Global Knowledge®

Expert Reference Series of White Papers

Cisco Data Center Ethernet

Cisco Data Center Ethernet

Dennis Hartmann, Global Knowledge Senior Instructor, CCIE, CCVP, CSI, CCNP, CCIP, MCSE

Introduction

Cisco's Catalyst 6500 (Cat6K) series platform is one of the most widely deployed switches used for data center and enterprise network connectivity. The Cat6K platform has been in production for nearly a decade and the longevity of the platform is a testament to Cisco's value proposition of investment protection. Cisco has introduced many chassis, power supply, supervisor and line card modules to assist with ever-changing network capacity requirements and technology evolution.

VMware and other virtualization products are revolutionizing network requirements of the data center as dedicated application servers are migrated to virtual machines (VMs). These larger servers can accommodate the computational requirements of running multiple server-based operating systems on one hardware platform. Traditional network switch platforms can only apply network policies statically on a per port basis. VMware virtualization has created a requirement for network policies to be performed on a per virtual machine basis. VMware technologies like VMotion and Distributed Resource Scheduling (DRS) enable automatic virtual machine migration across physical servers, further increasing the complexity of switch requirements. Network policies need to be able to migrate between switches (or ports on the same switch) to assure proper network policy for migrating virtual machines.

Cisco has worked very closely with VMware to create new technologies that enable virtual machine aware policies in data center switches that have migratory properties. This white paper will provide an overview of some of Cisco's data center platforms and their virtualization aware features; specifically the Nexus product line of switches (1000V, 2000, 5000, 7000) and the Cisco Unified Computing System (UCS).

Data Center Evolution

Data centers began as centralized mainframe-based systems that had the benefit of centralized administration, but the disadvantage of limited flexibility and vendor choice. Most of the applications maintained in mainframe environments were migrated to server-based systems in the late 1990s to early 2000s due to lower costs, open standards, the Internet revolution, and a diverse talent pool. The exponential growth of server-based systems has resulted in a condition referred to as "server sprawl" where large enterprises have hundreds of different servers dedicated to certain applications. Network managers have been tasked to minimize the costs associated with operations, administration, maintenance, and provisioning (OAM&P) of these vast distributed systems, while also maximizing the utilization of computing resources.

Computing Resource Optimization

When put into service, servers and clustered server systems are sized for peak traffic loads to accommodate the busiest hour of the busiest day of the year. For instance, if a retail organization did not have enough resources to accommodate the holiday shopping rush, the cost of lost business could be very detrimental. Servers sized for these maximum utilization periods will typically run at a very low utilization for eleven months of the year, however. Under-utilized servers are very costly to IT budgets. VMware virtualization technology allows multiple operating systems to share the same hardware resources, while also providing fault isolation. One misbehaving operating system (OS) should not be able to affect another OS running on the same hardware server platform unless the two servers have dependencies on each other. Virtualization software allows the OS of individual virtual machines to share the multi-core processors and large random access memory (RAM) pools to maximize the utilization and efficiency of server solutions.

Virtualization

Virtualization technology operating systems are referred to as a "hypervisor". A hypervisor is a thin operating system that recognizes the hardware resources of a server and makes the resources available to the operating systems running on top of the hypervisor in virtual machines. VMware's hypervisor (VMkernel) uses virtual network interface cards (vNICs) that direct traffic to the physical NIC of the server hardware platform. vNICs allow many virtual machines to map network traffic to the same physical server NIC.

Network switches forward Ethernet traffic based on the destination media access control (MAC) address of the network traffic. vNICs use vMAC (virtual MAC addresses) to individually identify a virtual machine using the same physical NIC. Traditional network switches are not able to forward traffic received on a switch port back to the same port (hair pinning), which is a requirement for applications communicating between VMs that are running on the same physical server. Virtual machines located on the same server communicate through VMware software vSwitches without sending traffic through any physical NIC cards. vSwitches enable communication between two VMs on the same server, something that would otherwise be impossible using a physical network switch.

vSphere 4 (ESX and ESXi) introduced support of distributed virtual switches (DvSwitchs). A DvSwitch is very similar to the vSwitch concept, but it allows multiple physical servers running VMware to share the same virtual switch defined in the VMware vCenter management platform. DRS is an important virtualization technology that allows the VMware virtualization infrastructure to meet the needs of busy periods without the waste associated with dedicating hardware platforms for worst-case scenario utilization. Workloads are distributed amongst hardware servers utilizing VMotion to copy VMs between the physical servers. It's important that network policies follow VMs, but traditional switches and VMware's software based DvSwitches do not perform this operation. VMware created an application programming interface (API) with the help of Cisco Systems to allow third-party vendors to write their own switching software to tie in with VMware's virtualization infrastructure. At the time of this writing, Cisco Systems is the only vendor to take advantage of the VMware network API.

