Cisco Data Center Interconnect Design and Implementation Guide

System Release 1.0

 September 25, 2009

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Cisco Data Center Interconnect Design and Implementation Guide

CONTENTS

Preface

This Data Center Interconnect (DCI) Design and Implementation Guide (DIG) describes a portion of Cisco's system for interconnecting multiple data centers for Layer 2-based business applications.

Given the increase in data center expansion, complexity, and business needs, DCI has evolved to support the following requirements:

- Business Continuity
- Clustering
- Virtualization
- Load Balancing
- Disaster Recovery

There is a strong need to expand the application domain beyond a single data center. DCI is driven by the business requirements shown in Table i-1.

Table i-1 DCI Business Drivers

Business	IT Solutions
Disaster Prevention	Active/Standby Migration
Business Continuance	Server HA clusters, "Geo-clustering"
Workload Mobility	Move, consolidate servers, "Vmotion"

Several Applications are available for IT solutions to solve these business requirements.

HA Clusters/Geoclusters

- Microsoft MSCS
- Veritas Cluster Server (Local)
- Solaris Sun Cluster Enterprise
- VMware Cluster (Local)
- VMware VMotion
- Oracle Real Application Cluster (RAC)

- IBM HACMP
- EMS/Legato Automated Availability Manager
- NetApp Metro Cluster
- HP Metrocluster

Active/Standby Migration, Move/Consolidate Servers

- VMware Site Recovery Manager (SRM)
- Microsoft Server 2008 Layer 3 Clustering
- VMware VMotion

The applications above drive the business and operation requirement for extending the Layer 2 domain across geographically dispersed data centers. Extending Layer 2 domains across data centers present challenges including, but not limited to:

- Spanning tree isolation across data centers
- Achieving high availability
- Full utilization of cross sectional bandwidth across the Layer 2 domain
- Network loop avoidance, given redundant links and devices without spanning tree

Additional customer demands, such as Quality-of-Service (QoS) and encryption, may be required on an as needed basis.

Cisco's DCI solution satisfies business demands, while avoiding challenges, by providing a baseline for the additional enhancements cited above.

Cisco's DCI solution ensures Cisco's leadership role in the Enterprise/Data Center marketplace, securing their position as the primary competitive innovator.

Audience

This document serves as an introduction to DCI capabilities; providing configurations with basic guidelines and convergence results. This DIG will assist Network Architects, Network Engineers, and Systems Engineers in understanding various Cisco solution recommendations to geographically extend Layer 2 networks over multiple distant data centers, while addressing high performance and fast convergence requirements across long distances.

Motivation

Cisco recommends isolating and reducing Layer 2 networks to their smallest diameter, limited to the access layer. Server-to-server communication, High Availability clusters, networking, and security all require Layer 2 connectivity. In many instances, Layer 2 functionality must extend beyond a single data center, particularly when a campus framework extends beyond its original geography, spanning multiple long distance data centers. This is more pervasive as high-speed service provider connectivity becomes more available and cost effective.

High-availability clusters, server migration, and application mobility warrant Layer 2 extensions across data centers. To simplify data center deployment, this system level design and configuration guide provides the appropriate governance in configuring, provisioning and scaling DCI by:

- Enabling data center expansion with a Cisco approved solution
- Building new capabilities on top of an existing deployment base, extending the Catalyst 6500 footprint, and positioning high density platforms such as Nexus 7000 series switch
- Extending existing operational capabilities

References

The following document accompanies this *Cisco Data Center Interconnect Design and Implementation Guide, System Release 1.0*:

- *Cisco Data Center Interconnect Test Results and Configurations, System Release 1.0*

Best Practices

The following Cisco ESE Data Center team best practices document is available:

http://www.cisco.com/en/US/netsol/ns748/networking_solutions_design_guidances_list.html

The following disaster prevention and recovery best practices document is available:

http://www.cisco.com/en/US/netsol/ns749/networking_solutions_sub_program_home.html

The following Catalyst 6500 series switch best practices document is available:

http://preview.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SX/best/practices/recommendations.html

Data Center Architecture

The following data center multi-tier architecture document is available:

- *Cisco Data Center Infrastructure 2.5 Design Guide*

http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DC_Infra2_5/DCInfra_4.html#wp1094957

