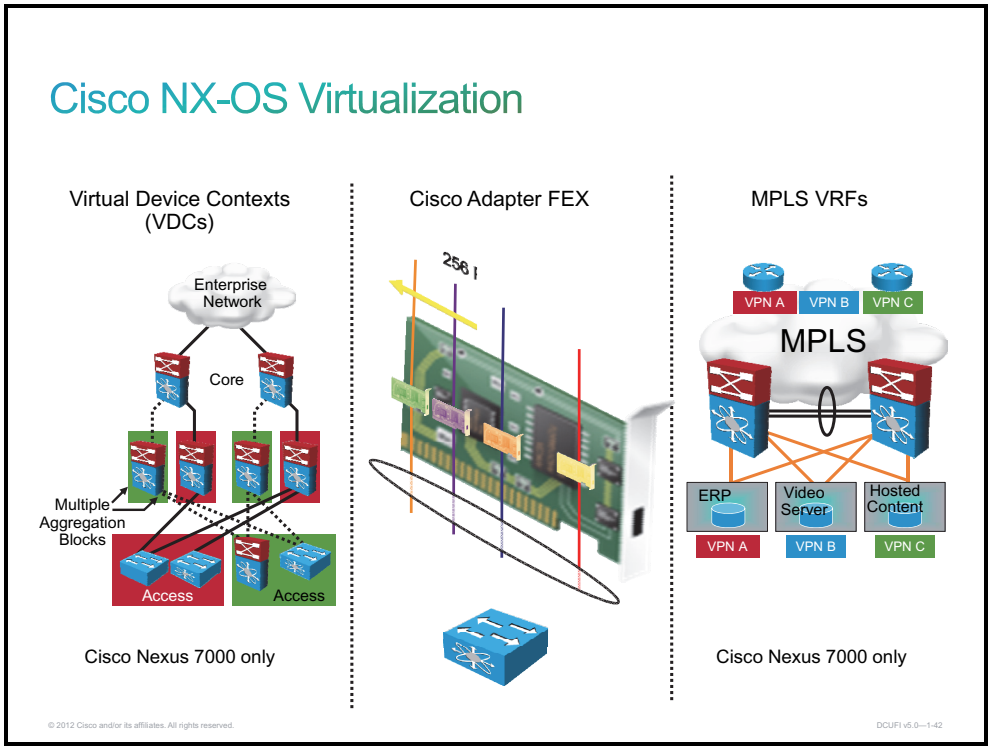
https://t.me/learningnets



Cisco NX-OS includes these efficiency mechanisms:

- With Cisco N-Port Virtualizer (Cisco NPV), Cisco Nexus switches relay the fabric login (FLOGI) and discover fabric service parameters (fabric discovery [FDISC]) to the upstream Fibre Channel switch. In this mode, a Cisco Nexus switch operates as an N-Port proxy mode (NP Port). It does not do any Fibre Channel switching itself. There are no local switching and zoning checks on the Cisco Nexus switches. The Cisco Nexus switch appears as a host to the upper layer switches and as a Fibre Channel switch to the server that is attached to it. In this mode, the switch does not provide Fibre Channel services, and therefore does not need a Fibre Channel domain ID. As a result, the switches increase the scalability and eliminate the switch-to-switch interoperability issues. They also make management easier as they do not get involved in the FSPF and Fibre Channel policy. The ports that connect to the upstream Fibre Channel switch need to support N-Port ID Virtualization (NPIV).
- Cisco Fabric Extender Link (FEX-Link) technology, in which the Cisco Nexus 2200 Platform fabric extenders operate as external line modules. This functionality provides data centers with massive scalability to manage the combination of an increasing number of servers and a higher demand for bandwidth from each server. The Cisco Nexus 2000 Series increases the scalability of the access layer to accommodate both sets of demands without increasing management points within the network.
- A built-in protocol analyzer that is based on the popular open source Wireshark software.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



Cisco NX-OS includes these virtualization features:

- VDCs allow a single physical Cisco Nexus 7000 Series switch to be partitioned into multiple logical switches. This partitioning enables the consolidation of different logical zones onto a single physical infrastructure. VDCs are supported only on the Cisco Nexus 7000 Series.
- Cisco Adapter Fabric Extender (FEX) is a virtualization-optimized FCoE PCI Express (PCIe) 2.0 x8 10-Gb/s adapter. The virtual interface card is a dual-port 10 Gigabit Ethernet PCIe adapter that can support up to 256 PCIe standards-compliant virtual interfaces, which can be dynamically configured so that both their interface type (NIC or HBA) and identity (MAC address and world wide name [WWN]) are established using just-in-time provisioning. In addition, the Cisco Adapter Fabric Extender can support network interface virtualization and Cisco Virtual Machine Fabric Extender (VM-FEX) technology.
- MPLS VRF, along with related technologies such as MPLS traffic engineering, enables you to provide VPN connectivity over a common MPLS infrastructure. A VRF is a separated logical entity within the device that is dedicated to a given VPN. MPLS is supported only on the Cisco Nexus 7000 Series.

https://t.me/learningnets

Summary

This topic summarizes the key points that were discussed in this lesson.

Summary

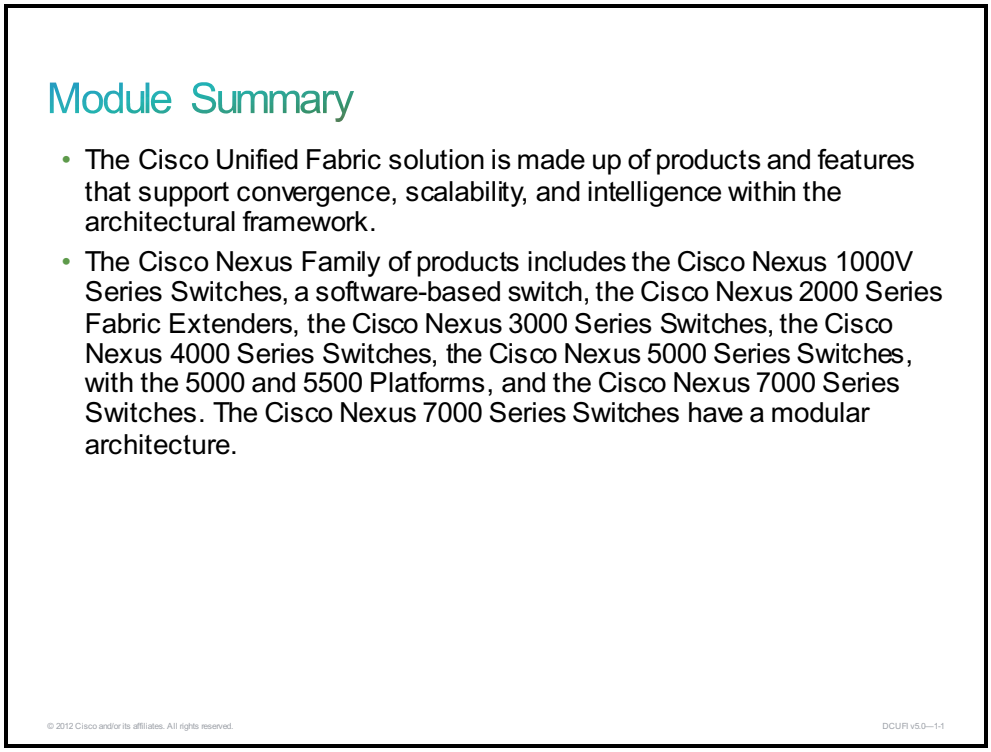
- The Cisco Nexus Family of switches is made up of a number of products ranging from the Cisco Nexus 1000V Series Switches to the Cisco Nexus 7000 18-Slot Switch. The Cisco Nexus 2000 Fabric Extender can be used as a remote line module for the Cisco Nexus 5000 and 5500 Platform switches, and for the Cisco Nexus 7000 Series Switches.
- There are a number of line modules that are available for the Cisco Nexus 7000 Series Switches to support up to 100 Gigabit Ethernet interfaces.
- Cisco NX-OS offers a range of features that can be grouped into several categories, such as resiliency, extensibility, security, efficiency, and virtualization.

© 2012 Cisco and/or its affiliates. All rights reserved.

https://t.me/learningnets

Module Summary

This topic summarizes the key points that were discussed in this module.



The Cisco Nexus Family of products is made up of a number of switches. The Cisco Nexus 1000V Series Switches are software-based and reside on a VMware ESXi server, allowing policies to be applied at a virtual machine (VM) level. The Cisco Nexus 3000 Series Switches are targeted at ultrafrequency environments with extreme low latency requirements. The Cisco Nexus 4000 Series Switch is a blade switch for the IBM environment. The Cisco Nexus 2000 Series Fabric Extenders are remote line modules for the Cisco Nexus 5000 and 7000 Series Switches, supporting Gigabit Ethernet and 10 Gigabit Ethernet expansion for server connectivity. The Cisco Nexus 5000 and 5500 Platform switches (within the 5000 series) support Fibre Channel over Ethernet (FCoE). The Cisco Nexus 7000 Series Switches are modular chassis supporting 1, 10, 40, and 100 Gigabit Ethernet line modules. They are multilayer switches that are designed for the high availability and performance requirements of the data center, enterprise, and ISP environments.

The Cisco Unified Fabric solution is supported by all the Cisco Nexus Family of products. Features such as Cisco Overlay Transport Virtualization (OTV), Cisco FabricPath, and I/O consolidation support the convergence, scalability, and intelligence requirements of the business network infrastructures of today.

Businesses can choose to integrate services into the data center by using Integrated Services Modules in the Cisco Catalyst 6500 Series Switches or by using service appliances. A Cisco Catalyst 6500 Series switch can be connected to a Cisco Nexus 7000 Series switch as a services chassis, or you can attach the Cisco ASA adaptive security appliance directly to the Cisco Nexus switch.

This module identified the architectural components, the Cisco Nexus Family of products, and the important features of the Cisco Unified Fabric solution and Cisco Nexus Operating System (NX-OS).

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

Q1) Which multifunction adapter integrates with the Cisco Nexus 5000 and 5500 Platform switches to provide Cisco Unified Fabric convergence? (Source: Describing the Cisco Data Center Network Architecture)

A) network interface card
B) host bus adapter
C) consolidated network adapter
D) converged network adapter

Q2) What is the advantage of having a core layer? (Source: Describing Cisco Data Center Network Architecture)

A) ability to replace the core
B) high-speed interfaces
C) cost savings
D) test bed

Q3) Which Cisco product supports the integration of service modules to create a services chassis? (Source: Describing Cisco Data Center Network Architecture)

A) Cisco Catalyst 4000 Series Switch
B) Cisco Catalyst 6000 Series Switch
C) Cisco Nexus 7000 Series Switch
D) Cisco Catalyst 6500 Series Switch

Q4) Which technology enables data center architects to gain new design flexibility while simplifying cabling infrastructure and management complexity? (Source: Identifying Cisco Nexus Products)

A) Cisco OTV
B) Cisco FabricPath
C) Cisco FEX-Link
D) vPC

Q5) Which Cisco Nexus switch is supported on the IBM BladeCenter H and HT chassis? (Source: Identifying Cisco Nexus Products)

A) Cisco Nexus 4001I Switch Module
B) Cisco Nexus 2000 Series Fabric Extender
C) Cisco Nexus 1000V Switch
D) Cisco Nexus 5000 Series Switch

Q6) Which feature is supported on the Cisco Nexus 5500 Platform switches to provide native Layer 2 multipathing? (Source: Identifying Cisco Nexus Products)

A) IS-IS
B) Extended Port Channels
C) Spanning Tree Protocol
D) Cisco FabricPath

https://t.me/learningnets

Q7) Which feature would be used to extend the Layer 2 domain between multiple network locations? (Source: Identifying Cisco Nexus Products)

A) vPC
B) Cisco OTV
C) Cisco FabricPath
D) TRILL

Q8) Which feature would be used to virtualize the Cisco Nexus 7000 Series Switches hardware? (Source: Identifying Cisco Nexus Products)

A) virtual device contexts
B) VLANs
C) VRFs
D) security contexts

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Module Self-Check Answer Key

Q1) D

Q2) A

Q3) D

Q4) C

Q5) A

Q6) D

Q7) B

Q8) A

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Module 2

Cisco Nexus Switch Feature Configuration

Overview

The Cisco Nexus switches provide highly available, high-performance Layer 2 and Layer 3 switching for both unicast and multicast traffic. It is important for data center engineers to understand how to configure the various Layer 2 and Layer 3 switching functions and associated control protocols. These protocols include Spanning Tree Protocol (STP), PortChannels, and unicast and multicast routing protocols. In addition to describing the implementation of standard Layer 2 and Layer 3 switching, this module also covers the configuration of features that are specific to the Cisco Nexus switches. These features include virtual device contexts (VDCs), PortChannels, virtual port channels (vPCs), and enhanced VPCs. These features enable unique data center network designs. Therefore, a thorough understanding of the operation and configuration of these features is essential to data center network engineers.

Module Objectives

Upon completing this module, you will be able to select and configure the distinctive Cisco Nexus switch features in order to meet the implementation requirements and expectations in the Cisco Data Center Network Architecture. You will be able to meet these objectives:

- Evaluate the service-level and network-level high availability of the Cisco Nexus switches
- Identify how to plan and implement VDCs into the data center solution when given a certain requirement
- Configure the Layer 2 switching features in order to support network requirements when given an implementation plan
- Evaluate how PortChannels and vPCs should be used to improve a particular solution, and configure these features
- Implement and verify Cisco FabricPath on the Cisco Nexus switch
- Implement and verify Layer 3 switching features on the Cisco Nexus switch
- Implement multicast functionality in a Cisco Data Center Network Architecture

https://t.me/learningnets

 © 2012 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 1

Understanding High Availability and Redundancy

Overview

High availability is of critical importance to every data center solution. To prevent or minimize traffic disruption during hardware or software failures, the Cisco Nexus Operating System (NX-OS) software provides a number of features that are based on physical and software redundancy at every component level. The Cisco NX-OS in-service software upgrade (ISSU) feature allows the administrators to perform non-disruptive upgrades. This lesson discusses the high-availability and redundancy features on the Cisco Nexus switches.

Objectives

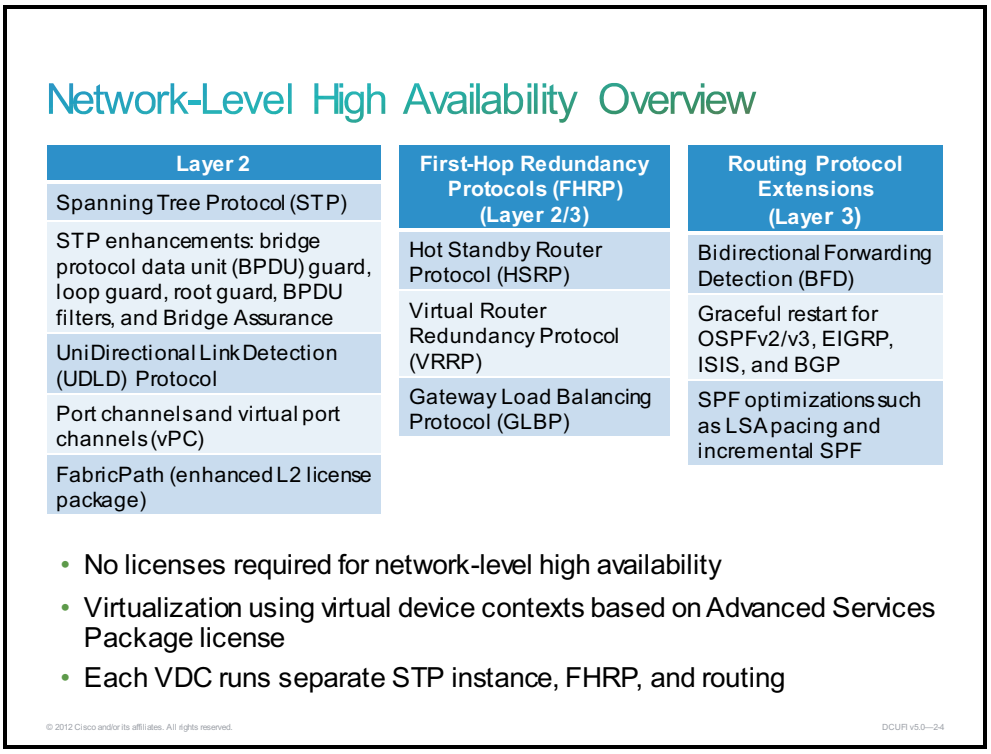
Upon completing this lesson, you will be able to evaluate the service-level and network-level high availability of the Cisco Nexus switches. You will be able to meet these objectives:

- Identify the network-level high-availability features available and how to configure those features on the Cisco Nexus switches
- Identify the system-level high-availability features available and how to configure those features on the Cisco Nexus switches
- Identify how to perform a Cisco NX-OS ISSU

https://t.me/learningnets

Network-Level High Availability

This topic identifies the network-level high-availability features available and how to configure those features on the Cisco Nexus switches.



Network-level high-availability features of Cisco NX-OS Software comprise a set of protocols that span several layers of the Open Systems Interconnection (OSI) model. The features can be grouped into these categories:

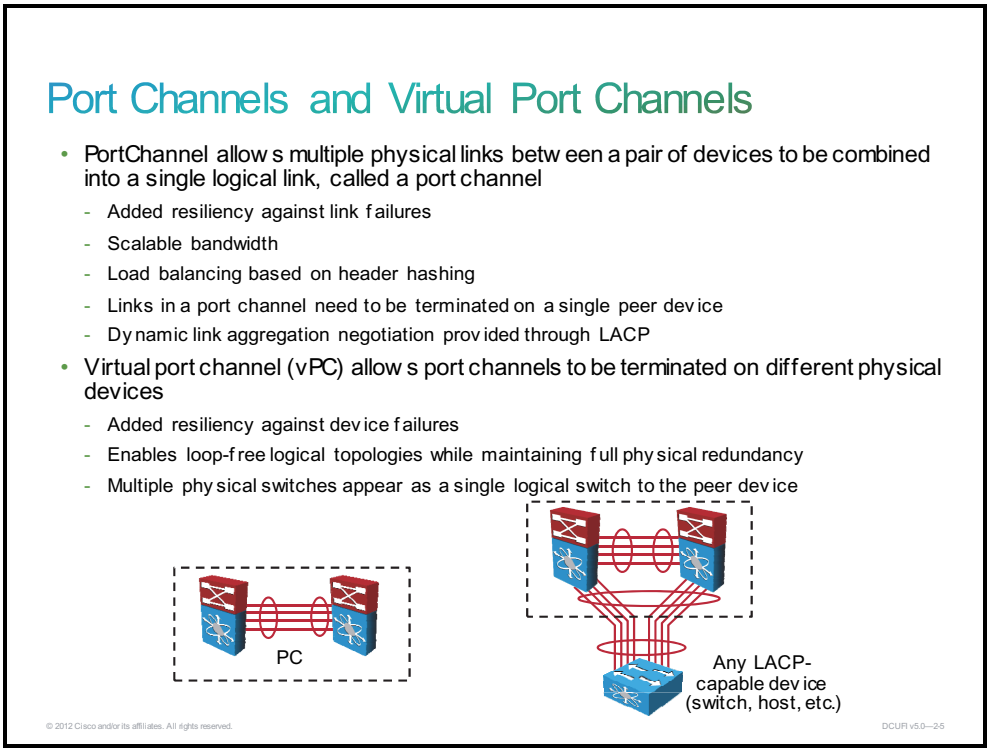
- Layer 2 high-availability features that include the following:
  - Spanning Tree Protocol (STP) and its enhancements, such as bridge protocol data unit (BPDU) guard, loop guard, root guard, BPDU filters, and Bridge Assurance, all of which guarantee the health of the STP control plane
  - UniDirectional Link Detection (UDLD) protocol
  - IEEE 802.3ad Link Aggregation Control Protocol (LACP), the Cisco PortChannel feature, and virtual port channels (vPCs)
- Layer 2 and 3 high-availability features encompass First-Hop Redundancy Protocol (FHRP), Hot Standby Router Protocol (HSRP), Virtual Router Redundancy Protocol (VRRP), and Gateway Load Balancing Protocol (GLBP).
- Cisco FabricPath:
  - – provide ECMP and path convergence without the use of STP
  - requires a separate license for Cisco FabricPath on F-Series modules

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

- Layer 3 high-availability features consist of the following:
  - Bidirectional Forwarding Detection (BFD)
  - Cisco Nonstop Forwarding (NSF) graceful restart extensions for routing protocols Open Shortest Path First (OSPF) versions 2 and 3, Intermediate System-to-Intermediate System (IS-IS), Enhanced Interior Gateway Routing Protocol (EIGRP), and Border Gateway Protocol (BGP) to utilize graceful restart extensions to the base protocols in order to provide NSF and least obtrusive routing recovery for those environments
  - Shortest Path First (SPF) optimizations such as link-state advertisement (LSA) pacing and incremental SPF

https://t.me/learningnets



PortChannel is one of the core technologies that are used in Ethernet-based networks. To add resiliency against link failures and to increase the available bandwidth between two devices, multiple physical links can be provisioned between the devices. However, without PortChannel, control plane protocols, such as STP or routing protocols, treat those links as individual links. In the case of STP, the result is blocked ports, and although the additional links add resiliency, the available bandwidth between the two devices is not increased. In the case of routing protocols, the additional links could be used for load balancing. However, a routing adjacency would have to be formed for every link, which increases routing protocol overhead.

PortChannel combines the physical links into a single logical link, which is called a *port channel*. Control plane protocols, such as STP and routing protocols, treat the port channel as a single link. Spanning tree does not block the links that are part of the port channel, and routing protocols form only a single routing adjacency across the port channel.

Traffic that is switched or routed to a port channel interface is balanced across the individual physical links through a hashing mechanism. The hashing mechanism uses a selection of the fields in the packet headers as input. This process ensures that packets with the same header are forwarded on the same physical link in order to prevent packet reordering.

Link Aggregation Control Protocol (LACP), described in the 802.1AX standard, can be used to dynamically negotiate the aggregation of multiple links into a single port channel.

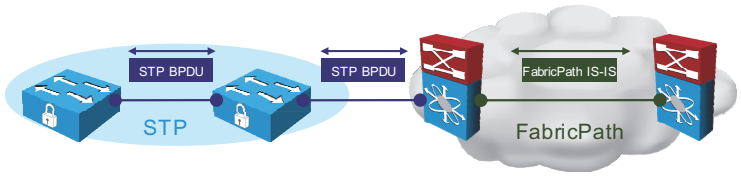
A major restriction of classic PortChannel technology is that it is basically limited to the aggregation of a number of links that run between the same two devices. Regular PortChannel technology does not allow the links of a port channel to be terminated on different devices.

To terminate the links of a port channel on different physical devices, it is necessary for the different physical devices to present themselves as a single logical device to the neighbor. One way this result can be accomplished is by using the vPC solution. vPC allows a pair of Cisco Nexus 5000 or 7000 Series switches to form a logical entity called a *vPC domain*, which presents itself as a single logical switch to devices connected to the vPCs.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Cisco FabricPath

- Cisco FabricPath IS-IS replaces STP as control plane protocol in a Cisco FabricPath network
- IS-IS-based link-state protocol with support for Layer 2 load balancing
- Exchanges reachability of switch IDs and builds forwarding trees
- Improves failure detection, network reconvergence, and high availability
- Minimal IS-IS knowledge required—no user configuration necessary



© 2012 Cisco and/or its affiliates. All rights reserved.

With Cisco FabricPath, you use the Layer 2 IS-IS protocol for a single control plane that functions for unicast, broadcast, and multicast packets. There is no need to run STP, although the network remains purely a Layer 2 domain. Cisco FabricPath Layer 2 IS-IS is a separate process from Layer 3 IS-IS.

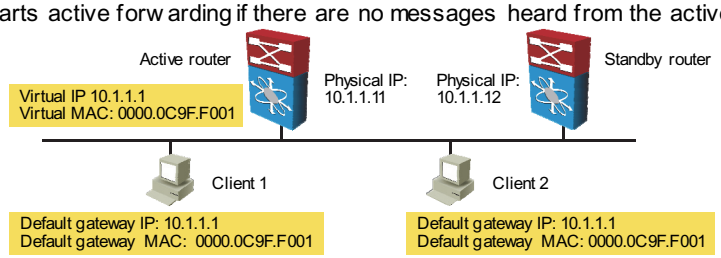
IS-IS provides these benefits:

- **Has no IP dependency:** There is no need for IP reachability in order to form adjacency between devices.
- **Is easily extensible:** Using custom type, length, values (TLVs), IS-IS devices can exchange information about virtually anything.
- **Provides SPF routing:** IS-IS has excellent topology building and reconvergence characteristics.

https://t.me/learningnets

HSRP

- HSRP group (standby group)
  - Set of HSRP devices emulating a virtual router
- Active router
  - Responds to ARP requests of default gateway with MAC of virtual router
  - Assumes the active forwarding of packets for the virtual router
  - Sends hello messages
- Standby router
  - Listens for periodic hello messages
  - Starts active forwarding if there are no messages heard from the active router



© 2012 Cisco and/or its affiliates. All rights reserved.

HSRP uses a group of routers to provide redundancy. Each router is configured with an individual router-specific IP address. All routers in the HSRP group are also configured with a shared group IP address known as the virtual IP (VIP) address. A shared virtual group MAC address is generated from the HSRP group number. This group IP address presents the image of a single, fault-tolerant router to the client network.

Members of the HSRP group send multicast packets that are called *hello packets* to other routers on the same segment. The hello packets are used to elect the active and standby routers as well as to monitor the health of the active router. When the active router ceases to send hello packets, the standby router takes over the active role. A new standby router is then elected if more routers are available. HSRP configuration statements control whether the currently active router is pre-empted by routers that are entering service or returning to service. HSRP also supports object tracking, which modifies the priority of a router based on the availability of an interface or other network object. This change in priority may also trigger a change from one active router to another.

The group IP address is used by clients who must send data out through the HSRP group. When clients use Address Resolution Protocol (ARP) for the MAC address of this default gateway IP address, the active HSRP router responds with the shared virtual MAC (also called VMAC) address. During failover, the standby HSRP router assumes this MAC address, thereby avoiding the need to refresh the ARP cache in client devices.

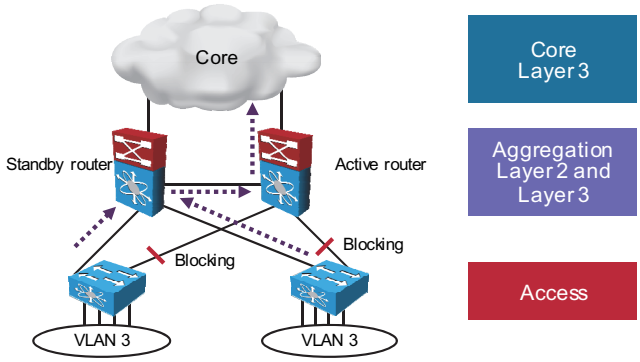
Multiple HSRP groups can be configured on one LAN segment, and different routers can be configured as the default active router for each group. This configuration can be used to provide some traffic load balancing through the routers.

HSRP supports virtual routing and forwarding (VRF), which exists within virtual device contexts (VDCs). Cisco NX-OS Software places you in the default VDC and default VRF unless you specifically configure another VDC and VRF. If you change the VRF membership of an interface, Cisco NX-OS Software removes all Layer 3 configurations, including HSRP.

https://t.me/learningnets

HSRP and STP

- Blocked uplinks may cause traffic to take less-than-optimal path
- Configured active router should be the same as STP root bridge



© 2012 Cisco and/or its affiliates. All rights reserved.

In a redundant spanning-tree topology, some links are blocked. The spanning-tree topology has no awareness of the HSRP configuration. There is no automatic relationship between the HSRP active router election process and the spanning-tree root bridge election process.

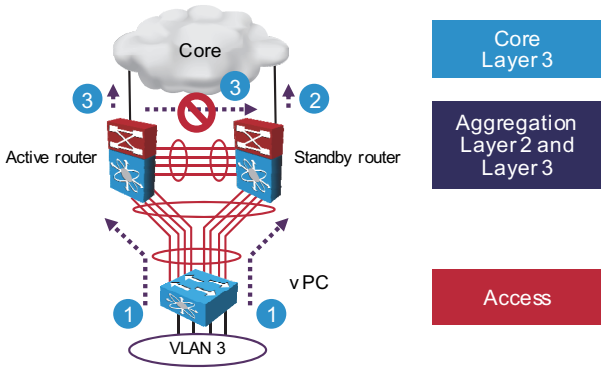
When configuring both the STP and HSRP, you should make sure that the active router is the same as the root bridge for the corresponding VLAN. When the root bridge is different from the HSRP active router, take some time to analyze the uplink path to the active router in order to make sure that no suboptimal path is used.

https://t.me/learningnets

HSRP and vPC

Cisco extension for HSRP/vPC enables optimal flows:

1. Clients load-balance traffic to both vPC uplink switches
2. Active router forwards traffic directly to the core
3. Standby router would have to forward traffic to active router but takes a shortcut



© 2012 Cisco and/or its affiliates. All rights reserved.

Whenever a vPC peer switch needs to forward traffic for a vPC, it will forward it to a local vPC port if possible. Only if it has no active vPC member ports for the vPC does it forward it across the vPC peer link to the other vPC peer switch.

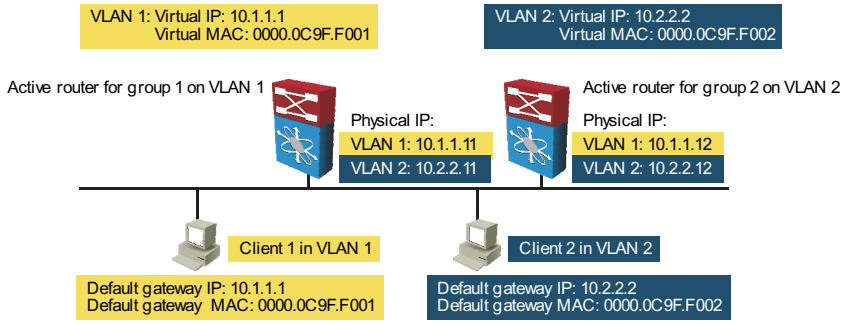
Aggregation switches using vPCs commonly use an FHRP, such as HSRP, as shown in the figure. Normally, only the active HSRP router forwards traffic that is received for the virtual default gateway MAC address. For vPCs, the forwarding rules have been enhanced to allow the standby router to forward frames that are destined for the VMAC address. However, the active device is still responsible for responding to ARP requests. The same modification has been implemented for VRRP and GLBP, which is discussed next.

The result of the enhanced vPC forwarding behavior is that the vPC peer link between the active and standby routers does not carry vPC traffic unless there is a failure.

https://t.me/learningnets

HSRP Load Balancing

- One HSRP group per VLAN
- Active routers from each group are on different aggregation switches
- HSRP active router and STP root for each VLAN should be on the same switch.



© 2012 Cisco and/or its affiliates. All rights reserved.

HSRP routers can simultaneously provide redundant backup and perform load sharing across different IP subnets.

In the figure, two HSRP-enabled routers participate in two separate VLANs. Running HSRP over trunking allows users to configure redundancy among multiple routers that are configured as front ends for VLAN IP subnets.

By configuring HSRP over trunks, you can eliminate situations in which a single point of failure causes traffic interruptions. This feature inherently provides some improvement in overall networking resilience by providing load balancing and redundancy capabilities between subnets and VLANs.

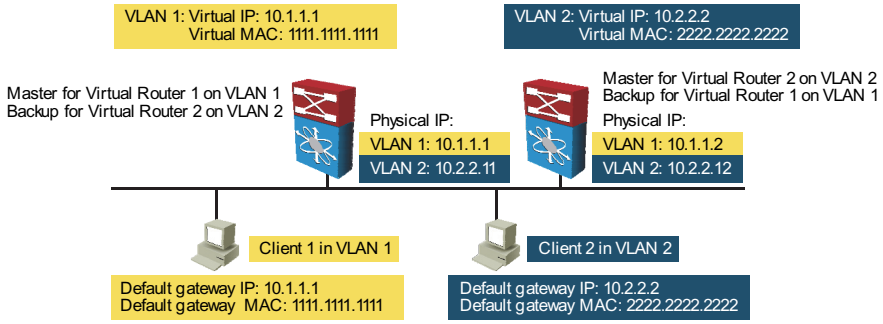
For a VLAN, configure the same device to be both the spanning-tree root and the HSRP active router. This approach ensures that the Layer 2 forwarding path leads directly to the Layer 3 active router and thereby achieves maximum efficiency of load balancing on the routers and the trunks.

For each VLAN, a standby group, an IP address, and a single well-known MAC address with a unique group identifier is allocated to the group. Although up to 255 standby groups can be configured in HSRPv1 (4095 with HSRPv2), the actual number of groups should be kept to a minimum.

https://t.me/learningnets

VRRP

- Standards-based alternative to HSRP
- Minor differences, such as:
  - Virtual router IP can be identical to physical IP address
  - More groups per interface



© 2012 Cisco and/or its affiliates. All rights reserved.

Like HSRP, VRRP allows a group of routers to form a single virtual router. In an HSRP or VRRP group, one router is elected to manage all requests that are sent to the virtual IP address. A VRRP group has one master router and one or more backup routers. Another difference between HSRP and VRRP is is that the VPPR group IP address can be the same as the router-specific IP address for one of the routers.

Despite some minor differences, HSRP and VRRP are very similar in their features and behaviors. The main difference is that HSRP is a Cisco proprietary implementation, whereas VRRP is an open standard. This means that HSRP is usually found in Cisco networks, whereas VRRP is used in multivendor implementations.

The LAN workstations are then configured with the address of the virtual router as their default gateway. This configuration occurs either manually or dynamically via DHCP.

https://t.me/learningnets

VRRP vs. HSRP

HSRP	VRRP
Cisco proprietary (1994)	IETF (1998–2005), RFC 3768
16 groups max	255 groups max
1 active router, 1 standby router, several candidates	1 active, several backups
Virtual IP address is different from active and standby real IP addresses	Virtual IP address can be the same as the real IP address of one of the group members
Uses 224.0.0.2	Uses 224.0.0.18
Can track interfaces or objects	Can track only objects
Default timers: hello, 3 sec; hold time, 10 sec	Default timers: hello, 1 sec; hold time, 3 sec
Authentication supported	Authentication no longer supported

© 2012 Cisco and/or its affiliates. All rights reserved.

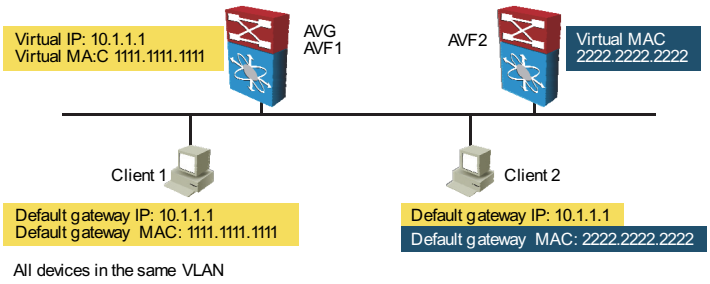
VRRP differs from HSRP in the following ways:

- VRRP is an IEEE standard (RFC 2338 in 1998, and then RFC 3768 in 2005) for router redundancy. HSRP is a Cisco proprietary protocol, created in 1994 and formalized with RFC 2281 in March 1998.
- In VRRP, the virtual router, representing a group of routers, is known as a *VRRP group*.
- In VRRP, the active router is referred to as the *master virtual router*.
- In VRRP, the master virtual router may have the same IP address as the virtual router group.
- In VRRP, multiple routers can function as backup routers.
- Intragroup communications use multicast IP address 224.0.0.2 for HSRP and 224.0.0.18 for VRRP.
- Both HSRP and VRRP can track objects. HSRP can also directly track an interface status, whereas VRRP cannot directly track an interface status. Interfaces can be tracked with VRRP through a tracked object.
- The default timers are shorter in VRRP than HSRP. This fact often gave VRRP a reputation of being faster than HSRP. However, the convergence speed for failover depends on the actual timer configuration.
- HSRP uses authentication within each group by default. When authentication is not configured, a default authentication, using “cisco” as the password, is actually used. Formerly, VRRP used to support plaintext and Hashed Message Authentication Code-Message Digest 5 (HMAC-MD5) authentication methods (RFC 2338). The new VRRP RFC (RFC 3768) removes support for these methods. Nevertheless, current Cisco IOS Software still supports the RFC 2338 authentication mechanisms.

https://t.me/learningnets

GLBP

- Allows use of all devices without creating multiple groups
- Provides a single virtual IP address and multiple virtual MAC addresses
- Routes traffic to a single gateway distributed across routers
  - Active virtual gateway—responds to ARP requests with AVF MAC addresses
  - Active virtual forwarder—actively forwards traffic



© 2012 Cisco and/or its affiliates. All rights reserved.

Although HSRP and VRRP provide gateway resiliency, the upstream bandwidth is not used while the device is in standby mode. Only the active router for HSRP and VRRP groups forwards traffic for the VMAC address. Resources that are associated with the standby router are not fully utilized. You can accomplish some load balancing with these protocols by creating multiple groups and assigning multiple default gateways, but this configuration creates an administrative burden.

Note	A vPC enhancement to HSRP and VRRP removes this constraint, but HSRP and VRRP do not allow load balancing within a single FHRP group.

GLBP is a Cisco proprietary solution that allows the automatic selection and simultaneous use of multiple available gateways in addition to automatic failover between those gateways. Multiple routers share the load of frames that, from a client perspective, are sent to a single default gateway address.

With GLBP, you can fully utilize resources without the administrative burden of configuring multiple groups and managing multiple default gateway configurations, as required with HSRP and VRRP.

There are three GLBP key functions, which are depicted in the figure:

- **Active virtual gateway (AVG)**: Members of a GLBP group elect one gateway to be the AVG for that group. Other group members provide backup for the AVG if the AVG becomes unavailable. The AVG assigns a VMAC address to each member of the GLBP group.
- **Active virtual forwarder (AVF)**: Each gateway assumes responsibility for forwarding packets that are sent to the VMAC address that is assigned to that specific gateway by the AVG. These gateways are known as AVFs for their specific VMAC address.
- **GLBP communication**: GLBP members communicate between each other through hello messages sent every three seconds to the multicast address 224.0.0.102, UDP port 3222.

https://t.me/learningnets

GLBP vs. HSRP

HSRP	GLBP
Cisco proprietary (1994)	Cisco proprietary (2005)
16 groups max	1024 groups max
1 active router, 1 standby router, several candidates	1 AVG, several AVFs; AVG load-balances traffic among AVFs and AVG
Virtual IP address is different from active and standby real IP addresses	Virtual IP is different from AVG and AVF real IP addresses
1 virtual MAC address for each group.	1 virtual MAC address per AVF or AVG in each group
Uses 224.0.0.2.	Uses 224.0.0.102
Can track interfaces or objects	Can track only objects
Default timers: hello, 3 sec; hold time, 10 sec	Default timers: hello, 3 sec; hold time, 10 sec

© 2012 Cisco and/or its affiliates. All rights reserved.

GLBP offers these features:

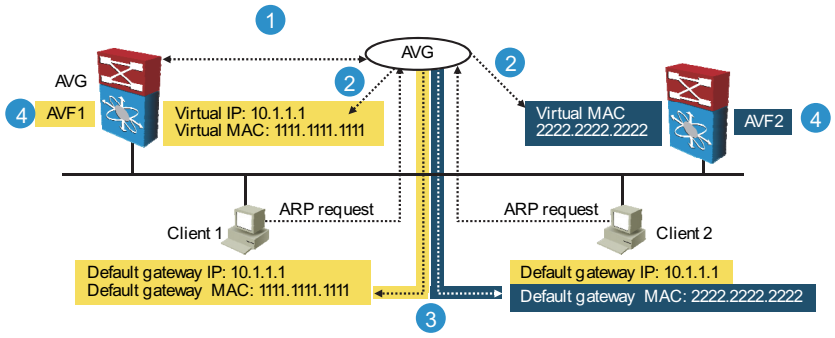
- **Load sharing**: You can configure GLBP in such a way that multiple routers can share traffic from LAN clients, thereby sharing the traffic load more equitably among available routers.
- **Multiple virtual routers**: GLBP supports up to 1024 virtual routers (GLBP groups) on each physical interface of a router and up to four virtual forwarders per group.
- **Pre-emption**: The redundancy scheme of GLBP enables you to pre-empt an AVG with a higher-priority backup virtual gateway that has become available. Forwarder pre-emption works in a similar way, except that forwarder pre-emption uses weighting instead of priority and is enabled by default.
- **Efficient resource utilization**: GLBP makes it possible for any router in a group to serve as a backup, which eliminates the need for a dedicated backup router since all available routers can support network traffic.

GLBP allows automatic selection and simultaneous use of all available gateways in the group. The members of a GLBP group elect one gateway to be the AVG for that specific group. Other members of the group provide backup for the AVG if the gateway becomes unavailable. The AVG assigns a virtual MAC address to each member of the GLBP group. All routers become AVFs for frames that are addressed to that specific virtual MAC address. As clients send ARP requests for the address of the default gateway, the AVG sends these virtual MAC addresses in the ARP replies. (A GLBP group can have up to four group members.)

https://t.me/learningnets

GLBP Operation

1. GLBP group members elect one AVG.
2. AVG assigns a virtual MAC address to each member of the group.
3. AVG replies to the ARP requests from clients with different virtual MAC addresses, thus achieving load balancing.
4. Each router becomes an AVF for frames that are addressed to that virtual MAC address.



© 2012 Cisco and/or its affiliates. All rights reserved.

The figure illustrates how GLBP attempts to balance traffic on a per-host basis by using the round-robin algorithm, which is the default GLBP method.

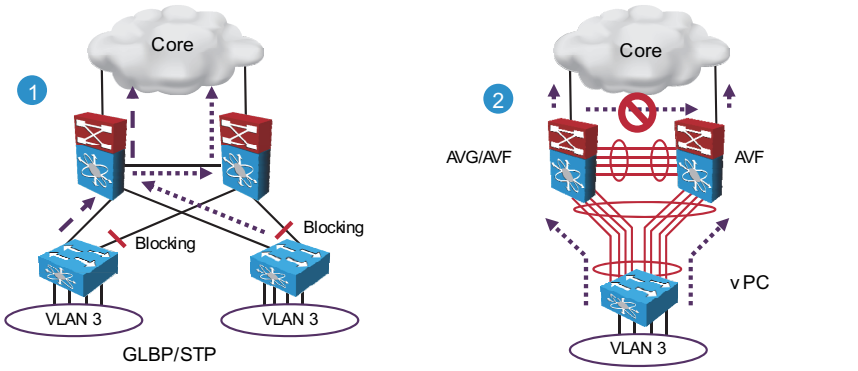
When a client sends an ARP message for the gateway IP address, the AVG returns the virtual MAC address of one of the AVFs. When a second client sends an ARP message, the AVG returns the next virtual MAC address from the list.

After the clients have resolved the default gateway IP address and obtained two different MAC addresses for the default gateway, client A and client B send their routed traffic to separate routers, although they both have the same default gateway address configured. Each GLBP router is an AVF for the virtual MAC address to which it has been assigned.

https://t.me/learningnets

GLBP in STP and vPC Environment

1. GLBP in STP-based Layer 2 domain
   - Both distribution switches act as a default gateway
   - Blocked uplinks may cause traffic to take a less-than-optimal path
2. GLBP in vPC environment
   - vPC enhancement for FHRP ensures optimal forwarding



© 2012 Cisco and/or its affiliates. All rights reserved.

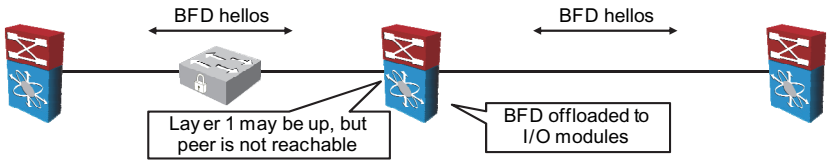
Topologies where STP has blocked one of the access uplinks may result in a two-hop path at Layer 2 for upstream traffic. In the figure, the left graphic shows an interface linking directly to the core in the blocking state. Although it is invisible and transparent to VLAN 2 clients, this state causes the frames that are coming from VLAN 3 attached to the access switch in the right graphic to transit both distribution switches before being sent to the core.

The right graphic shows a GLBP implementation in a vPC environment in which STP is not used at all. In the Cisco GLBP implementation on Nexus switches running vPC, the distribution switches send outbound packets directly to the core instead of forwarding them on to the respective AVFs, as the GLBP protocol would normally require.

https://t.me/learningnets

Bidirectional Forwarding Detection

- Fast, reliable detection of a link failure using frequent link hellos
- Useful for link failures that are not detectable through Layer 1 mechanisms
- Can be tied to Layer 3 control protocols
  - BGP, OSPF, EIGRP, IS-IS, HSRP, and PIM
  - More efficient than the fast hello mechanisms of the individual protocols
- On Cisco Nexus 7000 switches, BFD is run in a distributed manner
  - Offloads the BFD processing to the CPUs on the I/O modules



© 2012 Cisco and/or its affiliates. All rights reserved.

Many Layer 3 control protocols require a fast method of detecting link or node failures in order to achieve fast convergence. In many situations, a link or node failure can be detected through Layer 1 mechanisms. The loss of an optical or electrical signal indicates that a connection to a neighbor has failed. However, there are many other situations where Layer 1 mechanisms cannot be relied on to accurately detect the loss of a link or neighboring device. Therefore, most Layer 3 control protocols use a hello mechanism in order to detect the loss of a neighbor. To achieve fast convergence, network administrators often tune the hello timers of the different Layer 3 control protocols that are used on the network.

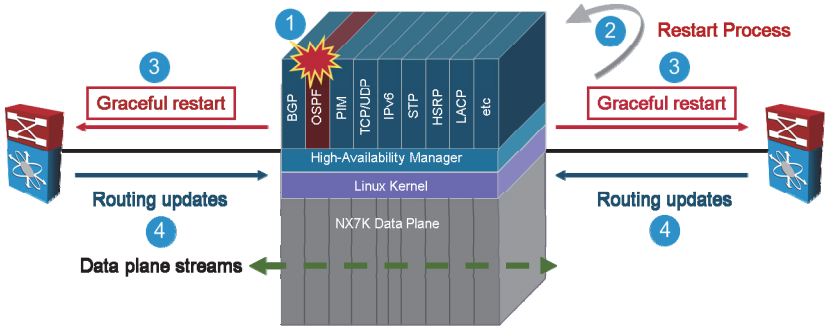
BFD is a detection protocol that is designed to provide fast forwarding path failure detection to Layer 3 protocols. Those protocols include BGP, OSPF, EIGRP, IS-IS, HSRP, Protocol Independent Multicast (PIM), and even static routes.

An advantage of using BFD for fast failure detection instead of tuning the hello timers of all of the Layer 3 protocols is that it allows the switch to detect forwarding path failures at a uniform rate rather than at the variable rates for different protocol hello mechanisms. BFD provides subsecond failure detection between two adjacent devices. BFD can also be less CPU-intensive than individual protocol hello messages because some of the BFD load can be distributed onto the data plane on supported I/O modules on the Cisco Nexus 7000 Series switch.

https://t.me/learningnets

Cisco Nonstop Forwarding (NSF)

- Often referred to as a “graceful restart”
- Uninterrupted traffic forwarding during restart of a control plane process:
  1. Process (such as OSPF) experiences a problem
  2. High-availability manager restarts the process
  3. Graceful restart messages sent to peers
  4. Control plane information received from peers and installed in all tables



© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus 7000 Series switch offers the Cisco NSF extension for the supported routing protocols. The function is also often referred to as a *graceful restart*. Specifically, several situations can occur:

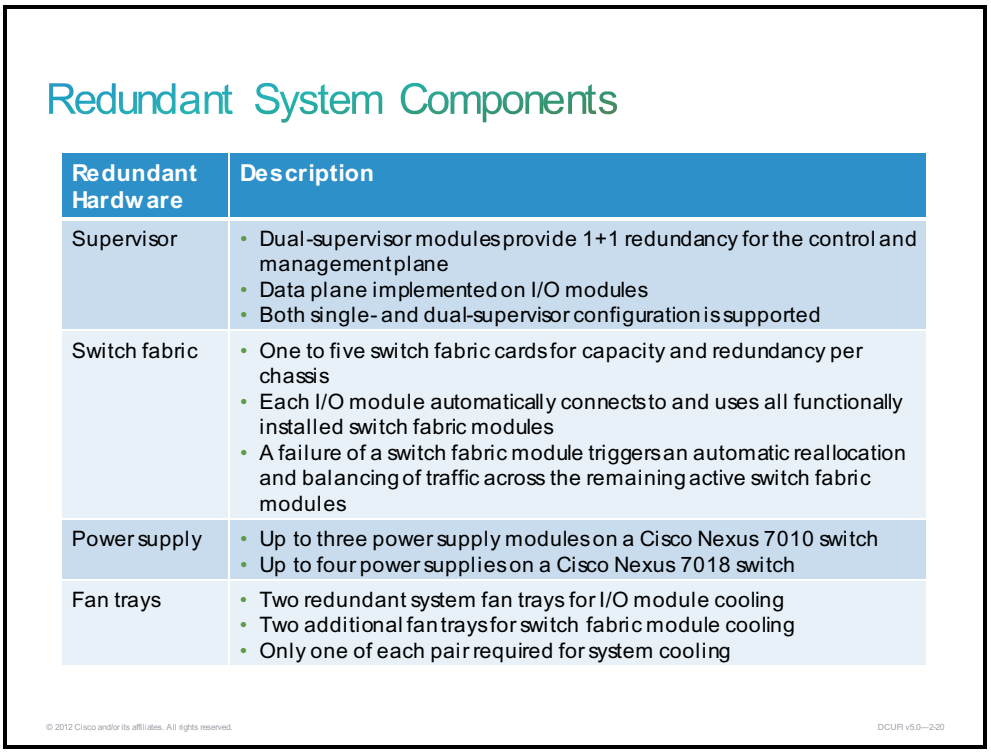
- **Stateless restart**: If a switch experiences a cold reboot, the network stops forwarding traffic to the system and removes the system from the network topology. In this scenario, the process undergoes a stateless restart and removes all neighbor adjacencies on the local system. Cisco NX-OS Software applies the startup configuration, and the routing process rediscovers the neighbors and establishes the adjacencies again.
- **Graceful restart on switchover**: When a supervisor switchover begins, the routing process initiates a graceful restart by announcing that it will be unavailable for a period of time. During the switchover, neighbor devices continue to forward traffic and keep the system in the network topology. After the switchover, Cisco NX-OS applies the running configuration and the routing process informs the neighbors that it is operational again. The neighbor devices help to reestablish adjacencies.
- **Graceful restart on routing process failure**: A routing process automatically restarts if it experiences problems. After the restart, the routing process initiates a graceful restart so that the platform is not taken out of the network topology. If you manually restart the routing process, it performs a graceful restart, which is similar to a Stateful Switchover (SSO). The running configuration is applied in both cases. The graceful restart allows the process to remain in the data forwarding path.

In a graceful restart, if the high-availability manager determines that the best recovery action is to restart, the process then restarts. The restart has no impact on the data plane, and state checkpointing allows instant, stateful process recovery.

https://t.me/learningnets

System-Level High Availability

This topic identifies the system-level high-availability features available and how to configure those features on the Cisco Nexus switches.



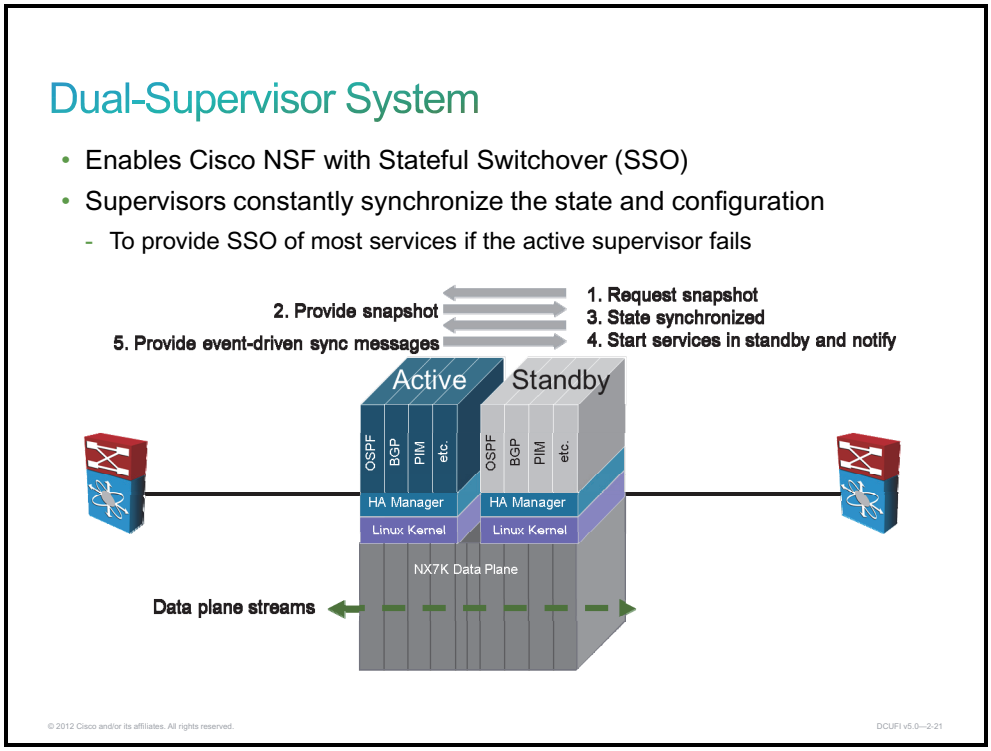
The Nexus 7000 Series switch has the following physical redundancies:

- **Supervisor module redundancy**: The Cisco Nexus 7000 Series switch chassis supports dual supervisor modules in order to provide redundancy for the control and management plane. A dual supervisor configuration operates in an active/standby capacity in which only one of the supervisor modules is active at any given time, while the other acts as a standby backup. The state and configuration remain constantly synchronized between the two supervisor modules so as to provide a statefu1 switchover if the active supervisor module fails.
- **Fabric redundancy**: Cisco NX-OS provides switching fabric availability through redundant switch fabric modules. You can configure a single Cisco Nexus 7000 Series switch chassis with one to five switch fabric cards for capacity and redundancy. Each I/O module installed in the system automatically connects to and uses all functionally installed switch fabric modules. The failure of a switch fabric module triggers an automatic reallocation and balancing of traffic across the remaining active switch fabric modules. Replacing the failed fabric module reverses this process. When you insert the fabric module and bring it online, traffic is again redistributed across all installed fabric modules, and redundancy is restored.
- **Power supply redundancy**: The Cisco Nexus 7000 Series switch chassis supports three power supply modules on a Cisco Nexus 7010 chassis and up to four power supplies on a Cisco Nexus 7018 chassis, each of which is composed of two internalized isolated power units. This design gives it two power paths per modular power supply and six paths in total, per chassis, when fully populated.
- **Fan tray redundancy:** The Cisco Nexus 7010 chassis contains two redundant system fan trays for I/O module cooling and two redundant fan trays for switch fabric module cooling.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

One of each pair of fan trays is sufficient to provide system cooling. There is no time limit for replacing a failed Cisco Nexus 7010 fan tray, but to ensure the proper air flow, you must leave the failed fan tray in place. The Cisco Nexus 7018 chassis contains two fan trays, each of which is required to cool the modules in the chassis. The upper fan tray cools slots 1 to 9 as well as the fabric modules. The lower fan tray cools slots 10 to 18. Each of these fan trays is hot-swappable, but you must replace a fan tray within three minutes of removal or the switch will shut down.



When two supervisors are installed in the Cisco Nexus 7000 Series switch, one is active and the other is a standby. This setup allows Cisco NSF with SSO when a supervisor-level failure occurs. The standby supervisor requests a snapshot of the configuration and services states. Once the state is synchronized between the active and standby supervisor, the standby supervisor starts the services in standby mode and notifies the active supervisor that it is in the hot standby state. During normal operation, the active supervisor provides event-driven synchronization messages to the standby supervisor. The two supervisors constantly synchronize the state and configuration in order to provide a seamless SSO of most services if the active supervisor module fails.

Restarts on Single Supervisors

In a system with only one supervisor, when all high-availability policies have been unsuccessful in restarting a service, the supervisor then restarts. The supervisor and all services then reset and start with no prior state information.

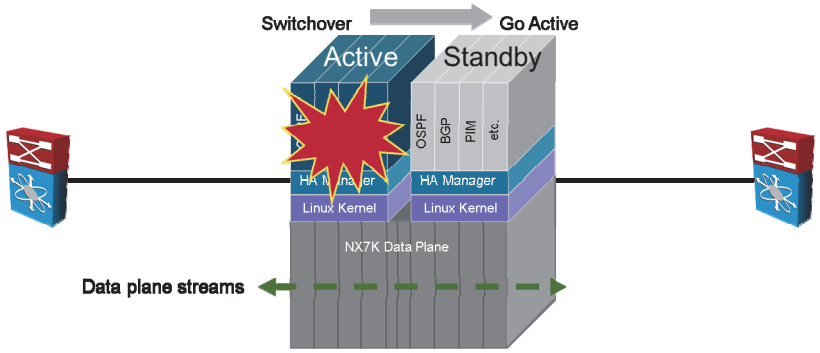
Restarts on Dual Supervisors

When a supervisor-level failure occurs in a system with dual supervisors, the the NX-OS System Manager performs a switchover rather than a restart in order to maintain stateful operation. In some cases, however, a switchover may not be possible at the time of the failure. For example, if the standby supervisor module is not in a stable standby state, a restart rather than a switchover is performed.

https://t.me/learningnets

Supervisor Switchover

- Can be manual or automatic (triggered by failure of the active supervisor)
- Stateful (nondisruptive) because control traffic is not affected
- Does not disrupt data traffic because I/O modules are not affected
  - Switching modules are not reset
- No reload of the Connectivity Management Processor (CMP)



© 2012 Cisco and/or its affiliates. All rights reserved.

Switchovers occur by one of the following two mechanisms:

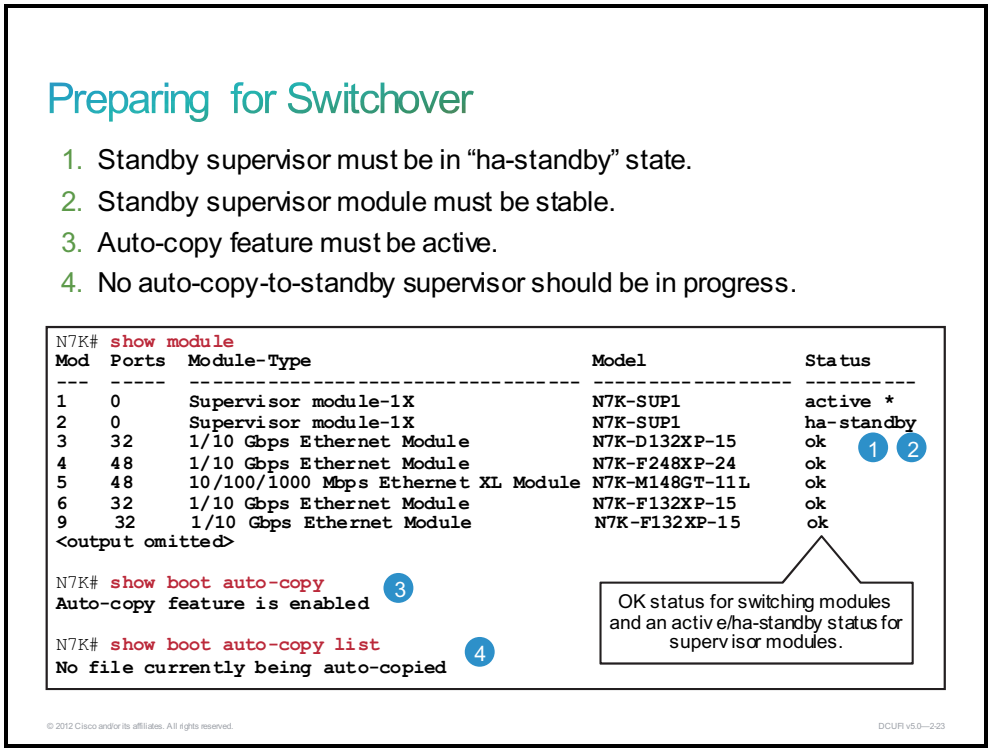
- The active supervisor module fails, and the standby supervisor module then automatically takes over.
- You manually initiate a switchover from an active supervisor module to a standby supervisor module.

When a switchover process begins, another switchover process cannot be started on the same switch until a stable standby supervisor module is available.

A high-availability switchover has the following characteristics:

- It is stateful (nondisruptive) because control traffic is not affected.
- It does not disrupt data traffic because the switching modules are not affected.
- Switching modules are not reset.
- It does not reload the Connectivity Management Processor (CMP).

https://t.me/learningnets

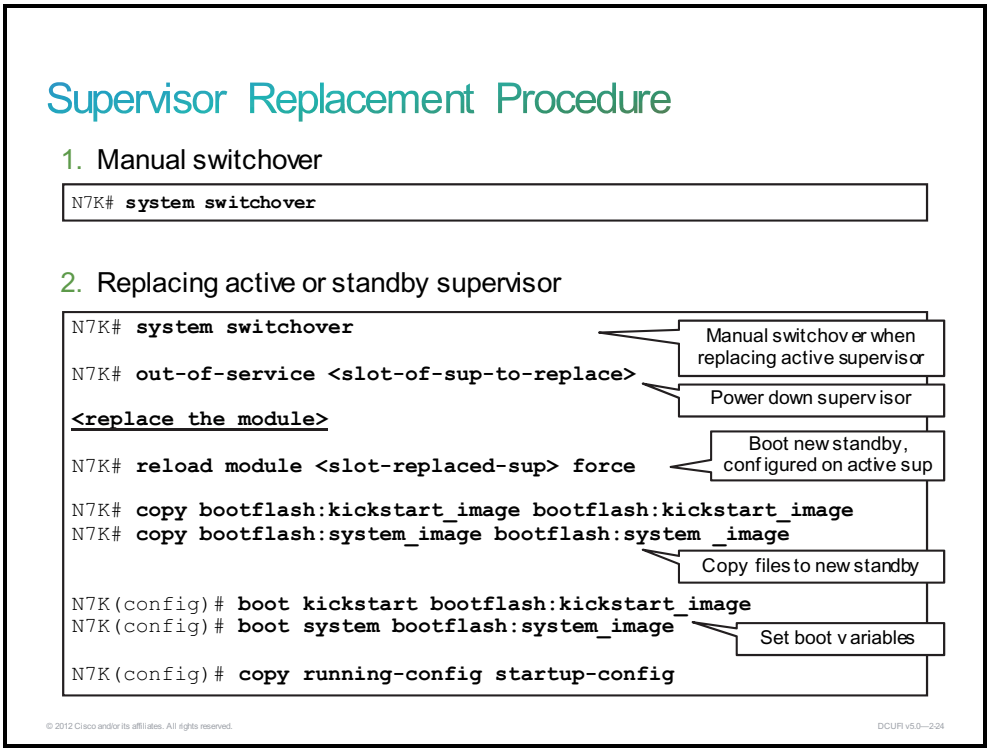


You should verify the status of the switch and the modules before a switchover—either manual or automatic. If the standby supervisor module is not in a stable ha-standby state, neither a manual nor an automatic switchover is performed.

You can use the **show system redundancy status** command to ensure that the system is ready to accept a switchover or the **show module** command to verify the status and presence of the installed modules. A sample output of the **show module** command is shown in the figure. The Status column in the output displays a status of "OK" for the switching modules, "active" for the active supervisor, and "ha-standby" for the standby supervisor module.

Furthermore, you should use the **show boot auto-copy** command to verify the configuration of the auto-copy feature and make sure that no auto-copy to the standby supervisor module is in progress. Sample outputs of the **show boot auto-copy** command are included in the figure.

https://t.me/learningnets



To manually initiate a switchover from an active supervisor module to a standby supervisor module, use the **system switchover** command. The switchover will start immediately. After you run this command, you cannot start another switchover process on the same system until a stable standby supervisor module is available.

Replacing the Active Supervisor

Use this procedure to nondisruptively replace the active supervisor module in a dual-supervisor system:

Step 1 Initiate a manual switchover to the standby supervisor by using the **system switchover** command. Wait until the switchover completes and the standby supervisor becomes active.

Step 2 Power down the supervisor module that you are replacing by using the **out-of-service** command.

Step 3 Replace the supervisor module.

Step 4 Boot the supervisor module replacement by using the **reload module** command. If you do not force the boot, the replacement supervisor module should be booted by the active supervisor module six minutes after insertion.

Step 5 Copy the kickstart and system images from the active supervisor module to the standby supervisor module.

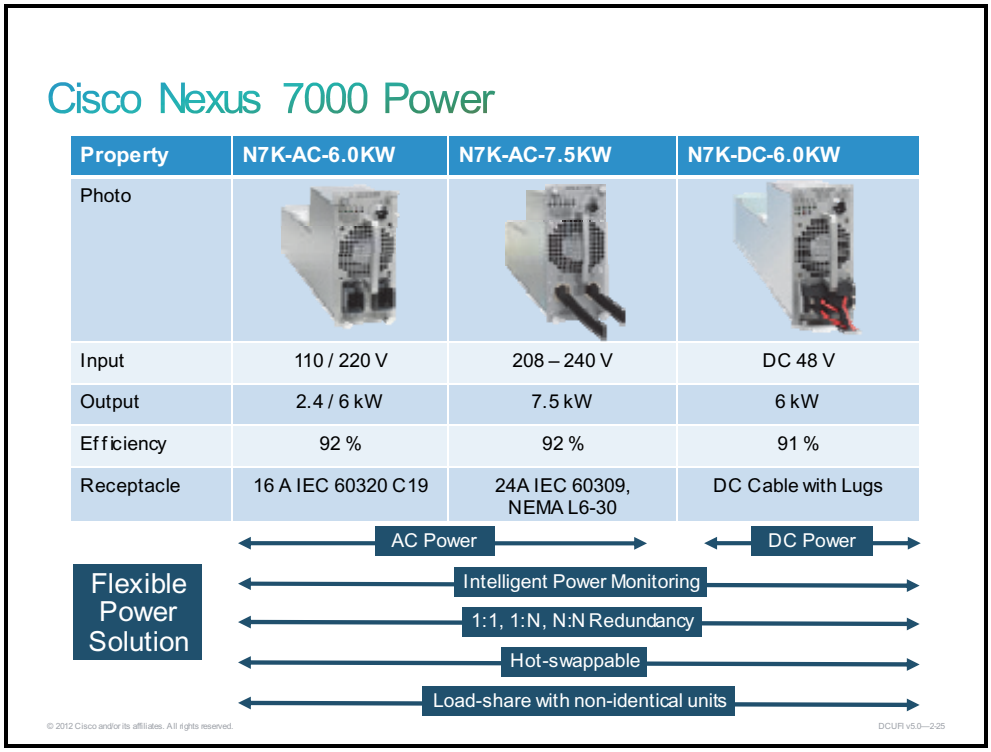
Step 6 Enter global configuration mode and configure the standby supervisor kickstart and system boot variables.

Step 7 Save the change so that it persists through reboots and restarts by copying the running configuration to the startup configuration.

Replacing the Standby Supervisor

The procedure to replace the standby supervisor is nearly identical to replacing the active supervisor except that the first step is not needed: You do not need to initiate a manual switchover.

https://t.me/learningnets



The 6-kW AC power supply module is the first of the Cisco Nexus 7000 Series switch power supplies. Each 10-slot chassis can hold up to three load-sharing, fault-tolerant, hot-swappable power supplies. The power supplies are located in the rear of the chassis for easy installation and removal without obstruction by the cables at the front.

The power supply module is a dual 20-A AC input unit providing the following:

- **Single input**: 220-V, 3000-W output, 110-V, 1200-W output
- **Dual input**: 220-V, 6000-W output, 110 V, 2400-W output
- **Dual input**: 110- and 220-V, 4200-W output

The power supply has four user-configurable power redundancy modes.

Key features of the power supply include the following:

- Multiple inputs providing redundancy if one input fails
- Universal input providing flexibility
- Compatibility with future Cisco Nexus 7000 Series Chassis
- Hot-swappable so that no downtime is needed when replacing power supplies
- Temperature sensor and instrumentation that shut down the power supply if the temperature exceeds the thresholds, thereby preventing damage due to overheating
- Internal fault monitoring so that if a short circuit and component failure is detected, the power supply unit can be shut down automatically
- Intelligent remote management so that users can remotely power-cycle one or all power supplies using the supervisor CLI
- Real-time power draw showing real-time actual power consumption (not available in the initial software release)
- Variable fan speed, allowing reduction in fan speed for lower power usage in well-controlled environments

https://t.me/learningnets

Cisco Nexus 7000 7.5-kW AC Power Supply Module

The Cisco Nexus 7000 7.5-kW AC Dual 30A power supply module that is shown in the figure delivers fault-tolerant, high-efficiency, load-sharing, and hot-swap features to the Cisco Nexus 7000 Series switch. Each Cisco Nexus 7000 Series Chassis can accommodate multiple power supplies, providing both chassis-level and facility power fault tolerance.

Cisco Nexus 7000 6.0-kW DC Power Supply Module

The Cisco Nexus 7000 Series switch 6.0-kW power supply is designed for DC environments. This power supply is a variable-output, high-capacity power supply scalable from 3000 to 6000 W, delivering fault-tolerant load-sharing capability. The power supplies are hot-swappable.

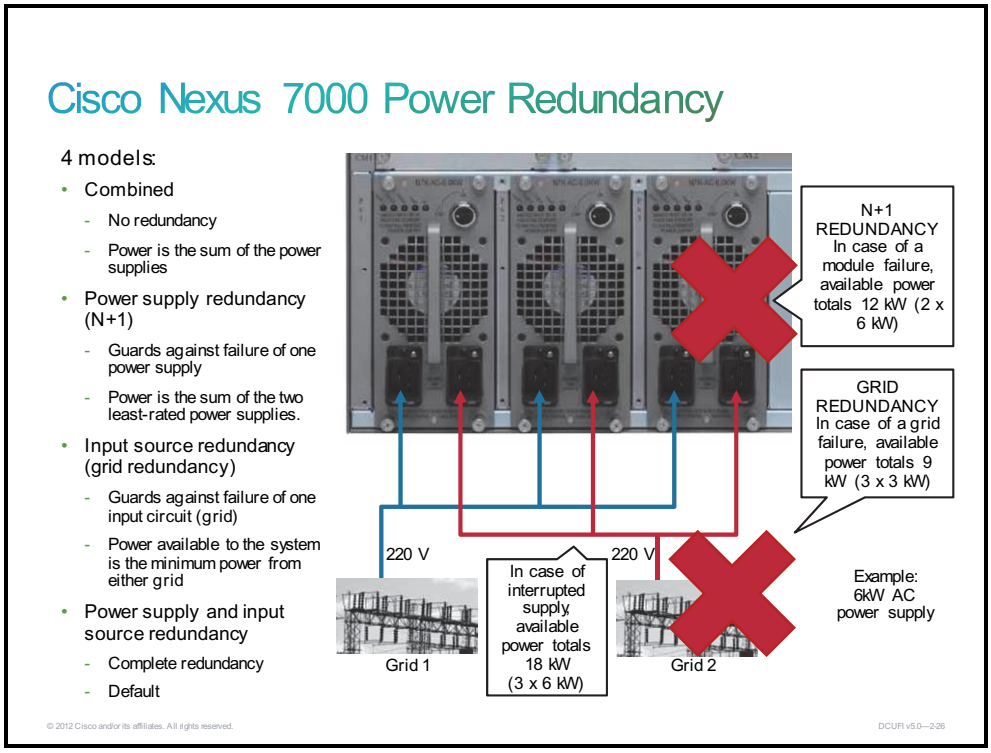
DC power connections are made using a hot-swap DC power cable that enables quick and easy installation of the power supplies without the need to disturb the DC terminal blocks. The DC cable supports both direct connection to DC power sources and connection to an intermediate power interface unit in situations in which connections to the source are beyond the cable length.

The Cisco Nexus 7000 Series switch DC power interface unit (PIU) is an optional element that is provided for environments in which the Cisco Nexus 7000 Series switch DC cable needs to connect to existing DC power cabling and provides 16 two-pole terminal connections. The PIU supports one or two Cisco Nexus 7000 6.0 kW DC power supply modules, with each power supply using two DC power cables for a total of four connections to the PIU.

The Cisco Nexus 7009 chassis supports up to three power supplies in a single chassis, the Cisco Nexus 7010 chassis supports up to three power supplies, and the Cisco Nexus 7018 chassis supports up to four power supplies.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



The Cisco Nexus 7000 Series switch is powered by three internally redundant power supplies. Each of these individual power supply units is composed of two internalized isolated power units, effectively giving it two power paths per modular power supply and six paths in total, per chassis, when fully populated. The power supplies use a proportional load-sharing method for power distribution in order to power system components, thus allowing the efficient use of dissimilar capacity power supplies in the same chassis. Therefore, all installed power supplies are active and share the overall load. Additionally, the power subsystem allows the three power supplies to be configured in any one of four redundancy modes:

- **Combined**: Combined mode has no redundancy, with the power available to the system being the sum of the power outputs of all of the power supplies in the chassis.
- **Power supply redundancy (N+1)**: N+1 redundancy guards against the failure of one of the power supplies. Power available to the system is the sum of the two least-rated power supplies.
- **Input source redundancy (grid redundancy)**: Grid redundancy guards against failure of one input circuit (grid). For grid redundancy, each input on the power supply is connected to an independent AC feed, and power available to the system is the minimum power from either of the input sources (grids).
- **Power supply and input source redundancy (complete redundancy)**: Complete redundancy is the system default redundancy mode. This mode guards against failure of either one power supply or one AC grid. The power available is always the minimum of input source and power supply redundancy.

https://t.me/learningnets

Cisco Nexus 7000 Cooling Redundancy

Redundant system fan trays provide cooling of I/O modules and supervisor engines.

N7K-C7010-FAN-S

N7K-C7010-FAN-F

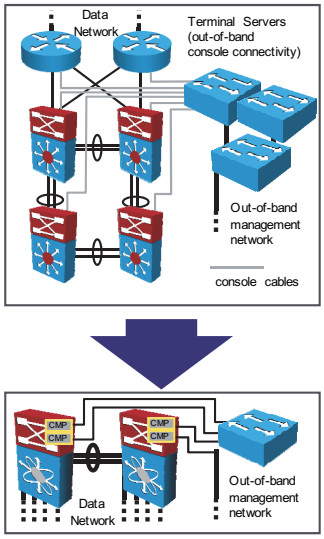
Redundant fabric fans provide cooling of crossbar fabric modules.

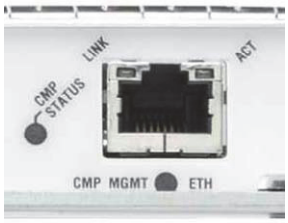
© 2012 Cisco and/or its affiliates. All rights reserved.

Variable-speed redundant fans provide complete system cooling. The fans are located in the rear of the chassis so that no disruption to cabling occurs if a fan module needs to be removed.

All fans are hot-swappable, with a blue beacon LED for easy identification.

Connectivity Management Processor





*CMP is available only on Sup1

© 2012 Cisco and/or its affiliates. All rights reserved.

The CMP provides out-of-band management and monitoring capability, which is independent from the primary operating system. The CMP enables lights-out Remote Monitoring (RMON) and management of the supervisor module, all other modules, and the Cisco Nexus 7000 Series switch system without the need for separate terminal servers.

https://t.me/learningnets

The CMP delivers the remote control through:

- A dedicated processor
- Its own memory
- Its own bootflash memory
- A separate Ethernet management port

In addition, the CMP can reset all system components, including power supplies, and it can reset the host supervisor module to which it is attached, thereby allowing a complete system restart if necessary.

CMP is available only on the Supervisor-1 module.

These are key features of the CMP:

- **Dedicated operating environment**: Independent remote system management monitoring capabilities
- **Monitoring of supervisor status and initiation of resets**: Removes the need for separate terminal server devices for out-of-band management
- **System reset while retaining out-of-band Ethernet connectivity**: Complete visibility during the entire boot process
- **Capability to initiate a complete system power shutdown and restart**: No local operator intervention required
- **Login authentication**: Provides secure access to the out-of-band management environment
- **Access to supervisor logs**: Access to critical log information enables rapid detection and prevention of potential system problems
- **Control capability**: Ability to take full console control of the supervisor
- **Dedicated front-panel LEDs**: CMP status clearly identified separately from the supervisor

https://t.me/learningnets

System Management Tools for Availability

Tool	Description
Cisco Generic Online Diagnostics (GOLD)	• Subsystem and additional monitoring processes on the supervisor • Triggers a stateful failover to the redundant supervisor upon the detection of unrecoverable critical failures, service restartability errors, kernel errors, or hardware failures
Cisco IOS Embedded Event Manager (EEM)	• Consists of Event Detectors, the Event Manager, and an Event Manager Policy Engine • Takes specific actions when the system software recognizes certain events through the Event Detectors. • Set of tools to automate many network management tasks. • Can improve availability, event collection, and notification
Smart Call Home	• Combines Cisco GOLD and Cisco EEM capabilities • Provides an e-mail-based notification of critical system events. • Method variety: pager, email, XML, and direct case to Cisco TAC

© 2012 Cisco and/or its affiliates. All rights reserved.

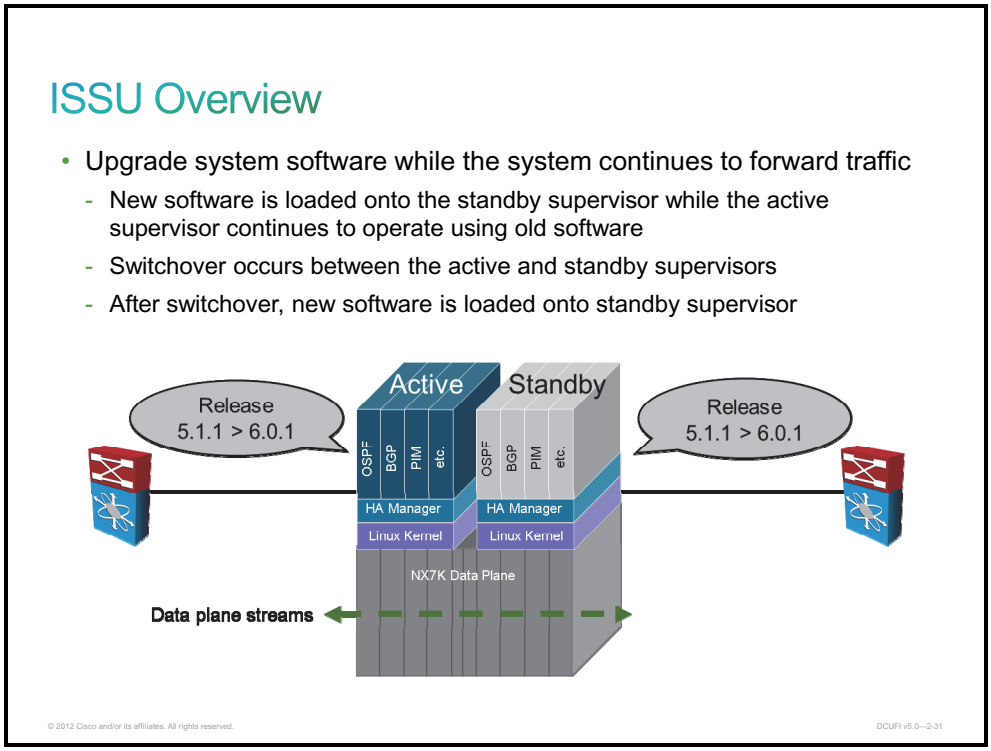
Cisco NX-OS incorporates several system management tools for monitoring and notification of system availability events.

- Cisco Generic Online Diagnostics (GOLD) subsystem and additional monitoring processes on the supervisor facilitate the triggering of a stateful failover to the redundant supervisor upon detection of unrecoverable critical failures, service restartability errors, kernel errors, or hardware failures.
- Cisco IOS Embedded Event Manager (EEM) consists of Event Detectors, the Event Manager, and an Event Manager Policy Engine. Using EEM, you can define policies to take specific actions when the system software recognizes certain events through the Event Detectors. The result is a flexible set of tools to automate many network management tasks and to direct the operation of Cisco NX-OS to increase availability, collect information, and notify external systems or personnel about critical events.
- Combining Cisco GOLD and Cisco EEM capabilities, Smart Call Home provides email-based notification of critical system events. Smart Call Home has message formats that are compatible with pager services, standard email, or XML-based automated parsing applications. You can use this feature to page a network support engineer, email a network operations center, or automatically generate a case with the Cisco Technical Assistance Center (TAC).

https://t.me/learningnets

Cisco IOS In-Service Software Upgrade

This topic identifies how to perform a Cisco IOS In-Service Software Upgrade (ISSU).



Cisco NX-OS provides the ability to perform Cisco IOS in-service software updates (ISSUs). A Cisco IOS ISSU allows you to perform complete image upgrades, nondisruptively and without impacting the data-forwarding plane. This capability allows for NSF during a software upgrade, including upgrades between full-image versions (for example, from 6.0 to 6.1). Sometime a graduate upgrade (from one software version to another) is needed.

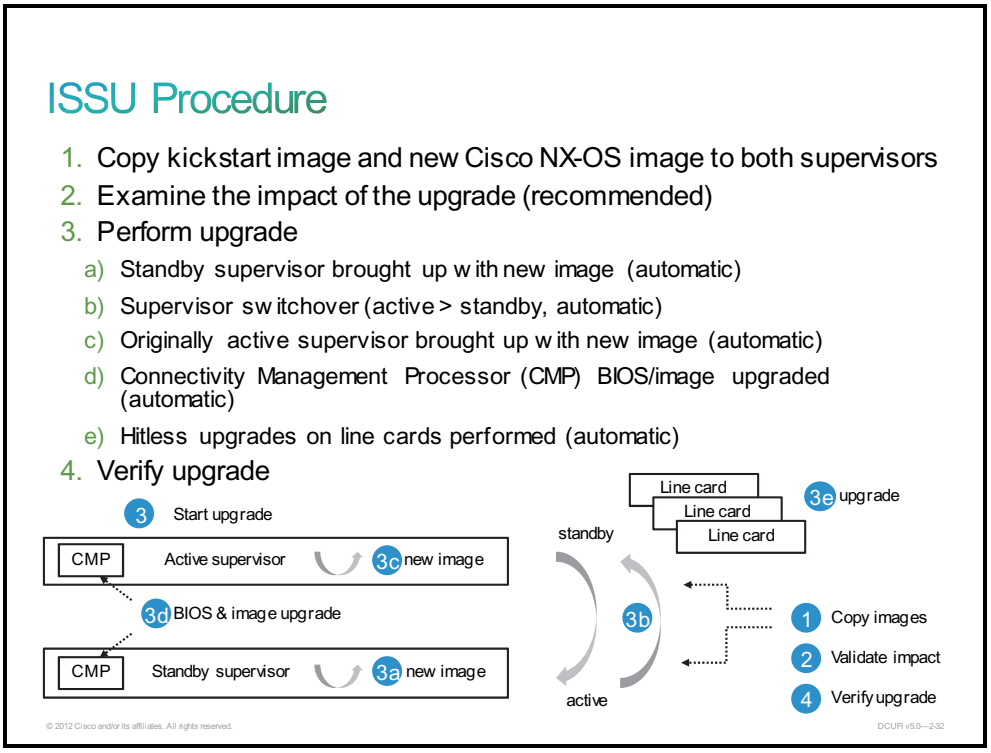
A Cisco IOS ISSU uses the existing features of NSF with Stateful Switchover (SSO).

Note	Prior to upgrade, it is strongly recommended to read the release notes.

The user is notified about the process but does not need to configure each step individually. During a Cisco IOS ISSU, the new software is loaded onto the standby supervisor while the active supervisor continues to operate using the old software. As part of the upgrade, a switchover occurs between the active and standby supervisors, and the standby supervisor then becomes active and begins running the new software. After the switchover, the new software is loaded onto the (formerly active) standby supervisor.

Note	It is still recommended to perform an upgrade in the dedicated maintenance period. Any topology change during a Cisco IOS ISSU will cancel the upgrade or, if it is already too far into the procedure, even continue with the disruptive upgrade.

https://t.me/learningnets



A Cisco IOS ISSU is initiated manually either through the CLI by an administrator, or via the management interface of the Data Center Network Manager (DCNM) software platform.

Follow this procedure to perform a Cisco IOS ISSU:

Step 1 Copy the kickstart image and new Cisco NX-OS image to both supervisors.

Step 2 Examine the impact of the upgrade (recommended).

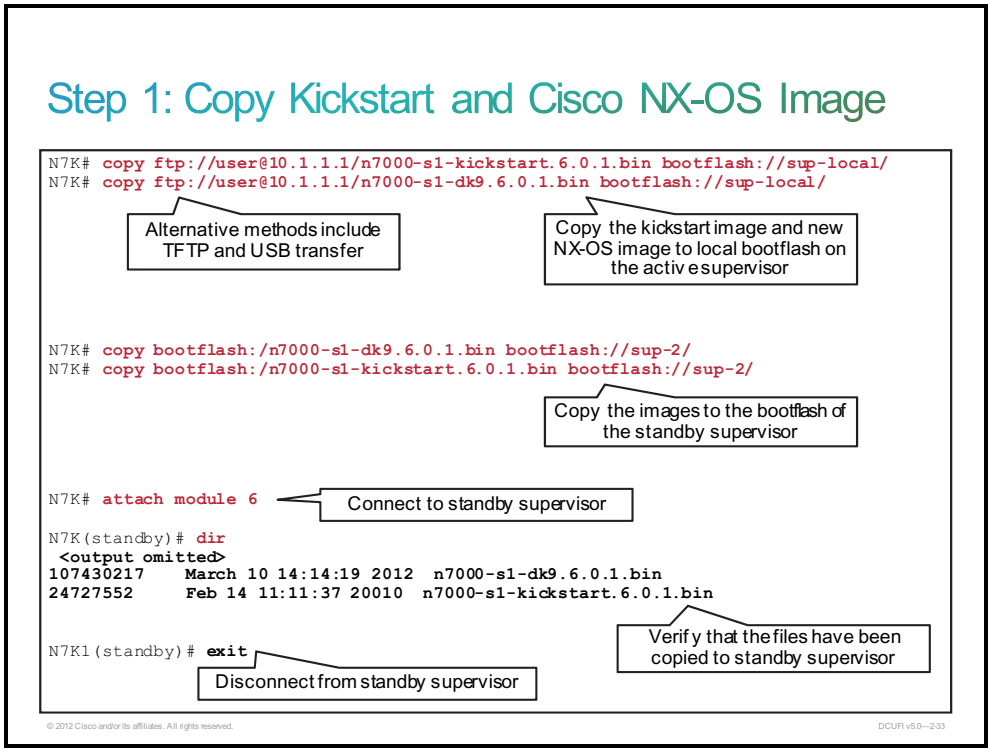
Step 3 Perform the upgrade. Once initiated, the ISSU Installer service begins the Cisco IOS ISSU cycle. The upgrade process is composed of several phased stages that are designed to minimize overall system impact, with no impact to data traffic forwarding. The upgrade encompasses these phases, which are transparent to the user:

1. Standby supervisor brought up with new image
2. Supervisor switchover (active to standby)
3. Originally active supervisor brought up with new image
4. CMP BIOS/image upgraded
5. Hitless upgrades on line cards performed

Step 4 Verify the upgrade.

These steps are described in more detail next.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



First, you need to download the new kickstart and Cisco NX-OS images. You can obtain these from Cisco.com and copy them to the local bootflash using FTP, TFTP, or USB-based storage. Both files need to be also copied to the bootflash of the standby supervisor, as shown in the figure.