Nexus 1000V

Cisco's Nexus 1000V is the first third-party software-based distributed virtual switch that can be used in place of the VMware DvSwitch. The Nexus 1000V is a Layer 2 software-based switch that allows the application of network policies at the virtual machine level, instead of the physical switch port level. Nexus 1000V is a software implementation of Cisco's VN-Link technology that uses a 6 byte tag field in the Ethernet header (VNTag) to identify the VM in which the traffic was received. One physical NIC on a VMware server may be carrying traffic for many VMs, but the VNTag field will indicate the VM in which the traffic was received. VNTagging allows Cisco data center equipment to create a logical point-to-point connection between devices on the unified fabric. The VNTag header is shimmed in between the source MAC address and the tag protocol identifier (TPID) field in IEEE 802.1Q trunk links.

VN-Link and VNTags are also implemented in hardware in the following platforms.

- Cisco UCS Servers
- Nexus 2000/5000/7000 switches
- Cisco UCS 6100 Fabric Interconnects
- Cisco UCS 2100 Fabric Extenders

Network policies currently supported by the Nexus 1000V include the following.

- Access Control Lists (ACLs)
- Private VLANs (PVLAN)
- NetFlow statistics
- Quality of Service (QoS)
- IGMP Snooping
- Port Channel
- Port mirroring (SPAN)

Nexus 1000V deployments utilize a server-based OS installed as a VM or physical server referred to as a virtual supervisor module (VSM). Virtual Ethernet Modules (VEMs) are created in the Nexus 1000V VSM that runs the NX-OS operating system. Port profiles are configured in the VSM and downloaded to the VEMs that operate as logical switches in the VMware hypervisor. VEMs contain the network policies that traditional Ethernet switches cannot apply on a per VM basis.

Nexus switches run Cisco's NX-OS variant of Cisco IOS. NX-OS is a trimmed-down version of Cisco IOS with new features for the data center environment. NX-OS allows the use of show commands in configuration mode and prefix length notation syntax for subnet masks instead of dotted decimal notation (e.g., /24 instead of 255.255.255.0). The references section at the end of this white paper has a link to some YouTube videos in which you can learn more about Cisco NX-OS. If you are already familiar with Cisco IOS, it is my belief that you will be comfortable with NX-OS. NX-OS 4.1(3)N1 introduced standards-based fibre channel over Ethernet (FCoE) support to the Nexus 5000 platform.

Unified Fabric

Server data center connectivity has leveraged many Gigabit Ethernet (GE) local area network (LAN) connections and multiple Fibre Channel (FC) connections to the storage area network (SAN) environment. Large servers typically have many GE NICs (LAN connectivity) and many host bus adaptors (HBA) for connectivity to the storage area network (SAN). Fibre Channel HBA functionality and LAN NIC functionality has been combined in the Fibre Channel over Ethernet (FCoE) standard. FCoE requires converged network adaptors (CNA) on the server platform and FCoE aware switch platforms.

Many components of the Cisco Nexus product line support FCoE technology, creating one Ethernet-based unified fabric for both LAN and SAN connectivity. Converged network adaptors leverage 10GE-based NICs to support the traffic requirements of both the LAN and SAN traffic. Emulex and QLogic (two CNA vendors) cards also include TCP offload engines (TOE) in the CAN adaptors which offload the TCP/IP header processing of IP packets from the server CPU to optimized hardware on the CNA. A unified fabric minimizes cabling requirements, switch interfaces, power requirements, and adapters.

Most TCP/IP-based network traffic is resilient and can accommodate packet loss by retransmitting the lost data. Fibre channel traffic has been designed around a lossless architecture that does not make any attempt at retransmitting traffic, resulting in a technology that cannot accommodate packet loss. It is very important that FCoE infrastructures are designed as a lossless infrastructure to ensure SAN traffic (fibre channel) is not dropped. The IEEE Data Center Bridging Task Force is in the process of standardizing the following Ethernet based prioritization technologies aimed at ensuring lossless fibre channel storage transmission over Ethernet.

