The Virtual Machine Aware SAN

What You Will Learn

Virtualization of the data center, which includes servers, storage, and networks, has addressed some of the challenges related to consolidation, space constraints, demand for high power, and cooling requirements. End-to-end virtualization helps increase efficiency and reduces overall total cost of ownership (TCO). Virtual machines make it possible to run multiple applications and operating systems in a single machine. However, the proliferation of virtual machines introduces new challenges for the SAN, including challenges related to loss of visibility, security, traffic isolation for applications, quality of service (QoS), performance monitoring, and management complexity.

The Cisco® MDS 9000 Family provides VN-Link Storage Services to support the Data Center 3.0 vision by enabling IT departments to dynamically respond to changing business demands, providing fabric scalability and performance, performance monitoring and trending, QoS, Virtual SANs (VSANs) for fault isolation, and mobility for virtual machines.

This document describes the innovative approach introduced by the Cisco MDS 9000 Family to support an end-to-end virtualized SAN environment that is flexible, secure, scalable, and mobile.

Challenges of the Virtualized Environment

Server virtualization technologies enable the consolidation of numerous application servers on a much smaller number of physical servers running a hypervisor, a specialized operating system capable of hosting multiple guest system images. While this solution dramatically reduces the administrative steps required to deploy and administer each individual application server, it is very demanding for the network infrastructure.

Challenges for the network infrastructure include:

- Change of communication patterns from a many-to-one model to a many-to-few (mesh) model
- Unpredictable traffic generated by the dynamic movement of virtual machines across physical servers
- Rapid deployment of virtual machines without complex management procedures and without affecting security

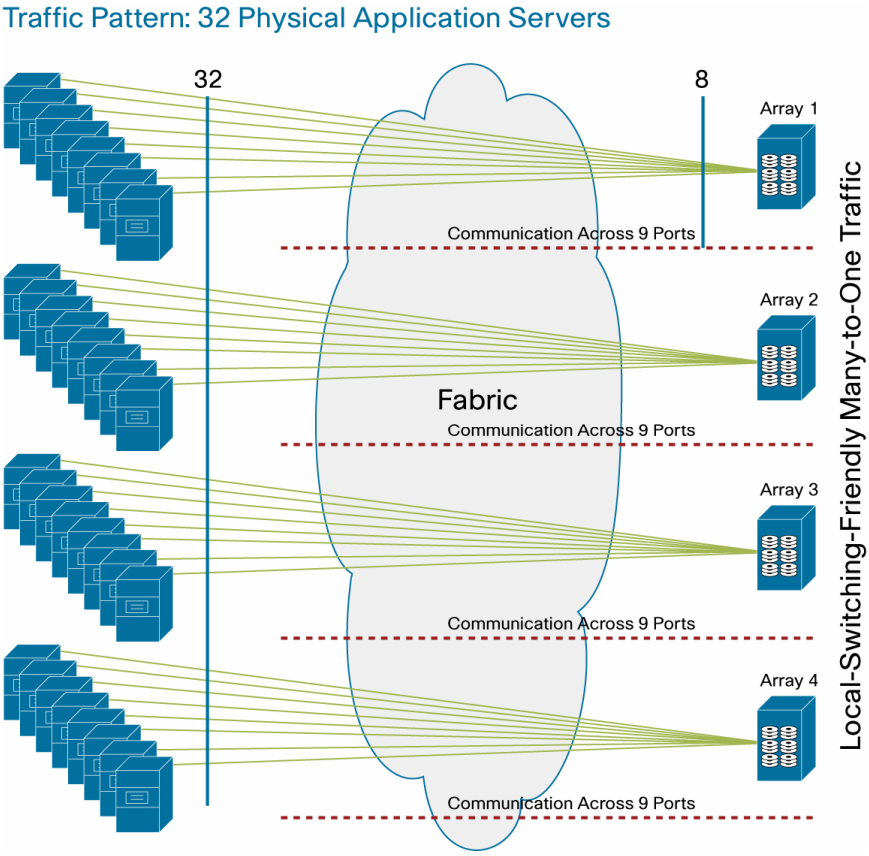
This document analyzes these challenges in detail.

Complex Communication Patterns

The fundamental building block of the virtualized data center is not the individual server, but a cluster of servers, possibly sharing access to a large amount of storage. The most commonly deployed hypervisor typically uses a cluster of 32 servers.