After you copy the files to both bootflash locations, you should verify that they have been copied to the bootflash of the standby supervisor by attaching to the standby supervisor module and viewing the directory content. You can then disconnect from the standby supervisor by using the **exit** command.

https://t.me/learningnets

Step 2: Examine Impact of Upgrade

```
N7K# show install all impact kickstart bootflash:n7000-s1-kickstart.6.0.1.bin
system bootflash:n7000-s1-dk9.6.0.1.bin

Verifying image bootflash:/n7000-s1-kickstart.6.0.1.bin for boot variable
"kickstart".
 [####################] 100% — SUCCESS

Verifying image bootflash:/n7000-s1-dk9.6.0.1.bin for boot variable "system".
 [####################] 100% — SUCCESS
<output omitted>

Compatibility check is done:
 Module  bootable  Impact  Install-type  Reason
 ------  --------  ------  ------------  ------
 2       yes       non-disruptive       rolling
 3       yes       non-disruptive       rolling
 4       yes       non-disruptive       rolling
 5       yes       non-disruptive         reset
 6       yes       non-disruptive         reset
 7       yes       non-disruptive       rolling
 8       yes       non-disruptive       rolling
 9       yes       non-disruptive       rolling
 10      yes       non-disruptive       rolling

<output omitted>
```

Only the supervisors will have to be reset (mod 5 and 6)

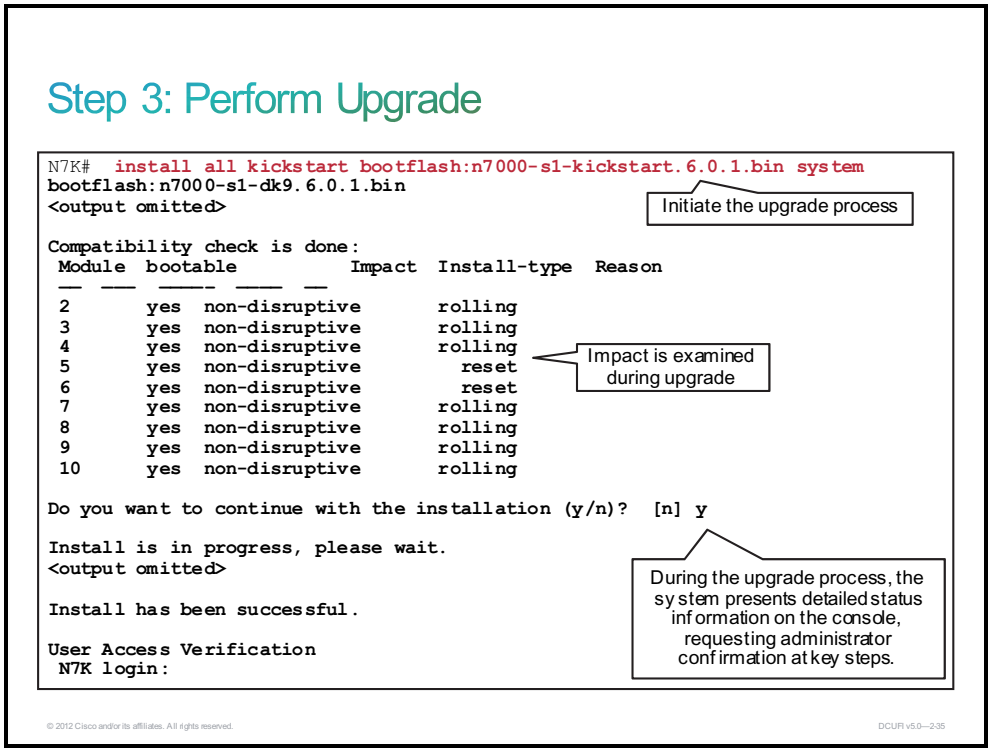
Upgrade is non-disruptive to traffic

© 2012 Cisco and/or its affiliates. All rights reserved.

A good practice is to examine the impact of the upgrade by using the **show install all impact** command. The output informs you about the actions that are required on each module within the chassis. In the Cisco IOS ISSU example in the figure, only the supervisors (in slots 5 and 6) will be automatically reset. No other modules will be affected by the upgrade.

The impact of the upgrade is also evaluated when you start the upgrade in the next step.

https://t.me/learningnets

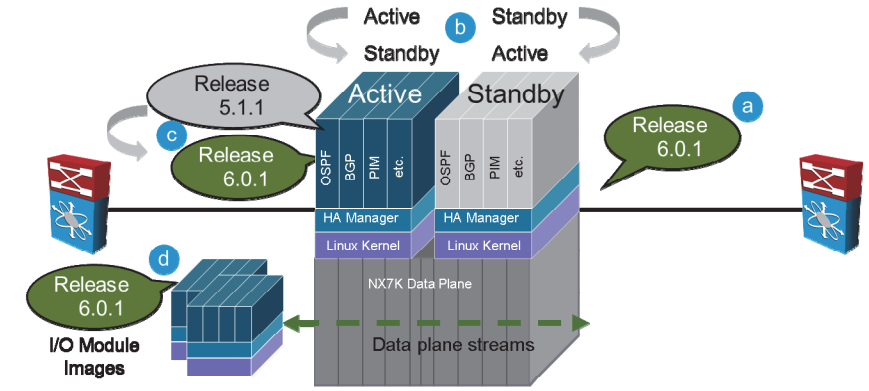


You perform the Cisco IOS ISSU from the CLI by using the **install all kickstart** command. The Cisco IOS ISSU performs a compatibility check first and then proceeds with several phases. During the upgrade process, the system presents detailed status information on the console, thus requesting administrator confirmation at critical steps.

https://t.me/learningnets

Automatic Steps at Upgrade

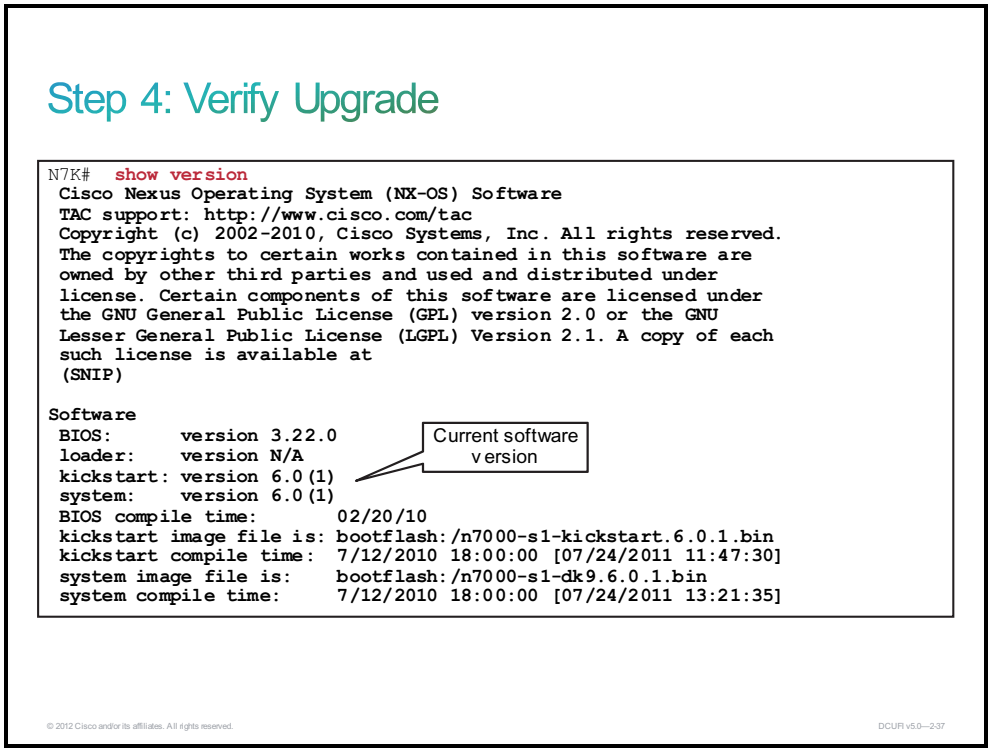
a. Standby supervisor brought up with new image
b. Supervisor switchover (active > standby, standby > active)
c. Originally active supervisor brought up with new image
d. CMP and I/O module images upgraded



A Cisco IOS ISSU process performs these tasks:

- Verifies the location and integrity of the new software image files
- Verifies the operational status and the current software versions of both supervisors and all switching modules to ensure that the system is capable of a Cisco IOS ISSU
- Forces a supervisor switchover
- Brings up the originally active supervisor with a new image
- Performs a nondisruptive upgrade of each switching module—one at a time
- Upgrades the CMP

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

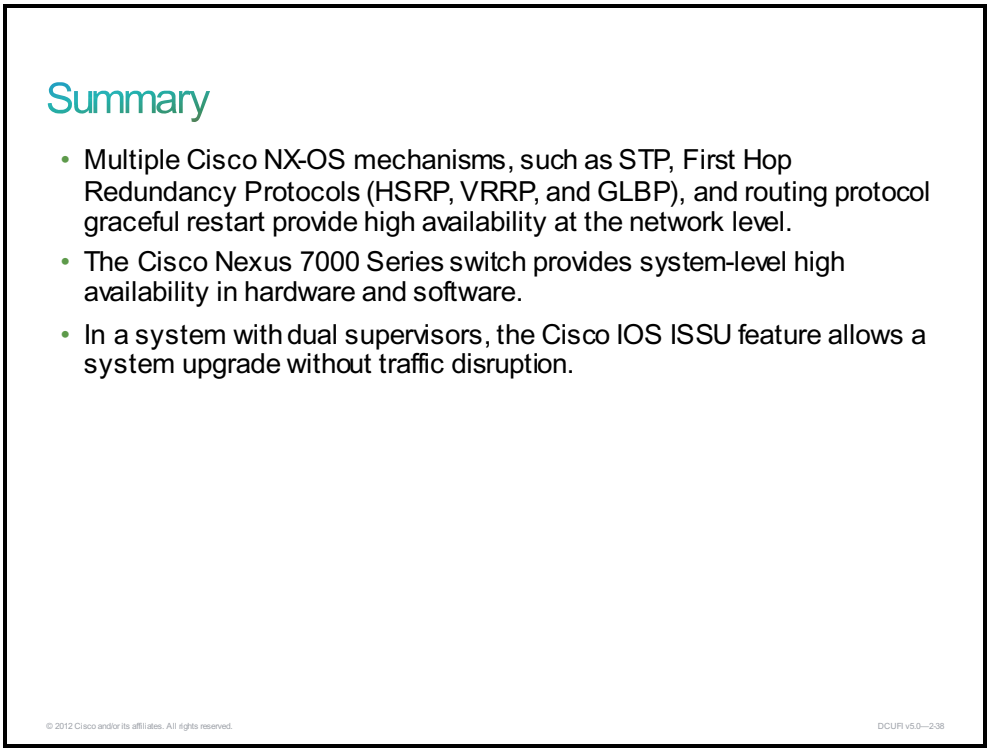


Finally, you can use the **show version** command to verify that the Cisco IOS ISSU has been successful and examining the current kickstart and system versions. In this example, the system is running Cisco NX-OS Release 6.0.1.

https://t.me/learningnets

Summary

This topic summarizes the key points that were discussed in this lesson.



References

For additional information, refer to this resource:

- To learn more about Cisco Nexus 7000 Series NX-OS high availability and redundancy, refer to *Cisco Nexus 7000 Series NX-OS High Availability and Redundancy Guide* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/sw/nx-os/high_availability/configuration/guide/b_Cisco_Nexus_7000_Series_NX-OS_High_Availability_and_Redundancy_Guide.html

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 2

Configuring Virtual Device Contexts

Overview

The Cisco Nexus Operating System (Cisco NX-OS) Software can divide a single physical switch into up to four virtual switches, which are referred to as *virtual device contexts*, or VDCs. Each VDC operates like a standalone switch, meaning that each has a distinct configuration file, a set of physical ports, and separate instances of control plane protocols such as routing and spanning-tree. This feature provides the option to use a single physical switch to serve multiple roles within a data center topology. Data center deployments can leverage this capability to provide service integration, enhanced security, administrative boundaries, or flexibility of hardware deployment as business needs change.

Objectives

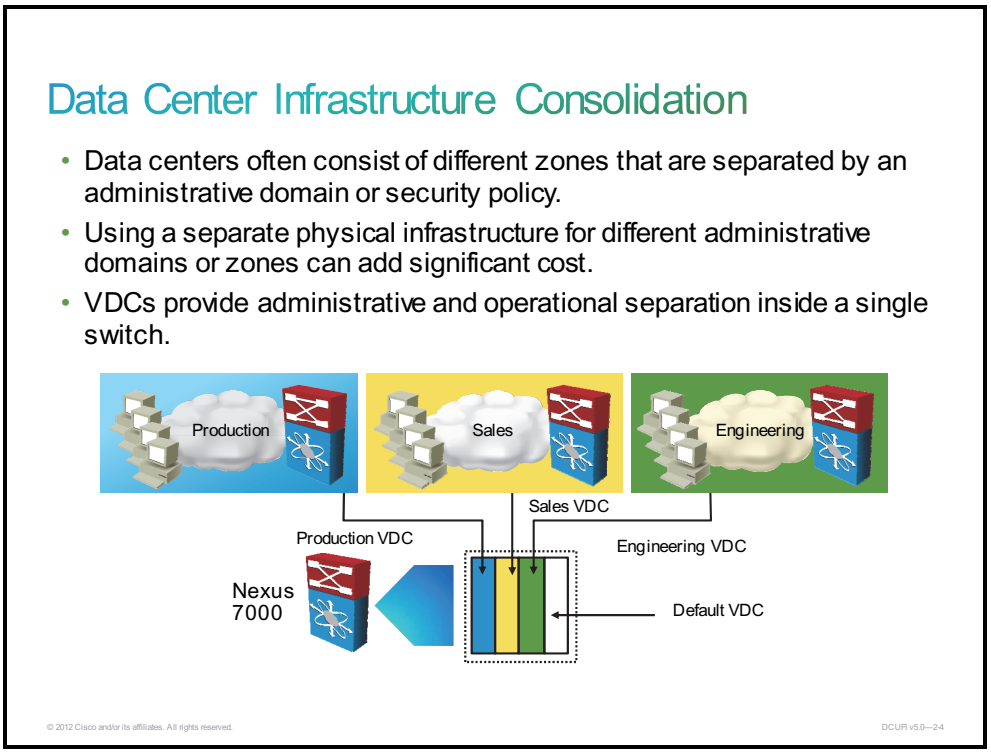
Upon completing this lesson, you will be able to identify how to plan and implement VDCs into the data center solution when given a certain requirement. You will be able to meet these objectives:

- Identify how VDCs could be used to consolidate the physical infrastructure
- Identify the architecture of VDCs, their use of resources on the physical switch, and how Cisco NX-OS supports VDCs
- Explain how to configure VDC resource templates
- Explain major new VDC features in Cisco NX-OS 6.1
- Explain how to configure VDCs on the Cisco Nexus 7000 Series switch
- Explain how to configure the management settings for VDCs
- Explain the concept of shared ports versus dedicated ports and how to configure a storage VDC

https://t.me/learningnets

Using VDCs in Data Centers

This topic identifies how virtual device contexts (VDCs) could be used to consolidate the physical infrastructure.



Data centers are often partitioned into separate domains or zones that are implemented on separate physical infrastructures. The creation of separate physical infrastructures is commonly driven by a need to separate administrative domains for security and policy reasons.

Although VLANs and virtual routing and forwarding (VRF) instances can be used to separate user traffic on the data plane, these technologies do not provide separation of administration and management functions or isolation of fault domains.

Building separate physical infrastructures to separate zones by administrative or security policy can add significant cost to the infrastructure. Depending on the port counts and the functions that are needed in the separate domains, the physical switches in each of the separate domains may be underutilized. However, consolidation of multiple logical switches on a single physical switch can improve hardware utilization. Consolidation can also provide additional flexibility to the data center design.

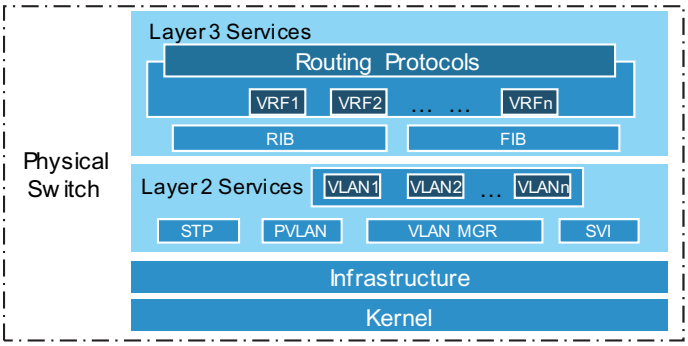
VDCs allow a single physical Cisco Nexus 7000 Series switch to be partitioned into multiple logical switches. This partitioning enables the consolidation of different logical zones onto a single physical infrastructure.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Virtualization Technologies

- VLANs and VRFs provide data plane separation and some degree of control plane separation
- Device contexts also provide management plane separation
- VDCs provide, in addition to the above:
  - Separation of resources and operating environment
  - Isolated fault domains



© 2012 Cisco and/or its affiliates. All rights reserved.

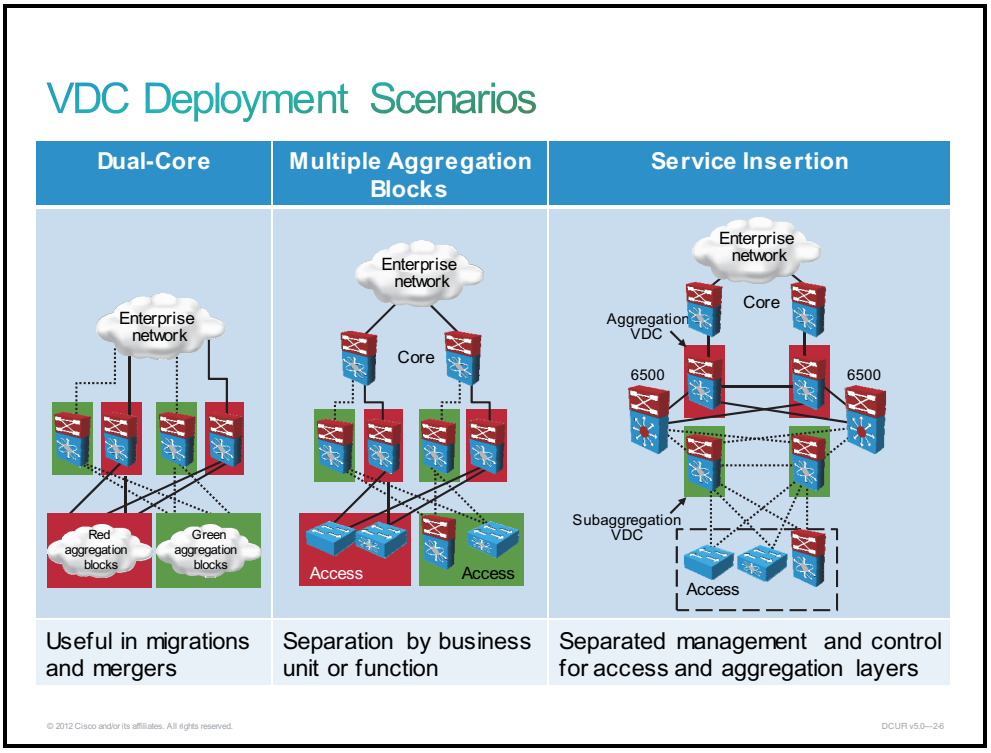
A device can be virtualized in a number of ways, each of which is defined by the level of fault containment and management separation provided. The main elements that are associated with virtualization include the following:

- **Control plane**: The ability to create multiple independent instances of the control plane processes and protocols allows the creation of multiple logical topologies and fault domains.
- **Data (or forwarding) plane**: Forwarding tables and other databases can then be partitioned to provide data segregation.
- **Management plane**: Well-delineated management environments can be provided independently for each virtual device.
- **Software partitioning**: Modular software processes can be grouped into partitions and dedicated to specific virtual devices in order to create well-defined, separated fault domains.
- **Hardware components**: Hardware components are partitioned and dedicated to specific virtual devices in order to allow predictable allocation of hardware resources.

VRFs and VLANs only provide a logical separation of the data plane and a certain amount of control plane functionality. The use of per-VLAN MAC address tables and per-VRF routing and forwarding tables separates the data plane functions within the switches. Control plane functions are only separated to an extent. Per-VLAN Spanning Tree Plus (PVST+), Rapid Per-VLAN Spanning Tree Plus (Rapid PVST+), and Multiple Spanning Tree (MST) all allow separate network topologies to be calculated for different VLANs. However, a single process maintains these topologies. The same principle applies to VRFs: A single routing process may distribute routing information for several VRFs.

In addition to separation of data, control, and management plane functions, the sharing of resources is an important aspect to consider. The VDCs that are used by the Cisco Nexus 7000 Series switches separate data plane, control plane, and management plane functions. VDCs combine these functions with resource management and process separation in order to provide isolated fault domains.

https://t.me/learningnets



Because a VDC has the same functional characteristics as a physical switch, it can be used in many different places in the overall data center network design.

A major advantage of using VDCs instead of separate physical switches is that physical ports can be easily reallocated to the various VDCs. This capability allows for easy changes and additions to the network design as the network grows and evolves. The following are some scenarios that could benefit from the utilization of VDCs. (Because a VDC has characteristics and capabilities that are similar to a separate physical switch, these are not VDC-specific topologies. Rather, they could also be built with separate dedicated switches in the roles that are occupied by VDCs. However, VDCs can provide additional design flexibility and efficiency in these scenarios.)

- **Split-core topology**: VDCs can be used to build two separate redundant data center cores using only a pair of Cisco Nexus 7000 Series switches. This technique can be useful to facilitate migration when the enterprise network needs to expand in order to support mergers and acquisitions. If sufficient ports are available on the existing data center core switches, then two additional VDCs can be created for a separate data center core. This approach allows a second data center network to be built alongside the original one. This second network can be built without any impact on the existing network. Eventually, aggregation blocks could be migrated from one core to the other by reallocating interfaces from one VDC to the other.
- **Multiple aggregation blocks**: At the aggregation layer of the data center network, a single aggregation block consists of a pair of aggregation switches for redundancy and their associated access layer switches. If an enterprise has a business requirement to deploy separate aggregation blocks for different business units or functions, the use of VDCs may accomplish this logical segregation without the need to deploy separate physical switches. Administration and management can be delegated to different groups. Configuration changes in the VDC of one aggregation block cannot affect the VDCs of the other aggregation blocks. For example, a separate production and development aggregation block could be built by using a single pair of aggregation switches.

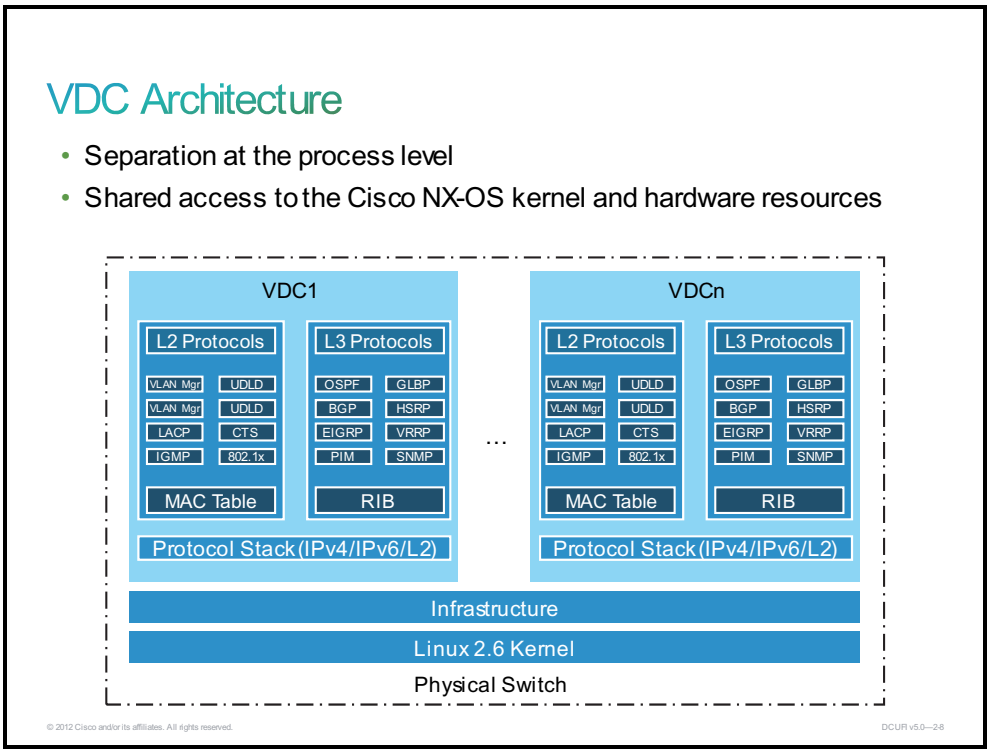
https://t.me/learningnets

- **Service insertion**: In some data center deployments, VRFs are used to create a Layer 3 hop that separates the servers in the access network from the services in the service chain as well as the aggregation layer. This approach creates a “sandwich” consisting of two VRFs, with the services chain in between. Instead of VRFs, two VDCs could be used to create this “services sandwich.” In addition to the control plane and data plane separation that is provided by VRFs, a VDC provides management plane separation and fault isolation. The VDC services sandwich design increases security by logically separating the switches on the inside and the outside of the services chain.

https://t.me/learningnets

Virtual Device Contexts

This topic identifies the architecture of VDCs, their use of resources on the physical switch, and how Cisco NX-OS supports VDCs.



The Cisco Nexus 7000 Series switch extends the VLAN and VRF virtualization concept by using the VDCs that virtualize the device itself. The VDCs split the physical switch into multiple logical devices, each of which is independent of one another.

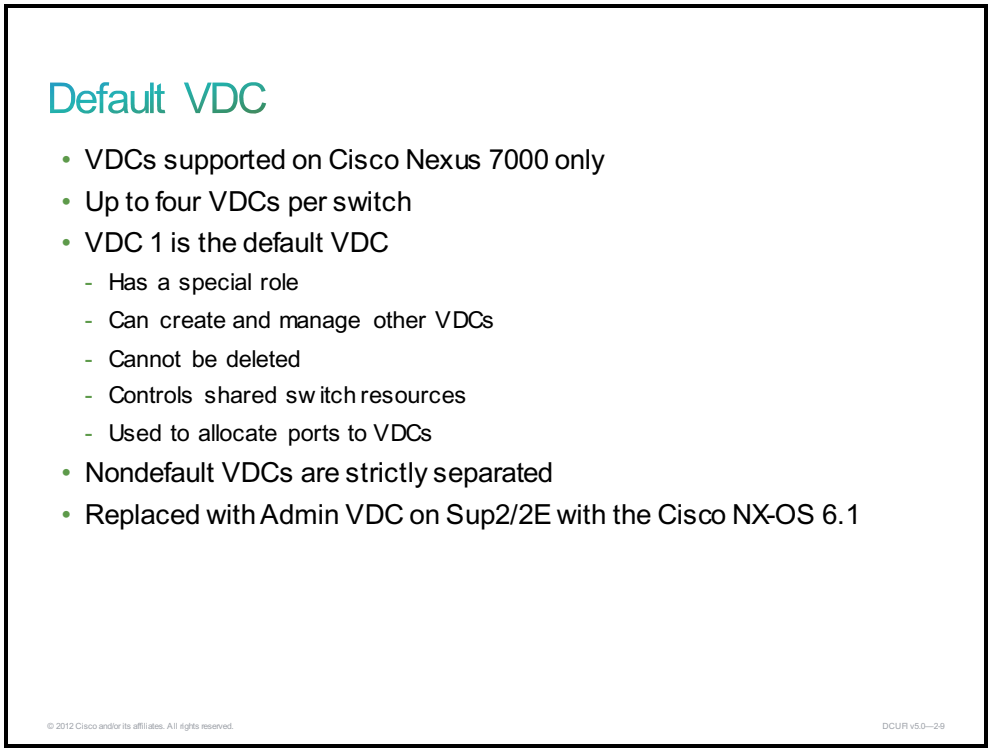
Within each VDC, there is a set of unique and independent VLANs and VRFs, with each VDC having physical ports assigned to it. This design also allows the hardware data plane to be virtualized, along with a separate management domain that can manage the VDC, which allows the management plane to be virtualized as well.

In its default state, the switch control plane runs as a single device context called VDC 1, which will run approximately 80 processes. Some of these processes will have other threads spawned. The result could be as many as 250 processes actively running on the system at any given time. This collection of processes constitutes what is seen as the control and management plane for a single physical device without any other VDCs enabled. The default VDC 1 is always active—always enabled—and can never be deleted. Even if no other VDC is created, support for virtualization through VRFs and VLANs within VDC 1 is available.

The Cisco Nexus 7000 Series switch supports multiple VDCs. The creation of additional VDCs takes these processes and replicates them for each device context that is created. The kernel and infrastructure modules of Cisco NX-OS Software are shared between the processes of the different VDCs, but the processes within each VDC are entirely independent. The hardware resources on the supervisor and I/O modules are also shared between the VDCs.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



The use of VDCs currently allows a single Cisco Nexus 7000 Series switch to be partitioned into up to four logical switches: the default VDC and three additional VDCs. Initially, all hardware resources of the switch belong to the default VDC. When you first configure a Cisco Nexus 7000 Series switch, you are effectively configuring the default VDC: VDC 1. The default VDC has a special role in that it controls all hardware resources and has access to all other VDCs. VDCs are always created from the default VDC. Hardware resources, such as interfaces and memory, are also allocated to the other VDCs from the default VDC. The default VDC can access and manage all other VDCs, while the additional VDCs only have access to the resources that are allocated to them and cannot access any other VDCs.

VDCs are truly separate virtual switches. They do not share any processes or data structures, and traffic can never be forwarded from one VDC to another VDC inside the chassis. Any traffic that needs to be passed between two VDCs in the same chassis will first have to leave the originating VDC through one of the ports that are allocated to it. The traffic will then enter the destination VDC through one of the ports that are allocated to that VDC. VDCs are separated on the data plane, control plane, and management plane. The only exception is the default VDC, which can interact with the other VDCs on the management plane. Control and data plane functions of the default VDC are still separated from the other VDCs.

The default VDC has several other unique and critical roles in the function of the switch:

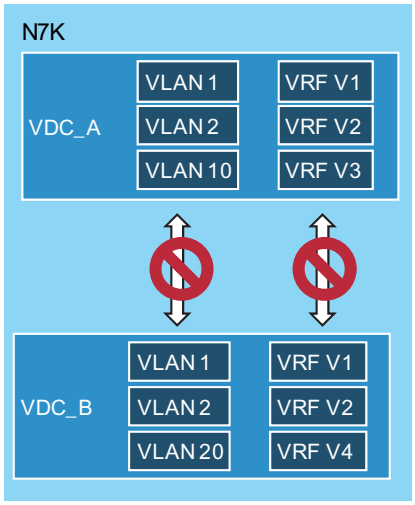
- Systemwide parameters such as Control Plane Policing (CoPP), VDC resource allocation, and Network Time Protocol (NTP) may be configured from the default VDC.
- Licensing of the switch for software features is controlled from the default VDC.
- Software installation must be performed from the default VDC. All VDCs run the same version of the software.
- Reloads of the entire switch may only be issued from the default VDC. Nondefault VDCs may be reloaded independently of other VDCs.

https://t.me/learningnets

If it is anticipated that a switch may be used in a multi-VDC configuration, it is recommended to reserve the default VDC for administrative functions and to configure production network connections in nondefault VDCs. This approach will provide flexibility and higher security. Administrative access into the nondefault VDCs to perform configuration functions may easily be granted without exposing access for reloading the entire switch or change software versions. No Layer 3 interfaces in the default VDC need to be exposed to the production data network. Only the management interface needs to be accessible through an out-of-band (OOB) management path. Unused interfaces may be retained in a shutdown state in the default VDC as a holding area until they are needed in the configuration of one of the nondefault VDCs. In this way, the default VDC may be maintained in an administrative context, thus requiring console access or separate security credentials. Following this guideline effectively allows a single Cisco Nexus 7000 Series switch to perform the functional roles of up to three production switches.

VLAN and VRF Scope

- Within each VDC, VLANs and VRFs can be used to provide additional levels of virtualization.
- VLANs and VRFs in different VDCs are strictly isolated.
- VLAN numbers and VRF names can be reused within different VDCs.
  - No connectivity between VLANs with same ID.
- External connections are necessary to forward traffic between VDCs.



© 2012 Cisco and/or its affiliates. All rights reserved.

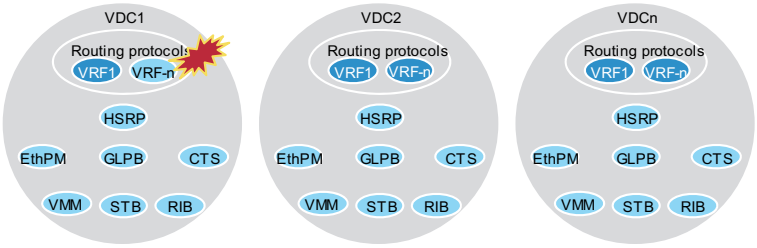
The use of VDCs does not preclude the use of VLANs and VRFs. Within each VDC, you can create VLANs and VRFs. The VLANs and VRFs inside a VDC are entirely independent of the VLANs and VRFs in any other VDC.

Because VDCs are independent, VLAN numbers and VRF names can be reused in different VDCs. However, VLANs and VRFs in one VDC are completely isolated from VLANs and VRFs in other VDCs. There is no internal connection between the VLANs and VRFs in different VDCs. To connect a VLAN or VRF in one VDC to a VLAN or VRF in a different VDC, an external connection is required. In this way, VDCs behave as completely separate logical switches.

https://t.me/learningnets

VDC Fault Domains

- Each VDC runs separate processes for control plane and management plane functions, thereby creating a separate fault domain.
- When a process crashes in a VDC, the processes in the other VDCs are not affected and continue to run unimpeded.



© 2012 Cisco and/or its affiliates. All rights reserved.

When multiple VDCs are created in a physical switch, the architecture of the VDC feature provides a means to prevent failures occurring within any one VDC from affecting other VDCs. For example, if a spanning-tree recalculation is started in one VDC, it does not affect the spanning-tree domains of other VDCs in the same physical chassis because it is an entirely independent process. The same applies to other processes, such as the Open Shortest Path First (OSPF) process in that network topology changes in one VDC do not affect other VDCs on the same switch.

Because Cisco NX-OS Software uses separate processes in each VDC, fault isolation is even extended to potential software process crashes. If a process crashes in one VDC, then that crash is isolated from other VDCs. The Cisco NX-OS high-availability features, such as stateful process restart, can be applied independently to the processes in each of the VDCs. Process isolation within a VDC is important for fault isolation and serves as a major benefit for organizations that implement the VDC concept.

In addition, fault isolation is enhanced with the ability to provide per-VDC debug commands and per-VDC logging of messages from syslog. These features provide administrators with the ability to locate problems within their own VDCs.

https://t.me/learningnets

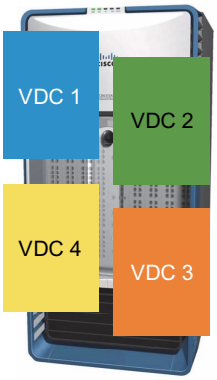
Resource Allocation

This topic explains how to configure VDC resource templates.

Resource Types

- VDCs share the Cisco NX-OS kernel and infrastructure resources
- Resources can be classified in three groups:

1. Global resources:
   - Allocated to all VDCs together
   - Examples: boot image configuration, the switch name, NTP servers, CoPP configuration, and in-band span sessions
2. Shared resources:
   - Resources that are shared between VDCs
   - Example: OOB Ethernet management port
3. Dedicated resources:
   - Resources that are allocated to a particular VDC
   - Examples: physical switch ports, VLAN/VRF limits



© 2012 Cisco and/or its affiliates. All rights reserved.

VDCs are isolated at the process level, but they share the Cisco NX-OS kernel and the hardware resources of the Cisco Nexus 7000 Series switch. As a result, some resources are completely dedicated to a single VDC. Other resources are shared between VDCs, and some resources can only be managed from the default VDC. There are three types of VDC resources:

- **Global resources**: Resources that can only be allocated, set, used, or configured globally for all VDCs. These include boot image configuration, the switch name, NTP servers, CoPP configuration, and Switched Port Analyzer (SPAN) sessions.
- **Dedicated resources**: Resources that are allocated to a particular VDC, such as physical switch ports.
- **Shared resources**: Resources that are shared between VDCs, such as the OOB Ethernet management port.

An example of a global resource is the boot string that specifies the version of software that should be used upon booting up the device. It is not possible to run different versions of Cisco NX-OS Software in different VDCs.

An example of a shared resource on the switch is the OOB Ethernet management interface on the supervisor. If multiple VDCs are configured and accessible from the management interface, then they must share it. The OOB management interface cannot be allocated to a VDC as other regular switch ports can be. Because the management interface does not support IEEE 802.1Q, the management interfaces of the VDCs should be configured with IP addresses on the same IP subnet for the management VRF.

https://t.me/learningnets

Resource Templates

- Resource templates limit impact of a VDC on supervisor resources
- Some constraints set indirectly by VLAN and VRF limits
- Supervisor CPU and memory shared between the VDCs

Resource	Configurable Range	Default Value	
		Default VDC	Nondefault VDC
Module type	M1, F1	-	M1
IPv4 unicast route memory	1–256 MB	min=max=96 MB	min=max=8 MB
IPv4 multicast route memory	1–90 MB	min=max=58 MB	min=max=8 MB
IPv6 unicast route memory	1–100 MB	min=max=24 MB	min=max=4 MB
IPv6 multicast route memory	1–20 MB	min=max=8 MB	min=max=2 MB
Port channels	0–768	min=0, max=768	min=0, max=768
SPAN sessions	0–2	min=0, max=2	min=0, max=2
ERSPAN sessions	0–24	min=0, max=24	min=0, max=24
VLANs	16 to 4094	min=16, max=4094	min=16, max=4094
VRFs	2 to 1000	min=16, max=1000	min=16, max=1000

© 2012 Cisco and/or its affiliates. All rights reserved.

Access to the CPU and memory on the supervisor is shared by the processes that are running in the different VDCs. VDC resource templates can be used to limit the impact of a VDC on the CPU and memory consumption on the supervisor. VDC resource templates can also control access to other limited resources, such as SPAN sessions or port channels.

VDC resource templates set the minimum and maximum limits for shared physical device resources when you create the VDC. Cisco NX-OS Software reserves the minimum limit for the resource to the VDC. Any resources that are allocated to the VDC beyond the minimum are based on the maximum limit and availability on the device.

You can explicitly specify a VDC resource template, or you can use the default VDC template that is provided by the Cisco NX-OS Software. VDC templates set limits on the following resources:

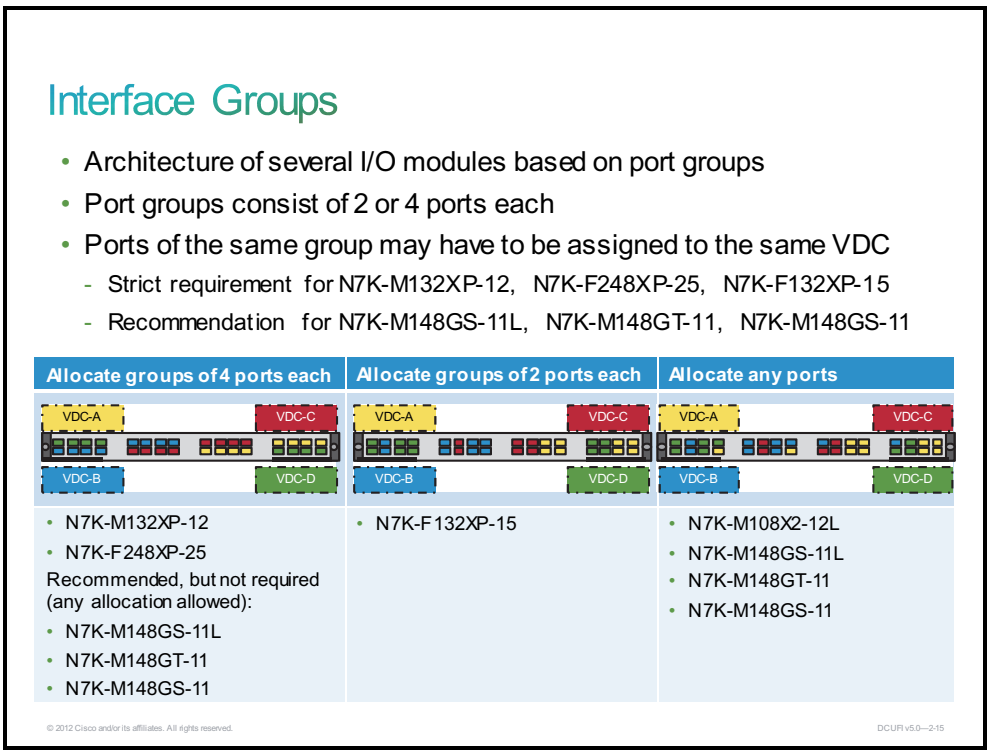
- IPv4 unicast route memory
- IPv4 multicast route memory
- IPv6 unicast route memory
- IPv6 multicast route memory
- Number of port channels
- Number of SPAN sessions
- Number of Encapsulated Remote Switched Port Analyzer (ERSPAN) sessions
- Number of VLANs
- Number of VRFs

Furthermore, you can restrict a VDC to the use of a specific card type. This restriction is configurable per each VDC and not on the VDC resource templates. By default, both the M1 and F1 types of line cards are supported in a VDC. You can restrict a VDC in these ways:

- Restriction to M1 modules only (the default setting)
- Restriction to F1 modules only
- No restriction

https://t.me/learningnets

The consumption of supervisor CPU resources is not directly limited through the resource template. However, by limiting the number of VLANs and VRFs that can be assigned to a VDC, you can indirectly influence the amount of control plane overhead that is generated by that VDC.



Physical ports are allocated to different VDCs from the default VDC. Logical interfaces, such as switch virtual interfaces (SVIs), subinterfaces, or tunnel interfaces cannot be assigned to a VDC. Logical interfaces are always created in the VDC to which they belong. Once a physical port is assigned to a VDC, all subsequent configuration of that port is performed within that specific VDC. Within a VDC, both physical and logical interfaces can be assigned to VLANs or VRFs.

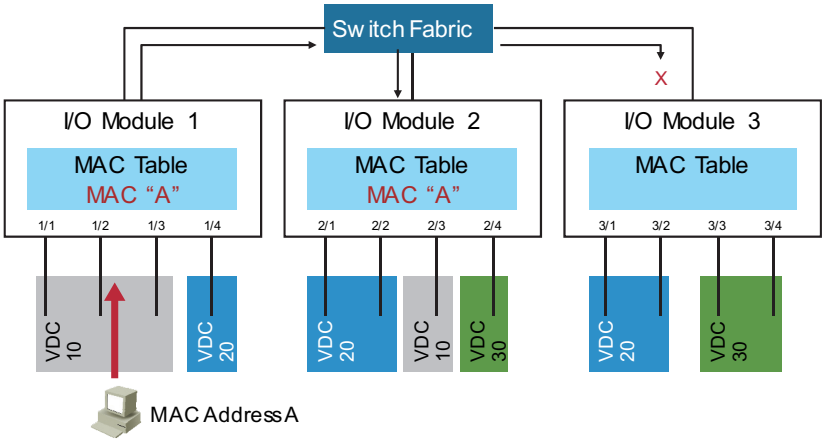
On many I/O modules, any port can be individually assigned to a VDC. The exceptions to this rule include modules whose architecture uses port groups. The ports within the same group share some common hardware elements. This architecture may necessitate the allocation of all ports within a group to the same VDC. Such modules include the following:

- N7K-M108X2-12L (1 interface * 8 port groups = 8 interfaces). There are no restrictions on the interface allocation between VDCs.
- N7K-M148GS-11L, N7K-M148GT-11, and N7K-M148GS-11 (12 interfaces * 4 port groups = 48 interfaces). There are no restrictions on the interface allocation between VDCs, but it is recommended that interfaces that belong to the same port group be in a single VDC.
- N7K-M132XP-12 (4 interfaces * 8 port groups = 32 interfaces). Interfaces belonging to the same port group must belong to the same VDC.
- N7K-F132XP-15 module, with 16 port groups that consist of two ports each (2 interfaces * 16 port groups = 32 interfaces). Interfaces belonging to the same port group must belong to the same VDC.
- N7K-F248XP-25 module, with 12 port groups that consist of four ports each (4 interfaces * 12 port groups = 48 interfaces). Interfaces belonging to the same port group must belong to the same VDC.

https://t.me/learningnets

Resource Utilization: MAC Address Tables

- Forwarding engines on the I/O modules only contain MAC address entries for VDCs that have a port on the I/O module
- Table distribution between multiple I/O modules



© 2012 Cisco and/or its affiliates. All rights reserved.

The allocation of ports on an I/O module to the various VDCs directly affects the utilization of the hardware resources on the forwarding engine on the I/O module.

The forwarding engine on each I/O module is responsible for Layer 2 address learning and maintains a local copy of the Layer 2 Forwarding table. The MAC address table on each I/O module supports 128,000 MAC addresses. When an I/O module learns a new MAC address, a copy is forwarded to other I/O modules. This enables the Layer 2 address-learning process to be synchronized across all I/O modules. Layer 2 learning is a VDC local process and has a direct effect on the addresses that are placed on an I/O module.

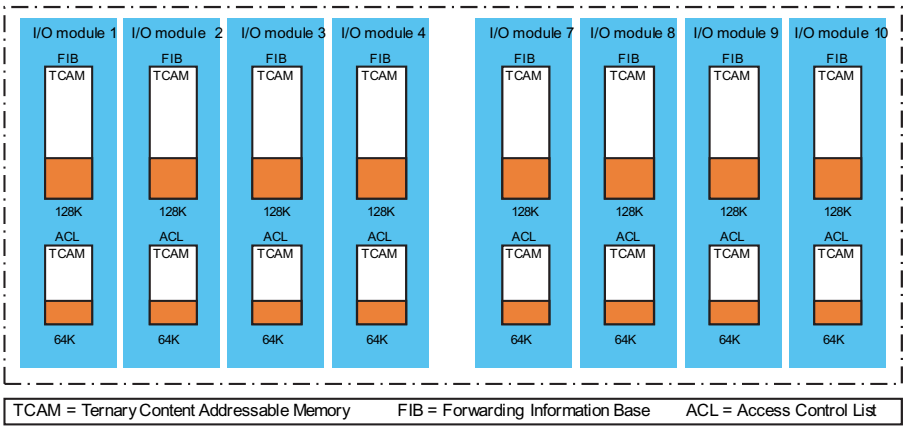
The following example is illustrated in the figure:

1. On I/O module 1, MAC address A is learned from port 1/2.
2. The address is installed in the local Layer 2 Forwarding table of I/O module 1.
3. The MAC address is then forwarded to I/O modules 2 and 3.
4. I/O module 3 has no ports that belong to VDC 10, so it does not install any MAC addresses that are learned from that VDC.
5. I/O module 2 does have a local port in VDC 10, so it installs MAC address A into its local forwarding tables. As a result, the forwarding engine on an I/O module only contains the MAC address entries for VDCs that have a port that is allocated on that I/O module.

https://t.me/learningnets

Layer 3 Resource Utilization Without VDCs

- Without VDCs, the TCAMs on all I/O modules contain the same set of routes and access list entries
- Non-XL I/O modules support 128,000 IPv4 FIB entries, 64,000 IPv6 FIB entries, and 64,000 ACL entries



© 2012 Cisco and/or its affiliates. All rights reserved.

The forwarding engine on each I/O module supports 128,000 entries in the IPv4 forwarding information base (FIB) and 64,000 entries in the IPv6 FIB. Additionally, the I/O modules are capable of 64,000 access control entries (ACEs) and 512,000 ingress and 512,000 egress NetFlow entries.

Note These numbers apply to non-XL I/O modules or XL I/O modules that are used without an XL license. XL I/O modules with an appropriate license installed support higher numbers of entries.

When the default VDC is the only active VDC, all the learned routes and access lists are loaded into the ternary content addressable memory (TCAM) tables of each I/O module. This information means that the I/O module has all the necessary information to make a correct forwarding decision, as can be seen in the figure. The routes for the default (red) VDC are present in the FIB and access control list (ACL) TCAMs on all I/O modules.

https://t.me/learningnets

Layer 3 Resource Utilization with VDCs

- Proper allocation of VDCs to I/O modules can improve hardware resource utilization.
- Avoid combining VDCs with high numbers of routes or access list entries on a single I/O module
- Example:
  - VDC 10 and VDC 30 should not share a I/O module
  - Combine VDC 10 and 20, or VDC 20 and 30 on a I/O module

VDC Number	Number of Routes	Number of ACEs	Allocated I/O Module
10	100K	50K	I/O module 1 and 2
20	10K	10K	I/O module 1, 2, 3, 7
30	90K	40K	I/O module 3 and 7

ACE = Access Control Entry

© 2012 Cisco and/or its affiliates. All rights reserved.

When physical port resources are split between VDCs, only the I/O modules that are associated with a specific VDC are required to store the forwarding information and associated ACLs for that VDC. This allocation method allows the resources to be scaled beyond the default system limits.

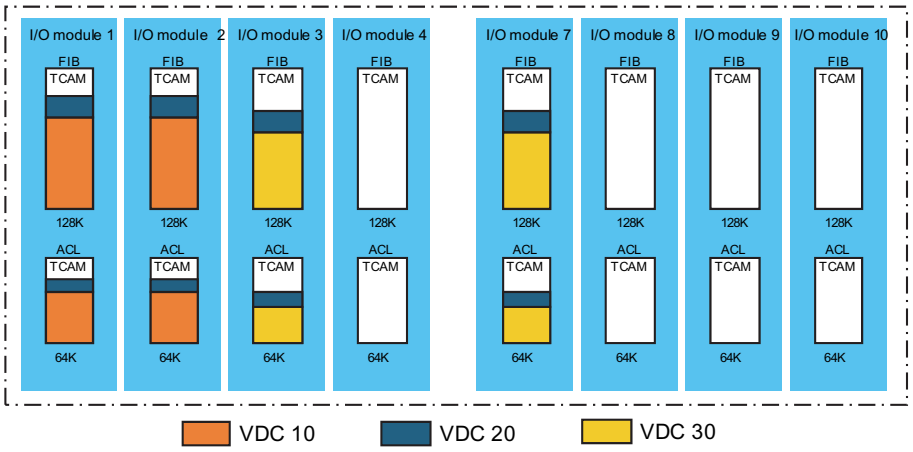
The figure shows a resource allocation example.

Each of the individual VDCs stays within the 128,000 IPv4 routes and 64,000 ACE limits. The total number of routes and ACEs combined exceeds the limits of the forwarding engines on non-XL I/O modules. Therefore, a single I/O module should never be shared among all three VDCs.

https://t.me/learningnets

Layer 3 Resource Utilization Example

- None of the TCAMs are overloaded
- Total number of routes and ACL entries exceeds the I/O module limits



© 2012 Cisco and/or its affiliates. All rights reserved.

The effect of allocating a subset of ports to a given VDC results in the FIB and ACL TCAMs for the respective I/O modules to be programmed with the forwarding information and ACLs for that VDC. Using the previous figure, a total of 180,000 IPv4 forwarding entries have been installed in a switch that, without VDCs, would have a system limit of 128,000 forwarding entries. Likewise, a total of 100,000 ACEs have been installed, whereas a single VDC would only allow 64,000 ACEs.

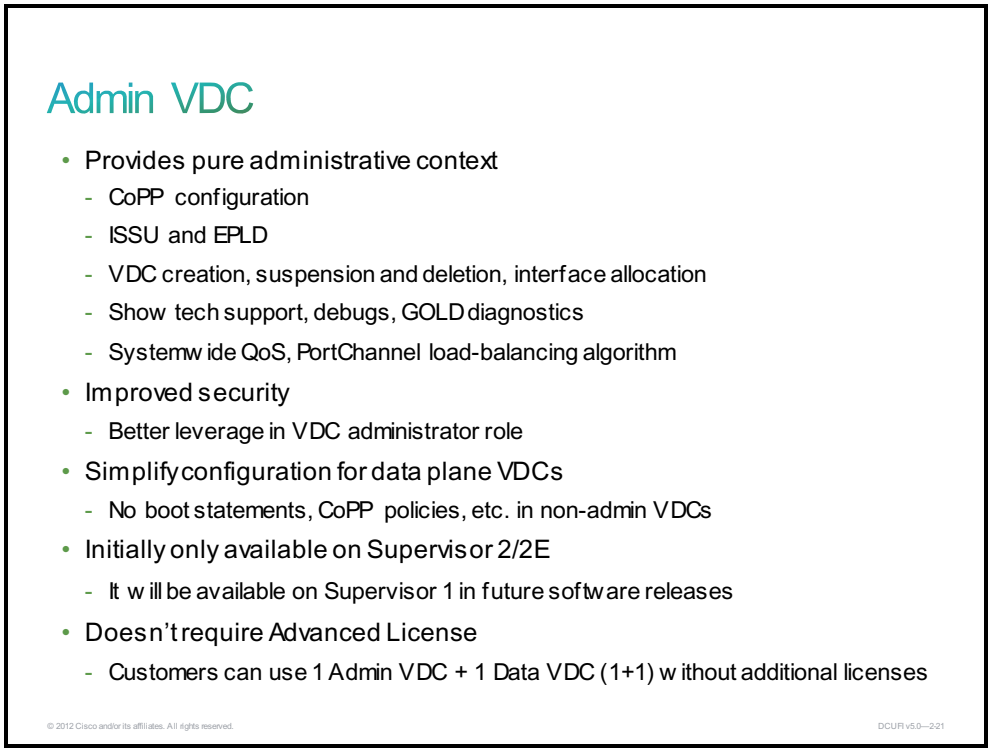
More importantly, the FIB and ACL TCAM space on I/O modules 4, 6, 7, and 8 is free for use by additional VDCs that might be created. This allows resources to be extended beyond the system limits.

As with the TCAMs for the FIB and ACLs, the use of the NetFlow TCAM is also more granular when multiple VDCs are active. When a flow is identified, a flow record is created on the local NetFlow TCAM resident on that particular I/O module. Both ingress and egress NetFlow are performed on the ingress I/O module, so it is the NetFlow TCAM of the ingress I/O module where the flow is stored. The collection and export of flows is always performed on a per-VDC basis. However, no flow in VDC 10 is exported to a collector that is part of VDC 20. After the flow is created in a NetFlow TCAM on I/O module 2, it is not to be replicated to NetFlow TCAMs on other I/O modules that are part of the same VDC. This design optimizes the use of the TCAM.

https://t.me/learningnets

New VDC Features in Cisco NX-OS 6.1

This topic explains major new VDC features in Cisco NX-OS 6.1.



With introduction of the new supervisor module Sup2/2E and the new Cisco NX-OS version 6.1, two major features are introduced to the VDC feature:

- **Admin VDC**: Instead of default VDC, Admin VDC is now used for system and VDC administration only. No user traffic exists on Admin VDC. On Sup2 4+1, VDCs are available, and on Sup2E 8+1, VDCs are available (with an additional license).
  In the Admin VDC, you can manage the system aspects of the switch—such as control plane protection—perform software and EPLD upgrades, manage VDCs, perform diagnostics, and so on.

  The Admin VDC greatly improves security since no production traffic forwarding is done in the VDC, so theoretically, no one can breach the security protocols and try to manage the switch in an unauthorized fashion.

Note	Initially, the Admin VDC feature is available only for the second-generation supervisor engines (Supervisor Engine 2/2E). Support for the Supervisor Engine 1 will be added in future Cisco NX-OS versions.

- **CPU shares**: Provides means of prioritization of VDC by allocating more CPU shares during congestion.

The Admin VDC itself does not require the Advanced Services Package license (LAN_ADVANCED_SERVICES_PKG). There is one administration and one data VDC available.

https://t.me/learningnets

Admin VDC (Cont.)

- All CoPP configuration is performed in Admin VDC
  - Hardware rate limiters in Admin VDC
- Module control is performed in Admin VDC
  - Poweroff and out-of-service
- License management is performed in Admin VDC
- **Cannot** perform the following in Admin VDC mode
  - Enable L2/L3 features, including routing protocols
  - Limited feature support

```
N7K-1(config)# feature ?
          ldap          Enable/Disable ldap
          ntp           Enable/Disable NTP
          password      Credential(s) for the user(s)/device(s)
          privilege     Enable/Disable IOS type privilege level support
          scheduler     Enable/Disable scheduler
          scp-server    Enable/Disable SCP server
          sftp-server   Enable/Disable SFTP server
          ssh           Enable/Disable ssh
          tacacs+       Enable/Disable tacacs+
          telnet        Enable/Disable telnet
```

© 2012 Cisco and/or its affiliates. All rights reserved.

Only a limited set of features is available in Admin VDC. You can configure administrative functions, such as role-based access control (RBAC), Control Plane Policing, management access, and so on. Modules can be powered on and off or put out of service for maintenance purposes only from Admin VDC.

Licenses are installed and activated in the Admin VDC.

Creation and deleting of VDC is done in Admin VDC.

The Admin VDC cannot perform any traffic forwarding. Layer 2 and Layer 3 functionality is disabled. Only administrative Cisco NX-OS features are available.

https://t.me/learningnets

CPU Shares

- Enables per-VDC CPU access and prioritization
  - Provides more control and protection per VDC for users
  - Network administrator controls each VDC priority
- CPU share is controlled by VDC priority
- CPU is shared equally among VDCs
- User can control allocation—priorities are linear in effect
  - The more VDCs configured, the lower the overall percentage per VDC
- Comes into use when CPU utilization increases
- Controlled by Cisco NX-OS scheduler in the kernel
- Available on SUP2/2E only

© 2012 Cisco and/or its affiliates. All rights reserved.

In VDC configuration in Cisco NX-OS 6.1, it is possible to define CPU priority. According to that priority, CPU cycles are allocated to processes in different VDCs.

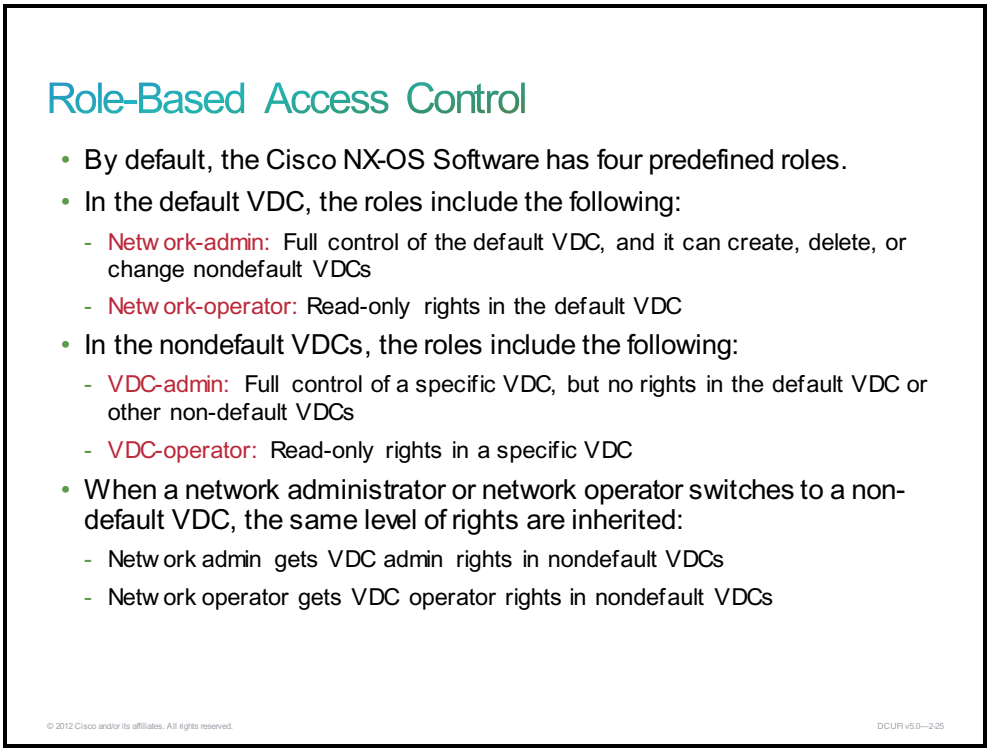
CPU shares are not used to limit the available CPU to VPC, but rather to allocate CPU cycles to VDC with higher priority during time of congestion. The allocation of the CPU shares is controlled by the kernel.

Note The CPU shares feature is available on Sup2/2E only.

https://t.me/learningnets

Configuring VDCs

This topic explains how to configure VDCs on the Cisco Nexus 7000 Series switch.



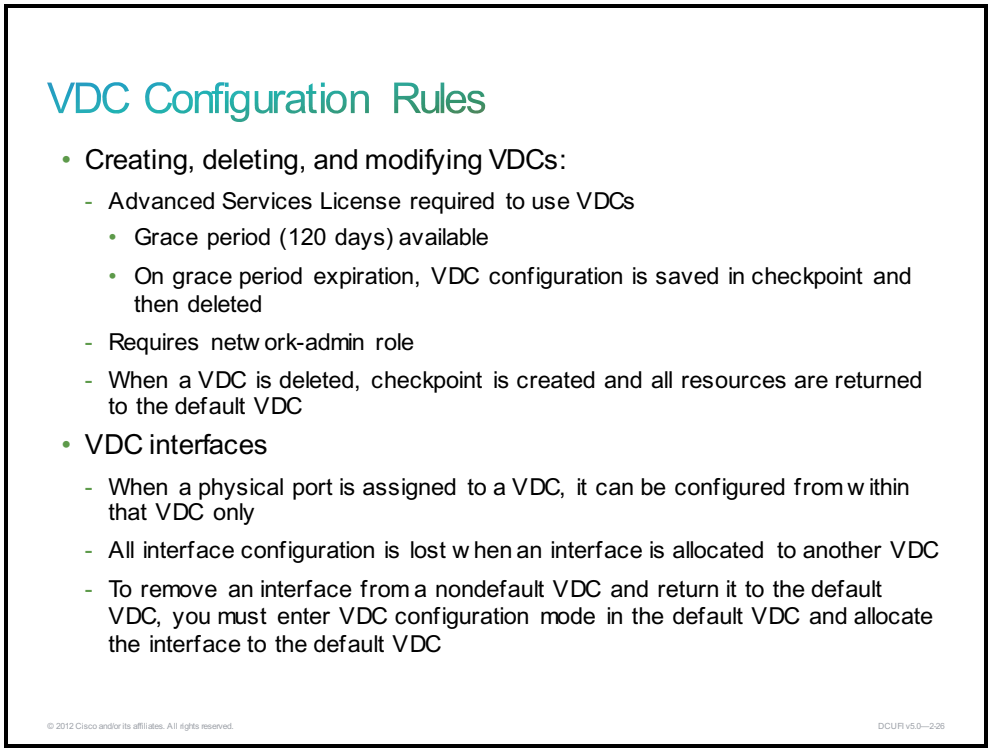
Cisco NX-OS Software uses RBAC to control the access rights of users. By default, the Cisco Nexus 7000 Series switches recognize four roles:

- **Network-admin**: The first user account that is created on a Cisco Nexus 7000 Series switch in the default VDC is the user “admin.” This user is automatically assigned the network-admin role. The network-admin role gives a user complete control over the default VDC of the switch. This role includes the ability to create, delete, or change nondefault VDCs.
- **Network-operator**: The second default role that exists on Cisco Nexus 7000 Series switches is the network-operator role. This role allows the user read-only rights in the default VDC. The network-operator role includes the right to issue the **switchto** command, which can be used to access a nondefault VDC from the default VDC. By default, there are no users that are assigned to this role. The role has to be specifically assigned to a user by a user who has network-admin rights.
- **vdc-admin**: When a new VDC is created, the first user account on that VDC is the user “admin.” This process is similar to the way that the admin user for a physical switch is created. The admin user on a nondefault VDC automatically gets the vdc-admin role assigned to it. This role gives a user complete control over the specific nondefault VDC. However, this user does not have any rights in any of the other VDCs and cannot access them through the **switchto** command.
- **vdc-operator**: The vdc-operator role has read-only rights for a specific VDC. This user has no rights in any of the other VDCs..

When a user who has the network-admin or network-operator role accesses a nondefault VDC by using the **switchto** command, that user will be mapped to a role of the same level within that VDC. A user with the network-admin role gets the vdc-admin role in the nondefault VDCs. A user with the network-operator role gets the vdc-operator role in the nondefault VDCs.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



To use VDCs, the Advanced Services Package needs to be installed on a Cisco Nexus 7000 Series switch. You can try out the feature during a 120-day grace period. However, when the grace period expires, any nondefault VDCs will be removed from the switch configuration. VDC configuration will be preserved in a system-created checkpoint. Any existing processes for those VDCs will be terminated.

VDCs can only be created, deleted, or changed from the default VDC. It is not possible to configure VDCs from a nondefault VDC. To configure VDCs, a user needs to have network-admin rights in the default VDC.

Physical interfaces and other resources are always assigned to nondefault VDCs from the default VDC. Once a physical interface has been assigned to a specific VDC, the configuration for that interface is performed from that nondefault VDC. It is not possible to configure an interface from any other VDC than the VDC that it is allocated to.

Initially, all physical resources are assigned to the default VDC. When interfaces are reallocated to a different VDC, any existing configuration on the interface is removed.

When a VDC is removed, Cisco NX-OS creates a checkpoint, and all resources that are associated with that VDC are returned to the default. All processes that belong to the VDC are terminated, and forwarding information for the VDC is removed from the forwarding engines.

It is not possible to move interfaces from a nondefault VDC to the default VDC from within the nondefault VDC itself. To remove a physical interface from a nondefault VDC, you must enter configuration mode in the default VDC and reallocate the interface to the default VDC.

Note	When you configure different VDCs from the default VDC, it is very important to verify that you are configuring the correct VDC. Accidentally making changes to the wrong VDC can have serious consequences.

https://t.me/learningnets

Configuring VDCs

- Nondefault VDCs are created from within the default VDC global configuration context:

```
N7K-1(config)# vdc RED
N7K-1(config-vdc)#

N7K-1# show vdc
vdc_id          vdc_name        state         mac
------          --------        -----         ----------
1               N7K-1           active        00:1b:21:09:3f:18
2               RED             active        00:1b:21:09:3f:19
```

- Nondefault VDCs are deleted from within the default VDC global configuration context:

```
N7K-1(config)# no vdc RED
Deleting this vdc will remove its config. Continue deleting this
vdc? [no] yes
Note: VDC deletion is a time consuming process, please wait until
the command completes
```

© 2012 Cisco and/or its affiliates. All rights reserved.

The example in the figure shows how to create a VDC named "RED" by using the **vdc** command. To remove an active and current VDC, use the **no** form of this command.

Allocating Interfaces

- Allocate a single Ethernet interface to a VDC:

```
N7K-1(config)# vdc RED
N7K-1(config-vdc)# allocate interface ethernet 2/1
Moving ports will cause all config associated to them in source
vdc to be removed. Are you sure you want to move the ports? [yes] yes
```

- Allocate a range of Ethernet interfaces to a VDC:

```
N7K-1(config)# vdc RED
N7K-1(config-vdc)#  allocate  interface ethernet 2/1 - 8
Moving ports will cause all config associated to them in source
vdc to be removed. Are you sure you want to move the ports? [yes] yes
```

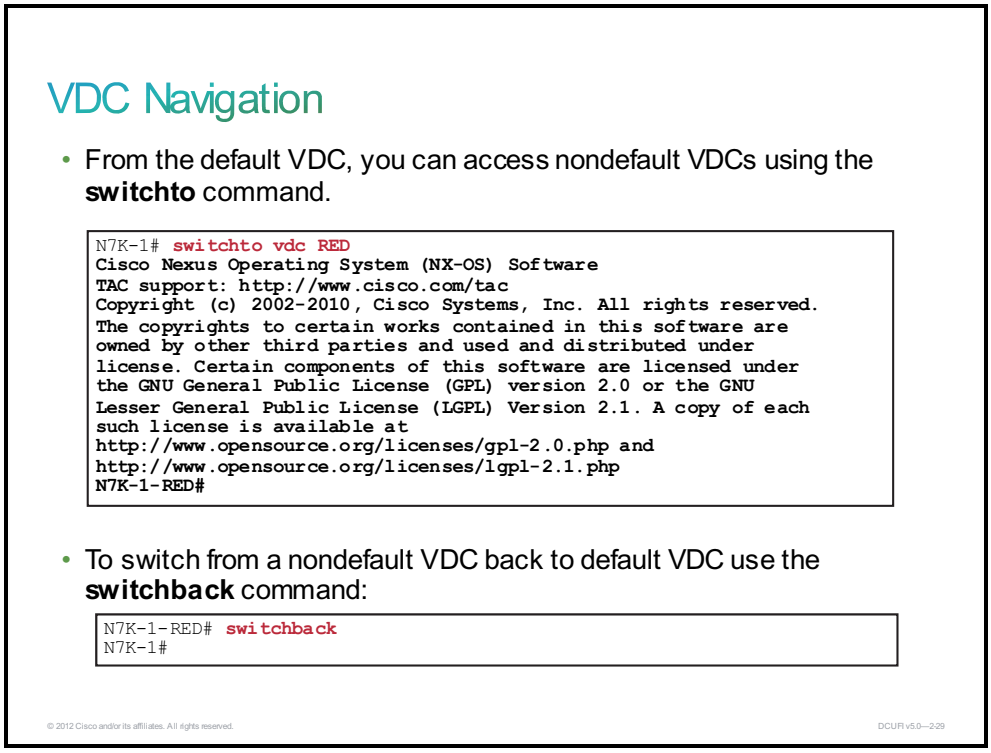
- Allocate multiple interfaces to a VDC:

```
N7K-1(config)# vdc RED
N7K-1(config-vdc)# allocate interface ethernet 2/1, ethernet 2/3,
ethernet 2/5, ethernet 2/7
Moving ports will cause all config associated to them in source
vdc to be removed. Are you sure you want to move the ports? [yes] yes
```

© 2012 Cisco and/or its affiliates. All rights reserved.

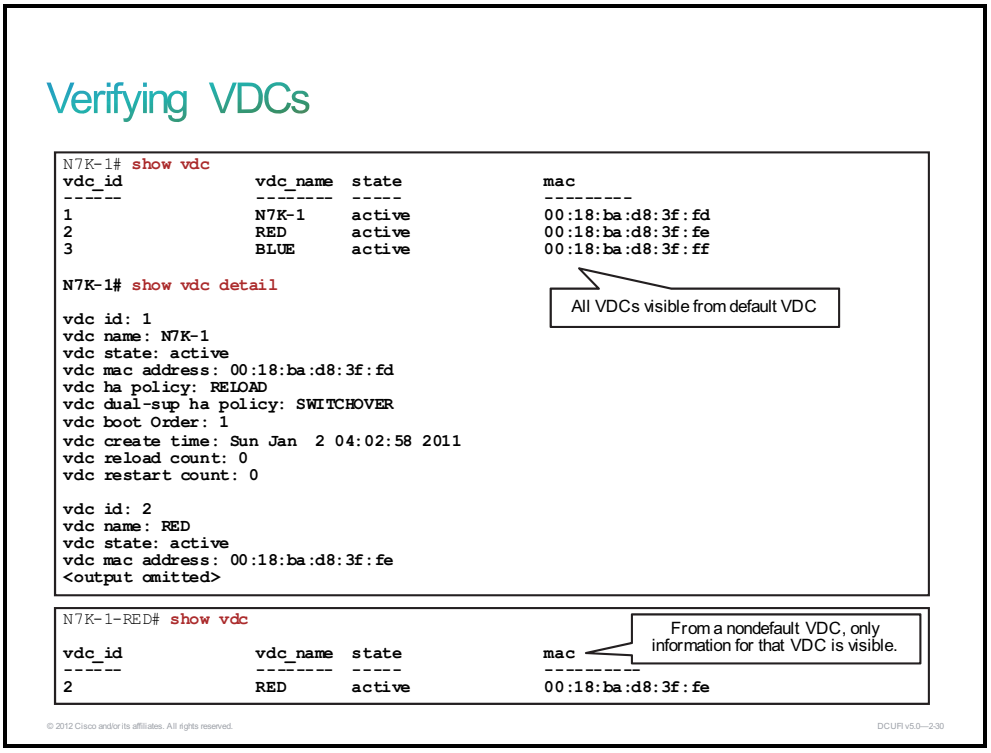
Interfaces are assigned to a VDC from the default VDC by using the **allocate-interface** command in VDC configuration mode for that specific VDC. Multiple interfaces can be assigned with a single command by specifying an interface range. When assigning multiple interfaces with a single command, make sure you are assigning a whole interface port group or else the assignment will fail.

https://t.me/learningnets



It is possible to navigate between the default and nondefault VDCs by using the **switchto vdc** command. This command will change the context from the default to the specified nondefault VDC. This command cannot be used to navigate directly between nondefault VDCs. To navigate from one nondefault VDC to another, the **switchback** command must first be issued to return to the default VDC. That command can then be followed by a **switchto** command in order to enter the configuration context for the desired nondefault VDC. This command is necessary to perform the initial setup of the VDCs. Once user accounts and IP connectivity have been properly configured, the VDC can be accessed over the network by using Secure Shell (SSH) or Telnet.

https://t.me/learningnets



The scope of the **show vdc** commands depends on the VDC in which it is executed. When these commands are executed in a nondefault VDC, the information that is displayed is restricted to that VDC only. If these commands are executed in the default VDC, they display information on all VDCs unless a specific VDC is entered as a command option.

As can be seen in the figure, issuing the **show vdc** command from within the default VDC lists all the active and current VDCs. The default VDC has visibility over all nondefault VDCs.

Issuing the **show vdc** command within a nondefault VDC only provides information about that particular VDC. Nondefault VDCs have no visibility to each other or to the default VDC.

The **show vdc detail** command provides more detailed information about the VDCs, including the name, state, and MAC address.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Verifying Interface Allocation

```
N7K-1# show vdc membership

vdc_id: 1 vdc_name: N7K-1 interfaces:
     Ethernet2/1  Ethernet2/2  Ethernet2/3  Ethernet2/4  Ethernet2/5
     Ethernet2/6  Ethernet2/7  Ethernet2/8  Ethernet2/9  Ethernet2/10
     Ethernet2/11 Ethernet2/12 Ethernet2/13 Ethernet2/14 Ethernet2/15
     Ethernet2/16 Ethernet2/17 Ethernet2/18 Ethernet2/19 Ethernet2/20
     Ethernet2/21 Ethernet2/22 Ethernet2/23 Ethernet2/24 Ethernet2/25
     Ethernet2/26 Ethernet2/27 Ethernet2/28 Ethernet2/29 Ethernet2/30
     Ethernet2/31 Ethernet2/32 Ethernet2/33 Ethernet2/34 Ethernet2/35
     Ethernet2/36 Ethernet2/37 Ethernet2/38 Ethernet2/39 Ethernet2/40
     Ethernet2/41 Ethernet2/42 Ethernet2/43 Ethernet2/44 Ethernet2/45
     Ethernet2/48

vdc_id: 2 vdc_name: RED interfaces:
     Ethernet2/47

vdc_id: 3 vdc_name: BLUE interfaces:
     Ethernet2/46
```

VDC interface allocation viewed from the default VDC

© 2012 Cisco and/or its affiliates. All rights reserved.

The **show vdc membership** command can be used to display the interfaces that are allocated to the VDCs.

Verifying Resource Restrictions

- When a VDC is created, default resource limits are imposed
- Resource limits are based on a default VDC template

```
N7K# show vdc resource
port-channel            0 used       0 unused      192 free      192 total
monitor-session         0 used       0 unused        2 free        2 total
vlan                   14 used      34 unused    16370 free    16384 total
u4route-mem            48 used       0 unused      208 free      256 total
vrf                     6 used      42 unused     8186 free     8192 total

N7K# show vdc resource detail

  port-channel           0 used       0 unused      192 free      192 total
  --------------
      Vdc         Min       Max      Used    Unused     Avail
     -----       -----     -----    ------  --------   -------
   switch          0       192         0         0       192
  Payroll          0       192         0         0       192
    MyVDC          0       192         0         0       192

  monitor-session        0 used       0 unused        2 free        2 total
  -----------------
      Vdc         Min       Max      Used    Unused     Avail
     -----       -----     -----    ------  --------   -------
   switch          0         2         0         0         2
  Payroll          0         2         0         0         2
    MyVDC          0         2         0         0         2
<output omitted>
```

© 2012 Cisco and/or its affiliates. All rights reserved.

When a VDC is first created, resources are allocated to it based on a default resource template. The settings that are applied by this template can be adjusted afterward to meet the specific requirements for that particular VDC. You can verify the current resource allocation by using the **show vdc resource** and **show vdc resource detail** commands.

https://t.me/learningnets

Modifying Resource Limits

- Resource limits can be applied directly to VDCs
- Module type (F1, M1) can be set only in this way

```
N7K-1(config)# vdc RED
N7K-1(config-vdc)# limit-resource vlan minimum 32 maximum 4094
N7K-1# show running-config vdc | begin RED
vdc RED id 2
  allocate interface Ethernet1/1,Ethernet1/3,Ethernet1/5,Ethernet1/7,
Ethernet1/17,Ethernet1/19,Ethernet1/21,Ethernet1/23
  boot-order 1
  limit-resource vlan minimum 32 maximum 4094
  limit-resource monitor-session minimum 0 maximum 2
  limit-resource vrf minimum 2 maximum 1000
  limit-resource port-channel minimum 0 maximum 768
  limit-resource u4route-mem minimum 8 maximum 8
  limit-resource u6route-mem minimum 4 maximum 4
  limit-resource m4route-mem minimum 8 maximum 8
  limit-resource m6route-mem minimum 2 maximum 2
```

Resource limit applied directly to VDC

© 2012 Cisco and/or its affiliates. All rights reserved.

The **limit-resource** command can be used to change the minimum and maximum resource limits for shared physical device resources for the VDC.

Setting the minimum number will reserve the specified resources for the VDC. When more resources are needed, they can be assigned to the VDC from a shared pool. Setting the maximum number for resources sets an upper limit to the amount of resources that can be assigned to the VDC from the shared pool.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Modifying VDC Resource Limits Using Templates

- Scalable method
- Templates reusable for multiple VDCs

```
N7K-1(config)# vdc resource template PRODUCTION
N7K-1(config-vdc-template)# limit-resource vlan minimum 32 maximum 256
N7K-1(config-vdc-template)# limit-resource vrf minimum 32 maximum 64
N7K-1(config-vdc-template)# exit
N7K-1(config)# vdc RED
N7K-1(config-vdc)# template PRODUCTION
N7K-1(config-vdc)# show vdc resource template PRODUCTION

   PRODUCTION
  -------------
     Resource               Min        Max
    ----------             -----      -----
     vlan                    32        256
     vrf                     32         64
<output omitted…>
```

Template configuration

Template assigned to VDC

© 2012 Cisco and/or its affiliates. All rights reserved.

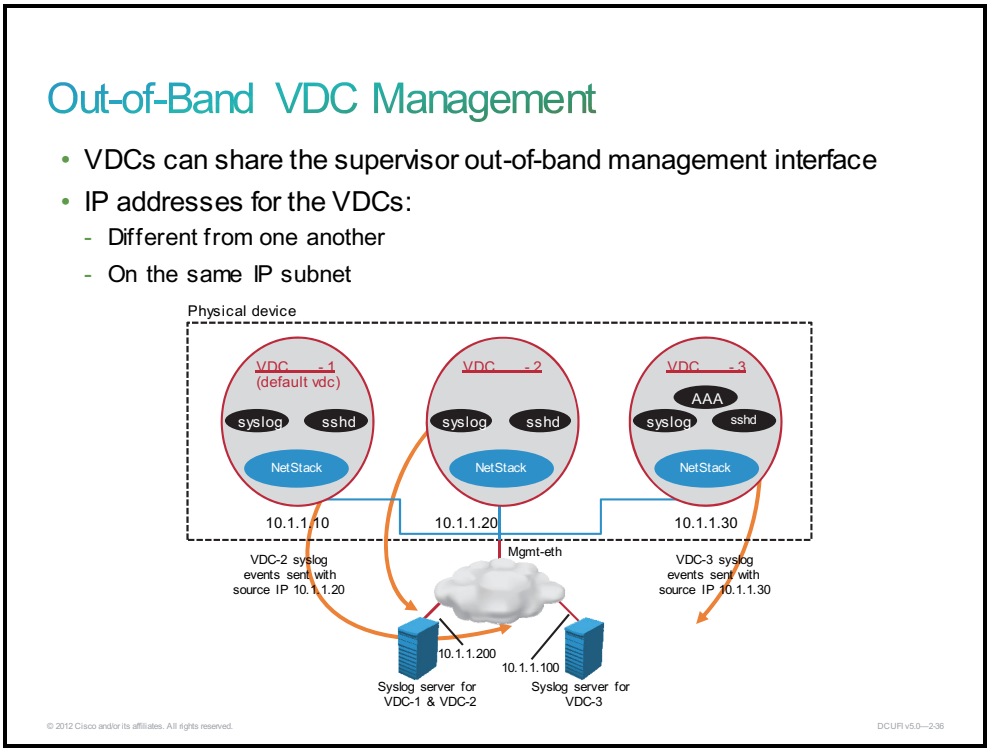
To optimize the process of assigning resources to VDCs, you can create resource templates. A resource template can then be applied to a VDC in order to change the resource allocations for that VDC to match the values in the template.

A VDC resource template is not a “live” template, meaning changes that are made to a VDC resource template do not affect any VDCs that were created by using that VDC resource template. To update a VDC with the new limits from the changed VDC resource template, you must explicitly reapply the template to the VDC.

https://t.me/learningnets

Management Settings

This topic explains how to configure the management settings for VDCs.



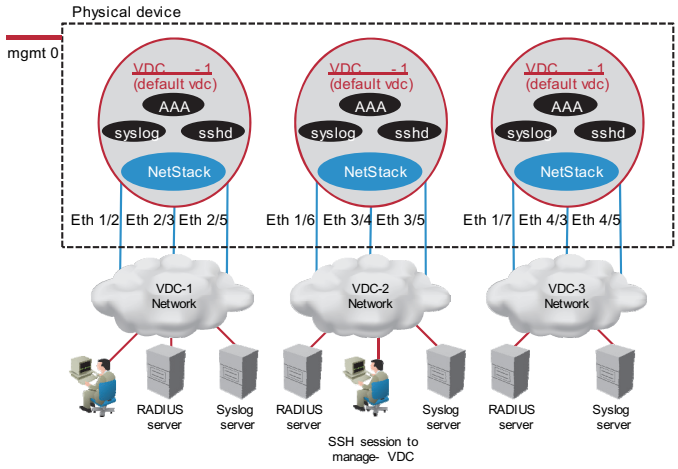
The OOB Ethernet management interface on the active supervisor of the Cisco Nexus 7000 Series switch is shared among the various VDCs. Cisco NX-OS Software provides a virtual management interface—mgmt 0—for OOB management for each VDC. You can configure this interface with a separate IP address that is accessed through the physical mgmt 0 interface on the supervisor. The virtual management interface allows you to connect to a single management network, which can share the authentication, authorization, and accounting (AAA) servers, as well as the syslog servers, among the VDCs.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

In-Band VDC Management

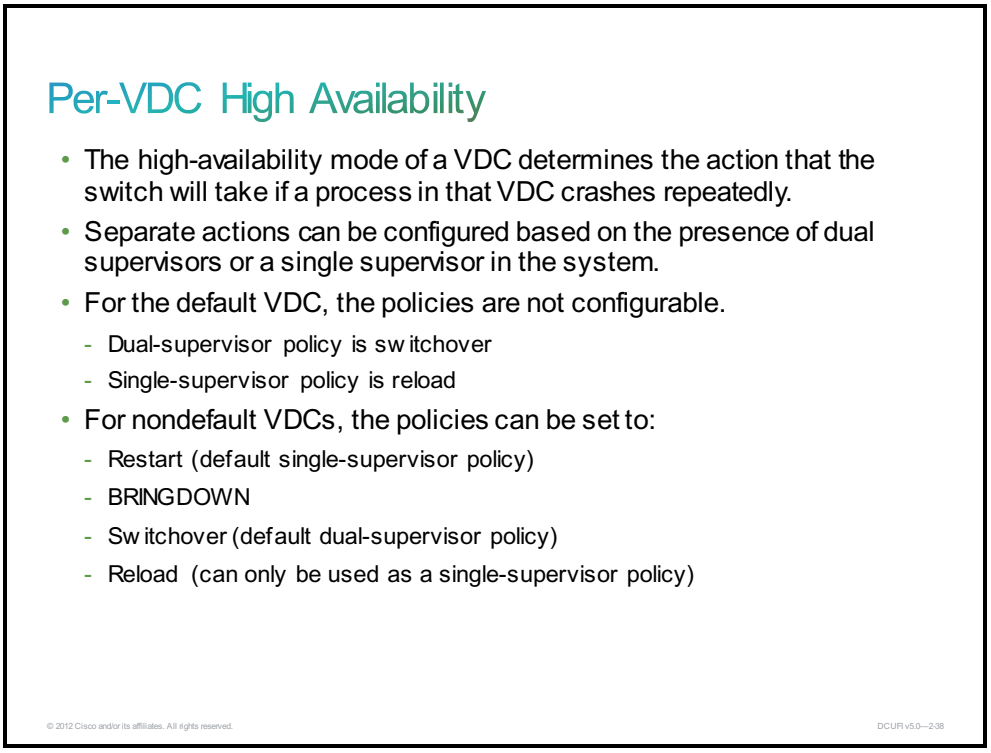
- In-band management access for each separate VDC
- Independent paths of management traffic



© 2012 Cisco and/or its affiliates. All rights reserved.

VDCs also support in-band management. You can access the VDC by using one of the Ethernet interfaces that are allocated to the VDC. The in-band management option allows you to implement strictly separated management networks. This method provides separation of the AAA, syslog, and other network management services for each of the VDCs.

https://t.me/learningnets



The high-availability policies for a VDC define the action that the Cisco NX-OS Software takes when an unrecoverable VDC fault occurs.

You can specify the high-availability policies for single-supervisor module and dual-supervisor module configurations when you create the VDC. The high-availability policy options are as follows:

- Single supervisor module configuration
  - **Bringdown**: Puts the VDC in the failed state. To recover from the failed state, you must reload the VDC or the physical device.
  - **Reload**: Reloads the supervisor module.
  - **Restart**: Deletes the VDC and recreates it by using the startup configuration.
- Dual supervisor module configuration
  - **Bringdown**: Puts the VDC in the failed state. To recover from the failed state, you must reload the VDC or the physical device.
  - **Restart**: Deletes the VDC and recreates it by using the startup configuration.
  - **Switchover**: Initiates a supervisor module switchover.

The default high-availability policy for a nondefault VDC that you create is restart for single-supervisor mode and switchover for dual-supervisor mode. The default high-availability policy for the default VDC is reloaded for a single-supervisor module configuration and switchover for a dual-supervisor module configuration. The policies for the default VDC cannot be changed.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Configuring Per-VDC High Availability

```
N7K-1(config)# vdc RED
N7K-1(config-vdc)# ha-policy dual-sup restart single-sup restart

N7K-1(config)# vdc BLUE
N7K-1(config-vdc)# ha-policy dual-sup bringdown single-sup bringdown

N7K-1# show vdc detail
<output omitted>
vdc name: RED
vdc state: active
vdc mac address: 00:18:ba:d8:3f:fe
vdc ha policy: RESTART
vdc dual-sup ha policy: RESTART
<…output omitted…>
vdc name: BLUE
vdc state: active
vdc mac address: 00:18:ba:d8:3f:ff
vdc ha policy: BRINGDOWN
vdc dual-sup ha policy: BRINGDOWN
```

Both high-availability options (single-sup and dual-sup) are configured in one command line

High-availability method displayed in the detailed verification

© 2012 Cisco and/or its affiliates. All rights reserved.

The example in the figure shows how to configure different high-availability policies for different nondefault VDCs by using the **ha-policy** command.

The **show vdc detail** command can be used to verify the configured high-availability policies for the VDCs.

VDC Running and Startup Configurations

1. Saves the running configuration of all VDCs to NVRAM

```
N7K-1# copy running-config startup-config vdc-all
```

2. Displays running configuration of the default VDC

```
N7K-1# show running-config vdc
```

3. Displays running configuration of all VDCs

```
N7K-1# show running-config vdc-all
```

4. Displays startup configuration of all VDCs

```
N7K-1# show startup-config vdc-all
```

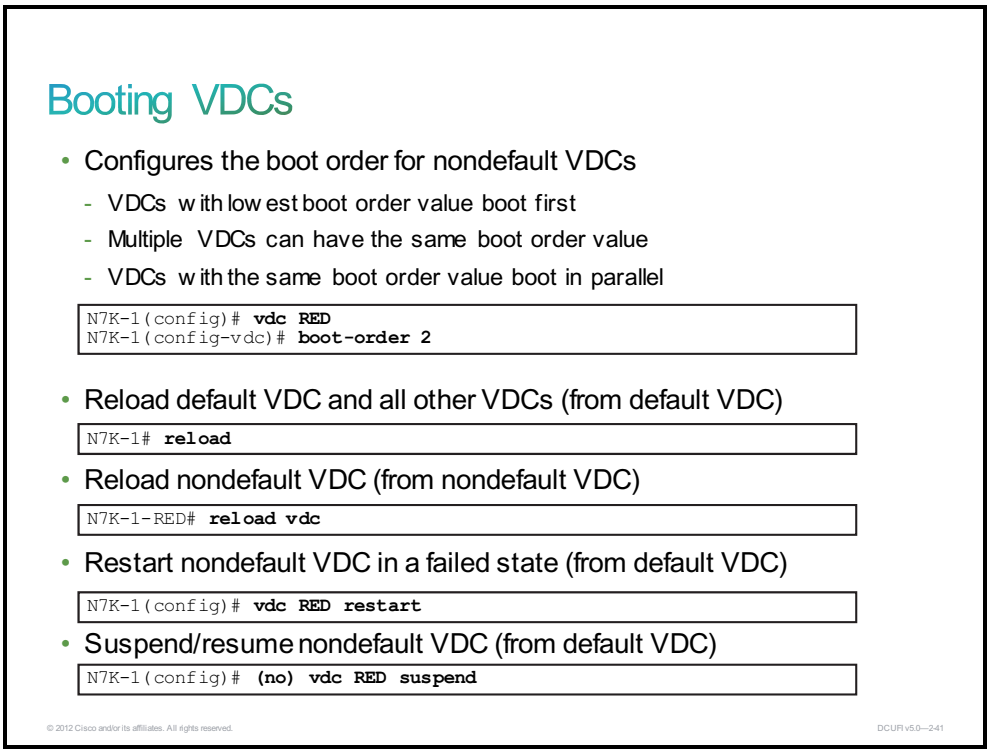
These commands are issued from the default VDC

© 2012 Cisco and/or its affiliates. All rights reserved.