ONS

The following document details ONS 15454 DWDM procedures discussed in the "Cisco ONS 15454" section on page 1-20:

- *Cisco ONS 15454 DWDM Procedure Guide, Release 9.0*

http://www.cisco.com/en/US/docs/optical/15000r9_0/dwdm/procedure/guide/454d90_procedure.html

The following documents detail Cisco optical technology regarding ONS 15454 MSTP and ONS 15454 OTU2-XP XPonder modules, and provisioning discussed in the "Cisco ONS 15454" section on page 1-20:

http://www.cisco.com/go/optical

http://www.cisco.com/en/US/prod/collateral/optical/ps5724/ps2006/data_sheet_c78-500937.html

http://www.cisco.com/en/US/docs/optical/15000r9_0/dwdm/procedure/guide/454d90_provisiontmxpcards.html#wp554539

Storm Control

The following documents detail storm control on the Catalyst 6500 and Nexus 7000 platforms, respectively, discussed in the "Storm Control and Flooding" section on page 1-16:

http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SXF/native/configuration/guide/storm.html

http://www.cisco.com/en/US/docs/switches/datacenter/sw/4_1/nx-os/security/configuration/guide/sec_storm.html

CoPP

The following documents detail control plane protection on the Catalyst 6500 and Nexus 7000 platforms, respectively, discussed in the "Protection of the Control Plane" section on page 1-17:

http://www.cisco.com/en/US/prod/collateral/switches/ps5718/ps708/prod_white_paper0900aecd802ca5d6.html

http://www.cisco.com/en/US/docs/switches/datacenter/sw/4_1/nx-os/security/configuration/guide/sec_cppolicing.html

Acronyms

The following acronyms and respective technologies were used and discussed in this Data Center Interconnect Design and Implementation Guide.

Table i-2 DCI DIG Acronyms

AAA	Authentication, Authorization and Accounting
ARP	Address Resolution Protocol
ASIC	Application Specific Integrated Circuits
BPDU	Bridge Protocol Data Units
CFS	Cisco Fabric Service
CoPP	Control Plane Policing
CWDM	Course Wave Division Multiplexing
DCI	Data Center Interconnect
DIG	Design and Implementation Guide
DWDM	Dense Wave Division Multiplexing
ECMP	Equal Cost MultiPathing
EOBC	Ethernet Out-of-Band Channel
EoMPLS	Ethernet over Multiprotocol Label Switching
HA	High Availability
HSRP	Hot Standby Router Protocol
ICV	Integrity Check Value
ISL	Inter-Switch Links

Table i-2 DCI DIG Acronyms

LACP	Link Aggregation Control Protocol
LR	Long Range
MACsec	MAC Security
MEC	Multichassis EtherChannel
MPLS	Multiprotocol Label Switching
MSTP	Multiservice Transport Platform
MUX	Multiplex
NIC	Network Interface Card
OEO	Optical-to-Electrical-to-Optical
ONS	Optical Networking System
OOB	Out-of-Band
P2P	Point-to-Point
PAgP	Port Aggregation Protocol
PKL	Peer Keepalive Link
RAC	Real Application Cluster)
ROADM	Reconfigurable Optical Add/Drop Multiplexer
rPVST+	Rapid spanning-tree protocol
SAN	Storage Area Network
SR	Short Range
SRM	Site Recovery Manager
SSO	Stateful switchover
STP	Spanning Tree Protocol
VM	Virtual Machine
vPC	Virtual PortChannel
VSL	Virtual Switch Link
VSS	Virtual Switching System

Document Version

Document Version 1.0 published September 25, 2009.

CHAPTER 1

Cisco DCI Design Architecture

Data Center Interconnect (DCI) System Release is a solution often used to extend subnets beyond the traditional Layer 3 boundaries of a single site data center. Stretching the network space across two or more data centers can accomplish many things. Many clustering applications, both commercial and those developed in-house, require Layer 2 connectivity between cluster nodes. Putting these nodes in separate data centers can help build resilience into a system. If one data center is lost, the backup resources in the second data center can take over services with a much smaller impact, or a shorter downtime. DCI can also facilitate and maximize a company's server virtualization strategy, adding flexibility in terms of where compute resources reside physically and being able to shift them around geographically as needs dictate. DCI also presents challenges. How does an extended spanning-tree environment avoid loops and broadcast storms? How does a provider router know where an active IP address or subnet exists at any given time? It is also important to know which products in the Cisco portfolio strategically work together in support of the DCI solution. This is the focus of the DCI solution.