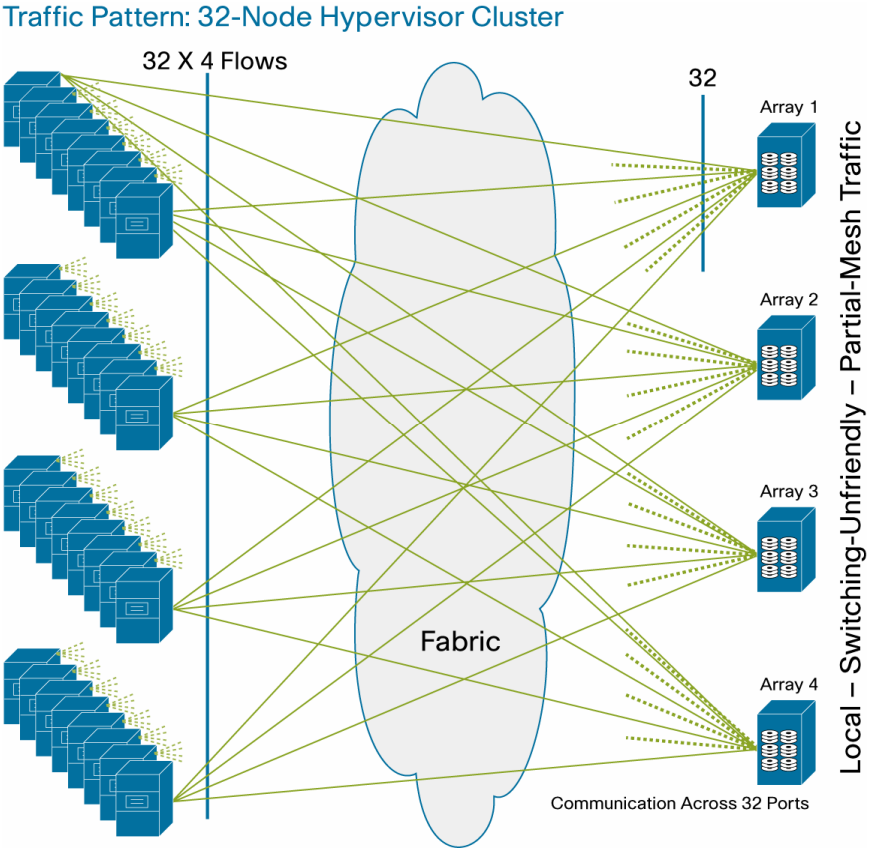
In the traditional configuration, a small group of servers access a given storage port, but in the hypervisor cluster, all the servers share all the storage ports. As a consequence, traffic patterns have evolved from many-to-one configurations (Figure 1) to many-to-few, or mesh, configurations (Figure 2). Note that in Figure 1 each physical server is an application server, while in Figure 2 each physical server hosts a number of virtual application servers.

Figure 1. Traditional Many-to-One Storage Access



The mesh configuration requires a switching infrastructure capable of delivering any traffic pattern with consistent performance.

Figure 2. Mesh Storage Access



In a virtualized environment, each physical server hosts a number of application servers. The total storage traffic, now generated by the hosting physical server, depends on the number of application transactions performed by each virtualized application server. The number of application transactions remains the same as the transactions pass from the discrete physical servers to the consolidated virtual environment, keeping the total amount of storage traffic basically unchanged.

Even when deploying a simple fabric, the resulting design leads to challenging traffic conditions, with many-to-many communication and unpredictable traffic bursts, pushing the capabilities of the switching infrastructure to the limit.

Unpredictable and Dynamic Traffic

The transparent mobility of virtual machine across physical servers is one of the compelling benefits of a server virtualization solution, but it prevents any attempt to lay out the storage network on the basis of static, predefined traffic relationships.

In a full-featured cluster, virtual machine mobility is triggered automatically by the physical server load or by application availability considerations, contributing to the creation of totally random patterns of traffic between all the cluster members and all the associated storage ports.

Virtual machine mobility requires the switching infrastructure to provide even and homogeneous performance across any port in a large deployment. As a consequence, the value of any traffic engineering based on locality, which assumes the static association of the application server and

the storage port, is nullified. In this context, a sound switching fabric architecture shows its potential, and the capabilities built in to the hardware components work at their upper limits.

Rapid Deployment and Management of Virtual Resources

Designed to consolidate hundreds of heterogeneous application servers, an enterprise-class deployment of virtual machines is composed of multiple clusters, necessarily with different security requirements and scope. This context demands the capability to create tight fabric segments and to assign a server to a given segment in very little time and with total flexibility.

A cluster is associated with a specific network segment and administration group. A spare server is usually prewired to the fabric; when a spare server is associated with a cluster, the fabric itself must be able to immediately associate the spare server interface with the network segment to which the destination cluster belongs.

The different user communities require each virtual machine to be managed and protected in the same way as the original physical application server. To meet this requirement, fabric administration must be based on roles, with granular control that matches the capabilities of the virtual machine management infrastructure.

Competitive Solutions

A large Fibre Channel switch can be developed by using a large number of smaller switching elements. The elements are then interconnected to create an ASIC lattice that is similar to a network of discrete switches, but that is assembled as a single large box.

While this approach drastically reduces research and development costs, the resulting device provides greatly differing paths to a given destination port from different source ports. Frames originated and received on the same line card or port group can traverse a single switching element and in most cases experience a better-than-average latency, but frames that traverse more switching elements experience worse-than-average and highly variable delay. For complex traffic patterns, the internal congestion in the ASIC lattice can lead to a dramatic performance jitter and a severe reduction of the actual bandwidth available to applications.

Some vendors in the Fibre Channel market have suggested to their customers that they connect the application servers to switch ports local to the corresponding storage ports. The goal of this approach is to avoid a significant amount of traffic leaving the local switching element and causing unpredictable performance.

However, this design requirement is not applicable to physical servers that are members of a hypervisor's cluster. One reason is that the number of ports involved exceeds the requirements for locality, but a more fundamental reason is that connecting a large number of cluster servers to the same ASIC dramatically reduces the availability of the entire cluster, eliminating the main purpose of deploying the cluster. For similar reasons, and also for effective load sharing, the storage ports must be spread across different line cards or switching ASICs. If the locality requirements are not respected, the physical servers in the cluster experience diverse paths and diverse performance.

Security is another critical factor: the consolidation of different business functions on the same physical infrastructure must provide the same level of isolation as provided by discrete servers and disconnected networks.

Although hypervisor vendors are making a strong effort to achieve the desired level of isolation between virtual machines, this goal is not pursued with the same energy by all storage networking

vendors. Isolation is delegated to an add-on to the zoning infrastructure and functions by limiting the privileges of a given administrator to a subset of the zones in the fabric.

This solution, vendor specific and not based on any standard, does not provide separate fabric services and hardware-based isolation. All data share the same fabric topology and are subjected to the same malicious attacks and to the same service interruptions from configuration errors and protocol violations.

The Cisco MDS 9000 Family Advantage

The Cisco MDS 9000 Family of Fibre Channel switches and directors is designed with the benefit of Cisco's broad view as a networking company experienced in very large networks with complex and unpredictable traffic profiles.