It is possible to save the running configuration for all VDCs by using the **copy running-config startup-config vdc-all** command. The **show running-config vdc** command displays the current configuration file for the default VDC. The **show running-config vdc-all** command displays the current configuration files for all VDCs. Similarly, the **show startup-config vdc-all** command displays the startup configuration files for all VDCs.

https://t.me/learningnets

All four commands must be issued from the default VDC because it has visibility for all VDCs. It is not possible to view the configurations of other VDCs from a nondefault VDC.



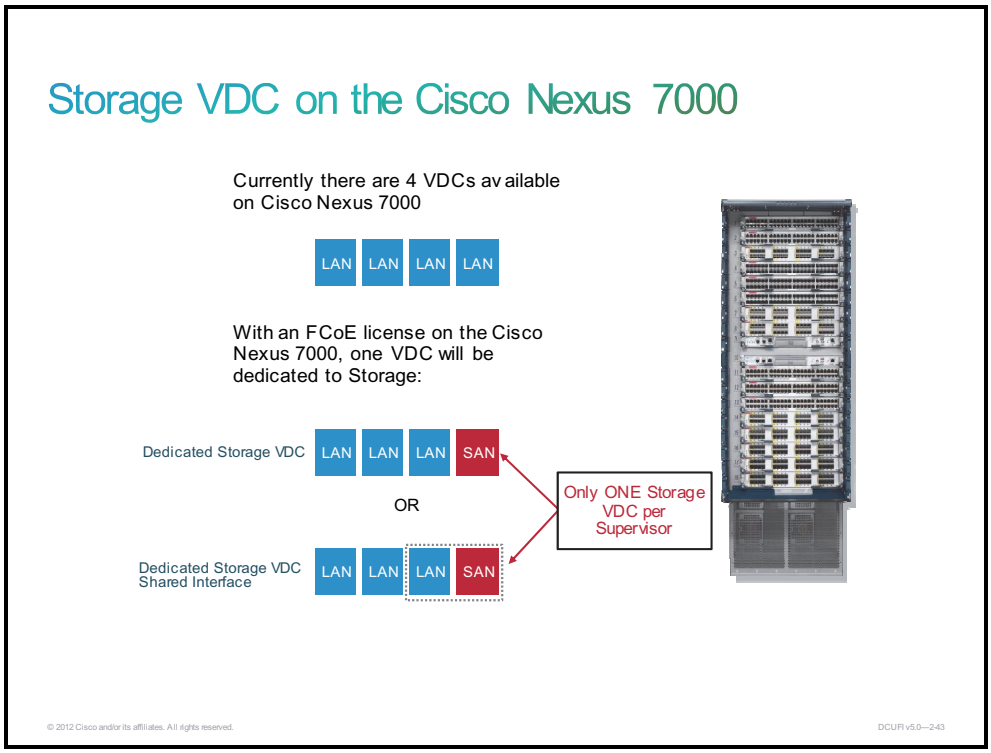
You can manage the VDC using these additional methods:

- **Boot order**: All VDCs will start booting at the same time, but there is no guarantee which one will start actual operations first since boot time may be different for different VDCs. Using the boot order value, the Cisco NX-OS Software starts the VDCs in a predictable sequence. The boot order feature has the following characteristics:
  - More than one VDC can have the same boot order value. By default, all VDCs have the boot-order value of 1.
  - The VDCs with the lowest boot order value boot first.
  - The Cisco NX-OS Software starts all VDCs with the same boot order value, followed by the VDCs with the next boot order value.
  - The Cisco NX-OS Software starts VDCs that have the same boot order value in parallel.
  - You cannot change the boot order for the default VDC; you can change the boot order only for nondefault VDCs.
- **Reload a default VDC**: Use the **reload** command to reload the default VDC. Reloading the default VDC reloads all VDCs on the Cisco NX-OS device.
- **Reload a nondefault VDC**: You can reload an active nondefault VDC that is in any state by using the **reload vdc** command from the nondefault VDC. The impact of reloading a nondefault VDC is similar to reloading a physical device. The VDC reloads using the startup configuration. Reloading a VDC disrupts all traffic on the VDC.
- **Restart a nondefault VDC**: To restart a VDC that is in the failed state due to a high-availability failure, use the **vdc restart** command from the default VDC.
- **Suspend or resume a nondefault VDC, or both**: To suspend VDC operation, use the **vdc suspend** command from the default VDC. To resume the VDC operation, use the **no** form of this command.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Storage VDCs

This topic explains the concept of shared ports versus dedicated ports and how to configure a storage VDC.



Cisco Nexus 7000 Series switch supports Fibre Channel over Ethernet (FCoE) in the Cisco NX-OS Release 5.2(1) and later, and it uses a special VDC type, called a *storage VDC*, to provide the FCoE functionality. Only one storage VDC can exist on the system, and it must be a nondefault VDC. The storage should be dedicated to providing FCoE connectivity and should not fulfill unrelated tasks.

https://t.me/learningnets

Storage VDCs

- Fiber Channel over Ethernet (FCoE) supported on the Cisco Nexus 7000 Series devices
  - Cisco NX-OS release 5.2(1) and later
- Storage VDC
  - Required to run FCoE
  - Cannot be the default VDC
  - Maximum of one storage VDC on the device
- Shared interfaces
  - Shared interfaces carry both Ethernet and Fibre Channel traffic
  - A shared interface allocated to both single Ethernet and a storage VDC
- FCoE supported on both Nexus 7000 F1 and F2 Series modules

© 2012 Cisco and/or its affiliates. All rights reserved.

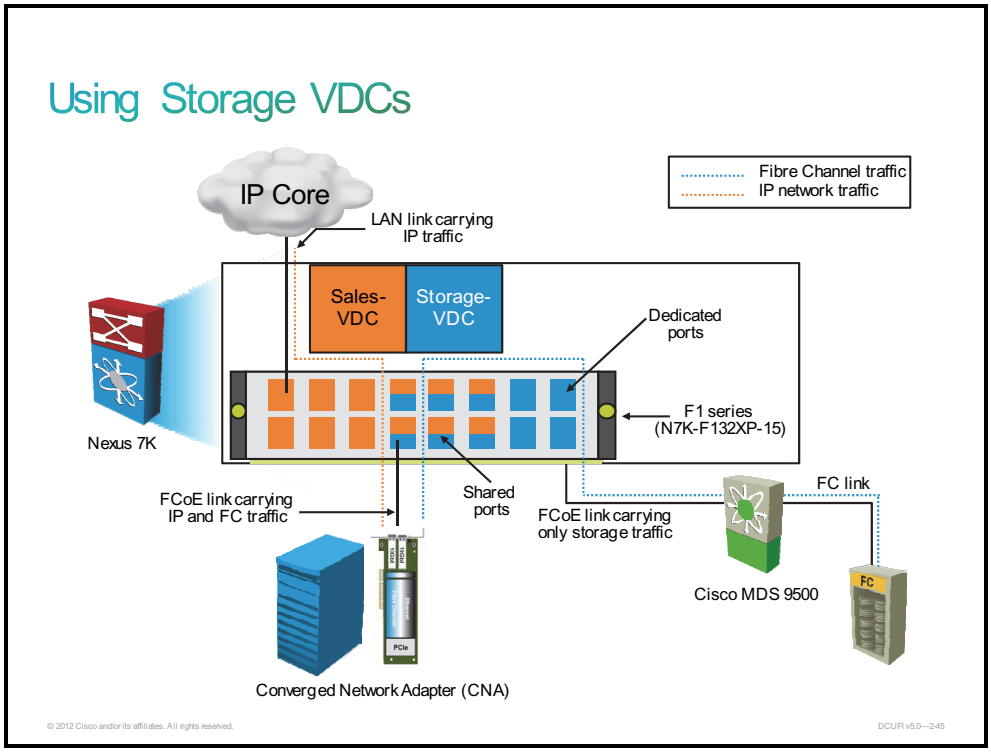
FCoE allows Fibre Channel traffic to be encapsulated over a physical Ethernet link. FCoE and FCoE Initiation Protocol (FIP) frames use a unique EtherType so that FCoE traffic and standard Ethernet traffic can be carried on the same link.

Storage VDC uses two types of interfaces—interfaces that are dedicated to it and interfaces that are shared by the storage VDC and one other VDC. The shared interfaces carry Ethernet and Fibre Channel traffic. They can be connected to hosts equipped with converged network adapters (CNAs). Traffic that is exchanged on shared ports is tagged, and the tagging indicates the type of traffic and allows the switch to direct it to the appropriate VDC, storage, or Ethernet.

Currently, FCoE is supported on Cisco Nexus 7000 F1-Series I/O modules and requires an FCoE Services Package license to be installed for each module.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



FCoE interfaces can be connected to various devices, most notably to servers equipped with CNAs and Cisco MDS 9500 switches. The figure illustrates a Cisco Nexus 7000 Series switch that is equipped with an F1 Series module, such as N7K-F132XP-15, and two nondefault VDCs, Sales-VDC and Storage-VDC. Storage-VDC is defined as the only storage VDC within the system.

The server CNA is connected to a port that is shared between both VDCs. Based on the tagging information, the switch distinguishes the type of traffic and directs the frames to the IP core of Sales-VDC, or to Storage-VDC toward the MDS switch and the storage disks that are attached to it. The port connecting to the Cisco MDS switch is dedicated to the storage VDC because it does not need to carry any network traffic.

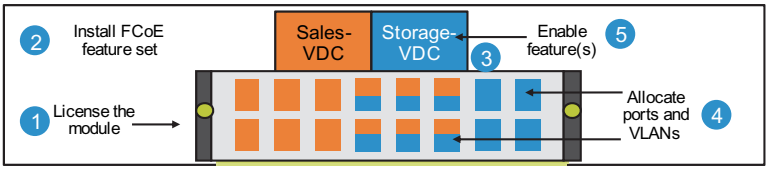
Interestingly, three types of links and traffic can be identified in this scenario:

- The link between the server CNA and the F1 module port is an FCoE link and carries two types of traffic—network and Fibre Channel.
- The link between the F1 module port and the MDS switch is an FCoE link but carries only Fibre Channel traffic. A Cisco Nexus 7000 Series switch does not provide any modules with native Fibre Channel ports. Such ports would present an alternative connection type to the MDS switch.
- The link between the Cisco MDS and the disk array is a native Fibre Channel link that carries only Fibre Channel traffic.

https://t.me/learningnets

Storage VDC Configuration Procedure

1. License each module configured for FCoE
2. Install FCoE feature set and enable features in default VDC
3. Create a dedicated storage VDC
4. Allocate ports and VLANs
   - Allocate dedicated FCoE ports to storage VDC (optional)
   - Allocate shared FCoE ports to storage VDC and another VDC (optional)
   - Allocate VLANs that can be used for FCoE and mapped to a VSAN
5. Enable features in storage VDC
   - Mandatory: Link Layer Discovery Protocol (LLDP)
   - Optional: Link Aggregation Control Protocol (LACP)



© 2012 Cisco and/or its affiliates. All rights reserved.

Follow this procedure to configure a storage VDC on a Cisco Nexus 7000 Series switch:

Step 1 License each module that is configured for FCoE.

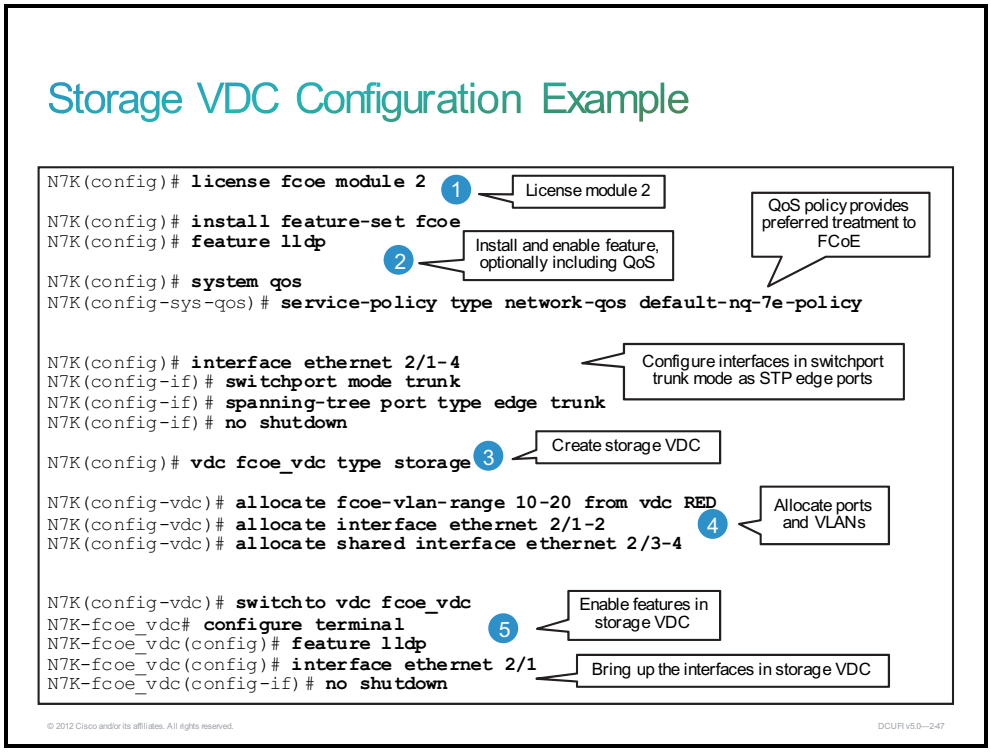
Step 2 Install the FCoE feature set and enable features in the default VDC. The mandatory feature is Link Layer Discovery Protocol (LLDP), which is required for FCoE negotiation. Optional features include Link Aggregation Control Protocol (LACP), which is used for port channels.

Step 3 Create a dedicated storage VDC.

Step 4 Allocate ports and VLANs. You may allocate dedicated or shared FCoE ports, or both, to the storage VDC. Furthermore, you must allocate VLANs that will be used for transporting FCoE traffic. Those VLANs will be mapped to virtual storage area networks (VSANs).

Step 5 Enable features in the storage VDC. LLDP must be enabled, while other features, such as LACP, are optional.

https://t.me/learningnets



This configuration example presents the Cisco NX-OS commands that are required in order to implement a storage VDC.

Apart from the necessary features and resource allocation that were discussed so far, you can see a quality of service (QoS) policy configuration and Spanning Tree Protocol (STP) configuration for the interfaces (Ethernet 2/1-4).

The QoS configuration enables a predefined QoS policy that grants the FCoE traffic the required preferred treatment.

Interfaces Ethernet 2/1-4 are put into trunk mode and configured as STP-type edge ports in order to support STP Lite for loop prevention. Ports 2/1-2 are dedicated to the storage VDC, while ports 2/3-4 are shared between the storage VDC and another VDC (RED).

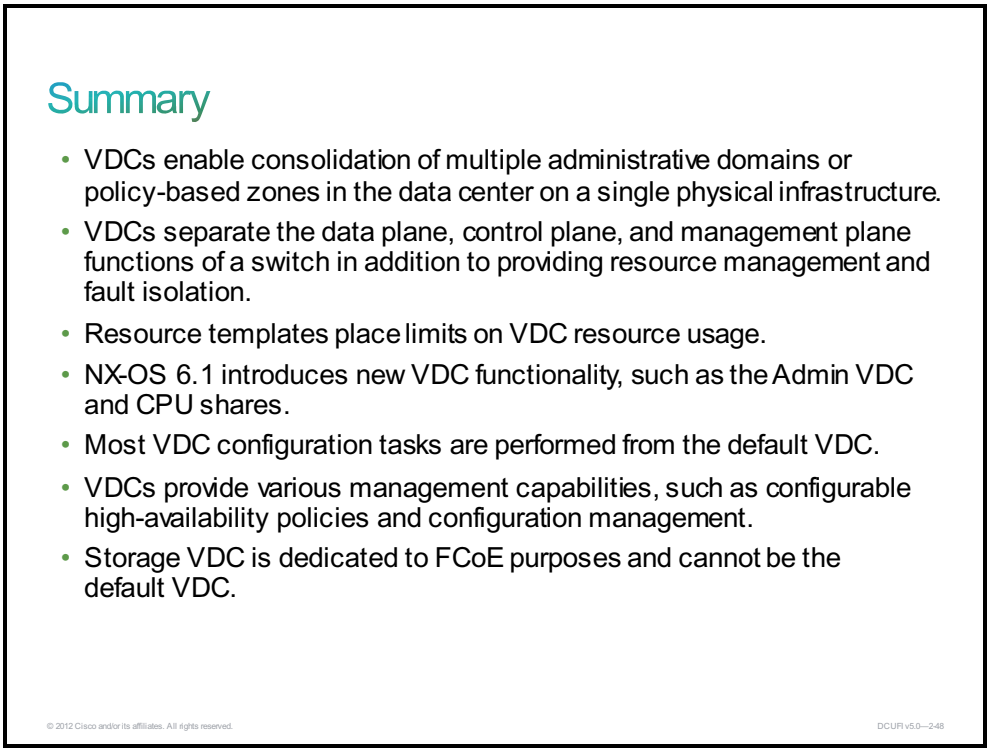
Apart from interface allocation, the storage VDC needs also to have VLANs allocated to it. This is done by using the **allocate fcoe-vlan-range** command. The VLANs can be shared with another VDC (RED in this example).

Note	Shutdown of the physical interface will also shut down the virtual interface in storage VDC (in current software version).

https://t.me/learningnets

Summary

This topic summarizes the key points that were discussed in this lesson.



References

For additional information, refer to these resources:

- To learn more about configuring VDCs on Cisco Nexus 7000 Series NX-OS, refer to *Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/sw/nx-os/virtual_device_context/configuration/guide/vdc_nx-os_cfg.html
- To learn more about configuring FCoE and storage VDCs on Cisco Nexus 7000 Series NX-OS, refer to *Cisco NX-OS FCoE Configuration Guide for Nexus 7000 and MDS 9500* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/sw/nx-os/fcoe/configuration/guide/b_Cisco_NX-OS_FCoE_Configuration_Guide.html

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 3

Configuring Layer 2 Switching Features

Overview

Layer 2 switching is a critical aspect of the data center network. To support the requirements of high-availability clusters and workload mobility, VLANs often need to be stretched across many different switches. To ensure that the foundation of the data center infrastructure is sound, Cisco Nexus switches support a wide range of features that help to scale, manage, and secure the Layer 2 switched network.

Objectives

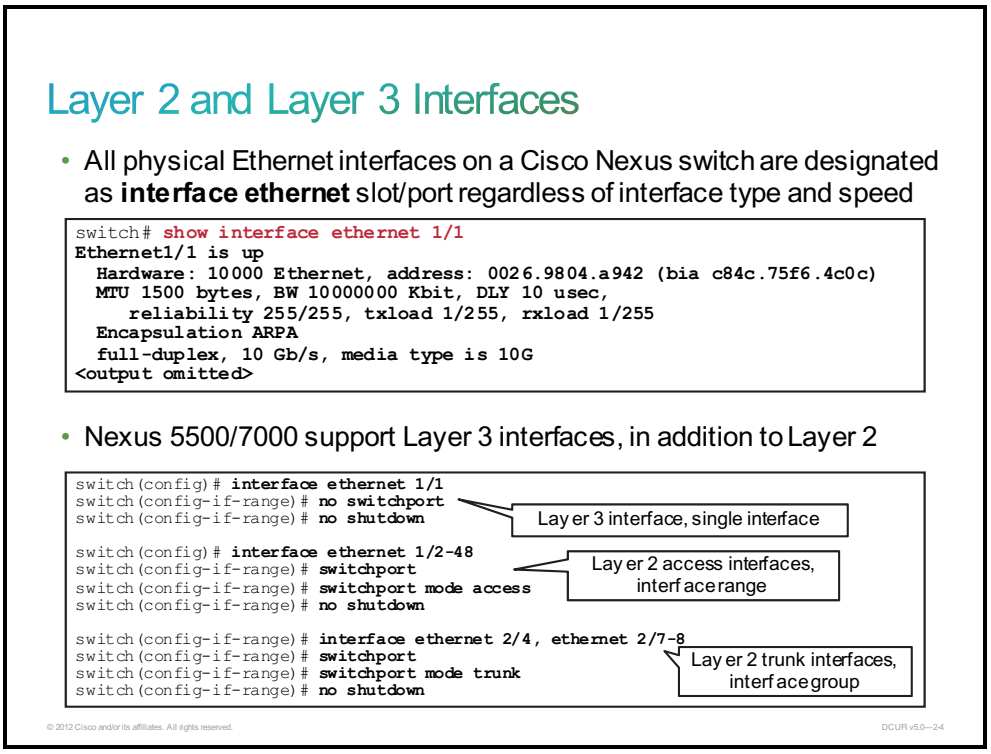
Upon completing this lesson, you will be able to configure Layer 2 switching features to support network requirements when given an implementation plan. You will be able to meet these objectives:

- Identify how to configure basic interface parameters on the Cisco Nexus 5000 and 7000 Series switch interfaces and Cisco Nexus 5500 Platform switch interfaces
- Identify the differences between the Layer 2 switching features of the Cisco Nexus 5000 and 7000 Series switches and the Cisco Nexus 5500 Platform switches
- Identify how to configure VLANs on Cisco Nexus switches
- Identify how to use and configure the STP extensions on Cisco Nexus switches

https://t.me/learningnets

Basic Interface Parameters

This topic identifies how to configure basic interface parameters on the Cisco Nexus 5000 and 7000 Series switch interfaces as well as the Cisco Nexus 5500 Platform switch interfaces.



Cisco Nexus Operating System (NX-OS) Software supports the following types of interfaces:

- **Physical**: Ethernet (10/100/1000/10G)
- **Logical**: PortChannel, loopback, null, switch virtual interface (SVI), tunnel, subinterface
- **In-Band**: Sup-eth0, Sup-core0
- **Management**: Management, Connectivity Management Processor (CMP)

All Ethernet interfaces are named “Ethernet.” There is no differentiation in the naming convention for different speeds.

The **show interface** command displays the operational state of any interface, including the reason why that interface might be down.

Interface Ranges and Groups

When configuring multiple interfaces with the same parameters, you can use the interface range feature rather than configuring each interface singularly.

The interface range configuration mode allows you to configure multiple interfaces with the same configuration parameters. After you enter interface range configuration mode, all command parameters that you enter are attributed to all interfaces within that range until you exit interface range configuration mode.

You enter a range of interfaces using hyphens (-) and commas (,). Hyphens separate contiguous interfaces, and commas separate discontiguous interfaces. When you enter discontiguous interfaces, you must enter the media type for each interface.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

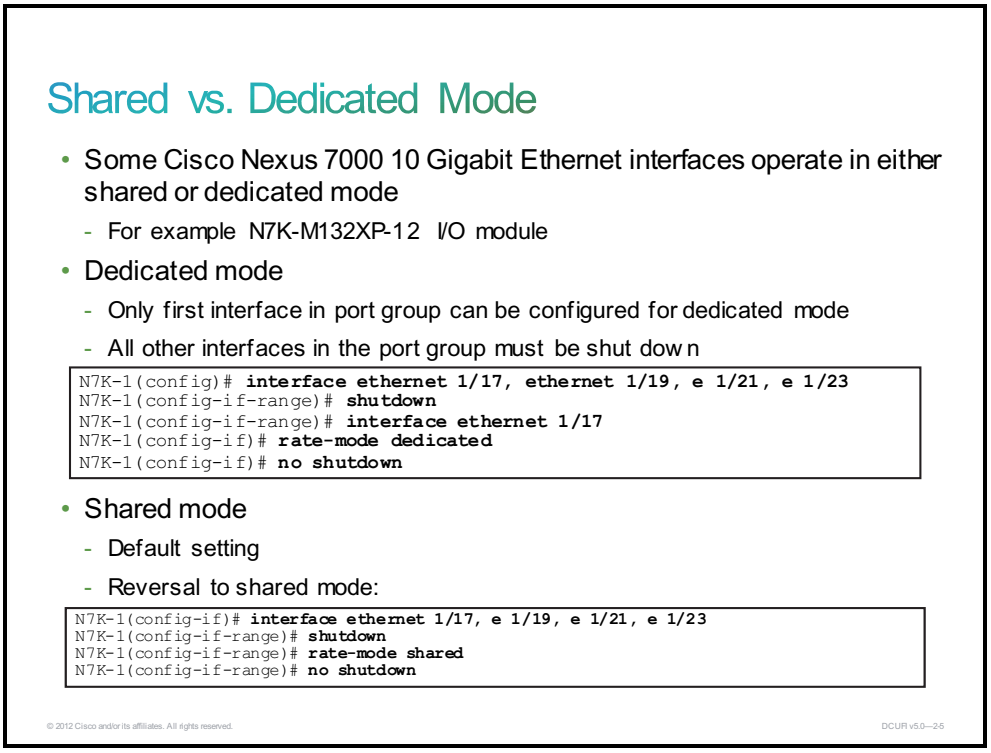
Cisco Nexus 5500 Platform switch and Cisco Nexus 7000 Series switch interfaces may operate as either Layer 2 switch ports or Layer 3 routed ports. Using the **no switchport** command while in interface configuration mode sets the interface or range of interfaces for Layer 3 operation.

Issuing the **switchport** command followed by the **switchport mode access** or **switchport mode trunk** commands sets the interface for Layer 2 operation.

Note The default mode of operation for all ports on a Cisco Nexus 7000 Series switch is Layer 3 mode. If you prefer that port default to be Layer 2 mode, use the **system default switchport** command to change this behavior.

---

https://t.me/learningnets



Cisco Nexus 7000 Series switch 10 Gigabit Ethernet interfaces on the N7K-M132XP-12(L) I/O modules are arranged into port groups that are serviced by a port group ASIC. There are eight port groups on a N7K-M132XP-12(L) I/O module, arranged as shown in the table.

Port Group	Interfaces
Port Group 1	Interfaces 1, 3, 5, and 7
Port Group 2	Interfaces 2, 4, 6, and 8
Port Group 3	Interfaces 9, 11, 13, and 15
Port Group 4	Interfaces 10, 12, 14, and 16
Port Group 5	Interfaces 17, 19, 21, and 23
Port Group 6	Interfaces 18, 20, 22, and 24
Port Group 7	Interfaces 25, 27, 29, and 31
Port Group 8	Interfaces 26, 28, 30, and 32

The port group ASIC provides 10 Gb/s of throughput to each port group. The interfaces in these port groups may operate in either a shared or dedicated mode.

When they operate in shared mode, all four interfaces within the port group are active and share the 10 Gb/s of throughput. When they operate in dedicated mode, only the first interface within each port group is active, and the other three are disabled.

Shared mode is typically used for server access, where full and continuous 10 Gb/s of uplink bandwidth may not be required. Dedicated mode is typically used for switch-to-switch uplinks and connections.

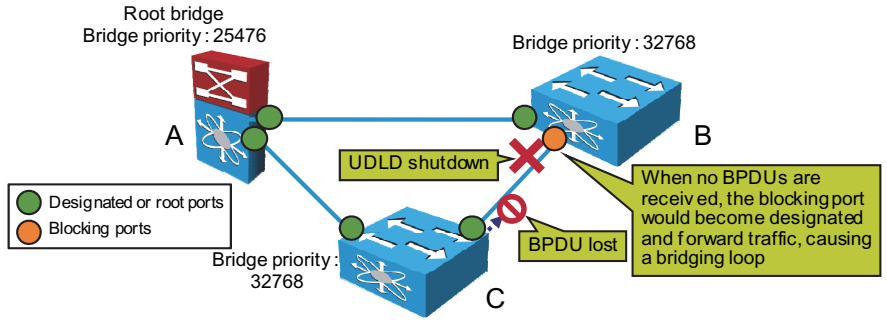
The bottom configuration in the figure shows the configuration steps to revert a range of interfaces to the shared mode.

Note The **show interface ethernet X/Y capabilities** command shows you the port group membership.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Unidirectional Link Detection (UDLD)

- Detects unidirectional links by combining Layer 1 and Layer 2 mechanisms
- When detected: shut down port and (optionally) generate syslog message
- If not detected: risk of bridging loop as two adjacent ports would be designated



© 2012 Cisco and/or its affiliates. All rights reserved.

Unidirectional Link Detection (UDLD) gives devices the ability to detect unidirectional links within the network. When a unidirectional link is detected, UDLD shuts down the affected LAN port and alerts the user. Unidirectional links can cause various problems, including spanning-tree topology loops.

UDLD works with Layer 1 protocols to determine the physical status of a link. At Layer 1, autonegotiation manages physical signaling and fault detection. At Layer 2, UDLD performs tasks that autonegotiation cannot perform. These tasks include detecting the identities of neighbors and shutting down misconnected LAN ports. When autonegotiation and UDLD are both enabled, Layer 1 and Layer 2 detection functions work together to prevent physical and logical unidirectional connections and the malfunctioning of other protocols.

A unidirectional link occurs when two-way traffic is suddenly reduced to traveling in a single direction. If a strand from a fiber pair is disconnected, autonegotiation ensures that the link becomes suspended. In this case, the logical link is undetermined, and UDLD takes no action. If both fibers are working normally at Layer 1, UDLD determines whether both fibers are connected correctly and whether traffic is flowing bidirectionally between the two neighbors. This task cannot be performed by autonegotiation because autonegotiation is restricted to Layer 1.

The switches periodically transmit UDLD packets to neighbor devices on LAN ports with UDLD enabled. If the packets are echoed back without a specific acknowledgment (echo), the link is then marked as unidirectional and the port is shut down. Devices on both ends of the link must support UDLD for the protocol to successfully identify and disable unidirectional links.

UDLD uses a special MAC address: 0100.0CCC.CCCC.

https://t.me/learningnets

Configuring UDLD

1. Enable UDLD in normal mode for all fiber-optic interfaces.
2. Enable aggressive mode for all fiber-optic interfaces (optional)
   - When a port stops receiving UDLD frames, it tries to reestablish the UDLD connection 8 times, then the port is disabled.
3. Modify individual interfaces (optional)
   - Disable, re-enable, or enable using aggressive mode
4. View UDLD neighbors (optional)

```
switch(config)# feature udld                                          1

switch(config)# udld aggressive                                       2

switch(config)# interface ethernet 2/2, ethernet 2/4                 3
switch(config-if)# udld disable

switch# show udld neighbors                                           4
Port             Device Name     Device ID   Port ID         Neighbor State
---------------------------------------------------------------------------
Ethernet2/1      TBM12234230     1           Ethernet2/1     bidirectional
Ethernet2/3      TBM12234230     1           Ethernet2/3     bidirectional
```

© 2012 Cisco and/or its affiliates. All rights reserved.

To use UDLD on the Cisco Nexus switches, enable the UDLD feature by using the **feature udld** command.

After globally enabling all 10-Gb (fiber) interfaces, run UDLD automatically. However, for the 1-Gb (copper) interfaces, UDLD must be manually enabled per each interface.

UDLD supports two operational modes—normal mode, which is the default, and aggressive mode, which must be specifically enabled.

UDLD aggressive mode can only be used on point-to-point links between network devices that are capable of supporting this mode. When a port on a bidirectional link stops receiving UDLD packets, UDLD tries to reestablish the connection with the affected neighbor. UDLD disables the port after eight failed retries.

UDLD configuration commands are as follows:

- **feature udld**: Enables the UDLD feature
- **udld aggressive**: Enables aggressive mode globally
- **interface** *type slot/port*: Enters the interface subconfiguration mode
- **udld** {enable | disable | aggressive}: Configures the UDLD mode on the interface

When UDLD is configured globally, the following must be taken into consideration:

- All 10-Gb (fiber) interfaces run UDLD automatically.
- For 1-Gb (copper) interfaces, you must manually enable UDLD per each interface.

https://t.me/learningnets

Port Profiles

- Groups of commands can be configured on interfaces through a port profile.
- Separate types of profiles exist for Ethernet, VLAN, and PortChannel interfaces.
- Port profiles can be inherited in a hierarchical manner.
- Port profile type and interface mode (Layer 2 or Layer 3) need to match in order to inherit a profile.

```
switch(config)# port-profile type ethernet SERVERS
switch(config-port-prof)# switchport
switch(config-port-prof)# no shutdown
switch(config-port-prof)# spanning-tree port type edge
switch(config-port-prof)# switchport mode access

switch(config)# port-profile type ethernet WEB-SERVERS
switch(config-port-prof)# switchport
switch(config-port-prof)# switchport access vlan 10
switch(config-port-prof)# inherit port-profile SERVERS
switch (config-port-prof)# state enabled

switch(config)# interface ethernet 1/1-8
switch(config-if)# inherit port-profile WEB-SERVERS
```

Profile inherits profile SERVERS

Port inherits profile WEB-SERVERS and SERVERS

© 2012 Cisco and/or its affiliates. All rights reserved.

On Cisco Nexus switches, you can create a port profile that contains many interface commands and then apply that port profile to a range of interfaces. Each port profile can be applied only to a specific type of interface. The supported interface types are Ethernet, VLAN, or PortChannel interfaces.

Note When you choose Ethernet as the interface type, the port profile is in the default mode, which is Layer 3. Enter the **switchport** command to change the port profile to Layer 2 mode.

You inherit the port profile when you attach the port profile to an interface or range of interfaces. When you attach—or inherit—a port profile to an interface or range of interfaces, the system applies all the commands in that port profile to the interfaces.

Note To apply the commands in the port profile to the interface, the port profile needs to be enabled through the **state enabled** command. By default, port profiles are not enabled.

Additionally, you can have one port profile inherit another port profile, which allows the initial port profile to assume all of the commands of the second inherited port profile that do not conflict with the initial port profile. Four levels of inheritance are supported, except for the **switchport private-vlan mapping** and **private-vlan mapping** commands, which support only one level of inheritance.

https://t.me/learningnets

Verifying Port Profiles

Verify port profile configuration, inheritance, and evaluated configuration

```
N7K-1# show port-profile name WEB-SERVERS

port-profile WEB-SERVERS
 type: Ethernet
 description:
 status: enabled
 max-ports: 512
 inherit:
  SERVERS
 config attributes:
  switchport
  switchport access vlan 10
 evaluated config attributes:
  switchport
  switchport mode access
  spanning-tree port type edge
  switchport access vlan 10
  no shutdown
 assigned interfaces:
  Ethernet1/1
  Ethernet1/2
  Ethernet1/3
  Ethernet1/4
<output omitted>
```

Profile inherits profile SERVERS

Configuration attributes

Interfaces to which the profile has been applied

© 2012 Cisco and/or its affiliates. All rights reserved.

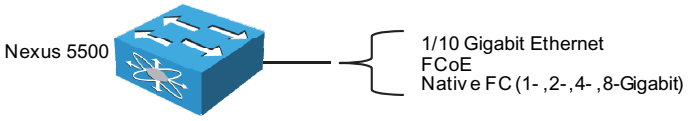
Use the **show port-profile** command to display information about the configured port profiles on the device. If the command is used without any additional parameters, it displays all configured port profiles. Further command options can be used to gather more specific information. The following options can be used:

- **show port-profile expand-interface**: This option shows the expanded interface configuration for each interface that has a port profile applied to it. Output can be limited to a specific port profile.
- **show port-profile usage**: This option shows which interfaces have a specific port profile applied to them. The name keyword can be used to limit the output to a specific port profile.

https://t.me/learningnets

Cisco Nexus Unified Port Technology

- Configure a physical port as one of:
  - 1/10-Gigabit Ethernet
  - Fibre Channel over Ethernet (FCoE)
  - 1-, 2-, 4-, 8-Gigabit native Fibre Channel port
- Available on Cisco Nexus 5500 Platform switches
  - Cisco NX-OS Release 5.0(3)N1(1b) or later
  - Cisco Nexus 5548UP and 5596UP Switches and expansion modules
- Aspects of unified fabric:
  - Unified platform (same platform architecture and software for LAN and SAN)
  - Unified device (cabling the same device)
  - Unified wire (convergence on a single CNA and cable)



© 2012 Cisco and/or its affiliates. All rights reserved.

Beginning with Cisco NX-OS Release 5.0(3)N1(1b), Cisco introduced the Cisco Nexus unified port technology. Cisco Nexus unified ports allow you to configure a physical port on a Cisco Nexus 5500 Platform switch as a 1/10-Gigabit Ethernet, Fibre Channel over Ethernet (FCoE), or 1-, 2-, 4-, or 8-Gigabit native Fibre Channel port.

Currently, most networks have two types of switches for different types of networks. For example, LAN switches carry Ethernet traffic up to Cisco Catalyst switches, and SAN switches carry Fibre Channel traffic from servers to Cisco MDS switches. With unified port technology, you can deploy a unified platform, unified device, and unified wire approach. Unified ports allow you to move from an existing segregated platform approach—where you choose LAN and SAN port options—to a single unified fabric that is transparent and consistent with existing practices and management software. A unified fabric includes the following:

- **Unified platform**: Uses the same hardware platform and the same software code level and certifies it once for your LAN and SAN environments.
- **Unified device**: Runs LAN and SAN services on the same platform switch. The unified device allows you to connect your Ethernet and Fibre Channel cables to the same device.
- **Unified wire**: Converges LAN and SAN networks on a single Converged Network Adapter (CNA) and connects them to your server.

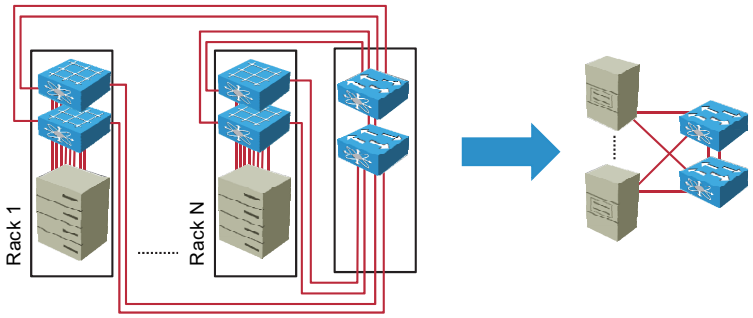
A unified fabric allows you to manage Ethernet and FCoE features independently by using the existing Cisco tools.

The new Cisco Nexus 5548UP Switch and the Cisco Nexus 5596UP Switch provide built-in unified port technology. In addition, a new unified port expansion module and two Layer 3 modules increase the benefits of a deployed unified fabric.

https://t.me/learningnets

Fabric Extenders

- The Cisco Nexus 2000 Series Fabric Extenders serve as remote I/O modules of a Cisco Nexus 5000/5500 or 7000 Switch:
  - Managed and configured from parent switch
- Together, parent switches and Cisco Nexus 2000 Series Fabric Extender combine benefits of top-of-rack cabling with end-of-row management



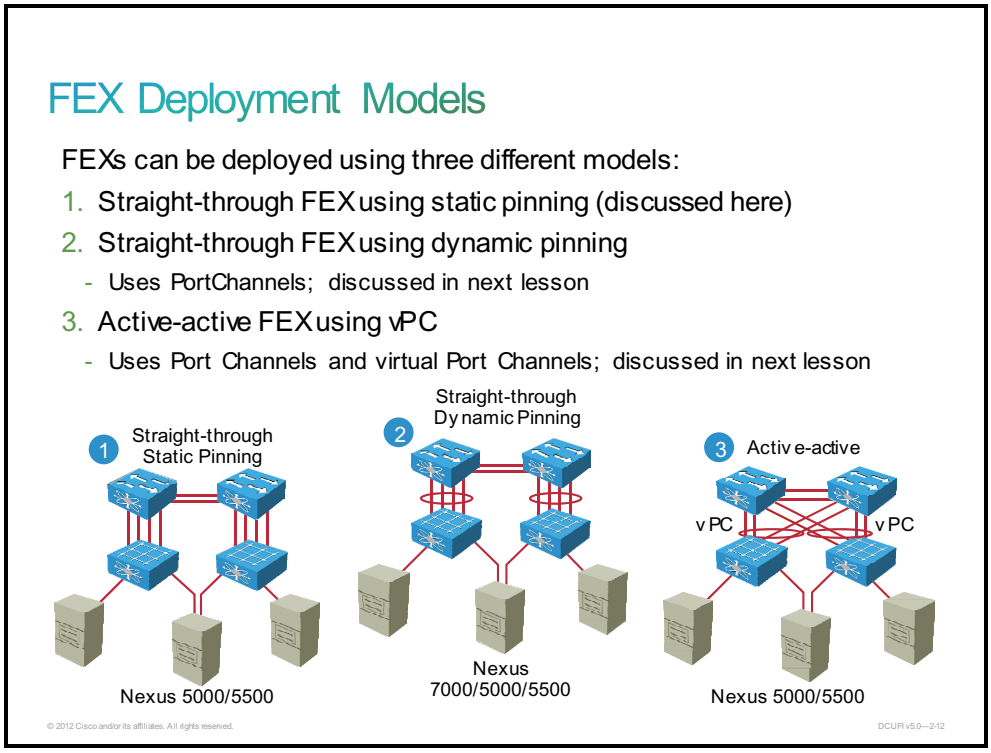
© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco Nexus 2000 Series Fabric Extenders (FEXs) can be deployed together with Cisco Nexus 5000 or Cisco Nexus 7000 Series switches to create a data center network that combines the advantages of a top-of-rack (ToR) design with the advantages of an end of row (EoR) design.

Dual-redundant Cisco Nexus 2000 Series FEXs are placed at the top of each rack. The uplink ports on the Cisco Nexus 2000 Series FEXs are connected to a Cisco Nexus 5000 or 7000 Series switch that is installed in the EoR position. From a cabling standpoint, this design is a ToR design. The cabling between the servers and the Cisco Nexus 2000 FEX is contained within the rack. Only a limited number of cables need to be run between the racks to support the 10 Gigabit Ethernet connections between the Cisco Nexus 2000 Series FEXs and the Cisco Nexus switches in the EoR position.

From a network deployment standpoint, however, this design is an EoR design. The FEXs act as remote I/O modules for the Cisco Nexus switches, which means that the ports on the Cisco Nexus 2000 Series FEX act as ports on the associated switch. In the logical network topology, the FEXs disappear from the picture, and all servers appear as directly connected to the Cisco Nexus switch. From a network operations perspective, this design has the simplicity that is normally associated with EoR designs. All the configuration tasks for this type of data center design are performed on the EoR switches. There are no configuration or software maintenance tasks that are associated with the FEXs.

https://t.me/learningnets



There are three deployment models that are used to deploy FEXs together with the Cisco Nexus 5000 and Cisco Nexus 7000 Series switches:

- **Straight-through using static pinning**: In the straight-through model, each FEX is connected to a single Cisco Nexus switch. The single switch that the FEX is connected to exclusively manages the ports on that FEX. Static pinning means that each downlink server port on the FEX is statically pinned to one of the uplinks between the FEX and the switch. Traffic to and from a specific server port always uses the same uplink. This model is discussed in this lesson.
- **Straight-through using dynamic pinning**: This deployment model also uses the straight-through connection model between the FEXs and the switches. However, there is no static relationship between the downlink server ports and the uplink ports. The ports between the FEX and the switch are bundled into a port channel, and traffic is distributed across the uplinks based on the port channel hashing mechanism. Port channels and this FEX deployment model are discussed in the "Configuring Port Channels" lesson.
- **Active-active FEX using virtual port channel (vPC)**: In this deployment model, the FEX is dual-homed to two Cisco Nexus switches. vPC is used on the link between the FEX and the pair of switches. Traffic is forwarded between the FEX and the switches based on vPC forwarding mechanisms. vPC and this FEX deployment model are discussed in the "Configuring Port Channels" lesson.

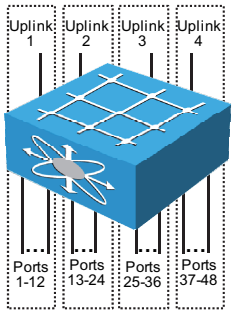
Note	Cisco Nexus 7000 Series switches currently support only straight-through deployment using dynamic pinning. Static pinning and active-active FEX are currently supported only on Cisco Nexus 5000 Series switches.

https://t.me/learningnets

FEX with Static Pinning

- Static pinning statically maps server-FEX downlink ports to the uplink ports that connect the FEX to the parent switch
- Port mapping depends on the number of uplink ports that are used and the number of downlinks on the FEX
- When an uplink port fails, all downlink ports pinned to it are disabled
  - Oversubscription ratio is preserved
  - Single-homed servers lose connectivity
  - Dual-homed servers fail over to the other NIC

Example:
Cisco Nexus
2248TP GE



© 2012 Cisco and/or its affiliates. All rights reserved.

In static pinning mode, the server ports on the Cisco Nexus 2000 Series FEX are statically pinned to one of the uplink ports. For example, when a Cisco Nexus 2000 Series FEX with 48 Gigabit Ethernet server ports is deployed in static pinning mode using four 10 Gigabit Ethernet uplink ports to the Cisco Nexus 5000 Series switch, 12 server ports will be pinned to each uplink port. Ports 1–12 are pinned to the first uplink, ports 13–24 to the second uplink, ports 25–36 to the third uplink, and ports 37–48 to the fourth uplink. This results in an oversubscription ratio of 1.2:1, because a group of 12 Gigabit Ethernet server ports shares the bandwidth of one 10 Gigabit Ethernet uplink.

If one of the uplinks between the Cisco Nexus 2000 Series FEX and the Cisco Nexus 5000 Series switch fails, the FEX will disable the server ports that are pinned to that uplink port. For example, if the fourth uplink fails, server ports 37–48 will be disabled. Servers that are connected to these ports will see the associated Ethernet link go down. If the servers are dual-homed and use some form of network interface card (NIC) redundancy, then this mechanism will be triggered, and the server will fail over to the other NIC. A single-homed server will simply lose connectivity if it is connected to one of the ports that are pinned to the failed uplink port. The oversubscription ratio on the other ports remains unchanged. The oversubscription rate remains unchanged because each of the other three groups of 12 Gigabit Ethernet ports is still sharing the same 10 Gigabit Ethernet uplink port as before.

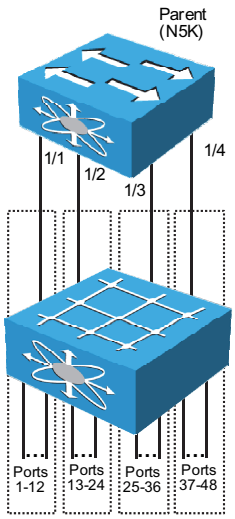
https://t.me/learningnets

Configuring FEX with Static Pinning

All FEX configuration performed on parent switch:

1. Enable the FEX feature
2. Configure the FEX instance number
3. Define number of uplinks used for static pinning
4. Set FEX-Fabric mode
5. Associate the ports with the FEX

```
N5K(config)# feature fex                                    1
N5K(config)# fex 111                                        2
N5K(config-fex)# description "FEX 111, rack 1"
N5K(config-fex)# pinning max-links 4                        3
Change in Max-links will cause traffic disruption
N5K(config)# interface ethernet 1/1-4
N5K(config-if-range)# switchport mode fex-fabric            4
N5K(config-if-range)# fex associate 111                     5
```



© 2012 Cisco and/or its affiliates. All rights reserved.

All configuration and discovery for the Cisco Nexus 2000 Series FEX is performed from the parent switch and involves these steps:

Step 1 Enable the FEX feature (for Cisco Nexus 5000 Series switch and 5500 Platform switch). The equivalent installation and enabling occurs on Cisco Nexus 7000 Series switches in a virtual device context (VDC).

Step 2 Create the FEX instance. To create an FEX instance, issue the **fex** *chassis-number* command from within the global configuration context. The chassis number may be any integer from 100–199.

Step 3 Once the FEX instance has been created, the configuration context changes to FEX configuration mode, where a description may be added. The **pinning max-links** *1–4* command binds the 48 server-facing ports to the uplink ports (up to four static ports may be activated), according to the following **max-links** argument.

pinning max-links 1	All 48 server-facing ports will be pinned to a single active uplink port (interface Ethernet CN/1/1-48), where CN = chassis number.
pinning max-links 2	All 48 server-facing ports will be pinned to two active uplink ports (interface Ethernet CN/1/1-24 assigned to the first active uplink port and interface Ethernet CN/1/25-48 assigned to the second active uplink port).
pinning max-links 3	All 48 server-facing ports will be pinned to three active uplink ports (interface Ethernet CN/1/1-16 assigned to the first active uplink port, interface Ethernet CN/1/17-32 assigned to the second active uplink port, and interface Ethernet CN/1/33-48 assigned to the third active uplink port).
pinning max-links 4	All 48 server-facing ports will be pinned to all four uplink ports (Ethernet CN/1/1-12 assigned to first active uplink port, Ethernet CN/1/13-24 assigned to the second active uplink port, Ethernet CN/1/25-36 assigned to third active uplink port, and Ethernet CN/1/37-48 assigned to fourth active uplink port).

Step 4 Configure the interface mode by using the **switchport mode fex-fabric** command.

Step 5 Associate a parent switch interface to a Cisco Nexus 2000 Series FEX uplink port for a specific FEX by using the **fex associate** *chassis-number* command.

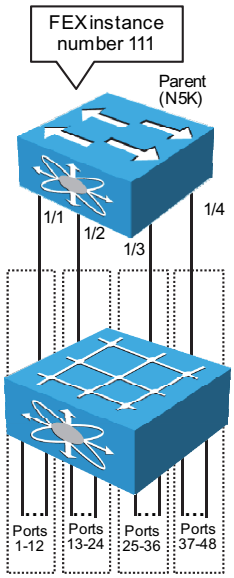
https://t.me/learningnets

Auto-Created FEX Interfaces

- Downlink ports visible on parent switch
- FEX number used as a "virtual slot" number

```
N5K# show running-config | begin "interface Ethernet111"
interface Ethernet111/1/1

interface Ethernet111/1/2

interface Ethernet111/1/3
<output omitted>

N5K# show fex 111
FEX: 111 Description: FEX 111, rack 1, top   state: Online
  FEX version: 5.0(2)N2(1) [Switch version: 5.0(2)N2(1)]
  Extender Model: N2K-C2248TP-1GE,  Extender Serial:
JAF1420AHPE
  Part No: 73-12748-05
 pinning-mode: static    Max-links: 4
  Fabric port for control traffic: Eth1/1
  Fabric interface state:
    Eth1/1 - Interface Up. State: Active
    Eth1/2 - Interface Up. State: Active
    Eth1/3 - Interface Up. State: Active
    Eth1/4 - Interface Up. State: Active
```

FEX interfaces on parent switch



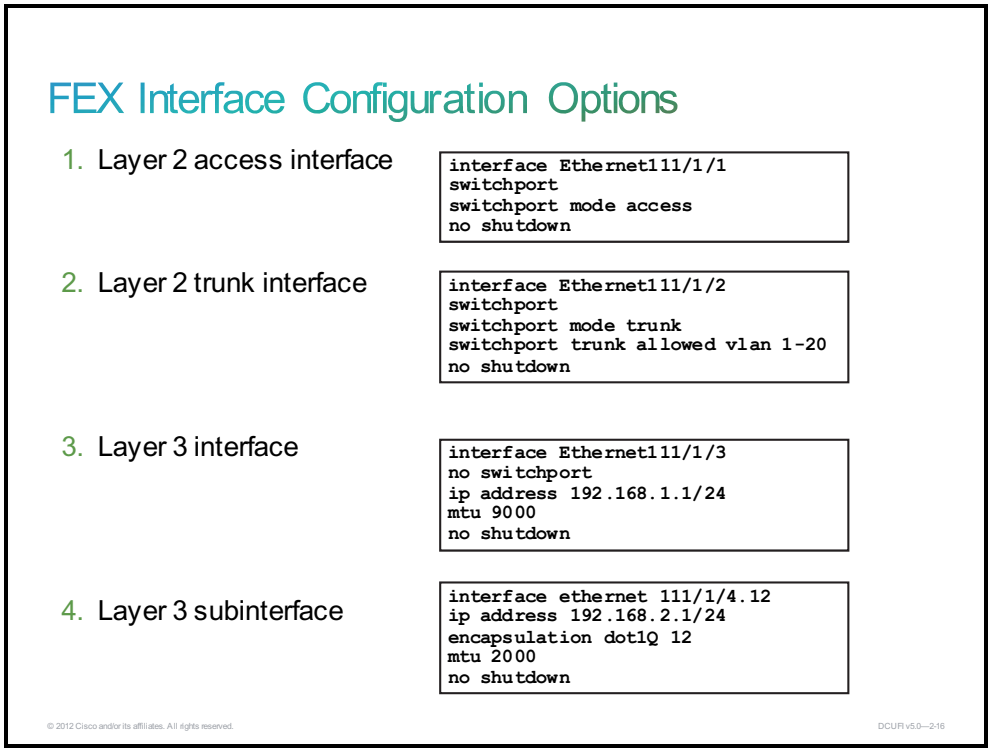
© 2012 Cisco and/or its affiliates. All rights reserved.

Once a single uplink interface has been associated to the FEX chassis, the FEX becomes active. The switch performs a software check to compare the software on the FEX to the software on the switch. If it turns out that the software on the Cisco Nexus switch is more recent than the software on the switch, the switch will trigger a download of the latest software to the FEX.

When the FEX comes online, the ports on the FEX are visible as ports on the switch. The ports can then be configured from the switch as if they were local ports.

To get a brief overview of the status of the FEX, use the **show fex** *chassis-number* command. The **show fex** *chassis-number* command displays the state of the FEX, the pinning mode, max-links, and the state of the physical and logical uplink interfaces.

https://t.me/learningnets



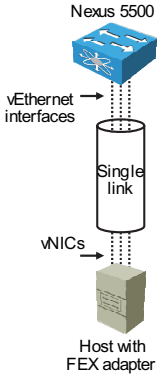
Each autoconfigured FEX interface represents a logical attachment of a host NIC to the parent switch. The parent switches offer the full range of Layer 2 and Layer 3 options to be configured on the autoconfigured interfaces. The four common cases include:

1. Layer 2 access interfaces.
2. Layer 2 trunk interfaces. You may configure additional settings, such as allowed VLAN ranges.
3. Layer 3 interfaces.
4. Layer 3 subinterfaces. The subinterfaces require configuration of the 802.1Q tag in addition to an IPv4 or IPv6 address.

https://t.me/learningnets

Adapter FEX

- Single physical link split into multiple virtual channels
- Channel
  - Identified by a unique channel number
  - Channel scope limited to the physical link
  - Connects a server vNIC with a switch vEthernet interface
  - Uses tagging with VNTag identifiers
- Support
  - Switch-side:
    - Nexus 5500
    - Nexus 2200 connected to Nexus 5500
  - Server-side:
    - Cisco UCS P81E Virtual Interface Card for UCS C-Series
    - Third-party adapters that support the VNTag technology
      - For example, Broadcom BCM57712 Convergence NIC



© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco NX-OS Adapter FEX feature combines the advantages of the FEX link architecture with server I/O virtualization to create multiple virtual interfaces over a single Ethernet interface. You can deploy a dual-port NIC on the server and configure more than two virtual interfaces that the server recognizes as a regular Ethernet interface. The advantage of this approach is that it allows you to reduce power and cooling needs and to reduce the number of network ports.

Cisco Adapter FEX can be thought of as a way to divide a single physical link into multiple virtual links or channels. Each channel is identified by a unique channel number, and its scope is limited to the physical link.

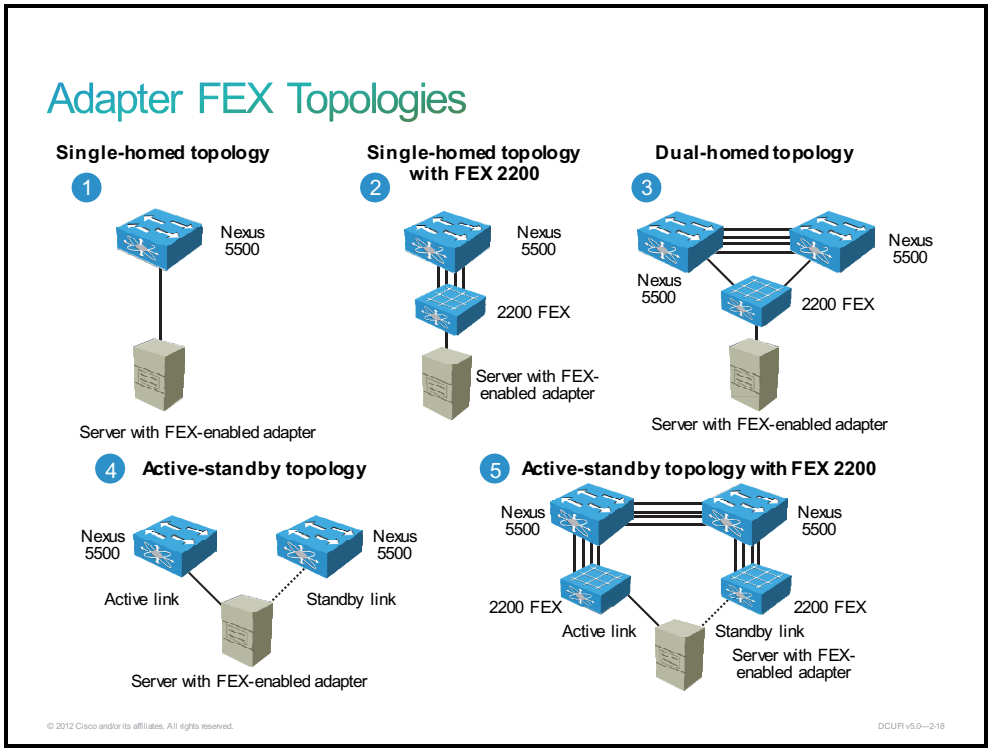
The physical link connects a port on a server network adapter with an Ethernet port on the switch. This allows the channel to connect a virtual network interface card (vNIC) on the server with a virtual Ethernet (vEthernet) interface on the switch.

Packets on each channel are tagged with a virtual network tag (VNTag) that has a specific source virtual interface identifier (VIF). The VIF allows the receiver to identify the channel that the source used to transmit the packet.

Cisco Adapter FEX requires a server network adapter that is connected to a parent switch that supports Cisco Adapter FEX functionality. Cisco Adapter FEX support is available with Cisco Nexus 5500 Platform Switches and with Cisco Nexus 2200 FEXs that are connected to a Cisco Nexus 5500 parent Platform switch.

This implementation is designed to work with server network adapters, such as the Cisco UCS P81E Virtual Interface Card for the Cisco UCS C-Series Rack-Mount Server (UCS P81E VIC) or third-party adapters that support the VNTag technology, such as the Broadcom BCM57712 Convergence Network Interface Card.

https://t.me/learningnets



The figures show examples of Cisco Nexus 5500 Platform switch Adapter FEX topologies with server network adapters.

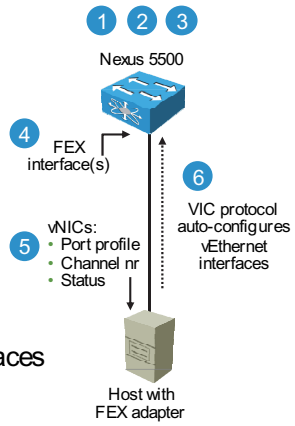
Numbers 4 and 5 show topologies that support active/standby teaming of uplinks. The active/standby topologies shown here have one uplink as active and the other uplink as a standby. With some server network adapters, you can select the active and standby uplinks per each vNIC. In this case, each uplink is an active uplink for a specific vNIC and becomes a standby for the remaining uplinks. Selecting the active and standby uplinks per vNIC is a recommended practice.

Note	The Cisco UCS P81E Virtual Interface Card supports active/standby uplink teaming. The Cisco UCS P81E allows each vNIC to choose the active uplink. The other uplink is configured as a standby uplink.

https://t.me/learningnets

Adapter FEX Configuration Procedure

1. Install and enable virtualization feature
2. Enable automatic creation of vEthernet interfaces (optional)
3. Configure port profiles
4. Configure FEX interface(s)
5. Configure vNICs
   - Port-profile name
   - Channel number
   - Active/standby status
6. Let Virtual Interface Configuration (VIC) protocol auto-configure the vEthernet interfaces or configure them manually
   - For automatic creation, step 2 is necessary



© 2012 Cisco and/or its affiliates. All rights reserved.

Follow this procedure to configure Cisco Adapter FEX:

Step 1 Install and enable the virtualization feature on the parent switch.

Step 2 Enable automatic creation of vEthernet interfaces. This step is optional and allows the switch to respond to the Virtual Interface Configuration (VIC) protocol requests by automatically configuring vEthernet interfaces.

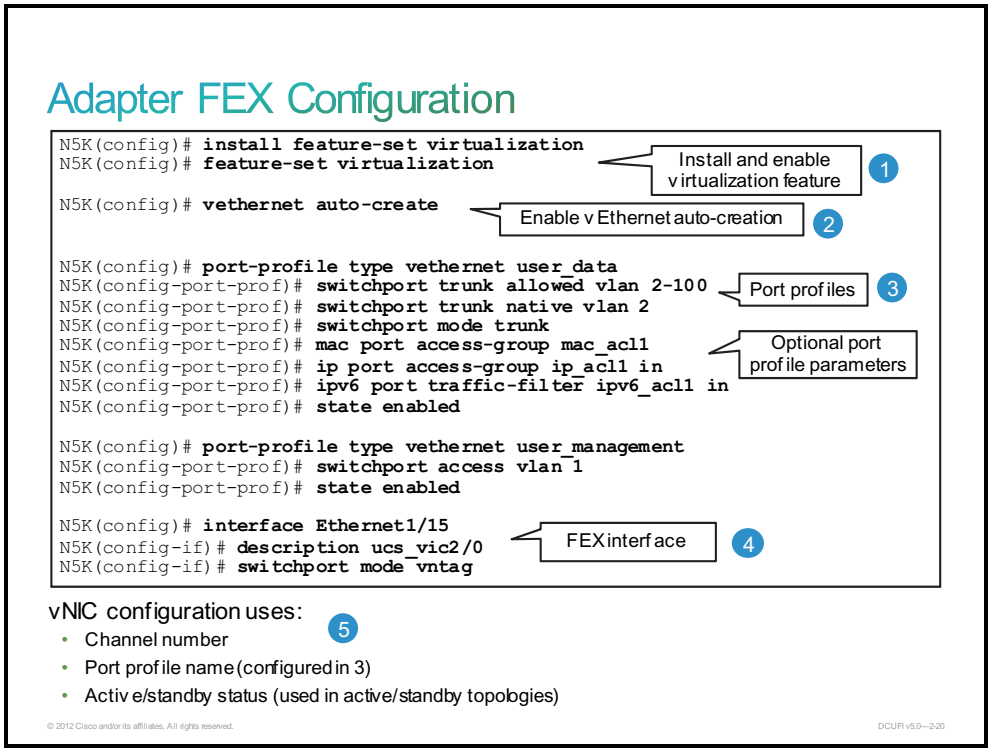
Step 3 Configure port profiles. Port profiles act as parameter containers for the autoconfigured vEthernet interfaces. They define relevant properties and policies, such as VLAN, bandwidth, quality of service (QoS), and access control lists (ACLs).

Step 4 Configure the Cisco FEX interface or interfaces that connect the parent switch to either the FEX 2200 or directly to the server with an FEX-enabled NIC. These interfaces need to be configured as switch ports for VNTag mode.

Step 5 Using the network adapter configuration utility on the server, create the appropriate number of vNICs. Create each vNIC with the appropriate properties, such as a unique channel number, MAC address, uplink failover properties, and port profile names. Each vNIC has a unique channel number associated with it. A vNIC is identified on the switch by the **bind** command, which associates a physical port and the vNIC channel number to a vEthernet interface.

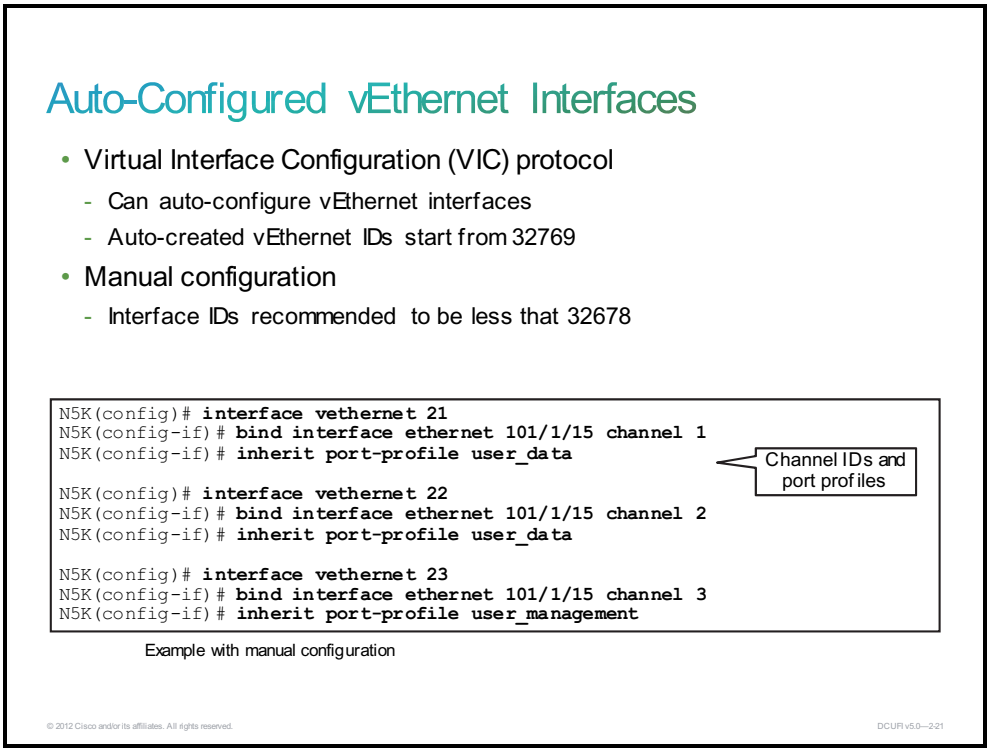
Step 6 Allow VIC protocol provision vEthernet interfaces on the parent switch. When the configuration is complete, the server network adapter and the switch re-establish a link and perform the initial handshake and negotiation process. The server network adapter and the switch establish higher-level control plane connectivity using the VIC protocol. When VIC connectivity is established, the server network adapter requests that the switch create a vEthernet interface for each vNIC that is configured on the server network adapter. The server network adapter passes the port profile name, channel number, and the active/standby status over the uplink in addition to the request to create a vEthernet interface. The switch responds by creating a vEthernet interface for each vNIC on the server network adapter and associates the port profile and channel number with the vEthernet interface.

https://t.me/learningnets



The configuration example illustrates the configuration steps that were described.

https://t.me/learningnets



The vEthernet interfaces can either be autoprovisioned by the VIC protocol or configured manually.

vEthernet interfaces that are created by the switch are numbered automatically as they are created. These vEthernet numbers start from 32769. The switch picks the lowest unused number when creating a vEthernet interface.

When you manually create vEthernet interfaces, you may select any number for the vEthernet. However, as a best practice, you should choose a number that is less than 32678.

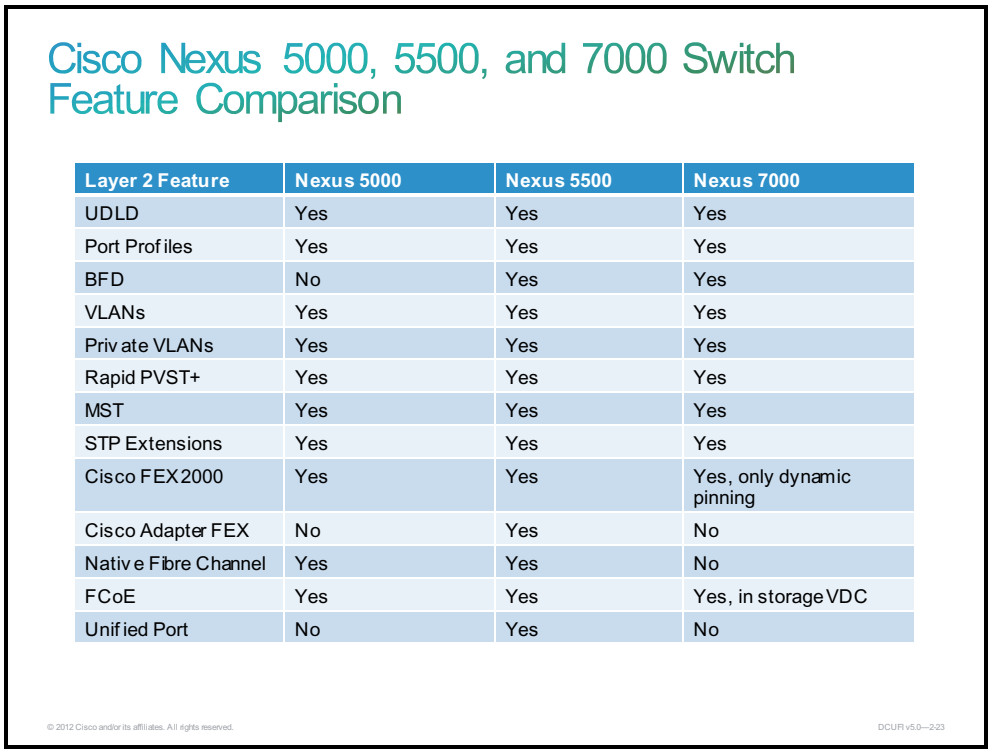
The vEthernet interfaces have two commands:

- The **bind** command specifies the channel ID and the Cisco FEX interface.
- The **inherit** command indicates the port profile from which the interface obtains its settings.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Cisco Nexus 7000 and Cisco Nexus 5000 Switch Feature Comparison

This topic identifies the differences between the Layer 2 switching features of the Cisco Nexus 5000 and 7000 Series switches and the Cisco Nexus 5500 Platform switches..



Layer 2 Feature	Nexus 5000	Nexus 5500	Nexus 7000
UDLD	Yes	Yes	Yes
Port Profiles	Yes	Yes	Yes
BFD	No	Yes	Yes
VLANs	Yes	Yes	Yes
Private VLANs	Yes	Yes	Yes
Rapid PVST+	Yes	Yes	Yes
MST	Yes	Yes	Yes
STP Extensions	Yes	Yes	Yes
Cisco FEX 2000	Yes	Yes	Yes, only dynamic pinning
Cisco Adapter FEX	No	Yes	No
Native Fibre Channel	Yes	Yes	No
FCoE	Yes	Yes	Yes, in storage VDC
Unified Port	No	Yes	No

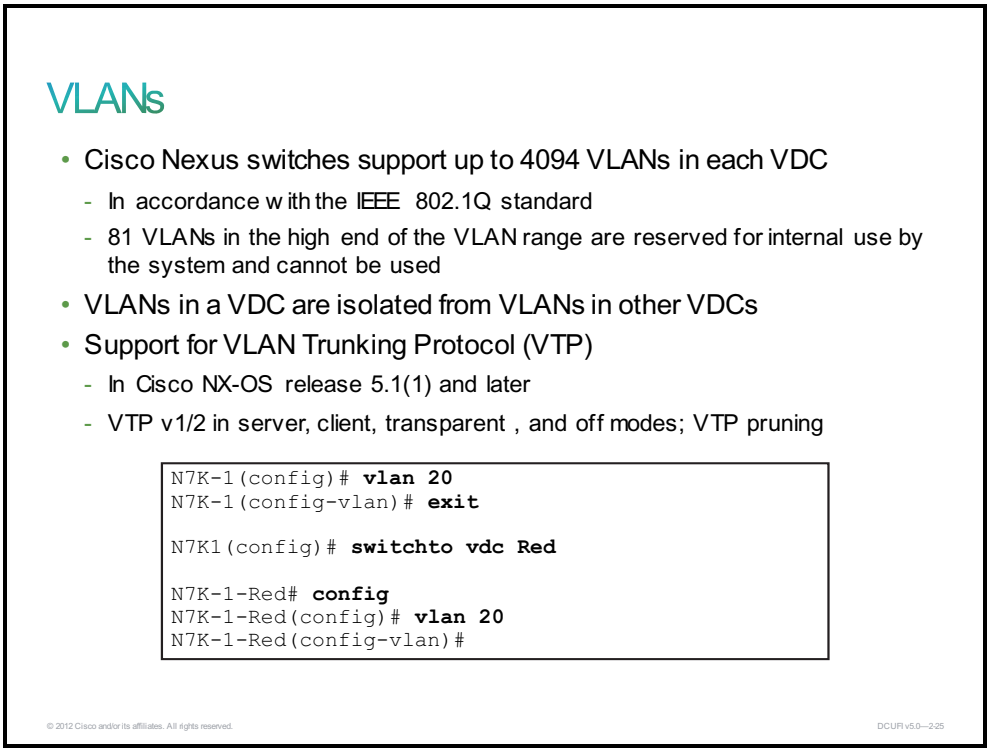
The Cisco Nexus 5000 and 7000 Series switches and Cisco Nexus 5500 Platform switches run Cisco NX-OS Software and share many features and functions. However, because these switches are positioned for different roles, the supported feature sets on the platforms differ. Some features are specific to only one of the platforms. Also, the software releases for both platforms are on independent release cycles, meaning that features that have been released for one of the platforms may not yet be available for another.

All of the platforms support an extensive set of Layer 2 switching features. The figure highlights some of the major differences and similarities between them.

https://t.me/learningnets

VLAN Configuration

This topic identifies how to configure VLANs on Cisco Nexus switches.



Layer 2 ports on Cisco Nexus switches can be configured as access ports or 802.1Q trunk ports. By default, they are configured as access ports.

A switch port belongs to a VLAN. Unicast, broadcast, and multicast packets are forwarded and flooded only to end stations in that VLAN. Each VLAN is considered a logical network. Packets that are destined for a station that does not belong to the same VLAN must be forwarded through a router.

The Cisco Nexus 7000 Series switches support up to 4094 VLANs, which are organized into ranges:

- **VLAN 1**: The default VLAN cannot be modified or deleted.
- **VLAN 2–1005**: Normal VLANs that can be created, used, modified, and deleted.
- **VLAN 1006–4094**: Extended VLANs that can be created, named, and used. The state of these VLANs is always active, and the VLAN is always enabled and cannot be shut down.
- **VLAN 3968–4047 and 4094**: Allocated for internal use only.

VLANs 3968–4047 and 4094 are reserved for internal use in each VDC for features that need to use internal VLANs for their operation—for example, multicast and diagnostics.

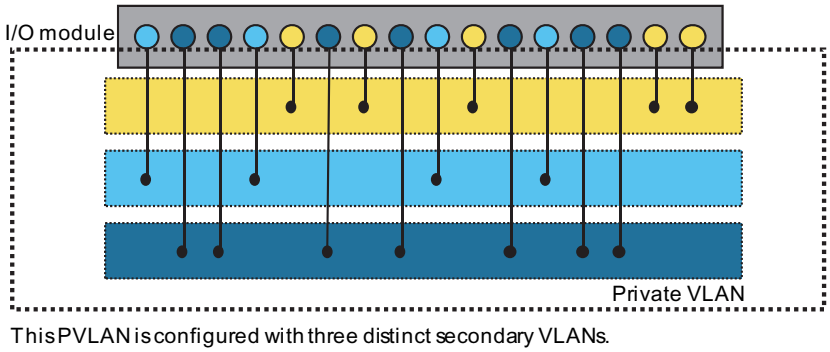
Due to the use of VDCs, a VLAN number can be reused in different VDCs because each VDC is a separate virtual device. The maximum number of VLANs in all VDCs is 16,000.

VLAN Trunking Protocol (VTP) is supported in Cisco NX-OS Release 5.1(1) and later. Supported VTP features include VTP v1/2 in the server, client, transparent, and off modes, as well as VTP pruning.

https://t.me/learningnets

Private VLANs

- Private VLANs can be used to implement Layer 2 isolation within a single VLAN and associated subnet.
- The primary VLAN represents the VLAN and associated subnet to the rest of the network.
- Secondary VLANs isolate hosts within the VLAN.



This PVLAN is configured with three distinct secondary VLANs.

© 2012 Cisco and/or its affiliates. All rights reserved.

Deploying private VLANs (PVLANs) in an enterprise data center environment provides an effective means of sparing IP address space and controlling Layer 2 access to servers and devices residing within the server farm. The Layer 2 isolation that is provided by PVLANs is an excellent way to supplement Layer 3 security that is already used to protect a particular server farm subnet.

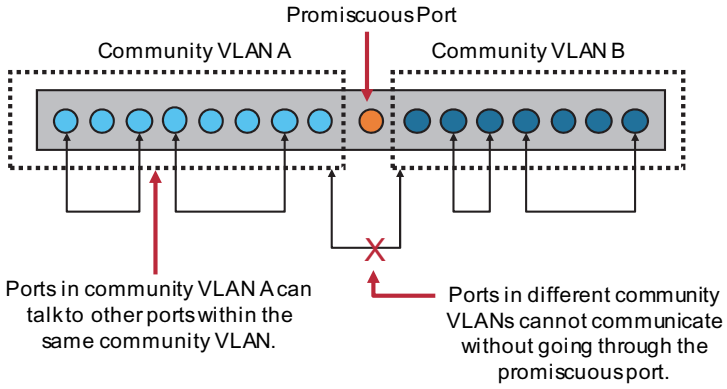
Two major benefits of deploying PVLANs are conserving IP address space and providing isolation for servers residing in the same subnet. Two VLAN concepts that are associated with PVLAN configuration are primary and secondary VLANs. Secondary VLANs consist of isolated VLANs and community VLANs. Servers residing in an isolated VLAN can only communicate through the primary VLAN are isolated at Layer 2 from any other servers that are configured in the same or any other isolated VLANs. Servers that are part of a community VLAN can communicate at Layer 2 with all other servers residing in the same community VLAN. However, they must still communicate with other devices or servers through the primary VLAN.

Any servers or applications that communicate using Layer 2 protocols such as multicast should be placed in the same community VLAN. As previously stated, all traffic to and from the isolated and community VLANs is first forwarded through the primary VLAN. Each primary VLAN is associated with a promiscuous port. Therefore, each isolated and community VLAN must be mapped to a primary VLAN. A promiscuous port can be configured either as a standard promiscuous port, which is the PVLAN equivalent of an access port, or as a promiscuous trunk port.

https://t.me/learningnets

Community VLAN

- Secondary community VLANs can be used to create subgroups within the primary VLAN.
- There is Layer 2 connectivity within each community VLAN, but not between community VLANs.



© 2012 Cisco and/or its affiliates. All rights reserved.

In cases where similar systems do not need to interact directly, PVLANs provide additional protection at a Layer 2 level. PVLANs are an association of primary and secondary VLANs.

A primary VLAN defines the broadcast domain to which secondary VLANs are associated. The secondary VLANs can be either isolated VLANs or community VLANs. Hosts on isolated VLANs communicate only with the associated promiscuous ports in a primary VLAN, while hosts on community VLANs communicate among themselves and with the associated promiscuous ports.

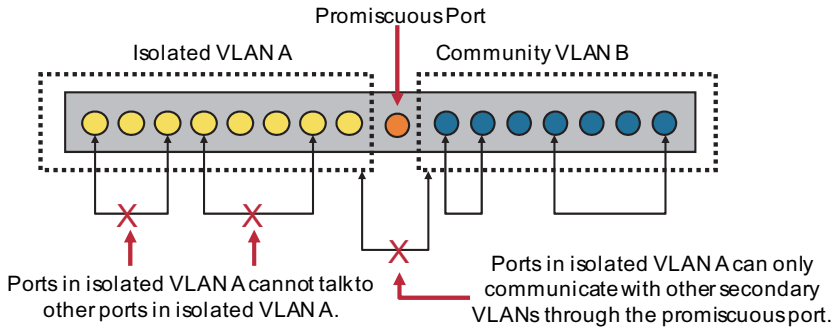
To use PVLANs, the private VLAN feature must first be enabled. After there are operational ports in a PVLAN, that feature cannot then be disabled. The private VLAN feature permits partitioning of a Layer 2 broadcast domain on a VLAN into subdomains while still using the same Layer 3 subnet.

Community VLANs are ports within a community VLAN that can communicate with each other but cannot communicate with ports in other community VLANs or any isolated VLANs at the Layer 2 level. A PVLAN host port is either a community PVLAN port or an isolated PVLAN port, depending on the type of secondary VLAN with which it is associated.

https://t.me/learningnets

Isolated VLAN

- An isolated secondary VLAN creates a subgroup within the primary VLAN in which all hosts are isolated from each other at Layer 2.



A PVLAN can only have one isolated VLAN.

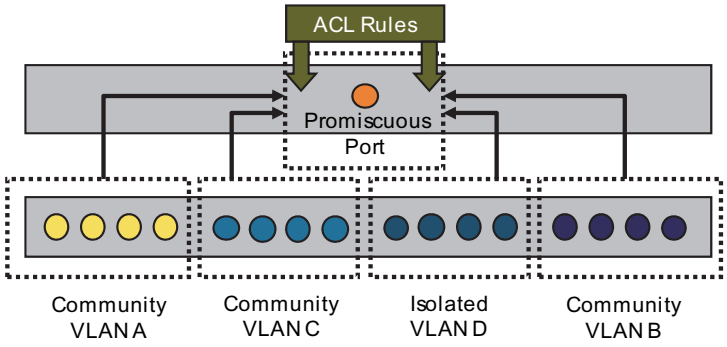
© 2012 Cisco and/or its affiliates. All rights reserved.

A secondary isolated VLAN creates a subgroup within the primary VLAN that isolates hosts from each other within that secondary VLAN. Ports within an isolated VLAN cannot communicate with each other at a Layer 2 level. Any port that is associated with the isolated VLAN has complete Layer 2 isolation from other ports within the same PVLAN domain, except that it can communicate with associated promiscuous ports. PVLANs block all traffic to isolated ports except traffic from promiscuous ports.

https://t.me/learningnets

Promiscuous Port

- Promiscuous ports provide outside connectivity for the secondary VLANs.
- Traffic from a promiscuous port is sent to all ports in the associated secondary VLANs, and traffic from all ports in the secondary VLANs is sent to the promiscuous port.



© 2012 Cisco and/or its affiliates. All rights reserved.

Promiscuous ports belong to the primary VLAN. The promiscuous port can communicate with all ports, including the community and isolated host ports that belong to the secondary VLANs associated with the promiscuous port and ports that are associated with the primary VLAN. Within a primary VLAN, there can be several promiscuous ports. Each promiscuous port can have several secondary VLANs, or no secondary VLANs, associated with that port. A secondary VLAN can be associated with more than one promiscuous port as long as the promiscuous port and secondary VLANs are within the same primary VLAN. This option might be used for load-balancing or redundancy purposes. If you have secondary VLANs that are not associated with any promiscuous port, these secondary VLANs cannot communicate with the outside world.

PVLANs only control Layer 2 connectivity within the VLAN. ACLs can be used to control the traffic that passes between these VLANs at Layer 3.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Configuring Private VLANs

1. Enable the PVLAN feature
2. Create primary VLAN
3. Create secondary VLANs of the appropriate types
4. Associate secondary PVLANs with the primary PVLAN

```
switch(config)# feature private-vlan                    1
switch(config)# vlan 142
switch(config-vlan)# private-vlan primary               2

switch(config-vlan)# vlan 100-102
switch(config-vlan)# private-vlan community
                                                        3
switch(config-vlan)# vlan 103
switch(config-vlan)# private-vlan isolated

switch(config-vlan)# vlan 142
switch(config-vlan)# private-vlan association 100-103   4

switch(config-vlan)# show vlan private-vlan
Primary  Secondary  Type            Ports
-------  ---------  --------------  -------------
 42      100        community
142      101        community
142      102        community
142      103        isolated
```

You can add or remove associated VLANs by using add and remove keywords

© 2012 Cisco and/or its affiliates. All rights reserved.

When configuring a PVLAN, the private VLAN feature must first be enabled. You can then start using the commands to configure primary and secondary VLANs.

To configure a VLAN as a primary VLAN, first create the VLAN, and then configure it as a primary VLAN.

Next, the secondary VLANs must be created and designated as secondary PVLANs. A secondary PVLAN must be configured either as type **isolated** or type **community**.

To configure a range of VLANs as secondary PVLANs, use the **vlan** *vlan-range* command.

The secondary PVLANs must be associated with the primary PVLAN in VLAN configuration mode.

Use the following guidelines when associating secondary VLANs with a primary VLAN:

- The *secondary_vlan_list* parameter can contain multiple community VLAN IDs.
- The *secondary_vlan_list* parameter can contain multiple isolated VLAN IDs, although it is common to have only a single isolated VLAN.
- Enter a *secondary_vlan_list* value or use the **add** keyword with a *secondary_vlan_list* value to associate secondary VLANs with a primary VLAN.
- Use the **remove** keyword with a *secondary_vlan_list* value to clear associations between secondary VLANs and a primary VLAN.

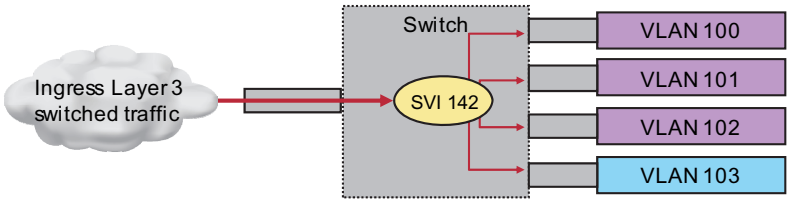
https://t.me/learningnets

Associating Layer 3 Interface

- Promiscuous ports can provide outside connectivity for Layer 2 switched traffic.
- To provide outside connectivity for Layer 3 switched traffic from the PVLAN, associate the secondary VLANs with the SVI for the primary VLAN.

```
switch(config)# feature interface-vlan
switch(config)# interface vlan 142
switch(config-if)# private-vlan mapping 100-103
```

You can add or remove associated VLANs by using add and remove keywords



© 2012 Cisco and/or its affiliates. All rights reserved.

Promiscuous ports or promiscuous trunk ports provide Layer 2 switched connectivity to the outside world, either to a network device—such as a switch, router, or firewall—or to a specific host, such as a backup server. When PVLANs are implemented on a Layer 3 switch, such as the Cisco Nexus 7000 Series switch, it is also possible to provide Layer 3 switched connectivity to the rest of the network via an SVI. To allow the secondary VLANs to use the SVI for the primary VLAN as a Layer 3 gateway to other subnets, it is necessary to associate the secondary VLANs with the SVI for the primary VLAN.

Consider the following guidelines when mapping secondary VLANs to the Layer 3 VLAN interface of a primary VLAN:

- The **private-vlan mapping** command only affects PVLAN ingress traffic that is Layer 3 switched.
- Enter a *secondary_vlan_list* parameter, or use the **add** keyword with a *secondary_vlan_list* parameter to map the secondary VLANs to the primary VLAN.
- Use the **remove** keyword with a *secondary_vlan_list* parameter to clear the mapping between secondary VLANs and the primary VLAN.

The example shows how to permit routing of secondary VLAN ingress traffic from PVLANs 100–103 and VLAN 142.

https://t.me/learningnets

Configuring Layer 2 Host Port in PVLAN

- Configure a Layer 2 port as a member of a community or isolated VLAN.

```
switch(config)# interface ethernet 2/3
switch(config-if)# switchport
switch(config-if)# switchport mode private-vlan host

switch(config-if)# switchport private-vlan host-association 142 101

switch(config-if)# show interface ethernet 2/3 switchport

Name: Ethernet 2/3
Switchport: Enabled
Administrative Mode: private-vlan host
Operational Mode: up
Administrative Trunking Encapsulation: negotiate
Negotiation of Trunking: Off
Access Mode VLAN: 1 (default)
Trunking Native Mode VLAN: 1 (default)
Voice VLAN: none
Administrative private-vlan host-association: 142 (VLAN0142) 101 (VLAN0101)
Administrative private-vlan mapping: none
Operational private-vlan: none
```

SVI ID

Community VLAN

© 2012 Cisco and/or its affiliates. All rights reserved.

To configure a Layer 2 port as a host port, use the **switchport mode private-vlan host** command. To associate the port with a primary and secondary VLAN, use the **switchport private-vlan host association** command. Whether this port is a community port or an isolated port is determined by the PVLAN type of the secondary VLAN that is assigned to the port.

This figure shows how to configure interface Ethernet 2/3 as a Layer 2 host port in a PVLAN.

https://t.me/learningnets

Configuring Layer 2 Port as Promiscuous Port

- Configure a Layer 2 port as a promiscuous port.

```
N7K-1(config)# interface ethernet 2/4
N7K-1(config-if)# switchport
N7K-1(config-if)# switchport mode private-vlan promiscuous
N7K-1(config-if)# switchport private-vlan mapping 142 100-103


N7K-1(config-if)# show interface ethernet 2/4 switchport

Name: Ethernet 2/4
Switchport: Enabled
Administrative Mode: promiscuous
Operational Mode: up
Administrative Trunking Encapsulation: negotiate
Negotiation of Trunking: Off
Access Mode VLAN: 1 (default)
Trunking Native Mode VLAN: 1 (default)
Voice VLAN: none
Administrative private-vlan host-association: none
Administrative private-vlan mapping: 142 (VLAN0142) 100 (VLAN0100) 101
Operational private-vlan: none
```

© 2012 Cisco and/or its affiliates. All rights reserved.