- Priority-based Flow Control (PFC) – IEEE 802.1Qbb
- Enhanced Transmission Selection (ETS) – IEEE 802.1Qaz
- Congestion Notification – IEEE 802.1Qau

Nexus 5000

Cisco Nexus 5000 switches (5010 and 5020) are Layer 2 switches that provide (20) or (40) 10 Gigabit Ethernet/FCoE SFP+ ports respectively. The platform supports low latency forwarding with only 3.2 microseconds of switching delay between ingress and egress ports with the full feature set applied. A Nexus 5010 switch supports one expansion module, while the Nexus 5020 has two expansion modules. Expansion modules have the following hardware options.

- 8-port 1/2/4 Gbps Fibre Channel
- 6-port 1/2/4/8 Gbps Fibre Channel
- 4-port 10 Gigabit Ethernet (DCB and FCoE) and 4-port 1/2/4 Gigabit Fibre Channel
- 6-port 10 Gigabit Ethernet (DCB and FCoE)

The Nexus 5000 platforms uses the data center bridging exchange (DCBX) protocol that extends the capabilities of link layer discovery protocol (LLDP). DCBX also communicates priority flow control (PFC) information across switch platforms which aids in the lossless forwarding of FCoE-based storage traffic. The Nexus 5000 platform also supports a hardware-based Cisco VN-Link support that has been submitted to the IEEE for consideration as a network interface virtualization (NIV) standard.

Short Reach Optics

The Cisco SFP+ CX1 solution includes two very short reach connectivity solutions using Twinax cabling that allows 10GE connections at lengths of 1, 3, or 5 meters. The SFP+ CX1 Twinax solution has low latency (.25us) and very low power requirements (0.1 Watts). The pre-terminated SFP+ CX1 solution has lower complexity and much lower cost than traditional short-reach multi-mode fiber (MMF) connectivity solutions.

Nexus 2000

The end of rack (EoR) and top of rack (ToR) models are two well known switch design models in data center environments. The EoR design model typically has a large switch (e.g., Catalyst 6500) located at the end of a row. All devices in the row of equipment racks are cabled to the switch at the end of the row.

The Nexus 2000/5000 (2K/5K) solution follows a top of rack (ToR) design philosophy in which Nexus 2000 switches are located at the top of each 19" rack and cabled back to a centralized Nexus 5000 in the rack row. The ToR has great cabling advantages when compared to EoR designs, but the ToR philosophy has not been popular due to the extra overhead of administering and managing a large number of distributed switches in the data center. ToR design models have also had the disadvantage of configuration complexity and multiple fault domains in Layer 2 IEEE 802.1d spanning tree protocol (STP) environments.

The Nexus 2K/5K model avoids both of the disadvantages of the traditional ToR design approach by leveraging the following protocols to eliminate the administration overhead of distributed systems and the spanning tree protocol loop avoidance mechanism.

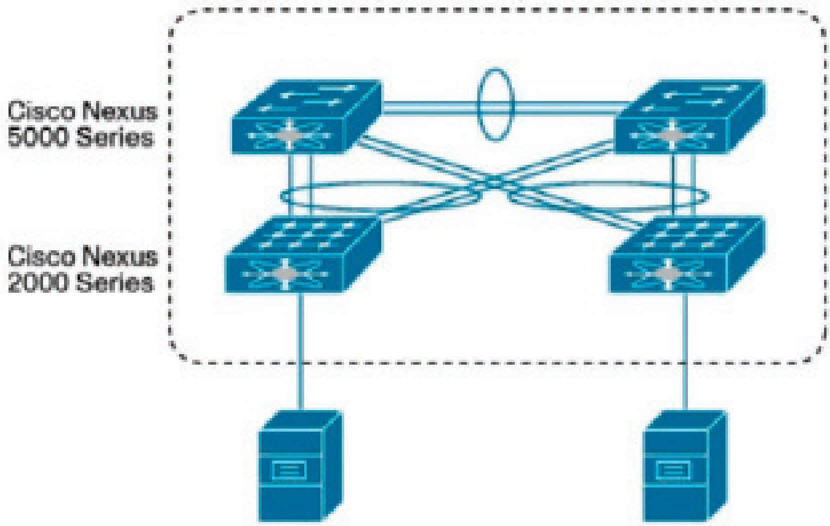
- Cisco Fabric Extender (FEX) – IEEE 802.1Qbh
- Virtual Port Channel (vPC) – IEEE 802.1Qbh

The Nexus 2000 switches are one rack unit (1RU) in height and have the following port configuration options.