The Cisco MDS 9000 Family was designed to deliver the best performance from any port to any port, making it ready to support the demands of the virtualized data center. In addition, virtualization was built in to Cisco MDS 9000 Family from the start, with the introduction of such innovative features as virtual SANs (VSANs) and role-based management.

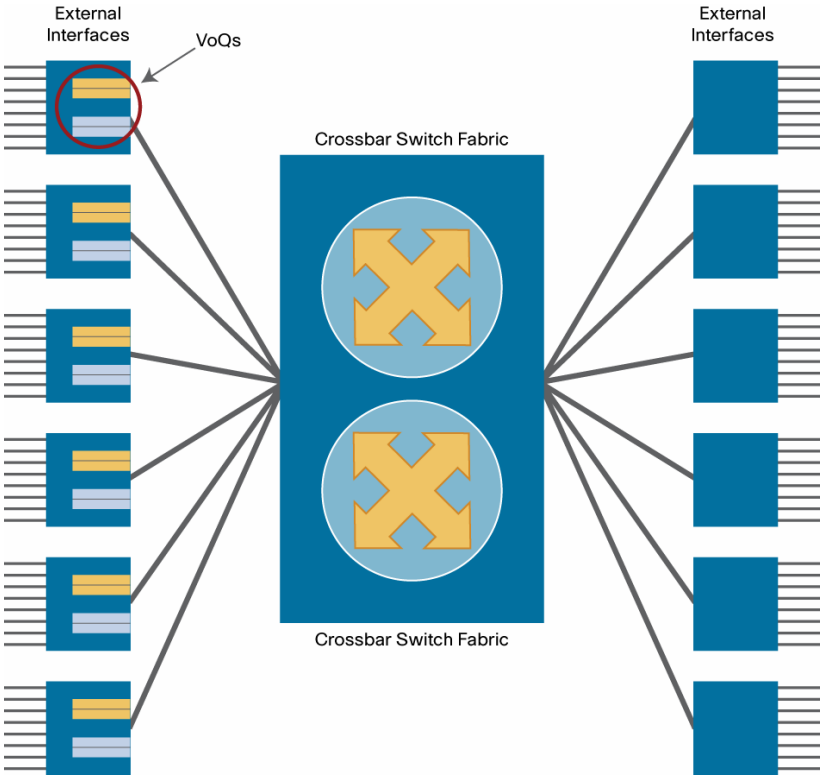
The Cisco MDS 9000 Family provides the following advantages:

- High-performance switch architecture
- Virtual SAN infrastructure to simplify cluster deployments
- Extension of virtual infrastructure management to SANs
- Security in a consolidated environment
- Use of NPIV to provide virtual machine identity in the SAN
- QoS
- Performance monitoring per virtual machine
- Troubleshooting tools
- Virtual machine deployment on blade servers
- F-Port Trunking feature
- Support for large-scale Small Computer System Interface over IP (iSCSI) deployments

High-Performance Switch Architecture

The Cisco MDS 9000 Family switches and directors are based on a centralized crossbar architecture (Figure 3) that is designed to provide the best performance in the most difficult traffic conditions. Virtual output queues (VOQs) at the input of the crossbar help ensure line-rate performance on each port, independent of traffic patterns, by eliminating head-of-line blocking. This architecture provides even and predictable performance for both large and small frames in the presence of many-to-one; many-to-few; and meshed, unpredictable, dynamic traffic patterns.

Figure 3. Cisco MDS 9000 Family Centralized Crossbar Architecture



Virtual SAN Infrastructure to Simplify Cluster Deployment

The virtual SAN (VSAN) technology, an ANSI T11 standard, provides secure hardware-based network segmentation of a single physical SAN fabric or switch.

Each individual VSAN is regulated by an independent set of fabric services (including zone server, name server, domain manager, and Fabric Shortest Path First [FSPF] routing services) in such a way that each VSAN can contain any configuration operation and choice and so is protected with respect to management, configuration, and protocol errors from what happens in a different VSAN.

For instance, zoning is a fabric service running on the top of the VSAN infrastructure: a configuration error or a protocol violation, which might lead to a catastrophic zoning misconfiguration, has no effect outside the VSAN where the problem occurred.

Any interface in an entire Cisco MDS 9000 fabric can be assigned to any VSAN in seconds with a simple configuration command.

Within a virtual machine infrastructure, individual physical servers are grouped into clusters, and a given virtual machine can be easily moved to any server in the cluster. Clusters can be grouped into an upper-level management entity called the “data center”. This management entity will be referred to as the virtual data center in the following part of this paper.

A virtual machine moves from one physical server to another to balance the load across the physical servers or, in case of failure of the hosting physical server, to preserve availability of the virtual machine itself.

To achieve the best level of availability, physical servers that are members of the same cluster should be spread across multiple SAN switches across the fabric or, at least, across multiple line cards, if a single switch is used.

Using the Cisco MDS 9000 Family infrastructure, physical servers belonging to the same cluster, and their associated storage devices, can be deployed in the same VSAN, but at the same time they can be spread across any port of any switch. This configuration maintains a tight isolation, but effectively shares the physical network infrastructure without the need of any preplanning.

Adding or removing a physical server from a given virtual data center or from a given cluster is easily achieved by changing the server's properties in the virtual infrastructure administration interface and reassigning the switch port to a different VSAN. This approach is functionally equivalent to relocating the Fibre Channel cable from one isolated physical fabric to another, but because of the adoption of VSANs, it is totally virtualized. Notice that, since any fabric port can belong to any VSAN, no advance planning is needed to specify which physical server will be associated with which virtual data center or cluster; after the cabling is in place, the assignment can be performed on demand using software.

