Large SAN Design Best Practices using MDS 9710

Introduction

As storage area networks (SANs) continue to grow in size, many factors need to be considered to help scale and manage them. This paper focuses on large SAN deployments within a data center, and provides best practices and design considerations when designing a large physical fabric. It does not address networks implementing Inter-VSAN Routing (IVR), FC or FCIP based SAN extension, or intelligent fabric applications (for example, Data Mobility Manager, or IO Acceleration (IOA)).

Design Parameters

In SAN environments there are many design criteria that needs to be addressed, such as how many servers access a shared storage frame, network topology, fabric scaling, etc. This document focuses on the following design parameters:

- 1000 or more end devices (servers, storage, and tape devices)
- Majority of end devices with connection speeds of 8G and 16G
- Identical dual physical fabrics (Fabric-A, Fabric-B)

Cisco MDS 9700 Multilayer Director

The Cisco MDS 9710 is the newest generation director-class multilayer series switch. It can support up to 384 line-rate 16-Gbps Fibre Channel or 10G FCoE ports. The Cisco MDS 9710 comes with dual supervisor modules and with six fabric modules provides up to 24-Tbps chassis throughput.

The Cisco MDS 9700 48-Port 16-Gbps Fibre Channel switching module delivers line rate non-blocking 16-Gbps FC performance to enable scalability in the virtualized data centers. Line-rate 16-Gbps performance provides high bandwidth throughput for consolidation of workloads from thousands of virtual machines while reducing SAN components, providing scaling with future SAN growth at the same time. These line card modules are hot-swappable and continue to provide all previous Cisco MDS features like predictable performance, high availability, advanced traffic management capabilities, Integrated VSANS, high-performance ISLs, fault detection, isolation of errored packets and sophisticated diagnostics. This module offers new hardware based slow drain, real time power consumption reporting and improved diagnostics capabilities.

Part Number	Product Description
MDS 9700 Component	
DS-C9710	MDS 9710 Chassis, No Power Supplies, Fans Included
DS-X97-SF1-K9	MDS 9700 Series Supervisor-1
DS-X9710-FAB1	MDS 9710 Crossbar Switching Fabric-1 Module
DS-X9448-768K9	48-Port 16-Gbps Fibre Channel Switching Module
Optional Licensed Software	
M97ENTK9	Enterprise package license for 1 MDS9700 switch
DCNM-SAN-M97-K9	DCNM for SAN License for MDS 9700