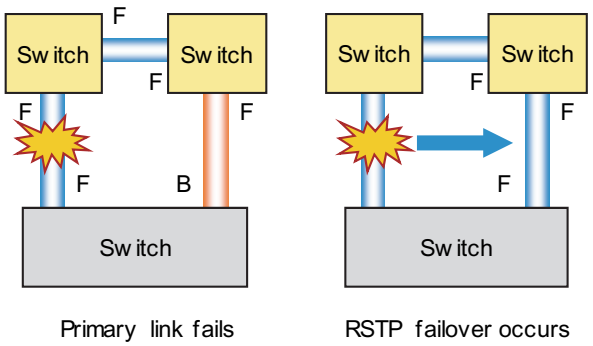
To configure a port as a promiscuous port, use the **switchport mode private-vlan host** command. To map the primary and secondary VLANs to the promiscuous port, use the **switchport private-vlan mapping** command.

The figure shows how to configure interface Ethernet 2/4 as a promiscuous port in a PVLAN.

https://t.me/learningnets

Rapid Per VLAN Spanning Tree Plus (Rapid PVST+)

- Cisco Nexus switches run the Rapid Per VLAN Spanning Tree Plus (Rapid PVST+) protocol by default for all VLANs.
- Rapid PVST+ uses a separate instance of the 802.1w RSTP protocol for each VLAN.



© 2012 Cisco and/or its affiliates. All rights reserved.

Spanning Tree Protocol (STP) 802.1D was designed at a time when recovering within a minute after an outage was considered adequate. However, with the advent of Layer 3 switching in LAN environments, bridging and switching methods are now competing with routed solutions such as Open Shortest Path First (OSPF) and Enhanced Interior Gateway Routing Protocol (EIGRP) to provide alternate paths more quickly than was previously possible.

Cisco has enhanced the original 802.1D specification with extensions such as UplinkFast, BackboneFast, and PortFast to accelerate the convergence time of a bridged network. The disadvantage of these solutions is that they are proprietary solutions and require additional configuration to tune their performance.

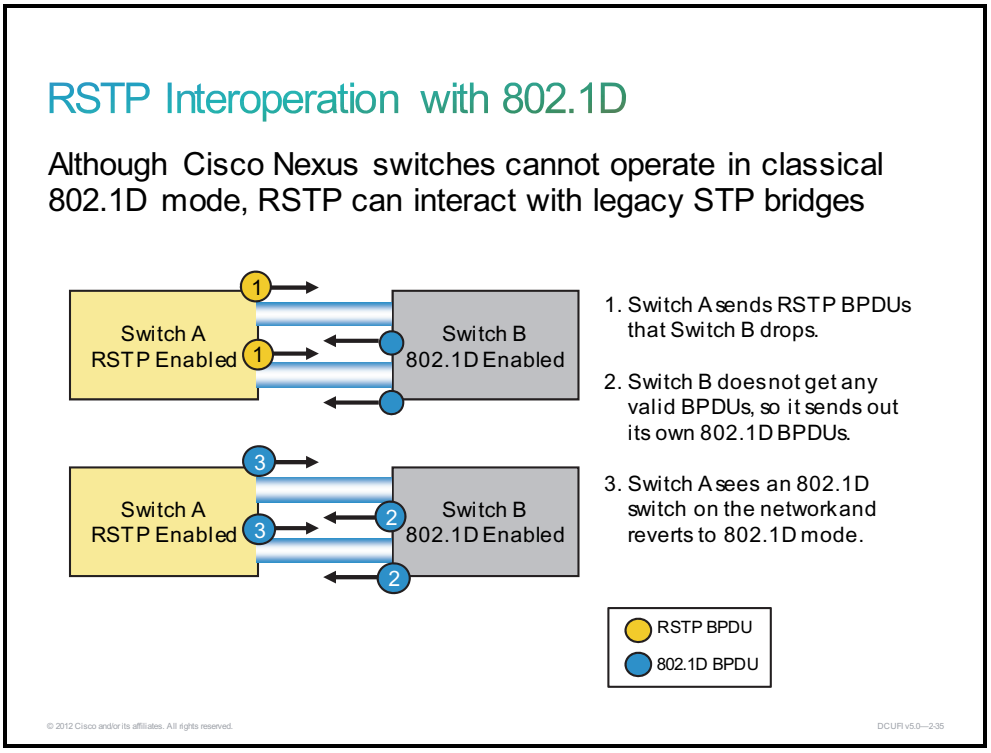
Rapid Spanning Tree Protocol (RSTP) IEEE 802.1w represents an evolution of the 802.1D standard. The 802.1D terminology remains basically unchanged in 802.1w, as do most parameters, thereby making it easier for users to configure the new protocol. In most cases, RSTP performs better than Cisco proprietary extensions without necessary additional configuration. RSTP 802.1w is also capable of reverting to 802.1D in order to interoperate with legacy bridges on a per-port basis. Reversion for legacy bridges loses the convergence benefits that were introduced by 802.1w.

Per VLAN Spanning Tree Plus (PVST+) allows the definition of one spanning-tree instance per VLAN. Normal PVST+ relies on the use of the older 802.1D STP to reconverge the STP domain in the case of link failures. Rapid Per VLAN Spanning Tree (Rapid PVST) allows the use of 802.1w with Cisco PVST in order to provide a much faster convergence per VLAN.

With Rapid Per VLAN Spanning Tree Plus (Rapid PVST+), each STP instance uses the 802.1w algorithm to reconverge the network following link failure.

Note	Within a VDC, you can run either Rapid PVST+ or Multiple Spanning Tree (MST) but not both simultaneously.

https://t.me/learningnets



Although Cisco Nexus switches cannot run in classical 802.1D mode, RSTP can interoperate with legacy STP protocols. However, the fast convergence benefits of RSTP are lost when interacting with legacy bridges.

Each port maintains a variable defining the protocol or mode in order to run on a corresponding segment. A migration delay timer of three seconds is also started when the port comes up. When this timer is running, the current mode (STP or RSTP) that is associated with the port is locked. After the migration delay expires, the port adopts the mode of the next bridge protocol data unit (BPDU) that it receives. If the port changes its operating mode as a result of receiving a BPDU, the migration delay is restarted to limit the frequency of possible mode changes.

Legacy STP bridges ignore RSTP BPDUs and drop them. The legacy STP bridge assumes that there are no other bridges on the segment and starts sending out inferior 802.1D-format BPDUs. Upon receiving these legacy BPDUs, RSTP bridges wait for twice the hello time before changing to 802.1D mode on that port only. As a result, the legacy 802.1D bridge begins to receive BPDUs that it can understand.

Note If the legacy STP bridge is removed from the segment, the RSTP bridge continues to run legacy STP on that port. This situation occurs because the RSTP bridge has no way of knowing that the legacy bridge has been removed from the segment. Manual intervention is required to restore the ability of a port to detect the current protocol.

When a port is in legacy 802.1D mode, it is also able to process topology change notification (TCN) BPDUs and BPDUs with the topology change (TC) or topology change acknowledgment (TCA) bit set.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Configuring Rapid PVST+

A Rapid PVST+ instance is automatically created when a VLAN is configured.

```
switch(config)# vlan 10
switch(config-vlan)# name Sales

switch(config-vlan)# show spanning-tree vlan 10 brief

VLAN0010
  Spanning tree enabled protocol rstp
  Root ID     Priority    32778
              Address     0026.9804.a942
              This bridge is the root
              Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

  Bridge ID   Priority    32778  (priority 32768 sys-id-ext 10)
              Address     0026.9804.a942
              Hello Time  2  sec  Max Age 20 sec  Forward Delay 15 sec

Interface        Role Sts Cost      Prio.Nbr Type
---------------- ---- --- --------- -------- --------------------------
Eth1/1           Desg FWD 2         128.129  P2p
Eth1/3           Desg FWD 2         128.131  P2p
Eth1/17          Desg FWD 2         128.145  P2p
```

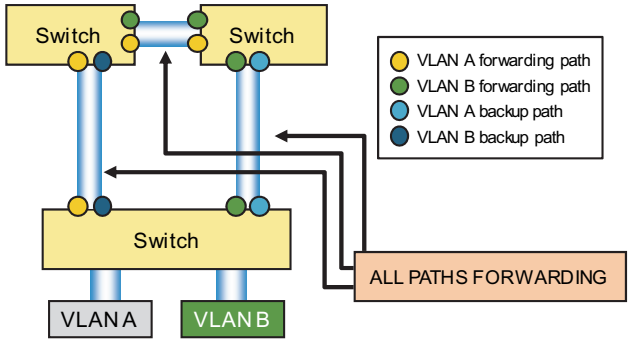
© 2012 Cisco and/or its affiliates. All rights reserved.

The figure shows the configuration steps to create and name a VLAN. The **show** command output displays that the spanning-tree version is RSTP. The output also displays the root ID, bridge ID, and port state for each VLAN. RSTP is the default spanning-tree version for Cisco NX-OS Software.

https://t.me/learningnets

Multiple Spanning Tree Protocol

- MST allows VLANs to be load-balanced across different spanning-tree topologies.
- MST calculates the spanning-tree topology for a group of VLANS rather than per VLAN, making it more scalable than Rapid PVST+.



© 2012 Cisco and/or its affiliates. All rights reserved.

The problem with running a single instance of STP is that any blocked link is unable to actively participate in the forwarding of data. The blocked link then becomes a wasted resource that is used for redundancy purposes only. Rapid PVST+ solves this issue by running a separate spanning-tree instance for each VLAN.

MST is defined in 802.1s and is designed to support multiple instances of spanning tree over VLAN trunks. MST permits the mapping of multiple VLANs into a single spanning-tree instance, with each instance supporting a spanning-tree topology independent of other spanning-tree instances. This architecture provides multiple forwarding paths for data traffic and enables load balancing while simultaneously reducing the number of spanning-tree instances required to support many VLANs. MST further improves the fault tolerance of the network, as a failure in one instance or forwarding path does not affect other instances or forwarding paths.

MST uses the RSTP mechanisms for each instance to provide rapid spanning-tree convergence through explicit handshaking, thereby eliminating the 802.1D forwarding delay while quickly transitioning root bridge ports and designated ports to the forwarding state.

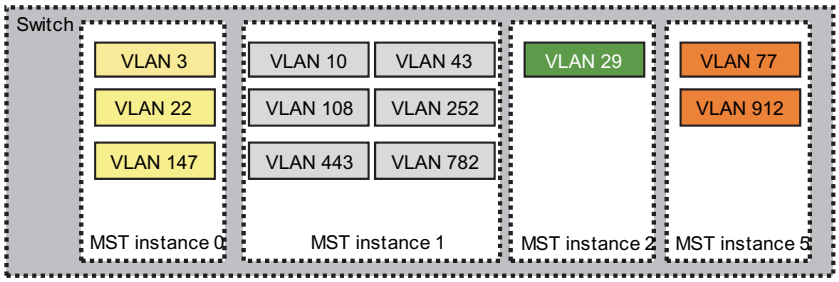
MST improves spanning-tree operation and maintains backward compatibility with the following:

- The original 802.1D STP
- Existing Cisco proprietary Multi-Instance STP (MISTP)
- Existing Cisco PVST+
- Rapid PVST+

https://t.me/learningnets

MST Regions and Instances

- MST region is a collection of interconnected switches with the same MST configuration.
- The following should be the same for all switches in an MST region:
  - Region name
  - Revision number
  - VLAN-to-instance mappings



© 2012 Cisco and/or its affiliates. All rights reserved.

For switches to participate in MST instances, their MST configuration information must be consistent. A collection of interconnected switches with the same MST configuration constitutes an MST region.

The configuration includes the name of the region, the revision number, and the VLAN-to-MST instance assignment mapping.

A region can have one or multiple members with the same MST configuration. Each member must be capable of processing 802.1w BPDUs. There is no limit to the number of MST regions in a network.

Note	Although multiple MST regions can interact with each other, it is not recommended to partition the network into many regions.

Each device can support up to 65 MST instances (MSTIs)—including Instance 0—in a single MST region. Instances are identified by any number in the range from 1 to 4094. The system reserves Instance 0 for a special instance, which is the Internal Spanning Tree (IST). By default, all VLANs are assigned to this instance. You can assign a VLAN to only one MST instance at a time.

The MST region appears as a single bridge to adjacent MST regions and to other Rapid PVST+ regions and 802.1D STPs.

https://t.me/learningnets

MST Configuration

1. Configure MST region parameters.
2. Exit from the configuration context to make changes effective
3. Change the spanning-tree mode to MST.

```
N7K-1(config)# spanning-tree mst configuration
N7K-1(config-mst)# name MST-DC-1                    1
N7K-1(config-mst)# revision 37
N7K-1(config-mst)# instance 1 vlan 100-199
N7K-1(config-mst)# instance 2 vlan 200-299

N7K-1(config-mst)# exit    2

N7K-1(config)# spanning-tree mode mst    3
```

© 2012 Cisco and/or its affiliates. All rights reserved.

This figure describes the proper configuration steps to set the MST configuration parameters and then to enable MST as the active spanning-tree mode.

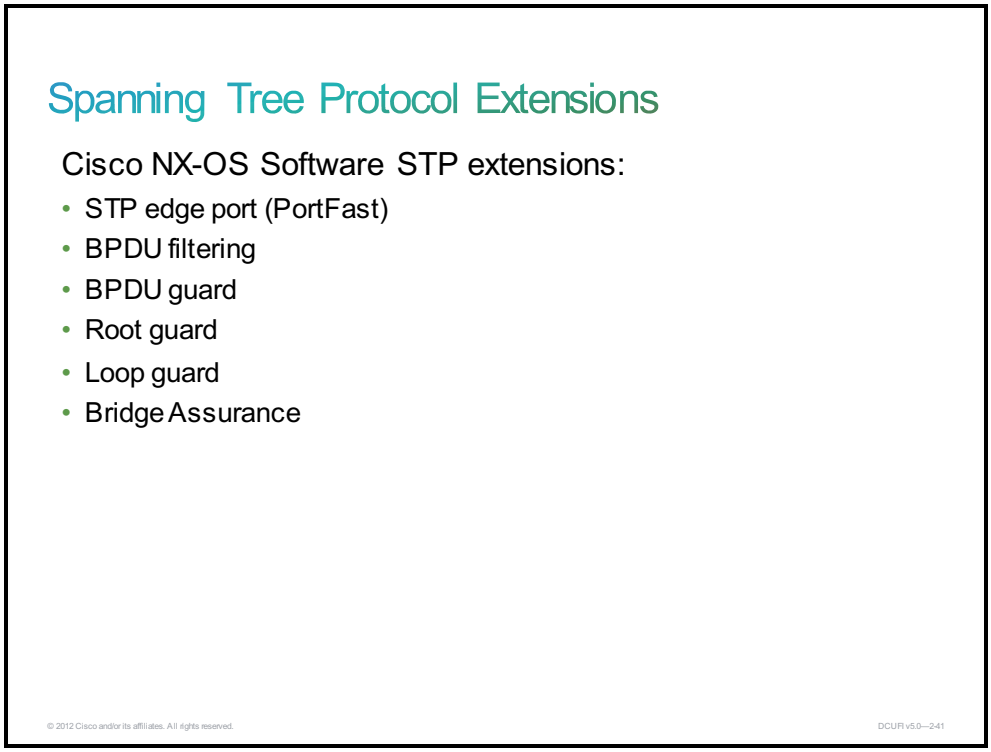
Note	Changes that are made in spanning-tree MST configuration mode are not applied until the **exit** command is issued. To exit MST configuration mode without applying the changes, use the **abort** command.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

STP Extensions

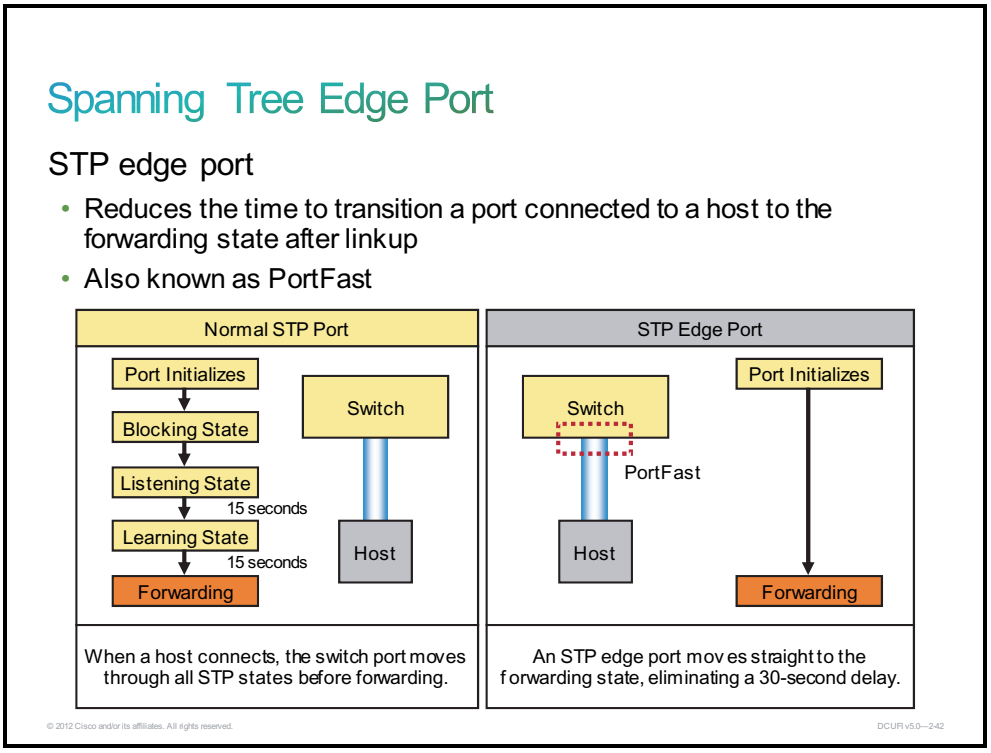
This topic identifies how to use and configure the STP extensions on the Cisco Nexus switches.



Cisco has added extensions to STP that enhance loop prevention, protect against user configuration errors, and provide better control over the protocol parameters.

The available extensions are spanning-tree edge ports (previously known as PortFast), BPDU filtering, BPDU guard, loop guard, root guard, and Bridge Assurance. All of these extensions can be used with both Rapid PVST+ and MST. Many of these features can be applied either globally or on specified interfaces.

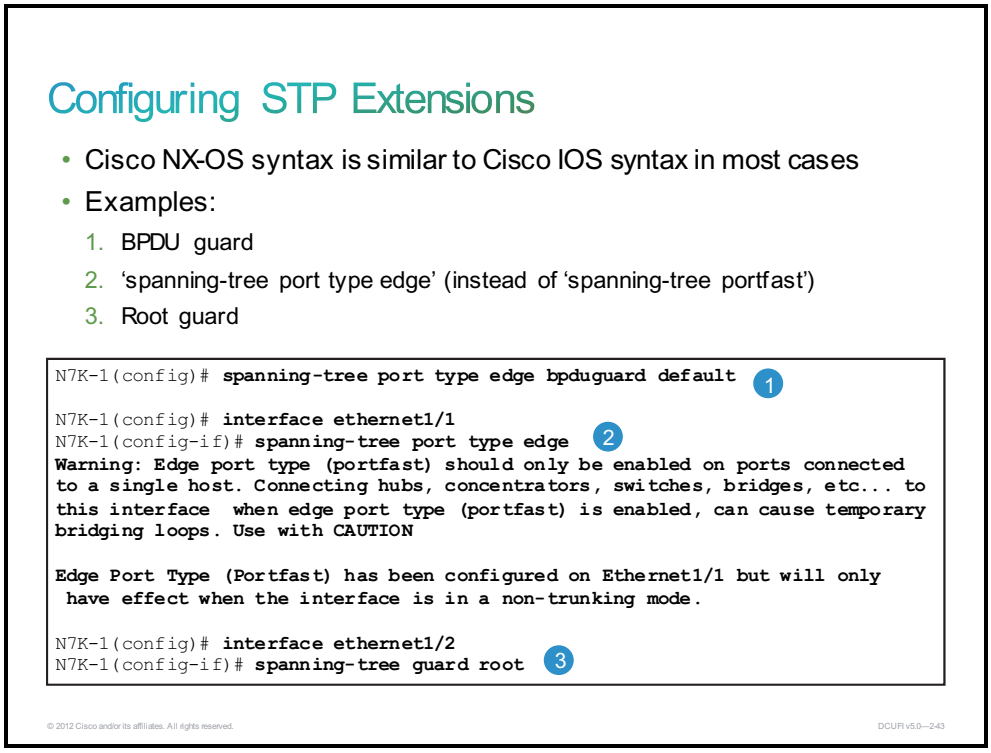
https://t.me/learningnets



Configuring a Layer 2 access port as a spanning-tree edge port causes the port to bypass the listening and learning states and enter the forwarding state immediately. This feature was formerly known as PortFast, but the name was changed to spanning-tree edge port in order to conform to the RSTP standard naming convention for this feature. Spanning-tree edge ports are typically deployed on Layer 2 access ports that are connected to a single workstation or server. This design allows those devices to connect to the network immediately without waiting for STP convergence to take place. Interfaces that are connected to a single workstation or server are not expected to receive BPDUs, and it should be safe to transition these ports to the forwarding state. When configured as a spanning-tree edge port, a port is still running STP. A spanning-tree edge port can immediately transition to the blocking state if necessary—for example, upon receipt of a BPDU.

Note Spanning-tree edge port configuration is used to minimize the time that access ports must wait for STP convergence to occur and, therefore, should only be used on access ports. If you enable the spanning-tree edge port feature on a port that is connected to a switch, you might inadvertently create a temporary bridging loop.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



The syntax that is used to configure the spanning-tree extensions in Cisco NX-OS Software that runs on Cisco Nexus switches is very similar to the syntax that is used in the Cisco IOS Software that runs on Cisco Catalyst switches.

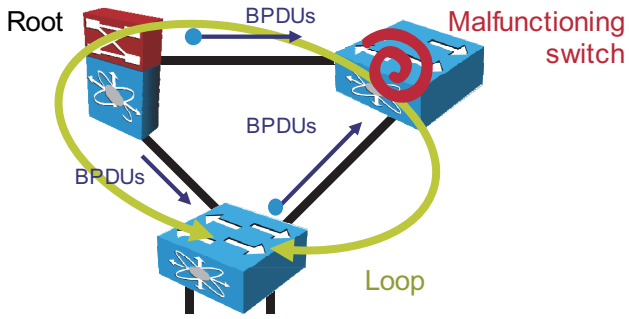
The most important exception is the spanning-tree edge port feature, which was formerly known as PortFast. This change in naming is reflected in the command syntax. The Cisco NX-OS syntax to enable this feature is **spanning-tree port type edge**, while the Cisco IOS syntax to enable this feature is **spanning-tree portfast**.

The figure shows an example configuration that enables the BPDU guard feature for all spanning-tree edge ports, configures interface Ethernet 1/1 as a spanning-tree edge port, and enables the root guard feature on interface Ethernet 1/2.

https://t.me/learningnets

Why Bridge Assurance?

- Normally, STP BPDUs flow from the root to the leaves of the tree.
- When a non-designated, non-root port stops receiving BPDUs, it will become designated and transition to the forwarding state.
- A switch might stop sending BPDUs due to a control plane failure condition while the data plane is still active.
- This can cause a bridging loop.



© 2012 Cisco and/or its affiliates. All rights reserved.

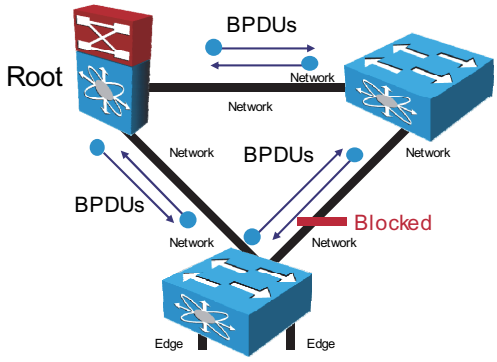
The figure shows a normal STP topology and normal STP behavior, including a root bridge. A malfunctioning switch that stopped sending any BPDUs (shown at the upper right of the graphic) could cause the neighboring switches to move a blocking port to non-blocking. In this situation, the malfunctioning switch can create a bridging loop in the network.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Bridge Assurance

- Bridge Assurance prevents bridging loops caused by STP failures.
- Bridge Assurance alters the behavior of STP.
- BPDUs are sent on all ports that have Bridge Assurance enabled.
- BPDUs are used as a hello protocol to detect protocol failure.



© 2012 Cisco and/or its affiliates. All rights reserved.

Bridge Assurance is used to protect against certain problems that can cause bridging loops in the network. Specifically, Bridge Assurance can be used to protect against unidirectional link failure or other software failure. Bridge Assurance can also be used to protect against situations where a device continues to forward data traffic when it is no longer running STP.

Note	Bridge Assurance is supported only by Rapid PVST+ and MST.

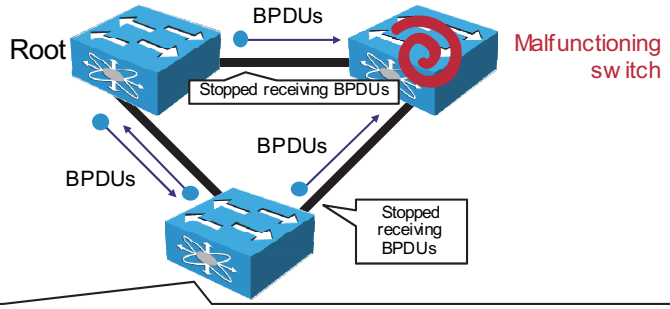
Bridge Assurance is enabled by default in Cisco NX-OS and can only be disabled globally. Bridge Assurance can only be enabled on point-to-point STP links: Both ends of the link must be enabled for Bridge Assurance. If they are not, the adjacent port is blocked.

When Bridge Assurance is enabled, BPDUs are sent on all operational network ports, including alternative and backup ports, for each hello time period. If the port does not receive a BPDU for a specified period, then the port moves into the blocking state and cannot be used for the root port calculation. After it receives a BPDU, it resumes normal spanning-tree operation.

https://t.me/learningnets

Bridge Assurance Operation

- When a port that has Bridge Assurance enabled stops receiving BPDUs, it will mark the port as inconsistent instead of moving to forwarding
  - No bridging loop occurs
  - The function of Bridge Assurance is similar to loop guard
    - Only use one of the two mechanisms on the same port



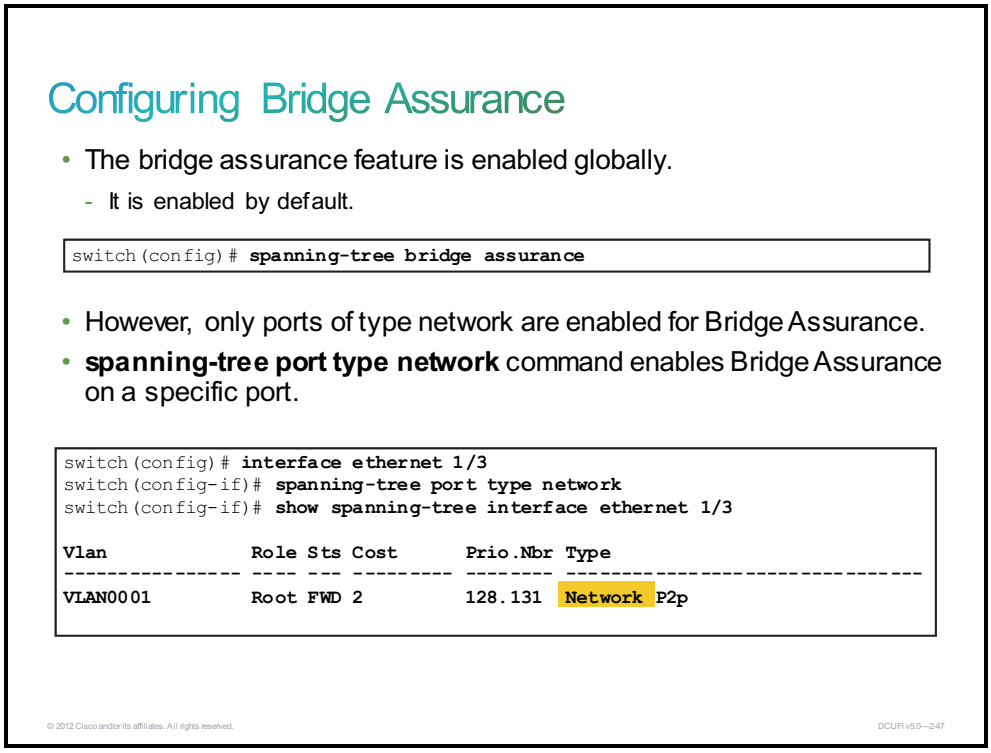

```
%STP-2-BRIDGE_ASSURANCE_BLOCK: Bridge Assurance blocking port Ethernet2/48 VLAN0700.
switch# show spanning-tree vlan 700 | include -i bkn
Eth2/48          Altn BKN*4         128.304  Network P2p *BA_Inc
```

© 2012 Cisco and/or its affiliates. All rights reserved.

With Bridge Assurance enabled, even a malfunctioning switch in the network does not create a bridging loop. When the potential loop is identified, Bridge Assurance puts the port into a Bridge Assurance inconsistent state.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



Bridge Assurance needs to be enabled globally in order to enable it at the interface level. However, it is already enabled globally by default. To disable Bridge Assurance, you can use the **no spanning-tree bridge assurance** command.

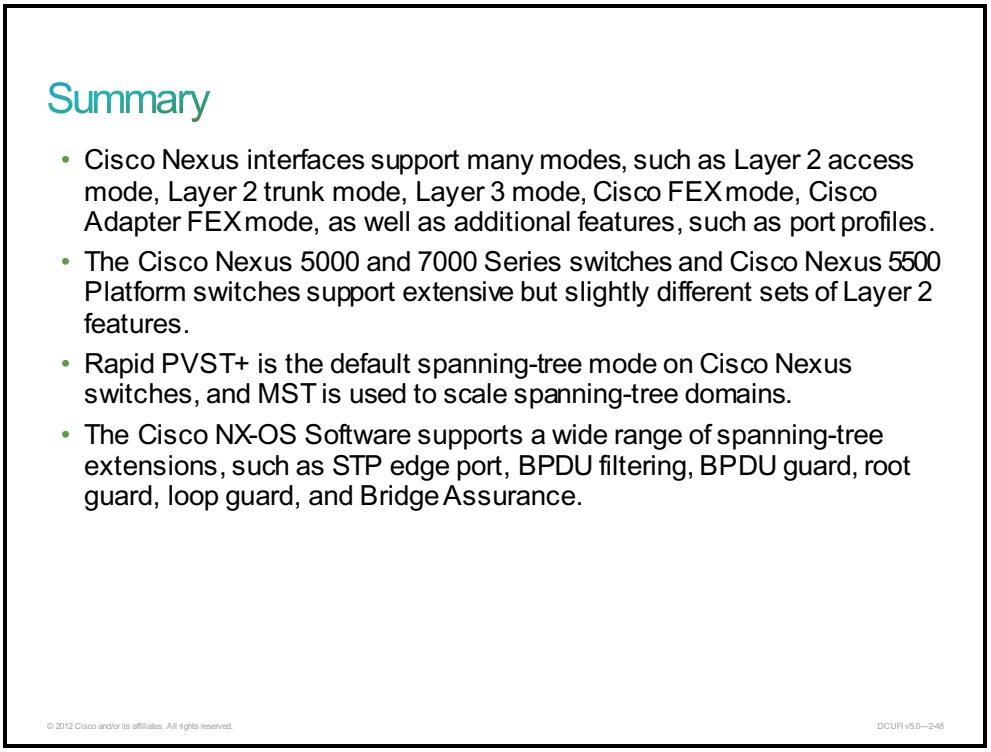
When Bridge Assurance is enabled globally, all interfaces included in spanning-tree will have BA enabled. However, the default port type of an interface is **normal**. To enable Bridge Assurance on an interface, use the **spanning-tree port type network** command.

Note	The **spanning-tree port type network** command should always be configured on both sides of a link to prevent a port from going into blocking because it is not receiving BPDUs from the neighbor.

https://t.me/learningnets

Summary

This topic summarizes the key points that were discussed in this lesson.



References

For additional information, refer to these resources:

- To learn more about configuring Cisco Nexus 2000 FEX, refer to *Cisco Nexus 2000 Series Fabric Extender Software Configuration Guide* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/nexus2000/sw/configuration/guide/rel_6_0/b_Configuring_the_Cisco_Nexus_2000_Series_Fabric_Extender_rel_6_0.html
- To learn more about configuring Cisco Adapter FEX, refer to *Cisco Nexus 5000 Series NX-OS Adapter-FEX Software Configuration Guide, Release 5.1(3)N1(1)* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/nexus5000/sw/adapter-fex/513_n1_1/b_Configuring_Cisco_Nexus_5000_Series_Adapter-FEX_rel_5_1_3_N1.html

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 4

Configuring PortChannels

Overview

Cisco PortChannel is one of the core technologies that are used in Ethernet-based networks. Cisco PortChannel is used to bundle multiple physical links into a single logical link, which improves resiliency and optimizes bandwidth utilization on the links.

A limitation of regular port channel is that it only allows the aggregation of links between two devices. The virtual port channel (vPC) technology that is used by the Cisco Nexus 5000 and 7000 Series switches enables Multichassis EtherChannels (MECs) to be formed between a network device and two separate physical chassis. vPC technology allows logical loop-free Layer 2 topologies to be created, which prevents Spanning Tree Protocol (STP) from blocking any of the ports in the network topology. This type of design combines high availability with increased bandwidth between the access and aggregation layers.

Cisco Nexus 2000 Fabric Extender (FEX) technology can be deployed together with Cisco Nexus 5000 or Cisco Nexus 7000 Series switches in order to create a data center network that combines the advantages of a top-of-rack (ToR) design with the advantages of an end-of-row (EoR) design.

Enhanced vPC combines two vPC topologies: hosts dual-homed to two FEXs and FEXs dual-homed to two Nexus 5500 Switches.

Objectives

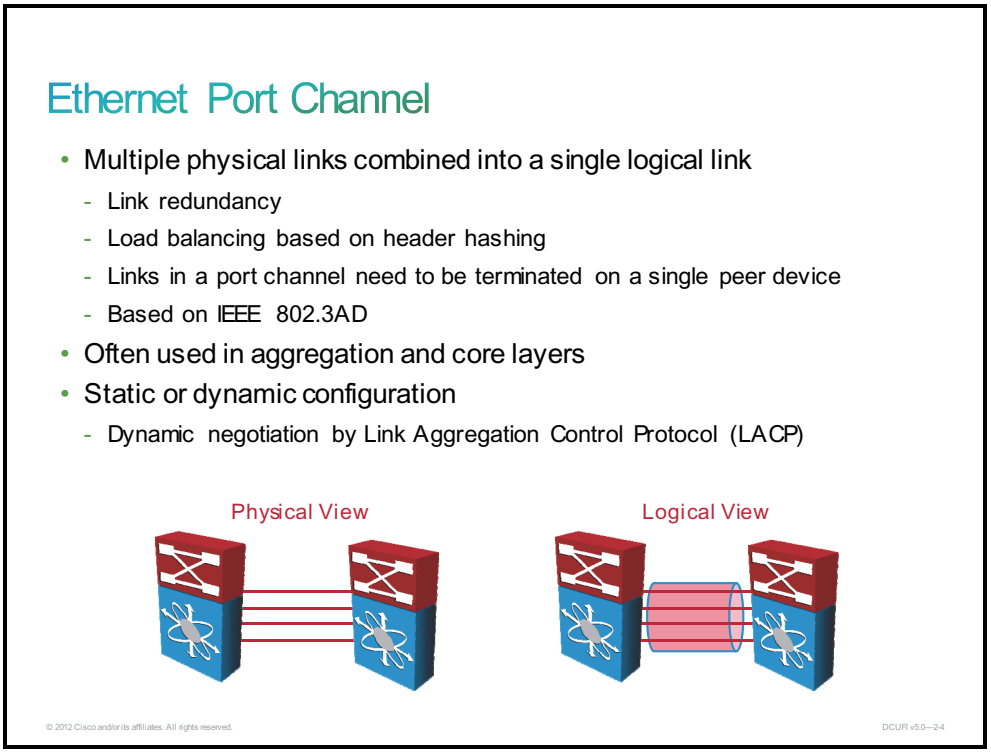
Upon completing this lesson, you will be able to evaluate how port channels and vPCs should be used to improve the solution and then configure the features. You will be able to meet these objectives:

- Identify where port channels and vPCs could be used to improve reliability
- Identify how to configure port channels on the Cisco Nexus switches
- Identify the architecture and components of vPCs
- Explain how to configure vPCs on the Cisco Nexus switches
- Explain how to configure the Cisco Nexus 2000 Series FEX connected to a Cisco Nexus 5000 or 7000 Series switch
- Explain how to configure the Enhanced vPCs on a Cisco Nexus 5000 Series switch

https://t.me/learningnets

Using Port Channels and vPCs

This topic identifies where port channels and vPCs could be used to improve reliability.



PortChannel is one of the core technologies that are used in Ethernet-based networks. To add resiliency against link failures and to increase the available bandwidth between two devices, multiple physical links can be provisioned between the devices. However, without PortChannel, control plane protocols, such as STP, or routing protocols will treat the links as individual links. In the case of STP, the result is blocked ports. Although the additional links add resiliency, the available bandwidth between the two devices is not increased.

PortChannel technology combines the physical links into a single logical link, which is called a *port channel*. Control plane protocols, such as STP and routing protocols, treat the port channel as a single link. Spanning tree will not block the links that are part of the port channel, and routing protocols only form a single routing adjacency across the port channel.

Traffic that is switched or routed to a port channel interface is balanced across the individual physical links through a hashing mechanism. The hashing mechanism uses a selection of the fields in the packet headers as input. This process ensures that packets with the same header will be forwarded on the same physical link to prevent packet reordering.

A port channel can either be defined statically or negotiated dynamically by using Link Aggregation Control Protocol (LACP). Cisco Nexus Operating System (NX-OS) Software performs a compatibility check when adding ports to a port channel so as to ensure that the port can participate in the port channel aggregation. Therefore, it is important that all physical ports that participate in a port channel are configured identically. LACP, which is described in the 802.1AX standard, can be used to dynamically negotiate the aggregation of multiple links into a port channel and to detect failure conditions.

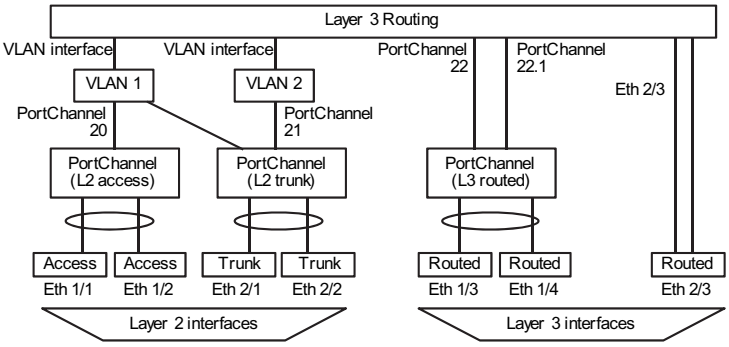
A major restriction of classical PortChannel technology is that it is inherently limited to the aggregation of a number of links that run between the same two devices.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Port Channel Layer 2 and Layer 3 Interfaces

- Layer 2 port channels in access or trunk mode
- Layer 3 port channel interfaces
  - May have subinterfaces
  - May have a static MAC address configured
    - Otherwise, the MAC of the first channel member to come up



© 2012 Cisco and/or its affiliates. All rights reserved.

You can classify port channel interfaces as Layer 2 interfaces or, in the case of the Cisco Nexus 5500 Platform and 7000 Series switches, Layer 3 interfaces. In addition, you can configure Layer 2 port channels in either access or trunk mode. Layer 3 port channel interfaces have routed ports as channel members and may have subinterfaces.

You can configure a Layer 3 port channel with a static MAC address. If you do not configure this value, the Layer 3 port channel then uses the router MAC of the first channel member to come up.

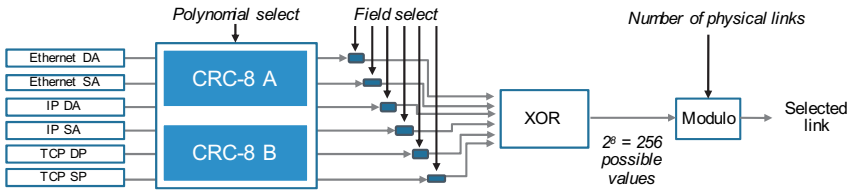
On the Cisco Nexus 7000 Series switches, all ports in a port channel must be in the same virtual device context (VDC).

https://t.me/learningnets

Port Channel Load Balancing

Load-Balancing Methods
Destination MAC address
Source MAC address
Source and destination MAC address
Destination IP address
Source IP address
Source and destination IP address
Destination TCP/UDP port number
Source TCP/UDP port number
Source and destination TCP/UDP port number

Platform	Maximum links in PC
Nexus 5000/5500	16 active links
Nexus 7000 F Series modules	16 active links
Nexus 7000 M Series modules	8 active and 8 standby links (Cisco NX-OS Release 5.1 and later)



© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco Nexus switches support the bundling of up to 16 ports into a port channel. The maximum number of ports in a channel depends on the exact switch hardware and software combination. On the M1-Series modules on the Cisco Nexus 7000 Series switches, the maximum is eight active links per port channel. Beginning with Cisco NX-OS Release 5.1, you can bundle up to 16 active ports simultaneously into a port channel on the F1 series modules on the Cisco Nexus 7000 Series switch. On the Cisco Nexus 5000 Series switches, you can bundle up to 16 active links into a port channel.

The Cisco Nexus switch load-balances all traffic that is switched or routed to a port channel interface across all operational individual physical links by hashing the various header fields in a frame into a numerical value that selects one of the links in the channel. This process ensures that packets with the same header will be forwarded on the same physical link in order to prevent packet reordering. The load-balancing mechanism is performed in the hardware and enabled by default. The load-balancing method can either be applied to all port channels on a specified module (Cisco Nexus 7000 Series switch) or to the entire switch (Cisco Nexus 5000 and 7000 Series switches and Cisco Nexus 5500 Platform switch). If a per-module load-balancing method is configured, it takes precedence over the switchwide setting.

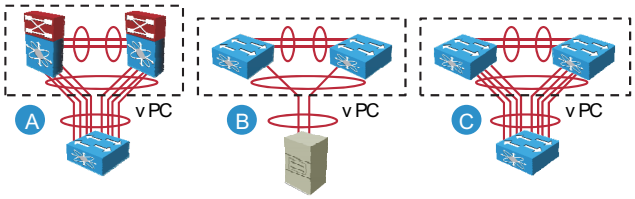
You can configure the switch to use one of the following load-balancing methods:

- Destination MAC address
- Source MAC address
- Source and destination MAC addresses
- Destination IP address
- Source IP address
- Source and destination IP addresses
- Source TCP or UDP port number
- Destination TCP or UDP port number
- Source and destination TCP or UDP port numbers

https://t.me/learningnets

Virtual Port Channel (vPC)

- Port channel extension
  - Port channels terminated on different physical devices
  - Resiliency against device failures
  - Multiple physical switches appear as single logical switch to the peer device
- Loop-free logical topologies with full physical redundancy
- Use cases:
  - A. Dual-uplink Layer 2 access
  - B. Server dual-homing
  - C. Active-active Fabric Extenders (FEX)



© 2012 Cisco and/or its affiliates. All rights reserved.

With the increased use of virtualization technologies in data centers, and even across data center locations, organizations are shifting from a highly scalable Layer 3 network model to a highly scalable Layer 2 model. This shift is causing changes in the technologies that are used to manage large Layer 2 network environments. These changes include migration away from STP as a primary loop-management technology and toward new technologies, such as vPCs.

The biggest limitation of classic PortChannel is that the port channel operates only between two devices. In large networks, the support of multiple devices together is often a design requirement to provide some form of hardware failure alternate path. This alternate path is often connected in a way that would cause a loop, thereby limiting the benefits that are gained with port channel technology to a single path. To address this limitation, the Cisco NX-OS Software platform provides a technology called *vPC*. Although a pair of switches acting as a vPC peer endpoint looks like a single logical entity to the port channel-attached devices, the two devices that act as the logical port channel endpoint are still two separate devices.

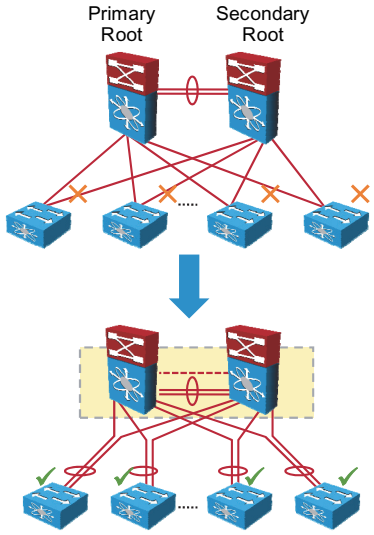
The vPC solution combines the benefits of hardware redundancy with the benefits of port channel loop management. The three main use cases of the vPC technology are as follows:

- **Dual-uplink Layer 2 access**: In this scenario, an access switch such as a Cisco Nexus 5000 Series switch is dual-homed to a pair of distribution switches, such as Cisco Nexus 7000 Series switches.
- **Server dual-homing**: In this case, a server is connected via two interfaces to two separate access switches.
- **Active-active Fabric Extenders**: In this topology, a Cisco Nexus 2000 Fabric Extender is dual-homed to a pair of Nexus switches.

https://t.me/learningnets

vPC Enhancement vs. STP

- Without vPC
  - STP blocks redundant uplinks
  - VLAN-based load balancing
  - Loop resolution relies on STP
  - Protocol failure can cause complete network meltdown
- With vPC
  - No blocked uplinks
  - Lower oversubscription
  - Hash-based EtherChannel load balancing
  - Loop-free topology
  - STP is used in case of keepalive and VPC peer link simultaneous failure



© 2012 Cisco and/or its affiliates. All rights reserved.

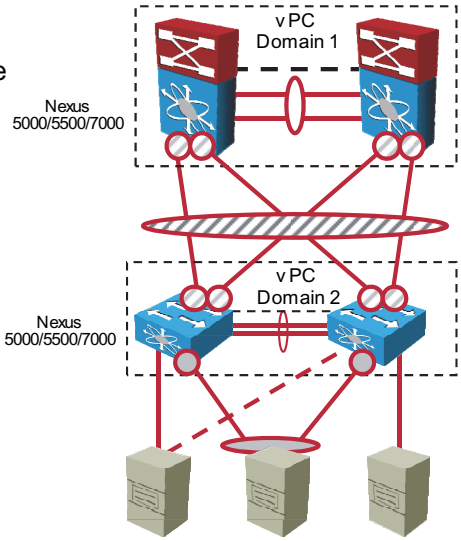
In early Layer 2 Ethernet network environments, it was necessary to develop protocol and control mechanisms that limited the disastrous effects of a topology loop in the network. STP was the primary solution to this problem, providing a loop detection and loop management for Layer 2 Ethernet networks. This protocol has gone through a number of enhancements and extensions. While STP scales to very large network environments, it still has one suboptimal principle: To break loops in a network, only one active path is allowed from one device to another. This principle is true regardless of how many actual connections might exist in the network.

The other main benefit of migration to an entirely port channel-based loop-management mechanism is that link recovery is potentially much faster. STP can recover from a link failure in approximately six seconds, while an entirely port channel-based solution has the potential for failure recovery in less than one second.

https://t.me/learningnets

vPCs at Multiple Layers

- vPC supported on Cisco Nexus 5000/5500/7000 switches
- vPC can be deployed in multiple layers of the data center simultaneously
  - Server to access
  - Access to aggregation
- Separate vPC configured at each level
- Known as dual-sided vPC



© 2012 Cisco and/or its affiliates. All rights reserved.

vPC is supported on Cisco Nexus 5000 and 7000 Series switches as well as Cisco Nexus 5500 Platform switches. The benefits that are provided by the vPC technology apply to any Layer 2 switched domain. Therefore, vPC is commonly deployed in both the aggregation and access layers of the data center.

vPC can be used to create a loop-free logical topology between the access and aggregation layer switches, which increases the bisectional bandwidth and improves network stability and convergence. vPC can also be used between servers and the access layer switches in order to enable server dual-homing with dual-active connections.

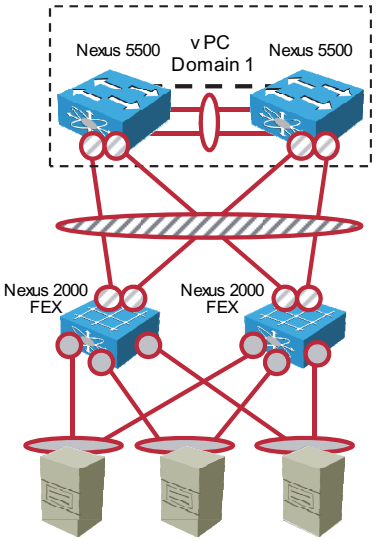
When the switches in the access and aggregation layers both support vPC, a unique vPC can be created at each layer. To implement this environment, you need to configure two separate vPC domains at the access and distribution layers. The access layer would typically consist of Cisco Nexus 5000 Series switches or Cisco Nexus 5500 Platform switches and the distribution layer of the Cisco Nexus 5500 Platform switch or the Cisco Nexus 7000 Series switch.

This scenario is commonly referred to as “dual-sided vPC.”

https://t.me/learningnets

vPC with Multiple Dual Homing

- Combination:
  - Dual-homed connection of a host to two FEXs
  - Dual-homed connection of an FEX to two switches
- Single vPC domain
  - Active-active setup at each layer
- Support on Nexus 5500
- Known as extended vPC



© 2012 Cisco and/or its affiliates. All rights reserved.

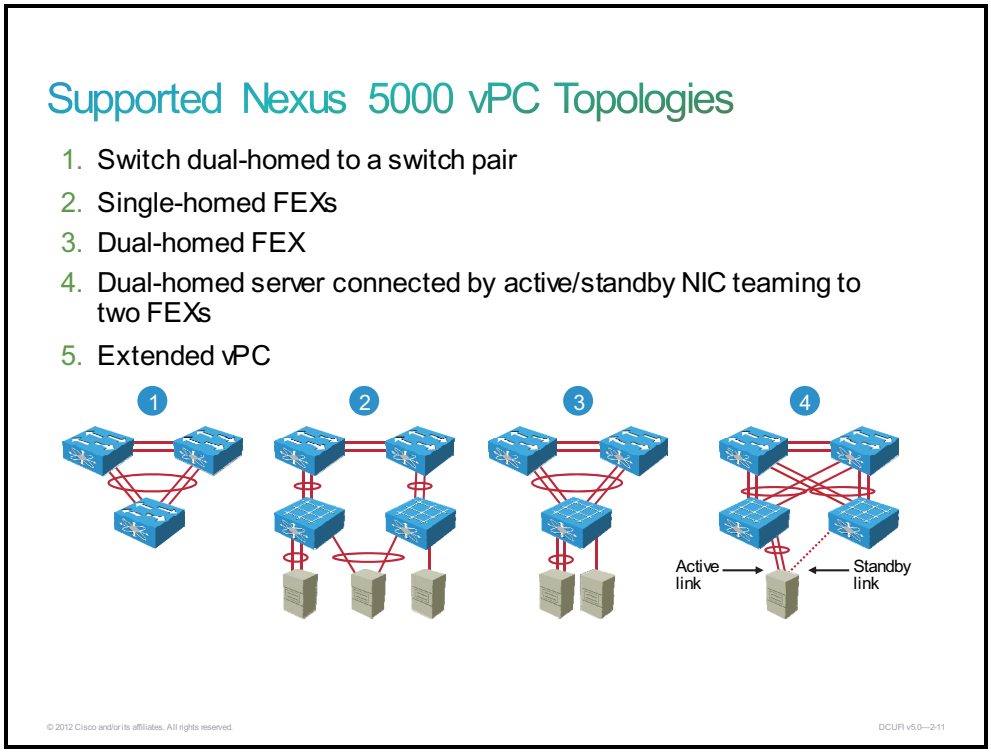
This figure illustrates another vPC solution, which is called extended vPC. This solution combines two layers of dual homing:

- Dual-homed connection of a host to two FEXs
- Dual-homed connection of an FEX to two switches

All links are active in the system. Extended vPC requires the configuration of only a single vPC domain, which is defined on the parent switch. Extended vPC is supported on Cisco Nexus 5500 Platform switches and any Cisco Nexus 2000 Fabric Extender.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



There are several supported vPC topologies with Cisco Nexus 5000 Series switches and Cisco Nexus 5500 Platform switches:

1. **Switch dual-homed to a switch pair**: This topology allows you to connect a pair of Cisco Nexus 5000 Series switches or a pair of Cisco Nexus 5500 Platform switches in a vPC directly to another switch or to a server. vPC peer switches must be of the same type. For example, you can connect a pair of Cisco Nexus 5000 Series switches or a pair of Cisco Nexus 5500 Platform switches, but you cannot connect a Cisco Nexus 5000 Series switch to a Cisco Nexus 5500 Platform switch in a vPC topology. Up to eight interfaces could be connected to each Cisco Nexus 5000 Series switch, providing 16 interfaces bundled for the vPC pair.

2. **Single-homed FEXs**: In this topology, you connect a server with dual or quad or more network adapters that are configured in a vPC to a pair of FEXs that are connected to the Cisco Nexus 5000 Series switches. Depending on the FEX model, you may be able to connect one or more network adapter interfaces to each fabric extender. As an example, with the Cisco Nexus 2148T Fabric Extender, the server has one link only to each fabric extender. A topology with Cisco Nexus 2248TP-E Fabric Extender or with Cisco Nexus 2232PP FEX could consist of more links from the server to a single FEX.

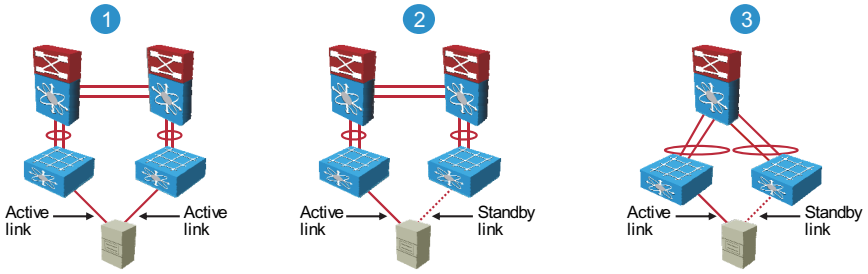
3. **Dual-homed FEX**: In this topology, you connect the FEX to two upstream Cisco Nexus 5000 Series switches and downstream to a number of single-homed servers. The topology shown in the following figure provides the vPC functionality to singly or dually connected servers.

4. **Dual-homed server connected by active/standby NIC teaming to two FEXs**: In this topology, host-side network interface card (NIC) teaming software is required to form the port channel.

https://t.me/learningnets

Supported Nexus 7000 vPC Topologies

1. Dual-attached active/active teaming host with port channel FEX uplink pinning
2. Dual-attached active/standby teaming host with port channel FEX uplink pinning
3. Dual-attached active/standby teaming host with port channel FEX uplink pinning to a single switch



Fewer options due to the redundant Cisco Nexus 7000 architecture
No need for dual-attached FEX

© 2012 Cisco and/or its affiliates. All rights reserved.

The supported Cisco Nexus 7000 Series switch vPC topologies include:

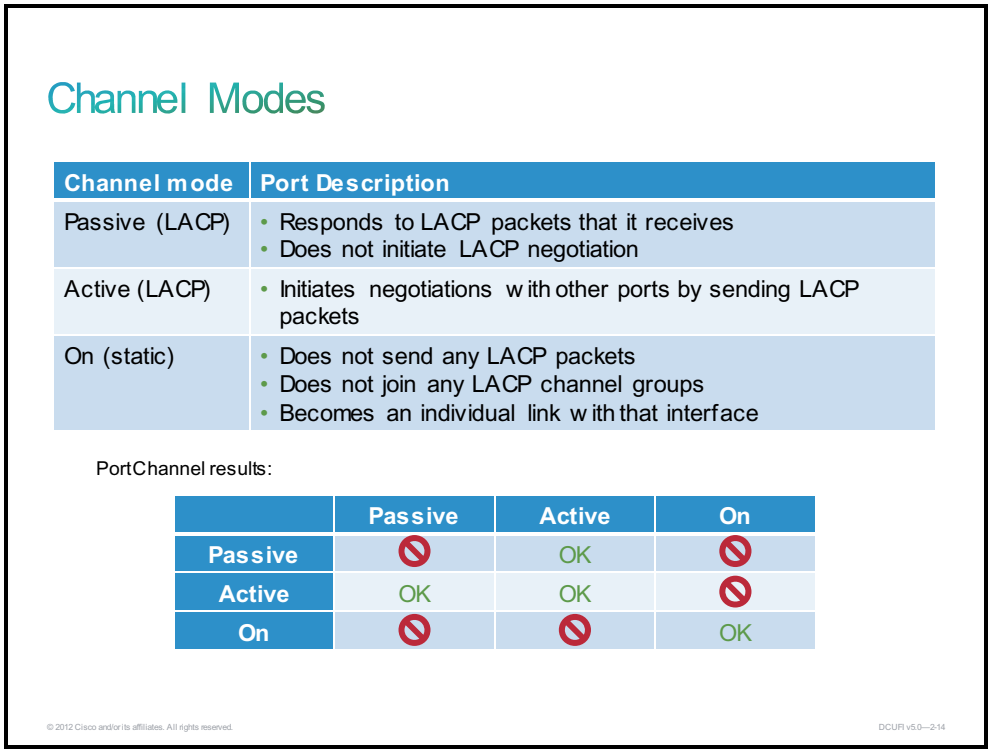
1. Dual-attached active/active teaming host with port channel FEX uplink pinning
2. Dual-attached active/standby teaming host with port channel FEX uplink pinning
3. Dual-attached active/standby teaming host with port channel FEX uplink pinning to a single switch

Compared to Cisco Nexus 5000 Series switches and Cisco Nexus 5500 Platform switches, Cisco Nexus 7000 Series switches offer fewer options. This results from the redundant architecture of the Cisco Nexus 7000 Series switch: There is no need for dual-attached FEXs.

https://t.me/learningnets

Configuring Port Channels

This topic identifies how to configure port channels on the Cisco Nexus switches.



Individual interfaces in port channels are configured with channel modes. When you run static port channels, with no protocol, the channel mode is always set to on. After you enable LACP globally on the device, you enable LACP for each channel by setting the channel mode for each interface to active or passive. You can then configure either channel mode for individual links in the LACP channel group.

The following table describes the channel modes.

Channel Mode	Description
Passive	This LACP mode places a port into a passive negotiating state, in which the port responds to LACP packets that it receives but does not initiate LACP negotiation.
Active	This LACP mode places a port into an active negotiating state, in which the port initiates negotiations with other ports by sending LACP packets.
On	All static port channels—that is, those that are not running LACP—remain in this mode. If you attempt to change the channel mode to active or passive before enabling LACP, the device returns an error message. You enable LACP on each channel by configuring the interface in that channel for the channel mode as either active or passive. When LACP attempts to negotiate with an interface in the “on” state, it does not receive any LACP packets and becomes an individual link with that interface. It does not join the LACP channel group.

Both the passive and active modes allow LACP to negotiate between ports in order to determine if they can form a port channel. This is based on criteria such as the port speed and the trunking state. The passive mode is useful when you do not know whether the remote system, or partner, supports LACP.

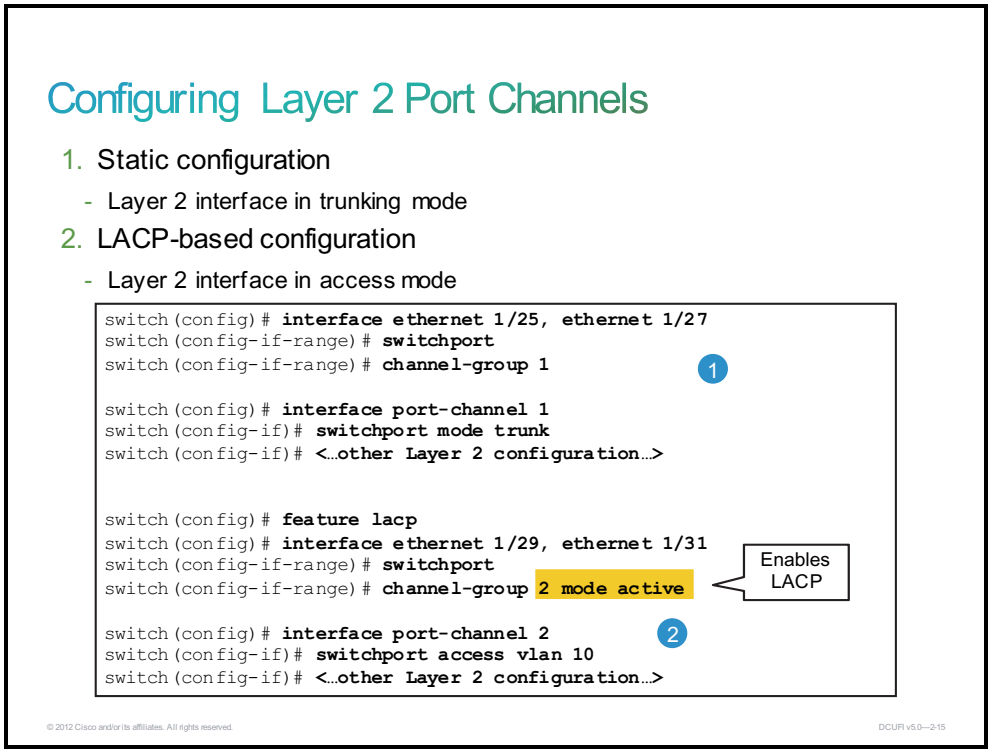
https://t.me/learningnets

Ports can form an LACP port channel when they are in different LACP modes as long as the modes are compatible, as in these examples:

- A port that is in active mode can form a port channel successfully with another port that is in active mode.
- A port that is in active mode can form a port channel with another port that is in passive mode.
- A port that is in passive mode cannot form a port channel with another port that is also in passive mode because neither port will initiate negotiation.
- A port that is in “on” mode is not running LACP.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



Configuration of port channels commonly consists of two elements—configuring the physical ports and configuring the port channel interface. The physical ports need to be assigned to a channel group, which then bundles the ports together. A channel group always has an associated port channel interface, which has the same number as the channel group number. You can create the port channel interface before you assign the physical interfaces to a channel group. If you do not create the port channel interface beforehand, it is automatically created when you assign the first physical interface to the channel group.

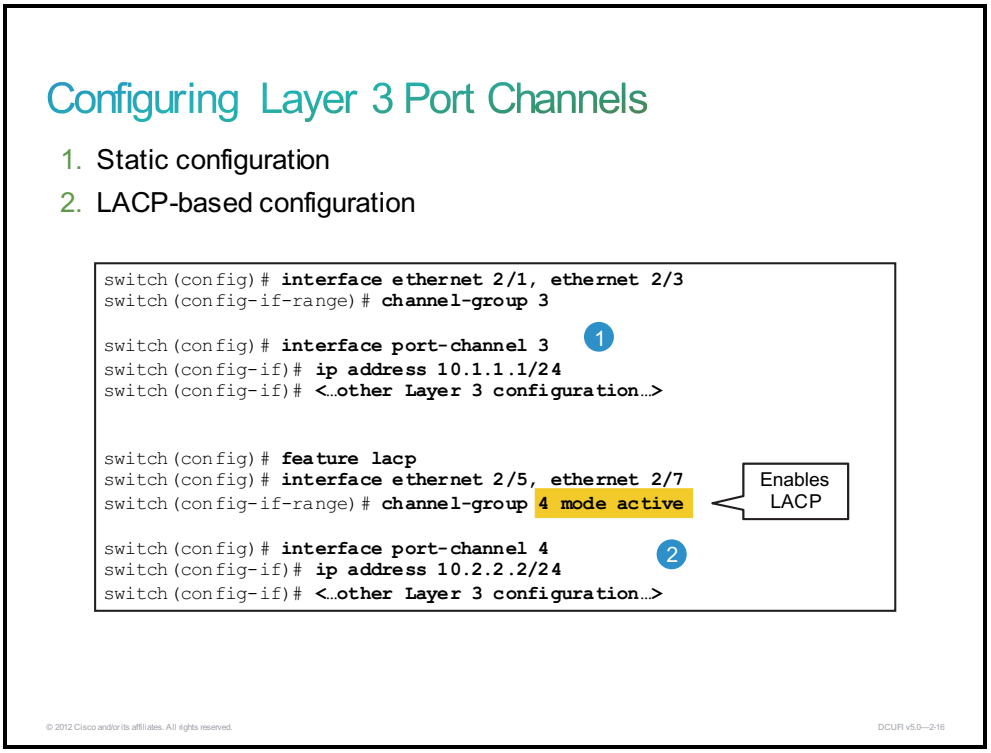
After the interfaces have been successfully bundled into a channel group with an associated port channel number, you can then configure Layer 2 or Layer 3 settings on the port channel interface. Settings that are configured on the port channel will be inherited by the physical interfaces that are members of the associated channel group.

If you create a port channel interface before assigning any interfaces to a channel group, it is important that the configuration of the port channel interface be compatible with the configuration of the physical member interfaces that you assign to the channel group.

The figure shows two examples of Layer 2 port channel configuration:

- The first example shows how to create a static port channel. In this case, the ports are configured for the “on” mode. The port channel is created statically without any negotiation.
- The second example shows how to configure a port channel that uses LACP to aggregate the links. The ports are configured for active mode and therefore initiate negotiations with other ports by sending LACP packets.

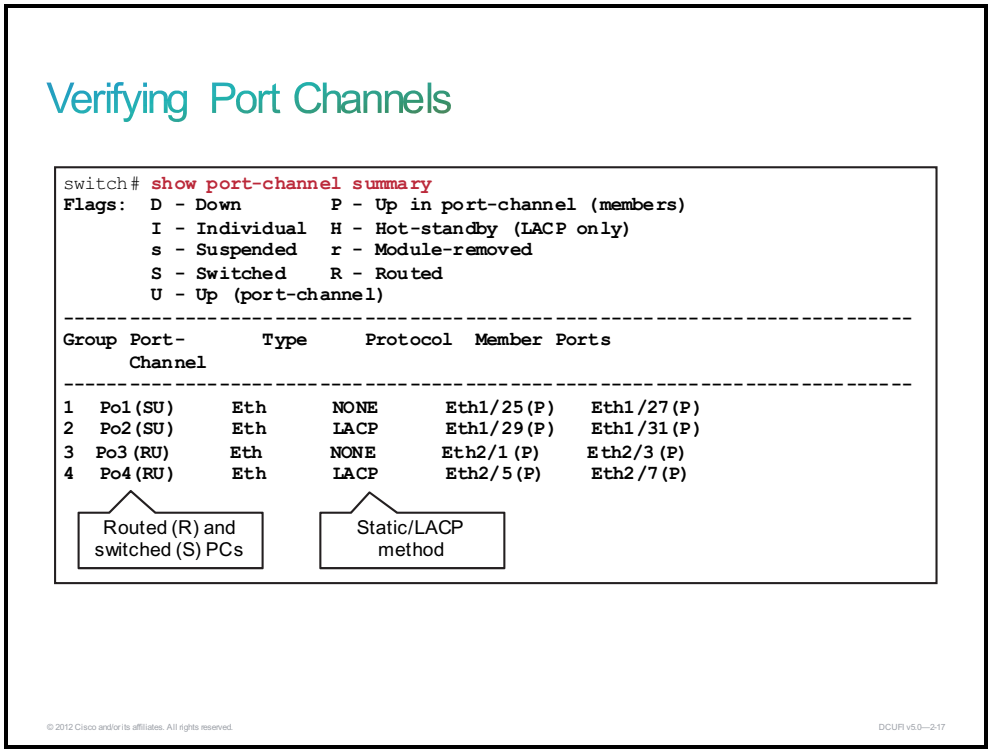
https://t.me/learningnets



The examples in the figure show how to configure Layer 3 port channels. The IP address for a port channel should always be configured on the port channel interface.

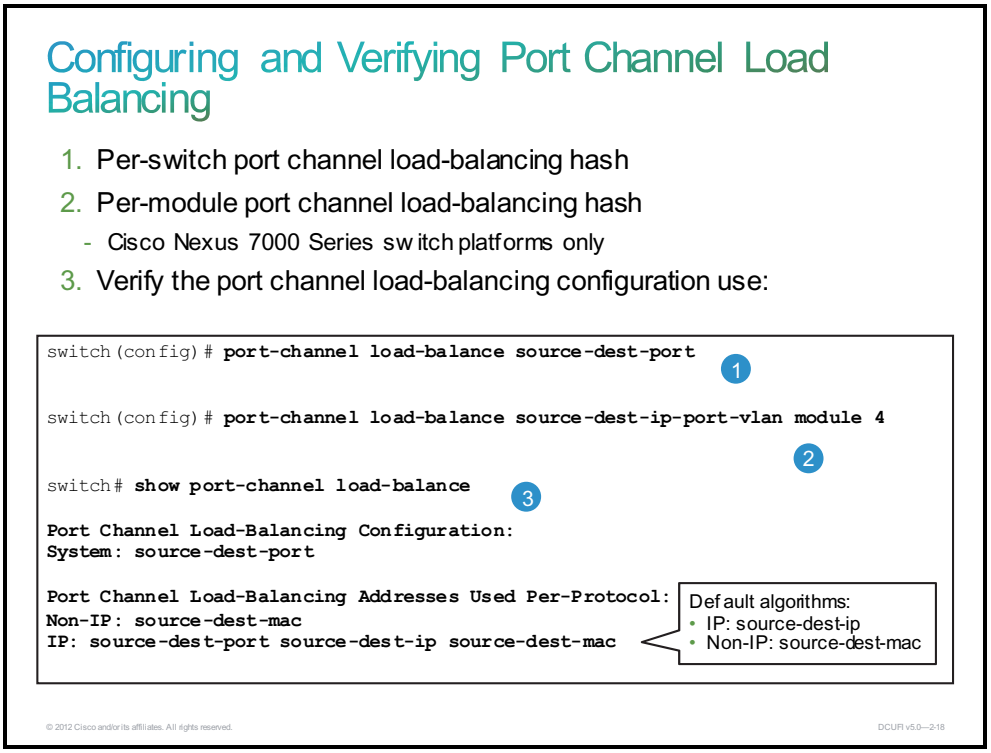
When you add an interface to a channel group, the software checks certain interface attributes to ensure that the interface is compatible with the channel group. For example, you cannot add a Layer 3 interface to a Layer 2 channel group. The Cisco NX-OS Software also checks a number of operational attributes for an interface before allowing that interface to participate in the port channel aggregation. Use the **show port-channel compatibility-parameters** command, which is described later, to see the complete list of compatibility checks that Cisco NX-OS Software uses.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



The **show port-channel summary** command can be used to verify the status of the configured port channels on the switch.

https://t.me/learningnets



To configure the port channel load-balancing options, use the **port-channel load-balance ethernet** command. The exact options for this command can vary by platform and Cisco NX-OS Software release.

The first example shows how to set the per-switch load-balancing option to include UDP and TCP port numbers. This command could be configured on a Cisco Nexus 5000 or 7000 Series switch of on a Cisco Nexus 5500 Platform switch.

The second example shows how to set the load-balancing hash on a specific module of a Cisco Nexus 7000 Series switch.

The **show port-channel load-balance** command, which is shown in the third example, can be used to verify the load-balancing hash that is currently used.

Note	Per-module PC load balancing is platform-specific. Please check the release notes or a configuration guide.

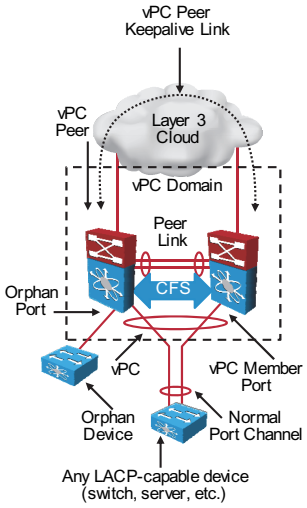
© 2012 Cisco Systems, Inc.
https://t.me/learningnets

vPC Architecture

This topic identifies the architecture and components of vPCs.

vPC Components

Component	Description
vPC	Combined port channel between the vPC peers and a vPC-capable downstream device
vPC peers	A pair of vPC-enabled switches
vPC peer link	Carries control traffic between vPC peers
Cisco Fabric Services	Protocol for state synchronization and configuration validation between vPC peers.
vPC peer keepalive link	Routed link carrying heartbeat packets for active-active detection.
vPC member port	One of the ports that forms a vPC
vPC domain	Pair of vPC peers and associated vPC components
Orphan device	Device connected to vPC peer on non-vPC link.
Orphan port	Port on a vPC peer that connects to an orphan device. Also used for a vPC member port on vPC peer that has lost connectivity to the other peer



© 2012 Cisco and/or its affiliates. All rights reserved.

A pair of Cisco Nexus switches that uses vPC present themselves to other network devices as a single logical Layer 2 switch. However, the two switches remain two separately managed switches with independent management and control planes. The vPC architecture includes modifications to the data plane of the switches in order to ensure optimal packet forwarding. vPC architecture also includes control plane components to exchange state information between the switches and allow the two switches to appear as a single logical Layer 2 switch to the downstream devices.

The vPC architecture consists of the following components:

- **vPC peers**: The core of the vPC architecture is a pair of Cisco Nexus switches. This pair of switches acts as a single logical switch, which allows other devices to connect to the two chassis using MEC.
- **vPC peer link**: The vPC peer link is the most important connectivity element in the vPC system. This link is used to create the illusion of a single control plane by forwarding bridge protocol data units (BPDUs) and LACP packets to the primary vPC switch from the secondary vPC switch. The peer link is also used to synchronize MAC address tables between the vPC peers and to synchronize Internet Group Management Protocol (IGMP) entries for IGMP snooping. The peer link provides the necessary transport for multicast traffic and for the traffic of orphaned ports. In the case of a vPC device that is also a Layer 3 switch, the peer link also carries Hot Standby Router Protocol (HSRP) packets.
- **Cisco Fabric Services**: The Cisco Fabric Services protocol is a reliable messaging protocol that is designed to support rapid stateful configuration message passing and synchronization. The vPC peers use the Cisco Fabric Services protocol to synchronize data plane information and implement necessary configuration checks. vPC peers must synchronize the Layer 2 Forwarding table between the vPC peers. This way, if one vPC peer learns a new MAC address, that MAC address is also programmed on the L2F table of

https://t.me/learningnets

the other peer device. The Cisco Fabric Services protocol travels on the peer link and does not require any configuration by the user. To help ensure that the peer link communication for Cisco Fabric Services over Ethernet is always available, spanning tree has been modified to keep the peer-link ports always forwarding. The Cisco Fabric Services over Ethernet protocol is also used to perform compatibility checks in order to validate the compatibility of vPC member ports to form the channel, to synchronize the IGMP snooping status, to monitor the status of the vPC member ports, and to synchronize the Address Resolution Protocol (ARP) table.

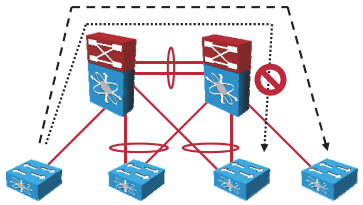
- **vPC peer keepalive link**: The peer keepalive link is a logical link that often runs over an out-of-band (OOB) network. The peer keepalive link provides a Layer 3 communications path that is used as a secondary test in order to determine whether the remote peer is operating properly. No data or synchronization traffic is sent over the vPC peer keepalive link—only IP packets that indicate that the originating switch is operating and running vPC. The peer keepalive status is used to determine the status of the vPC peer when the vPC peer link goes down. In this scenario, it helps the vPC switch to determine whether the peer link itself has failed or whether the vPC peer has failed entirely.
- **vPC**: A vPC is an MEC, a Layer 2 port channel that spans the two vPC peer switches. The downstream device that is connected on the vPC sees the vPC peer switches as a single logical switch. The downstream device does not need to support vPC itself. The downstream device then connects to the vPC peer switches using a regular port channel, which can either be statically configured or negotiated through LACP.
- **vPC domain**: The vPC domain includes both vPC peer devices, vPC peer keepalive link, vPC peer link, and all port channels in the vPC domain that are connected to the downstream devices. A numerical vPC domain ID identifies the vPC. You can have only one vPC domain ID on each device.
- **vPC member port**: This is a port on one of the vPC peers that is a member of one of the vPCs configured on the vPC peers.
- **Orphan device**: The term “orphan device” refers to any device that is connected to a vPC domain using regular links instead of connecting through a vPC.
- **Orphan port**: The term “orphan port” refers to a switch port that is connected to an orphan device. The term is also used for vPC ports whose members are all connected to a single vPC peer. This situation can occur if a device that is connected to a vPC loses all its connections to one of the vPC peers.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

vPC Data Plane for Local Traffic

- vPC peer link carries only:
  - vPC control traffic
  - Flooded traffic (broadcast, multicast, unknown unicast)
  - Traffic for orphan ports
- MAC address learning replaced with Cisco Fabric Services-based MAC address learning
  - Only for vPCs
  - Non-vPC ports use regular MAC address learning
- Frames arriving at peer switch on peer link cannot exit on vPC member port



© 2012 Cisco and/or its affiliates. All rights reserved.

vPCs are specifically designed to limit the use of the peer link to switch management traffic as well as the occasional traffic flow from a failed network port. To begin, the peer link carries vPC control traffic, such as Cisco Fabric Services over Ethernet, BPDUs, and LACP messages. In addition, the peer link carries traffic that needs to be flooded, such as broadcast, multicast, and unknown unicast traffic. The peer link also carries traffic for orphan ports.

The term “orphan port” is used for two types of ports. One type of orphan port is any Layer 2 port on a vPC peer switch that does not participate in vPC. These ports use normal switch forwarding rules, and traffic from these ports can use the vPC peer link as a transit link to reach orphan devices that are connected to the other vPC peer switch. The other type of orphan port is a port that is a member of a vPC but for which the peer switch has lost all the associated vPC member ports. When a vPC peer switch loses all member ports for a specific vPC, it forwards traffic that is destined for that vPC to the vPC peer link. In this special case, the vPC peer switch will be allowed to forward the traffic that is received on the peer link to one of the remaining active vPC member ports.

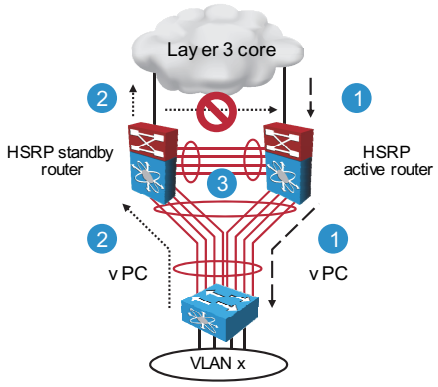
To implement the specific vPC forwarding behavior, it is necessary to synchronize the Layer 2 Forwarding tables between the vPC peer switches through Cisco Fabric Services instead of depending on the regular MAC address learning. Cisco Fabric Services-based MAC address learning applies to vPC ports only and is not used for non-vPC ports.

One of the most important forwarding rules for vPC is that a frame that enters the vPC peer switch from the peer link cannot exit the switch from a vPC member port. This principle prevents frames that are received on a vPC from being flooded back onto the same vPC by the other peer switch. The exception to this rule is traffic that is destined for an orphaned vPC member port.

https://t.me/learningnets

vPC Data Plane for External Traffic

1. Inbound traffic destined for a vPC is forwarded on a local vPC member port whenever possible
2. Outbound traffic actively forwarded by all FHRP routers whenever possible
3. Benefits:
   - Traffic avoids peer link if possible, which creates a scalable solution
   - Peer link capacity does not need to scale linearly with the number of vPCs



© 2012 Cisco and/or its affiliates. All rights reserved.

The use of the vPC bandwidth is also minimal when traffic is exchanged with external networks.

Whenever a vPC peer switch needs to forward inbound traffic for a vPC, it forwards it to a local vPC port if possible. Only if it has no active vPC member ports for the vPC does it then forward it across the vPC peer link to the other vPC peer switch.

Aggregation switches using vPCs commonly use a First Hop Redundancy Protocol (FHRP), such as HSRP, Gateway Load Balancing Protocol (GLBP), or Virtual Router Redundancy Protocol (VRRP) for default gateway redundancy. The normal forwarding behavior of these protocols has been enhanced with the peer gateway feature in order to allow them to interoperate with vPCs.

Normally, only active FHRP routers forward traffic for the virtual default gateway MAC address. For vPCs, the forwarding rules have been enhanced to allow a nonactive FHRP router to forward frames that are destined for the FHRP virtual MAC address. However, the primary FHRP device is still responsible for responding to ARP requests, even though the secondary FHRP device forwards the data traffic.

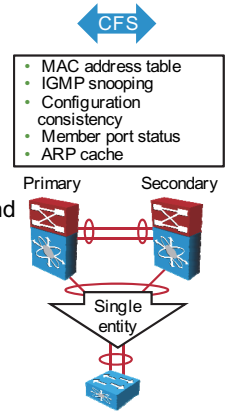
The result of the enhanced vPC forwarding behavior is that the vPC peer link does not carry vPC traffic unless a vPC has lost all its ports on one of the peer devices. Thus, there is no direct need to scale the bandwidth on the vPC peer link as you deploy more vPCs on a pair of vPC switches. However, the operation of the vPC peer link is vital to the operation of vPC. Therefore, the vPC peer link should consist of at least two dedicated 10 Gigabit Ethernet links. These two links should be terminated on different I/O modules if possible.

It is also recommended to avoid the use of orphan devices with vPC, if possible. Traffic from orphan ports may need to be forwarded across the peer link and must be taken into account when scaling peer link capacity. Also, orphan devices may experience traffic disruption in specific vPC failure scenarios.

https://t.me/learningnets

vPC Control Plane

- Cisco Fabric Services over Ethernet (FSoE) is used to synchronize vPC control plane information:
  - MAC address learning
  - IGMP snooping
  - Configuration consistency checking
  - vPC member port status
  - ARP cache (configurable)
    - Disables by default
    - Enable synchronization with **ip arp synchronize** command
- One switch is elected primary, other secondary
  - Role determines behavior during peer link failure
  - Primary switch is leading for STP on vPCs
  - Non-pre-emptive election
- Single logical entity
  - In LACP and STP
  - To neighbor devices connected on a vPC



© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco Fabric Services over Ethernet is used as the primary control plane protocol for vPC. Cisco Fabric Services over Ethernet performs several functions:

- vPC peers must synchronize the Layer 2 MAC address table between the vPC peers. If one vPC peer learns a new MAC address on a vPC, that MAC address is also programmed on the Layer 2 Forwarding table of the other peer device for that same vPC. This MAC address learning mechanism replaces the regular switch MAC address learning mechanism and prevents traffic from being forwarded across the vPC peer link unnecessarily.
- The synchronization of IGMP snooping information is performed by Cisco Fabric Services. Layer 2 Forwarding of multicast traffic with vPC is based on modified IGMP snooping behavior that synchronizes the IGMP entries between the vPC peers. In a vPC implementation, IGMP traffic entering a vPC peer switch through a vPC triggers hardware programming for the multicast entry on both vPC member devices.
- Cisco Fabric Services is also used to communicate essential configuration information to ensure configuration consistency between the peer switches. Similar to regular port channels, vPCs are subject to consistency checks and compatibility checks. During a compatibility check, one vPC peer conveys configuration information to the other vPC peer in order to verify that vPC member ports can actually form a port channel. In addition to compatibility checks for the individual vPCs, Cisco Fabric Services is also used to perform consistency checks for a set of switchwide parameters that need to be configured consistently on both peer switches.
- Cisco Fabric Services is used to track vPC status on the peer. When all vPC member ports on one of the vPC peer switches go down, Cisco Fabric Services is used to notify the vPC peer switch that its ports have become orphan ports and that traffic that is received on the peer link for that vPC should now be forwarded to the vPC.
- Layer 3 vPC peers may be configured to synchronize their respective ARP tables. This feature is disabled by default and can be enabled by using the **ip arp synchronize** command. If enabled, this feature helps ensure faster convergence time upon a vPC switch reload. When two switches are reconnected after a failure, they use Cisco Fabric Services to perform bulk synchronization of the ARP table.

https://t.me/learningnets

Between the pair of vPC peer switches, an election is held to determine a primary and secondary vPC device. This election is non-pre-emptive. The vPC primary or secondary role is primarily a control plane role that determines which of the two switches will primarily be responsible for the generation and processing of spanning-tree BPDUs for the vPCs.

Note The vPC peer-switch option allows both the primary and secondary devices to generate BPDUs for vPCs independently. The two switches will use the same spanning-tree bridge ID to ensure that devices connected on a vPC still see the vPC peers as a single logical switch. This option is discussed later in the lesson.

Both switches actively participate in traffic forwarding for the vPCs. However, the primary and secondary roles are also important in certain failure scenarios, most notably in a peer link failure. When the vPC peer link fails but the vPC peer switches determine through the peer keepalive mechanism that the peer switch is still operational, then the operational secondary switch suspends all vPC member ports. The secondary role also shuts down all switch virtual interfaces (SVIs) associated with any VLANs that are configured as allowed VLANs for the vPC peer link.

For LACP and STP, the two vPC peer switches present themselves as a single logical switch to devices connected on a vPC. For LACP, this result is accomplished by generating the LACP system ID from a reserved pool of MAC addresses, which are then combined with the vPC domain ID. For STP, the behavior depends on the use of the peer-switch option. If the peer-switch option is not used, the vPC primary is responsible for generating and processing BPDUs and uses its own bridge ID for the BPDUs. The secondary role relays BPDU messages but does not generate BPDUs itself for the vPCs. When the peer-switch option is used, both the primary and secondary switches send and process BPDUs. However, they use the same bridge ID to present themselves as a single switch to devices connected on a vPC.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

vPC Guidelines

Attribute	Description
Switch types	The switch type must be the same. For example pair 5000-5000 or 5500-5500 is supported, but not 5000-5500.
	The vPC has to be built between the same line card modules
Link speed	vPC peer links must consist of 10/40/100 Gigabit Ethernet ports.
vPC keepalive	Avoid running vPC keepalive over vPC peer link
vPC peer link	At least two 10-Gigabit Ethernet interfaces.
VDC	• vPC cannot stretch across multiple VDCs on a single switch. • Each VDC with vPC requires its own vPC peer link and vPC peer keepalive link.
Number of vPC peers	A vPC domain cannot consist of more than two peer switches.
Number of vPCs per switch	You cannot configure more than one vPC domain per switch or VDC.
Routing	• Dynamic routing to vPC peers across a vPC or across the vPC peer link is not supported. • Static routing across a vPC to an FHRP addresses is supported. • Dynamic routing across a vPC between two Layer 3 switches that are not participating in vPC is supported.
vPC member ports	• vPC member ports must be on same line card type on both switches, i.e. M1-M1, F1-F1, M2-M2 etc.

© 2012 Cisco and/or its affiliates. All rights reserved.

Consider these guidelines and limitations when deploying vPCs:

- You must pair Cisco Nexus switches of the same type. For example, you can deploy vPC on a pair of Cisco Nexus 5000 Series switches or Cisco Nexus 5500 Platform switches but not on a combination of them.
- A vPC peer link must consist of Ethernet ports with an interface speed of 10 Gb/s or higher. It is recommended to use at least two 10 Gigabit Ethernet ports in dedicated mode on two different I/O modules.
- vPC keepalive should not run across a vPC peer link.
- A vPC is a per-VDC function on the Cisco Nexus 7000 Series switches. A vPC can be configured in multiple VDCs, but the configuration is entirely independent. A separate vPC peer link and vPC peer keepalive link are required for each of the VDCs. vPC domains cannot be stretched across multiple VDCs on the same switch, and all ports for a given vPC must be in the same VDC.
- A vPC domain by definition consists of a pair of switches that are identified by a shared vPC domain ID. It is not possible to add more than two switches or VDCs to a vPC domain.
- Only one vPC domain ID can be configured on a single switch or VDC. It is not possible for a switch or VDC to participate in more than one vPC domain.
- A vPC is a Layer 2 port channel. vPC does not support the configuration of Layer 3 port channels. Dynamic routing from the vPC peers to routers connected on a vPC is not supported. It is recommended that routing adjacencies are established on separate routed links.
- Static routing to FHRP addresses is supported. The FHRP enhancements for vPC enable routing to a virtual FHRP address across a vPC.
- A vPC can be used as a Layer 2 link to establish a routing adjacency between two external routers. The routing restrictions for vPCs only apply to routing adjacencies between the vPC peer switches and routers that are connected on a vPC.