The new F-Port Trunking feature, in conjunction with adequate hypervisor drivers, provides additional flexibility, with the option to connect multiple virtual machines, located on the same physical server, to multiple VSANs. This feature is described more fully later in this document.

Virtual Infrastructure Management for Virtual Machines, SANs, and Storage

The virtual infrastructure management offers several levels of role-based access control (RBAC): for instance, an administrator can be in charge of one or more virtual data centers, or one or more clusters, or one or more physical servers.

The Cisco MDS 9000 Family management architecture offers an equally sophisticated RBAC infrastructure, enabling administrators to easily assign administrative rights to groups or individuals and to map the scope to a virtual data center or a physical server cluster.

A possible approach to administration is shown in Figure 4, in which all the physical servers in a virtual data center are assigned to the same VSAN, and administration is structured per VSAN and virtual data center.

Another option would be to assign to the same VSAN all the physical servers belonging to the same cluster, structuring administration per VSAN and cluster.

Figure 4. Administration Based on VSAN and Virtual Data Center

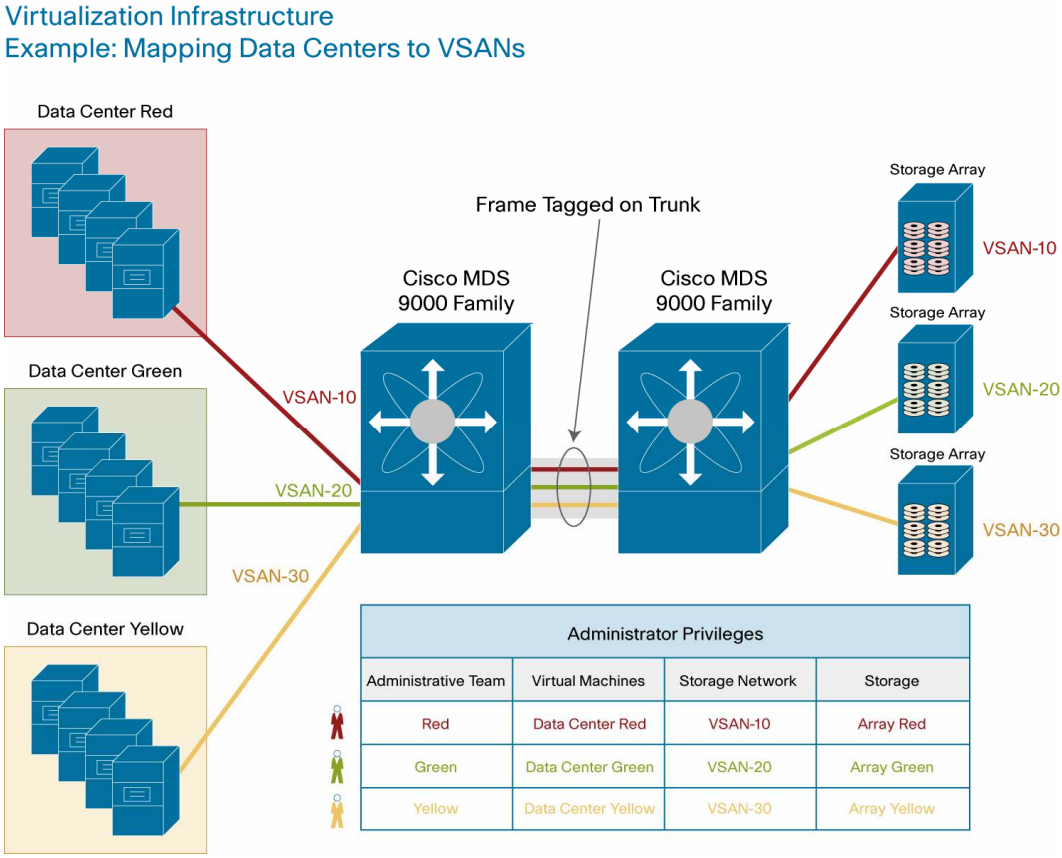


Figure 4 shows the operation of a single physical data center, structured in three virtual data centers: the Red, the Green and the Yellow. Administrative team Red, for example, is authorized to configure resources in virtual data center Red and in VSAN-10 and to set up storage volumes in storage array Red. Even if the physical network resources, the switches and the links between switches, are consolidated, administrative team Red and the virtual resources are fully isolated.

Security in a Consolidated Environment

Consolidation of different business functions on the same physical infrastructure must help ensure the same level of security and isolation as provided by a solution based on discrete servers or by unconnected networks.

While the foundation of fabric security is embedded in the VSAN-capable hardware, the holistic approach to security provided by the Cisco MDS 9000 Family provides peace of mind, even in a highly consolidated environment.

The Cisco MDS 9000 Family offers these main security features:

- Management performed through secure protocols such as HTTPS, Simple Network Management Protocol Version 3 (SNMPv3), Secure Shell (SSH), and Secure File Transfer Protocol (SFTP); both the command-line interface (CLI) and the Cisco Fabric Manager management tools use secure protocols and are fully integrated into the RBAC infrastructure

- Authentication, Authorization, and Accounting (AAA), which can be plugged into the enterprise infrastructure using standard protocols (RADIUS and TACACS+); using an external AAA server such as the Cisco Secure Access Control Server (ACS), the user authentication exchange also provides the role information to support RBAC
- Configuration management, distribution, and consistency analysis
- Fabric access security (fabric binding and port security), Port security can address the physical hypervisor server, or it can address the individual virtual machine when used in conjunction with the NPIV technology, as described in the following.
- Support for the Fibre Channel Security Protocol (FC-SP; a standard mostly developed by Cisco)
- Enhanced support for zoning (logical unit number [LUN] zoning and read-only zone)
- Integrated, high-performance network services for protecting data in flight (SAN extensions over IP, namely Fibre Channel over IP [FCIP] and iSCSI, can use IP Security [IPsec]) and at rest (Cisco Storage Media Encryption [SME])
- Hardware enforced, standards-based VSAN segmentation to isolate data traffic and secure the fabric services and protocols

The Cisco MDS 9000 Family provides all the tools needed to implement the required level of security in an enterprise-class virtual machine deployment.