Table 1: Cisco Product IDs and MDS 9700 component details

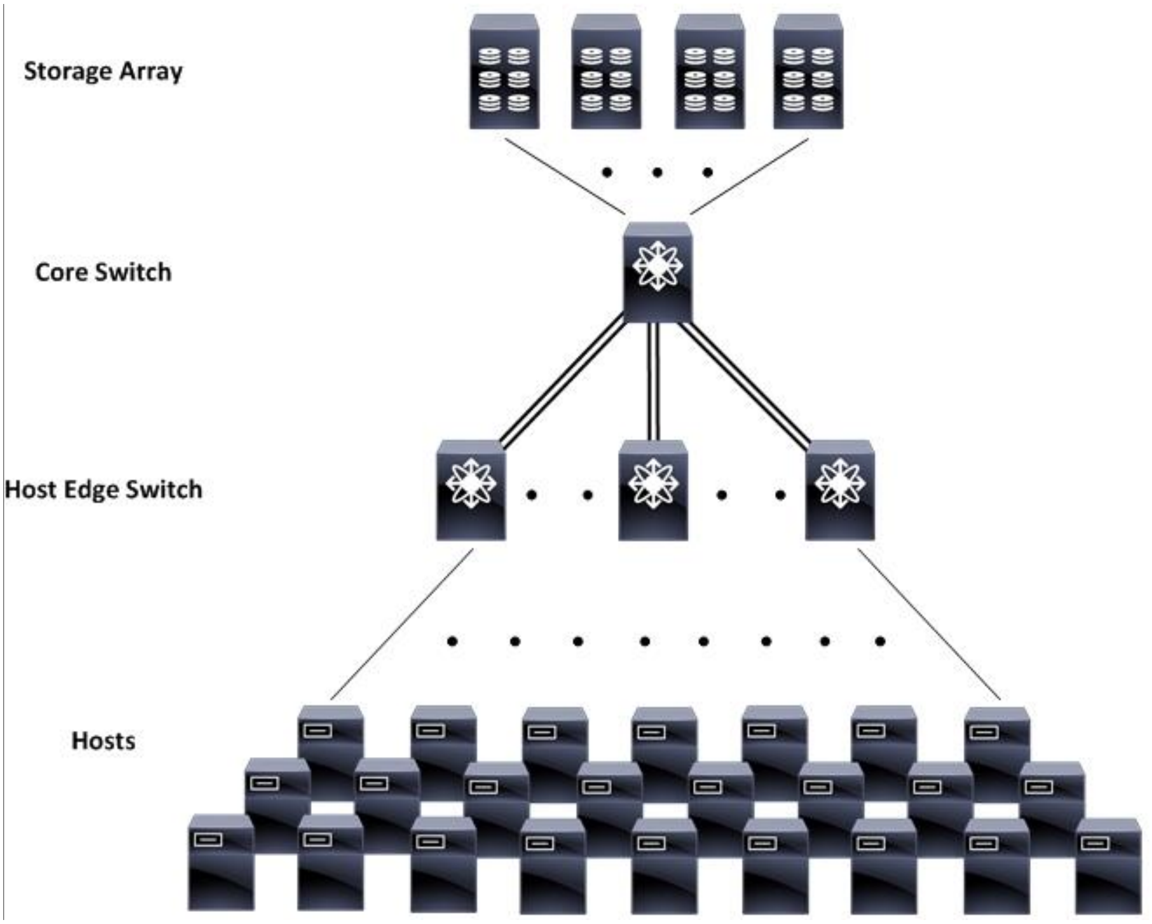
SAN Topology Considerations

It is common practice in SAN environments to build two separate, redundant physical fabrics (Fabric A and Fabric B) in case a single physical fabric fails. For this document, we are showing a single fabric in the topology diagrams however customers would deploy two identical fabrics for redundancy. When designing for large networks, most environments will fall into two types of topologies within a physical fabric:

1. Two-tier: Core-edge design
2. Three-tier: Edge-core-edge design

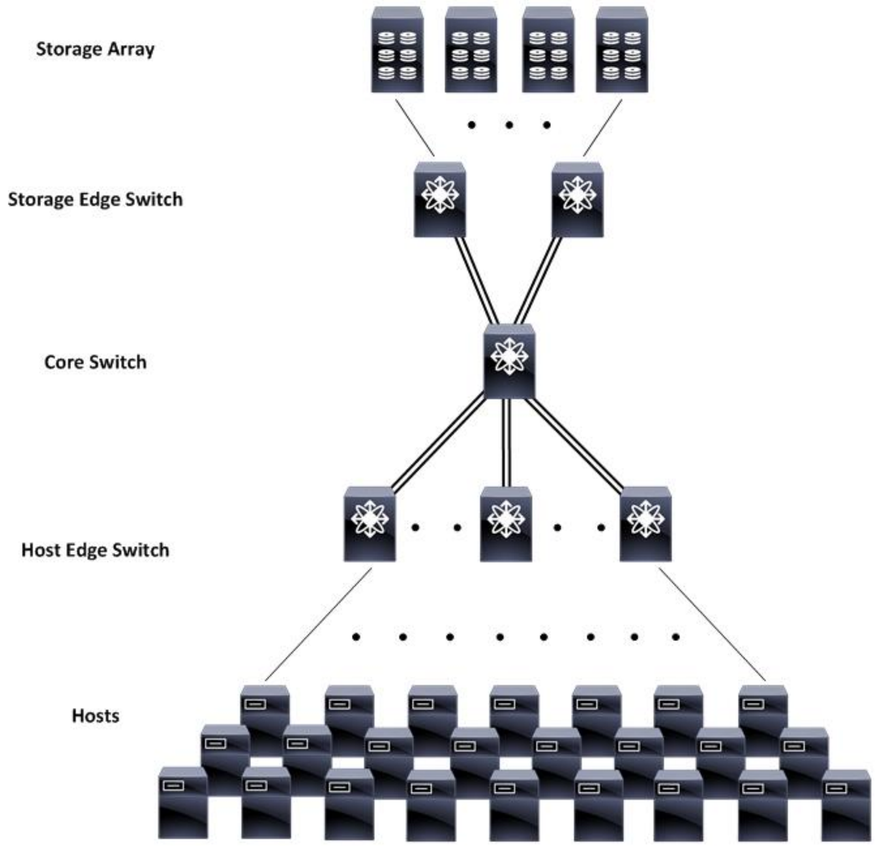
Within the two-tier design, servers connect to the edge switches, and storage devices connect to one or more core switches (Figure 1). This allows the core switch to provide storage services to one or more edge switches, thus servicing more servers in the fabric.