- 2148T (48 1000Base-T / 4 10GE SFP+ uplinks)
- 2248TP (48 100/1000Base-T / 4 10GE SFP+ uplinks
- 2232PP (32 10GE/FCoE SFP+ ports / 8 10GE/FCoE uplinks)

Cisco Fabric Extender (FEX)

Nexus 2000 switches cannot be used without the Nexus 5000 series switches. The Nexus 2000 switches do not have an option to configure them directly. Nexus 2000 switches communicate with the Nexus 5000 over fabric extender (FEX) ports to download the NX-OS operating system from the Nexus 5000. The 2148 and 2248 have four 10GE interfaces that can be used as fabric extenders, while the 2232 has up to eight 10GE FEX uplinks. In the following example scenario, I will use the 2148 to illustrate the FEX functionality. A redundant data center design with the Nexus 2K/5K combination would bundle two links using link aggregation control protocol (LACP) to provide 20Gbps throughput to Nexus 5K switch 1, while providing a second 20Gbps connection to Nexus 5K switch 2 as shown in the figure below.



Notice in the diagram above that there are two port channel groups (LACP) configured on each of the Nexus 5000 switches. The Nexus 2000 at the bottom of the picture has 4 uplinks between the two Nexus 5000 switches in the same port channel, while a unique port channel interface is used between the Nexus 5000 switches.

Cisco Virtual Port Channel (vPC)

The design in the figure above creates a looped topology in which spanning tree protocol (STP) normally runs to provide Layer 2 loop avoidance. Cisco virtual port channel (vPC) technology creates an all links forwarding environment, eliminating the complexity and convergence time of STP, while also providing twice the throughput between the platforms (40Gbps). The Nexus 2000 logically sees a Nexus 5000 switch pair as one Nexus 5000 and prevents the challenges normally associated with a redundant, looped Layer 2 environment.

vPC technology is near identical to the virtual switching system (VSS) 1440 on the Catalyst 6500 introduced with the VSS SUP 720 – 10G supervisor module. Both vPC and VSS are commonly referred to as multi-chassis Etherchannel (MCEC).

Nexus 7000

The Nexus 7010 and 7018 are modular chassis based data center Ethernet (DCE) switches with Layer 3 routing capabilities aimed at providing highly scalable Gigabit and 10Gigabit Ethernet modules with future expansion to 40GE and 100GE. The first generation Nexus 7000 line cards connect to each fabric module at 46Gbps per connection (92Gbps full duplex). Each line card can connect to up to five redundant fabric modules delivering up to 230Gbps (460Gbps full duplex) of bandwidth per line card module. First generation line cards connecting to the switch fabric at these speeds include the following.

- 32 port 10 Gigabit Ethernet module
- 48 port Gigabit Ethernet module

Second-generation line cards leverage fabric connections at up to 80Gbps (160Gbps full duplex) per line card. Most N7K line cards have an XL option that allows each line card to accommodate the full Internet routing table (up to 1 million IPv4 and 350K IPv6 routes). Second generation line cards include the following.

- 8 port 10 Gigabit Ethernet module
- 32 port 10 Gigabit Ethernet module

Security features added to the N7K include security group tags (SGT) that allow access control lists (ACL) to be based on metadata instead of IP address and port information. Security group access control lists (SGACL) are processed in hardware on the N7K platform. The N7K also supports the IEEE 802.1AE (MACsec) security standard that leverages the Advanced Encryption Standard (AES-128) cipher to encrypt and decrypt Layer 2 connections between servers and the N7K switch ports.

Virtual Device Contexts (VDC)

MPLS VPNs (Multi-protocol label switching virtual private networks) brought virtualization to Layer 3 routing tables with the introduction of VPN routing and forwarding (VRF) tables in the late 1990s. MPLS VPN routing is supported on the Nexus 7000, but the Nexus 7000 has taken this virtualization idea to the next level with virtual device contexts (VDC). Up to four VDCs can be created on the N7K platform allowing segregation of traffic at Layer 1 and 2. Two ports in different VDCs cannot communicate directly through the Nexus 7000 switch. Per-VDC administration and fault isolation allows customers requiring separated networks to collapse their network infrastructures into one device while maintaining the autonomy of separate administration and fault isolation.