Use of NPIV to Provide Virtual Machine Identity in the SAN

N-Port ID Virtualization (NPIV) is a Fibre Channel protocol feature that allows individual virtual machines to assume a full identity on the SAN. Using NPIV, the hypervisor can create a virtual HBA for each virtual machine. The virtual HBA is identified by the Fibre Channel port World Wide Name (pWWN) in the same way of a physical HBA, and this identity is an attribute of the given virtual machine that is preserved even when the virtual machine moves across physical servers. The Cisco MDS 9000 Family provides complete and scalable support for NPIV virtual machines. Each feature available for a physical server is available for an individual virtual machine addressed by NPIV. Examples are such basic Fibre Channel services as zoning and more advanced features such as troubleshooting tools, QoS, and performance monitoring.

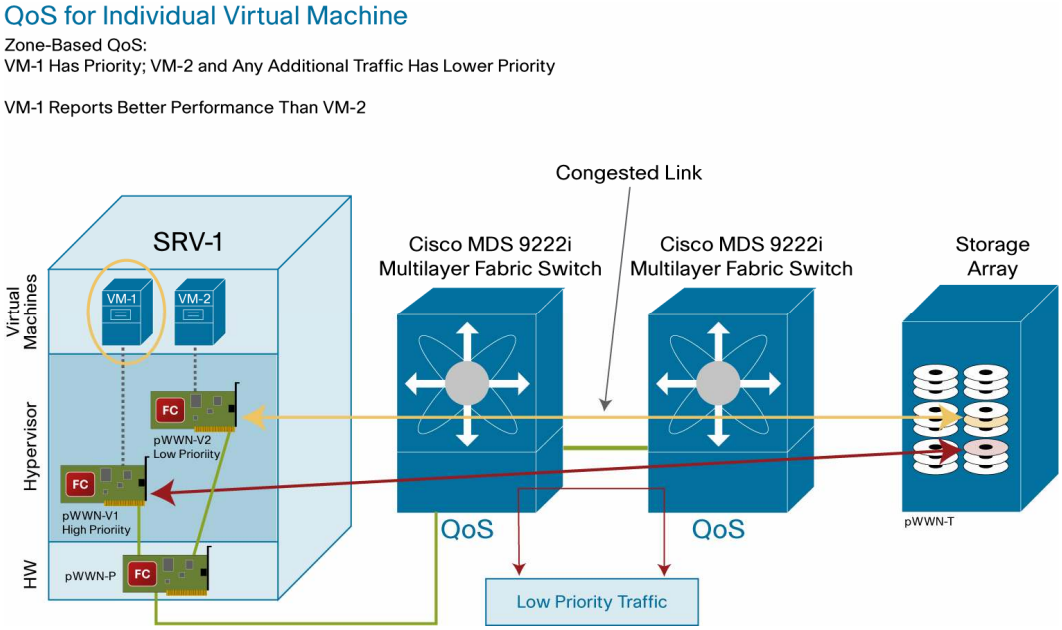
The Cisco MDS 9000 Family is ready to support any enterprise-class virtual machine deployment using the current hypervisor's NPIV features, and it is ready to enable the additional NPIV-based features that hypervisor vendors will make available in the future.

QoS

The Cisco MDS 9000 Family provides numerous options for QoS management, including VSAN-based and zone-based QoS.

QoS per VSAN can be useful when an entire virtual data center or an entire cluster has been mapped to a VSAN. Zone-based QoS can be used to enhance the performance of a specific physical server or of group of physical servers, and when used in combination with NPIV, it has been proved to be effective even for individual virtual machines (Figure 5).

Figure 5. QoS Level Can Be Assigned to the Individual NPIV Virtual HBA



Troubleshooting Tools

The Cisco MDS 9000 Family provides a set of troubleshooting tools to isolate connectivity problems. Fibre Channel ping (FC-ping), based on the Fibre Channel echo packets, is used to verify that a device is connected to the fabric, and FC-traceroute allows tracing of the network route between two devices. These tools can be used in the same way to troubleshoot physical servers, physical servers running a hypervisor, and NPIV-capable virtual machines.

Other troubleshooting tools useful in a virtualized environment include local and remote built-in protocol analysis and mirroring of data to a network analyzer port.

Performance Monitoring

Cisco Fabric Manager is the GUI-based management infrastructure for the Cisco MDS 9000 Family SAN. Cisco Fabric Manager provides a full set of tools for fabric configuration and performance monitoring.

The same performance monitoring capabilities available for the physical devices are available for collecting statistical data for the individual NPIV-enabled virtual machines, providing a single monitoring point across the entire end-to-end storage infrastructure (Figure 6).