Figure 1: Sample Core-Edge Design



In environments where future growth of the network has the number of storage devices exceeding the number of ports available at the core switch, a three-tier design may be ideal (Figure 2). This type of topology still uses a set of edge switches for server connectivity, but adds another set of edge switches for storage devices. Both sets of edge switches connect into a core switch via ISLs.

Figure 2: Sample Edge-Core-Edge Design



Network Considerations

When designing a large Cisco MDS storage area network, the following should be taken into consideration:

- Fan-in, Fan-out and Oversubscription ratio
- VSANs
- Port Channels
- ISLs
- Effects of fabric logins
- Type of zones
- Smart Zoning

Fan-in, Fan-out and Oversubscription ratios:

To efficiently and optimally use resources and to save deployment and reduce management costs, SANs are designed to share Array ports, ISL and line card bandwidth. The terms used to describe this sharing include Fan-in ratio, Fan-out ratio and Oversubscription ratio. Which term is used depends on the point of reference being described. Generally speaking, Fan-in ratio is calculated as the ratio of host port bandwidth to storage array port bandwidth and Fan-out ratio is calculated as the ratio of storage array port bandwidth to host port bandwidth. Oversubscription is a networking term that is generally defined as the overall bandwidth ratio between host and storage array ports. See Figure 3 below for more details.

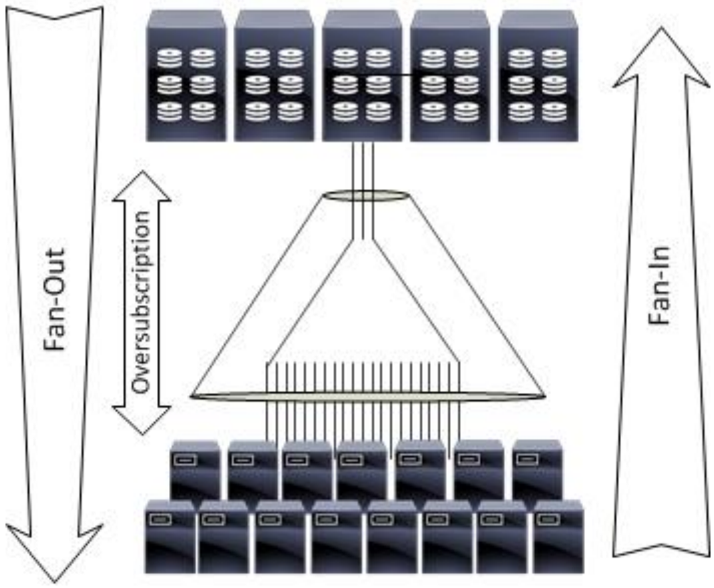


Figure 3: Fan-In; Fan-Out and Oversubscription ratios

Virtual SANs (VSANs)