Figure 1-1 DCI Connectivity Overview

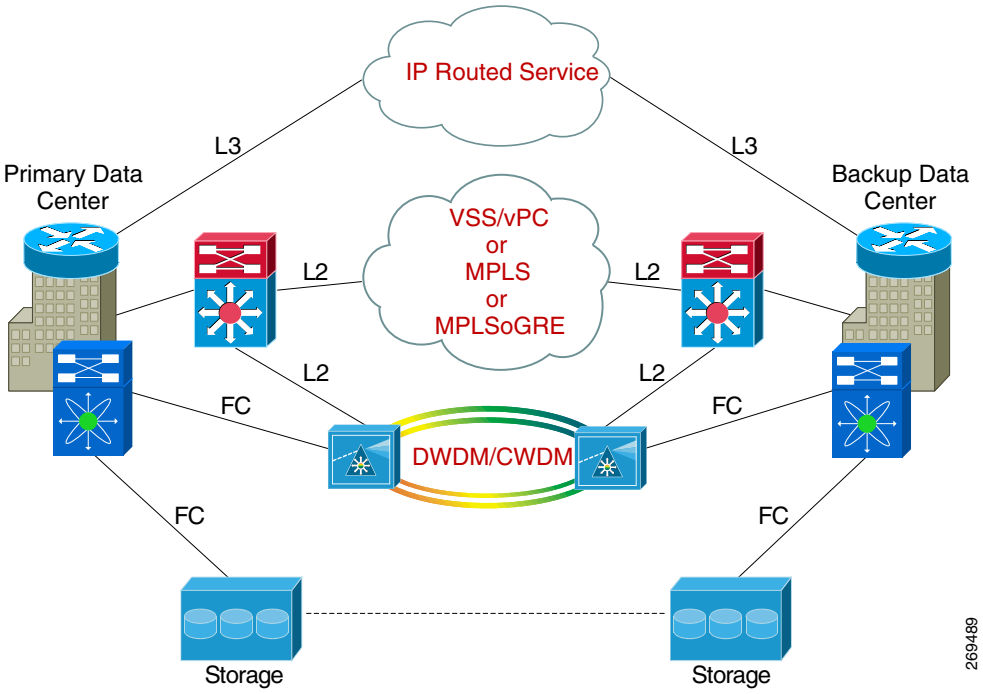


Figure 1-1 summarizes general DCI connectivity. The following data center extension requirements are dependent upon the application types.

- **Layer 2 Extensions:** Provide a single Layer 2 domain across data centers. The data center applications are often legacy or use embedded IP addressing that drives Layer 2 expansion across data centers.
- **Layer 3 Extensions:** Provide routed connectivity between data centers used for segmentation/virtualization and file server backup applications. This may be Layer 3VPN-based connectivity, and may require bandwidth and QoS considerations.
- **SAN Extensions:** Storing and replicating data for business continuity, disaster recovery, and/or regulatory compliance.

The DCI Release 1.0 solution focuses on the Layer 2 extension of the DCI network. There are multiple technical alternatives to provide LAN extension functionality:

- Point-to-point or point-to-multipoint interconnection, using Virtual Switching System (VSS) and virtual Port Channel (vPC) and optical technologies
- Point-to-point interconnection using Ethernet over Multiprotocol Label Switching (EoMPLS) natively (over an MPLS core) and over a Layer 3 IP core
- Point-to-multipoint interconnections using virtual private LAN services (VPLS) natively (over an MPLS core) or over a Layer 3 IP core

Note The DCI System Release 1.0 solution system release addresses Dark Fiber/DWDM based VSS and vPC deployment options. Other technology options will be covered in subsequent DCI system releases.