Figure 6. Performance Monitoring of an Individual Virtual Machine Using NPIV

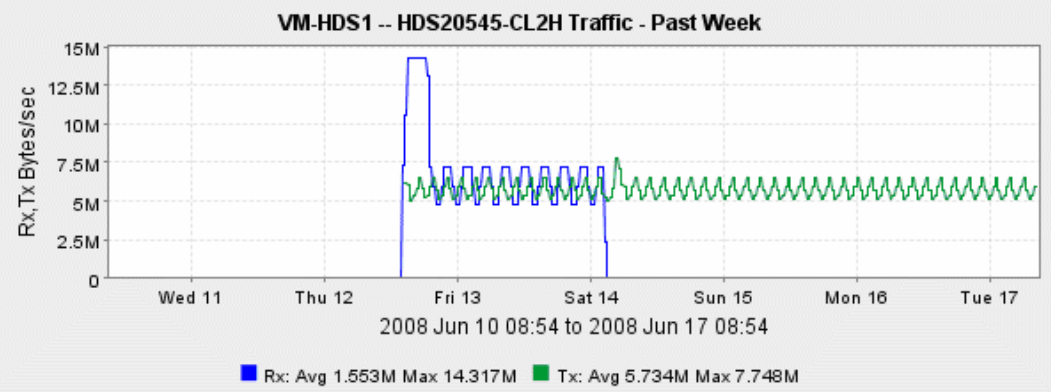


Figure 6 shows trend statistics collected for an individual NPIV-enabled virtual machine, with detailed information about the read and write operations.

Virtual Machines on Blade Servers

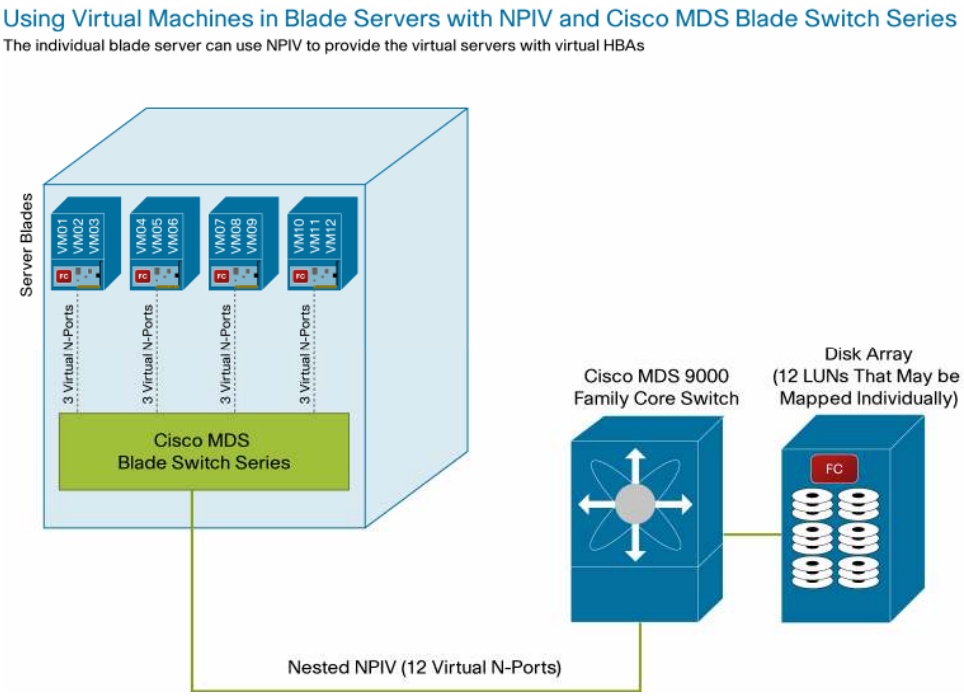
Full support for NPIV to provide virtual machine identity in the SAN is provided even for the Cisco MDS 9000 Family switches available in the blade server form factor.

The Cisco N-Port Virtualizer (NPV) feature facilitates deployment of blade switches in large-scale SANs and also their operation with any vendor's SAN core switch, with no additional hardware or software required. When NPV is engaged, the blade switch is seen as a host bus adaptor (HBA) aggregator, not as a switch.

All server blades, using their own physical identity, perform a fabric login to the blade switch, which in turn proxies each server blade, using the physical identity of each blade as an NPIV virtual identity.

If the server blade is running a hypervisor that is using NPIV, the individual virtual HBA identity of each virtual machine is preserved by the Cisco MDS Blade Switch Series and proxied to the core (Figure 7). This function is also known as nested NPIV, and it is essential to offering the same functions to virtual machines instantiated on either rack-mounted servers or blade servers.