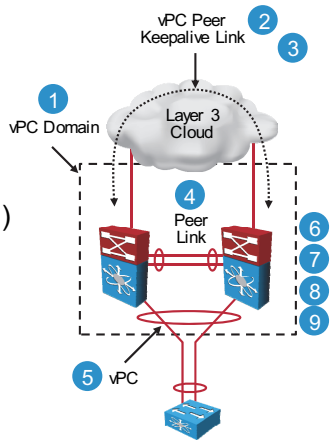
https://t.me/learningnets

Configuring vPC

This topic explains how to configure vPCs on Cisco Nexus switches.

vPC Configuration Procedure

1. Configure vPC domain
2. Choose peer keepalive option
3. Configure peer keepalive link
4. Configure vPC peer link
5. Configure vPCs
6. Optimize vPC—peer gateway (optional)
7. Optimize vPC—peer switch (optional)
8. Verify brief vPC (optional)
9. Verify vPC consistency parameters (optional)



© 2012 Cisco and/or its affiliates. All rights reserved.

Follow these steps to implement a vPC:

Step 1 Enable the vPC feature and configure the vPC domain ID on both switches.

Step 2 Choose a peer keepalive deployment option.

Step 3 Establish the vPC peer keepalive link.

Step 4 Configure the vPC peer link. This step completes the global vPC configuration on both vPC peer switches.

Step 5 Configure individual vPCs to downstream devices.

Step 6 Optionally, enable the peer gateway feature to modify the FHRP operation.

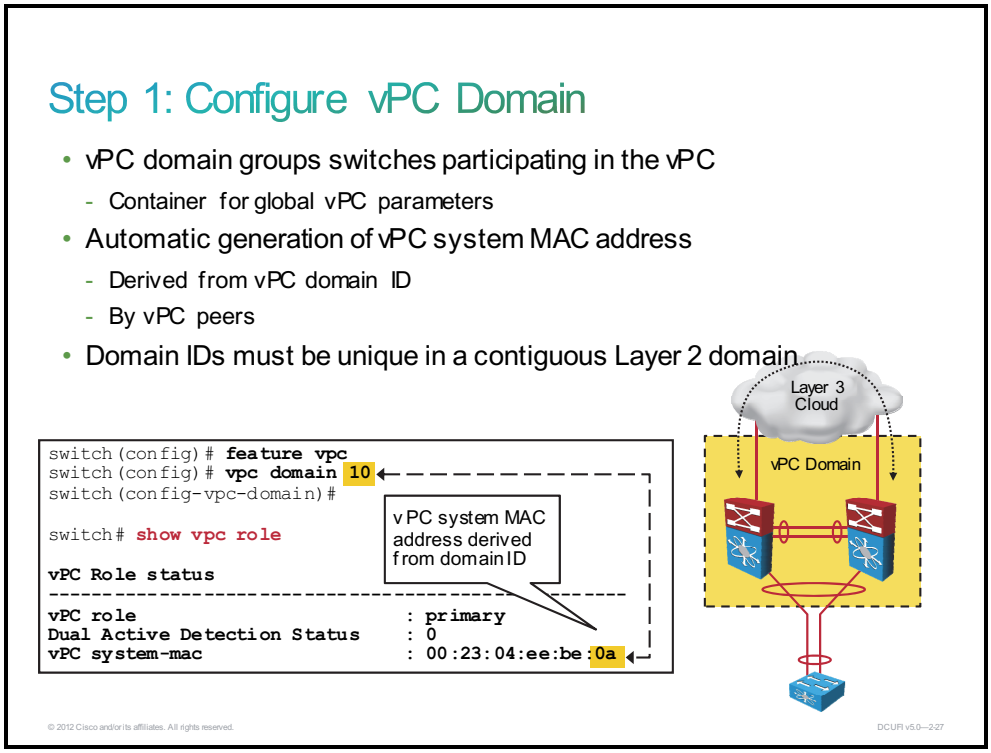
Step 7 Optionally, enable the peer switch feature to optimize the STP behavior with vPCs.

Step 8 Optionally, verify operation of the vPC.

Step 9 Optionally, verify vPC consistency parameters.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



The vPC domain defines the pair of vPC peer switches that participate in vPC. When you enter the vPC domain ID, you enter a subconfiguration mode where you can then configure additional global parameters for the vPC domain.

The vPC domain ID is a value between 1 and 1000 that uniquely identifies the vPC switch pair. The vPC peer devices use the vPC domain ID that you configure to automatically assign a unique vPC system MAC address. Each vPC domain has a unique MAC address that is used as a unique identifier for the specific vPC-related operation. Although the devices use the vPC system MAC addresses only for link-scope operations, such as LACP, it is recommended that you create each vPC domain within the contiguous Layer 2 network with a unique domain ID. You can also configure a specific MAC address for the vPC domain rather than having Cisco NX-OS Software assign the address.

The example in the figure shows how to configure and verify the vPC domain. These commands are used:

- **feature vpc**: This command enables the vPC feature.
- **vpc domain** *domain-id*: This command configures the domain ID. The same domain ID must be used on both vPC peer switches in the vPC domain.
- **show vpc role**: This command shows the result of the vPC role election, and it also shows the system MAC address that is derived from the vPC domain ID.

https://t.me/learningnets

Step 2: Choose Peer Keepalive Link Deployment

Peer keepalive link

- Detects and resolves roles if a dual-active condition occurs
- Out-of-band heartbeat between vPC peers

Deployment Option	Nexus 5000/5500	Nexus 7000
Dedicated non-mgmt port	Use a dedicated port and VLAN	Use a dedicated routed port in a separate VRF (a Gigabit Ethernet port is sufficient)
OOB mgmt	Use the OOB management interface mgmt0	Use the OOB management interface mgmt0 *
In-band	Use an in-band Layer 3 network	Use an upstream Layer 3 network

* Do not use crosscables to connect mgmt0

© 2012 Cisco and/or its affiliates. All rights reserved.

The peer keepalive link provides an OOB heartbeat between the vPC peer switches, which is used to detect and resolve a dual-active condition when the vPC peer link fails. The peer keepalives are IP-based and can be routed across an IP network if required.

Because it is vital that peer keepalives never be carried on the vPC peer link, these recommendations are made for deployment of the peer keepalive infrastructure:

- For Cisco Nexus 7000 Series switches, it is recommended that you create a separate virtual routing and forwarding (VRF) instance specifically for the vPC peer keepalives. By assigning a specific routed port to this VRF, you can ensure that the peer keepalive traffic is always carried on that link and never carried on the peer link. Because this link carries only keepalives, a Gigabit Ethernet port is sufficient for this link. Also, the port that is used for the peer keepalive link should ideally be terminated on a different I/O module than the links that form the peer link. If it is not possible to allocate a dedicated port for peer keepalives, the OOB management network can be used. However, in this case, it is important that the management ports on both supervisors are connected to the OOB management network. Do not use Ethernet crossover cables to connect the management ports on the vPC peers to each other back-to-back. To do so will cause the peer keepalive link to fail on supervisor switchover. If neither of these options is available, an upstream Layer 3 network in the core or aggregation layer of the data center could be used for the peer keepalives.
- For the Cisco Nexus 5000 Series switches, the recommendations are slightly different: It is recommended to use the OOB management interface mgmt 0 for peer keepalives if possible. If this option is not available, a dedicated port with an associated VLAN and SVI should be used. If it is also not possible to dedicate a separate port for the peer keepalives, then an in-band Layer 3 network can be used. However, you should take care that the VLAN associated with the peer keepalive connection is not allowed on the vPC peer link if this option is used.

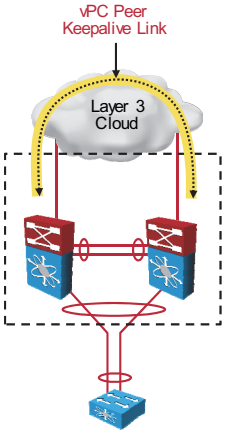
https://t.me/learningnets

Step 3: Configure Peer Keepalive Link

- Should never be in VLAN carried on the vPC peer link
- The management VRF is used by default

```
switch(config-vpc-domain)# peer-keepalive destination
192.168.1.2 source 192.168.1.1 vrf VPC-KEEPALIVE

switch# show vpc peer-keepalive

vPC keep-alive status            : peer is alive
--Peer is alive for              : (231) seconds, (92) msec
--Send status                    : Success
--Last send at                   : 2011.01.31 22:05:24 874 ms
--Sent on interface              : Eth1/27
--Receive status                 : Success
--Last receive at                : 2011.01.31 22:05:25 155 ms
```

Nexus 7000 example using VRFs



© 2012 Cisco and/or its affiliates. All rights reserved.

By default, the vPC peer keepalive packets are routed in the management VRF and use the OOB mgmt 0 interface. You can configure the vPC peer link to use a different VRF, but you should take care that the peer keepalive traffic is not routed across the vPC peer link.

The example in the figure shows how to configure and verify the vPC peer keepalive link. These commands are used:

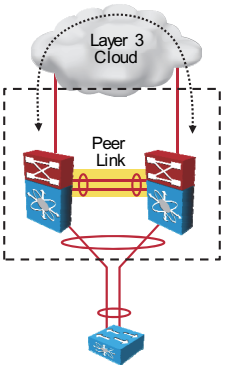
- **peer-keepalive destination** *ip-address* [**source** *ip-address*] [**vrf** {*name* | **management**}]: This command specifies the destination IP address for the vPC peer keepalive link. By default, this IP address is resolved in the management VRF, but other VRFs can be specified. If a VRF other than the management VRF is used, the source IP address should also be specified because the source IP address defaults to the management IP address. Additional options can be added to this command to change the timers and quality of service (QoS) values that are used by default.
- **show vpc peer-keepalive**: This command can be used to verify the status of the vPC peer keepalive link.

https://t.me/learningnets

Step 4: Configure Peer Link

- vPC peer link
  - Carries data and control traffic between peer switches
- Recommendations:
  - Port channel of at least two dedicated 10 Gigabit Ethernet ports
  - Trunk mode
  - Only transport of vPC VLANs

```
switch(config)# interface port-channel 20
switch(config-if)# switchport mode trunk
switch(config-if)# switchport trunk allowed vlan 100-105
switch(config-if)# switchport trunk native vlan 100
switch(config-if)# vpc peer-link
switch(config-if)# spanning-tree port type network
```



© 2012 Cisco and/or its affiliates. All rights reserved.

The vPC peer link carries essential vPC traffic between the vPC peer switches. The vPC peer link is a port channel, which should consist of at least two dedicated 10 Gigabit Ethernet links. These links should be terminated on two different I/O modules if at all possible.

The vPC peer link should be configured as a trunk. The allowed VLAN list for the trunk should be configured in such a way that only vPC VLANs (VLANs that are present on any vPCs) are allowed on the trunk. It is not recommended to carry non-vPC VLANs on the vPC peer link, because this configuration could cause severe traffic disruption for the non-vPC VLANs if the vPC peer link fails.

It is recommended that you enable Bridge Assurance on the vPC peer link and use Unidirectional Link Detection (UDLD) to protect against unidirectional link failures.

The example in the figure shows how to configure the vPC peer link. The primary command that is used in this example is the **vpc peer-link** command, which assigns an existing port channel interface as the vPC peer link. The example also shows the configuration of the recommended best practices.

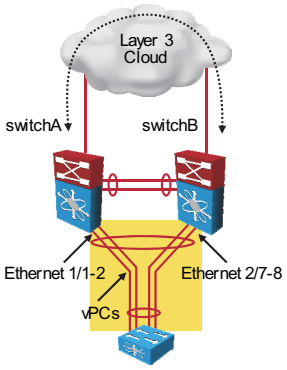
https://t.me/learningnets

Step 5: Configure vPCs

- Port channel on both vPC peer switches
- Port channels and physical interfaces compatible on both peers
- Binding through vPC number
  - Must be unique for the vPC domain
  - Several vPC numbers can exist per domain
  - Combines port channels on peer switches into a vPC

```
switchA(config)# interface ethernet 1/1-2
switchA(config-if-range)# channel-group 7 mode active
switchA(config-if-range)# interface port-channel 7
switchA(config-if)# switchport mode trunk
switchA(config-if)# vpc 7
```

```
switchB(config)# interface ethernet 2/7-8
switchB(config-if-range)# channel-group 7 mode active
switchB(config-if-range)# interface port-channel 7
switchB(config-if)# switchport mode trunk
switchB(config-if)# vpc 7
```



© 2012 Cisco and/or its affiliates. All rights reserved.

Once the vPC domain has been properly established, the individual vPCs can be configured. To configure a vPC, a port channel must be configured on both vPC peer switches. These two port channels must then be associated with each other by assigning a vPC number to the port channel interfaces. The vPC port number is unique for the vPC within the vPC domain and must be identical on the two peer switches.

As with regular port channels, vPC member ports should have a compatible and consistent configuration. You should ensure that the configurations on vPC member ports are not only compatible on a single switch but also between peer switches.

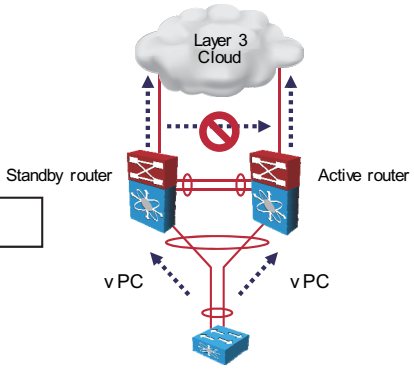
The example in the figure shows how to use the **vpc** *number* command to combine two existing port channel interfaces on the two vPC peer switches into a single vPC. The example also shows the use of the **channel-group** command to create the port channels that are combined into a vPC. In the example, the vPC is configured as a trunk. This configuration is optional. A vPC could also be configured as an access port, for example, if it is connected to a dual-homed server.

https://t.me/learningnets

Step 6: Optimize vPC Peer Gateway

- Peer gateway feature
  - Allows a vPC switch to act as active gateway for traffic to peer router MAC
  - Forwards local traffic to vPC node and avoids use of the peer link
  - Interoperable with NAS and load balancers
- ICMP redirects are disabled for SVIs associated with vPC VLANs

```
switch(config)# vpc domain 10
switch(config-vpc-domain)# peer-gateway
```



© 2012 Cisco and/or its affiliates. All rights reserved.

You can configure vPC peer devices to act as the gateway, even for packets that are destined to the vPC peer device MAC address.

The vPC peer gateway capability allows a vPC switch to act as the active gateway for packets that are addressed to the router MAC address of the vPC peer. This feature enables local forwarding of such packets without the need to cross the vPC peer link. In this scenario, the feature optimizes the use of the peer link and avoids potential traffic loss in FHRP scenarios.

The peer gateway feature must be configured on both primary and secondary vPC peers and be nondisruptive to the operations of the device or to the vPC traffic. The vPC peer-gateway feature can be configured globally under the vPC domain submode.

When this feature is enabled, IP redirects are automatically disabled on all interface VLANs that are associated with a vPC VLAN. This avoids generation of IP redirect messages for packets that are switched through the peer gateway router.

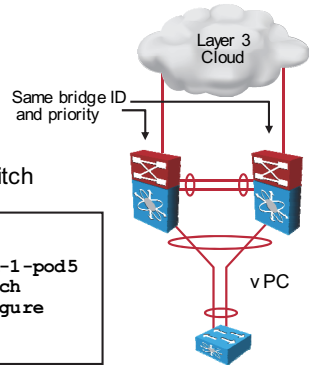
Note Packets arriving at the peer gateway vPC device will have their Time to Live (TTL) decremented, so packets carrying TTL = 1 may be dropped in transit because of TTL expiration. This fact needs to be taken into account when the peer gateway feature is enabled and particular network protocols sourcing packets with TTL = 1 operate on a vPC VLAN.

https://t.me/learningnets

Step 7: Optimize vPC Peer Switch

- vPC peer switch feature
  - vPC primary and secondary are both root devices
  - Different STP behavior on vPC and non-vPC ports
- On vPC ports
  - BPDUs originated by primary and secondary devices with same designated bridge ID
- On non-vPC ports
  - Maintain local bridge ID instead of the vPC bridge ID
  - Advertise Bridge ID of the vPC system as the root
- Better convergence
  - During vPC primary switch failure and recovery
  - Avoids RSTP sync
- No need for pinning the STP root to vPC primary switch

```
switch(config)# vpc domain 10
switch(config-vpc-domain)# peer-switch
switch(config-vpc-domain)# 2012 Jan 31 22:46:08 N7K-1-pod5
%STP-2-VPC_PEERSWITCH_CONFIG_ENABLED: vPC peer-switch
configuration is enabled. Please make sure to configure
spanning tree "bridge" priority as per recommended
guidelines to make vPC peer-switch operational.
```



© 2012 Cisco and/or its affiliates. All rights reserved.

The peer switch option optimizes the behavior of spanning tree with vPCs:

- The vPC primary and secondary are both root devices and both originate BPDUs.
- The BPDUs originated by both the vPC primary and the vPC secondary have the same designated bridge ID on vPC ports.
- The BPDUs originated by the vPC primary and secondary devices on non-vPC ports maintain the local bridge ID instead of the vPC bridge ID and advertise the bridge ID of the vPC system as the root.

The peer switch option has these advantages:

- It reduces the traffic loss upon restoration of the peer link after a failure.
- It reduces the disruption that is associated with a dual-active failure, whereby both vPC members become primary. Both devices keep sending BPDUs with the same bridge ID information on vPC member ports. This prevents the port channel STP consistency feature from potentially disabling the port channel on an attached device.
- It reduces the potential loss of BPDUs if the primary and secondary roles change.

The example in the figure shows how to configure the peer-switch feature by using the **peer-switch** command. In addition to enabling the peer-switch feature, you should also set the best possible spanning tree bridge priority value on both peer switches. This setting forces the vPC switch pair to become the root of the spanning tree for the vPC VLANs.

https://t.me/learningnets

Step 8: Verify Basic vPC

```
switch# show vpc brief
Legend:
                (*) - local vPC is down, forwarding via vPC peer-link

vPC domain id                     : 10
Peer status                       : peer adjacency formed ok
vPC keep-alive status             : peer is alive
Configuration consistency status: success
Type-2 consistency status         : success
vPC role                          : primary
Number of vPCs configured         : 1
Peer Gateway                      : Enabled
Dual-active excluded VLANs        : -

vPC Peer-link status
---------------------------------------------------------------------
id   Port   Status Active vlans
--   ----   ------ -----------------------------------------------
1    Po20   up     100-105

vPC status
----------------------------------------------------------------------------
id   Port   Status Consistency Reason                     Active vlans
--   ----   ------ ----------- ------------------------ -----------
7    Po7    up     success     success                  100-105
```

Domain ID

Peer gateway enabled. Default setting would show "Disabled."

vPC ID

© 2012 Cisco and/or its affiliates. All rights reserved.

Several commands can be used to verify the operation of vPC. The primary command to be used in initial verification is the **show vpc brief** command. This command displays the vPC domain ID, the peer-link status, the keepalive message status, whether the configuration consistency is successful, and whether a peer link has formed. The command also displays the status of the individual vPCs that are configured on the switch, including the result of the consistency checks.

 © 2012 Cisco Systems, Inc.

https://t.me/learningnets

Step 9: Verify vPC Consistency Parameters

```
switch# show vpc consistency-parameters global
Legend:
        Type 1 : vPC will be suspended in case of mismatch

Name                        Type  Local Value            Peer Value
-------------               ----  ---------------------- -----------------------
STP Mode                    1     Rapid-PVST             Rapid-PVST
STP Disabled                1     None                   None
STP MST Region Name         1     ""                     ""
STP MST Region Revision     1     0                      0
<output omitted>

switch# show vpc consistency-parameters vpc
    Legend:
        Type 1 : vPC will be suspended in case of mismatch

Name                        Type  Local Value            Peer Value
-------------               ----  ---------------------- -----------------------
STP Port Type               1     Default                Default
STP Port Guard              1     None                   None
STP MST Simulate PVST       1     Default                Default
lag-id                      1     [7f9b,0-23-4-ee-be-ac] [7f9b,0-23-4-ee-be-ac]
mode                        1     active                 active
Speed                       1     10 Gb/s                10 Gb/s
Duplex                      1     full                   full
Port Mode                   1     trunk                  trunk
Native Vlan                 1     1                      1
MTU                         1     1500                   1500
Allowed VLANs               -     1-3967,4048-4093       1-3967,4048-4093
Local suspended VLANs       -     1,10                   -
```

Local and peer values must match

© 2012 Cisco and/or its affiliates. All rights reserved.

If the **show vpc brief** command displays failed consistency checks, you can use the **show vpc consistency-parameters** command to find the specific parameters that caused the consistency check to fail. The **global** option on this command allows you to verify the consistency of the global parameters between the two peer switches. The **vpc** or **interface** options can be used to verify consistency between the port channel configurations for vPC member ports.

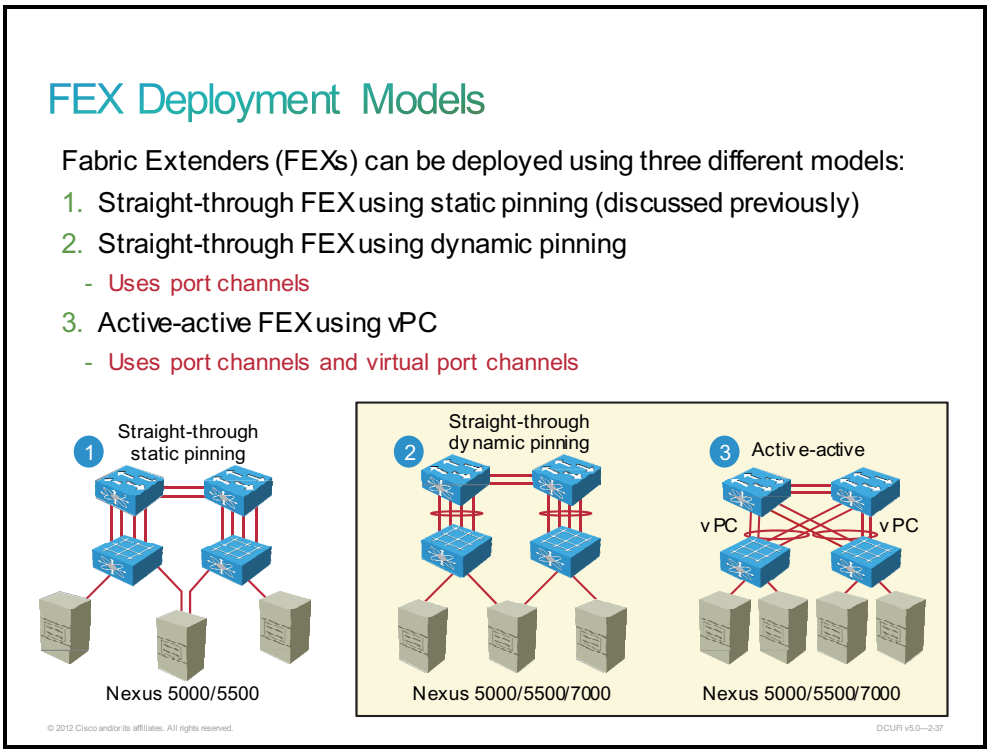
After you enable the vPC feature and configure the peer link on both vPC peer devices, Cisco Fabric Services messages provide a copy of the configuration on the local vPC device configuration to the remote vPC peer device. The system determines whether any of the crucial configuration parameters differ on the two devices. The parameters must be configured identically or the vPC moves into “suspend” mode. The per-interface parameters must be consistent per interface, and the global parameters must be consistent globally:

- Port channel mode: on, off, or active
- Link speed per channel
- Duplex mode per channel
- Trunk mode per channel, including native VLAN, VLANs allowed on trunk, and the tagging of native VLAN traffic
- STP mode
- STP region configuration for Multiple Spanning Tree (MST)
- Enabled or disabled state per VLAN
- STP global settings, including Bridge Assurance setting, port type, and loop guard settings
- STP interface settings, including port type, loop guard, and root guard
- Maximum transmission unit (MTU)

https://t.me/learningnets

Configuring the FEX

This topic explains how to configure the Cisco Nexus 2000 Series FEX connected to a Cisco Nexus 5000 or 7000 Series switch.



There are three deployment models that are used to deploy FEXs together with the Cisco Nexus 5000 and 7000 Series switches:

- **Straight-through using static pinning**: In the straight-through model, each FEX is connected to a single Cisco Nexus switch. The single switch that the FEX is connected to exclusively manages the ports on that particular FEX. Static pinning means that each downlink server port on the FEX is statically pinned to one of the uplinks between the FEX and the switch. Traffic to and from a specific server port always uses the same uplink. This model uses neither port channels nor vPCs and was explained in the "Configuring Layer 2 Switching Features" lesson.
- **Straight-through using dynamic pinning**: This deployment model also uses the straight-through connection model between the FEXs and the switches. However, there is no static relation between the downlink server ports and the uplink ports. The ports between the FEX and the switch are bundled into a port channel, and traffic is distributed across the uplinks based on the port channel hashing mechanism.
- **Active-active FEX using vPC**: In this deployment model, the FEX is dual-homed to two Cisco Nexus switches. vPC is used on the link between the FEX and the pair of switches. Traffic is forwarded between the FEX and the switches based on vPC forwarding mechanisms.

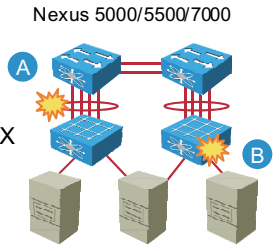
Note	The second and third models are discussed in this topic.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Dynamic Pinning

- Port channel between Cisco Nexus switch and FEX 2000
- Traffic distribution across the uplinks determined through port channel hashing
- Failure scenarios:
  - A. One or few uplinks fail
    - Traffic is rehashed onto the remaining links
    - Server downlinks are not disabled
    - Oversubscription ratio changes
    - Single-homed servers retain connectivity
    - Dual-homed servers do not fail over to other NIC/FEX
  - B. Access switch or FEX fails
    - Single-homed servers lose connectivity
    - Dual-homed servers fail over to other NIC/FEX
    - Oversubscription ratio changes



© 2012 Cisco and/or its affiliates. All rights reserved.

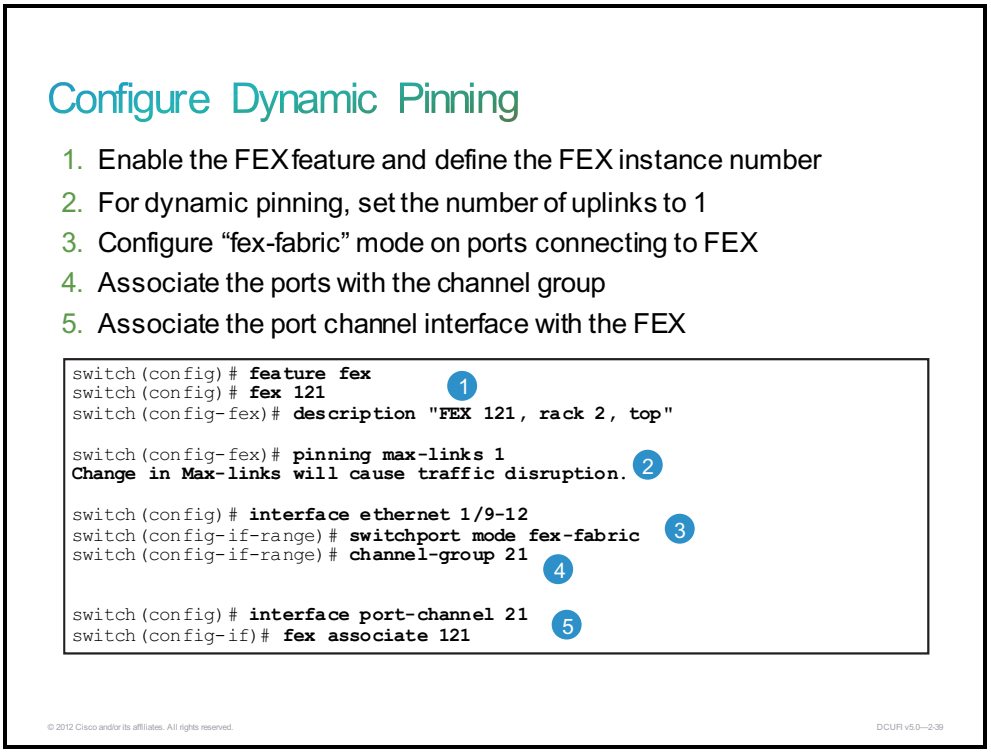
In dynamic pinning mode, the server ports are not statically pinned to an uplink port, but all server ports share the combined bandwidth of the uplink ports. This is achieved by configuring a port channel between the Cisco Nexus 2000 Series FEX and the Cisco Nexus 5000 Series switch. Instead of statically assigning traffic from specific server ports to the uplink port that they are pinned to, traffic is now distributed over the uplinks based on the port channel load-balancing hash algorithm.

If one of the uplinks between the Cisco Nexus 2000 Series FEX and the Cisco Nexus 5000 Series switch fails, the FEX will not disable any server ports because there is no longer a direct relationship between the uplink ports and the server ports. Instead, the traffic of the 48 server ports is now distributed over the remaining three 10 Gigabit Ethernet uplinks. The servers that are connected to the Cisco Nexus 2000 Series FEX will not register the failure and will keep forwarding traffic on the NIC that is connected to the FEX.

Single-homed servers will not lose connectivity when using dynamic pinning. Their traffic is simply redistributed over the remaining uplinks.

Dual-homed servers will not fail over to the redundant NIC. However, this means that the oversubscription ratio for the remaining uplink ports changes. The oversubscription ratio for the remaining ports in the example is 48:30 = 1.6:1, which represents a 33-percent increase in traffic on each uplink port. If the uplinks are already running close to maximum utilization, it may cause traffic from all servers to be dropped, thereby degrading performance for all servers.

https://t.me/learningnets



Follow these steps when implementing dynamic pinning:

Step 1 Enable the FEX feature and define the FEX instance number.

Step 2 For dynamic pinning, set the number of uplinks to 1. The pinning max-links parameter is set to one, because all of the ports are pinned to the port channel interface instead of the individual physical interfaces.

Step 3 Configure fex-fabric mode on ports connecting to a FEX.

Step 4 Associate the ports connecting to an FEX with the channel group.

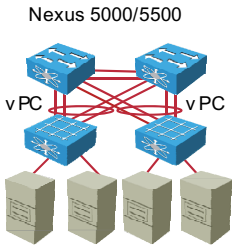
Step 5 Associate the port channel interface with the FEX.

Note The **pinning max-links** command is not required on the Cisco Nexus 7000 Series switches because only dynamic pinning is supported.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Active-Active FEX

- FEX dual-homed to two Cisco Nexus 5000/5500 switches
- Highest availability of FEX-based solutions
  - Protection against failures of uplinks, FEX, and switch
- vPC as FEX uplink connection
  - vPC ports on FEX configured consistently on both switches
  - Consistency checks as for regular vPCs
- Configuration synchronization feature
  - Automatic configuration synchronization
  - Available on Cisco Nexus 5000 switches



© 2012 Cisco and/or its affiliates. All rights reserved.

In the active-active FEX deployment, the Cisco Nexus 2000 Series FEX is controlled and configured by two Cisco Nexus 5000 Series switches. Both switches must be configured in a consistent manner, and vPC has to be set up to combine the FEX uplinks into a single port channel.

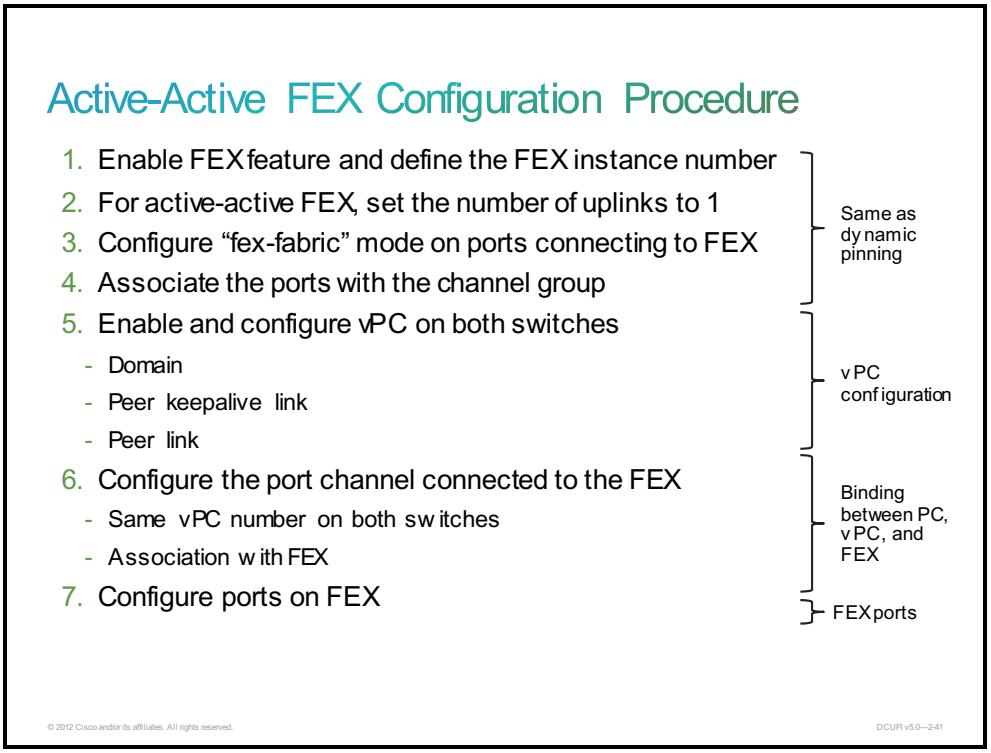
Because a vPC-based port channel is used between the FEXs and the switches, dynamic pinning is automatically used. Traffic is balanced across the FEX uplinks based on port channel load balancing. When one of the FEX uplinks is lost, traffic will be balanced across the remaining uplinks.

Because vPC is used between the FEXs and the switches, vPC cannot be used between the FEXs and the servers. This means that port channels cannot be configured between the server and the access switches. Active/standby and transmit load balancing can still be used as NIC teaming options to attain high availability and some extent of load balancing.

In the active-active FEX model, single-homed servers will maintain connectivity if one of the Cisco Nexus switches fails. Thus, this design increases the availability of single-homed servers to a level that is comparable to that of a single-homed server that is connected to a chassis-based switch with dual-redundant supervisors.

Note The two Cisco Nexus switches are configured independently. Therefore, configuration changes that are made to ports on a dual-homed Cisco Nexus 2000 Series FEX have to be manually synchronized between the two Cisco Nexus switches. To ease this administrative burden, the Cisco Nexus 5000 Series switch supports a configuration synchronization feature called switch profiles, which were discussed earlier in the lesson.

https://t.me/learningnets



The initial part of an active-active FEX configuration is similar to a straight-through configuration with dynamic pinning. The FEX is created and the fabric ports are configured and combined in a channel group. All these commands, shown in Steps 1–4, have to be executed on both switches.

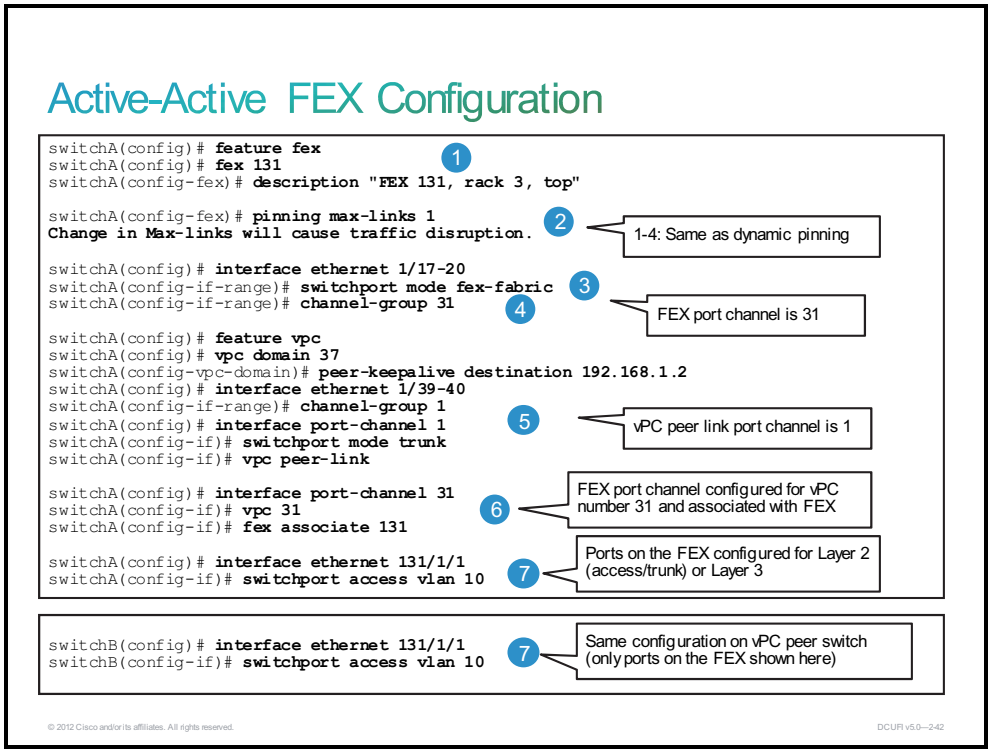
Next, vPCs should be configured. A vPC domain, including a peer keepalive link and peer link, should be created. This is shown in Step 5.

Next, the port channels that are associated with the FEX fabric ports are combined into a vPC, shown in Step 6.

Once the vPC between the FEX and the switches has successfully formed, the ports on the FEX will be visible on both switches and can be configured in Step 7.

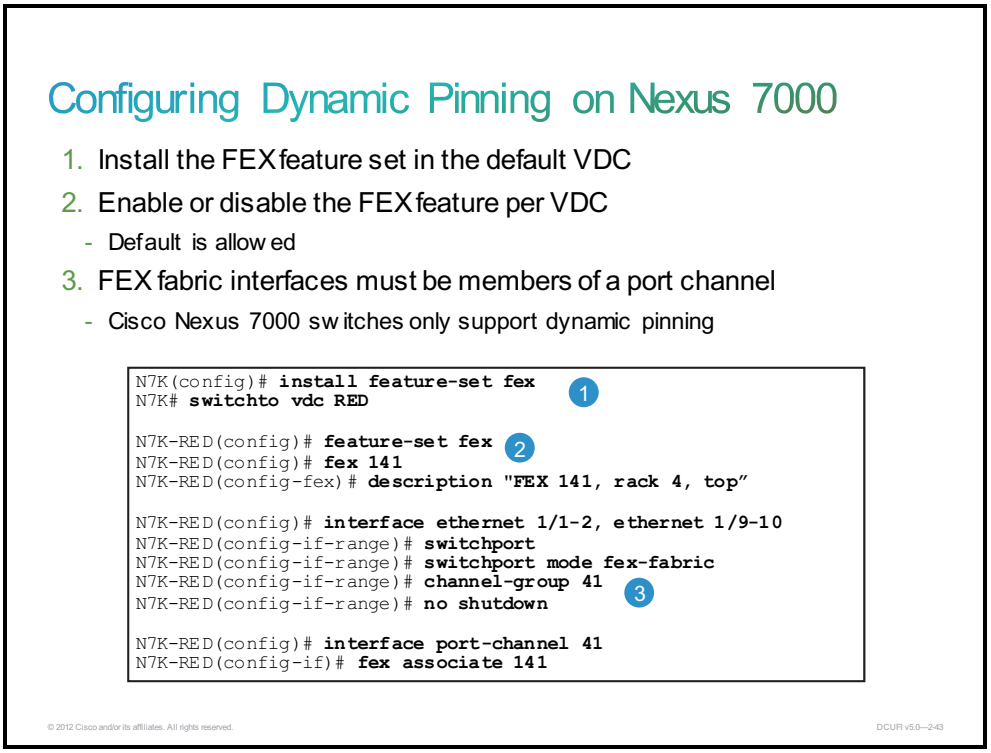
Effectively, each of the individual ports on the FEX is treated as a vPC on the Cisco Nexus switches. Therefore, the same consistency checks are applied to the ports on the FEX as for regular vPCs. This means that any configuration change that is made on a port on the FEX should be made consistently on both of the vPC peer switches.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



This figure illustrates an active-active FEX configuration example that was described in the previous procedure.

https://t.me/learningnets



Configuration of an FEX on a Cisco Nexus 7000 Series switch is slightly different from the configuration on a Cisco Nexus 5000 Series switch. Partially, this is caused by the VDC-based architecture of the Cisco Nexus 7000 Series switches. Before a FEX can be configured in a VDC, the services that are required by the FEX feature need to be installed in the default VDC. To enable the use of the FEX feature set, use the **install feature-set fex** command in the default VDC. After the FEX feature set has been installed in the default VDC, the feature set can be enabled in any VDC by using the **feature-set fex** command.

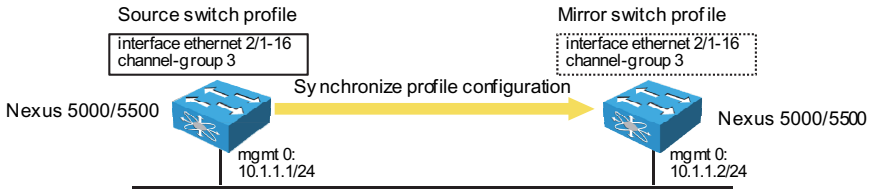
It is possible to restrict the use of the FEX feature set to specific VDCs only. By default, all VDCs can enable the FEX feature set once it has been installed in the default VDC. If you want to disallow the use of FEXs in a specific VDC, you can use the **no allow feature-set fex** command in VDC configuration mode for that particular VDC.

Another difference with the FEX configuration on the Cisco Nexus 5000 Series switches is that the Cisco Nexus 7000 Series switches only support dynamic pinning, which makes it unnecessary to specify the maximum number of pinning interfaces by using the **pinning max-links** command.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Switch Profiles

- Enable consistent configurations across multiple switches
  - Examples: identical configurations with PortChannel/vPortChannel
- Switch profile automatically synchronized to peer switch
  - Leverages configuration synchronization (config-sync) feature
  - Cisco NX-OS Release 5.0(2)N1(1) and later on Nexus 5000/5500 switches
- Provides control of exactly which configuration gets synchronized



© 2012 Cisco and/or its affiliates. All rights reserved.

Several applications, such as port channels and vPCs, require consistent configuration across multiple switches in the network. Mismatched configurations can cause errors or misconfigurations that can result in service disruptions.

The configuration synchronization feature in Cisco NX-OS Release 5.0(2)N1(1) and later versions allows you to configure one switch profile and have the configuration be automatically synchronized to the peer switch. The feature is supported on Cisco Nexus 5000 Series switches and Cisco Nexus 5500 Platform switches. A switch profile provides these benefits:

- Allows configurations to be synchronized between switches
- Merges configurations when connectivity is established between two switches
- Provides control of exactly which configuration gets synchronized
- Ensures consistency across peers through merge and mutual-exclusion checks
- Provides verify and commit semantics
- Supports configuring and synchronizing port profile configurations
- Provides an import command to migrate existing vPC configurations to a switch profile

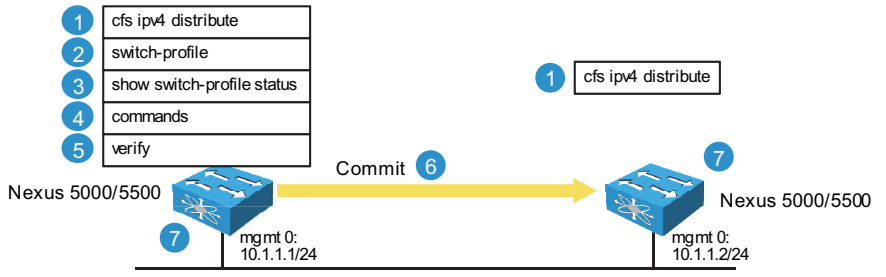
The switch profile feature includes the following configuration modes:

- **Configuration synchronization mode** (config-sync) allows you to create switch profiles. After entering the **config sync** command, you can create and name the switch profile that displays the switch profile mode. You must enter the **config sync** command on the local switch and on the peer switch that you want to synchronize.
- **Switch profile mode** allows you to add supported configuration commands to a switch profile that is later synchronized with a peer switch. Commands that you enter in switch profile mode are buffered until you enter the **commit** command.
- **Switch profile import mode** offers you the option to copy supported running configuration commands to a switch profile.

https://t.me/learningnets

Switch Profile Configuration Procedure

1. Enable Cisco Fabric Services distribution between the peer switches
2. Create switch profile and define the peer switch
3. Verify initial synchronization (optional)
4. Configure commands in the switch profile
5. Verify commands in the switch profile buffer (optional)
6. Commit changes
7. Verify switch profile synchronization



© 2012 Cisco and/or its affiliates. All rights reserved.

Follow these steps to configure Cisco Nexus 5000 Series switch and Cisco Nexus 5500 Platform switches by using the switch profiles:

Step 1 Enable Cisco Fabric Services distribution between the peer switches.

Step 2 Create the switch profile and define the peer switch.

Step 3 Verify initial synchronization (optional).

Step 4 Configure commands in the switch profile.

Step 5 Verify commands in the switch profile buffer (optional).

Step 6 Commit the changes.

Step 7 Verify switch profile synchronization.

https://t.me/learningnets

Switch Profile Configuration

```
switch(config)# cfs ipv4 distribute                                  1 Enable CFSoIP distribution over mgmt0 (both switches)

switch(config-sync)# switch-profile PC-profile
switch(config-sync-sp)# sync-peers destination 10.1.1.2              2 Switch profile and target switch
switch(config-sync-sp)# show switch-profile PC-profile status
Start-time:  15801 usecs after Mon March 26 06:21:08 2012
End-time:   6480 usecs after Mon March 26 06:21:13 2012              3 Initial verification
...
Status: Commit Success

switch(config-sync-sp)# interface ethernet 1/1-16                    4 Commands
switch(config-sync-sp-if-range)# switchport
switch(config-if-range)# channel-group 1
switch(config-sync-sp)# interface port-channel 1
switch(config-sync-sp-if)# <…Layer 2/3 configuration…>

                                                                     5 View buffer
switch(config-sync-sp-if)# show switch-profile switch-profile buffer
<output omitted>

switch(config-sync-sp-if)# commit                                    6 Push configuration
Commit Successful

                                                                     Verify results
switch(config-sync)# show switch-profile switch-profile status       7
<output omitted>
```

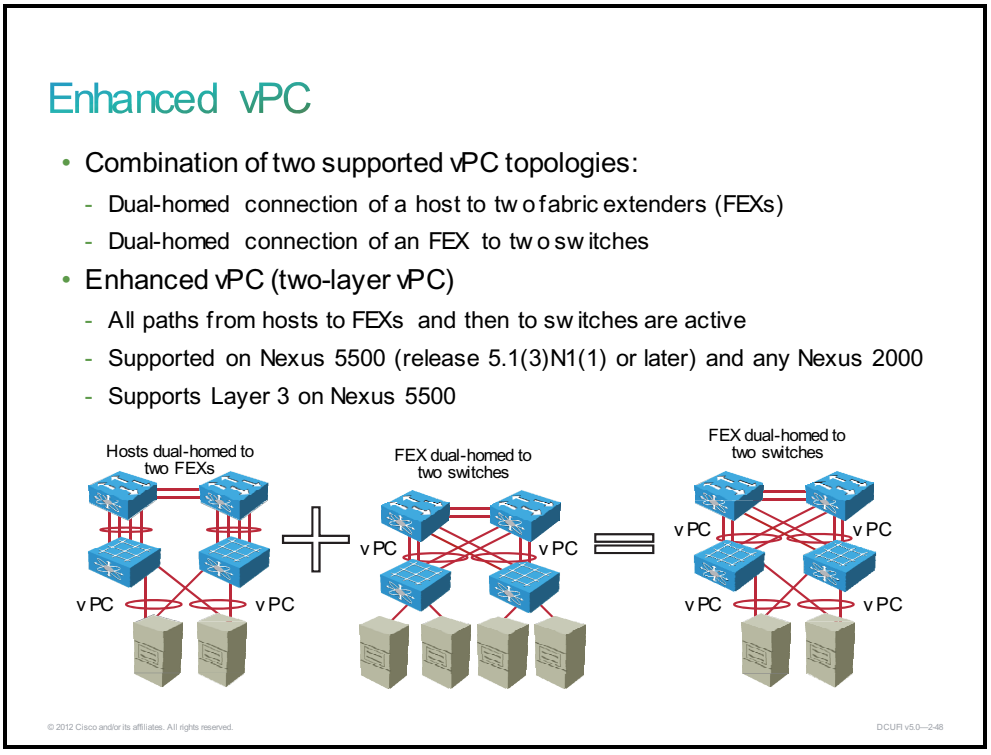
© 2012 Cisco and/or its affiliates. All rights reserved.

This configuration example illustrates the use of switch profiles.

https://t.me/learningnets

Configuring Enhanced vPCs

This topic explains how to configure the Enhanced vPCs on a Cisco Nexus 5000 Series switch.



The Enhanced vPC feature, known as two-layer vPC, combines two dual-homing in one solution:

- Dual-homed connection of a host to two FEXs
- Dual-homed connection of an FEX to two switches

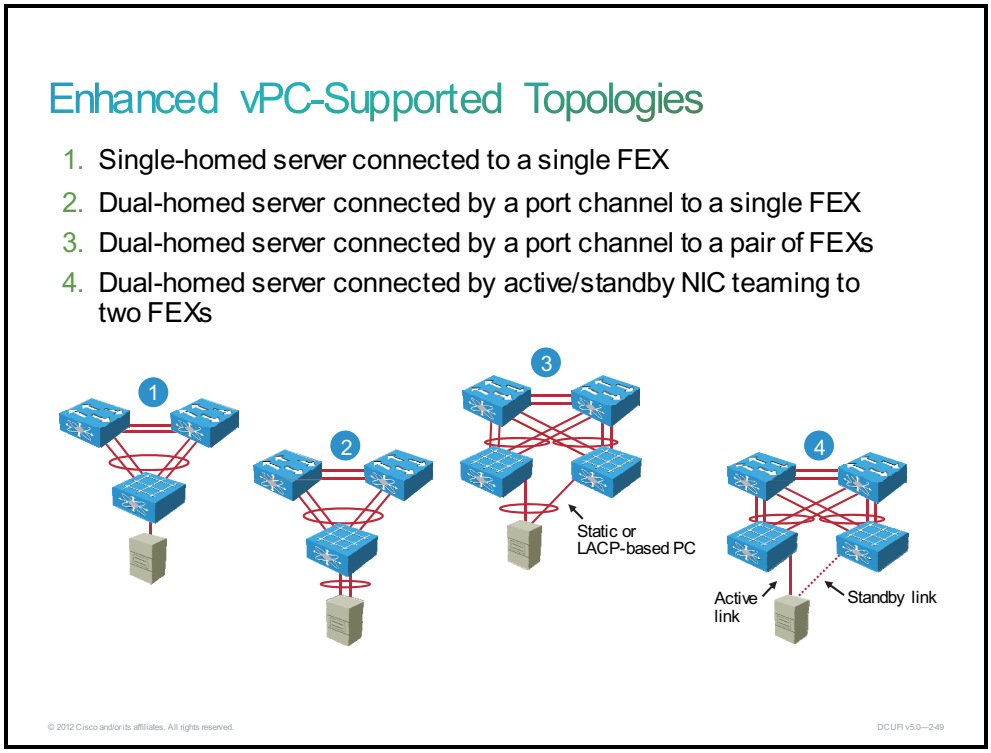
The combined topology is shown in the figure.

With Enhanced vPC, all available paths from the hosts to the FEXs and from the FEXs to the switches are active and carry Ethernet traffic, maximizing the available bandwidth and providing redundancy at both levels.

Enhanced vPC is supported on the Cisco Nexus 5500 Platform switch running NX-OS Release 5.1(3)N1(1) or a later release. Enhanced vPC can be deployed with any Cisco Nexus 2000 Series Fabric Extender. Enhanced vPC is compatible with Layer 3 features on the switch.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

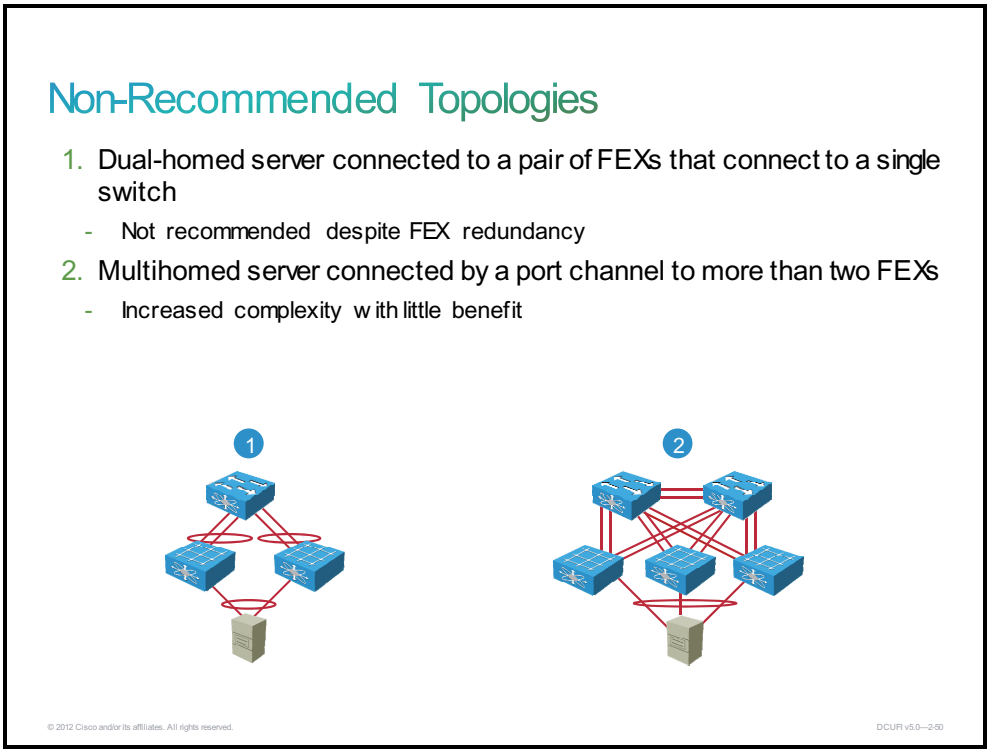


Enhanced vPC supports the following topologies:

- A single-homed server that is connected to a single FEX.
- A dual-homed server that is connected by a port channel to a single FEX.
- A dual-homed server that is connected by a port channel to a pair of FEXs. This topology allows connection to any two FEXs that are connected to the same pair of switches in a vPC domain. Static port channel and LACP-based port channel are both supported.
- A dual-homed server that is connected by active/standby NIC teaming to a pair of FEXs.

Listed topologies relate to Ethernet-only traffic.

https://t.me/learningnets



Enhanced vPC does not support the following topologies:

- **A dual-homed server that is connected to a pair of FEXs that connects to a single switch**: Although this topology becomes a functioning system when one switch has failed, it is not recommended in normal operation.
- **A multihomed server that is connected by a port channel to more than two FEXs**: This topology results in increased complexity with little benefit.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Enhanced vPC Scalability

- Scalability of Enhanced vPC is similar to that of the dual-homed FEX
- Enhanced vPC
  - Does not increase number of FEXs
  - Doubles the bandwidth in normal operation
  - Provides resilience

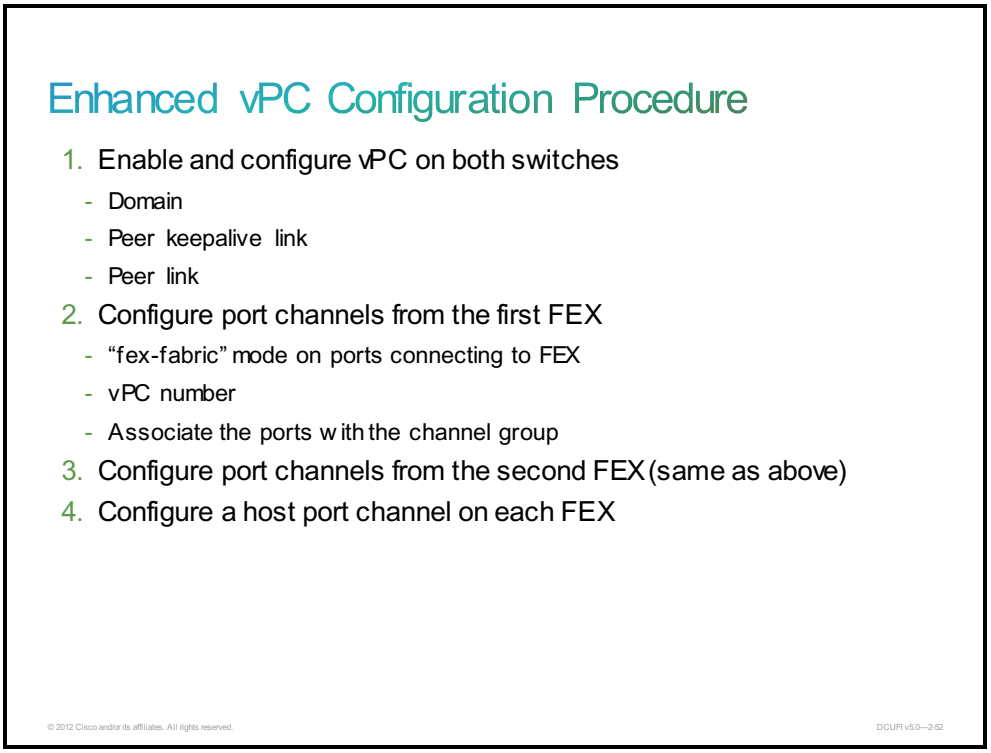
System	Number of FEXs supported with Layer 2-only configuration	Number of FEXs supported with Layer 2/3 configuration
Nexus 5500	24	8
Dual-homed FEX (Enhanced vPC)	24 (managed by a pair of 5500)	8 (managed by a pair of Nexus 5500)

© 2012 Cisco and/or its affiliates. All rights reserved.

The scalability of Enhanced vPC is similar to that of the dual-homed FEX topology.

Each Cisco Nexus 5500 Platform switch supports up to 24 FEXs with no Layer 3 configuration or 8 FEXs with Layer 3 configurations. In a dual-homed FEX topology, such as that in Enhanced vPC, each FEX is managed by two switches so that the pair together can support 24 or 8 FEXs as needed.

https://t.me/learningnets



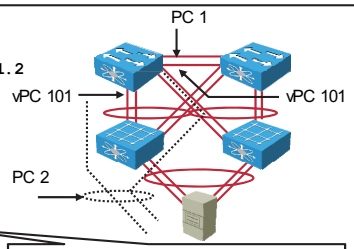
Perform these tasks to implement Enhanced vPC on Cisco Nexus 5500 Platform switches:

1. Enable and configure vPC on both switches. The vPC parameters include the domain ID, peer keepalive link, and peer link.
2. Configure port channels from the first FEX. Within this task, you need to configure fex-fabric mode on ports connecting to the FEX, define the vPC number, and associate the ports with the channel group.
3. Configure port channels from the second FEX. The individual steps are identical to those in Task 2. If you configure the enhanced vPC for Fibre Channel over Ethernet (FCoE) traffic, associate the first FEX to one switch, then associate the second FEX to the other switch.
4. Configure a host port channel on each FEX.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Enhanced vPC Configuration

```
N5500-1(config)# feature vpc
N5500-1(config)# feature lacp
N5500-1(config)# vpc domain 123
N5500-1(config-vpc)# peer-keepalive destination 10.1.1.2
N5500-1(config)# interface eth1/1-2
N5500-1(config-if)# channel-group 1 mode active
N5500-1(config-if)# interface Po1
N5500-1(config-if)# switchport mode trunk
N5500-1(config-if)# vpc peer-link


N5500-1(config)# fex 101
N5500-1(config-fex)# interface eth1/3-4
N5500-1(config-if)# channel-group 101
N5500-1(config-if)# interface po101
N5500-1(config-if)# switchport mode fex-fabric
N5500-1(config-if)# vpc 101
N5500-1(config-if)# fex associate 101

N5500-1(config)# fex 102
N5500-1(config-fex)# interface eth1/5-6
N5500-1(config-if)# channel-group 102
N5500-1(config-if)# interface po102
N5500-1(config-if)# switchport mode fex-fabric
N5500-1(config-if)# vpc 102
N5500-1(config-if)# fex associate 102

N5500-1(config)# interface eth101/1/1, eth101/1/2
N5500-1(config-if)# channel-group 2 mode active
N5500-1(config-if)# interface eth102/1/1, eth102/1/2
N5500-1(config-if)# channel-group 2 mode active
N5500-1(config-if)# int po2
N5500-1(config-if)# switchport access vlan 10
```



1

2

Configure port channel from the first FEX

First FEX port channel is 101

FEX port channel configured for vPC number 101 and associated with FEX

3

Second FEX port channel is 102

4

Configure a host port channel (PC 2) on each FEX

© 2012 Cisco and/or its affiliates. All rights reserved.

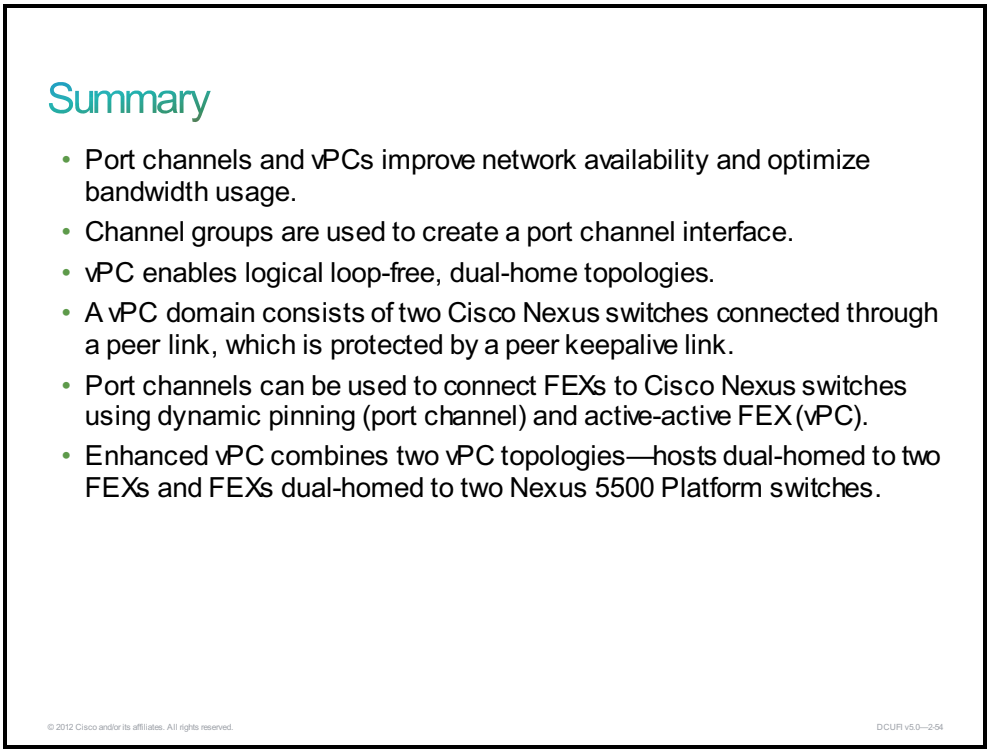
The configuration example depicts some of the components that are required for an Enhanced vPC solution:

- **The port channel that is used for the peer link**: In this case, its ID is “1.”
- **The port channel that is used for the links connecting the first parent switch to the first FEX**: In this case, its ID is “101,” and it is configured for a vPC with the same number and associated with an FEX of the same number.
- **The port channel that is used for the links connecting the first parent switch to the second FEX**: In this case, its ID is “102,” and it is configured for a vPC with the same number and associated with an FEX of the same number.
- **The port channel that groups the links connecting the host to the switch fabric**: In this case, the host has two interfaces that are connected to the first two Ethernet ports on the first FEX (Ethernet 101/1/1-2) and two interfaces that are connected to the first two Ethernet ports on the second FEX (Ethernet 102/1/1-2). The port channel grouping these links has an ID of “2.”

https://t.me/learningnets

Summary

This topic summarizes the key points that were discussed in this lesson.



References

For additional information, refer to these resources:

- To learn more about configuring Cisco Nexus 2000 Series FEX, refer to *Cisco Nexus 2000 Series Fabric Extender Software Configuration Guide* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/nexus2000/sw/configuration/guide/rel_6_0/b_Configuring_the_Cisco_Nexus_2000_Series_Fabric_Extender_rel_6_0.html
- To learn more about configuring port channels, vPCs, and enhanced vPCs on Cisco Nexus 5000 Series switches and Cisco Nexus 5500 Platform switches, refer to *Cisco Nexus 5000 Series NX-OS Layer 2 Switching Configuration Guide, Release 5.1(3)N2(1)* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/nexus5000/sw/layer2/513_N2_1/b_Cisco_n5k_Layer2_Config_513_N2_1.html
- To learn more about configuring port channels and vPCs on Cisco Nexus 7000 Series switches, refer to *Cisco Nexus 7000 Series NX-OS Interfaces Configuration Guide, Release 6.x* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/sw/6_x/nx-os/interfaces/configuration/guide/if_preface.html

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 5

Implementing Cisco FabricPath

Overview

Traditional network architectures have been designed to provide high availability for static applications. Industry trends such as server virtualization and massively scalable distributed applications require more flexibility and scalability. Flexibility is needed in order to be able to move freely between physical data center zones, and greater bandwidth scalability is required to support any-to-any communication. Cisco FabricPath is an innovative Cisco Nexus Operating System (NX-OS) Software technology that can transform the way data center networks are envisioned. Cisco FabricPath brings the benefits of Layer 3 routing to Layer 2 switched networks so as to build a highly resilient and scalable Layer 2 fabric. In this lesson, you will learn how to implement and troubleshoot Cisco FabricPath on the Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches.

Objectives

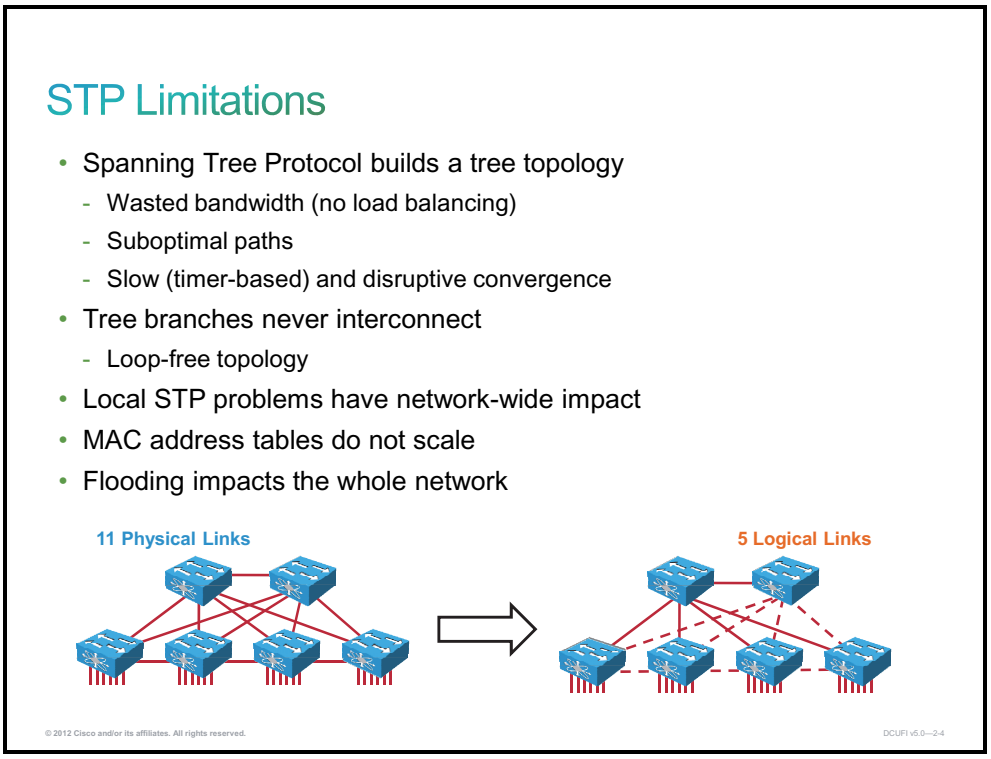
Upon completing this lesson, you will be able to implement and verify Cisco FabricPath on the Cisco Nexus switch. You will be able to meet these objectives:

- Explain how to deploy Cisco FabricPath in Cisco Data Center Network Architecture
- Verify Cisco FabricPath on Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches

https://t.me/learningnets

Implement Cisco FabricPath

This topic explains how to deploy Cisco FabricPath in Cisco Data Center Network Architecture.



Within the data center, most applications require some form of Layer 2 connectivity. Layer 2 is simple to implement and effectively provides a “plug-and-play” scenario for devices being connected to the infrastructure.

The Spanning Tree Protocol (STP) runs on the switches to create a tree-like structure that is loop-free. To provide a loop-free topology, spanning tree builds the tree and then blocks certain ports in order to ensure that traffic cannot loop around the network endlessly. This tree topology implies that certain links are unused, that traffic does not necessarily take the optimal path, and when a failure does occur, that the convergence time is based around timers.

So, current data center designs are a compromise between the flexibility that is provided by Layer 2 and the scaling that is offered by Layer 3:

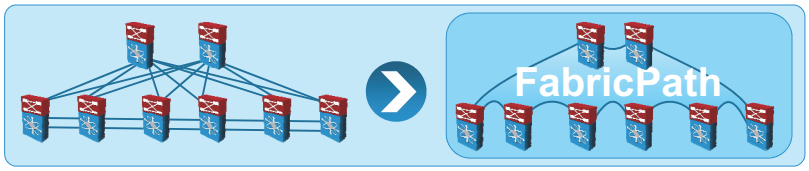
- **Limited scalability**: Layer 2 provides flexibility but it cannot scale. Bridging domains are thus restricted to small areas, strictly delimited by Layer 3 boundaries.
- **Suboptimal performance**: Traffic forwarding within a bridged domain is constrained by spanning-tree rules, thereby limiting bandwidth and enforcing inefficient paths between devices.
- **Complex operation**: Layer 3 segmentation makes data center designs static and prevents them from matching the business agility that is required by the latest virtualization technologies. Any change to the original plan is complicated, the configuration is intensive, and the change is disruptive.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

FabricPath: Ethernet Fabric

- Group of switches using an arbitrary topology
- Externally, the fabric looks like a single switch
  - Switching using the shortest path available
  - No STP inside
  - Single lookup at the ingress identifies the exit point
- Equal Cost Multipathing (ECMP)
  - Up to 256 active links
  - In case of failure traffic redistributed across active links
- Support on Cisco Nexus 5500 and 7000 Series switches
  - Requires Enhanced Layer 2 license
  - On Nexus 7000 available only on F1/F2 series modules



© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco FabricPath is an innovative Cisco NX-OS feature designed to bring the stability and performance of routing to Layer 2. Cisco FabricPath brings the benefits of Layer 3 routing to Layer 2-switched networks in order to build a highly resilient and scalable Layer 2 fabric.

Cisco FabricPath switching allows multipath networking at the Layer 2 level. The Cisco FabricPath network still delivers packets on a best-effort basis (which is similar to the Classic Ethernet network), but the Cisco FabricPath network can use multiple paths for Layer 2 traffic. In a Cisco FabricPath network, you do not need to run the STP. Instead, you can use Cisco FabricPath across data centers—some of which have only Layer 2 connectivity—without the need for Layer 3 connectivity and IP configurations.

Externally, a fabric looks like a single switch, but internally there is a protocol that adds fabric-side intelligence. This intelligence ties the elements of the Cisco FabricPath infrastructure together.

Frames are forwarded along the shortest path possible to their destination, thereby reducing the latency of the exchanges between end stations when compared to a spanning tree-based solution.

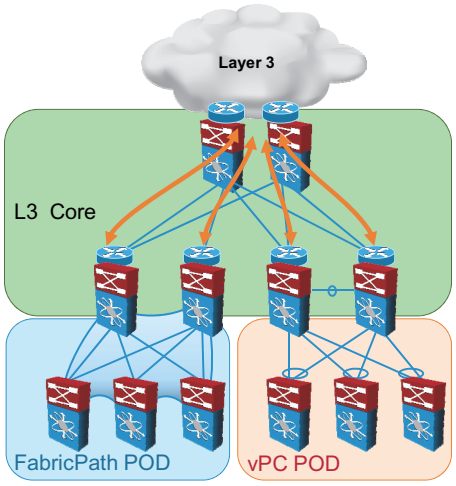
MAC addresses are learned selectively at the edge, thus allowing the network to scale beyond the limits of the MAC address table of individual switches.

Because Equal-Cost Multipath (ECMP) can be used at the data plane, the network can use all of the links available between any two devices. Cisco FabricPath can perform 16-way ECMP, which, in cases of port channels consisting of 16 10-Gb/s links each, represents a connection of 2.56 Tb/s between switches.

https://t.me/learningnets

Classical POD with FabricPath

- Simple configuration
  - No peer link
  - No switch pairs
  - No port channels
- Design flexibility
  - Easily extendible
- No STP
  - No traditional bridging
  - No topology changes
  - No risk of loops



© 2012 Cisco and/or its affiliates. All rights reserved.

You can deploy the Cisco FabricPath solutions in selected parts of the network and achieve migration scalability by migrating one part of the network after another. The figure illustrates a classic pod that is migrated to the Cisco FabricPath technology. This topology offers these benefits for the pod design:

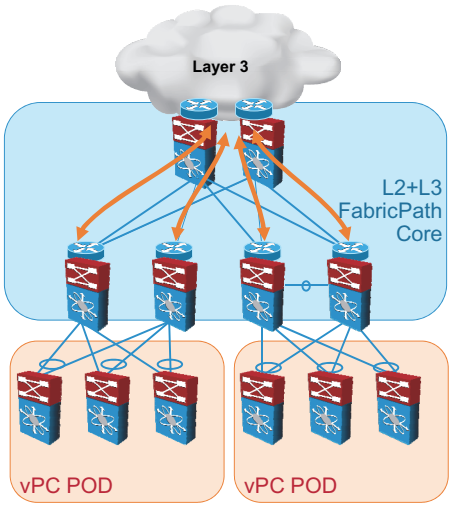
- Because of its simple configuration, Cisco FabricPath removes the requirement for deploying peer links, switch pairs, and port channels in order to achieve scalable bandwidth.
- Design flexibility allows the solution to be easily extendible.
- STP is not used within the Cisco FabricPath cloud. Traditional bridging features, such as mechanisms related to topology changes, are not deployed. The risk of loops is thereby mitigated.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

FabricPath in Core

- Efficient pod interconnect
- VLANs can terminate at the distribution or extend between PODs
- STP is not extended between PODs
- Remote PODs or even remote data centers can be aggregated
- Bandwidth or scale can be introduced in a non-disruptive way



© 2012 Cisco and/or its affiliates. All rights reserved.

You can deploy Cisco FabricPath in the network core. There are two main design options:

- The core can route between the pod subnets. In this case, VLANs terminate at the distribution. The distribution switches that route the traffic into the core can be configured with the vPC enhancement for Cisco FabricPath, which is called virtual port channel+ (vPC+). vPC+ is explained later in this lesson.
- The core can provide transparent Layer 2 connectivity between the pods. In this case, the VLANs are extended between the pods. The boundary routers connecting the core to the external network will route the traffic.

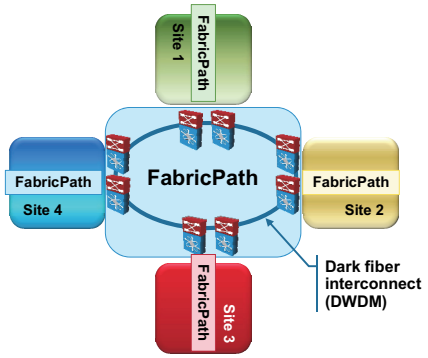
Either option offers an efficient pod interconnect with these characteristics:

- STP is not extended between PODs
- Remote PODs or even remote data centers can be aggregated
- Bandwidth or scale can be introduced in a nondisruptive way

https://t.me/learningnets

FabricPath as Site Interconnect

- Uses dark fiber
- Arbitrary interconnect topology
  - Not dependent of port channels
  - Any number of sites
- High bandwidth, fast convergence
- STP isolation
  - STP running within sites
  - STP terminated on edge ports
- MAC address scalability
  - No MAC learning
  - On-demand learning
- VLANs can be selectively extended while others can be terminated and routed over the interconnect



© 2012 Cisco and/or its affiliates. All rights reserved.

An ideal use case for the Cisco FabricPath technology is a site interconnect, where a high-bandwidth and low-delay fabric links multiple sites or networks. Typically, dark fiber is deployed as the transport medium. Cisco FabricPath supports any arbitrary interconnect topology that does not require port channels for bandwidth scalability.

The STP domains in each site are terminated on Cisco FabricPath edge ports and are thus isolated from one another.

Cisco FabricPath provides a scalable MAC address learning scheme that is demand-driven and creates smaller MAC address tables that are comparable to traditional learning.

Similarly to the core Cisco FabricPath design, the interconnect can route the traffic between the sites or provide transparent Layer 2 bridging. You can combine both approaches by selectively extending some VLANs while terminating and routing others.

https://t.me/learningnets

FabricPath Components

- IS-IS
- Interaction with STP
- FabricPath and Classic Ethernet VLANs
- Traffic encapsulation
  - Send/receive traffic with FabricPath header
- FabricPath routing
  - Forwarding based on 'Switch ID Table'
- Conversational MAC learning
- Multidestination trees
- vPC+
- Layer 3 Integration
- Multicast

© 2012 Cisco and/or its affiliates. All rights reserved.

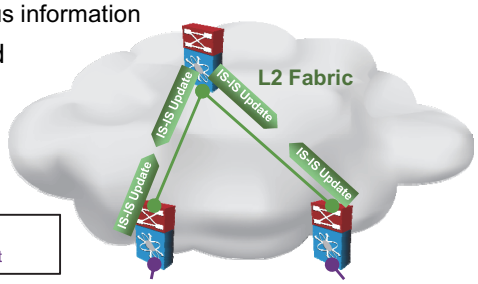
This lesson discusses these Cisco FabricPath components:

- IS-IS
- Interaction with STP
- Cisco FabricPath and Classic Ethernet VLANs
- Traffic encapsulation
- Send/receive traffic with Cisco FabricPath header
- Cisco FabricPath routing
- Forwarding based on "Switch ID Table"
- Conversational MAC learning
- Multidestination trees
- vPC+
- Layer 3 integration
- Multicast

https://t.me/learningnets

IS-IS

- IS-IS replaces STP as control-plane protocol
- Link-state protocol with support for Layer 2 ECMP
- Exchanges reachability of switch IDs and builds forwarding trees
  - SPF routing
- No IP dependency
  - No need for IP reachability to form adjacencies
- Easily extensible
  - Custom TLVs can exchange various information
- Minimal IS-IS knowledge required
  - No user configuration necessary
  - Maintains plug-and-play nature of Layer 2



© 2012 Cisco and/or its affiliates. All rights reserved.

For Cisco FabricPath, you will use the Layer 2 Intermediate System-to-Intermediate System (IS-IS) protocol for a single control plane that functions for unicast, broadcast, and multicast packets. This Cisco FabricPath Layer 2 IS-IS is a separate process from Layer 3 IS-IS.

IS-IS provides these main benefits:

- **Has no IP dependency**: No need for IP reachability in order to form adjacency between devices
- **Easily extensible**: Using custom TLVs, IS-IS devices can exchange information about virtually anything
- **Provides Shortest Path First (SPF) routing**: Excellent topology building and reconvergence characteristics

The interfaces in a Cisco FabricPath network run only the Cisco FabricPath Layer 2 IS-IS protocol. You do not need to run STP in the Cisco FabricPath network because Cisco FabricPath Layer 2 IS-IS discovers topology information dynamically.

Cisco FabricPath Layer 2 IS-IS is a dynamic link-state routing protocol that detects changes in the network topology and calculates loop-free paths to other nodes within the network. Each Cisco FabricPath device maintains a link-state database (LSDB) that describes the state of the network; each device updates the status of the links that are next to the device. The Cisco FabricPath device sends advertisements and updates to the LSDB through all of the existing adjacencies. Cisco FabricPath Layer 2 IS-IS protocol packets do not conflict with standard Layer 2 IS-IS packets because the Cisco FabricPath packets go to a different Layer 2 destination MAC address than that used by standard IS-IS for IPv4/IPv6 address families.

The system sends hello packets on the Cisco FabricPath core ports to form adjacencies. After the system forms IS-IS adjacencies, the Cisco FabricPath unicast traffic uses the Equal-Cost Multipathing (ECMP) feature of Layer 2 IS-IS to forward traffic, which provides up to 16 paths for unicast traffic.

https://t.me/learningnets

The control plane Layer 2 IS-IS comes up and runs automatically when you enable Cisco FabricPath on the device. The loop-free Layer 2 IS-IS protocol builds two trees for the topology. One tree carries unknown unicast, broadcast, and multicast traffic, while the second tree carries load-balanced multicast traffic. The system load-balances multicast traffic across both trees.

Cisco FabricPath Layer 2 IS-IS is based on the standard IS-IS protocol, with the following extensions for the Cisco FabricPath environment:

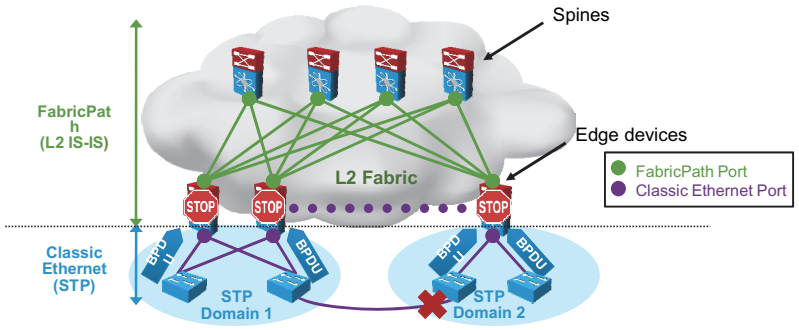
- Cisco FabricPath has a single IS-IS area with no hierarchical Layer 1/Layer 2 routing, as prescribed within the IS-IS standard. All devices within the Cisco FabricPath network are in a single Layer 1 area.
- The system uses a MAC address that is different from the MAC address used for Layer 3 IS-IS instances.
- The system adds a new sub-TLV that carries switch ID information, which is not in standard IS-IS. This feature allows Layer 2 information to be exchanged through the existing IS-IS protocol implementation.
- Within each Cisco FabricPath Layer 2 IS-IS instance, each device computes its shortest path to every other device in the network by using the SPF algorithm. This path is used for forwarding unicast Cisco FabricPath frames. Cisco FabricPath Layer 2 IS-IS uses the standard IS-IS functionality to populate up to 16 routes for a given destination device. The system uses multiple equal-cost available parallel links that provide ECMP.
- FabricPath IS-IS introduces certain modifications to the standard IS-IS in order to support the construction of broadcast and multicast trees (identified by the Forwarding Tags, or FTags). Specifically, using Cisco FabricPath, the system constructs two loop-free trees for forwarding multidestination traffic. Multidestination trees are explained later in the lesson.

By default, you can run Layer 2 IS-IS with Cisco FabricPath with no configuration. However, you can fine-tune some of the Layer 2 IS-IS parameters. Additionally, Cisco FabricPath IS-IS helps to ensure that each switch ID in steady state is unique within the Cisco FabricPath network. If Cisco FabricPath networks merge, switch IDs might collide. If the IDs are all dynamically assigned, Cisco FabricPath IS-IS ensures that this conflict is resolved without affecting any Cisco FabricPath traffic in either network.

https://t.me/learningnets

FabricPath Interaction with STP

- L2 Fabric appears as a single bridge to all connected Classic Ethernet devices
- L2 Fabric should be the root for all connected STP domains
  - Classic Ethernet ports will go into blocking state when 'better BPDU' is received (rootguard)
- No BPDUs are forwarded across the fabric
  - Terminated on Classic Ethernet ports



© 2012 Cisco and/or its affiliates. All rights reserved.

In Cisco FabricPath topologies, there are two types of functions:

- Edge (or leaf) devices: These devices have ports that are connected to Classic Ethernet devices (servers, router ports, and so on) and ports that are connected to the Cisco FabricPath cloud (or Cisco FabricPath ports). Edge devices are able to map a MAC address to the destination switch ID.
- Spine devices: These devices exclusively interconnect edge devices. Spine devices switch exclusively based on the destination switch ID, which is explained later.

STP domains do not cross into the Cisco FabricPath network.

You must configure the Cisco FabricPath edge switches to have the lowest STP priority of all the devices in the STP domain to which it is attached. This ensures that they become the root for any attached STP domains. You should also configure all the Cisco FabricPath edge switches with the same priority. The system assigns the bridge ID for the Layer 2 gateway devices from a pool of reserved MAC addresses.

Other than configuring the STP priority on the Cisco FabricPath Layer 2 gateway switches, you do not need to configure anything for the STP to work seamlessly with the Cisco FabricPath network. Only connected Classic Ethernet devices form a single STP domain. Those Classic Ethernet devices that are not interconnected form separate STP domains, as shown in the figure.

All Classic Ethernet interfaces should be designated ports, which occur automatically, or else they will be pruned from the active STP topology.

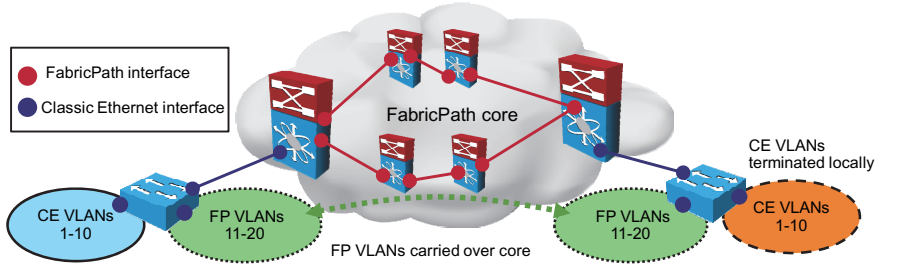
The Cisco FabricPath edge switches propagate the topology change notifications (TCNs) on all of the Classic Ethernet interfaces. The devices in the separate STP domains need to know the TCN information only for the domains to which they belong. You can configure a unique STP domain ID for each separate STP domain that connects to the same Cisco FabricPath network. The Layer 2 IS-IS messages carry the TCNs across the Cisco FabricPath network. Only those Cisco FabricPath Layer 2 gateway switches in the same STP domain as the TCN message need to act and propagate the message to connected Classic Ethernet devices.

https://t.me/learningnets

FabricPath and Classic Ethernet VLANs

VLAN property	Classic Ethernet VLAN (default)	FabricPath VLAN
Scope	Individual site	End-to-end (FabricPath core and sites)
Header	802.3 Ethernet	802.3 Ethernet in STP site Extended header when in FabricPath core

Interface property	Classic Ethernet Interface	FabricPath Interface
Placement	NICs and traditional network devices in STP sites	Internal to FabricPath cloud
Transported VLANs	Classic Ethernet and FabricPath VLANs	FabricPath VLANs



© 2012 Cisco and/or its affiliates. All rights reserved.

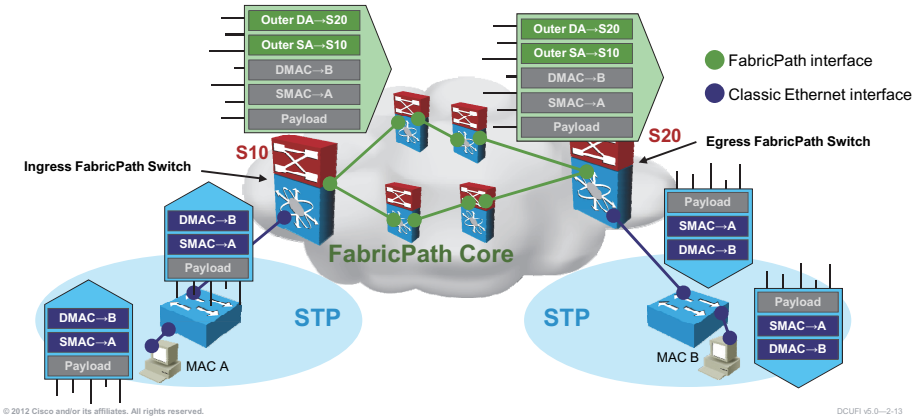
To interact with the Classic Ethernet network, you will set VLANs to either Classic Ethernet or Cisco FabricPath mode. The Classic Ethernet VLANs carry traffic from the Classic Ethernet hosts to the Cisco FabricPath interfaces, and the Cisco FabricPath VLANs carry traffic throughout the Cisco FabricPath topology. Only the active Cisco FabricPath VLANs configured on a switch are advertised as part of the topology in the Layer 2 IS-IS messages.

All VLANs that are meant to be forwarded over the Cisco FabricPath network must be created as Cisco FabricPath VLANs. A VLAN needs to be explicitly configured for Classic Ethernet mode or for Cisco FabricPath mode. By default, all VLANs are in Classic Ethernet mode.

https://t.me/learningnets

FabricPath Data Encapsulation

Element	Description
Ingress FabricPath switch	• Determines destination Switch ID • Encapsulates frame in FabricPath header
Outer destination/source address	• Contains destination/source switch ID • Destination address used for routing through Cisco FabricPath core
Forwarding in FabricPath core	No MAC learning or lookups required inside core



© 2012 Cisco and/or its affiliates. All rights reserved.