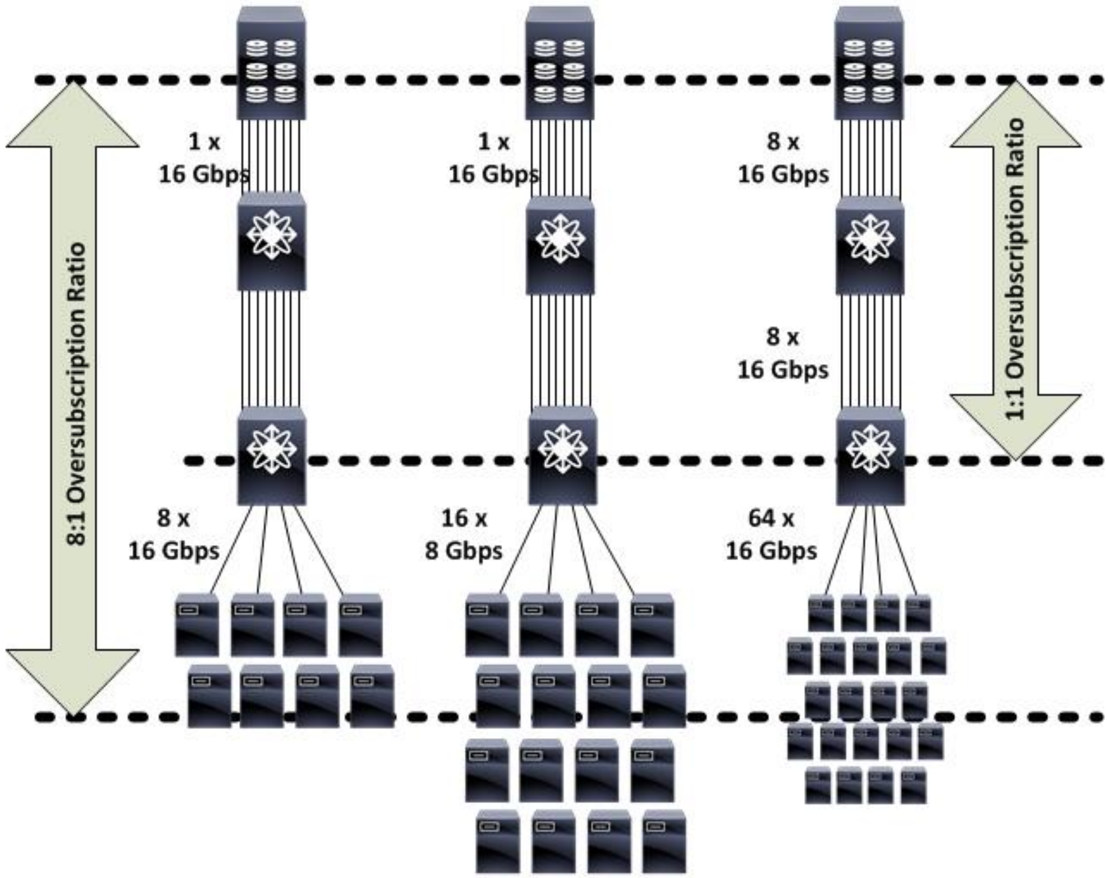
Cisco MDS switches offer VSAN technology, which is a simple and secure way to consolidate many SAN islands into a single physical fabric. Separate fabric services (i.e. per VSAN zoning, Name Services, Domains and separate role base management) are provided for each VSAN, providing separation of both the control plane and the data plane.

There are multiple use cases for VSANs, such as creating a VSAN for each type of operating system (i.e. VSAN for Windows or HP-UX), or utilizing them on the basis of business functions (i.e. a VSAN for development, for production, or for a lab). VSAN 1 is created on the Cisco MDS switch by default and cannot be deleted. As a best practice, VSAN 1 should be used as a staging area for un-provisioned devices and other VSAN(s) should be created for the production environment(s). With each VSAN having its own zones and zone-sets, Cisco MDS switches enable secure, scalable and robust networks.