Figure 7. Nested NPIV with Cisco MDS Blade Switch Series



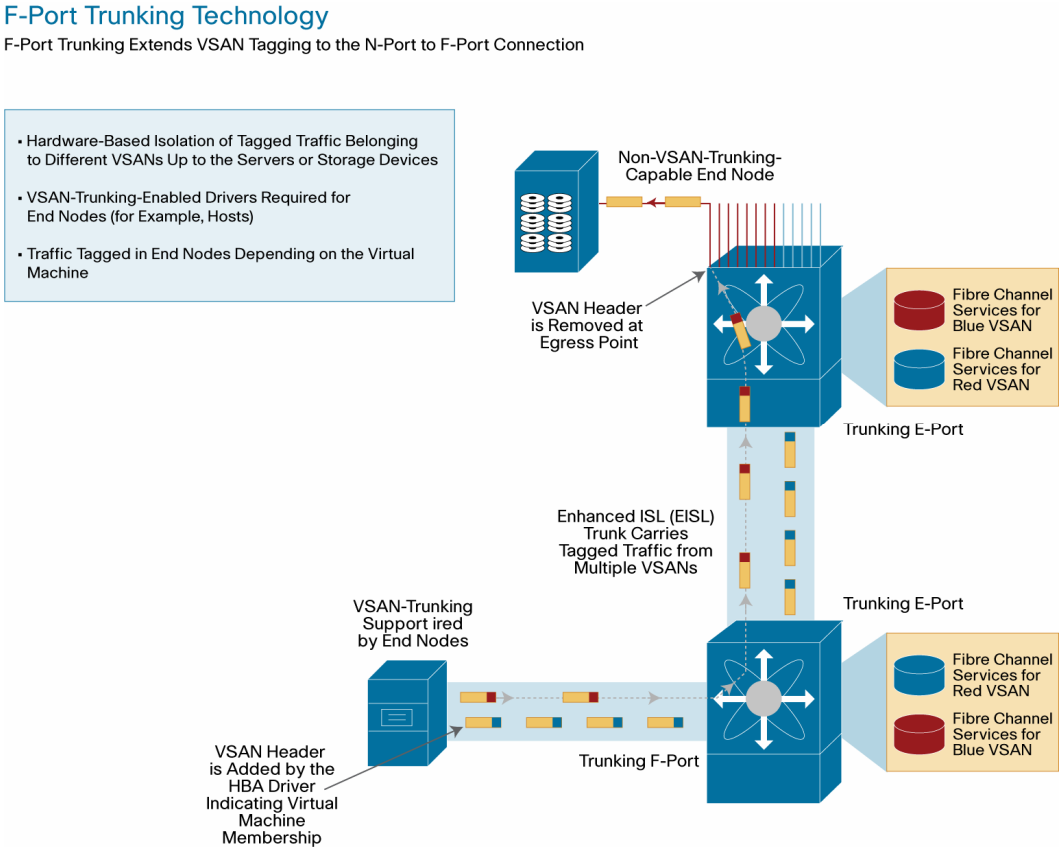
Cisco recently introduced the FlexAttach feature for blade chassis. The FlexAttach feature allows servers to be easily moved, added, or changed without reconfiguration of SAN switches or storage arrays. FlexAttach helps virtualize the SAN identity for a server, enabling it to retain that identity even when moved or replaced within or across the blade server chassis, giving server administrators more flexibility with no need for SAN reconfiguration.

The Cisco MDS Blade Switch Series, working in NPV mode, can bundle multiple connections to the core switch by enabling the F-Port PortChanneling feature. This feature associates multiple ports to create one logical link, to aggregate the bandwidth and increase the availability and resiliency of the fabric attachment. F-Port PortChanneling greatly enhances the overall robustness of the virtual machine deployment.

F-Port Trunking

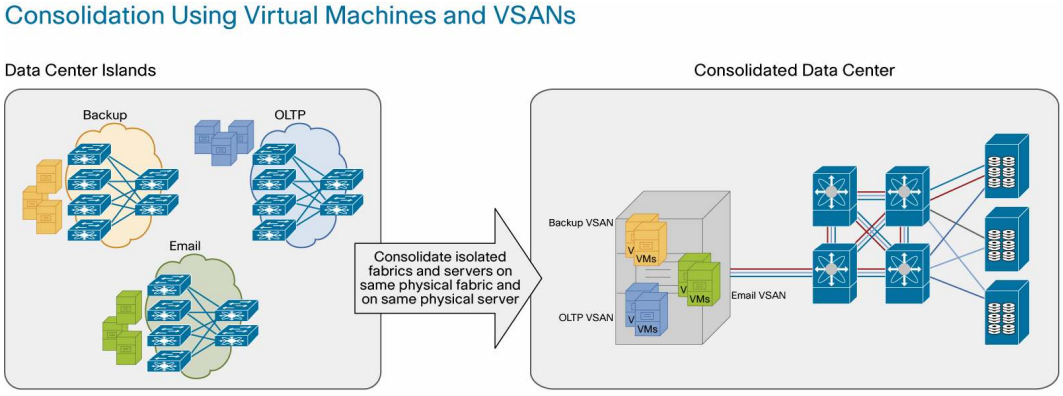
The F-Port Trunking feature allows virtual machines to belong to different virtual SANs while sharing the same physical server HBA (Figure 8). This technology is fully compliant with the relevant ANSI T11 Fibre Channel standards and requires analogous support in the HBA driver.

Figure 8. F-Port Trunking Technology



The F-Port Trunking feature allows a step forward in server and storage consolidation. The same physical server and the same HBA can host virtual machines belonging to different user groups, while maintaining a tight level of separation between storage resources (Figure 9).

Figure 9. Consolidation Using Virtual Machines and VSANs



	Attribute	
More	Number of SAN Switches and Servers	Fewer
No	Shared Disk or Tape	Yes
No	Shared Disaster Recovery Facilities	Yes
Complex	Server and SAN Management	Simple
High	Overall TCO	Low

Notice that, because the Fibre Channel services are instantiated per VSAN, this solution allows each user group to manage its own zones and zone sets.

Using F-Port Trunking in combination with NPIV allows creation of both independent zone sets per VSAN and independent zones per individual virtual machine within the given VSAN.