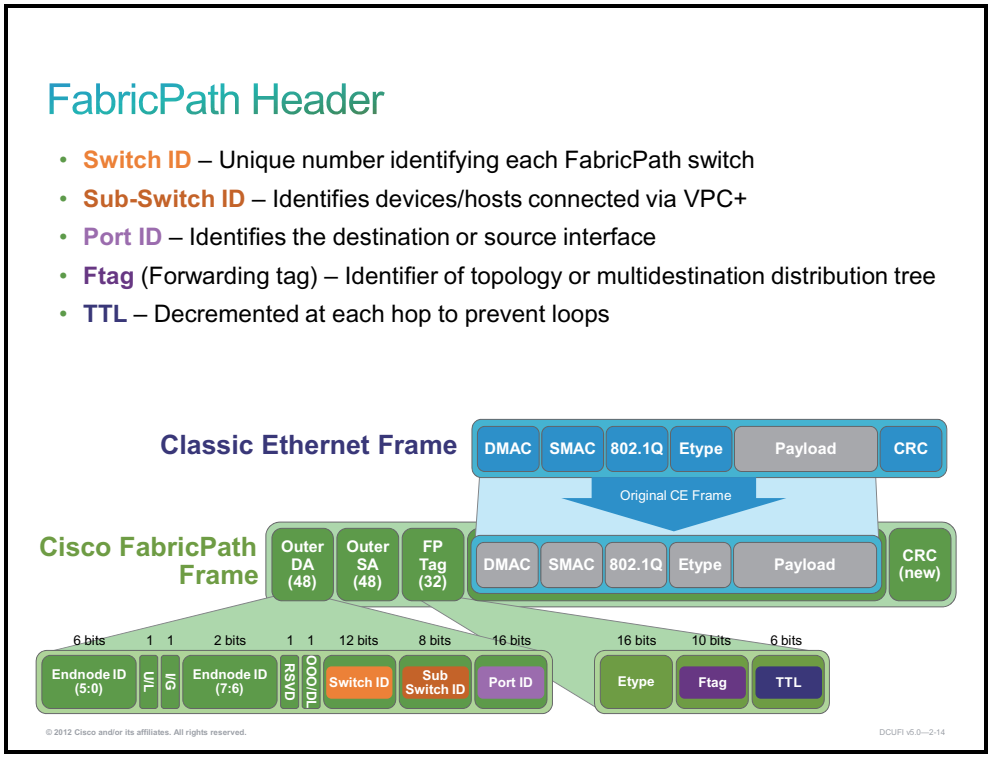
When a frame enters the Cisco FabricPath network on a Cisco FabricPath VLAN, the system encapsulates the Layer 2 frame with a new Cisco FabricPath header. The outer destination address (ODA) and outer source address (OSA) in the Cisco FabricPath header contain the switch IDs of the egress and ingress switch, respectively.

The system applies the encapsulation on the ingressing edge port of the Cisco FabricPath network and de-encapsulates the frame on the egressing edge port of the Cisco FabricPath network. All of the ports within the Cisco FabricPath network are Cisco FabricPath ports that use only the hierarchical MAC address. This feature greatly reduces the size of the MAC tables in the core of the Cisco FabricPath network.

The system automatically assigns each device in the Cisco FabricPath network with a unique switch ID. Optionally, you can configure the switch ID for the Cisco FabricPath device yourself.

The OSA is the Cisco FabricPath switch ID of the device where the frame ingresses the Cisco FabricPath network, and the ODA is the Cisco FabricPath switch ID of the device where the frame egresses the Cisco FabricPath network. When the frame egresses the Cisco FabricPath network, the Cisco FabricPath device strips the Cisco FabricPath header, and the original Classic Ethernet frame continues on the Classic Ethernet network. The Cisco FabricPath network uses only the OSA and ODA, with the Layer 2 IS-IS protocol transmitting the topology information. Both the Cisco FabricPath ODA and OSA are in a standard MAC format (xxxx.xxxx.xxxx).

https://t.me/learningnets



The figure illustrates the encapsulation process and the outer header that is used for transporting the frame through the Cisco FabricPath cloud. The Cisco FabricPath encapsulation uses a MAC address-in-MAC address encapsulation format. The original Ethernet frame, along with an IEEE 802.1Q tag, is prepended by a 48-bit outer source address (OSA), a 48-bit ODA, and a 32-bit Cisco FabricPath tag.

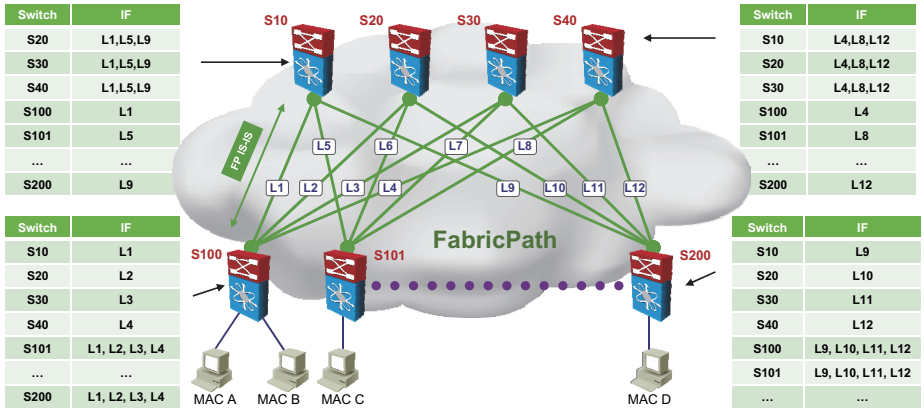
In addition to the switch ID, the Cisco FabricPath header addresses contain these fields:

- The subswitch ID (sSID) field identifies the source or destination vPC+ PortChannel interface associated with a particular vPC+ switch pair. Cisco FabricPath switches running vPC+ use this field to identify the specific vPC+ PortChannel upon which traffic is to be forwarded. The sSID value is locally significant to each vPC+ switch pair. In the absence of vPC+, this field is set to 0.
- The port ID, also known as the Local Identifier (LID), identifies the specific physical or logical interface on which the frame was sourced or to which it is destined. The value is locally significant to each switch. This field in the ODA allows the egress Cisco FabricPath switch to forward the frame to the appropriate edge interface without requiring a MAC address table lookup. For frames sourced from or destined to a vPC+ PortChannel, this field is set to a common value shared by both vPC+ peer switches, and the sSID is used to select the outgoing port instead.
- The EtherType value for Cisco FabricPath encapsulated frames is 0x8903.
- The function of the FTag depends on whether a particular frame is unicast or multidestination. In the case of unicast frames, the FTag identifies the Cisco FabricPath topology the frame is traversing. In the case of multidestination frames, the FTag identifies the multidestination forwarding tree that the frame should traverse.
- The Time to Live (TTL) field serves the same purpose as in traditional IP forwarding: Each switch hop decrements the TTL by 1, and frames with an expired TTL are then discarded. This prevents Layer 2 bridged frames from looping endlessly if a transitory loop occurs. Ingress Cisco FabricPath edge switches set the TTL to 32 for all frames.

https://t.me/learningnets

FabricPath Routing Table

- FabricPath IS-IS manages Switch ID (routing) table
- Equal-cost path selection based on ECMP hash function
  - Maximum 16 (default) next-hop interfaces for each destination Switch ID
  - Number controlled by maximum-paths command in FabricPath IS-IS process



© 2012 Cisco and/or its affiliates. All rights reserved.

The IS-IS protocol establishes switch ID tables that enable the routing of Cisco FabricPath frames through the Cisco FabricPath network. The tables describe all available shortest paths to a given switch ID. Frames that traverse the Cisco FabricPath network carry the destination switch ID in the ODA. The transit switches (spines) look up the destination switch ID in the switch ID table and forward the frame along the selected multidestination tree (identified with FTag) towards the destination edge switch.

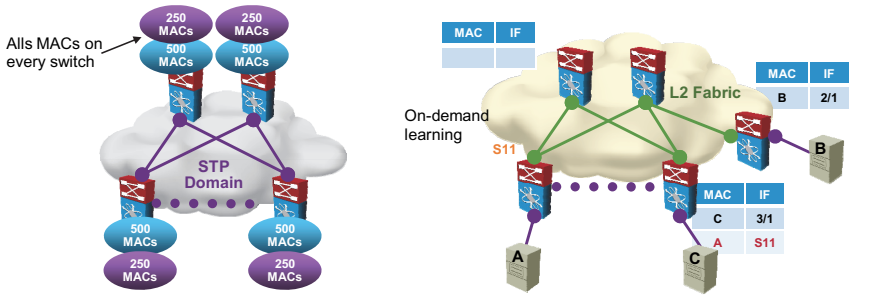
Cisco FabricPath, using Layer 2 IS-IS, can utilize up to 16 active Layer 2 paths for forwarding known unicast packets. Forwarding of broadcast and multicast packets is constrained to a specific multidestination tree.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Conversational MAC Learning

MAC address type	FabricPath learning	Traditional MAC learning
Local (in local site)	• When traffic received on Classic Ethernet ports • Learned from source MAC address	Learned when frame from the MAC address is seen
Remote (in remote site)	• When traffic received on FabricPath ports • Learned from source MAC only if destination MAC is already known as local • Broadcast and unknown unicast do not populate FabricPath MAC table	



© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco FabricPath edge switches maintain two important forwarding tables—a traditional MAC address table and a switch identifier (SID) table. The MAC address table in Cisco FabricPath edge switches resembles the MAC address table that is used in Classic Ethernet, but there are some important differences.

In Classic Ethernet, when a switch receives an Ethernet frame, it unconditionally populates the MAC address table with the source MAC address of the frame. Additionally, forwarding in a Classic Ethernet switch is always based on the destination MAC address. If the MAC address is already learned, the frame is then constrained to the port on which that MAC address was learned. If the MAC address is unknown, or if the frame is a broadcast frame, the frame is flooded to all ports in the VLAN on which it was received.

The side effect of this behavior in Classic Ethernet networks is that every switch that has a port in a particular VLAN will learn every MAC address within that VLAN. One potential downside of this behavior is that MAC address tables can become saturated with entries that are never used, and the overall MAC address scalability of the network can be limited by the size of the smallest MAC address table that is supported among all of the switches.

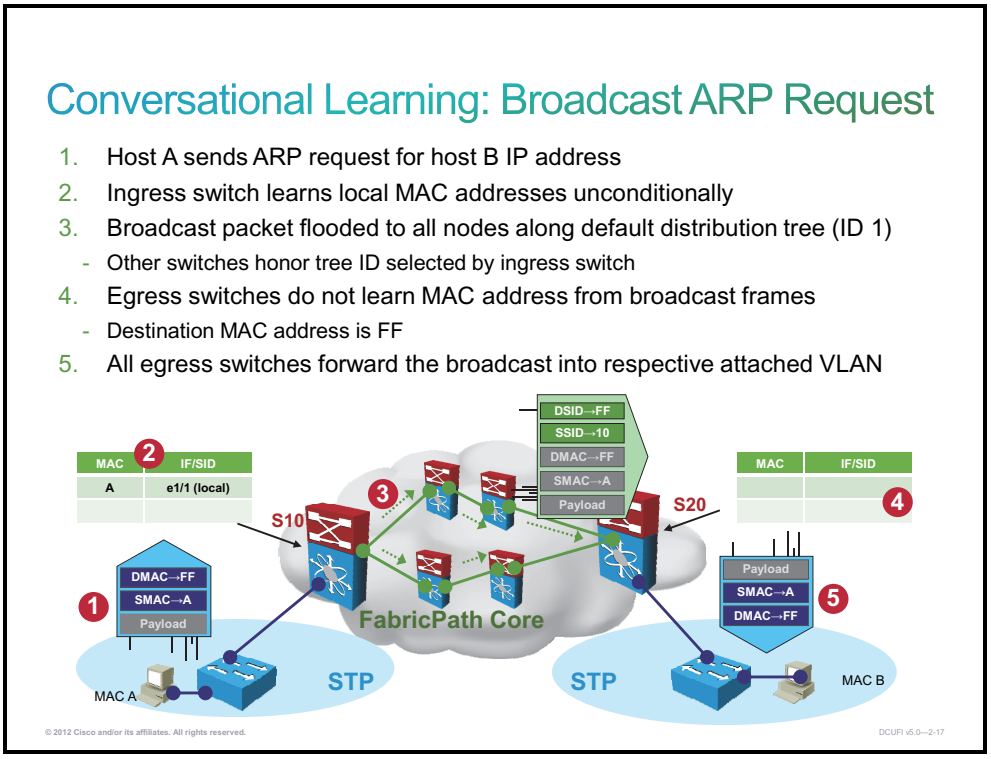
In contrast, Cisco FabricPath introduces new MAC address learning rules that optimize the learning process within the fabric and help conserve MAC address table space on the edge switches. This technique, which is known as conversational learning, occurs automatically in VLANs configured for Cisco FabricPath mode.

The first general rule of Cisco FabricPath MAC learning is that only Cisco FabricPath edge switches populate the MAC address table and use MAC address table lookups in order to forward frames. Cisco FabricPath core switches do not learn any MAC addresses at all. Rather, all frame forwarding within the fabric is based on the ODA of the Cisco FabricPath header.

Each Cisco FabricPath edge switch distinguishes between two types of MAC address entries:

- Local MAC address entries are created for devices that are directly connected to the switch.
- Remote MAC address entries are created for devices that are connected to a different Cisco FabricPath switch.

https://t.me/learningnets



Cisco FabricPath switches follow these MAC address learning rules:

1. For Ethernet frames received from a directly connected access or trunk port, the switch unconditionally learns the source MAC address as a local MAC address entry, much as a STP switch would.

2. For unicast frames received with Cisco FabricPath encapsulation, the switch learns the source MAC address of the frame as a remote MAC address entry only if the destination MAC address matches an already learned local MAC address entry. In other words, the switch learns remote MAC addresses only if the remote device is having a bidirectional conversation with a locally connected device. Unknown unicast frames being flooded in the Cisco FabricPath network do not necessarily trigger learning on edge switches.

3. In addition, broadcast frames do not trigger learning on edge switches. However, broadcast frames are used to update any existing MAC address entries already in the table. For example, if a host moves from one switch to another and sends a Gratuitous Address Resolution Protocol (ARP) message to update the Layer 2 Forwarding (L2F) tables, Cisco FabricPath switches receiving that broadcast will update an existing entry for the source MAC address.

4. Multicast frames (whether IP or non-IP multicast) also trigger learning on edge switches since several critical LAN protocols, such as Hot Standby Routing Protocol (HSRP), rely on source MAC address learning from multicast frames in order to facilitate proper forwarding.

This figure illustrates this process with an ARP request:

Step 1 Host A wants to communicate with Host B, another device within the same IP subnet. Host A therefore transmits an ARP request for the IP address of Host B. This frame is a standard Ethernet frame with the source MAC address of Host A and an all-ones broadcast destination MAC address (FFFF.FFFF.FFFF).

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Step 2 The Cisco FabricPath edge switch S10 receives the frame on the edge port e1/1 in VLAN 10, which is configured for Cisco FabricPath mode. S10 performs both a source and destination MAC address table lookup in VLAN 10 for the frame.

— The source MAC address lookup for {VLAN 10, MAC A} returns a miss, causing the forwarding engine on S100 to unconditionally learn MAC A as a local MAC address entry for port e1/1 in VLAN 10. MAC A is learned only on the forwarding engine ASIC associated with e1/1. Other forwarding engines in the system do not learn MAC A.

Step 3 The destination MAC address lookup indicates that the frame is broadcast, causing the forwarding engine to flood the frame in VLAN 10. Any additional edge ports in VLAN 10 on S10 receive the frame. In addition, S10 selects the first multidestination tree (Tree 1) to forward the broadcast frame. As a rule, Cisco FabricPath switches use Tree 1 to forward all broadcast, non-IP multicast and unknown unicast traffic. The Cisco FabricPath header for the frame consists of the following parameters:

— **Outer destination address**: The outer destination address for a broadcast frame uses the same MAC address as the inner frame, meaning the all-ones broadcast the MAC address.

— **Outer Source Address (OSA)**: The outer source address has the SID 10. The sSID is 0 and a local ID set to a locally significant value is associated with e1/1 on S10.

— **Cisco FabricPath tag**: The EtherType is Cisco FabricPath (0x8903), the FTag is 1 (identifying multidestination Tree 1), and the TTL is 32 (default).

Note Because the frame is broadcast, the spines in the core use the FTag value that is already populated by S10 to identify which multidestination tree the frame is traversing.

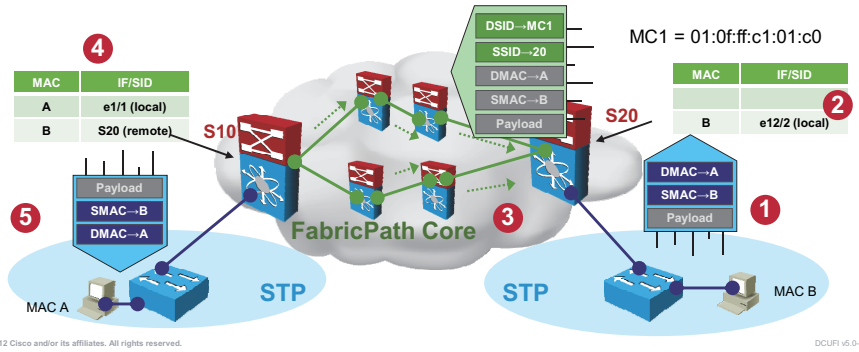
Step 4 The frame arrives on the egress switch S20. The switch then removes the Cisco FabricPath header and floods the original broadcast Ethernet frame on those ports. No MAC address learning occurs on S20 based on these actions.

Step 5 The broadcast ARP request from Host A to Host B is received by Host B. However, the only switch in the Cisco FabricPath network that the learned MAC A based on the SMAC address of the broadcast frame is S10.

https://t.me/learningnets

Conversational Learning: Unicast ARP Reply

1. Host B responds with ARP reply
2. Nearest switch (S20) learns local MAC address unconditionally
3. Unknown unicast flooded to all nodes along default distribution tree
   - Outer destination MAC set to well-known “flood to fabric” multicast address (MC1)
4. Remote switch (S10) learns source MAC address
   - Accepted because destination MAC (A) already in the MAC table
5. All egress switches flood unknown unicast into respective attached VLAN



© 2012 Cisco and/or its affiliates. All rights reserved.

The unicast ARP reply is processed in this way:

Step 1 Having received a broadcast ARP request from Host A, Host B replies with a unicast ARP reply. This frame is a standard Ethernet frame with the SMAC address of Host B and the unicast destination MAC address of Host A.

Step 2 S20 receives the frame on a Cisco FabricPath edge port in VLAN 10, which is configured for Cisco FabricPath mode. S20 performs both a source and destination MAC address table lookup in VLAN 10 for the frame.

— The source MAC address lookup for {VLAN 10, MAC B} returns a miss, causing the forwarding engine on S20 to unconditionally learn MAC B as a local MAC address entry for port e12/2 in VLAN 10.

Step 3 The destination MAC address lookup for {VLAN 10, MAC A} also returns a miss, causing the forwarding engine to flood the frame in VLAN 10 as an unknown unicast. Any additional edge ports in VLAN 10 on S20 receive the frame.

— S20 selects the first multidestination tree (Tree 1) to forward the unknown unicast frame. Having selected Tree 1, S20 performs a multidestination lookup for Tree 1 in order to determine on which interfaces the frame must be flooded. S20 floods the original unknown unicast frame on those links, encapsulated in a new Cisco FabricPath header. The outer destination address for an unknown unicast frame uses a reserved multicast MAC address (called MC1: 010F.FFC1.01C0). The FTag is 1.

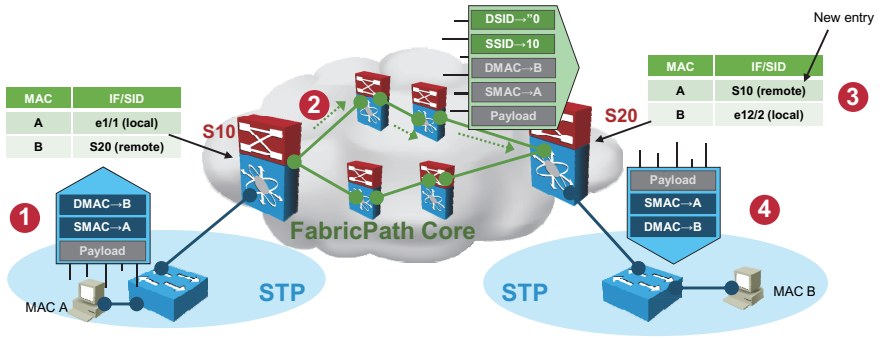
Step 4 S10 receives the encapsulated frame. On the forwarding engine ASIC on which host A is attached, since the inner DMAC address (MAC A) is already known as a local MAC address entry, the inner SMAC address (MAC B) is learned with an SID, sSID, and local ID.

Step 5 S10 and all other egress switches remove the Cisco FabricPath header and flood the original unicast Ethernet frame on the egress ports in VLAN 10.

https://t.me/learningnets

Conversational Learning: Unicast Traffic

1. Host A sends unicast packet to host B
2. Ingress switch forwards the unicast along the shortest path
   - FabricPath MAC table complete after seeing ARP reply
3. Egress switch learns source MAC address
4. Egress switch forwards packet to host B



© 2012 Cisco and/or its affiliates. All rights reserved.

Step 1 Having received a unicast ARP reply from Host B, Host A can now transmit a unicast data frame to Host B. This frame is a standard Ethernet frame with the SMAC address of Host A and the unicast DMAC address of Host B.

Step 2 S10 receives the frame on a Cisco FabricPath edge port in VLAN 10, which is configured for Cisco FabricPath mode. S10 performs both a source and destination MAC address table lookup in VLAN 10 for the frame.

— The source MAC address lookup for {VLAN 10, MAC A} returns a hit, causing the forwarding engine on S10 to update the aging timer of the local entry for MAC A.

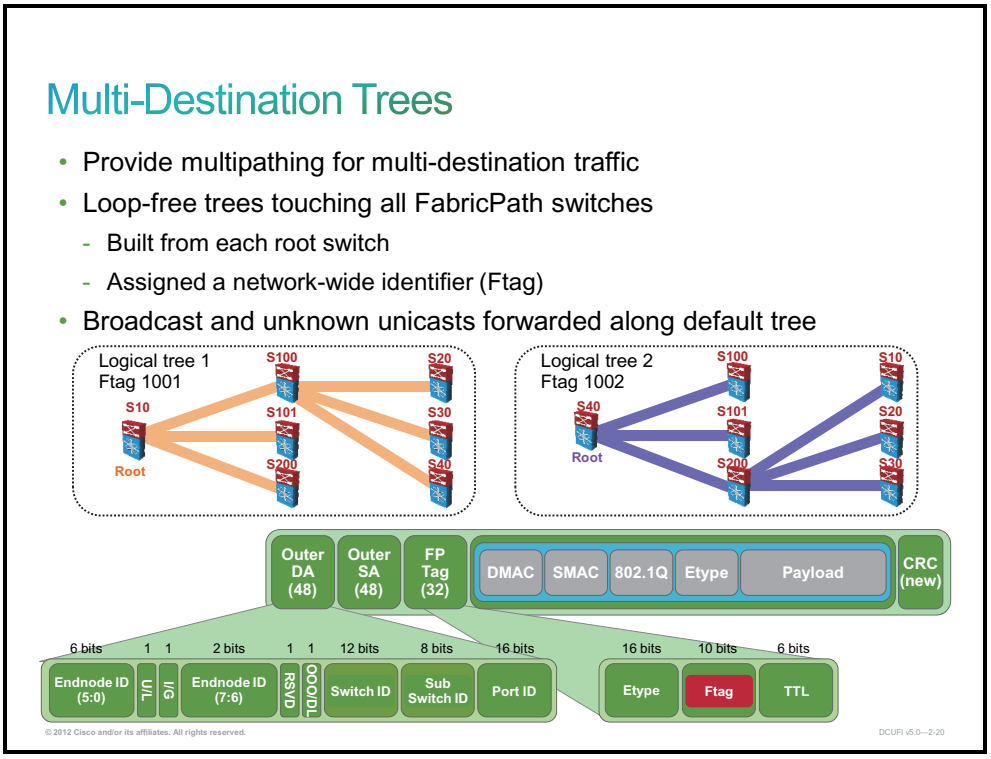
— The destination MAC address lookup for {VLAN 10, MAC B} also returns a hit, returning the SID, sSID, and local ID associated with MAC B.

— The forwarding engine performs a routing lookup for S20. If it has multiple next-hop interfaces through which S20 is reachable, the forwarding engine on S10 will then use a hash function to select one of the available paths. (The default is the source and destination IP addresses plus Layer 4 ports.) The frame will be sent in topology 1.

Step 3 S20 receives the Cisco FabricPath encapsulated frame and performs a routing lookup that is based on the destination SID contained in the ODA. Because the lookup indicates that S20 is the egress Cisco FabricPath switch, S20 uses the sSID and LID to determine on which physical edge interface the frame should be forwarded (in this case, the LID value identifies interface e12/2).

Step 4 On the forwarding engine ASIC to which Host B is attached, because the inner DMAC address (MAC B) is already known as a local MAC address entry, the inner SMAC address (MAC A) is learned with a SID, sSID, and local ID.

https://t.me/learningnets



When a Cisco FabricPath edge switch receives a multidestination frame on an edge port, it selects one of the available multidestination trees to forward the frame. The tree that is selected depends on the type of multidestination frame the switch is forwarding:

- **Broadcast, unknown unicast, non-IP multicast**: These frames are primarily forwarded on the first tree in the topology, Tree 1. There is, however, an exception to this rule: In a vPC+ environment, each tree has an affinity for one or the other peer switch. In this situation, broadcast, unknown unicast, and non-IP multicast frames traverse the first tree for which the particular peer switch has affinity.
- **IP multicast**: Cisco FabricPath edge switches use a hash function to select a multidestination tree for IP multicast frames. Therefore, IP multicast frames can traverse any of the available multidestination trees.

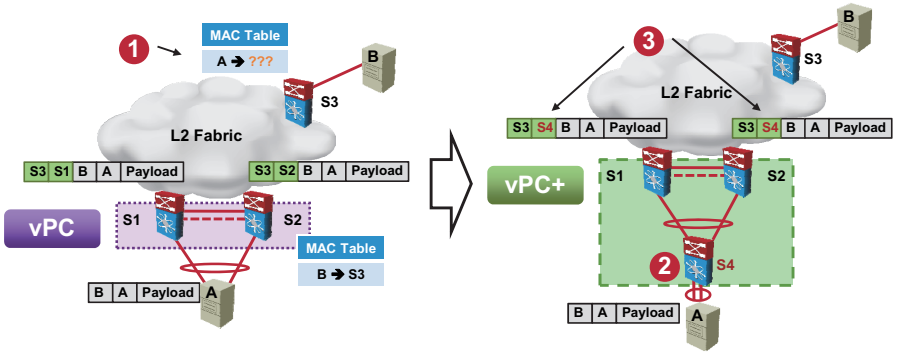
After the switch determines which multidestination tree a frame will traverse, it encapsulates the frame in a Cisco FabricPath header, populating the ODA with the appropriate address and populating the FTag field in the Cisco FabricPath tag with the unique value associated with that particular multidestination tree.

In a generic sense, the frame is then flooded on any Cisco FabricPath interfaces that belong to that tree. Other Cisco FabricPath switches receiving the frame will then further flood the frame that is based on the topology of that particular multidestination tree.

https://t.me/learningnets

vPC Enhancement for FabricPath

1. MAC table only allows 1-to-1 mapping between MAC and Switch ID
2. vPC+ introduces a 'virtual switch'
   - For each vPC domain
   - Represented by an unique 'Virtual Switch' ID to the rest of L2 Fabric
3. Virtual switch ID used as source in FabricPath encapsulation



© 2012 Cisco and/or its affiliates. All rights reserved.

The Cisco FabricPath MAC address table maps MAC addresses to switch IDs. This mapping allows a MAC address to be bound to a single switch ID only. With vPCs, the host MAC addresses are reachable via two egress switches configured as vPC peers. Such mapping cannot exist in the Cisco FabricPath MAC address table.

vPC+ is an extension of vPCs that provides the solution by creating a unique virtual switch that appears as a separate device to the rest of the Cisco FabricPath network. A vPC+ provides active-active Layer 2 paths for dual-homed Classic Ethernet devices or clouds, even though the Cisco FabricPath network allows only 1-to-1 mapping between the MAC address and the switch ID.

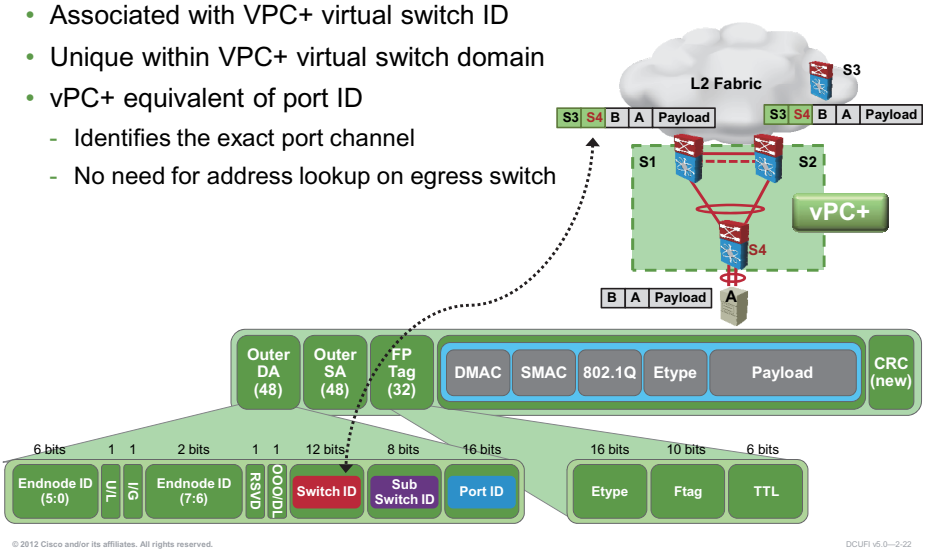
The Cisco FabricPath switch ID for the virtual switch becomes the MAC OSA in the Cisco FabricPath encapsulation header. Each vPC+ domain must have its own virtual switch ID. Layer 2 multipathing is achieved by emulating a single virtual switch. Packets that are forwarded from Host A to Host B are tagged with the MAC address of the virtual switch as the transit source, and traffic from Host B to Host A is now load-balanced. You must assign the same switch ID to each of the two vPC+ peer devices so that the peer link can form.

You must enable all interfaces in the vPC+ peer link as well as all the downstream vPC+ links for Cisco FabricPath. The vPC+ downstream links will be Cisco FabricPath edge interfaces, which then connect to the Classic Ethernet hosts.

https://t.me/learningnets

Sub-Switch ID

- vPC+ bundle identifier (8 bits)
- Associated with VPC+ virtual switch ID
- Unique within VPC+ virtual switch domain
- vPC+ equivalent of port ID
  - Identifies the exact port channel
  - No need for address lookup on egress switch



© 2012 Cisco and/or its affiliates. All rights reserved.

The sSID, carried in the outer encapsulation header, identifies the vPC+ bundle in a similar fashion to how the port ID identifies an individual edge link. The port ID is not used in a vPC+ scenario.

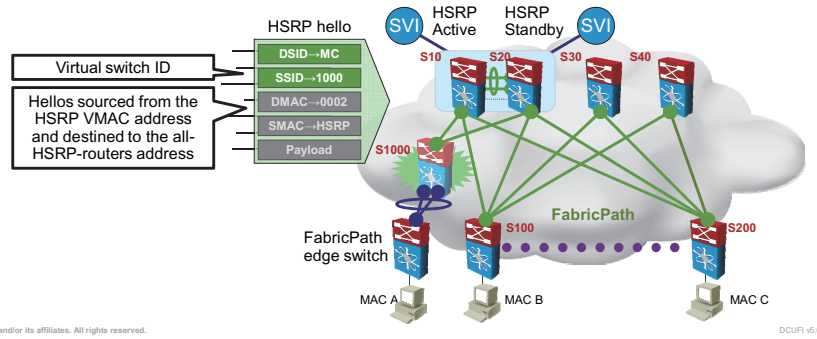
When the egress edge switch receives frame with the sSID set, it then uses this value to identify the outgoing bundle instead of looking it up from the MAC address table.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

vPC+ and Active/Active HSRP

- When HSRP hellos are sent on Cisco FabricPath core ports:
  - Outer SA field contains the vPC+ virtual switch ID
  - FabricPath edge switches learn the HSRP VMAC using the vPC+ virtual switch ID
- Active-active HSRP forwarding function of vPC+ for all devices:
  - Connected to vPC+ PortChannels
  - Connected to native Cisco FabricPath ports
- Traffic destined to HSRP MAC can leverage ECMP if available



© 2012 Cisco and/or its affiliates. All rights reserved.

The First Hop Routing Protocols (FHRP) and the Hot Standby Routing Protocol (HSRP) interoperate with a vPC+. You should dual-attach all Layer 3 devices to both vPC+ peer devices.

The primary FHRP device responds to ARP requests even though the secondary vPC+ device also forwards the data traffic. Both the primary and secondary vPC+ devices forward traffic, but only the primary FHRP device responds to ARP requests.

To simplify initial configuration verification and vPC+/HSRP troubleshooting, you can configure the primary vPC+ peer device with the FHRP active router highest priority.

When the primary vPC+ peer device fails over to the secondary vPC+ peer device, the FHRP traffic continues to flow seamlessly.

You should configure a separate Layer 3 link for routing from the vPC+ peer devices rather than using a VLAN network interface for this purpose.

https://t.me/learningnets

vPC vs. vPC+

- A given VDC can be part of vPC domain, or vPC+ domain, but not both
- vPC+ only works on F1 and F2 modules with FabricPath enabled in the VDC
- Conversion between vPC and vPC+ is disruptive

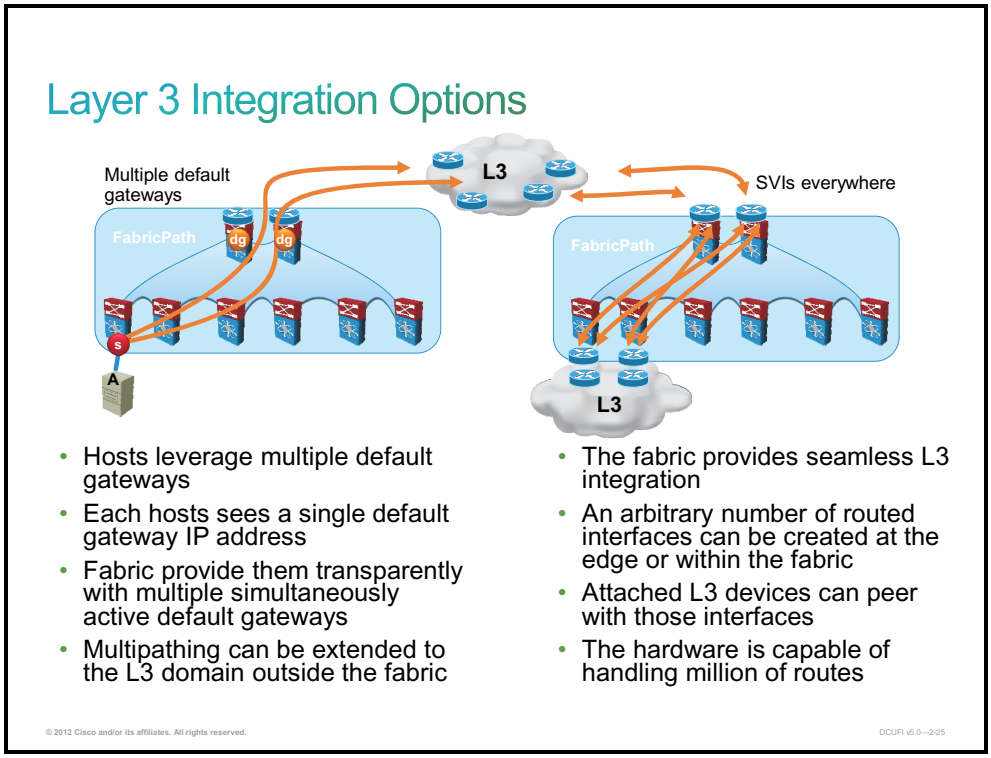
	vPC	vPC+
Peer-link	M1, F1 or F2 ports	F1/F2 ports
Member ports	M1 ports or F1 or F2 ports	F1/F2 ports
VLANs	Classic Ethernet or FabricPath VLANs	FabricPath VLANs only
Peer-link switchport mode	Classic Ethernet trunk port	FabricPath core port

© 2012 Cisco and/or its affiliates. All rights reserved.

The table describes several differences between vPC and vPC+.

On Cisco Nexus 7000 Series switches, Cisco FabricPath interfaces can only be configured on the F (F1 and F2) Series modules. Cisco Nexus F-Series modules have only Layer 2 interfaces. To use routing with a vPC+, you must have an M Series module that is inserted into the same Cisco Nexus 7000 Series chassis. The system then performs proxy routing using both the Cisco Nexus F-Series module and the M Series modules in the chassis.

https://t.me/learningnets



Cisco FabricPath provides a transparent Layer 2 infrastructure that can be integrated with Layer 3 routing in multiple ways. The figure illustrates two main cases of providing end-to-end connectivity:

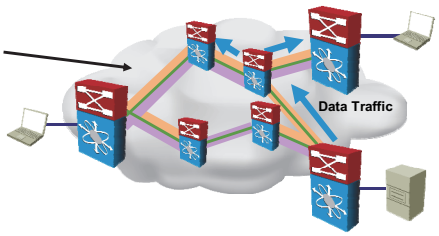
- **Multiple default gateways**: In this scenario, the hosts are attached to a Cisco FabricPath VLAN that provides Layer 2 connectivity to the remote Cisco FabricPath edge. The remote Cisco FabricPath edge provides default gateway functionality towards the external Layer 3 network. The remote Cisco FabricPath edge, if implemented on a Cisco Nexus 7000 Series switch, must have an F Series module for Cisco FabricPath links and an M Series module for routed interfaces. This solution leverages FHRPs, such as GLBP.
- **Multiple Switched Virtual Interfaces (SVIs)**: In this case, the individual sites are connected via Classic Ethernet VLANs to Cisco FabricPath edge devices that route the traffic into Cisco FabricPath VLANs. The traffic then traverses the Cisco FabricPath cloud and reaches Layer 3 next hops on the remote Cisco FabricPath edge. Multiple Layer 3 paths between the edges can be provided by multiple Cisco FabricPath VLANs.

https://t.me/learningnets

FabricPath Multicast

- IGMP snooping operates as usual in FabricPath edge switches
- Receivers are signaled using Group Membership Link State Packets (GM-LSP) in IS-IS
- Edge switch uses a hash function to pick a multidestination tree
  - Hash function is per flow, combining Layer 3, 4, and VLAN ID
- Multicast packets do not necessarily traverse every core port in the selected tree
  - Multicast traffic forwarded only to multicast receivers and routers within the selected tree

FabricPath IS-IS creates a "pruned tree" to constrain multicast data frames to just those switches that have interested receivers



© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco FabricPath enables Layer 2 multicast multipathing. Cisco FabricPath uses a hash-based system to assign each of the multicast flows to one of the two designated trees in order to ensure that the multicast traffic is load-balanced.

The system uses Cisco FabricPath Layer 2 IS-IS and Classic Ethernet Internet Group Management Protocol (IGMP) snooping to learn the multicast group information at the boundaries of the Cisco FabricPath/Classic Ethernet network. The system carries that information through the Cisco FabricPath network by using a new Layer 2 IS-IS link-state packet (LSP) called Group Membership LSP (GM-LSP). GM-LSPs carry multicast group/source membership information. This information is carried across the Cisco FabricPath network. All Cisco FabricPath switches maintain multicast routing information and forward multicast data packets only to switches that have interested receivers. Each node in each Cisco FabricPath topology shares the same view and has all of the same information.

The multicast traffic uses the per-VLAN source, multicast group, and flow information to compute the hash and allocate traffic to one or the other of the two trees. This system constrains multicast traffic based on the group IP address. The resulting multicast distribution tree does not include all ports in the selected topology, but rather only the ports that have receivers that are attached to it.

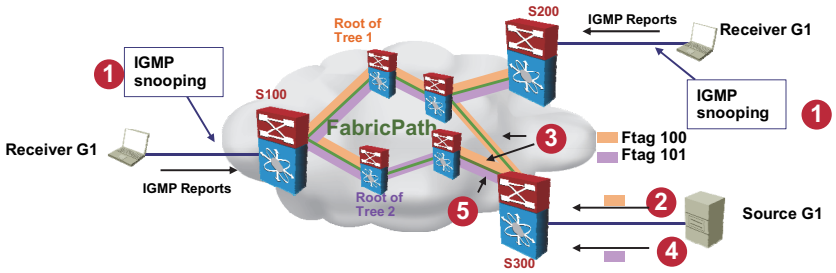
For Layer 2 multicast traffic, you do not need to run PIM at all. Within the Cisco FabricPath cloud, the source- and receiver-related information is transmitted by using the GM-LSPs in IS-IS.

For Layer 3 multicast packets, the system sets the ODA to a special multicast group that identifies all IP routers for that group and then forwards the traffic along the tree for that particular group.

https://t.me/learningnets

FabricPath Multicast Operation

1. Edge switches perform IGMP snooping and signal attached receivers
2. Source sends multicast traffic to group G1 over VLAN 10
3. Edge switch runs hash and determines tree with Ftag 100
   - Multicast forwarded along the tree only to ports with G1 receivers
4. Source sends different UDP traffic to group G1 over VLAN 10
5. Edge switch computes different hash and uses tree with Ftag 101
   - Multicast forwarded along the tree only to ports with G1 receivers



© 2012 Cisco and/or its affiliates. All rights reserved.

This figure explains the main aspects of multicast forwarding:

Step 1 Edge switches perform IGMP snooping and signal the receivers to the cloud via the GM-LSPs.

Step 2 The multicast source sends multicast traffic to group G1 over VLAN 10.

Step 3 The ingress Cisco FabricPath edge switch runs hash and determines the tree that has FTag 100. Multicast traffic is forwarded along the tree with FTag 100, but only to ports with G1 receivers or routers that are attached behind them. The pruning of the unnecessary interfaces in that tree has been done by IS-IS based on the information in the GM-LSPs.

Step 4 The multicast source sends a different UDP traffic to group G1 over VLAN 10. Although the destination group and the VLAN are the same, the hash will yield a different value due to the different Layer 4 parameters.

Step 5 The ingress Cisco FabricPath edge switch computes a different hash and uses the tree that has FTag 101. Multicast packets are forwarded along the tree only to ports with G1 receivers.

https://t.me/learningnets

Transparent Interconnection of Lots of Links (TRILL)

- IETF standard for Layer 2 multipathing
- Driven by multiple vendors, including Cisco
  - Cisco FabricPath capable hardware is also TRILL capable
  - TRILL mode will be provided with a software upgrade
  - Cisco will push FabricPath-specific enhancements to TRILL

Feature	FabricPath	TRILL
Frame routing (ECMP, TTL, RPFC etc…)	Yes	Yes
vPC+	Yes	No
FHRP active/active	Yes	No
Multiple topologies	Yes	No
Conversational learning	Yes	No
Inter-switch links	Point-to-point only	Point-to-point OR shared

© 2012 Cisco and/or its affiliates. All rights reserved.

Transparent Interconnection of Lots of Links (TRILL) is an IETF protocol that provides a Layer 2 multipathing functionality without having to use spanning tree as a mechanism to find loop-free trees within a Layer 2 broadcast domain.

TRILL computes the topology of the cloud and forwards Layer 2 frames by using the IS-IS protocol. TRILL is able to perform optimal forwarding for unicast frames and multipathing for both unicast and multicast traffic. TRILL interoperates with other TRILL-enabled devices as well as existing devices where the STP is running.

The purpose of utilizing TRILL and having Layer 2 multipathing is to eliminate spanning tree from the backbone. This function is important when there is effectively no differentiation between the speed of access and the backbone links. In addition to providing Layer 2 multipathing, TRILL also reduces the latency of traffic through the network.

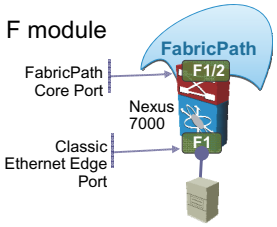
Cisco FabricPath and standards-based TRILL are very similar. The TRILL standard is driven by multiple vendors, including Cisco.

Cisco FabricPath-capable hardware is also TRILL-capable. (Software compatibility will be provided with a software upgrade.) In its long-term commitment for standards-based solutions, Cisco will push Cisco FabricPath-specific enhancements to the TRILL implementation.

https://t.me/learningnets

FabricPath Configuration Procedure

1. Install the Cisco FabricPath feature set in the default VDC
   - Make sure you have enhanced Layer 2 license
   - VDC-specific steps only on Nexus 7000, not Nexus 5500
2. Enable the feature set in any VDC (including the default VDC)
3. Set FabricPath switch ID
4. Configure STP priorities on FabricPath edge devices
   - Set to the lowest value in the Layer 2 network (should become root)
   - The priorities should match
   - Recommended value on all FabricPath edge devices is 8192
5. Configure FabricPath interfaces
   - On Nexus 7000 FP and CE edge ports must be on an F module
6. Define FabricPath VLANs
7. Configure virtual switch ID for vPC+ (optional)
8. Tune load balancing hash functions (optional)
   - Unicast/multicast



© 2012 Cisco and/or its affiliates. All rights reserved.

Follow these steps to implement Cisco FabricPath on Cisco Nexus 7000 Series switches and Cisco Nexus 5500 Platform switches:

Step 1 On Cisco Nexus 7000 Series switches, install the Cisco FabricPath feature that is set in the default VDC. Make sure that you have an enhanced Layer 2 license installed.

Step 2 On Cisco Nexus 7000 Series switches, enable the feature that is set in any VDC (including the default VDC). On Cisco Nexus 5500 Platform switches, enable the feature globally on the switch.

Step 3 Set the Cisco FabricPath switch ID.

Step 4 Configure STP priorities on Cisco FabricPath edge devices. Set them to the lowest value in the Layer 2 network so that the Cisco FabricPath edge switches become STP roots in their respective domains. (The priorities should match.) The recommended value on all Cisco FabricPath edge devices is 8192.

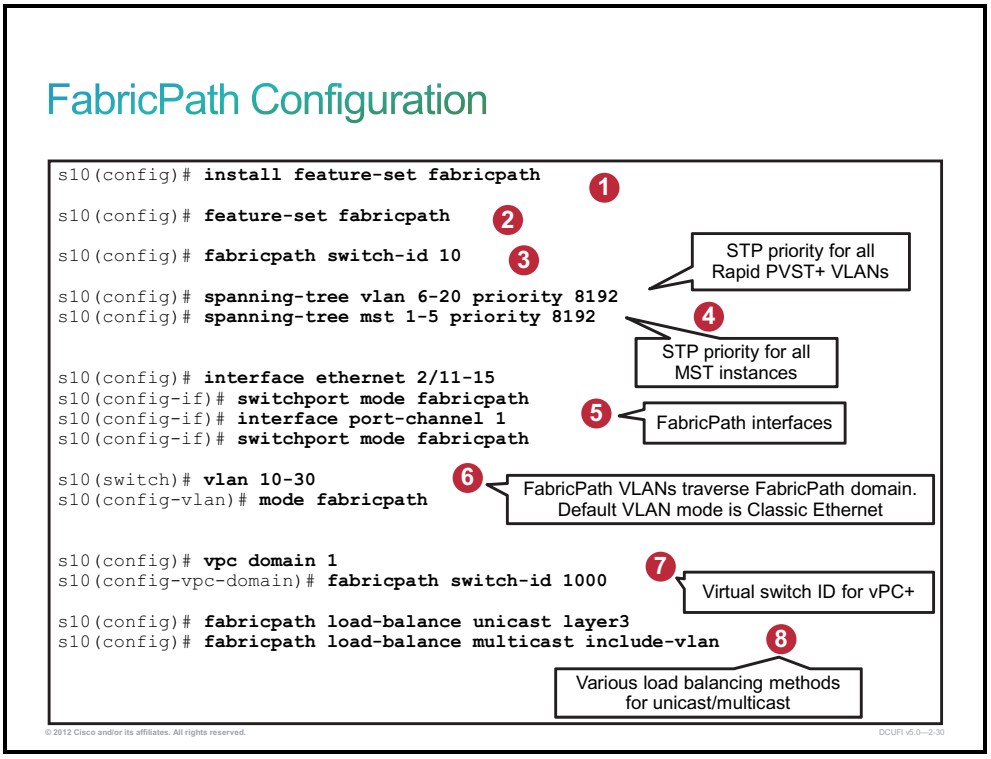
Step 5 Configure Cisco FabricPath interfaces. On Cisco Nexus 7000 Series switches, both the Cisco FabricPath and Classic Ethernet edge ports must be on an F module.

Step 6 Define FabricPath VLANs that will cross the FabricPath cloud.

Step 7 Configure virtual switch ID for vPC+, if you deploy this solution in your environment.

Step 8 Optionally, tune load-balancing hash functions that are separately configurable for unicast and multicast traffic.

https://t.me/learningnets



The configuration example illustrates the commands that are required in order to implement Cisco FabricPath. Setting the switch ID is optional but recommended. All VLAN priorities are set to the recommended value of 8192.

VLANs in the range 10–20 are configured as Cisco FabricPath VLANs. Other VLANs exist on this switch and use the default Classic Ethernet mode, but they are not shown in this example.

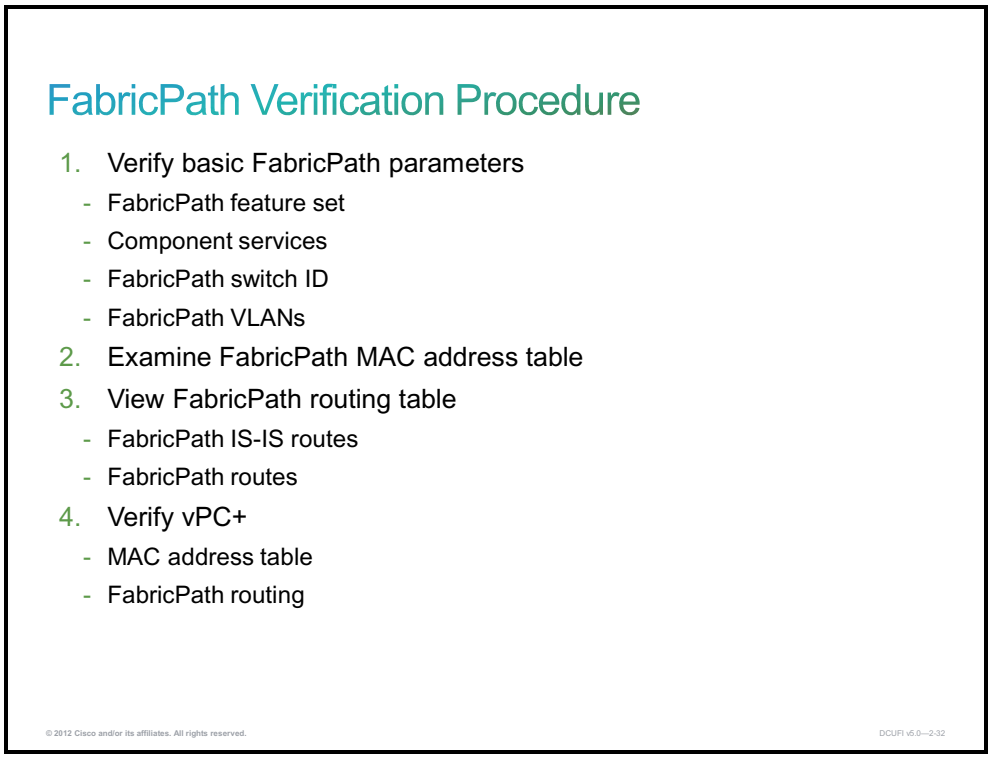
Note that the virtual switch ID is set to 1000 using the **fabricpath switch-id** command in the vpc domain configuration mode. You will verify this configuration in the next section.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Verify Cisco FabricPath

This topic explains how to verify Cisco FabricPath on Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches.



Perform these steps in order to verify Cisco FabricPath operation:

Step 1 Verify basic Cisco FabricPath parameters, such as the enabled Cisco FabricPath feature, its component services, Cisco FabricPath switch ID, and Cisco FabricPath VLANs.

Step 2 Examine the Cisco FabricPath MAC address table.

Step 3 View the Cisco FabricPath routing table. You can use various command options to view different information about the switch ID table.

Step 4 Verify vPC+. You can examine the MAC address table for entries that are related to vPC+ and search the switch ID table for entries with the virtual switch ID.

https://t.me/learningnets

Verify Basic FabricPath Settings

```
switch# show feature-set
Feature Set Name      ID        State
--------------------  --------  --------
fabricpath            2         enabled
fex                   3         disabled

switch# show feature-set services fabricpath
u2rib
drap
isis_fabricpath
3 services in feature set fabricpath

switch# show fabricpath switch-id

                        FABRICPATH SWITCH-ID TABLE
Legend: '*' - this system
================================================================================
SWITCH-ID      SYSTEM-ID      FLAGS        STATE     STATIC EMULATED
----------+----------------+------------+-----------+-------------
10          0018.bad8.12fd    Primary     Confirmed   Yes      No
*25         0018.bad8.12fe    Primary     Confirmed   Yes      No
30          0018.bad8.12ff    Primary     Confirmed   Yes      No

switch# show fabricpath topology vlan active
TPG-name TPG-  ID              Active VLAN List
-------- ---------  ---------------------------------------------
0            0                  10-30
```

© 2012 Cisco and/or its affiliates. All rights reserved.

Make sure that the Cisco FabricPath feature set has been enabled. The feature set comprises three services: Unicast Layer 2 Routing Information Base (U2RIB), Dynamic Resource Allocation Protocol (DRAP), and IS-IS_FabricPath. Examine the switch IDs by using the **show fabricpath switch-id** command. The switch IDs can be set manually or automatically provisioned by the system. View the active Cisco FabricPath VLANs by using the **show fabricpath topology vlan active** command.

https://t.me/learningnets

Verify FabricPath MAC Address Table

- Local MAC addresses denoted by attachment port
- Remote MAC addresses denoted by switch ID (SWID), sub-switch ID (SSID), and local ID (LID)
- Local ID
  - Identifies the exact source/destination port on the switch
  - No need for address lookup on egress switch

```
switch# show mac address-table dynamic vlan 10
Legend:
        * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
        age - seconds since last seen,+ - primary entry using vPC Peer-Link
   VLAN     MAC Address      Type      age     Secure NTFY Ports/SWID.SSID.LID
---------+-----------------+--------+---------+------+----+------------------
* 10       0000.0000.0001    dynamic   0          F    F  Eth1/15
* 10       0000.0000.0002    dynamic   0          F    F  Eth1/15
* 10       0000.0000.0008    dynamic   0          F    F  Eth1/15
* 10       0000.0000.0009    dynamic   0          F    F  Eth1/15
* 10       0000.0000.000a    dynamic   0          F    F  Eth1/15
  10       0000.0000.000b    dynamic   0          F    F  200.0.30
  10       0000.0000.000c    dynamic   0          F    F  200.0.30
  10       0000.0000.000d    dynamic   0          F    F  200.0.30
  10       0000.0000.000e    dynamic   0          F    F  200.0.30
```

MAC address table in FabricPath VLAN

Local address

Remote address

© 2012 Cisco and/or its affiliates. All rights reserved.

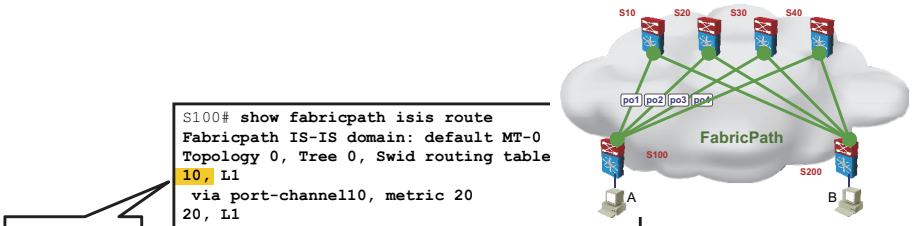
You can display the MAC address table by using the **show mac address-table** command and view the MAC addresses that are learned in a VLAN. VLAN 10 has been configured as a Cisco FabricPath VLAN, and therefore contains two types of MAC address entries: local and remote.

Local addresses are identified by the Classic Ethernet interface to which they are attached.

Remote addresses are denoted by these parameters: switch ID (SID), (sSID), and local ID (LID). The remote MAC addresses that are shown in the figure are attached behind the remote switch with ID 200. The sSID is set to 0 because there is no vPC+ bundle at the remote end. The remote interface has the local interface ID of 30. LID identifies the exact source/destination port on the switch, then simplifies forwarding by making an address lookup on the egress switch redundant.

https://t.me/learningnets

Verify FabricPath Routing Table (IS-IS)




```
S100# show fabricpath isis route
Fabricpath IS-IS domain: default MT-0
Topology 0, Tree 0, Swid routing table
10, L1
 via port-channel10, metric 20
20, L1
 via port-channel20, metric 20
30, L1
 via port-channel30, metric 20
40, L1
 via port-channel40, metric 20
200, L1
 via port-channel30, metric 40
 via port-channel40, metric 40
 via port-channel20, metric 40
 via port-channel10, metric 40
300, L1
 via port-channel30, metric 40
 via port-channel40, metric 40
 via port-channel20, metric 40
 via port-channel10, metric 40
```

Destination switch ID

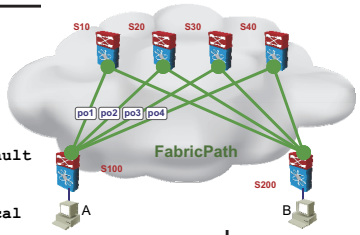
Metric to destination switch

Next hop interfaces

© 2012 Cisco and/or its affiliates. All rights reserved.

You can examine the switch ID table by using multiple command options. The **show fabricpath isis route** command displays the IS-IS topology by showing the available best paths to all switch IDs within the Cisco FabricPath domain.

Verify FabricPath Routing Table




```
S100# show fabricpath route
FabricPath Unicast Route Table
'a/b/c' denotes ftag/switch-id/subswitch-id
'[x/y]' denotes [admin distance/metric]
ftag 0 is local ftag
subswitch-id 0 is default subswitch-id

FabricPath Unicast Route Table for Topology-Default

0/100/0, number of next-hops: 0
        via ---- , [60/0], 0 day/s 04:43:51, local
1/10/0, number of next-hops: 1
        via Po10, [115/20], 0 day/s 02:24:02, isis_fabricpath-default
1/20/0, number of next-hops: 1
        via Po20, [115/20], 0 day/s 04:43:25, isis_fabricpath-default
1/30/0, number of next-hops: 1
        via Po30, [115/20], 0 day/s 04:43:25, isis_fabricpath-default
1/40/0, number of next-hops: 1
        via Po40, [115/20], 0 day/s 04:43:25, isis_fabricpath-default
1/200/0, number of next-hops: 4
        via Po10, [115/40], 0 day/s 02:24:02, isis_fabricpath-default
        via Po20, [115/40], 0 day/s 04:43:06, isis_fabricpath-default
        via Po30, [115/40], 0 day/s 04:43:06, isis_fabricpath-default
        via Po40, [115/40], 0 day/s 04:43:06, isis_fabricpath-default
1/300/0, number of next-hops: 4
        via Po10, [115/40], 0 day/s 02:24:02, isis_fabricpath-default
        via Po20, [115/40], 0 day/s 04:43:25, isis_fabricpath-default
        via Po30, [115/40], 0 day/s 04:43:25, isis_fabricpath-default
        via Po40, [115/40], 0 day/s 04:43:25, isis_fabricpath-default
```

Tree ID, Switch ID, Subswitch ID

Client protocol

Administrative distance, metric

© 2012 Cisco and/or its affiliates. All rights reserved.

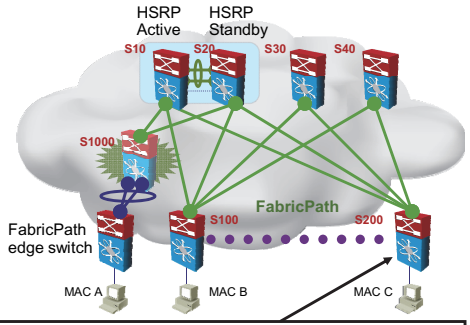
The **show fabricpath route** command offers more comprehensive information by adding details about the administrative distance, age, and client protocol.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Verify VPC+: MAC Address Table

- Notation: SWID/SSID/LID
- In vPC+:
  - SWID: Virtual switch ID (1000)
  - SSID: Sub-switch identifies exact port channel
  - LID: not used




```
S200# show mac address-table dynamic
Legend:
        * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
        age - seconds since last seen,+ - primary entry using vPC Peer-Link
   VLAN     MAC Address      Type      age     Secure NTFY Ports/SWID.SSID.LID
---------+-----------------+--------+---------+------+----+------------------
*10        0000.0000.000c    dynamic   1500       F    F    Eth1/30
*10        0000.0c07.ac0a    dynamic   0          F    F    1000.11.4513
```

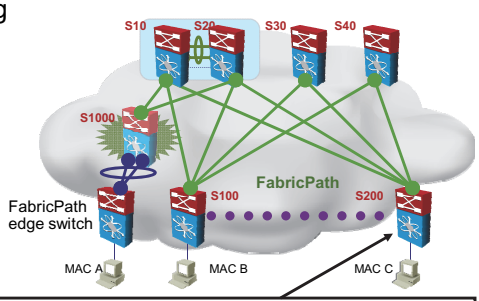
HSRP VMAC

© 2012 Cisco and/or its affiliates. All rights reserved.

You can verify the vPC+ operations by examining the MAC address tables on other switches within the Cisco FabricPath network. You will see one or more MAC addresses that are related to the virtual switch ID. The output in the figure displays the HSRP virtual MAC address. The switch ID is the virtual switch ID (1000), the SSID (11) identifies the vPC bundle, and the local ID is not used.

Verify vPC+: FabricPath Routing

1. Search the FabricPath routing table for the virtual switch ID (1000)
2. In this example, two parallel paths to the virtual switch in default tree




```
S200# show fabricpath route topology 0 switchid 1000
FabricPath Unicast Route Table
'a/b/c' denotes ftag/switch-id/subswitch-id
'[x/y]' denotes [admin distance/metric]
ftag 0 is local ftag
subswitch-id 0 is default subswitch-id


FabricPath Unicast Route Table for Topology-Default

1/1000/0, number of next-hops: 2
          via Po1, [115/10], 0 day/s 01:09:56, isis_l2mp-default
          via Po2, [115/10], 0 day/s 01:09:56, isis_l2mp-default
```

1

2

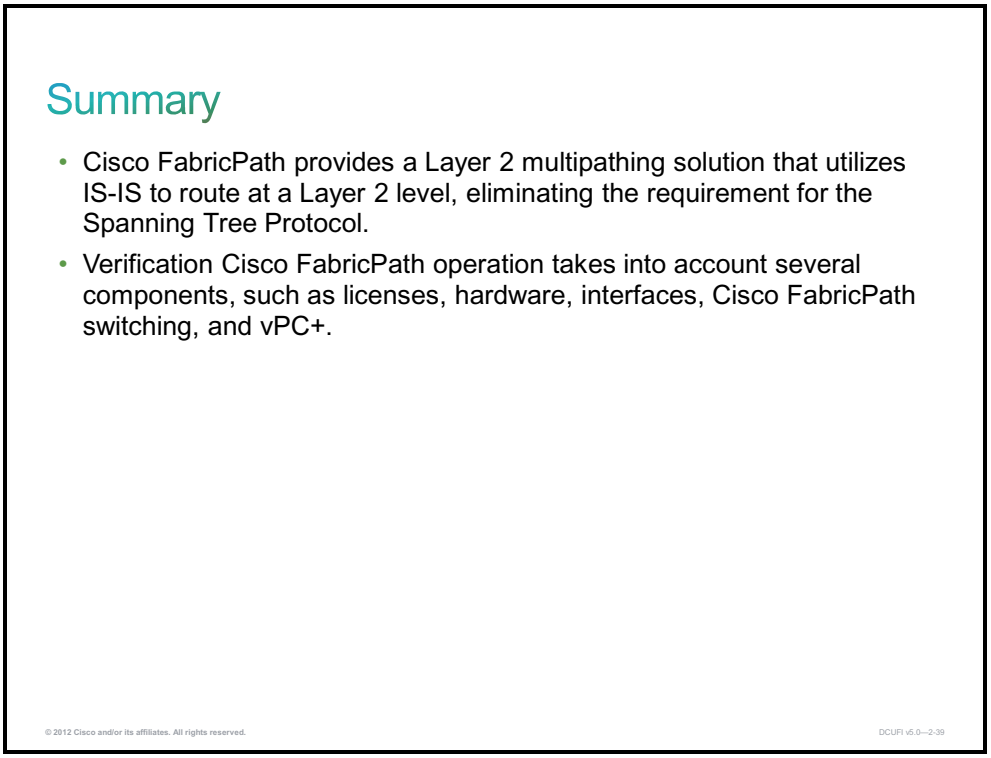
© 2012 Cisco and/or its affiliates. All rights reserved.

You can verify vPC+ operation by examining the switch ID table on the other switches. You can narrow down the contents by selecting a specific topology (FTag) and the desired switch ID, as in the command in the figure. The output displays two paths to the virtual switch ID 1000 in the default topology (ID 1).

https://t.me/learningnets

Summary

This topic summarizes the key points that were discussed in this lesson.



References

For additional information, refer to these resources:

- To learn more about configuring Cisco FabricPath on Cisco Nexus 7000 Series Switches, refer to *Cisco Nexus 7000 Series NX-OS FabricPath Configuration Guide* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/sw/6_x/nx-os/fabricpath/configuration/guide/fp_cli_Book.html
- To learn more about configuring Cisco FabricPath on Cisco Nexus 5500 Series Switches, refer to *Cisco Nexus 5000 Series NX-OS FabricPath Configuration Guide, Release 5.1(3)N1(1)* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/nexus5000/sw/fabricpath/513_n1_1/N5K_FabricPath_Configuration_Guide.html
- To learn more about Cisco FabricPath operation, refer to *Cisco FabricPath for Cisco Nexus 7000 Series Switches* at this URL: http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9402/white_paper_c11-687554.html
- To learn more about designing the Cisco FabricPath solution, refer to *Cisco FabricPath Design Guide: Using FabricPath with an Aggregation and Access Topology* at this URL: http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9670/guide_c07-690079.html

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 6

Configuring Layer 3 Switching Features

Overview

While the access and aggregation layers in the data center network commonly use Layer 2 switching, Layer 3 routing is typically used between the aggregation and core layers in the data centers. The use of Layer 3 routing helps in isolating failure domains and providing equal cost multipathing to optimize the use of bandwidth in the core of the data center network. Therefore, it is important that you understand all aspects of Layer 3 switching, including routing protocol implementation, Layer 3 virtualization, route filtering, policy-based routing (PBR), and IPv6.

Objectives

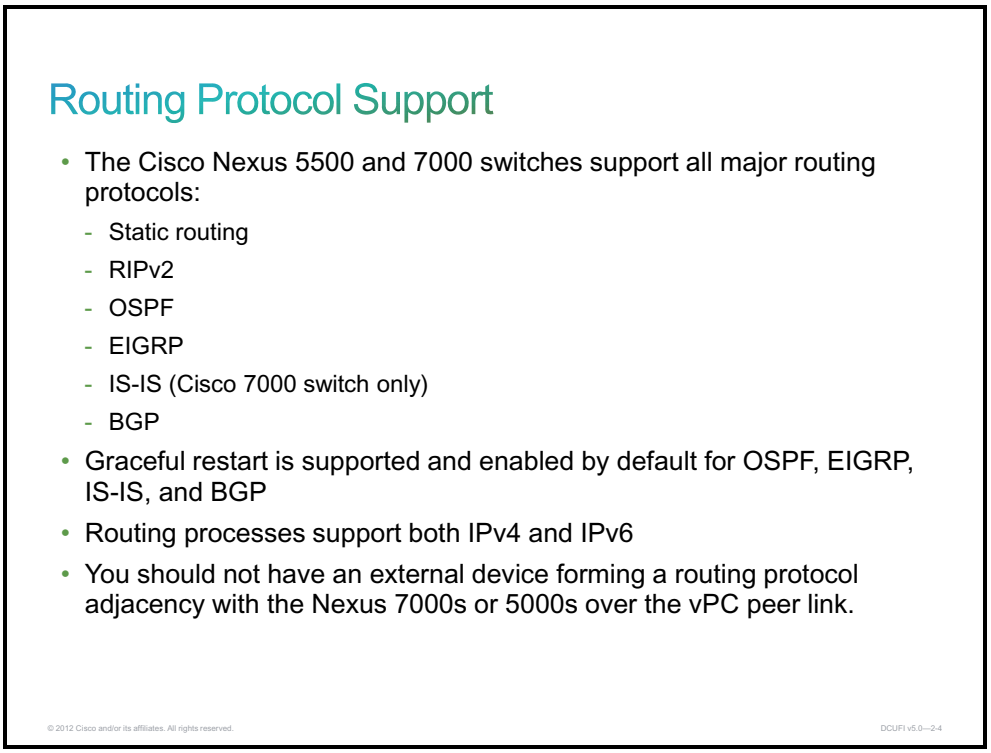
Upon completing this lesson, you will be able to implement and verify Layer 3 switching features on the Cisco Nexus switch. You will be able to meet these objectives:

- Identify how to configure the different routing protocols supported by the Cisco Nexus 7000 Series switch
- Identify how to configure FHRP on Cisco Nexus switches
- Identify how to configure bidirectional forwarding detection on the Cisco Nexus switches
- Identify the use and configuration of Layer 3 virtualization on the Cisco Nexus 7000 Series switch
- Identify how to manage the unicast RIB and FIB on the Cisco Nexus 7000 Series switch
- Identify the use and configuration of the Route Policy Manager on the Cisco Nexus switch
- Identify the use and configuration of policy-based routing on a Cisco Nexus switch
- Identify the implications of using IPv6 in the data center

https://t.me/learningnets

Routing Protocols

This topic identifies how to configure the different routing protocols supported by the Cisco Nexus 7000 Series switch.



The Cisco Nexus Operating System (NX-OS) Software for the Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches supports all of the major routing protocols: static routing, Routing Information Protocol version 2 (RIPv2), Open Shortest Path First (OSPF), Enhanced Interior Gateway Routing Protocol (EIGRP), Intermediate System-to-Intermediate System (IS-IS) (Cisco Nexus 7000 Series switch only), and Border Gateway Protocol (BGP).

OSPF, EIGRP, IS-IS, and BGP all support a graceful restart feature, which is enabled by default. Graceful restart routing protocol enhancements allow a graceful restart-capable device to notify neighboring graceful restart-aware devices that a restart is taking place during a supervisor switchover or stateless process restart.

Following a switchover, the graceful restart-capable device requests that the graceful restart-aware neighbor devices send state information so as to help rebuild the routing tables during a graceful restart. The graceful restart message requests that the neighbor relationship is not reset. As the graceful restart-capable device communicates with other devices on the network, it can then begin to rebuild its neighbor list. After neighbor relationships are reestablished, the graceful restart-capable device begins to resynchronize its database with all of its graceful restart-aware neighbors. Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches are capable of maintaining data plane forwarding functions while the control plane data structures are being synchronized.

Bidirectional Forwarding Detection (BFD) can be used to achieve fast convergence for OSPF, EIGRP, IS-IS, and BGP. The routing protocols that are listed in the figure support both IPv4 and IPv6.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Enabling Routing Protocols

- Routing protocols must be enabled using the **feature** command before they can be configured.
- Cisco Nexus 7000 Switch licensing:

Feature set	Description	License
Restricted Layer 3	Only RIPv2	No license
Unlimited Layer 3	All protocols	Enterprise Services License

- Cisco Nexus 5500 Switch licensing:

Feature set	Description	License
Base Layer 3	Connected, static, RIPv2, OSPF (restricted), EIGRP stub, HSRP, VRRP, IGMPv2/3, PIMv2, RACLs, uRPF	Layer 3 Base License (included)
Unlimited Layer 3	All base + full EIGRP, unrestricted OSPF, BGP, VRF-Lite	Layer 3 LAN-Enterprise License

© 2012 Cisco and/or its affiliates. All rights reserved.

On the Cisco Nexus 7000 Series switches, all routing protocols except RIPv2 require the LAN Enterprise package. Static routing and RIPv2 are included in the base Cisco NX-OS features on the Cisco Nexus 7000 Series switch and require no additional license.

On the Cisco Nexus 5500 Platform switches, the Layer 3 Base License enables you to use a basic set of Layer 3 processes, such as connected, static, RIPv2, OSPF (restricted), EIGRP stub, Hot Standby Router Protocol (HSRP), Virtual Router Redundancy Protocol (VRRP), Internet Group Management Protocol versions 2 and 3 (IGMPv2 and IGMPv3), Protocol Independent Multicast version 2 (PIMv2), router ACLs, and Unicast Reverse Path Forwarding (uRPF). To deploy advanced features that include full EIGRP functions, unrestricted OSPF, BGP, and Multi-VRF Customer Edge (VRF-Lite), you must first install the Layer 3 LAN-Enterprise license.

Routing protocols must be enabled by using the **feature** command for the respective routing protocol. When a feature is disabled, the corresponding routing protocol configuration commands are removed from the switch configuration. When a feature is disabled with the **no feature** command, the Cisco NX-OS automatically creates a checkpoint that can be used to roll the configuration back to the state before the feature was disabled.

To enable an interior gateway protocol (IGP) on a specific interface, the Cisco NX-OS software uses the **ip router** *igp* command in interface configuration mode. It is not possible to enable multiple interfaces at once in routing protocol configuration mode by using a **network** command. Similarly, passive interfaces are configured in interface configuration mode instead of routing protocol configuration mode. Multiple interfaces can be enabled at once for an IGP by using the Cisco NX-OS interface range feature.

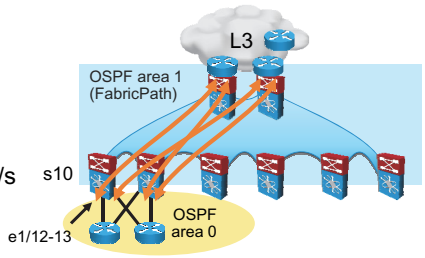
All IGPs support routing virtualization through the use of virtual routing and forwarding (VRF) instances.

Note	Configuration of network virtualization through the use of VRFs is covered in more detail later in this lesson.

https://t.me/learningnets

OSPF Configuration

1. Enable OSPF feature
2. Start OSPF process
3. Configure global parameters
   - Default auto-cost reference bw: 40 Gb/s
4. Enable process on interfaces
5. Configure per-interface settings



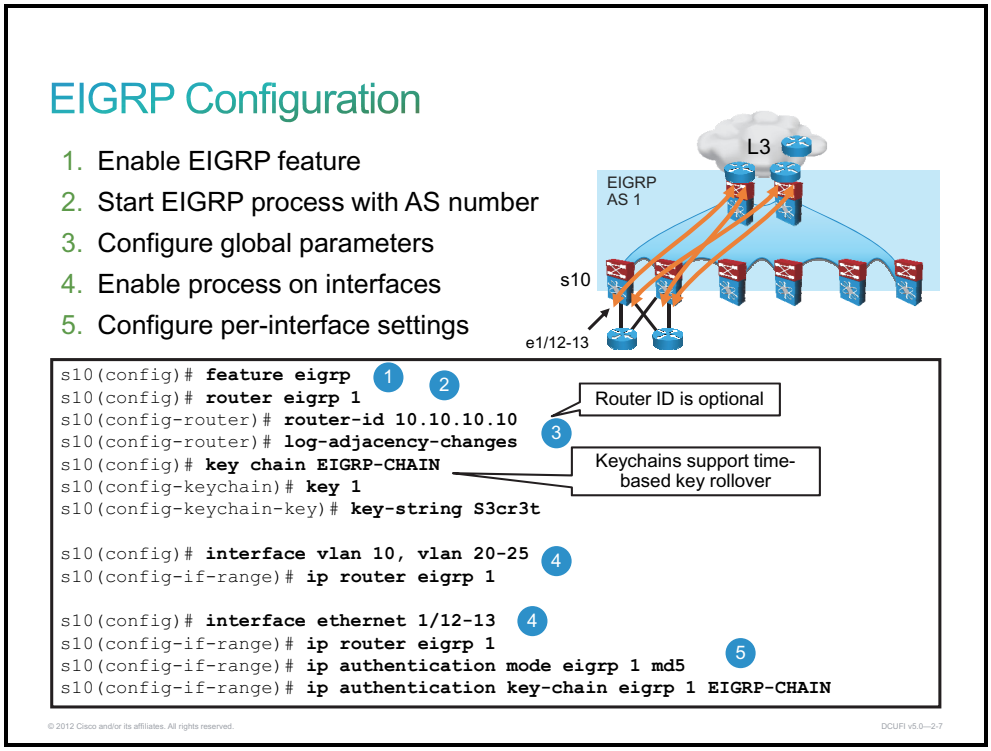

```
s10(config)# feature ospf                                   1
s10(config)# router ospf 1                                  2
s10(config-router)# router-id 10.10.10.10   Router ID is optional
s10(config-router)# log-adjacency-changes
s10(config-router)# auto-cost reference-bandwidth 100 Gbps  3

s10(config)# interface vlan 10, vlan 20-25
s10(config-if-range)# ip router ospf 1 area 1               4   Enable OSPF on SVIs

s10(config)# interface ethernet 1/12-13
s10(config-if-range)# ip router ospf 1 area 0               4   Enable OSPF on physical interfaces
s10(config-if-range)# ip ospf authentication message-digest
s10(config-if-range)# ip ospf message-digest-key 1 md5 S3cr3t   5
```

© 2012 Cisco and/or its affiliates. All rights reserved.

The figure shows an example of a basic OSPF configuration on a Cisco Nexus 5500 Platform switch and a Cisco Nexus 7000 Series switch. The following commands are used:

- **feature ospf**: This command enables the OSPF version 2 (OSPFv2) feature. This feature requires the Enterprise Services License.
- **router ospf** *instance-tag*: This command creates the OSPFv2 instance. The instance tag can be any alphanumeric string.
- **router-id** *ip-address*: This command configures the OSPFv2 router ID. This command is optional.
- **log-adjacency-changes**: This command generates a system message whenever a neighbor state changes. The command is optional and is not enabled by default.
- **auto-cost reference-bandwidth** *bandwidth* **[Gbps | Mbps]**: This command sets the reference bandwidth that is used to calculate the default metrics for an interface. This command is optional. The default value is 40 Gb/s.
- ip router ospf *instance-tag* area *area-id*: This command enables an OSPF instance on an interface for a specific area.
- **ip ospf passive-interface**: This command suppresses the sending of OSPF packets on the interface. This command is optional and is not shown here.
- **ip ospf authentication** [**key-chain** *key-name* | **message-digest** | **null**]: This command sets the OSPF authentication type for the interface. In addition, it can be used to specify a keychain, which contains the authentication keys. This command is optional.
- **ip ospf message-digest-key** *key-id* **md5** [**0** | **3**] *key*: This command defines the key to be used for OSPF Message Digest 5 (MD5) authentication. This command is optional.

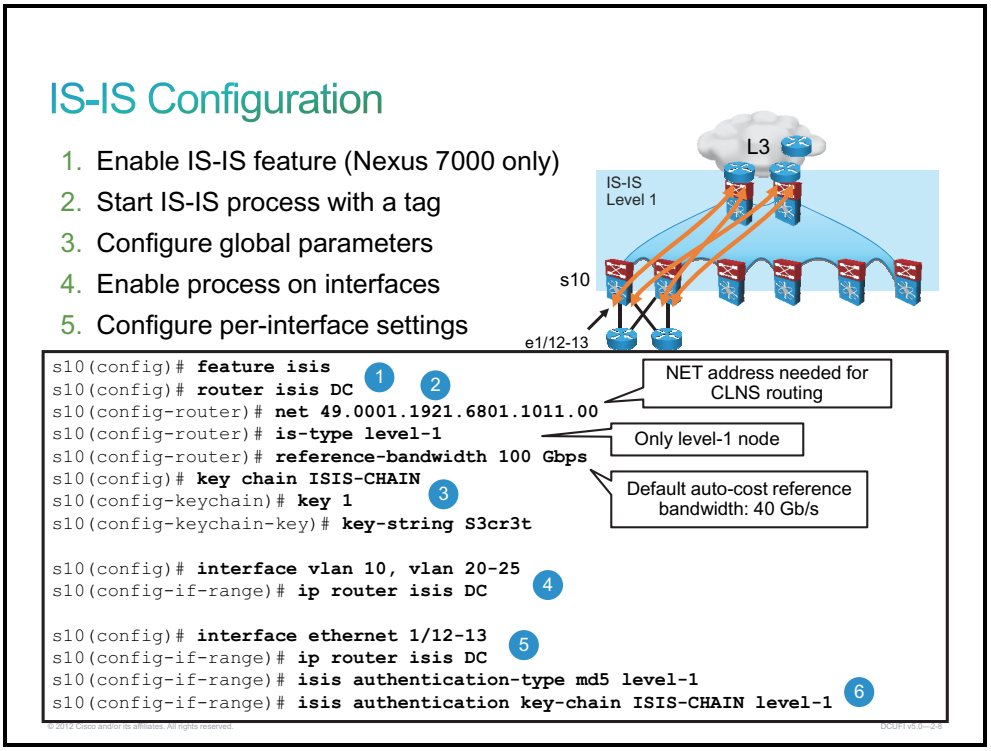
https://t.me/learningnets



The figure shows an example of a basic EIGRP configuration on a Cisco Nexus 5500 Platform switch and a Cisco Nexus 7000 Series switch. The following commands are used:

- **feature eigrp**: This command enables the EIGRP feature. This feature requires the Enterprise Services License.
- **router eigrp** *as-number*: This command begins the EIGRP process. The instance tag can be any case-sensitive alphanumeric string of up to 20 characters. However, if a nonnumeric instance tag is used, then the autonomous system number for EIGRP must be separately specified by using the **autonomous-system** command. If a numeric instance ID is used, the autonomous system number is equal to the instance tag.
- **router-id** *ip-address*: This command configures the EIGRP router ID. This command is optional.
- **eigrp log-neighbor-changes**: This command generates a system message when a neighbor state changes. This command is optional and not enabled by default.
- **ip router eigrp** *as-number*: This command configures the associated EIGRP process on an interface.
- **ip passive-interface eigrp** *instance-tag*: This command suppresses the sending of EIGRP packets on the interface. This command is optional and is not shown here.
- **ip authentication mode eigrp** *instance-tag* **md5**: This command enables MD5 authentication for EIGRP on the interface. This command is optional.
- **ip authentication key-chain eigrp** *instance-tag name-of-chain*: This command specifies the keychain to be used for EIGRP authentication on this interface. This command is optional.
- key chain *keychain-name*: This command creates a keychain, which can contain multiple authentication keys.
- key *key-ID*: This command creates a key with a specific key ID. This ID must be a whole number between 0 and 65535.
- key-string [*encryption-type*] *text-string*: This command defines the keystring for a specific key. The keystring can contain up to 63 case-sensitive, alphanumeric characters.

https://t.me/learningnets

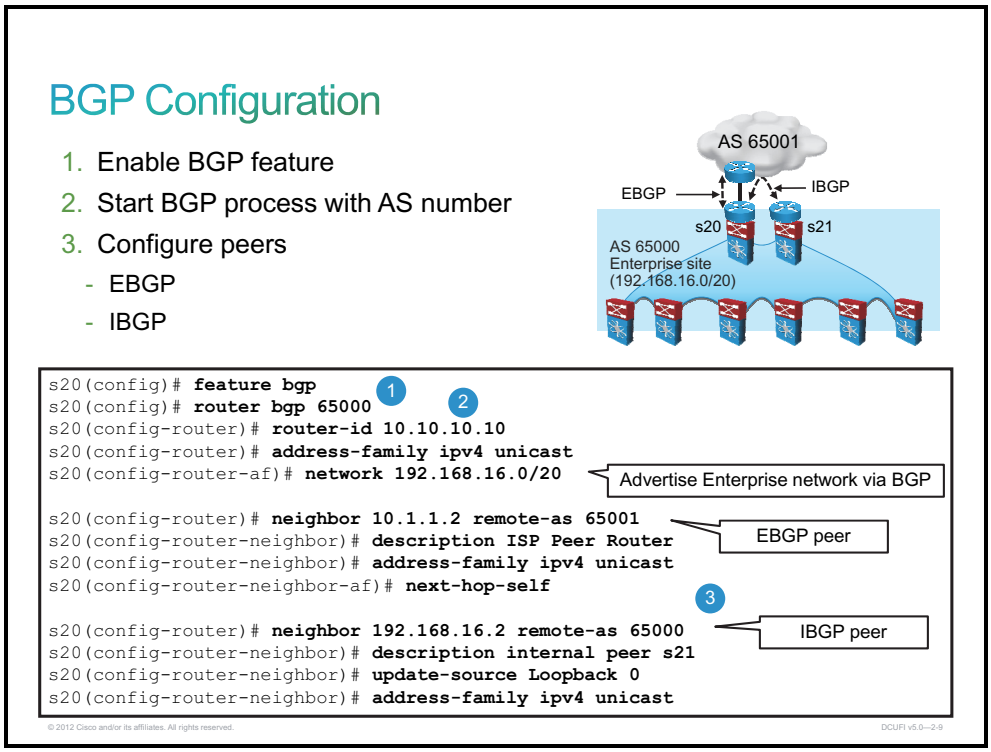


The figure shows an example of a basic IS-IS configuration on a Cisco Nexus 7000 Series switch. The following commands are used:

- **feature isis**: This command enables the IS-IS feature. This feature requires the Enterprise Services License.
- **router isis** *instance-tag*: This command creates a new IS-IS instance.
- **net** *network-entity-title*: This command configures the network entity title (NET) for this IS-IS instance.
- **is-type (level-1 | level-2 | level-1-2)**: This command can be used to configure the routing level for this IS-IS instance. The default level is level 1–2.
- **log-adjacency changes**: This command sends a system message when an IS-IS neighbor changes state. This command is optional and not enabled by default.
- **reference-bandwidth** *bandwidth-value* **(Mbps | Gbps)**: This command sets the default reference bandwidth that is used for calculating the IS-IS metric. The default value is 40 Gb/s.
- **ip router isis** *instance-tag*: This command associates the interface with an IS-IS instance for IP version 4 (IPv4) routing.
- **isis passive {level-1 | level-2 | level-1-2}**: This command prevents the interface from forming adjacencies but still advertises the prefix. This command is optional and is not shown here.
- **isis authentication type md5 {level-1 | level-2}**: This command enables MD5 authentication for IS-IS for the specified routing level on the interface. This command is optional.
- **isis authentication key-chain** *instance-tag name-of-chain* {level-1 | level-2}: This command specifies the keychain to be used for IS-IS authentication for the specified routing level on this interface. This command is optional.
- **key chain** ***keychain-name***: This command creates a keychain, which can contain multiple authentication keys.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

- **key *key-ID*:** This command creates a key with a specific key ID. This ID must be a whole number between 0 and 65535.
- **key-string [*encryption-type*] *text-string*:** This command defines the keystring for a specific key. The key-string can contain up to 63 case-sensitive, alphanumeric characters.