Inter-Switch Links (ISLs)

Inter-Switch Link (ISL) is the term to describe the connection between fibre channel switches. The number of ISLs required between Cisco MDS switches will depend on the desired end-to-end oversubscription ratio. The storage port oversubscription ratio from a single storage port to multiple servers can be used to help determine the number of ISLs needed for each edge-to-core connection. In the diagram below (Figure 4), there are 3 examples of different storage, server and ISL combinations, all with the same oversubscription ratio of 8:1. The first example has one 16 Gbps storage port with eight 16 Gbps server ports going over one 16 Gbps ISL. The second example has one 16 Gbps storage port with sixteen 8 Gbps server ports going over one 16 Gbps ISL. The third example has eight 16 Gbps storage port with sixty-four 16 Gbps server ports going over eight 16 Gbps ISLs . Maintaining a 1:1 ratio of storage bandwidth to ISL bandwidth is recommended for SAN design. Additional ISL bandwidth can be added with additional ISLs to provide greater availability in case of link failure.

Figure 4: Number of ISLs Needed to Maintain Oversubscription Ratio



PortChannels:

PortChannels refer to the aggregation of multiple physical interfaces into one logical interface to provide higher aggregated bandwidth, load balancing, and link redundancy while providing fabric stability in event of member failure. PortChannels can connect to interfaces across different switching modules, so a failure of a switching module cannot bring down the PortChannel link.

A PortChannel has the following functionality:

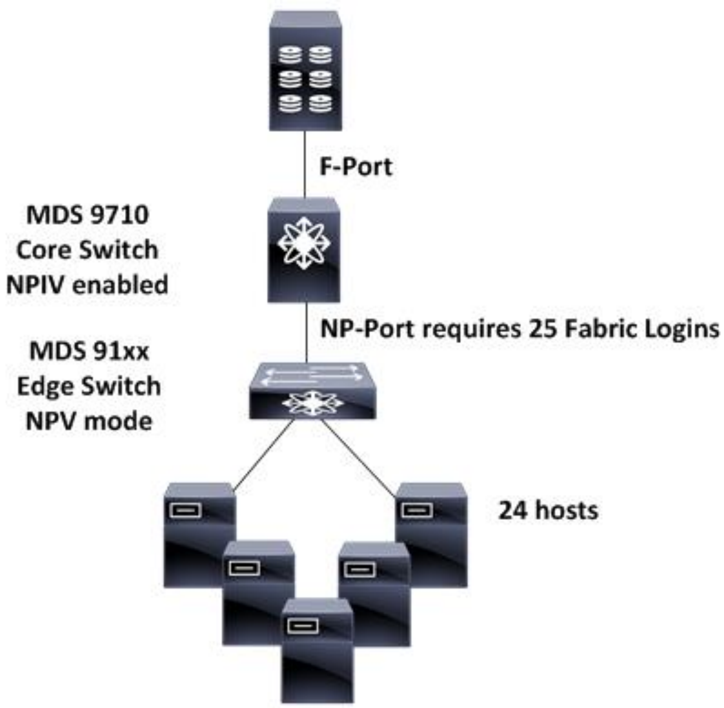
- Provides a single logical point-to-point connection between switches
- Provides a single VSAN ISL (E port) or trunking of multiple VSANs over an EISL (TE port). EISL ports are only between Cisco switches and carries traffic for multiple VSANs compared to ISL.
- Increases the aggregate bandwidth on an ISL by distributing traffic among all functional links in the channel. PortChannels may contain up to 16 physical links and may span multiple modules for added high availability. Multiple Port channels can be used if more than 16 ISLs are required between switches.
- Load balances across multiple links and maintains optimum bandwidth utilization. Load balancing is based on per VSAN configuration (Source ID / Destination ID (SID/DID) or Source ID / Destination ID / Exchange ID (SID/DID/OXID)).
- Provides high availability on an ISL. If one link fails, traffic is redistributed to the remaining links. If a link goes down in a PortChannel, the upper protocol is not aware of it. To the upper protocol, the link is still there, although the bandwidth is diminished. The routing tables are not affected by link failure.