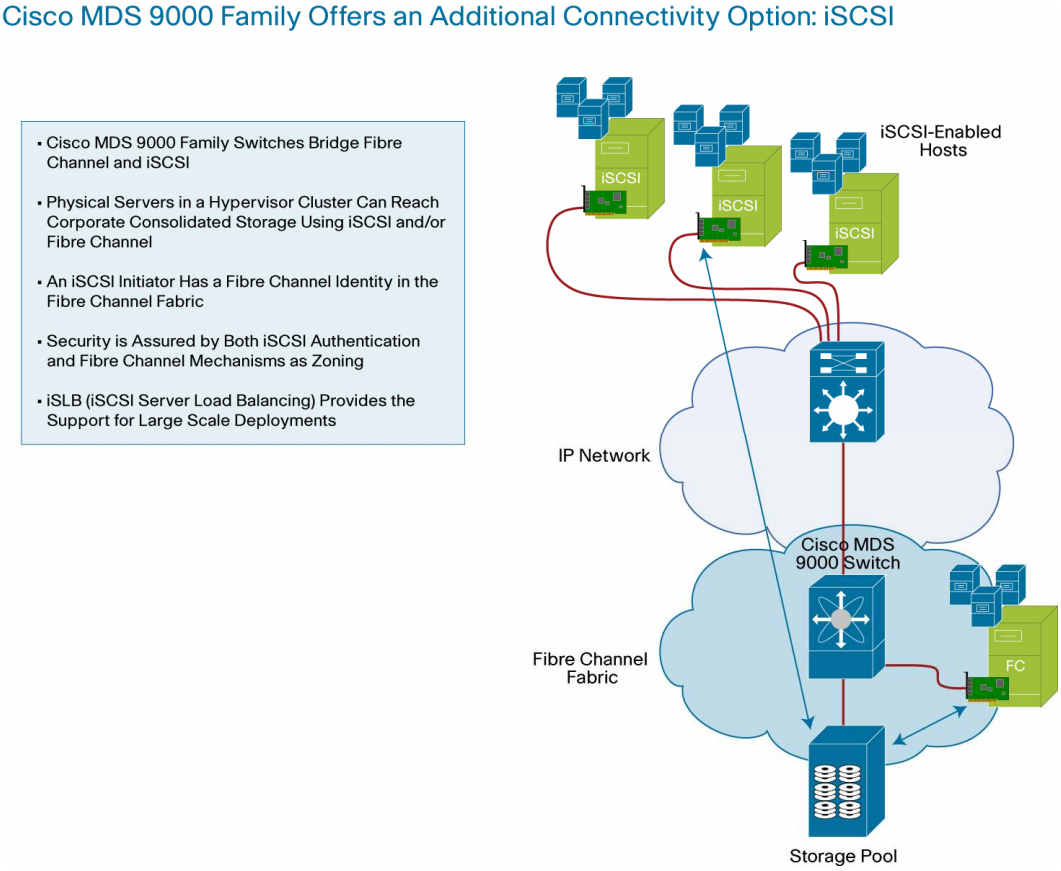
Support for Large-Scale iSCSI Deployment

The major hypervisor vendors offer iSCSI as a connectivity option to provide virtual machines with block-level access to storage volumes without the need to deploy high-performance Fibre Channel hardware.

iSCSI has been proven capable of supporting enterprise-class applications, assuming that the specific solution can reach the desired level of performance and scalability and that management is simple enough not to negatively affect TCO.

The Cisco MDS 9000 Family offers a fully integrated Fibre Channel to iSCSI gateway solution that allows iSCSI clients to access the consolidated Fibre Channel storage. The Cisco MDS 9000 Family iSCSI implementation provides all the capabilities available to a Fibre Channel initiator (including VSAN, advanced security, and zoning) to the iSCSI initiator, simplifying migration and enabling hybrid solutions (Figure 10).

Figure 10. Deployment of Virtual Machines Using iSCSI



To streamline the deployment of a large number of initiators, as is needed in enterprise-class virtual machine deployments, the Cisco MDS 9000 Family provides a unique feature called iSCSI Server Load Balancing (iSLB).

iSLB combines three main features:

- The initiator-oriented and wizard-based GUI simplifies the deployment of a large number (hundreds) of initiators.
- The iSLB portal is highly available. By configuring a single iSCSI portal, the initiators are automatically distributed across a large number of Ethernet interfaces, across multiple line cards, and across multiple Cisco MDS 9000 Family switches and directors.
- Automatic distribution of the iSCSI configuration across all the switches in the fabric provides iSCSI access, offers a single configuration point, and prevents configuration inconsistencies across the fabric.

These features make the Cisco MDS 9000 Family the ideal choice for large-scale iSCSI, enterprise-class, virtual machine deployments.

Conclusion

The increasing popularity of enterprise class virtual machines deployments has introduced new challenges to storage networking.

Since most deployments are organized into relatively large clusters, sharing a number of storage devices, the typical Fibre Channel traffic profile changes from many-to-one (for example, 10:1) to many-to-few, or meshed (for example, 32:8). Also, virtual machine mobility is randomizing the profile of the traffic between physical server and storage devices.

The resulting complex traffic distribution is pushing the limits of the switching infrastructure. SAN design practices based on server and storage locality cannot be applied any more, and the fabric itself must offer best performance under any unpredictable traffic distribution.

Simple switching architectures and switches based on a network of small switching elements have given way to more sophisticated architectures such as the Cisco MDS 9000 Family of Fibre Channel switches and directors, designed from the start to handle unpredictable, many-to-many network traffic.

A large physical data center, hosting multiple user communities controlling one or more hypervisor clusters each, has stringent requirements for deployment of physical servers, network segmentation, fault isolation, and security.

In an enterprise class deployment of virtual machines, the amount of time required to connect a spare physical server to the appropriate network segment and get it ready to join a cluster must be minimal. With the Fibre Channel standard VSAN technology implemented by the Cisco MDS 9000 Family, the server is prewired anywhere to a fabric interface, and assignment of the interface to the appropriate network segment and cluster is performed in seconds through a configuration command.

At the same time, the VSAN infrastructure provides the hardware foundation for tight security and fault isolation, and the same approach is used for Fibre Channel and for iSCSI. Competing technologies, based on limiting administrative access to Fibre Channel zoning, can barely protect against macroscopic human errors, and they do not provide any true network segmentation or dynamic management of network segmentation.

The Cisco MDS 9000 Family provides a complete toolset for network management, configuration control, troubleshooting, and traffic engineering. The RBAC infrastructure easily maps to the powerful functions offered by the enterprise-class tools for virtual machine management. Management and troubleshooting support extends to the hypervisor running in blade servers and to virtual machines using NPIV-based virtual HBAs.

For More Information

For more information please visit http://www.cisco.com/go/storage.

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.

Printed in USA C11-494982-01 09/08