Overlay Transport Virtualization (OTV)

VMware VMotion technology allows VMs to be moved transparently between hardware platforms based on platform utilization, but does not support operation across Layer 3 routing boundaries. Both source and destination server platforms must be on the same Layer 2 switched network, which makes it impossible to use VMotion between data centers that are connected across Layer 3 routed links.

Overlay transport virtualization (OTV) technology on the Nexus 7000 platform encapsulates MAC addresses between data centers with an IP header in a similar manner as tunneling technologies. Flooding of unknown Ethernet frames and transmission of bridge protocol data units (BPDUs) are suppressed from traveling between data centers for fault isolation purposes. OTV creates an environment where VMware VMotion can operate between data centers because the virtual machines are not aware of the IP encapsulation occurring over the IP WAN links.

Summary

DCE technologies are going through the standardization process to ensure solutions from different vendors are compatible in the future. Cisco and VMware are two of the leaders in the technologies being developed that will allow increased functionality, while also meeting the requirements of cost containment, simplified management, and feature parity compared to traditional switch infrastructures.

Learn More

Learn more about how you can improve productivity, enhance efficiency, and sharpen your competitive edge. Check out the following Global Knowledge courses:

DCNI1-BC – Data Center Networking Infrastructure Boot Camp

DCNI-2 – Cisco Data Center Network Infrastructure 2 v3.0 (Nexus 7000 and 5000)

ICNX5+7 – Implementing and Configuring the Cisco Nexus 5000 and 7000 v2.0

ICNX5 – Implementing and Configuring the Cisco Nexus 5000 v2.0

ICNX7 – Implementing and Configuring the Cisco Nexus 7000

For more information or to register, visit **www.globalknowledge.com** or call **1-800-COURSES** to speak with a sales representative.

Our courses and enhanced, hands-on labs and exercises offer practical skills and tips that you can immediately put to use. Our expert instructors draw upon their experiences to help you understand key concepts and how to apply them to your specific work situation. Choose from our more than 1,200 courses, delivered through Classrooms, e-Learning, and On-site sessions, to meet your IT and business training needs.

About the Author

Dennis Hartmann, CCIE No. 15651, is a Senior Unified Communications and Data Center Instructor at Global Knowledge. He is also the author of the Cisco Press book, *Implementing Cisco Unified Communications Manager, Part 1 (CIPT1)*. Dennis has various certifications, including CCVP, CCSI, CCNP, CCIP, and MCSE. Dennis has worked for various Fortune 500 companies, including AT&T, Sprint, Merrill Lynch, KPMG, and Cabletron Systems. Dennis lives with his wife and children in Hopewell Junction, New York.

Further Information

Global Knowledge Data Center Blogs

http://ciscoblog.globalknowledge.com/tag/data-center/

Cisco Data Center Ethernet (DCE)

http://www.cisco.com/en/US/solutions/collateral/ns340/ns517/ns224/ns783/at_a_glance_c45-460907.pdf

www.cisco.com/go/dce

Cisco Nexus Data Center Switches

www.cisco.com/go/nexus

Cisco NX-OS

http://www.networkworld.com/community/node/38587

Cisco UCS Servers

www.cisco.com/go/ucs

Cisco UCS C Series servers

http://www.cisco.com/en/US/products/ps10493/products_data_sheets_list.html

Fibre Channel over Ethernet (FCoE)

http://www.cisco.com/en/US/solutions/collateral/ns340/ns517/ns224/ns945/ns1060/at_a_glance_c45-578384.pdf

IEEE Data Center Bridging Task Force

http://www.ieee802.org/1/pages/dcbridges.html

Priority Flow Control (PFC)

http://www.ciscosystems.sc/en/US/prod/collateral/switches/ps9441/ps9670/white_paper_c11-542809_ns783_Networking_Solutions_White_Paper.html

Virtual Device Contexts (VDC)

http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9402/ps9512/White_Paper_Tech_Overview_Virtual_Device_Contexts_ps9372_Products_White_Paper.html

Virtual Port Channels (VPC)

http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9402/white_paper_c11-516396.html

http://ciscoblog.globalknowledge.com/tag/data-center/

Virtual Port Channel Configuration

http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9670/configuration_guide_c07-543563.html