Fabric Logins

The number of actual physical ports in the fabric is larger than the number of end devices (server, storage, and tape ports) in the physical fabric. The Cisco MDS Family supports up to 10,000 fabric logins in a physical fabric, independent of the number of VSANs in the network. Typically when designing a SAN, the number of end devices determines the number of fabric logins. The increase in blade server deployments and the consolidation of servers due to server virtualization technologies, will affect the design of the network. With features such as N_Port ID

Virtualization (NPIV) and Cisco N_Port Virtualization (NPV), the number of fabric logins has further increased (Figure 5). The proliferation of NPIV-capable end devices such as host bus adaptors (HBAs) and Cisco NPV-mode switches makes the number of fabric logins on a per-port, per-line-card, per-switch, and per-physical-fabric basis a critical consideration. These fabric login limits will determine the design of the current SAN, as well as future growth. The total number of hosts and NPV switches will determine the fabric logins required on the core switch.

Figure 5: Cisco NPV-Enabled Switches and Fabric Logins



Note: Prior to NPIV and NPV, a single port had a maximum of one fabric login. With NPIV and Cisco NPV-enabled switches, a single port can now support multiple fabric logins.

Zones

Within each VSAN, there is only one active zoneset that contains one or more zones. Each zone consists of one or more members to allow for communication between the members. The Cisco MDS SAN-OS and NX-OS provide multiple ways to identify zone members, but the commonly used ones are:

- pwwn: Port World-Wide Name of the device (most commonly used)
- device-alias: An easy-to-read name associated with a single device's Port World-Wide Name (pwwn)

Depending on the requirements of the environment, choosing the type of zone members is a matter of preference. A recommended best practice is to create a "device-alias" for end devices when managing the network. The device-alias provides an easy-to-read name for a particular end device. For example, a storage array with a pwwn 50:06:04:82:bf:d0:54:52 can be given a device-alias name of Tier1-arrayX-ID542-Port2. In addition, with device-alias, when the actual device moves from one VSAN (VSAN 10) to a new VSAN (VSAN 20) in the same physical fabric, the device-alias name will follow that device. So there is no need to re-enter the device-alias for each port of the moved array in the new VSAN.

Note: As a best practice for large SAN deployments, it is recommended to have more zones with two-member zones versus a single zone with three or more members. This is not a concern in smaller environments.

Smart Zoning:

Smart zoning supports zoning multiple devices in a single zone by reducing the number of zoning entries that needs to be programmed. This means that multiple member zones consisting of multiple initiators and multiple zones can be zoned together without increasing the size of the zone set. Smart zoning can be enabled at the zone level, zone set level, zone member, or at the VSAN level.

Sample Use Case Deployments

Below are two sample deployments of large-scale Cisco MDS fabrics, one with more than 1000 devices in the fabric.