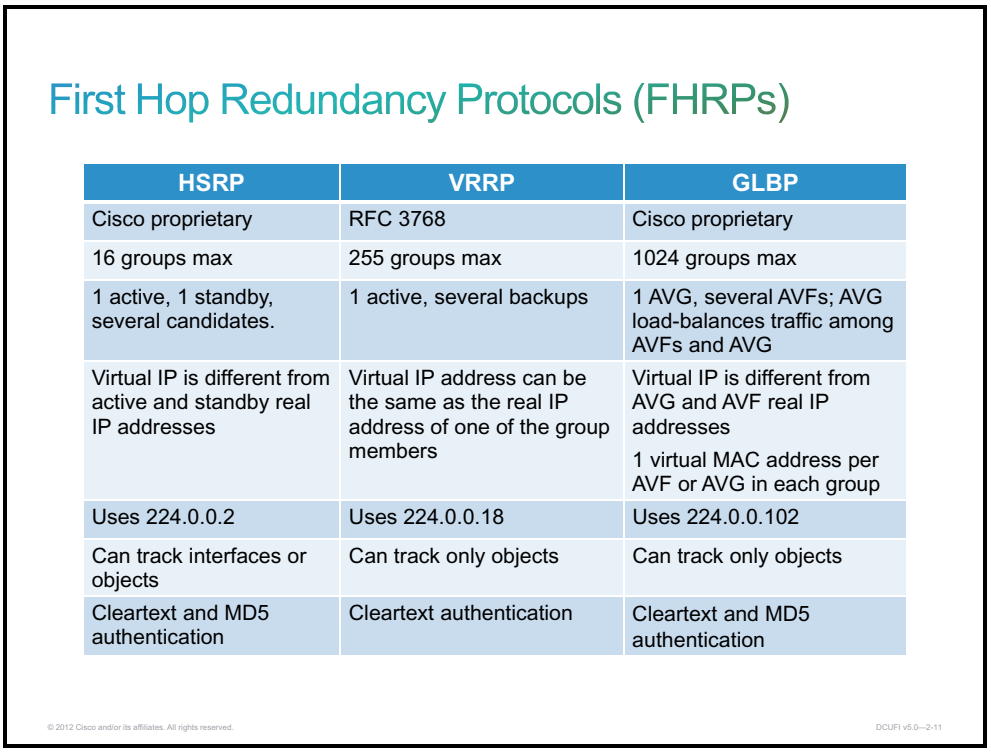
The figure shows an example of a basic BGP configuration on a Cisco Nexus 5500 Platform switch and a Cisco Nexus 7000 Series switch. The following commands are used:

- **feature bgp**: This command enables the BGP feature. This feature requires the Enterprise Services License.
- **router bgp** *autonomous-system-number*: This command enables BGP and assigns an autonomous system number (AS number) to the local BGP process.
- **router-id** *ip-address*: This command defines the BGP router ID.
- **address-family {ipv4 | ipv6} {unicast | multicast}**: This command enters the global address family configuration mode for an address family. In this mode, the specific options for the selected address family can be configured.
- **network** *ip-addr* | *ip-prefix*/*length* mask *mask-num* [route-map *name*]: This command specifies a network prefix as local to this AS number and then adds it to the BGP table.
- **neighbor** *ip-address* **remote-as** *as-number*: This command configures the IP address and AS number for the remote BGP peer.
- **description** *text*: This command adds a description for the neighbor.
- **address-family {ipv4 | ipv6} {unicast | multicast}**: This command enables the address family for a neighbor and enters the address family configuration mode for the specified neighbor. At least one address family must be enabled for a neighbor to enable the peering.
- **next-hop-self**: This command makes the router use the local BGP speaker address as the next-hop address in route updates. This command triggers an automatic soft clear or refresh of BGP neighbor sessions. This command is optional.
- **update-source** *interface*: This command specifies the source of the BGP session and updates.

https://t.me/learningnets

First Hop Redundancy Protocols (FHRPs)

This topic identifies how to configure First Hop Redundancy Protocols (FHRP) on Cisco Nexus switches.



FHRPs, such as Gateway Load Balancing Protocol (GLBP), HSRP, and VRRP, allow you to provide redundant connections to your hosts. If an active first-hop router fails, the FHRP automatically selects a standby router to take over. You do not need to update the hosts with new IP addresses since the address is virtual and shared between each router within the FHRP group.

Object tracking allows you to track specific objects on the network, such as the interface line protocol state, IP routing and route reachability, and take action when the state of the tracked object changes. This feature allows you to increase the availability of the network and shorten the recovery time if an object state goes down.

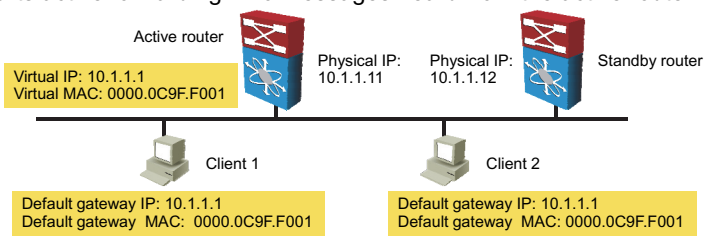
The table compares the features of the three common FHRPs.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Hot Standby Router Protocol (HSRP)

- HSRP group (standby group)
  - Set of HSRP devices emulating a virtual router
- Active router
  - Responds to ARP requests of the default gateway with MAC of virtual router
  - Assumes the active forwarding of packets for the virtual router
  - Sends hello messages
- Standby router
  - Listens for periodic hello messages
  - Starts active forwarding if no messages heard from the active router



© 2012 Cisco and/or its affiliates. All rights reserved.

All routers in the HSRP group are configured with a shared-group IP address known as the virtual IP (VIP) address. A shared, virtual group MAC address is generated from the HSRP group number. This group IP address presents the image of a single, fault-tolerant router.

The group IP address is used by clients that must route out through the HSRP group. When clients use the Address Resolution Protocol (ARP) for the MAC address of this default gateway IP, the active HSRP router responds with the shared VMAC address. During failover, the standby HSRP router assumes this MAC address, thereby avoiding the need to refresh the ARP cache in client devices.

Multiple HSRP groups can be configured on one LAN segment, and different routers can be configured as the default active router for each group. This configuration can be used to provide some traffic load balancing through the routers.

HSRP supports VRF, which exists within virtual device contexts (VDCs). If you change the VRF membership of an interface, the Cisco NX-OS Software removes all Layer 3 configurations, including HSRP.

https://t.me/learningnets

HSRP Versions

Property	Version 1	Version 2
Status	Active by default	Must be enabled
Supported groups	Group number from 0 to 255	Group number from 0 to 4095
Virtual MAC address	Virtual MAC address 0000.0C07.ACXX (XX = HSRP group)	Virtual MAC address 0000.0C9F.FXXX (XXX = HSRP group)
Hello multicast group	Hello packets sent to multicast address 224.0.0.2	Hello packets sent to multicast address 224.0.0.102
Authentication	Only cleartext	Cleartext and MD5
Compatibility	Different packet format than HSRPv2	The packet format uses a type, length, value (TLV) format; HSRP version 2 packets received by an HSRP version 1 router are ignored

© 2012 Cisco and/or its affiliates. All rights reserved.

Cisco NX-OS supports HSRP version 1 (HSRPv1) by default. You can configure an interface to use HSRP version 2 (HSRPv2). HSRPv2 features the following enhancements to HSRPv1:

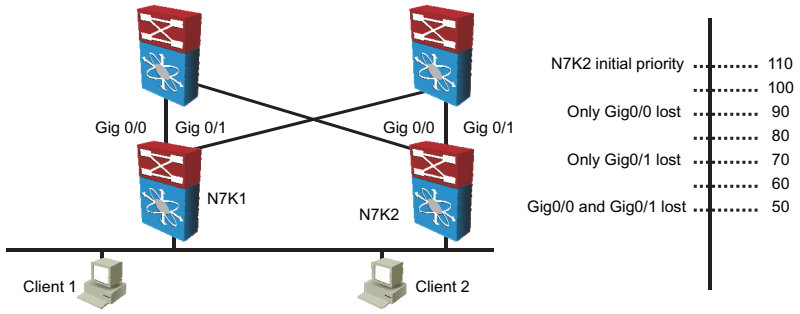
- HSRPv2 expands the group number range. HSRPv1 supports group numbers from 0 to 255; HSRP version 2 supports group numbers from 0 to 4095.
- For IPv4, HSRPv2 uses the IPv4 multicast address 224.0.0.102 or the IPv6 multicast address FF02::66 to send hello packets instead of the multicast address 224.0.0.2, which is used by HSRPv1.
- HSRPv2 uses the MAC address range from 0000.0C9F.F000 to 0000.0C9F.FFFF for IPv4 and 0005.73A0.0000 through 0005.73A0.0FFF for IPv6 addresses. HSRPv1 uses the MAC address range 0000.0C07.AC00 to 0000.0C07.ACFF.
- HSRPv2 adds support for MD5 authentication.
- HSRPv2 adds support for IPv6. HSRP for IPv6 uses these parameters:
  - UDP port 2029
  - Virtual MAC (VMAC) address range from 0005.73A0.0000 through 0005.73A0.0FFF
  - Multicast link-local IP destination address of FF02::66
  - Hop limit set to 255

When you change the HSRP version, Cisco NX-OS reinitializes the group because it now has a new VMAC address. HSRPv2 has a different packet format than HSRPv1. The packet format uses a type, length, value (TLV) format. HSRPv2 packets received by an HSRPv1 router are ignored.

https://t.me/learningnets

HSRP Priorities and Interface Tracking

- Device with the highest priority in HSRP group becomes active router.
- The default priority is 100.
- In case of a tie, the router with highest IP address becomes active.
- Pre-emption enables a higher-priority device to become active.
- Interface tracking modifies the current priority depending on uplink state.



© 2012 Cisco and/or its affiliates. All rights reserved.

When two routers participate in an election process, a priority can be configured to determine which router should be active. Without specific priority configuration, each router has a default priority of 100, and the router with the highest IP address is elected as the active router.

Regardless of other router priorities or IP addresses, an active router will stay active by default. A new election will occur only if the active router is removed. When the standby router is removed, a new election is made to replace the standby. You can change this default behavior by enabling the pre-empt option.

HSRP interface tracking enables the priority of a standby group router to be automatically adjusted based on the availability of the router interfaces. When a tracked interface becomes unavailable, the HSRP priority of the router is decreased. When properly configured, the HSRP tracking feature ensures that a router with an unavailable key interface will relinquish the active router role.

The HSRP group tracks the uplink interfaces. If the uplink to the core on the right switch fails, the router automatically decrements the priority on that interface and sends hello messages with the decremented priority. The switch on the left now has a higher priority and, with pre-emption enabled, becomes the active router.

A router can track several interfaces. In the figure, each of the access switches at the bottom tracks the uplink interfaces Gig0/0 and Gig0/1. The initial priority of the N7K2 switch is set to 110. If the switch loses the uplink Gig0/0, its priority is decreased by 20. If the switch loses the uplink Gig0/1, its priority is decreased by 40.

https://t.me/learningnets

HSRP Configuration

```
switch(config)# key chain hsrp-keys
switch(config-keychain)# key 0
switch(config-keychain-key)# key-string 7 VerySecret0
switch(config-keychain-key)# accept-lifetime 00:00:00 Apr 01 2012 23:59:59
Sep 12 2012
switch(config-keychain-key)# send-lifetime 00:00:00 Apr 01 2012 23:59:59 Aug
12 2012
switch(config-keychain-key)# key 1
switch(config-keychain-key)# key-string 7 VerySecret1
switch(config-keychain-key)# accept-lifetime 00:00:00 Aug 12 2012 23:59:59
Dec 12 2012
switch(config-keychain-key)# send-lifetime 00:00:00 Sep 12 2012 23:59:59 Nov
12 2012

switch(config)# feature hsrp

switch(config)# track 2 interface ethernet 2/2 ip

switch(config)# interface ethernet 1/2
switch(config-if)# ip address 192.0.2.2/8
switch(config-if)# hsrp 1
switch(config-if-hsrp)#  authenticate md5 key-chain hsrp-keys
switch(config-if-hsrp)#  priority 90
switch(config-if-hsrp)#  track 2 decrement 20
switch(config-if-hsrp)#  ip-address 192.0.2.10
switch(config-if-hsrp)# no shutdown
```

Keychain enables key rollover

Enable HSRP feature

Interface tracking

HSRP group with options

© 2012 Cisco and/or its affiliates. All rights reserved.

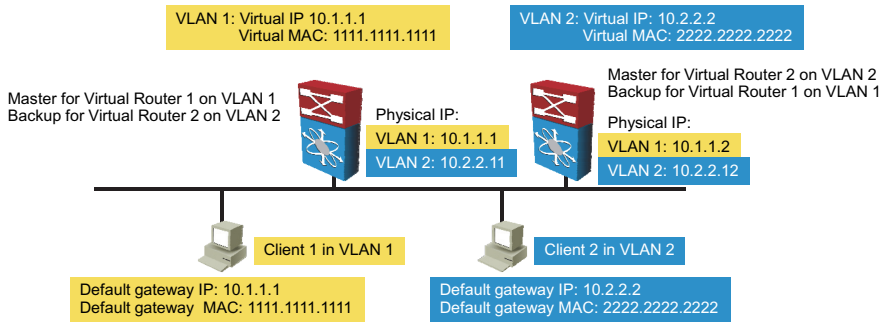
The figure illustrates a sample HSRP configuration that illustrates HSRP MD5 authentication that is based on keychains and interface tracking. The keychain feature enables an automated time-based key rollover. The interface tracking will decrease the router priority if the interface Ethernet 1/2 fails.

 © 2012 Cisco Systems, Inc.

https://t.me/learningnets

Virtual Router Redundancy Protocol (VRRP)

- Standards-based alternative to HSRP
- Minor differences, such as:
  - Virtual router IP can be identical to physical IP address
  - More groups per interface
  - Can track only objects

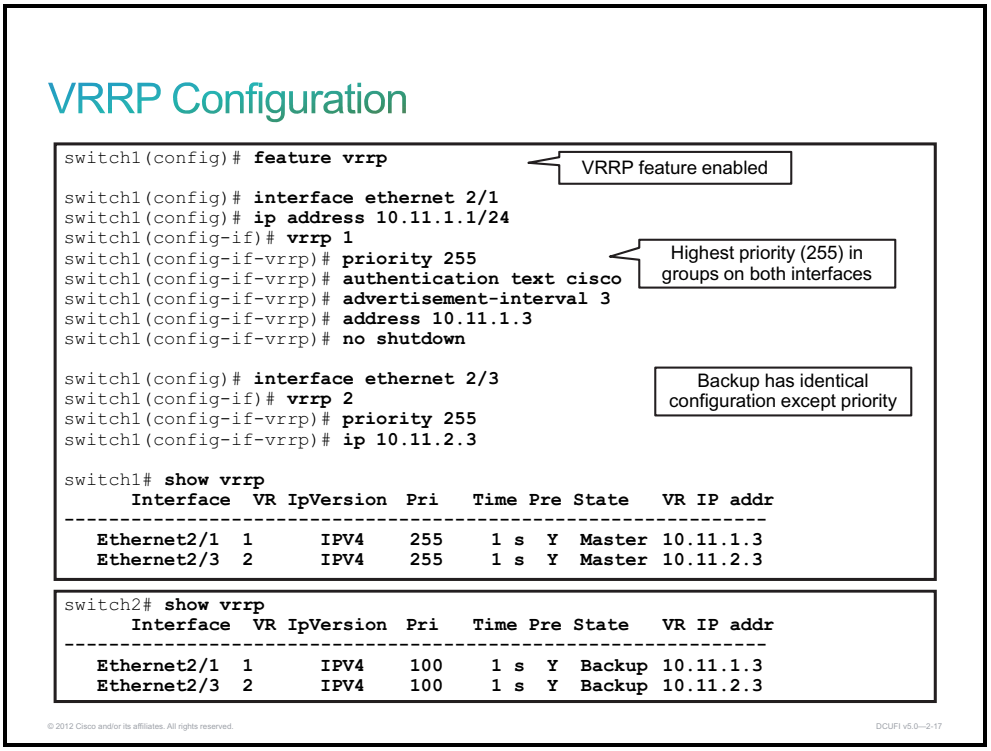


© 2012 Cisco and/or its affiliates. All rights reserved.

VRRP allows for transparent failover at the first-hop IP router by configuring a group of routers to share a VIP address. VRRP selects a master router from within that group to manage all packets for the VIP address. The remaining routers are in standby and take over if the master router fails.

VRRP provides a standards-based alternative to HSRP but supports fewer features than its Cisco-proprietary counterpart.

https://t.me/learningnets



This figure presents a sample of VRRP. You must globally enable the VRRP feature before you can configure and enable any VRRP groups.

You should create a VRRP group, assign the virtual IP address, and enable the group. You can configure one virtual IPv4 address for a VRRP group. By default, the master VRRP router drops the packets that are addressed directly to the virtual IP address because the VRRP master is only intended as a next-hop router to forward packets. Some applications require that Cisco NX-OS accept packets that are addressed to the virtual router IP. Use the secondary option for the virtual IP address in order to accept these packets when the local router is the VRRP master.

Once you have configured the VRRP group, you must explicitly enable the group before it becomes active.

The valid priority range for a virtual router is from 1 to 254 (1 is the lowest priority, and 254 is the highest). The default priority value for backups is 100. For devices whose interface IP address is the same as the primary VIP address (the master), the default value is 255.

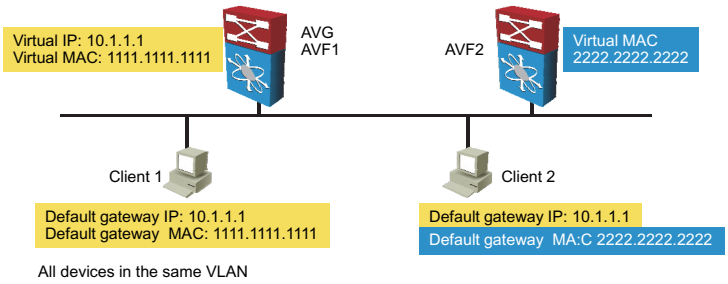
Interface state tracking changes the priority of the virtual router based on the state of another interface in the device. When the tracked interface goes down or the IP address is removed, Cisco NX-OS then assigns the tracking priority value to the virtual router. When the tracked interface comes up and an IP address is configured on this interface, Cisco NX-OS restores the configured priority to the virtual router.

If you configure VRRP on a vPC-enabled interface, you can optionally configure the upper and lower threshold values to control when to fail over to the virtual port channel (vPC) trunk. If the backup router priority falls below the lower threshold, VRRP sends all backup router traffic across the vPC trunk to then forward through the master VRRP router. VRRP maintains this scenario until the backup VRRP router priority increases above the upper threshold.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Gateway Load Balancing Protocol (GLBP)

- Allows use of all devices without creating multiple groups
- Provides a single virtual IP address and multiple virtual MAC addresses
- Routes traffic to single gateway distributed across routers
  - Active Virtual Gateway: responds to ARP requests with AVF MAC addresses
  - Active Virtual Forwarder: actively forwards traffic



© 2012 Cisco and/or its affiliates. All rights reserved.

GLBP provides path redundancy for IP by sharing protocol and MAC addresses between redundant gateways. Additionally, GLBP allows a group of Layer 3 routers to share the load of the default gateway on a LAN. A GLBP router can automatically assume the forwarding function of another router in the group if the other router fails.

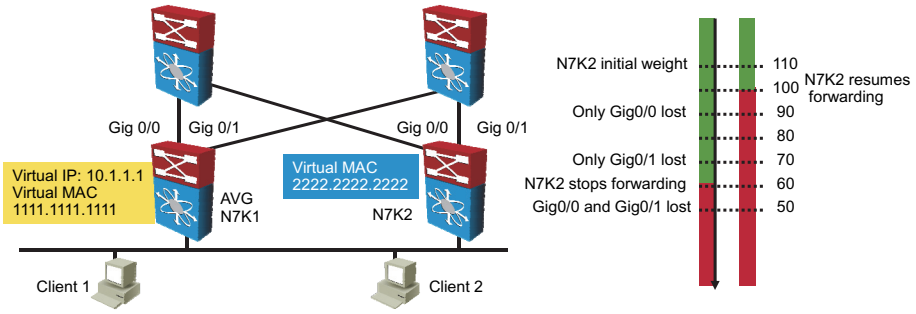
GLBP prioritizes gateways to elect an active virtual gateway (AVG). If multiple gateways have the same priority, the gateway with the highest real IP address becomes the AVG. The AVG then assigns a virtual MAC address to each member of the GLBP group. Each member is the active virtual forwarder (AVF) for its assigned virtual MAC address, forwarding packets sent to its assigned virtual MAC address.

The AVG also answers ARP requests for the virtual IP address. Load sharing is achieved when the AVG replies to the ARP requests with different virtual MAC addresses.

https://t.me/learningnets

GLBP Weights and Interface Tracking

- Weights determine the forwarding capacity of each router
  - The proportion of hosts for which it will forward packets
- Thresholds set to:
  - Disable forwarding when weight falls below a certain value
  - Re-enable forwarding and when weight rises above another threshold
- Weighting can be automatically adjusted by interface tracking



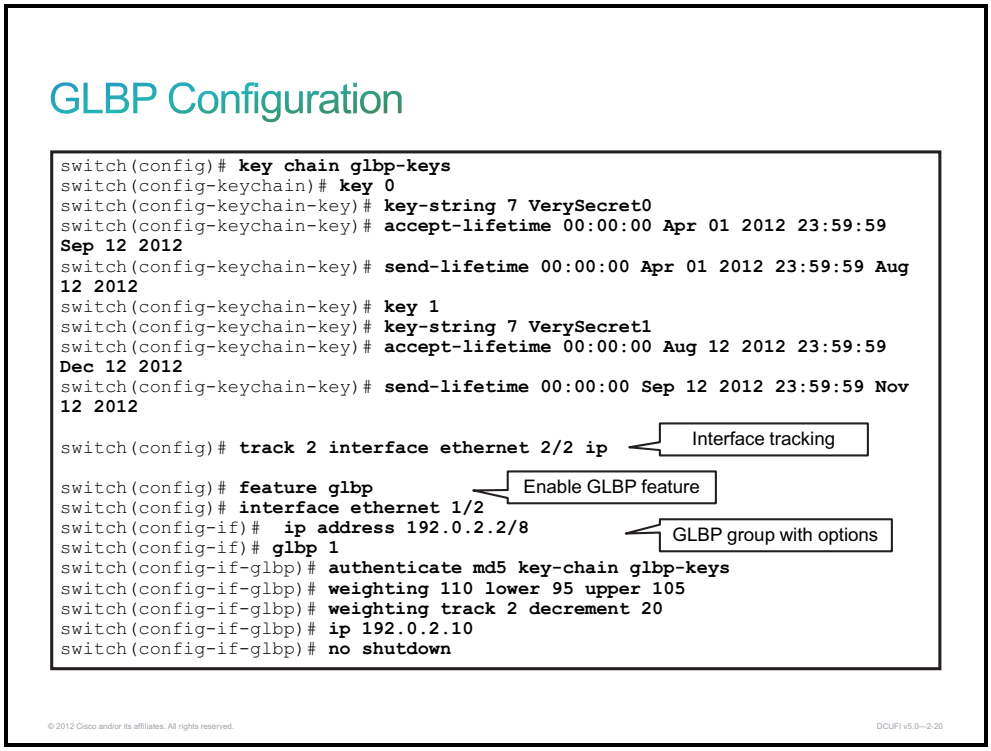
© 2012 Cisco and/or its affiliates. All rights reserved.

You can configure GLBP to track an interface or routes and enable the secondary virtual forwarder to take over if a tracked link goes down. GLBP tracking uses weighted load balancing to determine whether a GLBP group member acts as an AVF. You must configure the initial weighting values and optional thresholds to enable or disable this group member as an AVF. You can also configure the interface to track the value that reduces interface weighting if the interface goes down. When the GLBP group weighting drops below the lower threshold, the member is no longer an AVF, at which point a secondary virtual forwarder takes over. When the weighting rises above the upper threshold, the member can resume its role as an AVF.

The figure illustrates the process of weight modification when selected interface fails and the secondary virtual forwarder becomes active.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



The figure illustrates how to configure GLBP on Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 series Switches. This example uses keychains for automatic key rollover, interface tracking, and weighting thresholds.

https://t.me/learningnets

Bidirectional Forwarding Detection

This topic identifies how to configure bidirectional forwarding detection on the Cisco Nexus switches.

Bidirectional Forwarding Detection (BFD)

- BFD:
  - Uses frequent link hellos
  - Provides fast, reliable detection of a link failure
  - Useful for link failures that are not detectable through Layer 1 mechanisms
- BFD can be tied to Layer 3 control protocols
  - BGP, OSPF, EIGRP, IS-IS, HSRP, and PIM
  - Serves as a fast failure-detection mechanism
  - More efficient than hellos in individual protocols
- BFD on Cisco Nexus 7000 switches:
  - Runs in a distributed manner
  - Offloads the BFD processing to the CPUs on the I/O modules

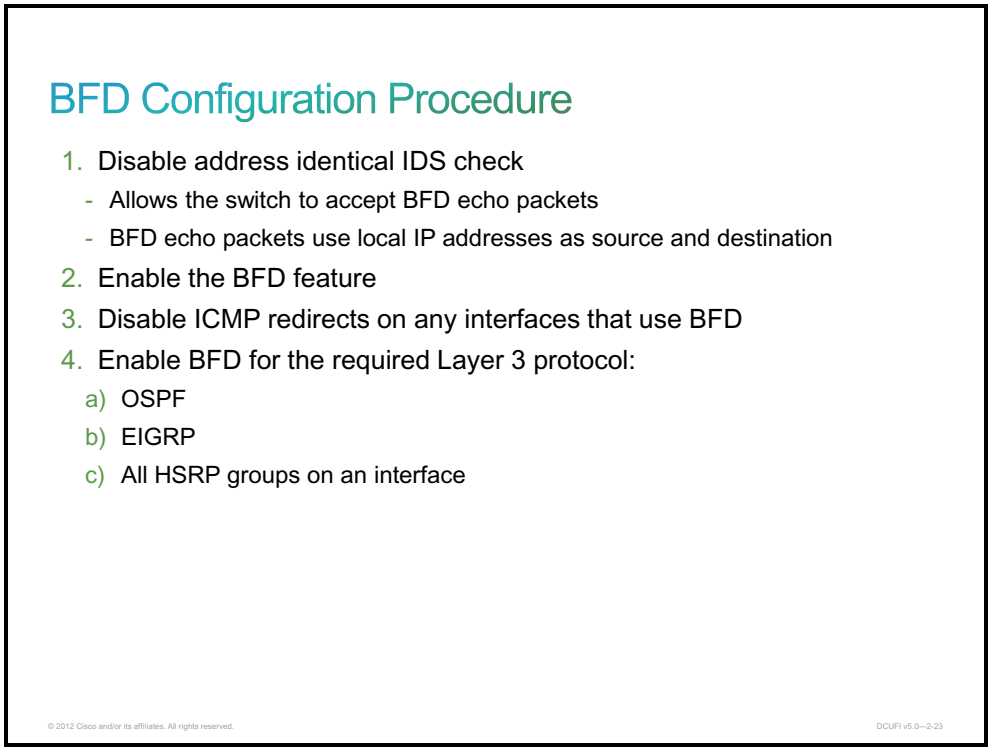
© 2012 Cisco and/or its affiliates. All rights reserved.

Many Layer 3 control protocols require a fast method of detecting link or node failures in order to achieve fast convergence. In many situations, a link or node failure can be detected through Layer 1 mechanisms. The loss of an optical or electrical signal indicates that a connection to a neighbor has failed. However, there are many other situations where Layer 1 mechanisms cannot be relied upon to accurately detect the loss of a link or neighboring device. Therefore, most Layer 3 control protocols use a hello mechanism in order to detect the loss of a neighbor. To achieve fast convergence, network administrators often tune the hello timers of the different Layer 3 control protocols that are used on the network.

BFD is a detection protocol that is designed to provide fast forwarding path failure detection to Layer 3 protocols. Those protocols include BGP, OSPF, EIGRP, IS-IS, HSRP, Protocol Independent Multicast (PIM), and even static routes.

An advantage of using BFD for fast failure detection instead of tuning the hello timers of all of the different Layer 3 protocols is that it allows the switch to detect forwarding path failures at a uniform rate rather than at variable rates for different protocol hello mechanisms. BFD provides subsecond failure detection between two adjacent devices. BFD can also be less CPU-intensive than individual protocol hello messages because some of the BFD load can be distributed onto the data plane on supported I/O modules on the Cisco Nexus 7000 Series switch.

https://t.me/learningnets



Follow these steps to implement BFD on Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches:

Step 1 Disable address identical IDS check. This allows the switch to accept BFD echo packets. BFD echo packets use local IP addresses as both source and destination.

Step 2 Enable the BFD feature.

Step 3 Disable ICMP redirects on any interfaces that use BFD.

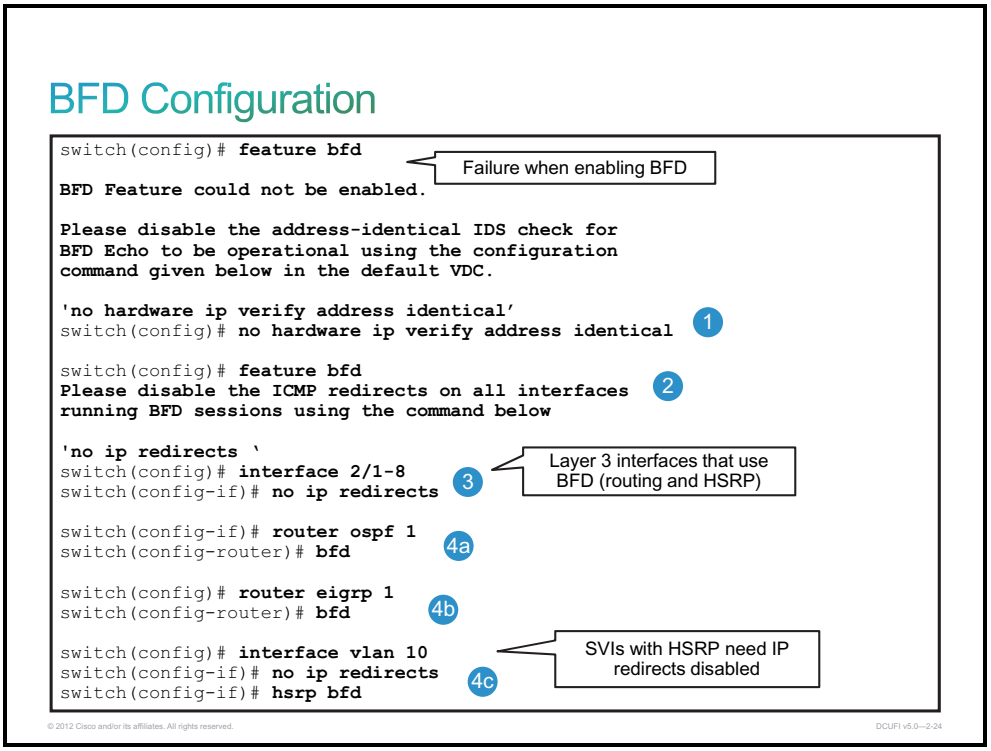
Step 4 Enable BFD for the required Layer 3 protocol, such as OSPF, EIGRP, or HSRP.

By default, BFD on the Cisco Nexus 7000 Series switches uses echo mode. In this mode, BFD echo packets are sent with the source and destination address of the echo packets that are set to the local IP address of the switch on the BFD-enabled interface. This allows the packets to be echoed back to the sender through the data plane of the neighboring switch without any interference from the BFD process, thereby reducing the CPU impact. However, the use of packets with identical source and destination addresses has some implications.

The Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches support an intrusion detection system (IDS) that verifies various fields in the IP packet header for anomalies. One of the default checks that is implemented is to verify that the source and destination IP address of a packet are not identical. However, because BFD echo mode, which is the default mode, uses packets with an identical source and destination address, BFD cannot be enabled unless this IDS check is disabled by first issuing the command **no hardware ip verify address identical** in the default VDC. After the IDS check for identical addresses has been disabled, BFD can then be enabled globally in any VDC through the **feature bfd** command.

Another issue that is caused by the use of identical source and destination addresses is that the neighboring router or switch will typically send an Internet Control Message Protocol (ICMP) redirect message back to the source when it loops the packet back through the interface upon which it was received. This mechanism can have an adverse effect on the CPU of the routers or switches. Therefore, it is recommended to disable the sending of ICMP redirects on any interface that is enabled for BFD. ICMP redirects can be disabled on an interface by using the **no ip redirects** command.

https://t.me/learningnets



The configuration in the figure illustrates how to configure the individual steps in order to implement BFD in combination with OSPF, EIGRP, and HSRP.

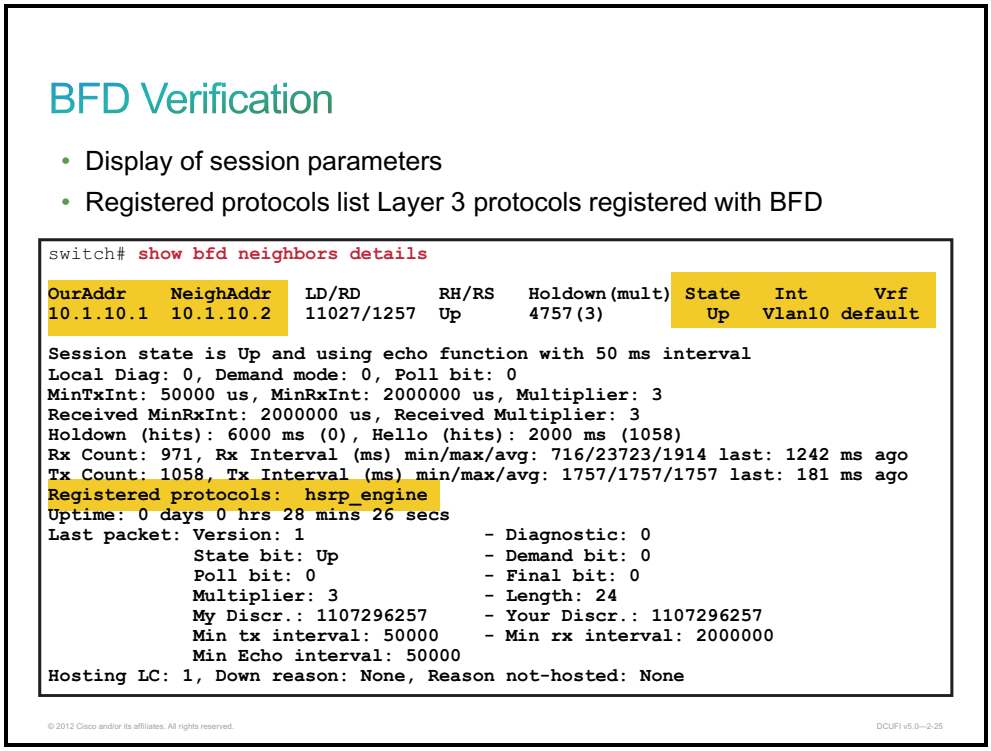
Enabling BFD globally does not start the BFD process on any interfaces until BFD is enabled for a specific Layer 3 control protocol. BFD on the Cisco Nexus 7000 Series switch can provide fast failure detection for the BGP, OSPF, EIGRP, IS-IS, HSRP, and PIM Layer 3 control protocols.

To enable BFD for the OSPF routing protocol, you should issue the **bfd** command in routing protocol configuration mode for OSPF. This enables BFD for all interfaces that are enabled for that specific OSPF process. Use the **no ip redirects** command in interface configuration mode in order to disable ICMP redirects for those interfaces.

Enabling BFD for EIGRP works in a similar way. As soon as BFD is configured under the routing process by using the **bfd** command, BFD will then be enabled on all interfaces that are enabled for that routing process.

The final example in the figure shows how to enable BFD for HSRP. In this case, BFD is not enabled under a globally configured process but rather under the Layer 3 interfaces that are enabled for HSRP. The command **hsrp bfd** enables BFD for all HSRP groups that are configured on that particular interface.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



Use the **show bfd neighbors** command in order to verify the operation of BFD. In the figure, the **detailed** option of the command is used. The output shows the relevant parameters and statistics necessary to verify the operation of the BFD process.

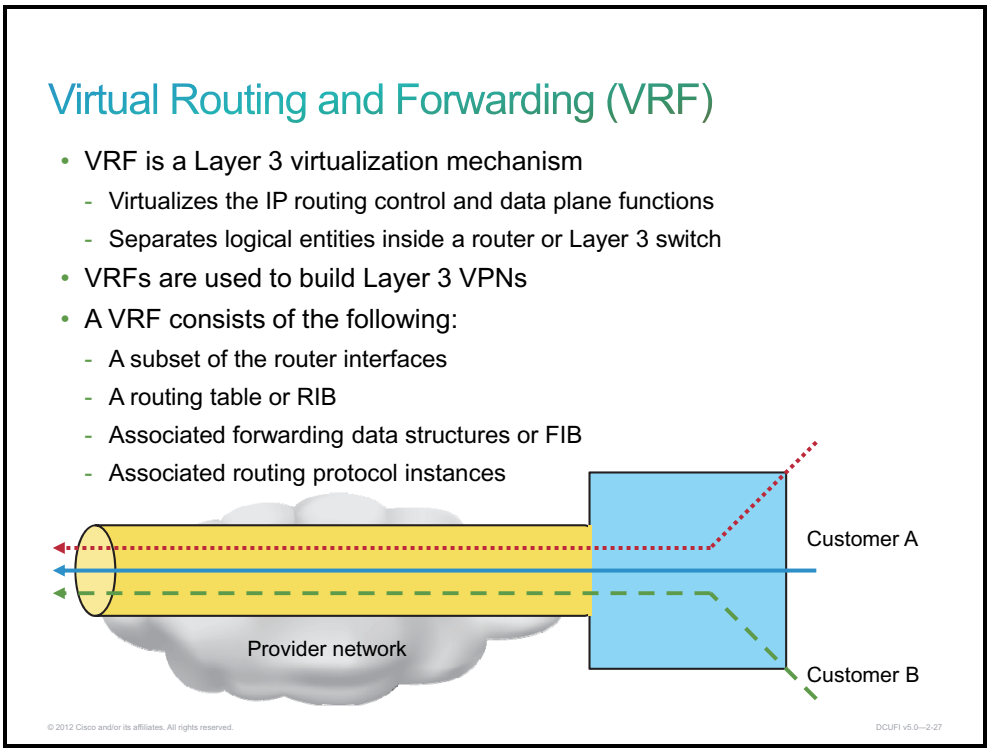
Some of the key fields in the output of this command are:

- **OurAddr**: This field lists the IP address of the interface for which the **show bfd neighbors** command was entered.
- **NeighAddr**: This field lists the IP address of the BFD adjacency or neighbor.
- **State**: This field lists the state of the interface as either “Up” or “Down.”
- **Int**: This field lists the interface type and slot/port.
- **Session state and mode**: This field identifies whether the BFD session is up and whether or not echo mode is used.
- **Registered protocols**: This field identifies the Layer 3 control protocols that have been registered with BFD.

https://t.me/learningnets

Layer 3 Virtualization

This topic identifies the use and configuration of Layer 3 virtualization on the Cisco Nexus 7000 Series switch.



To provide logical Layer 3 separation within a Layer 3 switch or router, the data plane and control plane functions of the device need to be segmented into different Layer 3 VPNs. This process is similar to the way that a Layer 2 switch segments the Layer 2 control plane and data plane into different VLANs.

The core concept in Layer 3 VPNs is a VRF instance. This instance consists of all of the data plane and control plane data structures and processes that together define the Layer 3 VPN.

A VRF includes the following components:

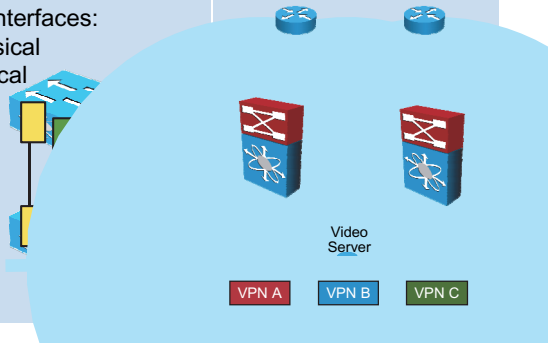
- **A subset of the Layer 3 interfaces on a router or Layer 3 switch**: Similar to how Layer 2 ports are assigned to a particular VLAN on a Layer 2 switch, the Layer 3 interfaces of the router are assigned to a VRF. Because the elementary component is a Layer 3 interface, this component includes software interfaces, such as subinterfaces, tunnel interfaces, loopback interfaces, and switch virtual interfaces (SVIs).
- **A routing table or routing information base (RIB)**: Because traffic between Layer 3 interfaces that are in different VRFs should remain separated, a separate routing table is necessary for each individual VRF. The separate routing table ensures that traffic from an interface in one VRF cannot be routed to an interface in a different VRF.
- **A forwarding information base (FIB)**: The routing table or RIB is a control plane at a structure. An associated FIB is calculated from it, which is then used in the actual packet forwarding. This also needs to be separated by a VRF.
- **Routing protocol instances**: To ensure control plane separation between the different Layer 3 VPNs, it is necessary to implement routing protocols on a per-VRF basis. To accomplish this task, you can run an entirely separate process for the routing protocol in the VRF. Alternately, you can use a subprocess or routing protocol instance within a global process that is in charge of the routing information exchange for the VRF.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

VRF Support

Platform	Cisco Nexus 5500	Cisco Nexus 7000
Supported feature	VRF-Lite	Full VRF functionality
Requirements	• Cisco NX-OS Release 5.0(3)N1(1b) • Layer 3 LAN-Enterprise License • Layer 3 Module	Enterprise Services License
Deployment scenarios	VRF interfaces: • Physical • Logical	



© 2012 Cisco and/or its affiliates. All rights reserved.

Beginning with Cisco NX-OS Release 5.0(3)N1(1b), the Cisco Nexus 5500 Platform switch supports VRF-Lite with a Layer 3 LAN Enterprise license, and you can also create a VRF and assign the interface to a VRF. Prior to this release, two VRFs were created by default: the VRF management and VRF default. The management interface (mgmt0) and all SVI interfaces resided in the VRF management and VRF default, respectively.

Cisco Nexus 7000 Switches support an extensive VRF and MPLS functionality and can be deployed in MPLS VPN environments. It supports Layer 3 VPNs and MPLS Traffic Engineering. It does not, however, support Layer 2 VPNs such as AToM (point-to-point) or Virtual Private LAN Service (VPLS, multipoint).

https://t.me/learningnets

VRF Configuration Procedure

1. Create VRF(s)
   - In default or nondefault VDC (Cisco Nexus 7000)
   - VRFs in different VDCs are completely independent
2. Assign Layer 3 interfaces to the VRF
3. Configure VRF static routes (optional)
   - In VRF configuration mode
4. Enable a routing process for the VRF (optional)
   - Associate VRF with the routing protocol

© 2012 Cisco and/or its affiliates. All rights reserved.

Follow these steps in order to implement VRFs on Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches:

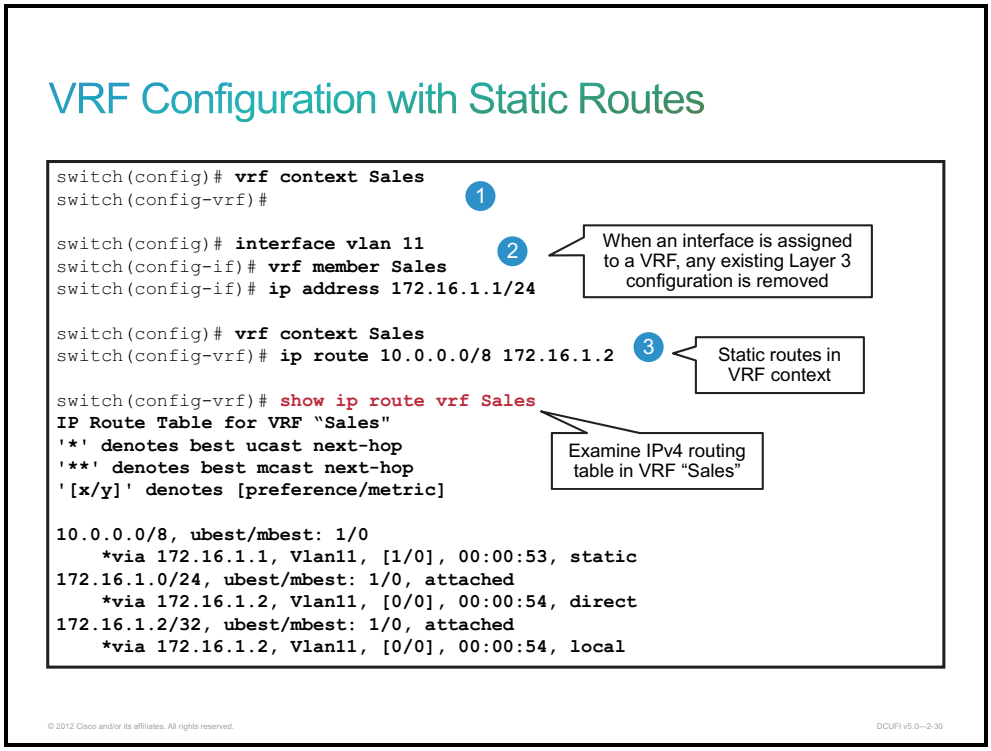
Step 1 Create VRF(s). On Cisco Nexus 7000 Series switches, you can define VRFs in default or nondefault VDCs. VRFs in different VDCs are completely independent from one another. Each VRF contains a separate address space with unicast and multicast route tables for IPv4 and IPv6, which makes routing decisions independent of any other VRF. A VRF name is local to a VDC. Two VRFs can be configured with the same name as long as they exist in different VDCs. At the very least, each VDC has two VRFs—a default VRF and a management VRF. All Layer 3 interfaces and routing protocols exist in the default VRF until they are assigned to another VRF. The mgmt0 interface exists in the management VRF and is accessible from any VDC.

Step 2 Assign Layer 3 interfaces to the VRF.

Step 3 Configure VRF static routes (optional). VRF-specific static routes are configured in VRF configuration mode.

Step 4 Enable a routing process for the VRF (optional). Depending on the routing protocol, you will use different methods to associate a VRF with the routing protocol.

https://t.me/learningnets



The first example illustrates a VRF scenario with static routing.

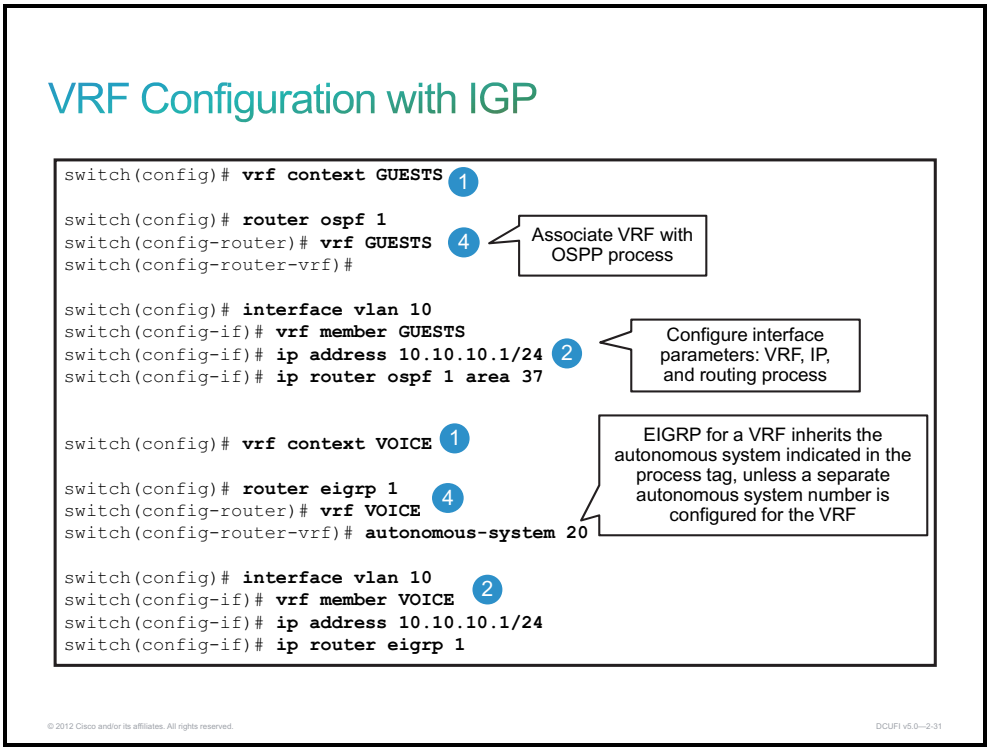
You create a VRF by using the **vrf context** command. After the VRF has been created, Layer 3 interfaces—such as SVIs, routed ports, routed port channels, tunnel interfaces, and loopback interfaces—can be assigned to it by using the **vrf member** command.

Note	When you change the VRF association of an interface, any existing Layer 3 configuration on that interface is removed.

Finally, you configure static routes. Static routes for nondefault VRFs are configured in **vrf context** configuration mode.

The example in the figure shows how to examine the VRF routing table by using the **show ip route vrf** command. The equivalent **show routing vrf** command can also be used to display the routes in the routing table.

https://t.me/learningnets



To enable a routing process for a VRF, the VRF must first be associated with a routing process in routing protocol configuration mode by using the **vrf** command. In the VRF submode of the specific routing protocol, you can configure the routing protocol parameters that are specific for the specific associated VRF.

The first example in the figure shows how to configure OSPF for a VRF, associate an interface with a VRF, and successively enable OSPF on the interface for that particular VRF.

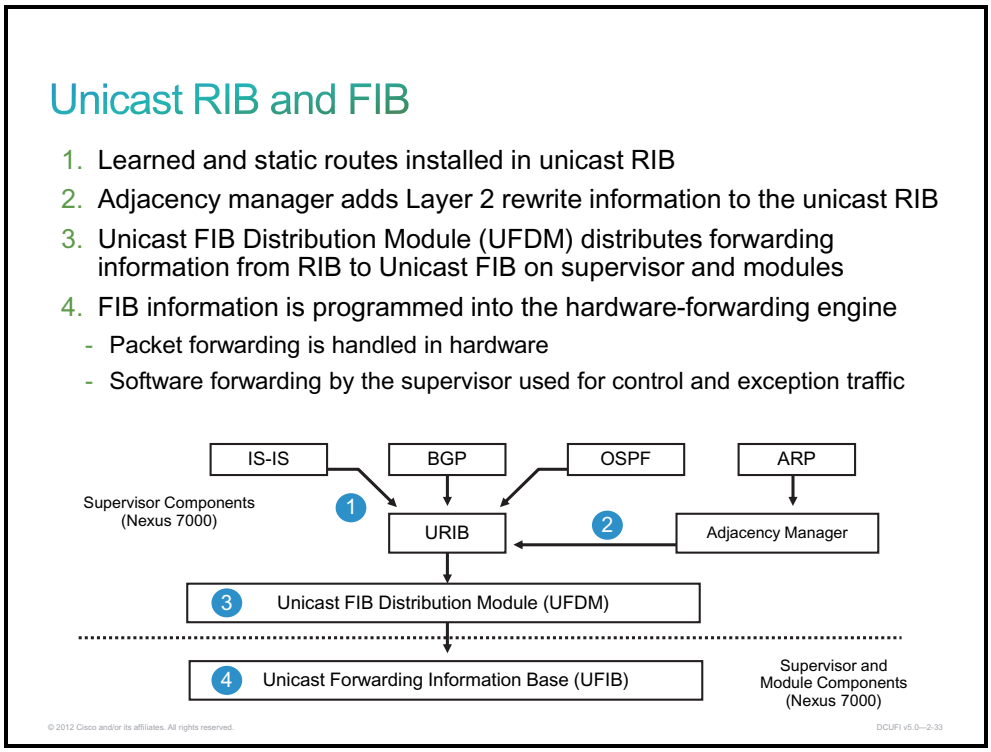
The configuration of all routing processes for a VRF follows a similar structure. The second example in the figure shows how to enable EIGRP for a VRF.

If a number is used for the EIGRP process instead of a nonnumeric tag, this number is then used as the default autonomous system number for all VRFs. To specify a separate autonomous system for a VRF, use the **autonomous-system** command in the VRF submode for the VRF associated with the EIGRP routing process.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Unicast RIB and FIB

This topic identifies how to manage the unicast RIB and FIB on the Cisco Nexus 7000 Series switch.



The Cisco Nexus 7000 Series switch forwarding architecture consists of multiple components. The unicast RIB exists on the active supervisor to maintain the routing table with directly connected routes, static routes, and routes that are learned from dynamic unicast routing protocols. The unicast RIB also collects adjacency information from sources such as the ARP. The unicast RIB determines the best next hop for a given route and populates the unicast FIB on the supervisors and modules by using the services of the unicast FIB distribution module (UFDM).

Cisco NX-OS Software supports distributed packet forwarding. The ingress port takes relevant information from the packet header and passes that information on to the local switching engine. The local switching engine performs the Layer 3 lookup and then uses this information to rewrite the packet header. The ingress module forwards the packet to the egress port. If the egress port is on a different module, the packet is then forwarded by using the switch fabric to the egress module. The egress module does not participate in the Layer 3 forwarding decision.

The software-forwarding path is used primarily to manage features that are not supported in hardware or to manage errors that are encountered during hardware processing. Typically, packets with IP options or packets that require fragmentation are passed to the CPU on the active supervisor. All packets that should be switched in software or terminated go on to the supervisor. The supervisor uses the information that is provided by the unicast RIB and the adjacency manager to make the forwarding decision. The module is not involved in the software-forwarding path.

The adjacency manager exists on the active supervisor and maintains adjacency information for different protocols, including ARP, Neighbor Discovery Protocol (NDP), and static configuration. Outgoing Layer 2 packets use the adjacency information to complete the Layer 2 header. The adjacency manager can trigger ARP requests to find a particular Layer 3-to-Layer 2 mapping.

https://t.me/learningnets

The UFDM exists on the active supervisor and distributes the forwarding path information from the unicast RIB and other sources. The unicast RIB generates forwarding information that the unicast FIB programs into the hardware-forwarding tables on the standby supervisor and the modules. The unicast RIB also downloads the FIB information to newly inserted modules.

The UFDM gathers adjacency information, rewrite information, and other platform-dependent information when updating routes in the unicast FIB. The adjacency and rewrite information consists of interface, next-hop, and Layer 3-to-Layer 2 mapping information. The interface and next-hop information is received in route updates from the unicast RIB. The Layer 3-to-Layer 2 mapping is received from the adjacency manager.

The unicast FIB exists on the supervisors and switching modules. The unicast FIB then builds the information that is used for the hardware-forwarding engine. The unicast FIB receives route updates from the UFDM and sends this information along to be programmed in the hardware-forwarding engine. The unicast FIB controls the addition, deletion, and modification of routes, paths, and adjacencies.

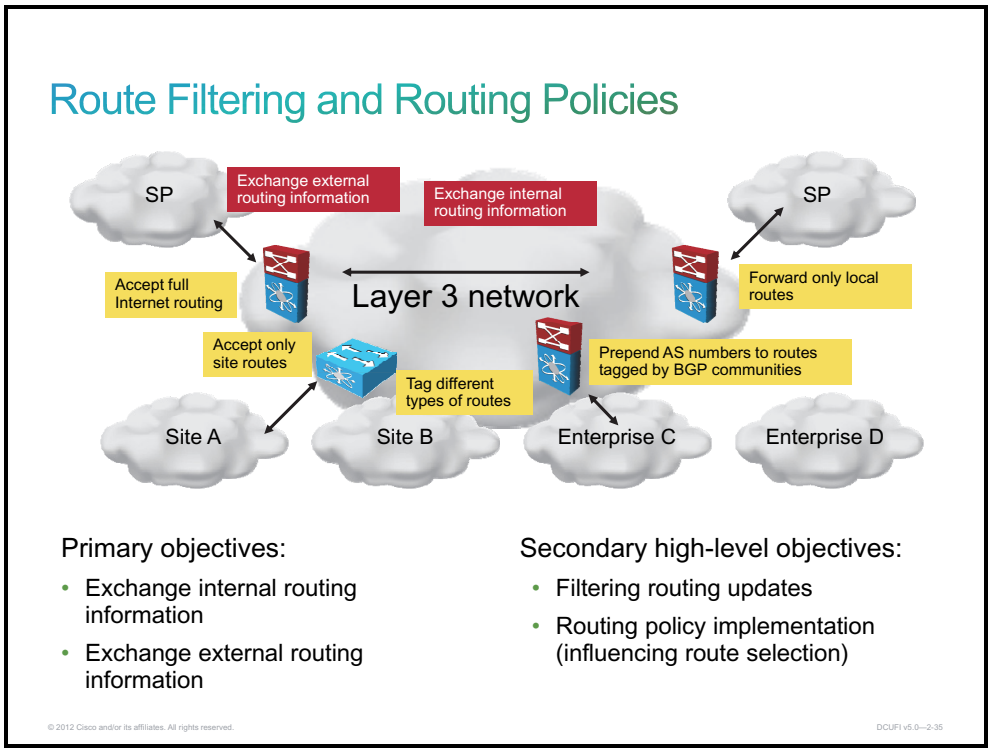
The unicast FIBs are maintained on a per-VRF and per-address family basis. There is one unicast FIB for IPv4 and IPv6 for each configured VRF. Based on route update messages, the unicast FIB maintains a per-VRF prefix and next-hop adjacency information database. The next-hop adjacency data structure contains the next-hop IP address and the Layer 2 rewrite information. Multiple prefixes could share a next-hop adjacency information structure.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Route Policy Manager

This topic identifies the use and configuration of the Route Policy Manager on the Cisco Nexus switch.



The figure illustrates various actions that can be applied to routing updates on Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches. The actions can be divided into two main categories:

- Exchanging routing information (the primary objective of routing protocols)
- Implementing a routing policy and filtering of routing information

To exchange routing information, a typical service provider would use two routing protocols:

- An IGP, such as OSPF, or IS-IS to exchange local routing information
- BGP to exchange external routing information (that is, customer routing information and complete Internet routing information from other service providers)

BGP will always be combined with advanced filtering and policy mechanisms for security and performance reasons.

https://t.me/learningnets

Tools for Route Filtering and Routing Policies

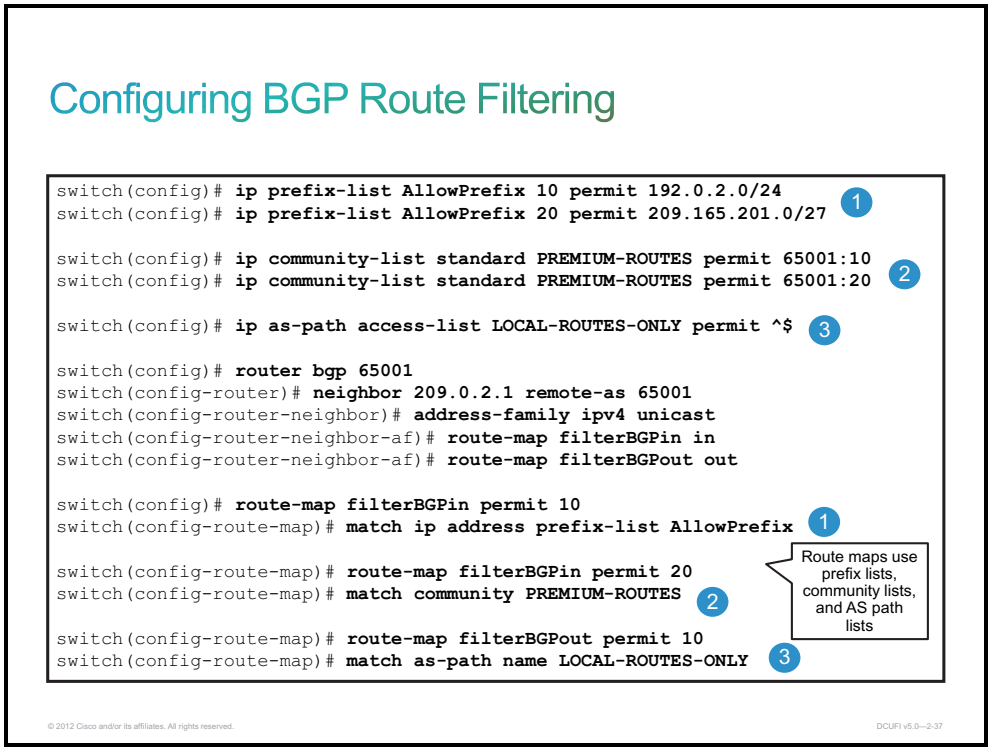
Tool	Description
Prefix lists	• Used for prefix-based filtering or matching of routes • Can be used to match on the prefix, route source or next-hop address • IPv4 and IPv6
AS path access lists	• Used in BGP for filtering or route matching based on AS path attribute • Uses regular expressions
Community lists	• Used in BGP for filtering or route matching based on standard or extended communities • Various matching options
Route maps	• Primarily used to implement complex routing policies • Secondary use as filtering tool • In BGP and at redistribution

© 2012 Cisco and/or its affiliates. All rights reserved.

The route policy manager provides route filtering and route manipulation capabilities that can be used to implement routing policies. The route policy manager provides the following route filtering options:

- **Prefix lists**: Prefix lists can be used to permit or deny ranges of IPv4 and IPv6 (Cisco Nexus 7000 Series switch) prefixes. Prefix lists can be used in route maps to match a set of IPv4 or IPv6 (Cisco Nexus 7000 Series switch) prefixes.
- **AS path lists**: AS path lists can be used to select BGP routes that are based on the BGP AS path attribute. Regular expressions are used to match patterns in the AS path string, which contains the AS numbers of the autonomous systems that the BGP update has passed through. AS path lists can be used in route maps, or they can be directly applied to BGP routes received from or sent to a neighbor.
- **Community lists**: Community lists can be used to select BGP routes that are based on BGP community attributes. There are two types of community lists—standard and expanded. *Standard community lists* specify simple lists of community values that are permitted or denied. *Expanded community lists* use regular expressions to match patterns in the community string. Community lists can be used in route maps.
- **Route maps**: Route maps can be used to permit or deny routes to match specific selection criteria. In addition, if the routes are permitted, the associated attributes of the route, such as metrics or BGP attributes, can be manipulated to implement a routing policy. Route maps can be applied to a route-redistribution process. Alternately, they can be used with BGP to filter or manipulate BGP routes that are received from or sent to a BGP neighbor. Route maps can also be used to implement PBR. However, in this particular case, the **match** statements in the route map should match packets instead of routes.

https://t.me/learningnets



The figure shows an example of a prefix list, AS path list, and community list that are tied into two route maps that are then used for BGP inbound and outbound filtering. The following commands are used:

- **ip prefix-list** *name* [**seq** *number*] {**permit** | **deny**} *prefix* [**eq** *length* | [**ge** *length] [le length*]]: This command configures IP prefix filtering. You can configure prefix lists to match an exact prefix or a range of prefixes that falls within the specified prefix. Use the ge and le keywords to specify a range of the prefix lengths to match, which provides more flexible configuration than can be achieved with only the network/length argument. Cisco NX-OS Software processes the prefix list using an exact prefix match when you do not configure either the ge or le keyword. If you configure both the ge *ge-length* and le *le-length* keywords and arguments, the allowed prefix length range falls between the values that are used for the ge-length and le-length arguments. In addition to the ge and le keywords, the Cisco NX-OS Software also provides an eq keyword that matches all of the prefixes within the main prefix that have the exact prefix length that is specified by the eq keyword.
- **ip as-path access-list** *name* {deny | permit} *regexp*: This command is used to configure BGP AS path filters, which permit or deny BGP routes that are based on the BGP AS path attribute. A regular expression is used to match patterns in the AS path string.
- **ip community-list standard** *list-name* {deny | permit} {*aa*:*nn* | internet | local-as | no-advertise | no-export}: This command is used to configure standard BGP community lists, which permit or deny BGP routes that are based on BGP community attributes. The communities are specified in the 4-byte, new community format, which consists of two decimal numbers in the range 1 to 65535 and separated by a colon. To match the well-known community values of Internet, local-AS, no-advertise, and no-export, special keywords are defined.

https://t.me/learningnets

Configuring Route Redistribution

1. Redistribute updates matched by PRIVATE-SUBNETS to OSPF with default cost (20)
2. Redistribute updates matched by CAMPUS-SUBNETS to OSPF with custom cost (10)
3. Redistribute all other updates with custom cost 1000
4. Link the route map to the redistribution

```
switch(config)# route-map EIGRP-TO-OSPF deny 10
switch(config-route-map)# match ip address prefix-list PRIVATE-SUBNETS      1

switch(config-route-map)# route-map EIGRP-TO-OSPF permit 20
switch(config-route-map)# match ip address prefix-list CAMPUS-SUBNETS       2
switch(config-route-map)# set metric 10

switch(config-route-map)# route-map EIGRP-TO-OSPF permit 30                  3
switch(config-route-map)# set metric 1000
                                  Filtering is mandatory in redistribution
switch(config)# router ospf 1
switch(config-router)# redistribute eigrp 200 route-map EIGRP-TO-OSPF       4
```

© 2012 Cisco and/or its affiliates. All rights reserved.

In addition to BGP route filtering, route maps are used to filter routes when redistributing routing information between routing protocols. The Cisco NX-OS Software running on the Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches does not allow unfiltered redistribution. The configuration of a route map is a mandatory component when configuring redistribution on a Cisco Nexus switch. The example in the figure shows how to configure route redistribution between EIGRP and OSPF by using a route map to filter the traffic and manipulate the OSPF cost. The following commands are used:

- **route-map** *map-tag* [**deny** | **permit**] [*sequence-number*]: This command creates a single entry in a route map. The permit or deny statement specifies whether the prefixes matched by this route map entry will be forwarded or dropped. Sequence numbers are used to determine the order in which the different route map entries will be evaluated.
- **match ip address prefix-list** *prefix-list-name* [*prefix-list-name...*]: This command specifies one or more prefix lists to be used to match the prefixes that are evaluated by the specified route map entry. If multiple prefix lists are specified, they are then filtered with “or” semantics. If any one of the specified prefix lists is matched, this match is treated as a successful match for the entire match statement.
- **set metric** *bandwidth-metric*: This command *s*ets the metric value for a routing protocol.
- **redistribute** {bgp *as-number* | direct | eigrp *id* | isis *instance-tag* | ospf *instance-tag* | rip *instance-tag* | static} route-map *map-name*: This command injects routes from the specified routing source into the routing protocol under which it is configured. A route map is used to filter the routes that are being redistributed. The configuration of the route map is mandatory: Cisco NX-OS Software does not allow unfiltered redistribution.

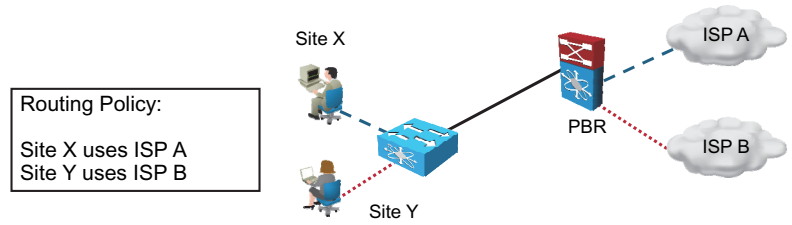
https://t.me/learningnets

Policy-Based Routing (PBR)

This topic identifies the use and configuration of policy-based routing on a Cisco Nexus switch.

Policy-Based Routing (PBR)

- Normal unicast routing is destination-based.
- PBR allows routing decisions to be based on different characteristics of the packets, such as the following:
  - Source IP address
  - TCP or UDP port numbers
  - Packet length
- PBR can be used to implement routing policies
- Available on Cisco Nexus 5500 and 7000 switches



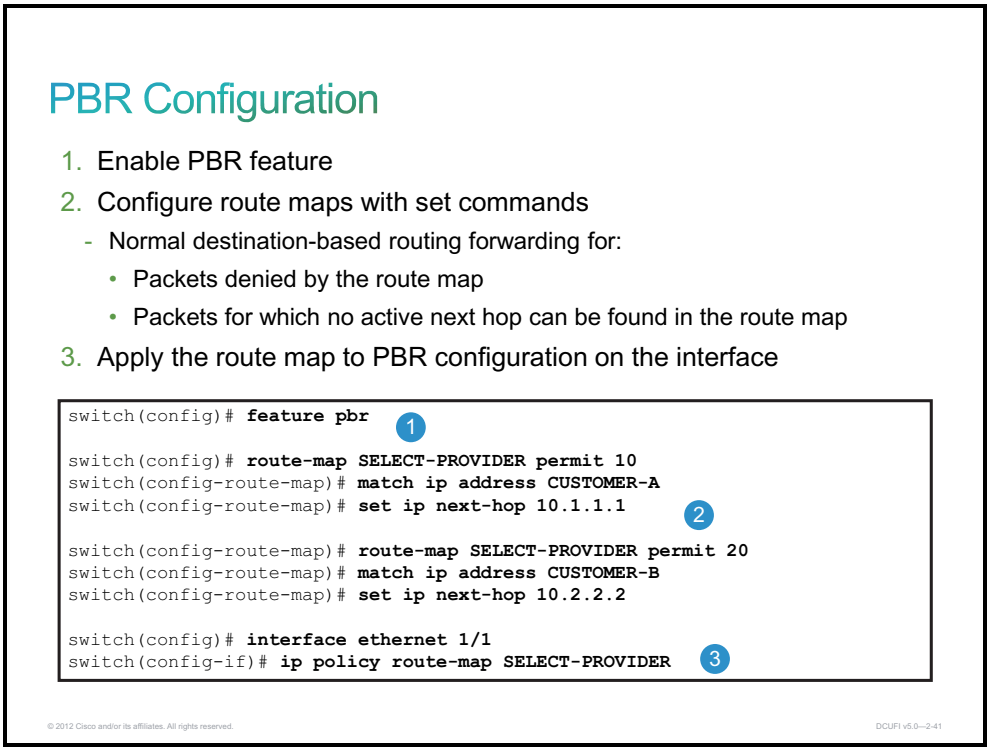
© 2012 Cisco and/or its affiliates. All rights reserved.

Normal unicast routing is based on the destination of a packet. The destination IP address in the packet header is used as the input for a longest match search through the FIB, which results in a next-hop IP address and an associated outbound interface. PBR allows routing decisions to be made based on different characteristics of the packet, such as the source IP address, protocol field, packet length, and even Layer 4 characteristics, such as source and destination TCP or UDP port numbers.

PBR can be used to implement routing policies. For example, PBR could be used for source-based provider selection: Packets are evaluated by a route map and if they come from a specific source address range they are forwarded to a specific ISP, while packets coming from other sources may be forwarded to other ISPs. It is also possible to forward packets that are based on application, as determined by the UDP or TCP port numbers in the packet header.

PBR is not a dynamic routing mechanism. Rather, the policy is implemented on a per-hop basis by using static route maps. However, multiple next hops can be specified, which will be evaluated until a valid next hop is found in order to provide redundancy. It is possible to specify load sharing among up to 16 next hops in order to provide both redundancy and load sharing.

https://t.me/learningnets



In order to configure PBR, you need to enable the feature and configure a route map. The route map consists of a series of entries that are evaluated in order based on their sequence number. Each entry in a route map contains a combination of **match** and **set** statements. The **match** statements define the criteria for packets to meet in order to be managed through the policy that is defined in the specified entry. The **set** clauses define how the packets should be routed if they match the specified criteria.

Route map statements can be marked as **permit** or **deny**. If the statement is marked as a **deny**, the packets that meet the **match** criteria are managed through the normal forwarding channels, which means that normal destination-based routing is performed. If the statement is marked as **permit** and the packets meet the match criteria, all the set clauses are evaluated in order until a valid next hop is found. If no valid next hop can be found, then those packets are also forwarded through the normal routing channel.

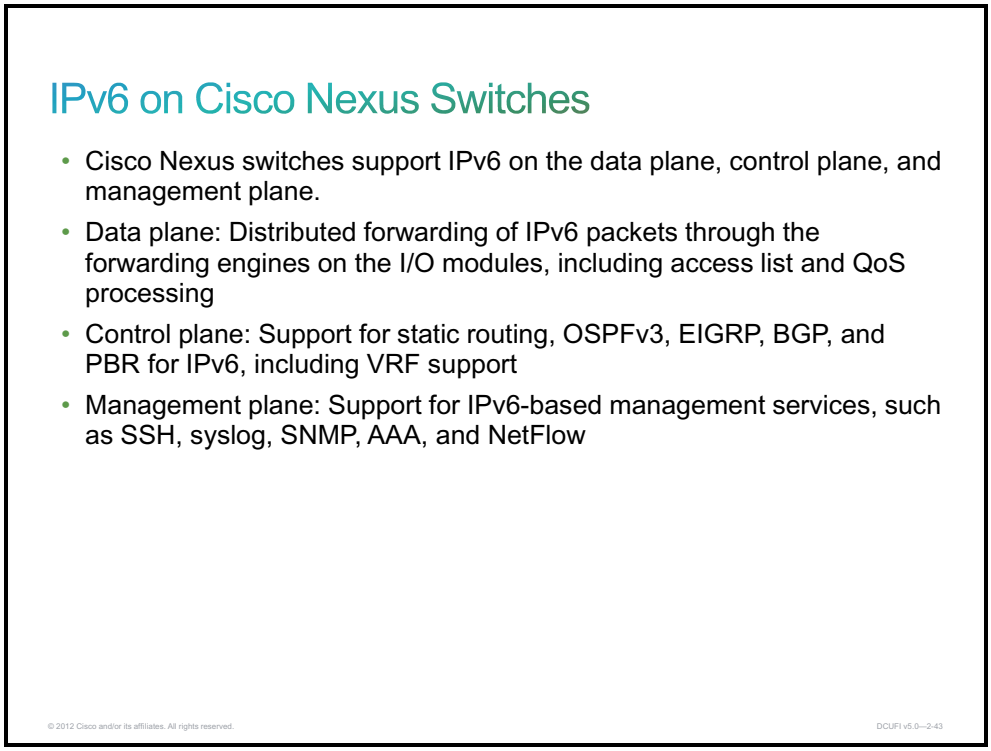
Route maps used for PBR must be associated with one or more ingress Layer 3 interfaces in order to apply the routing policy to the packets that are received on the specified interfaces.

The example in the figure shows how to configure a route map for PBR and apply it to an ingress Layer 3 interface.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

IPv6

This topic identifies the implications of using IPv6 in the data center.

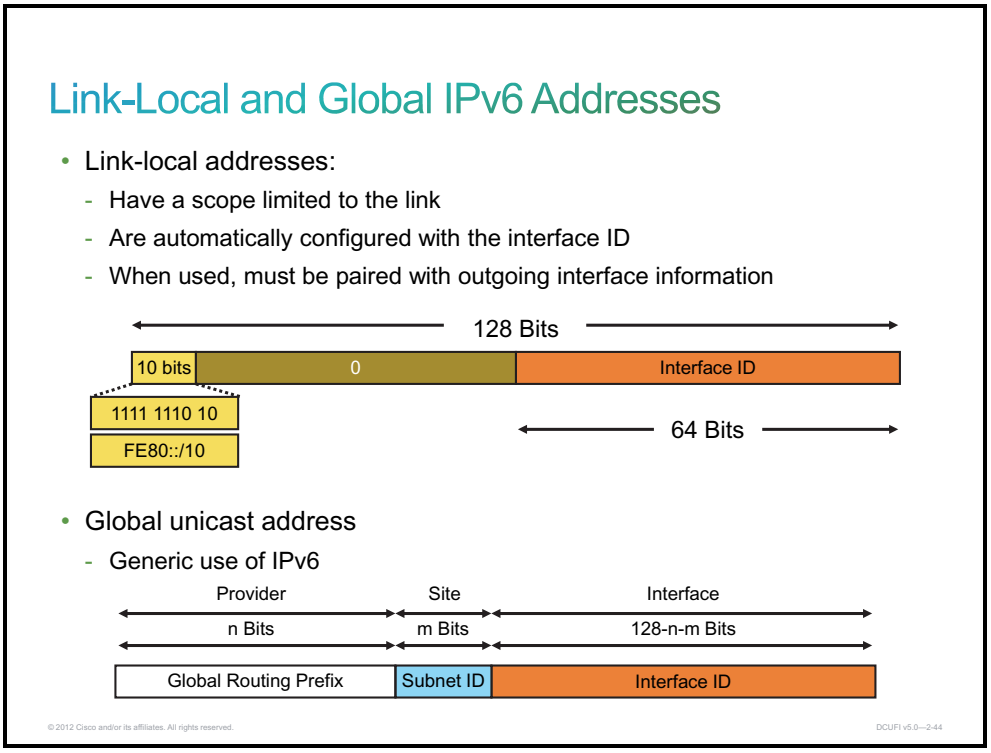


As global IPv4 address space becomes more difficult to obtain, there will be an increasing need to deploy services on IPv6 within the data center. In order to support these new services, the data center network infrastructure needs to be capable of supporting IPv6.

The Cisco NX-OS Software includes a wide range of features that are necessary to integrate IPv6 services in the data center. The IPv6 features in the Cisco NX-OS Software encompass the data plane, control plane, and management plane functions of the Cisco Nexus switches.

- **Data plane**: The M1 forwarding engines on the I/O modules of the Cisco Nexus 7000 Series switches are capable of forwarding IPv6 packets in hardware at 30 mpps, including security and QoS processing. The XL I/O modules are capable of storing up to 350,000 IPv6 routes in the ternary content addressable memory (TCAM) of the forwarding engine, while non-XL I/O modules can hold up to 64,000 IPv6 routes. The exact number of routes that can be stored in the TCAMs is dependent upon prefix distribution.
- **Control plane**: The Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches support the OSPFv3, EIGRP, and BGP routing protocols for IPv6, in addition to static routing and PBR. IPv4 and IPv6 routing can be virtualized using VRFs.
- **Management plane**: Cisco Nexus switches can use IPv6 as the transport protocol for various management services. For example, network management protocols and services such as SSH, syslog, Simple Network Management Protocol (SNMP), and authentication, authorization, and accounting (AAA) can be implemented based on IPv6. IPv6 support is also included in network management protocols, such as NetFlow.

https://t.me/learningnets



When configuring IP addresses on Cisco NX-OS router interfaces, some IPv6-specific aspects require special attention. These aspects are:

- **Link-local addresses**: The two approaches that are used for link-local addresses are the automatic IP address assignment when the interface is enabled for IPv6 and static configuration.
- **Global addresses**: The available options include EUI-64 autoconfiguration and static configuration.

Link-Local Addresses

All IPv6-enabled interfaces must have a link-local address. Link-local addresses are used for addressing on a single link, meaning that they have a scope that is limited to the link. Link-local addresses are created dynamically on all IPv6 interfaces by using a specific link-local prefix, FE80::/10, as well as a 64-bit interface identifier.

Link-local addresses are used for automatic address configuration, neighbor discovery, and router discovery. Many routing protocols also use link-local addresses. Link-local addresses can serve as a means to connect devices on the same local network without requiring global or unique local addresses. When communicating with a link-local address, you must specify the outgoing interface because every interface connects to FE80::/10.

Global Addresses

Global unicast addresses correspond to the principal use of IPv6 addresses for generic global IPv6 traffic. The structure of a global unicast address is as follows:

- A global routing prefix, typically a /48, is assigned to a site.
- A subnet identifier, typically 16 bits, is used to identify links within a site.
- A 64-bit interface identifier identifies the interface of the node.

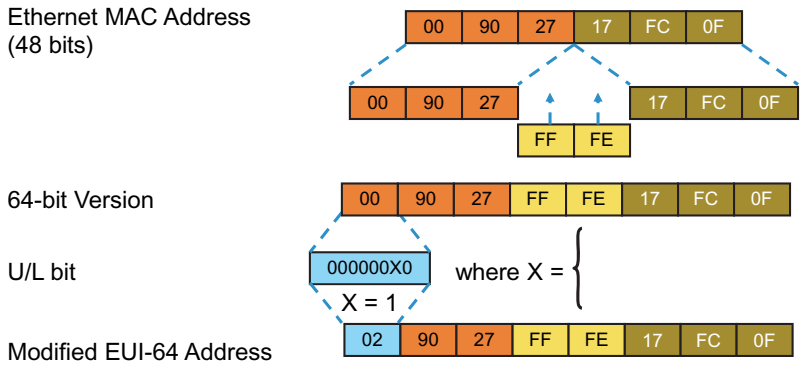
The interface identifier can be of any length but should be kept at 64 bits since the stateless autoconfiguration of hosts depends on the 64-bit length of the interface identifier.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Modified EUI-64 Format

Autogenerated interface ID:

- Inserted "FFFE"
- U/L bit identifies the uniqueness of the MAC address.



© 2012 Cisco and/or its affiliates. All rights reserved.

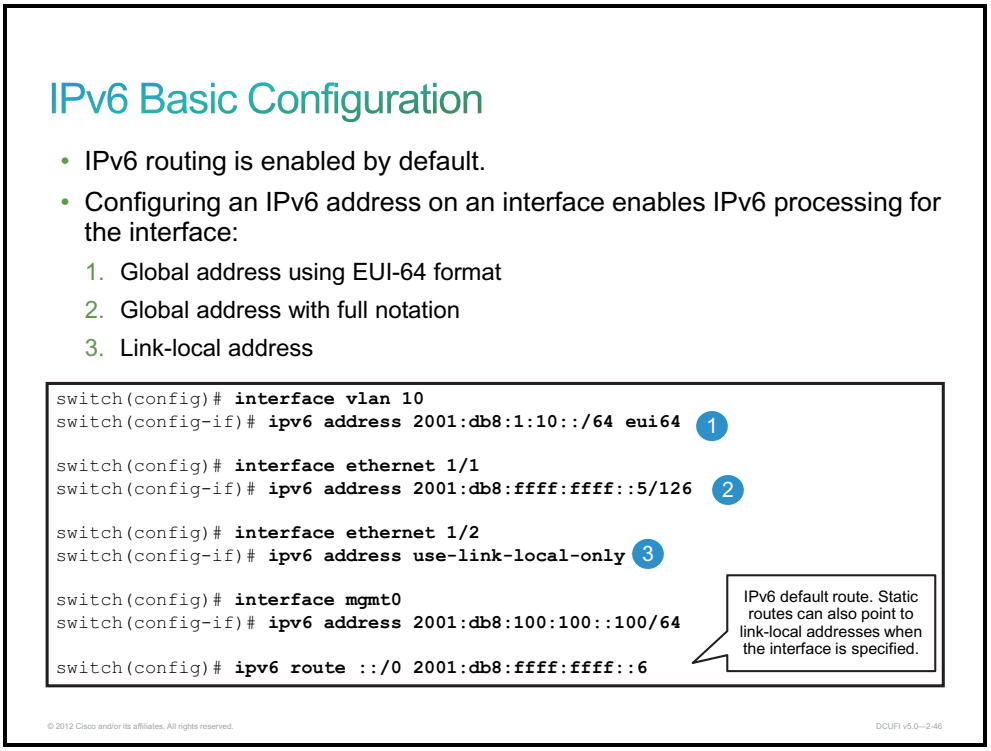
Having a much larger address space available, IPv6 engineers designed a way to enable autoconfiguration of the addresses while still keeping the global uniqueness. A host or router can autoconfigure itself by appending its data link layer address (in a special 64-bit extended unique identifier -64 format) to the local link prefix (64 bits). This autoconfiguration results in a complete 128-bit IPv6 address that is usable on the local link and is, most likely, globally unique.

The interface identifier for stateless autoconfiguration in an Ethernet environment uses the modified extended unique identifier -64 format. The extended unique identifier -64 format expands the 48-bit Ethernet MAC address format to a 64-bit version by inserting “FFFE” into the middle of the 48 bits.

The seventh bit (starting with the leftmost bit) in an IPv6 interface identifier is referred to as the Universal/Local (U/L) bit, which identifies whether this interface identifier is universally unique or is locally unique on the link. If the interface identifier was created from an Ethernet MAC address, it is assumed that the MAC address is universally unique, and thus, so is the interface identifier.

The U/L bit is for future use of the upper-layer protocols to uniquely identify a connection, even when there is a change in the leftmost part of the address. However, this feature is not yet in use.

https://t.me/learningnets

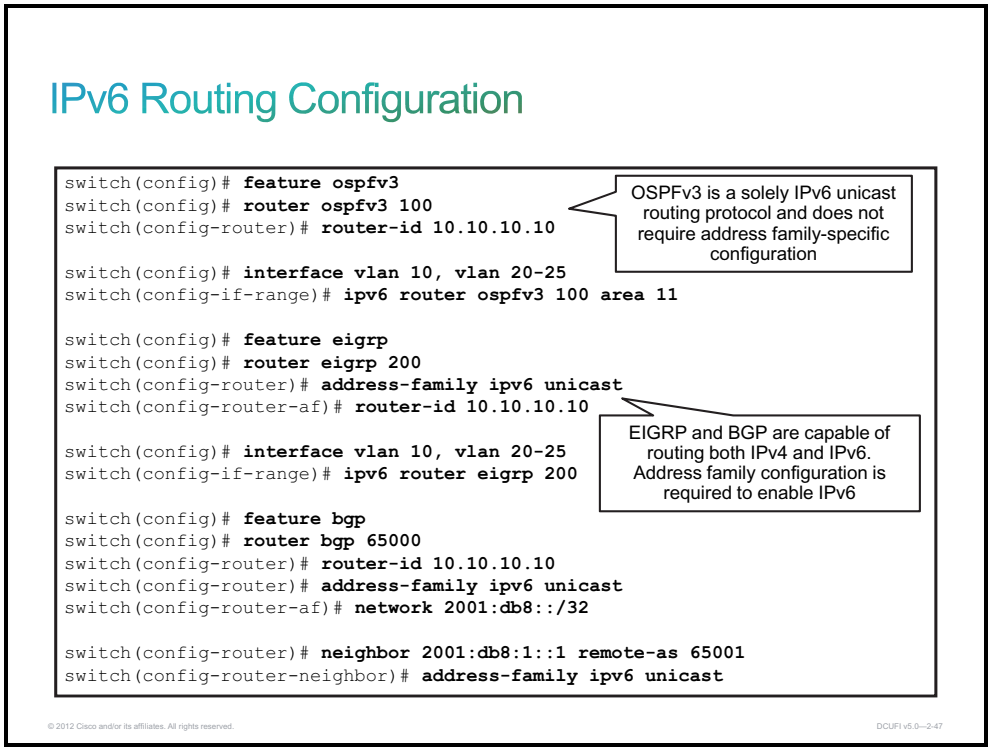


IPv6 unicast routing is enabled by default. To enable IPv6 on an interface, it is necessary to first configure an IPv6 address. Commonly, globally unique unicast IPv6 addresses are assigned to all interfaces, but for point-to-point interfaces, it is also possible to assign a link-local address only.

The example in the figure shows how to configure basic IPv6 routing. The following commands are used:

- **ipv6 address** {*address* [eui64] [secondary] | use-link-local-only]}: This command configures an IPv6 address and prefix for an interface. If the extended unique identifier -64 is used, only the first 64 bits of the prefix need to be specified. To enable IPv6 without specifying a globally routable IPv6 address, you can use the use-link-local-only option. To specify multiple IPv6 addresses on an interface, use the secondary keyword to configure additional addresses.
- **ipv6 route** *ipv6-prefix/length* {{*next-hop-addr* / *next-hop-prefix*} | *interface* / *link-local-addr*}: This command is used to configure a static IPv6 route. If this command is configured in global configuration mode, the route is installed in the routing table of the default VRF. To configure a static IPv6 route for a nondefault VRF, this command needs to be issued in VRF configuration mode for the specific VRF context.

https://t.me/learningnets



The figure shows examples of a basic EIGRP configuration and a basic BGP configuration for IPv6 on a Cisco Nexus 5500 Platform switch or a Cisco Nexus 7000 Series switch. The figure shows how address family configuration is used to specify the parameters that are specific to IPv6. The following commands are used:

- **feature eigrp**: This command enables the EIGRP feature. This feature requires the Enterprise Services License.
- **router eigrp** *as-number*: This command starts the EIGRP process. The instance tag can be any case-sensitive alphanumeric string of up to 20 characters. However, if a nonnumeric instance tag is used, then the autonomous system number for EIGRP must be separately specified using the **autonomous-system** command. If a numeric instance ID is used, then the autonomous system number is equal to the instance tag.
- **address-family {ipv4 | ipv6} {unicast | multicast}**: This command enters the global address family configuration mode for an address family and enables the address family for the routing protocol. In this mode, the specific options for the selected address family can be configured.
- **router-id** *ip-address*: This command configures the EIGRP router ID. This command is optional but strongly recommended.
- **ipv6 router eigrp** *as-number*: This command enables the associated EIGRP process for IPv6 on an interface.
- **feature bgp**: This command enables the BGP feature. This feature requires the Enterprise Services License.
- **router bgp** *autonomous-system-number*: This command enables BGP and assigns an AS number to the local BGP process.
- **router-id** *ip-address*: This command defines the BGP router ID. This command is optional but strongly recommended. If you do not configure the router ID using this command, then BGP may not be able to select a router ID if the switch does not have an active interface with an IPv4 address to use as the router ID.

https://t.me/learningnets

- **network** *ipv6-prefix*/*length* [**route-map** *name*]: This command specifies a network prefix as local to this AS and then adds it to the BGP table. This command is optional.

- **neighbor** *ip-address* **remote-as** *as-number*: This command configures the IP address and AS number for the remote BGP peer.