Sample Deployment 1

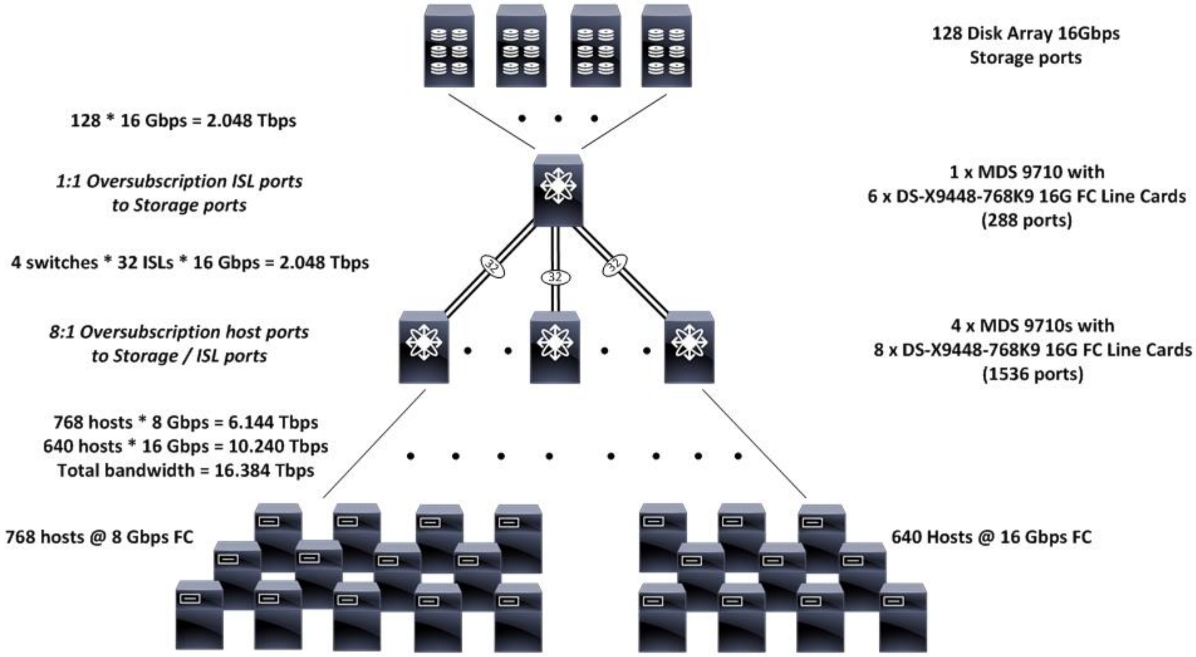


Figure 6: Use Case 1 Topology with mix of 8-Gbps and 16-Gbps hosts connected to 16-Gbps storage ports.

Note: This deployment allows scaling to over 1500 devices in a single fabric. The actual production environment has 128 storage ports running at 16-Gbps and roughly 1400 host ports. The environment requires a minimum of 8:1 oversubscription within the network, which requires each host edge switch to have a 2-Tbps of ISL bandwidth. Storage ports will not grow quite as rapidly and the core switch has room to grow to add more host edge switches.

In this environment, the following was used in managing the network:

- Total of four VSANs created
  - VSAN 1 used for staging new SAN devices
  - VSAN 100 for Development SAN
  - VSAN 200 for Lab SAN
  - VSAN 300 for Production SAN
- Use of TACACS+ for authorization and authentication of MDS switches
- Use of Role-Based Access Control (RBAC) to create separate administrative roles for VSANs
- Use of device-aliases logical device identification
- Use of two member zones with device-alias

Sample Deployment 2

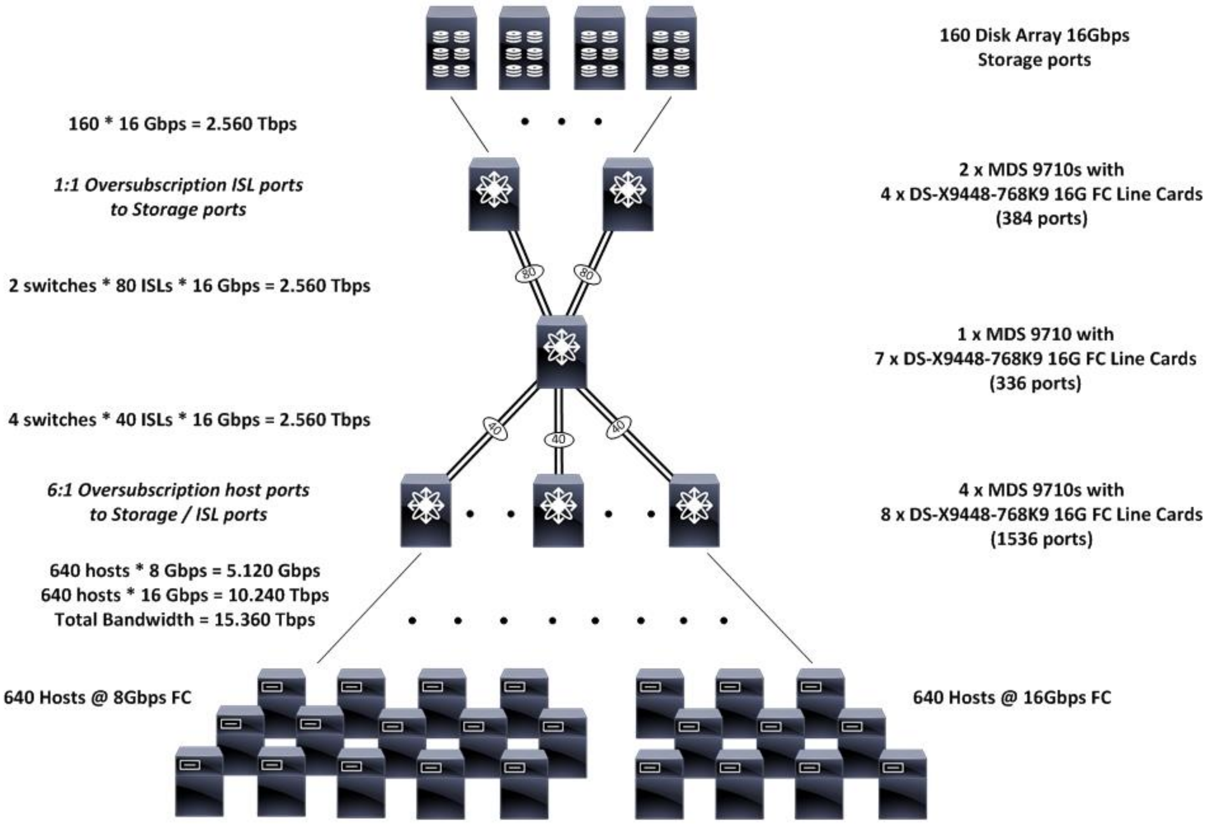


Figure 7: Use Case 2 Topology with mix of 8-Gbps and 16-Gbps hosts connected to 16-Gbps storage ports.

Note: This deployment scales to nearly 1500 devices in a single fabric. The actual production environment has 128 storage ports running at 16-Gbps and roughly 1300 host ports. The environment requires a minimum of 6:1 oversubscription within the network, which requires each host edge switch to have a 2.560-Tbps of ISL bandwidth. Again, storage ports will not grow quite as rapidly and the core switch has room to grow to add more host edge switches.

In this environment, the following was used in managing the network:

- Total of five VSANs created
    - VSAN 1 used for staging new devices
    - Four VSANs based upon business operations
- Used TACACS+ for authorization and auditing of Cisco MDS switches
- Created separate administrative roles for VSANs
- Created device-alias for environment
- Enabled Dynamic Port VSAN Membership (DPVM) feature. It dynamically assigns VSAN membership to ports by assigning VSANs based on the device WWN. DPVM eliminates the need to reconfigure the port VSAN membership to maintain fabric topology when a host or storage device connection is moved between two Cisco SAN switches or two ports within a switch. It retains the configured VSAN regardless of where a device is connected or moved.
- Mixture of two and three member zones

Summary

With data centers continually growing, SAN administrators must design networks that meet their current needs and can scale for demanding growth. Cisco MDS 9710 Multilayer Director class switches provide embedded features to help SAN administrators in these tasks. SAN administrators deploying large Cisco SAN fabrics can use the design parameters and best practices discussed in this paper to design optimized and scalable SANs.

Important links:

Cisco MDS 9700 Series Multilayer Directors
Cisco MDS 9710 Multilayer Director Data Sheet
Cisco MDS 9700 Series Supervisor-1 Module Data Sheet
Cisco MDS 9700 48-Port 16-Gbps Fibre Channel Switching Module Data Sheet
Compare Models: Learn about the similarities and differences of the models within this product series.
Data Sheets and Literature (4)
At-a-Glance Sheets
Data Sheets
Presentations
White Papers