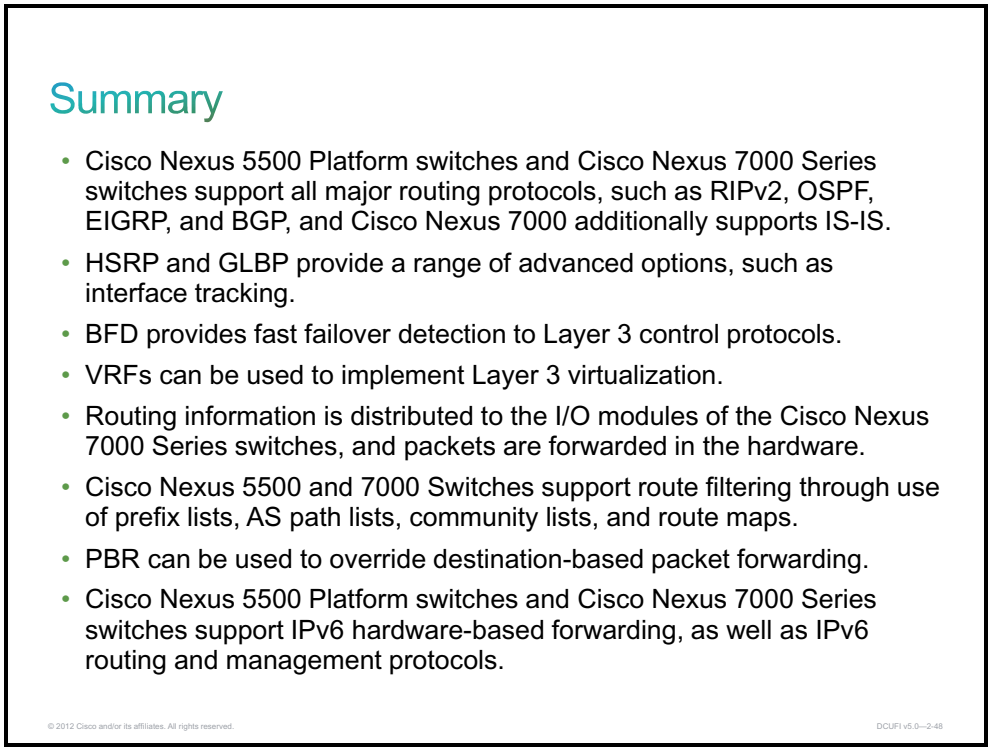
- **address-family {ipv4 | ipv6} {unicast | multicast}**: This command enables the address family for a neighbor and enters the address family configuration mode for the specified neighbor. In this mode, the specific options for the selected address family can be configured for that neighbor. At least one address family needs to be enabled for a neighbor to enable the peering.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Summary

This topic summarizes the key points that were discussed in this lesson.



References

For additional information, refer to these resources:

- To learn more about routing configuration on Cisco Nexus 7000 Switches refer to *Cisco Nexus 7000 Series NX-OS Unicast Routing Configuration Guide, Release 6.x* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/sw/6_x/nx-os/unicast/configuration/guide/l3_cli_nxos.html
- To learn more about routing configuration on Cisco Nexus 5500 Switches refer to *Cisco Nexus 5000 Series NX-OS Unicast Routing Configuration Guide, Release 5.0(3)N1(1)* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/nexus5000/sw/unicast/5_0_3_N1_1/Cisco_n5k_layer3_ucast_cfg_rel_503_N1_1.html

https://t.me/learningnets

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Lesson 7

Configuring IP Multicast

Overview

IP multicast is used as an efficient data-distribution mechanism in important enterprise applications. Common business applications using IP multicast are IP-based video applications, market data applications, desktop or application deployment systems, and other applications that need to deliver the same data at the same time to multiple receivers.

Components of these multicast-based applications may be hosted in the data center. Therefore, it is important that an enterprise data center is capable of supporting IP multicast.

The Cisco Nexus Operating System (NX-OS) Software that runs on the Cisco Nexus switches includes all of the major IP multicast protocols, such as the Internet Group Management Protocol (IGMP) that is used for multicast group management in IP version 4 (IPv4), Multicast Listener Discovery (MLD) for group management in IP version 6 (IPv6), Protocol Independent Multicast (PIM), PIM for IPv6 network (PIM6), and the Multicast Source Discovery Protocol (MSDP). In addition, the Cisco Nexus switches include key features to support IP multicast at Layer 2, such as IGMP and MLD snooping.

Objectives

Upon completing this lesson, you will be able to implement multicast functionality in a Cisco Data Center Network Architecture. You will be able to meet these objectives:

- Identify the components and architecture of IP multicasting
- Identify how to configure the IGMP and MLD on the Cisco Nexus switches
- Identify how to configure the PIM features on the Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches
- Identify how to configure the IGMP snooping on the Cisco Nexus switches
- Identify how to configure the MSDP on the Cisco Nexus 7000 Series switch

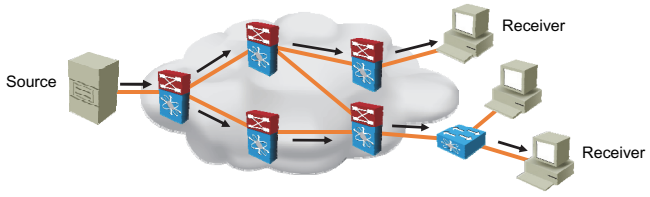
https://t.me/learningnets

IP Multicast

This topic identifies the components and architecture of IP multicasting.

IP Multicast

- IP multicast distributes data from multicast sources to a group of multicast receivers
- Sources are unaware of the client population
- Clients announce their interest in multicast groups to first-hop routers using IGMP (IPv4) or MLD (IPv6)
- Routers build a distribution tree to deliver the multicast data from the source to the receivers
- Switches can snoop IGMP/MLD messages to optimize Layer 2 forwarding of IP multicast traffic



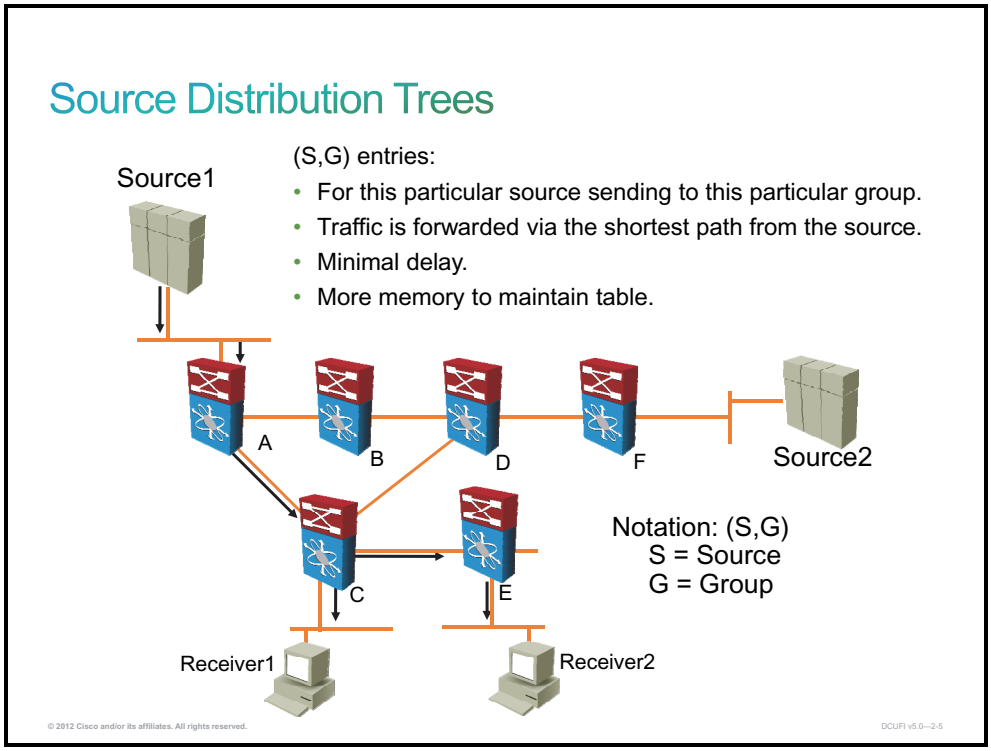
© 2012 Cisco and/or its affiliates. All rights reserved.

In multicast, the sender sends only one copy of a single data packet that is addressed to a group of receivers—a multicast group. Downstream multicast routers replicate and forward the data packet to all branches where receivers exist. Receivers show readiness for multicast traffic by registering at their first-hop routers by using IGMP for IPv4 multicast or MLD for IPv6 multicast.

The figure shows a multicast source host transmitting one copy of the data and a network then replicating the packet. Routers and switches are responsible for replicating the packet and forwarding it to multiple recipients. The network devices replicate the packet at any point where the network paths diverge and then use Reverse Path Forwarding (RPF) techniques to ensure that the packet is forwarded to the appropriate downstream paths without routing loops. Each packet exists as a single copy in any given network. The multicast source host may send to multiple receivers simultaneously because it is sending only one packet.

When the source becomes active, it begins sending the data without providing any indication. First-hop routers, to which the sources are directly connected, start forwarding the data to the network. Receivers that are interested in receiving IPv4 multicast data register to the last-hop routers by using IGMP membership messages. Last-hop routers are those routers that have directly connected receivers. Last-hop routers forward the group membership information of their receivers to the network. In this way, the other routers are informed as to which multicast flows are needed.

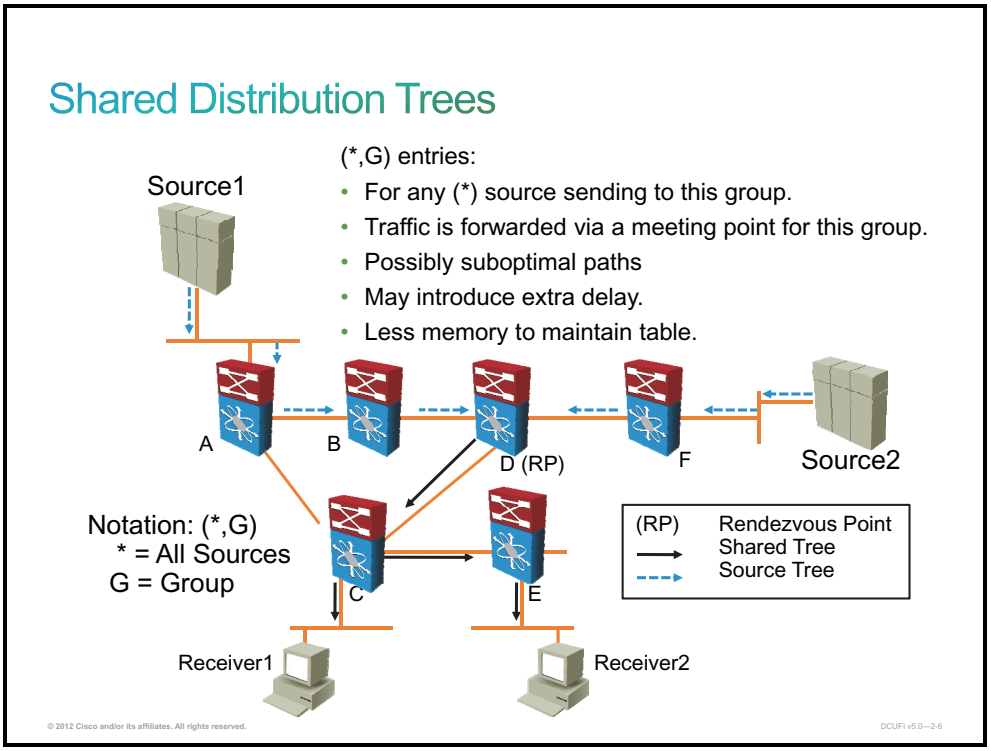
https://t.me/learningnets



The figure shows a common multicast distribution model called *source distribution tree*. In the figure, the source distribution tree builds a shortest path tree (SPT) between Source1 and Receiver1 and 2. The path between the source and receivers over routers A, C, and E is the path with the lowest cost. Packets are forwarded down the SPT according to the pairs of source and group addresses. The forwarding state is referred to by the notation (S,G) (pronounced *S comma G*). In this notation, S is the IP address of the source, and G is the multicast group address. A separate SPT is built for every source S sending to group G.

The multicast forwarding entries that describe a source distribution tree in the multicast forwarding tables use the (S,G) notation. This notation indicates that a source S is sending to the group G. SPT (S,G) state entries use a great deal of router memory because there is an entry for each sender and group pair. Because the traffic is sent over the optimal path to each receiver, the delay in packet delivery is minimized.

https://t.me/learningnets



In the shared distribution tree, which is shown in the figure, switch D is the root. The tree is built from switch D to switches C and E toward Receiver1 and Receiver2. In PIM, the root of the shared tree is called a rendezvous point (RP). Packets are forwarded down the shared distribution tree to the receivers. The notation (*,G) (pronounced *star comma G*) identifies the default forwarding state for the shared tree. The * symbol represents a wildcard entry, meaning that any source can be plugged into the notation, and G is the multicast group address.

The figure also shows traffic flow on two source-rooted trees in addition to the shared distribution tree. Source1 and Source2 are sending multicast packets toward an RP via the source-rooted trees. From the RP, the multicast packets are flowing via a shared distribution tree toward Receiver1 and Receiver2.

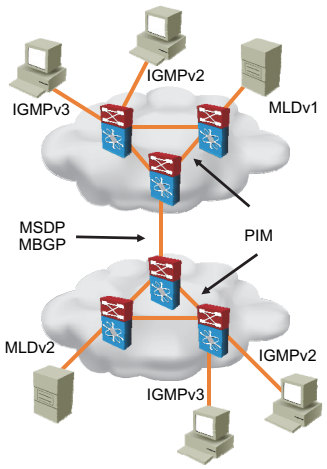
The multicast forwarding entries that describe a shared distribution tree in the multicast forwarding table use the (*,G) notation. This notation indicates that any source (*) is sending to the group G. These entries reflect the shared tree, but they are also created for any existing (S,G) entry. (*,G) state entries for shared distribution trees consume less router memory, but you may get suboptimal paths from a source to the receivers, thereby introducing an extra delay in packet delivery.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Multicast Control Protocols

- Group membership
  - IGMPv2/MLDv1
  - IGMPv3/MLDv2
- Multicast intradomain routing
  - PIM Sparse Mode
  - PIM Bidir
  - PIM SSM
- Multicast interdomain routing
  - MSDP
  - MBGP (Cisco Nexus 7000 Switch)



© 2012 Cisco and/or its affiliates. All rights reserved.

Multicast control protocols build and maintain the multicast distribution trees that are necessary to forward multicast data from the sources to the receivers.

A group membership protocol is necessary for receivers to announce their readiness to receive multicast traffic for a particular group. A Layer 3 switch running IGMP uses periodic querying on a segment in order to track the receivers and the multicast groups that are of interest. Receivers can send explicit join messages to signal their interest in a group or send explicit leave messages to announce that they are no longer interested in a group. There are two versions of IGMP and MLD that are commonly used for IP multicast:

- **IGMPv2** is the most common form of IGMP, used in Any Source Multicast (ASM) deployments. With IGMPv2, a host can signal its interest in a particular multicast group, but it cannot specify which source to receive it from. Traffic from any source that is sent to the specified group address will be delivered to the host.
- **MLDv1** is the equivalent of IGMPv2 in IPv6.
- **IGMPv3** is used in Source Specific Multicast (SSM) deployments. IGMPv3 allows a receiver to not only specify the group but also a list of sources that it is interested in. This ability requires the application on the client to know both the source IP address and the multicast group that is associated with a specific multicast stream.
- **MLDv2** is the equivalent of IGMPv3 in IPv6.

https://t.me/learningnets

Once the clients have signaled their interest in specific multicast groups, the routers are then responsible for building a distribution tree to forward the data from the sources on to the receivers. The protocol that is commonly used for this is PIM version 2 (PIMv2). PIM can be deployed in a number of different modes:

- **PIM sparse mode**: PIM sparse mode is commonly used in ASM deployments. PIM sparse mode uses an RP as a central point where routers can register multicast sources that are present on a directly connected subnet. When a receiver signals its interest in a specific multicast group using IGMP, the leaf router that is connected to the receiver builds a shared tree to the RP, which will then build a source-based tree to the source and connect the two trees so as to create a forwarding path to the receiver. Once the leaf router starts receiving the data, it may opt to build a source-based tree back to the source in order to optimize the flow of data.
- **BIDIR-PIM**: Bidirectional Protocol Independent Multicast (BIDIR-PIM) is typically deployed for many-to-many multicast applications. This type of application requires many multicast routes to be maintained if source-based trees are used. BIDIR-PIM only uses a shared tree that is rooted at the RP. BIDIR-PIM allows traffic to flow both up the tree from the sources to the RP as well as down the tree from the RP to the receivers.
- **PIM-SSM**: Protocol Independent Multicast Source Specific Mode (PIM-SSM) is typically used in SSM multicast deployments. While BIDIR-PIM only uses a shared tree, PIM-SSM is the exact opposite in that it uses only source-based trees that are built directly from the leaf router that is attached to the receiver back to the source. This mode requires the receiver to signal the source that it is interested in using IGMPv3.

PIM is typically used for multicast routing within a single routing domain or autonomous system (AS). To support multicast routing between two separate routing domains, two protocols can be used—the MSDP and Multicast Border Gateway Protocol (MBGP).

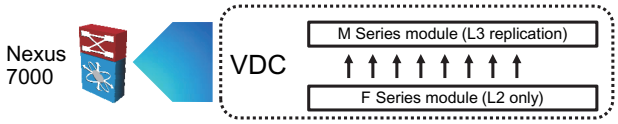
- **MSDP**: MSDP is used to announce the availability of multicast sources between RPs in the different domains. First-hop routers register sources through PIM with the RP within their domain. MSDP can then be used to advertise that information to RPs in different routing domains.
- **MBGP**: Multicast routing depends on unicast routing information for RPF, which is used for loop prevention and to build trees toward a source or RP. MBGP exchanges IP prefix information to be used for multicast RPF between two autonomous systems.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Multicast on Cisco Nexus Switches

- Multicast supported on Cisco Nexus 5500 and 7000
- License requirements for PIM, PIM6, and MSDP:
  - Nexus 5500: Layer 3 Base License
  - Nexus 7000: Enterprise Services License
- Hardware considerations on Cisco 7000:
  - F Series module is Layer 2-only
  - F Series requires M Series module in the same VDC
  - Packets entering through F Series module automatically forwarded to interface on M Series module
  - Interfaces on the M Series module perform egress replication for Layer 3 multicast packets
- vPC supports only Any Source Multicast (ASM) PIM, not Bidir or SSM



© 2012 Cisco and/or its affiliates. All rights reserved.

Multicasting is supported on Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches. You must license the switches to enable support for multicast routing protocols PIM, PIM6, and MSDP. The requirements are the Layer 3 Base license on the Cisco Nexus 5500 Platform switches and the LAN Enterprise Package on the Cisco Nexus 7000 Series switches.

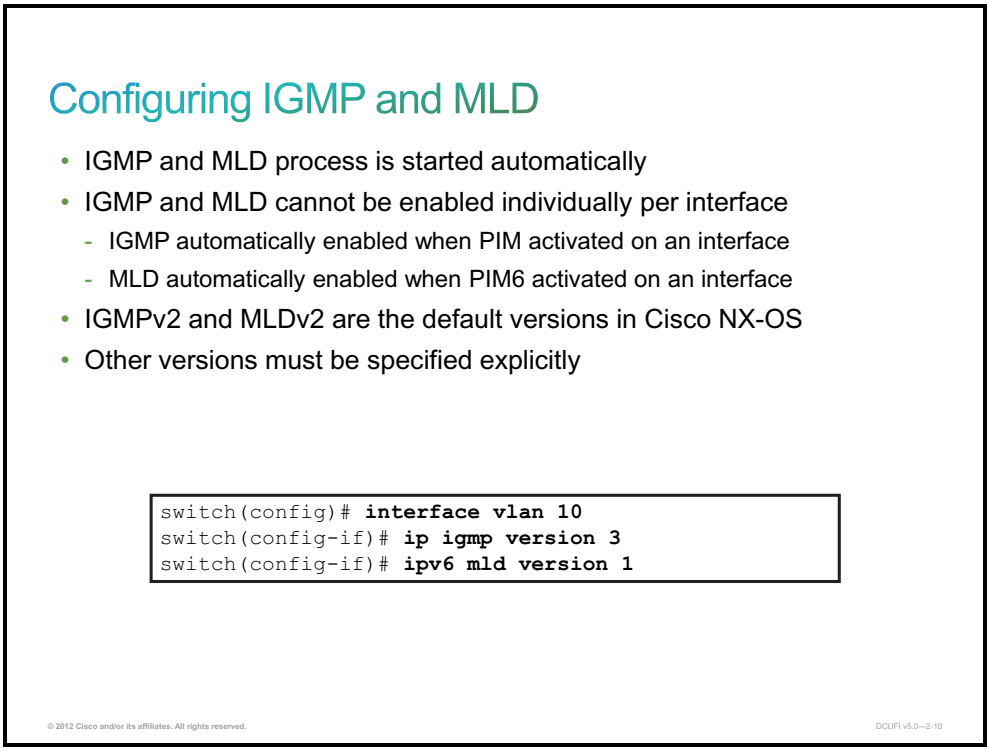
Beginning with Cisco NX-OS Release 5.1, you can add an F-Series module, which is a Layer 2-only module, into the Cisco Nexus 7000 Series chassis. When you add this module to a chassis that already contains M Series modules, you can then provision multicasting. You can position a chassis with both F- and M Series modules at the boundary between the Layer 2 and Layer 3 networks. Packets that enter an interface on the F-Series module are automatically forwarded to one of the interfaces on the M Series modules in the same virtual device context (VDC) to be routed. The interface on the M Series module performs egress replication for Layer 3 multicast packets that enter an interface on the F-Series module in the same VDC.

A virtual port channel (vPC) allows a single device to use a port channel across two upstream switches. Cisco NX-OS Software does not support PIM-SSM or BIDIR-PIM on a vPC. However, Cisco NX-OS Software fully supports PIM ASM on a vPC.

https://t.me/learningnets

Configuring IGMP and MLD

This topic identifies how to configure IGMP and MLD on the Cisco Nexus switches.



Receivers send IGMP or MLD messages to their adjacent Layer 3 device to indicate that they are interested in joining a group. Layer 3 devices listen to IGMP and MLD messages and periodically send out queries in order to discover which groups are active or inactive on a particular subnet. Members joining a group do not have to wait for a query to join; they can send an unsolicited report indicating their interest. Unsolicited queries are commonly referred to as "joins." The query and report mechanism is supported in all versions of IGMP and MLD.

IGMPv2 and its IPv6 equivalent, MLDv1, allow hosts to actively communicate to the local multicast router that they intend to leave the group by sending a leave group message.

IGMPv3 and its IPv6 equivalent, MLDv2, provide a further step in the evolution of multicast group management. They add support for "source filtering," which enables a multicast receiver host to signal to a router the groups from which it wants to receive multicast traffic and from which sources this traffic is expected. This membership information enables a router to then forward traffic from only those sources from which receivers requested the traffic.

There is no specific command to enable IGMP or MLD on an interface on a Cisco Nexus switch. IGMP is automatically enabled when PIM is enabled on an interface, and MLD is automatically enabled when PIM6 is enabled on an interface.

The default IGMP and MLD versions that are used by Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches are IGMPv2 and MLDv2. If IGMPv3 is required to support SSM, it must be specifically configured on every interface where it is required. Similarly, MLDv1 has to be specified in certain legacy environments.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Verifying IGMP

1. Verify interfaces enabled for IGMP and check multicast receivers
2. Verify receivers of specific groups
3. Verify receivers of specific groups and sources

```
switch# show ip igmp interface brief                                    1
IGMP Interfaces for VRF "default", count: 3
Interface            IP Address      IGMP Querier    Membership  Version
                                                     Count
Vlan10               172.16.10.75    172.16.10.75    0           v2
Vlan11               172.16.11.75    172.16.11.75    1           v3
loopback1            192.168.1.75    192.168.1.75    0           v2

switch# show ip igmp groups 239.1.1.1
IGMP Connected Group Membership for VRF "default" - matching
  Group "239.1.1.1"                                                     2
Type: S - Static, D - Dynamic, L - Local, T - SSM Translated
Group Address      Type Interface           Uptime    Expires   Last Reporter
239.1.1.1          D    Vlan11              00:00:22  00:04:02  172.16.11.76

switch# show ip igmp groups 239.1.1.1 192.168.1.1
IGMP Connected Group Membership for VRF "default" - matching
  Group "239.1.1.1" and Source "192.168.1.1"                            3
Type: S - Static, D - Dynamic, L - Local, T - SSM Translated
Group Address      Type Interface           Uptime    Expires   Last Reporter
239.1.1.1
  192.168.1.1      D    Vlan11              00:00:19  00:04:03  172.16.11.76
```

Use **show ipv6 mld interface** in IPv6 environment

Use **show ipv6 mld groups** in IPv6 environment

© 2012 Cisco and/or its affiliates. All rights reserved.

You can use several commands to verify that IGMP or MLD is enabled on all of the necessary interfaces.

The **show ip igmp interface brief** command can be used to verify the interfaces that are activated for IGMP, which version of IGMP is used, and the number of groups that are joined on that interface.

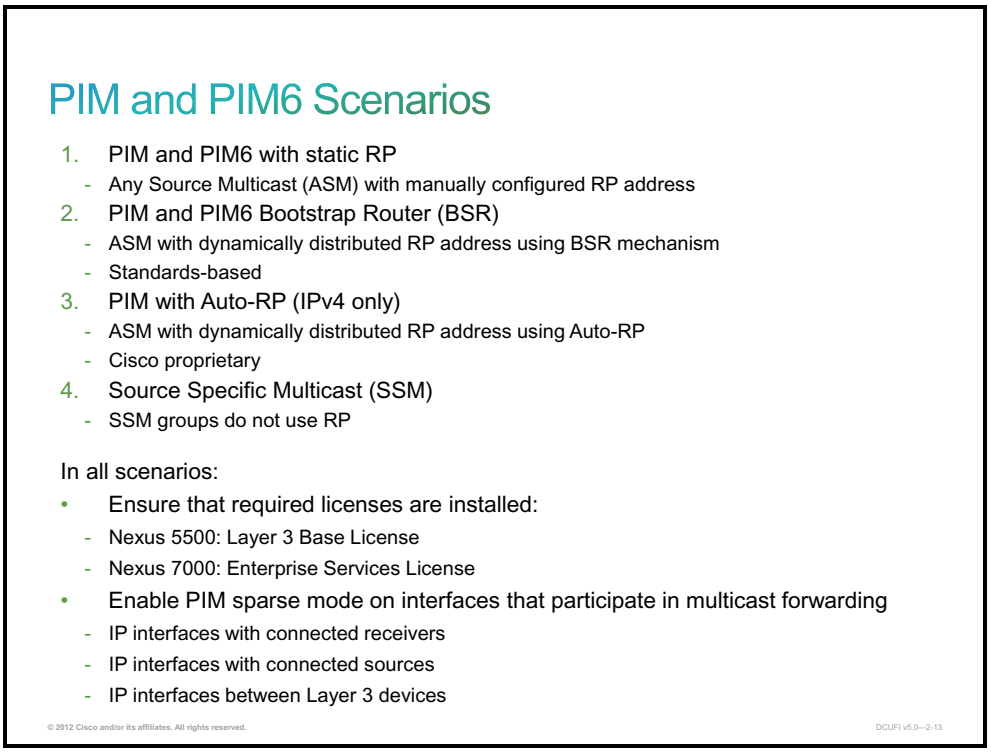
To verify whether the receiver has properly joined the multicast group through IGMP, you can use the **show ip igmp groups** command for the group that you are interested in. For IGMPv3, you can also specify the source to make sure that the client joined the correct source and group combination.

Although not shown in the figure, you can use similar commands, such as **show ipv6 mld interface** and **show ipv6 mld groups**, to check the MLD receivers.

https://t.me/learningnets

Configuring PIM

This topic identifies how to configure the PIM features on the Cisco Nexus 5500 Platform switches and Cisco Nexus 7000 Series switches.



This section presents the PIM and PIM6 configuration in four main scenarios:

1. **PIM and PIM6 with static RP**: This is an ASM deployment with a manually configured RP address on all Layer 3 devices except for the RP.

2. **PIM and PIM6 bootstrap router (BSR)**: This is an ASM scenario with a dynamically distributed RP address using the BSR mechanism. BSR is a standards-based mechanism that is offered by PIMv2 (IPv4) and PIM6 (IPv6).

3. **PIM with Auto-RP**: This is an ASM deployment with a dynamically distributed RP address that uses Auto-Rendezvous Point (Auto-RP). Auto-RP is a Cisco proprietary method that is available for IPv4 only.

4. **Source Specific Multicast (SSM)**: SSM groups do not use RP, so there is no need to configure or distribute RP information. One network can have a mix of groups that operate in SSM and ASM modes. The multicast groups using the ASM mode need RP information in order to work properly.

In all scenarios, you must take the following into consideration:

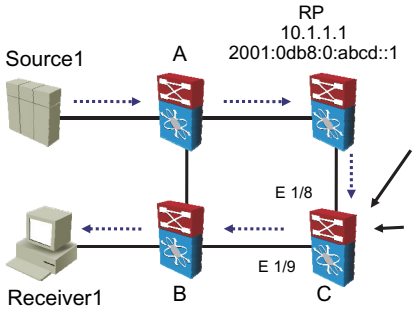
- Ensure that required licenses are installed. The Cisco Nexus 5500 Platform switches need the Layer 3 Base license; the Cisco Nexus 7000 Series switches require the Enterprise Services License.
- Enable PIM sparse mode on interfaces that participate in multicast forwarding. These interfaces include the Layer 3 interfaces with connected receivers, with connected sources, and interfaces between Layer 3 devices.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

PIM with Static RP

1. Enable PIM/PIM6 feature
2. Enable PIM on interfaces
3. Configure RP address (except on RP)
   - Could be configured in VRF context




```
C(config)# feature pim                                    1

C(config)# interface vlan 10
C(config-if)# ip pim sparse-mode                          2
C(config)# interface ethernet 1/8-9
C(config-if-range)# ip pim sparse-mode

C(config)# ip pim rp-address 10.1.1.1                     3
group-list 224.0.0.0/9
```

```
C(config)# feature pim6                                   1

C(config)# interface ethernet 1/8-9                       2
C(config-if-range)# ipv6 pim sparse-mode

C(config)# ipv6 pim rp-address
2001:0db8:0:abcd::1 group-list                            3
ff1e:abcd:def1::0/24
```

© 2012 Cisco and/or its affiliates. All rights reserved.

The figure shows an example of a basic PIM sparse mode configuration on a Cisco Nexus 5500 Platform switch and a Cisco Nexus 7000 Series switch. The following commands are used:

- **feature pim**: This command enables the PIM feature if the appropriate license is installed.
- **ip pim rp-address** *rp-address* [**group-list** *prefix]*: This command defines the RP address for a set of groups that are indicated by the multicast group prefix. If no prefix is specified, this command applies to all IP multicast groups, as indicated by the prefix 224.0.0.0/4.
- **ip pim sparse-mode**: This command enables PIM sparse mode on an interface. This command implicitly also enables IP multicast processing and IGMP for that particular interface.

Optionally, you may use the **ip pim log-neighbor-changes** command to generate syslog messages that list the PIM neighbor state changes. This command is optional and is not enabled by default.

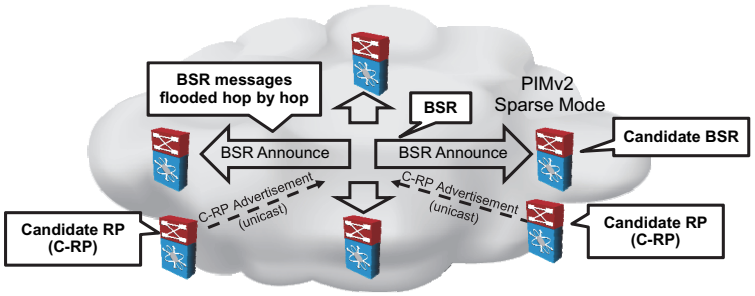
The second set of commands depicts the equivalent configuration for IPv6.

PIM can also be configured for virtual routing and forwarding (VRF) instances. The configuration is similar to the configuration for the default VRF except that global PIM parameters, such as the PIM RP configuration, are performed in VRF context mode for the specific VRF.

https://t.me/learningnets

PIM Bootstrap Router (BSR)

- A single BSR elected out of multiple candidate BSRs
- Candidate RPs send candidacy announcements to the BSR:
  - Unicast transport
  - BSR stores all candidate-RP announcements in the RP set
- BSR periodically sends BSR messages to all routers:
  - With the entire RP set and IP address of the BSR
  - Flooded hop by hop throughout the network
- All routers select the RP from the RP set



© 2012 Cisco and/or its affiliates. All rights reserved.

The bootstrap router (BSR) is an IETF standards mechanism available in PIMv2, both for IPv4 and IPv6. A single router is elected as the BSR from a collection of candidate BSRs. If the current BSR fails, a new election is then triggered. The election mechanism is preemptive based on candidate BSR priority.

Via unicast, candidate RPs send candidate-RP announcements directly to the BSR. Candidate RPs learn the IP address of the BSR via periodic BSR messages.

The BSR does not elect the best RP for every group range that it discovers. For every group range known, the BSR builds a set of candidate RPs, including all of the routers that advertised their willingness to service as the RP for the group range. The BSR stores the collection of candidate RP announcements in a database that is called the RP set. The BSR periodically sends out BSR messages to the routers in the network to let them know that the BSR is still active. BSR messages are flooded hop by hop throughout the network as multicasts to the all-PIM-routers group (224.0.0.13 in IPv4) with a TTL of 1. When a router receives a BSR message, it applies RPF check, which is based on the source IP address in the packet. If the RPF check succeeds, the message is then flooded out to all PIM-enabled interfaces.

BSR messages contain these elements:

- The RP set consisting of candidate RP announcements
- The IP address of the BSR so that candidate RPs know where to send their announcements

Because the packets are flooded throughout the network, the routers will receive the BSR messages. Each receiving router selects the active RP for each group range by using a common hash algorithm that is run against the RP set. The routers in the network will select the same RP for a given group range.

Candidate RPs periodically send candidate RP messages directly to the BSR via unicast. The candidate RPs know the BSR address because the BSR messages have been periodically flooded to the all-PIM-routers group (224.0.0.13 in IPv4). The default interval for the candidate RP messages is 60 seconds.

Candidate BSRs participate in the BSR-election mechanism using these rules:

https://t.me/learningnets

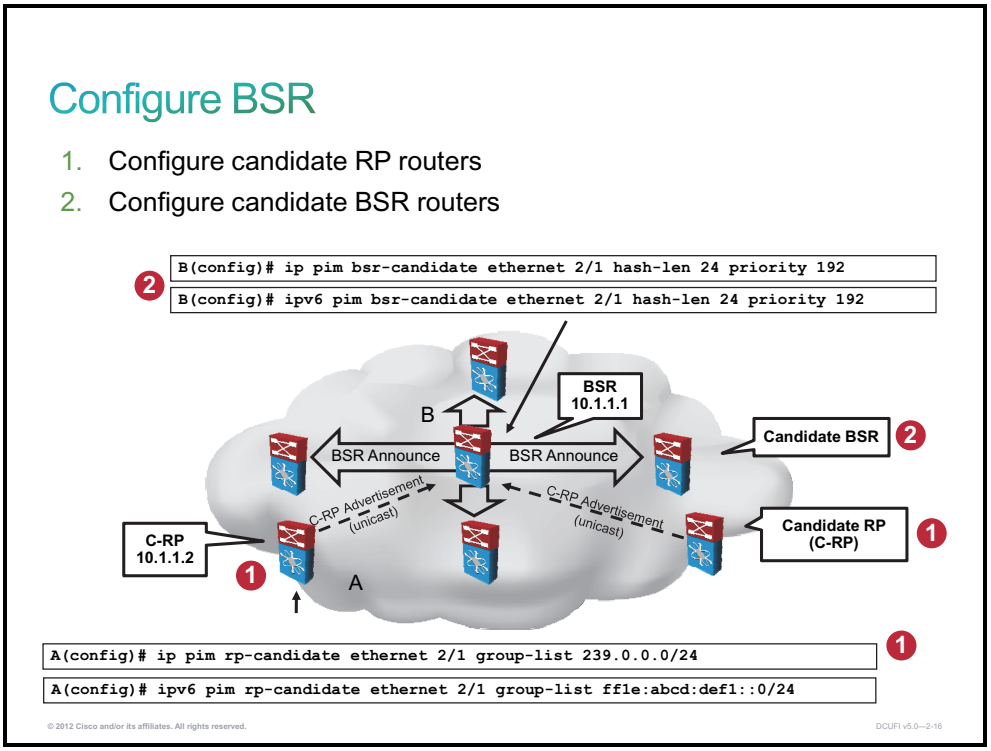
- The candidate BSR with the highest priority is elected as the BSR.
- The highest IP address of the candidate BSRs is used as a tiebreaker.
- The election mechanism is pre-emptive, meaning that if a new candidate BSR with a higher priority comes up, it then triggers a new election process.

All PIMv2 routers follow these specifications:

- PIMv2 routers accept BSR messages that are based on the rules that are described earlier in this lesson. When a BSR message is accepted:
  - The RP set in the BSR message is stored in the local group-to-RP mapping cache.
  - The BSR message is forwarded out the other interfaces, except for the one in which it was received.
- PIMv2 routers select an RP using a hash algorithm:
  - The RP for a group is selected from the set of candidate RPs that have advertised their candidacy for a matching group range.
  - The routers use the same hashing algorithm to select the RP from the set of candidate RPs in the RP set. Since the routers run the same algorithm on the same RP set, they will select the same RP for a given group.

The hashing algorithm permits multiple candidate RPs to load balance the duties of the RP across a range of groups. Only one candidate RP will be selected as the RP for any single group within the group range. However, the hash algorithm may select other candidate RPs as the RP for another group within the group range.

https://t.me/learningnets



Candidate RPs are configured by using the **ip pim rp-candidate** command for IPv4 and the **ipv6 pim rp-candidate** command for IPv6. Candidate BSRs are configured by using the **ip pim bsr-candidate** command and the **ipv6 pim bsr-candidate** command, respectively.

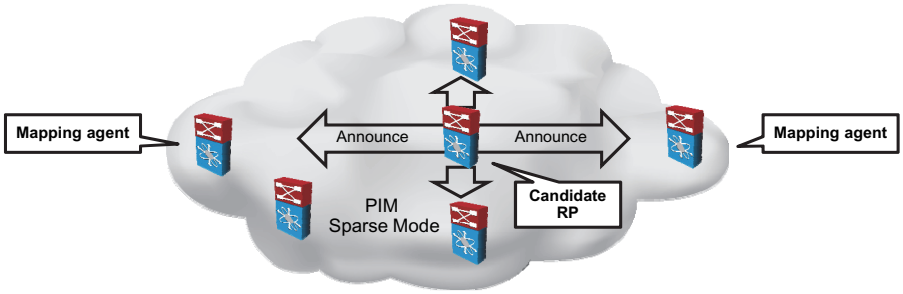
The **ip pim rp-candidate** command and its IPv6 equivalent cause the switch to send bootstrap messages to all its PIM neighbors, with the address of the designated interface as the BSR address. Each neighbor then compares the BSR address with the address it received from previous bootstrap messages (though not necessarily received on the same interface). If the current address is the same or a higher address, the PIM neighbor caches the current address and forwards the bootstrap message. Otherwise, the bootstrap message is dropped.

The **ip pim bsr-candidate** command and its IPv6 equivalent cause the switch to send a PIMv2 message advertising itself as a candidate rendezvous point to the BSR. The addresses that are allowed by the access list, together with the router identified by the IP address, constitute the RP and the range of addresses for which it is responsible.

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Auto-RP

- IPv4-only mechanism
- Candidate RPs advertise themselves using announcements
- Mapping agents act as relays:
  - Receive RP announcements
  - Store the Group-to-RP mapping in cache
  - Elect the highest C-RP address as RP for group range
  - Advertise RP-Discovery messages
- All Cisco routers join discovery group and receive discovery messages



© 2012 Cisco and/or its affiliates. All rights reserved.

Auto-RP allows all routers to automatically learn group-to-RP mappings. The Cisco proprietary mechanism is only available for IPv4 environments. There are special Auto-RP roles:

- Candidate RPs (C-RPs)
- Mapping agents

Multicast is used to distribute group-to-RP mapping information via two special multicast groups that are assigned with Internet Assigned Numbers Authority (IANA) addresses:

- Cisco announce group: 224.0.1.39.
- Cisco discovery group: 224.0.1.40.

Because multicast is used to distribute this information, the 224.0.1.39 and 224.0.1.40 groups are run in PIM dense mode so that this information is flooded throughout the network.

Multiple candidate RPs may be defined so that in the case of an RP failure, the other candidate RP can assume the responsibility of the RP. Auto-RP can be configured to support administratively scoped zones (unlike the BSR mechanism, which is explained later). Administratively scoped zones can be important when trying to prevent high-rate group traffic from leaving a campus and consuming too much bandwidth on the WAN links.

An Auto-RP candidate RP is sending RP-announcement messages to the Cisco announce group (224.0.1.39). These messages announce the router as being a candidate RP. By default, the messages are sent every 60 seconds. RP-announce messages contain:

- The group address range (the default is the all-multicast-groups address 224.0.0.0/4).
- The IP address of the candidate RP.
- A hold time, used to detect when the candidate RP has failed.

https://t.me/learningnets

An Auto-RP mapping agent joins the RP-announce group (224.0.1.39) in order to receive RP announcements from a C-RP. When a mapping agent receives an announcement:

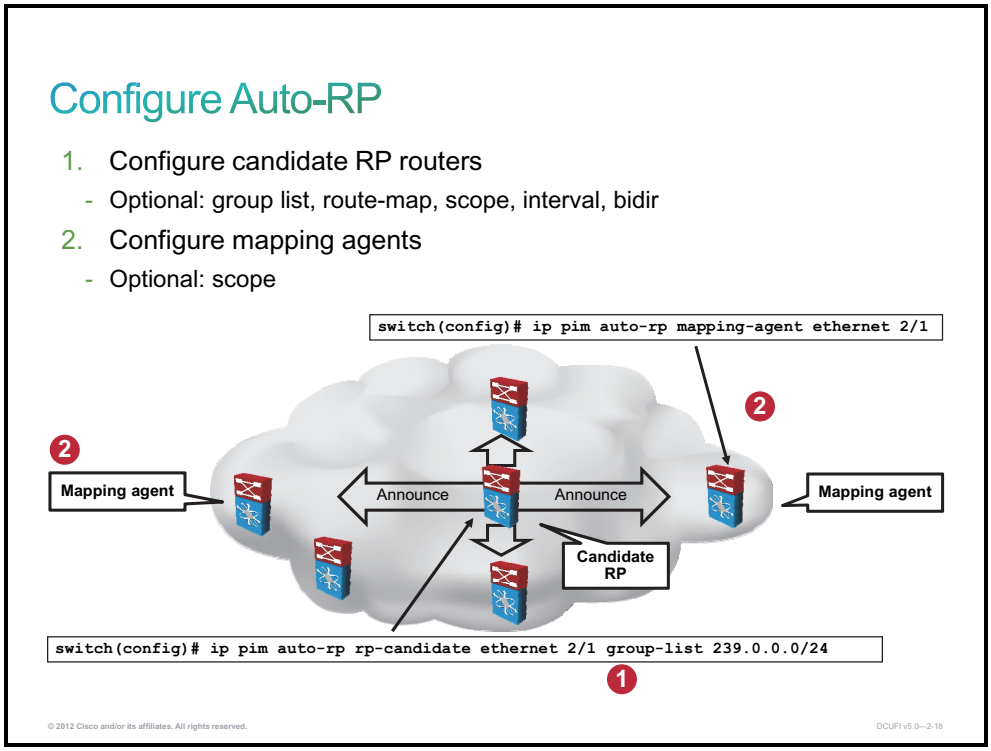
- It saves the announcement in the group-to-RP mapping cache.
- It selects the C-RP with the highest IP address as the RP for the group range.

Cisco routers automatically join the Cisco discovery group (224.0.1.40) in order to receive the group-to-RP mapping information being multicast by the mapping agent in the network. No configuration is required to join this group.

Group-to-RP mapping information that is contained in the RP-Discovery messages is stored in the local group-to-RP mapping cache of the router. The router then uses this information to map a group address to the IP address of the active RP for the group.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



To implement the Auto-RP mechanism, you must configure the following:

- **One or more candidate RPs**: This configuration is done with the **ip pim auto-rp rp-candidate** command, as shown in the figure.
- **One or more mapping agents**: This configuration is done with the **ip pim auto-rp mapping-agent** command, as shown in the figure.

Additionally, you may tune some Auto-RP options:

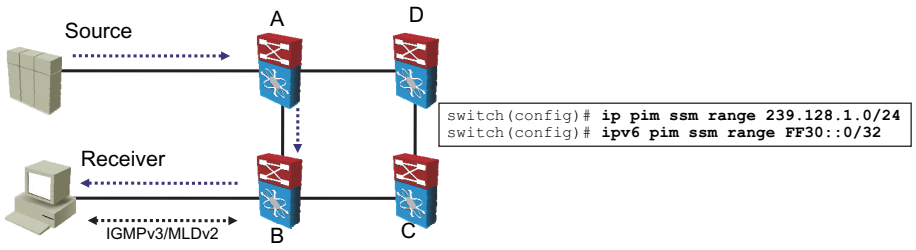
- Advertisement filtering
- Failover
- RP fallback
- Scope constraining

Option tuning is not explained in this course.

https://t.me/learningnets

Source Specific Multicast (SSM)

- Solution for well-known sources
- Also referred to as Single Source Multicast
- Immediate shortest path from source to receivers
  - Last-hop router sends (S,G) join directly to source
  - First-hop router responds to receiver-initiated join requests
  - No shared tree
- Typically combined with IGMPv3 / MLDv2



© 2012 Cisco and/or its affiliates. All rights reserved.

Source Specific Multicast (SSM) uses all of the benefits of sparse mode protocols but eliminates shared trees and only builds source-specific SPTs. These trees are built directly based upon the receipt of group membership reports that request a given source.

SSM is suitable for use when there are well-known sources. A dedicated multicast group address range of 232.0.0.0/8 is used exclusively for SPTs for SSM. Routers are prevented from building a shared tree for any of the groups from this address range. The address range 232.0.0.0/8 is assigned for global well-known sources.

SSM allows the last-hop router to immediately send an (S,G) join toward the source. Thus, the PIM spare mode (PIM-SM) (*,G) join toward the RP is eliminated, and the first-hop routers start forwarding the multicast traffic on the SPT from the very beginning. The SPT is built by receiving the first (S,G) join. Source-specific groups may coexist with other groups in PIM-SM domains.

The prerequisite for SSM deployment is a mechanism that allows hosts to report not only the group that they want to join but also the source for the group. This mechanism is built into the IGMPv3 and MLDv2 protocols. With IGMPv3 and LDPv2, last-hop routers may receive membership reports requesting a specific multicast source and group traffic flow. The router responds by simply creating an (S,G) state and triggering an (S,G) join toward the source.

Exactly how a host learns about the existence of sources may occur via directory service, session announcements directly from sources, or some out-of-band (OOB) mechanisms (for example, web pages).

You implement SSM by configuring the appropriate address range that is used by SSM multicast groups. This configuration is done with the **ip pim ssm range** and **ipv6 im ssm range** commands.

https://t.me/learningnets

Verify PIM

1. Verify PIM interfaces
2. Examine PIM neighbors
3. View RP information

```
switch# show ip pim interface brief                                    1
PIM Interface Status for VRF "default"
Interface            IP Address      PIM DR Address  Neighbor  Border
                                                     Count     Interface
Vlan10               172.16.10.75    172.16.10.76    1         no
Vlan11               172.16.11.75    172.16.11.75    0         no
loopback1            192.168.1.75    192.168.1.75    0         no


switch# show ip pim neighbor                                           2
PIM Neighbor Status for VRF "default"
Neighbor        Interface            Uptime    Expires   DR       Bidir-  BFD
                                                         Priority Capable State
172.16.10.76    Vlan10               04:30:38  00:01:18  1        yes     n/a

switch# show ip pim rp
PIM RP Status Information for VRF "default"                            3
<output omitted>

RP: 21.21.0.11, (0), uptime: 00:12:36, expires: never,
  priority: 0, RP-source: (local), group-map: rmap11, group ranges:
      231.0.0.0/8  231.128.0.0/9 (deny)
```

© 2012 Cisco and/or its affiliates. All rights reserved.

You can use multiple commands to verify PIM operation.

The **show ip pim interface brief** and **show ipv6 pim interface brief** commands can be used to verify the interfaces that are activated for PIM and PIM6, whether any neighbors are present on that interface, and which router is the PIM or PIM6 designated router (DR) for the segment.

To verify that the routers have established PIM or PIM6 adjacencies on all appropriate interfaces, you can use the **show ip pim neighbors** or **show ipv6 pim neighbors** commands. For correct multicast forwarding between the source and receiver, it is necessary that a contiguous chain of PIM neighbors exists on the RPF path between the receiver and the source.

The **show ip pim rp** and **show ipv6 pim rp** commands display information about the RPs, such as their distribution method and the served multicast groups.

https://t.me/learningnets

Verify Group Ranges and Multicast Routing Table

1. Verify multicast types used by group ranges
2. Examine multicast routing table

```
switch# show ip pim group-range
PIM Group-Range Configuration for VRF "default"                    1
Group-range        Mode      RP-address       Shared-tree-only range
232.0.0.0/8        SSM       -                -
231.0.0.0/8        ASM       21.21.0.11       -
231.128.0.0/9      ASM       21.21.0.22       -
231.129.128.0/17   Unknown   -                -

switch# show ip mroute 239.1.1.1
IP Multicast Routing Table for VRF "default"

(*, 239.1.1.1/32), uptime: 00:04:26, pim ip                       2
  Incoming interface: loopback1, RPF nbr: 192.168.1.75
  Outgoing interface list: (count: 1)
    Vlan10, uptime: 00:04:26, pim

(172.16.11.76/32, 239.1.1.1/32), uptime: 00:05:57, ip pim mrib
  Incoming interface: Vlan11, RPF nbr: 172.16.11.76, internal
  Outgoing interface list: (count: 1)
    Vlan10, uptime: 00:02:10, mrib, pim
```

Any Source Muticast (ASM) is the regular sparse mode using RPs

(*, G) entry describes forwarding of traffic from any source to group G

(S, G) entry describes forwarding of traffic from source S to group G

© 2012 Cisco and/or its affiliates. All rights reserved.

You can examine the mode of various multicast groups. The **show ip pim group-range** command displays the multicast group ranges and their mode: SSM, ASM, or unknown. The SSM groups do not use an RP. The ASM groups have an RP listed in the command output. The equivalent IPv6 command is **show ipv6 pim group-range.**

You can view the multicast routing table by using the **show ip mroute** and **show ipv6 mroute** commands. You can examine individual entries by specifying the multicast groups, as shown in the figure. The entries can describe a group (*,G) or a source-group pair (S,G). The entries include an incoming interface list (IIL) and outgoing interface list (OIL). The IIL and OIL define from which and to which interfaces the multicast streams are forwarded.

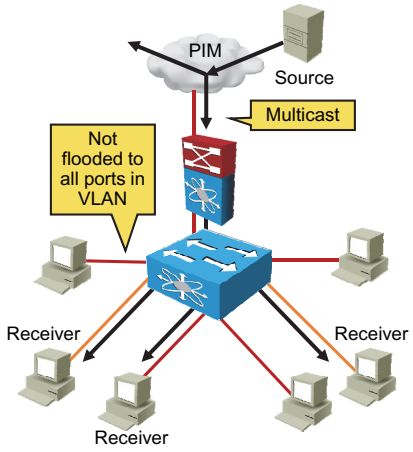
https://t.me/learningnets

Configuring IGMP Snooping

This topic identifies how to configure the IGMP snooping on the Cisco Nexus switches.

IGMP Snooping

- Switch examines content of IGMP messages to determine which ports need to receive multicast traffic for a group:
  - Examines IGMP membership reports to determine interested receivers
  - Examines IGMP leave messages to remove receivers
- Switch forwards multicast traffic more efficiently at Layer 2
- Does not require Layer 3 multicast routing on the switch
  - Supported also on Nexus 5000



© 2012 Cisco and/or its affiliates. All rights reserved.

IGMP snooping makes Layer 2 switches IGMP-aware. IGMP snooping constrains IPv4 multicast traffic at Layer 2 by configuring the Layer 2 LAN ports dynamically in order to forward IPv4 multicast traffic only to those ports that need to receive it. IGMP snooping requires the LAN switch to examine—or "snoop"—network layer information such as IGMP join and leave messages in the IGMP packets that are sent between the hosts and a router or multilayer switch.

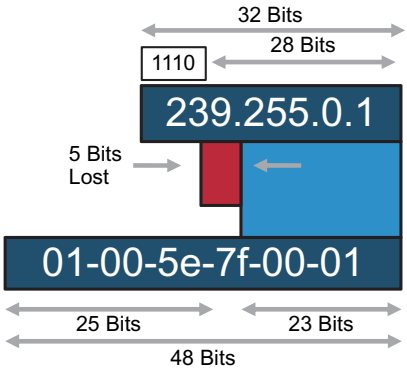
When the switch sees an IGMP host report from a host for a particular multicast group, the switch then adds the port number of the host to the associated multicast forwarding table entry. When the switch receives the IGMP leave group message from a host, the switch removes the table entry for the host. IGMP snooping also locates the multicast routers on a VLAN through the IGMP queries and PIM hellos that are sent by the routers. This is important because multicast routers should always receive multicast traffic for all groups. IGMP snooping is useful on any VLAN that contains multicast receivers and allows a Layer 2 switch to more efficiently forward IP multicast traffic on those VLANs.

IGMP snooping is supported on all Cisco Nexus switches (Cisco Nexus 5000 and 7000 Series and Cisco Nexus 5500 Platform switches). It does not require that Layer 3 multicast routing on the same platform, but it can coexist with it.

https://t.me/learningnets

IP Multicast MAC Address Mapping

- Without IGMP snooping
  - Switch floods multicast packets to all ports in VLAN
  - Non-receivers discard packets
- With IGMP snooping
  - Switch learns about connected receivers of a multicast group
  - Forwards corresponding multicast frames only to receivers
  - Forwarding decision based on destination MAC address



© 2012 Cisco and/or its affiliates. All rights reserved.

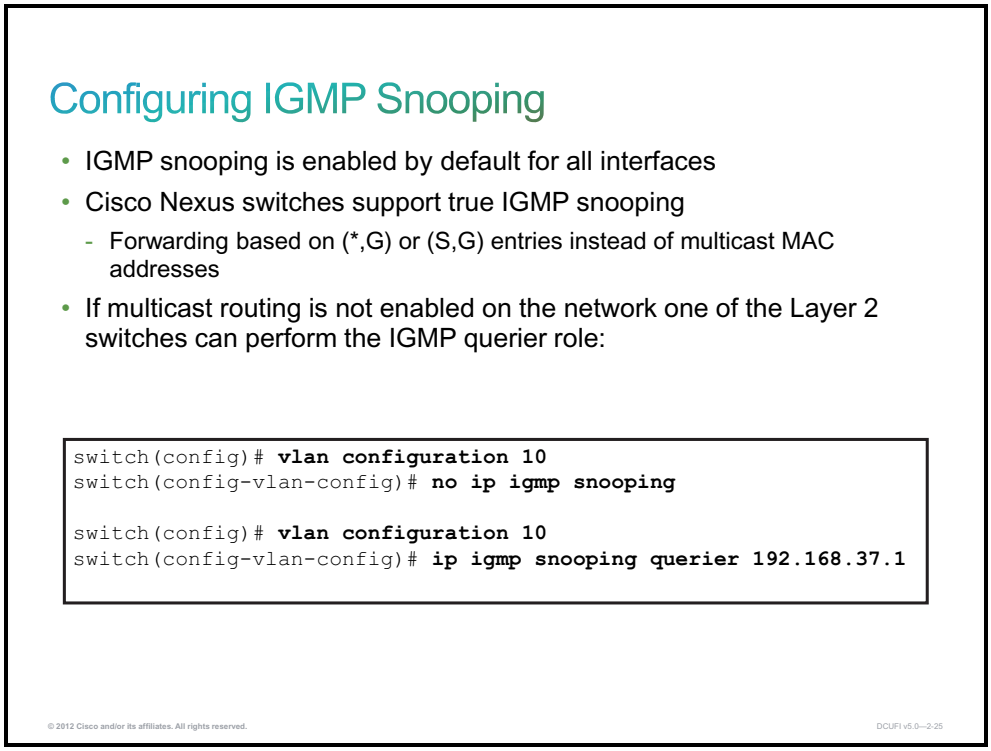
The translation between IP multicast and MAC address is achieved by the mapping of the low-order 23 bits of the IP (Layer 3) multicast address into the low-order 23 bits of the IEEE (Layer 2) MAC address. In the MAC address, the low-order bit (0x01) in the first octet indicates that this packet is a Layer 2 multicast packet. The 0x01005e prefix (vendor code) has been reserved for use in mapping Layer 3 IP multicast addresses into Layer 2 MAC addresses.

Because there are 28 bits of unique address space for an IP multicast address (32 minus the first four bits containing the 1110 Class D prefix) and there are only 23 bits mapped into the IEEE MAC address, there are five bits of overlap—or 28 – 23 = 5, and 2^5 = 32. So there is a 32:1 overlap of Layer 3 addresses to Layer 2 addresses. Therefore, be aware that several Layer 3 addresses may map to the same Layer 2 multicast address.

IGMP leverages this IP-to-MAC address mapping in order to forward multicast streams only to interested receivers. Without IGMP snooping, the switch would flood multicast packets to all ports in the VLAN. All hosts that are attached to the VLAN not listening to the multicast group would then discard the packets.

With IGMP snooping, the switch learns about connected receivers of a multicast group. The switch then leverages this information in order to forward corresponding multicast frames only to the interested receivers. The forwarding decision is based on the destination MAC address. The switch uses the IP-to-MAC address mapping to decide which frames (or destination MAC) go to which IP multicast receivers.

https://t.me/learningnets



IGMP snooping is enabled by default for all VLANs on a Cisco Nexus 5000 Series switch or Cisco Nexus 7000 Series switch. However, if required, IGMP snooping can be specifically disabled for a particular VLAN. To disable IGMP snooping for a VLAN, use the **no ip igmp snooping** command. Where precisely to configure this command is dependent upon the type of switch and Cisco NX-OS Software version. For Cisco NX-OS Software Release 5.1 and newer releases, for the Cisco Nexus 7000 Series switches, disabling IGMP snooping is configured by using the **vlan configuration** *vlan* command. However, in older Cisco NX-OS Software releases, disabling IGMP snooping is configured by using the **vlan** *vlan* command.

Generally, it is not necessary to configure IGMP snooping specifically in an IP multicast-enabled network because the feature is enabled by default. The exception is a network where IP multicast is used locally in a VLAN but IP multicast routing is not enabled. IGMP snooping needs an IGMP querier to function properly, but if there is no multicast router on the segment, this function is missing. In this scenario, you can configure one of the Layer 2 switches to perform the IGMP querier function.

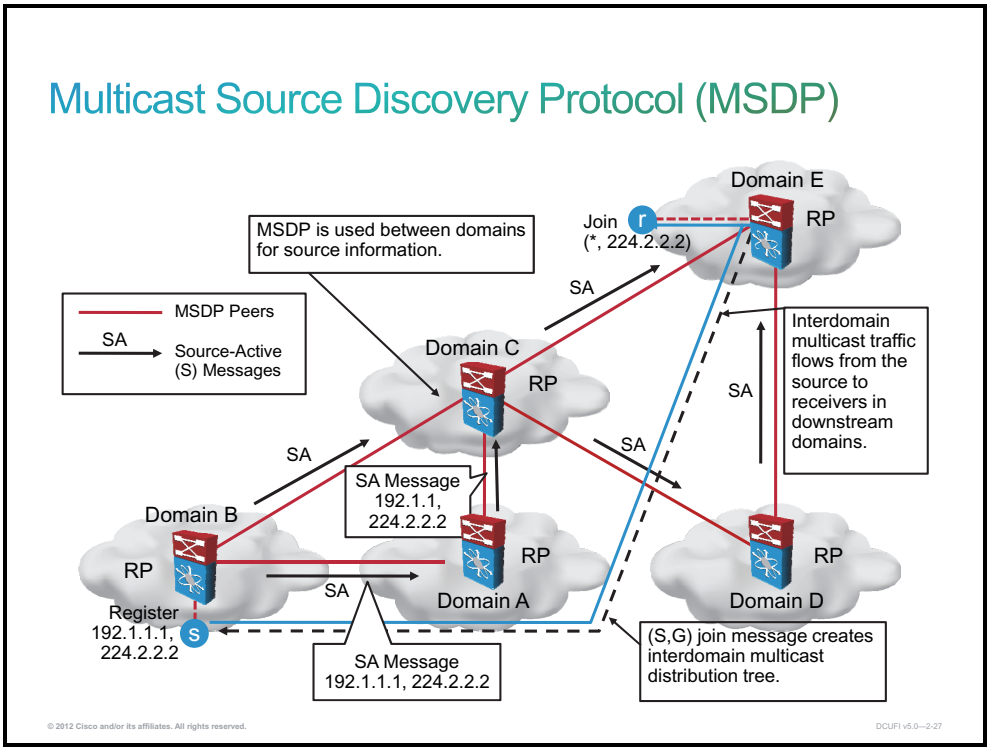
The example in the figure shows how to configure a Cisco Nexus 5500 Platform switch and a Cisco Nexus 7000 Series switch as the IGMP querier for a VLAN.

Note	Similar to other IGMP snooping-related commands, the **ip igmp snooping querier** command is configured by using the **vlan configuration** command in the latest versions of the Cisco NX-OS Software.

https://t.me/learningnets

Configuring MSDP

This topic identifies how to configure MSDP on the Cisco Nexus 7000 Series switch.



MSDP provides a solution for interdomain multicast routing. This protocol works with PIM-SM and allows RPs in one domain to announce their sources to other domains by using source-active (SA) messages. Cisco Nexus switches or routers that act as RPs establish MSDP peering and then exchange information on the active sources and groups to which they are sending.

When information is received in a domain with active receivers for the group, an appropriate (S,G) join is initiated (by the RP of the domain) across domains toward the source. With this mechanism, downstream domains pull the traffic from the sources in upstream domains. When (S,G) joins travel toward the source, the distribution tree is built across domains.

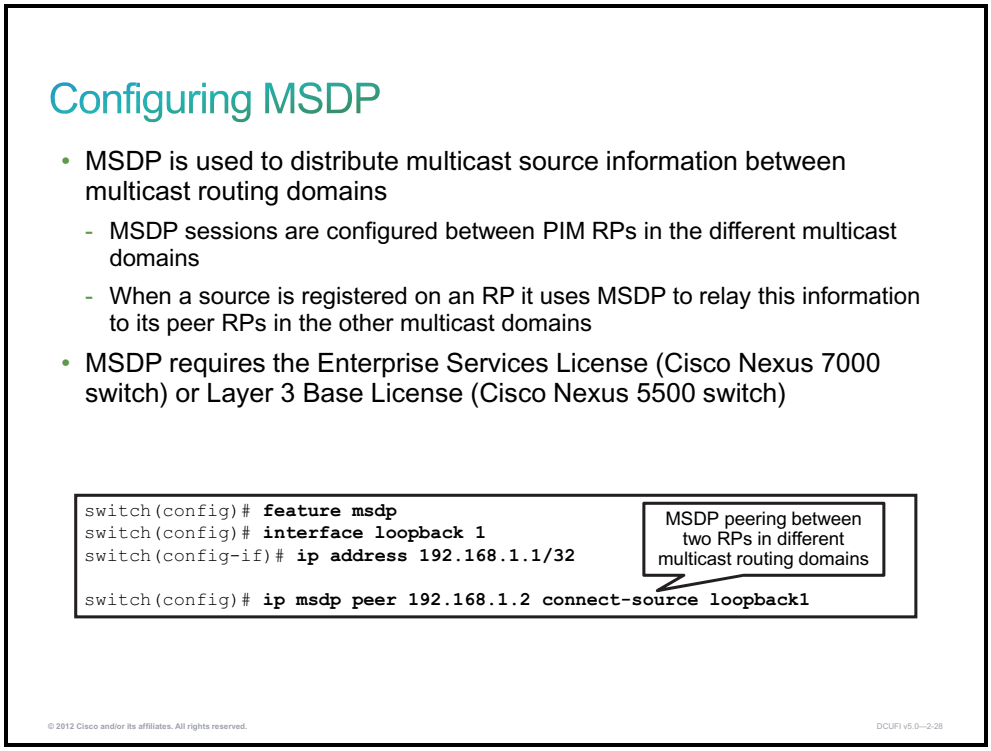
As soon as the distribution tree is built across domains, the traffic from the active source starts flowing to downstream receivers.

The only working solution today for interdomain multicast is PIM-SM with Multiprotocol Border Gateway Protocol (MP-BGP) and MSDP.

In the figure, the 192.1.1.1 source in domain B starts sending multicast traffic for the 224.2.2.2 group. The first-hop router in domain B sends a register message to the RP in domain B. The RP in domain B then sends a source-active message to all of its MSDP peers.

© 2012 Cisco Systems, Inc.

https://t.me/learningnets



You need the LAN Enterprise Package in order to implement MSDP on a Cisco Nexus 7000 Series switch and the Layer 3 Base License in order to deploy it on the Cisco Nexus 5500 Platform Switch.

The example in the figure shows a basic MSDP configuration for a single MSDP session between two MSDP peers. The example shows one side of the session. The other side should be configured similarly.

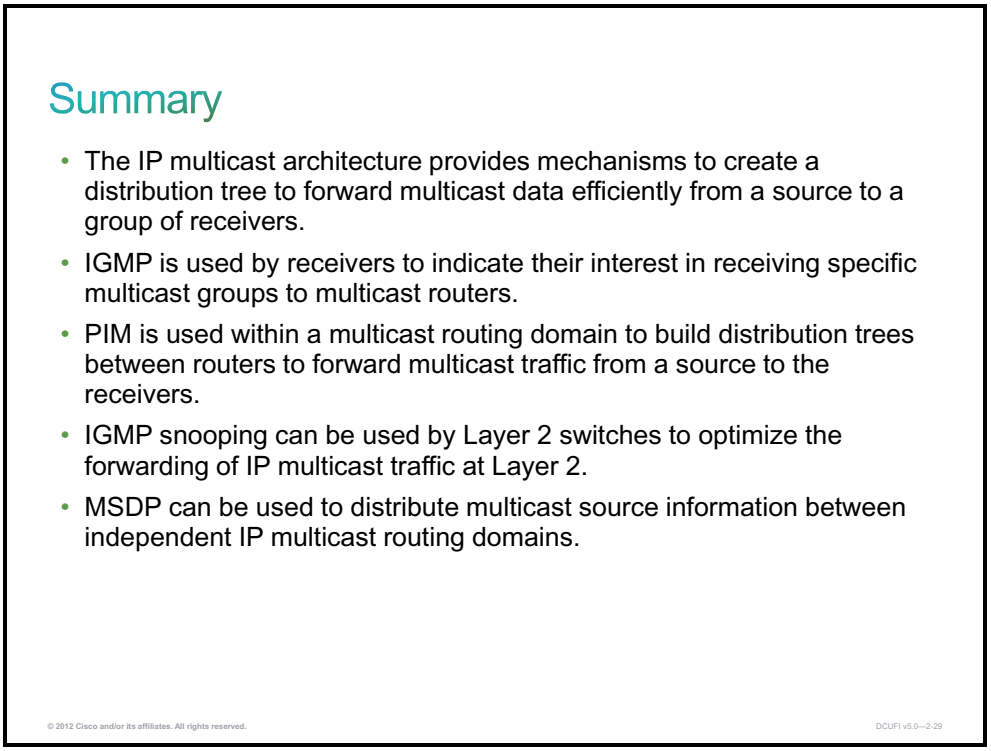
The example in the figure uses the following commands:

- **feature msdp**: This command enables the MSDP feature. This feature requires the Enterprise Services License.
- **ip msdp peer** *peer-address* **connect-source** *if-type if-number*: This command defines an MSDP peer. This command needs to be configured on both routers in order to establish an MSDP session. It is important that the peer address configured on one side matches the connect source on the other side and vice versa.

https://t.me/learningnets

Summary

This topic summarizes the key points that were discussed in this lesson.



References

For additional information, refer to these resources:

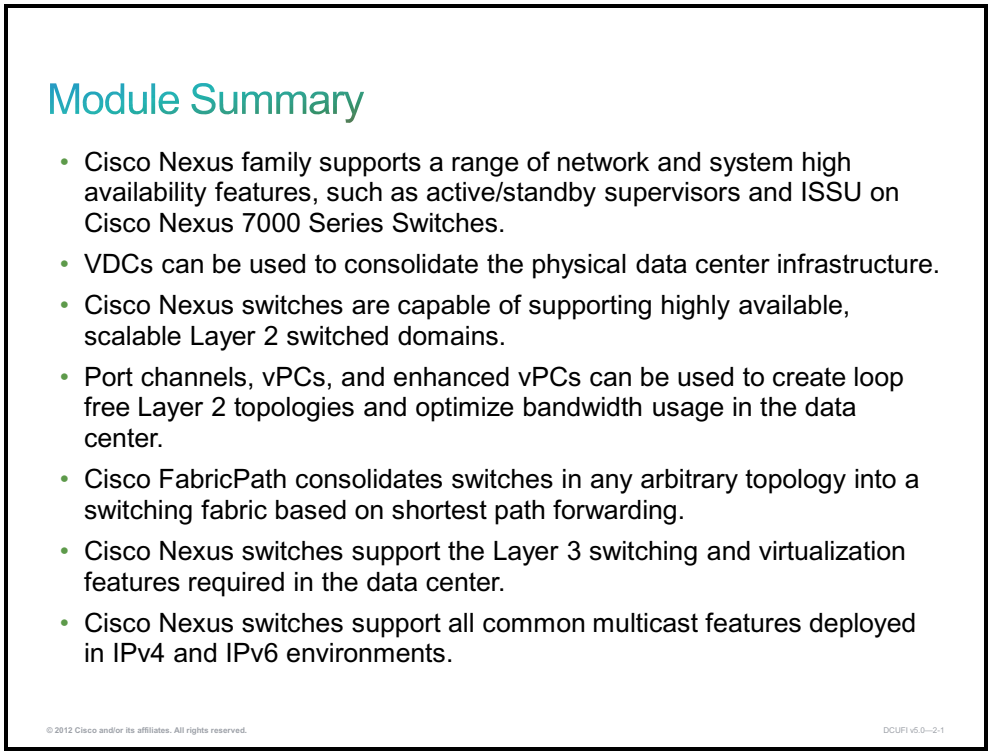
- To learn more about configuring multicast on Cisco Nexus 7000 Series Switches, refer to *Cisco Nexus 7000 Series NX-OS Multicast Routing Configuration Guide* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/sw/nx-os/multicast/configuration/guide/b_multicast.html
- To learn more about configuring multicast on Cisco Nexus 5500 Series Switches, refer to *Cisco Nexus 5000 Series NX-OS Multicast Routing Configuration Guide, Release 5.0(3)N1(1)* at this URL: http://www.cisco.com/en/US/docs/switches/datacenter/nexus5000/sw/multicast/5_0_3_N1_1/Cisco_n5k_layer3_multicast_config_gd_rel_503_N1_1.html

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Module Summary

This topic summarizes the key points that were discussed in this module.



To build a highly available and scalable data center infrastructure, it is necessary to support a broad range of technologies and features. The Cisco Nexus Operating System (NX-OS) Software running on the Cisco Nexus switches provides features that are specifically developed for use in the data center environment. Network and system high-availability features ensure uninterrupted operation. Virtual device contexts (VDCs) allow separated administrative and policy domains to be consolidated on a single physical infrastructure.

Advanced Layer 2 switching features are available to build a solid Layer 2 foundation in the data center access and aggregation layers. Port channels, virtual port channels (vPC), and enhanced vPCs allow loop-free logical Layer 2 topologies, which optimize the use of bisectional bandwidth and increase the availability of the data center infrastructure.

Cisco FabricPath is an innovative Cisco NX-OS Software technology that can transform the way data center networks are conceived. Cisco FabricPath brings the benefits of Layer 3 routing to Layer 2 switched networks in order to build a highly resilient and scalable Layer 2 fabric.

Advanced Layer 3 routing, switching, and virtualization features are available to provide isolated fault domains, increased scalability, and fast network convergence. The Cisco Nexus switches also support advanced IP multicast technologies and features to support critical business applications that use multicast-based data delivery in both IPv4 and IPv6 environments.

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets

Module Self-Check

Use the questions here to review what you learned in this module. The correct answers and solutions are found in the Module Self-Check Answer Key.

Q1) Which option provides the most effective load-balancing solution? (Source: Understanding High Availability and Redundancy)

A) HSRP with STP
B) GLBP with STP
C) HSRP with vPC
D) GLBP with PC

Q2) Which statement best describes Nonstop Forwarding (NSF)? (Source: Understanding High Availability and Redundancy)

A) It exchanges frequent hello messages to detect connectivity problems
B) It replaces hello messages of the routing protocols
C) It facilitates graceful restart when a process fails
D) It performs Stateful Switchover (SSO) when the active supervisor experiences a problem

Q3) Which component in the Cisco Nexus 7000 Series switch is upgraded last during a Cisco IOS ISSU process? (Source: Understanding High Availability and Redundancy)

A) active supervisor
B) CMP
C) I/O module
D) standby supervisor

Q4) Which of the following virtualization models best describes VDCs? (Source: Configuring Virtual Device Contexts)

A) Data plane virtualization
B) Data plane and control plane virtualization
C) Data plane, control plane, and management plane virtualization
D) Data plane, control plane, and operating environment virtualization with resource control
E) Data plane, control plane, management plane, and operating environment virtualization with resource control

Q5) Which of the following deployment scenarios are enabled by VDCs? (Choose three.) (Source: Configuring Virtual Device Contexts)

A) Split core for migrations and mergers
B) Access layer expansion for increased numbers of access switches
C) Multiple aggregation blocks for management and policy separation
D) Service insertion for increased security and fault isolation
E) Collapsed core for reduced management points

https://t.me/learningnets

Q6) How many VDCs are supported on a Cisco Nexus 7000 Series switch? (Source: Configuring Virtual Device Contexts)

A) Four equal VDCs, which control a subset of the resources on the switch
B) Eight VDCs, which control a subset of the resources on the switch
C) One default VDC, which controls the entire switch and three nondefault VDCs, with limited control over a subset of the switch resources
D) One default VDC, which controls the entire switch and seven nondefault VDCs, with limited control over a subset of the switch resources
E) Two VDCs, which control a subset of the resources on the switch

Q7) Which of the following resources can be assigned to a specific VDC on a Cisco Nexus 7000 Series switch? (Source: Configuring Virtual Device Contexts)

A) A percentage of the supervisor CPU capacity
B) Individual interfaces or port groups on I/O modules
C) A dedicated slice of the supervisor DRAM
D) The out-of-band (OOB) management interface

Q8) Which of the following resources can be limited through a VDC resource template? (Choose three.) (Source: Configuring Virtual Device Contexts)

A) Number of VLANs
B) Number of SPAN sessions
C) Amount of space on the supervisor CompactFlash in megabytes
D) Power consumption for the ports that are assigned to the VDC
E) Amount of supervisor memory that is assigned to IPv4 routes in megabytes
F) Amount of supervisor memory that is assigned to MAC addresses and access list entries

Q9) When you change a value in a resource template, it is automatically applied to all VDCs that use that resource template. (Source: Configuring Virtual Device Contexts)

A) true
B) false

Q10) Which of the following licenses needs to be installed in order to configure VDCs? (Source: Configuring Virtual Device Contexts)

A) Enterprise Services Package
B) Transport Services Package
C) Enhanced Layer 2 Package
D) Scalable Feature Package
E) Basic Storage Services Package
F) Advanced Services Package
G) Storage Protocols Services Package
H) Virtualization Services Package

Q11) Which feature must be enabled in order to implement a storage VDC? (Source: Configuring Virtual Device Contexts)

A) FCoE Initialization Protocol (FIP)
B) Link Layer Discovery Protocol (LLDP)
C) Link Aggregation Control Protocol (LACP)
D) FCoE license

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Q12) Which of the following commands is used to configure the first four Gigabit Ethernet ports on an I/O module of a Cisco Nexus 7000 Series switch? (Source: Configuring Layer 2 Switching Features)

A) `interface range ethernet 1/1-4`
B) `interface ethernet 1/1-4`
C) `interface range GigabitEthernet 1/1-4`
D) `interface GigabitEthernet 1/1-4`

Q13) Which of the following statements best describes the use of dedicated mode on the N7K-M132XP-12 I/O modules? (Source: Configuring Layer 2 Switching Features)

A) The first port in the port group can use up to 10 Gb/s. The other three ports in the port group are down and cannot be enabled.
B) One of the four ports in the port group can be configured to use up to 10 Gb/s. The other three ports in the port group are down and cannot be enabled.
C) One of the ports in the port group can use up to 10 Gb/s, but if you enable any of the other ports in the port group, they share the bandwidth.
D) One of the ports in the port group can use up to 10 Gb/s. If you enable any of the other ports in the port group, all ports are error-disabled.

Q14) When you change a command in a port profile, it is automatically applied to all ports that inherit that particular port profile. (Source: Configuring Layer 2 Switching Features)

A) true
B) false

Q15) What do you need to configure when deploying Cisco Adapter FEX? (Choose three.) (Source: Configuring Layer 2 Switching Features)

A) vEthernet interfaces
B) Cisco FEX interfaces
C) Virtual Interface Configuration (VIC) protocol
D) Port profiles
E) Interface mode—shared or dedicated
F) Auto-creation feature
G) Channel numbers

Q16) Which of the following are valid private VLAN types that can be configured for a VLAN? (Choose three.) (Source: Configuring Layer 2 Switching Features)

A) Primary
B) Secondary
C) Tertiary
D) Closed
E) Community
F) Secure
G) Isolated

Q17) Which of the following types of ports can a port in a private VLAN of type community communicate with? (Choose two.) (Source: Configuring Layer 2 Switching Features)

A) Isolated ports
B) Community ports for all secondary VLANs
C) Community ports in the same secondary VLAN
D) Promiscuous ports

https://t.me/learningnets

Q18) What is the default type of Spanning Tree Protocol (STP) used by Cisco Nexus switches? (Source: Configuring Layer 2 Switching Features)

A) PVRST+
B) PVST+
C) 802.1D-1998
D) MST
E) 802.1D-2004

Q19) Which types of ports send bridge protocol data units (BPDUs) when the bridge assurance feature is enabled on the port? (Source: Configuring Layer 2 Switching Features)

A) Designated ports only
B) Designated and root ports
C) Root ports and alternate ports
D) Any type of spanning-tree port

Q20) Which of the following header fields can be used in the port channel load-balancing hash algorithm? (Choose three.) (Source: Configuring PortChannels)

A) Destination IP address
B) Source MAC address
C) TCP flags
D) IP header length
E) ICMP type and code
F) TCP destination port

Q21) Match the components of the vPC architecture to their descriptions. (Source: Configuring PortChannels)

A) vPC peer link
B) vPC peer keepalive link
C) Cisco Fabric Services

_____ 1. Carries control traffic between vPC peer devices

_____ 2. Used to reliably synchronize vPC control plane information

_____ 3. Carries heartbeat messages to detect a dual-active condition

Q22) A vPC domain cannot consist of more than two switches or VDCs. (Source: Configuring PortChannels)

A) true
B) false

Q23) Which of the following commands allows a vPC switch to forward traffic for the vPC peer router MAC address? (Source: Configuring PortChannels)

A) peer-switch
B) peer-gateway
C) peer-link
D) peer-mac
E) peer-router

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Q24) Which of the following Cisco FEX deployment options is supported on the Cisco Nexus 7000 Series switches? (Source: Configuring PortChannels)

A) Active-active FEX
B) Straight-through with static pinning
C) Straight-through with dynamic pinning
D) Active-standby FEX

Q25) Match the correct Cisco NX-OS command to the function that it has in the vPC troubleshooting process. (Source: Configuring PortChannels)

A) **show port-channel summary**
B) **show vpc consistency parameters vpc**
C) **show vpc**
D) **show vpc peer-keepalive**

_____ 1. This command can be used to verify that the vPC peer link is operational in addition to allow viewing of the global vPC parameters.

_____ 2. This command can be used to verify the state of the port channel interfaces, both on the vPC peer switches and on the connected downstream device.

_____ 3. This command displays the state of the peer-keepalive link, which must be operational before the vPC peer link can come up.

_____ 4. This command can be used to verify that the configuration on the vPC port channels is consistent on both vPC peer switches.

Q26) The vPC peer-keepalive link must be operational before the vPC peer link can come up. (Source: Configuring PortChannels)

A) true
B) false

Q27) In which topology can you deploy enhanced vPC? (Source: Configuring PortChannels)

A) Server dual-homed to two Cisco Nexus 5500 Platform switches
B) Dual-homed server connected by a port channel to a single Cisco FEX, with the Cisco FEX dual-homed to two switches
C) Dual-homed server connected to two Cisco FEXs, with both Cisco FEXs dual-homed to two Cisco Nexus 7000 Series switches
D) Dual-homed server connected to a pair of Cisco FEXs that connects to a single switch

Q28) What is the default MAC address learning for Cisco FabricPath VLANs? (Source: Implementing Cisco FabricPath)

A) Source
B) Conversational
C) Destination
D) Source/Destination

https://t.me/learningnets

Q29) Which of the following protocols is used as the Cisco FabricPath control protocol? (Source: Implementing Cisco FabricPath)

A) OSPF
B) IS-IS
C) BGP
D) OTV
E) Cisco Fabric Services

Q30) Which of the following is true about Cisco FabricPath forwarding in the Cisco FabricPath core? (Source: Implementing Cisco FabricPath)

A) MAC addresses are learned using conversational MAC learning.
B) MAC addresses are learned using the regular Ethernet flooding process.
C) No MAC address lookups or MAC address learning is required.
D) MAC address reachability is advertised using Cisco FabricPath IS-IS.

Q31) The Cisco Nexus 7000 Series switch F1 I/O modules and Cisco Nexus 5500 Platform switch hardware are capable of running both Cisco FabricPath and TRILL modes. (Source: Implementing Cisco FabricPath)

A) true
B) false

Q32) Match the field in the Cisco FabricPath header with the correct explanation. (Source: Implementing Cisco FabricPath)

A) Switch ID
B) Subswitch ID
C) Port ID
D) FTag (Forwarding Tag)
E) TTL (Time to Live)

_____ 1. Identifies the destination or source interface

_____ 2. Decremented at each hop to prevent loops

_____ 3. Identifier of topology or multidestination distribution tree

_____ 4. Identifies devices/hosts connected via VPC+

_____ 5. Unique number identifying each Cisco FabricPath node

Q33) Which of the following is true about Cisco FabricPath vPC+? (Source: Implementing Cisco FabricPath)

A) It creates a logical switch ID for each switch in the vPC domain.
B) It can be deployed in Classic Ethernet VLANs and Cisco FabricPath VLANs.
C) On Cisco Nexus 7000 Series switches, it works only on F Series modules
D) It requires only one address lookup on the egress switch for traffic going out through vPC+.

Q34) All routing protocols require the Enterprise Services License to be installed on a Cisco Nexus 7000 Series switch. (Source: Configuring Layer 3 Switching Features)

A) true
B) false

 © 2012 Cisco Systems, Inc.

https://t.me/learningnets

Q35) Which of the following commands is used to configure a static route for a VRF? (Source: Configuring Layer 3 Switching Features)

A) `switch (config)# ip route 10.0.0.0/8 172.16.1.1 vrf RED`
B) `switch (config-if)# ip route 10.0.0.0/8 172.16.1.1`
C) `switch (config-if)# ip route 10.0.0.0/8 172.16.1.1 vrf RED`
D) `switch (config-vrf)# ip route 10.0.0.0/8 172.16.1.1`

Q36) Which command enables Hot Standby Router Protocol (HSRP) on a Cisco Nexus 7000 Series switch? (Source: Configuring Layer 3 Switching Features)

A) `switch(config)# enable hsrp`
B) `switch(config)# feature hsrp`
C) `switch(config)# hsrp enable`
D) `switch(config)# feature hsrp enable`

Q37) Which redundancy protocols provide the best protection against undesired peering with rogue gateways? (Choose two.) (Source: Configuring Layer 3 Switching Features)

A) FHRP
B) HSRPv1
C) HSRPv2
D) VRRP
E) GLBP

Q38) Match the components of the unicast routing architecture components on the Cisco Nexus 7000 Series switches with their descriptions. (Source: Configuring Layer 3 Switching Features)

F) UFDM
G) RIB
H) FIB
I) TCAM

_____ 1. Specialized hardware that contains forwarding information used in packet header lookups

_____ 2. Exists on the active supervisor and distributes the forwarding path information to the I/O modules

_____ 3. Builds the information that is used for the hardware-forwarding engine

_____ 4. Contains routing information learned through routing protocols and other sources

Q39) Which of the following commands is used to filter routes when redistributing between routing protocols? (Source: Configuring Layer 3 Switching Features)

A) `switch(config-router)# redistribute eigrp 200 route-map EIGRP-TO-OSPF`
B) `switch(config-router)# redistribute eigrp 200 prefix-list EIGRP-TO-OSPF`
C) `switch(config-router)# redistribute eigrp 200 access-list name EIGRP-TO-OSPF`
D) `switch(config-router)# distribute-list name EIGRP-TO-OSPF out eigrp 200`

https://t.me/learningnets

Q40) What happens to a packet that matches one of the match statements in a policy-based routing (PBR) route map entry with a **deny** action? (Source: Configuring Layer 3 Switching Features)

A) The packet is dropped.
B) The next entry in the route map is processed.
C) The packet is forwarded using normal destination-based routing.
D) The packet is treated according to the configured policy

Q41) Which of the following commands are used to start an OSPFv3 process and enable it on interface Ethernet 1/1? (Choose four.) (Source: Configuring Layer 3 Switching Features)

A) `switch(config)# feature ospfv3`
B) `switch(config)# router ospfv3 1`
C) `switch(config)# ipv6 router ospfv3 1`
D) `switch(config-router)# address-family ipv6 unicast`
E) `switch(config)# interface ethernet 1/1`
F) `switch(config-if)# ipv6 router ospfv3 1 area 0`
G) `switch(config-router-af)# network 0.0.0.0 255.255.255.255 area 0`
H) `switch(config-router-af)# interface ethernet 0/0 area 0`

Q42) Which of the following Cisco NX-OS commands configures an operational IPv6 static default route? (Source: Configuring Layer 3 Switching Features)

A) `switch(config)# ipv6 route ::/0 FE80::260:3EFF:FE47:1530`
B) `switch(config-if)# ipv6 route ::/0 FE80::1`
C) `switch(config)# ipv6 route ::/0 FE80::1 ethernet 2/1`
D) `switch(config)# ipv6 route ::/0 2001::ffff:ffff::6`

Q43) Which of the following protocols can be used for intradomain IP multicast routing on a Cisco Nexus 7000 Series switch? (Choose three.) (Source: Configuring IP Multicast)

A) MSDP
B) MBGP
C) MOSFP
D) PIM sparse mode
E) PIM dense mode
F) PIM SSM
G) BIDIR-PIM
H) IGMPv2
I) IGMPv3

Q44) Which of the following commands are used to enable IGMPv3 on an interface? (Choose two.) (Source: Configuring IP Multicast)

A) `switch(config-if)# ip pim sparse-mode`
B) `switch(config-if)# ip igmp`
C) `switch(config-if)# ip igmp version 3`
D) `switch(config-if)# ip igmp enable`
E) `switch(config-if)# ip igmp querier 10.1.1.1`

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Q45) Which of the following options can be used to distribute PIM RP information in a PIM sparse mode deployment? (Choose three.) (Source: Configuring IP Multicast)

A) Static RP configuration
B) MSDP
C) MBGP
D) Auto-RP
E) Cisco Fabric Services
F) BSR
G) MOSPF

Q46) What destination MAC address has IP multicast packets? (Source: Configuring IP Multicast)

A) Destination MAC address FFFF.FFFF.FFFF
B) Destination MAC address of the appropriate receiver
C) Destination MAC address corresponding to multicast group
D) Destination MAC address requested by IGMP messages

Q47) Which of the following commands can be used to view the multicast forwarding state for the multicast group 239.1.1.1? (Source: Configuring IP Multicast)

A) **`show ip igmp group 239.1.1.1`**
B) **`show ip pim topology 239.1.1.1`**
C) **`show ip mroute 239.1.1.1`**
D) **`show ip mcast 239.1.1.1`**
E) **`show forwarding ipv4 mcast 239.1.1.1`**

https://t.me/learningnets

Module Self-Check Answer Key

Q1) C

Q2) C

Q3) C

Q4) D

Q5) A, C, D

Q6) C

Q7) B

Q8) A, B, E

Q9) B

Q10) F

Q11) B

Q12) B

Q13) A

Q14) A

Q15) B, D, G

Q16) A, E, G

Q17) C, D

Q18) A

Q19) D

Q20) A, B, F

Q21) 1-A
2-D
3-B
4-C

Q22) A

Q23) B

Q24) C

Q25) 1- C
2-A
3-D
4-B

Q26) A

Q27) B

Q28) B

Q29) B

Q30) C

Q31) A

Q32) 1- C
2-E
3-D
4-B
5-A

Q33) C

© 2012 Cisco Systems, Inc.

https://t.me/learningnets

Q34) B

Q35) D

Q36) B

Q37) C, E

Q38) D, A, C, B

Q39) A

Q40) C

Q41) A, B, E, F

Q42) C

Q43) D, F, G

Q44) A, C

Q45) A, D, F

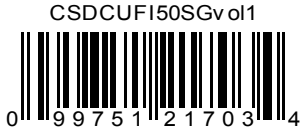
Q46) D

Q47) C

Q48) C

https://t.me/learningnets

© 2012 Cisco Systems, Inc.
https://t.me/learningnets



https://t.me/learningnets