Design Goals

All DCI designs should meet the basic guidelines and requirements set forth in this DIG. DCI compatible platforms meet and support the following requirements:

- Layer 2 transport mechanism (VSS, vPC)
- High Availability (HA)
- Spanning Tree Isolation
- Loop avoidance
- Multi Path load balancing
- End-to-End convergence of 2-4 seconds
- Aggregating 2-4 10GE/GE and forwarding at least 1-2 10Gbps across the WAN
- Providing basic queuing
- Interconnect traffic encryption

Virtualization

Demand for virtualization of networks, servers, storage, desktops and applications continues to increase. Cisco is assuming the virtualization of DC services is now the default building block of the data center, and has organized its portfolio to hasten this transition.

Virtualization optimizes resources and is applicable in all areas of a data center network.

According to Forrester Research, server virtualization usage and criticality in production environments has increased. Correspondingly, as the use of live migration and blade server technologies has become commonplace, firms are increasingly focused on guaranteeing the performance and availability of their virtual environments from an end-to-end perspective. Looking beyond the boundaries of the virtual machine itself, there are operational challenges to ensure that network and storage configuration policies are followed uniformly. Better management and virtualization technology will address these and other maturity issues over time. Forrester believes that server and network virtualization technologies will become increasingly integrated to produce more dynamic data centers where administrators have greater visibility and control over the quality of service provided by their virtual infrastructure.

Advantages of Network and Server Virtualization

- Run multiple operating systems on a single server including Windows, Linux and more.
- Reduce capital costs by increasing energy efficiency and requiring less hardware and increasing your server to admin ratio.
- Leverage your existing network devices for segmentation, security, storage mobility and even Application delivery.
- Ensure your enterprise applications perform with the highest availability and performance.
- Build up business continuity through improved disaster recovery solutions and deliver high availability throughout the data center.
- Improve enterprise desktop management and control with faster deployment of desktops and fewer support calls due to application conflicts.

The DCI Release 1.0 solution, however, focuses on extending the Layer 2 data center domain , providing network and server virtualization with applications like VMware's VMotion.

Planning, Designing, Deploying, Operating DCI

Depending on the number of data center sites a customer maintains, DCI is categorized as point-to-point (P2P) or multi-site.

- **P2P:** Refers to two data center sites connected to each other, deployed in an active/backup or an active/active state.
- **Multi-Site:** Refers to data centers with more than two sites, which may be deployed with all sites active, two sites active and one backup, or in a hierarchical data center design covering different geographical regions.

Transport Options

Data centers may be interconnected over the following DCI-classified transport options:

- Dark Fiber (Optical)
- MPLS
- IP

The DCI Release 1.0 solution described herein utilizes customer-owned optical transport equipment. Optical multiplexing can be used on the Cisco Catalyst 6500 switches using Dense Wave Division Multiplexing (DWDM)/Course Wave Division Multiplexing (CWDM) optics installed directly on the Catalyst 6500. The Nexus 7000 will also support DWDM/CWDM. Check latest roadmap for support.

DWDM lets you take a single fiber and multiplex (MUX) several different signals across it. It requires an Optical Networking System (ONS) to MUX/deMUX. CWDM is like DWDM, but doesn't require ONS. It is a passive prism and can MUX/deMUX.

Release 1.0 testing focused on using short range (SR) and long range (LR) optics for the Catalyst 6500 and Nexus 7000 and off loading the optical multiplexing and transport capabilities onto the Cisco ONS 15454 platform.

Generic Data Center Network Best Practices

The DCI Release 1.0 solution provides best practices for DCI using VSS and vPC. Generic data center best practices are documented and tested by the Cisco ESE Data Center team. The following best practices documents are available at

http://www.cisco.com/en/US/netsol/ns748/networking_solutions_design_guidances_list.html

- *Data Center Blade Server Integration Guide*
- *Data Center Design—IP Network Infrastructure*
- *Data Center Service Integration: Service Chassis Design Guide*
- *Data Center Service Patterns*
- *Integrating the Virtual Switching System in Data Center Infrastructure*

HA Scalability

The DCI Release 1.0 solution places special emphasis on a highly scalable solution.

Disaster Prevention and Recovery

Disaster can and does strike anytime, anywhere, often without warning. No business can afford to shut down for an extended period of time. An effective business continuity program is imperative to keep a company up, running, and functional.

Before disaster strikes, it is essential for businesses to ensure that the right failover mechanisms are in place, most often in geographically dispersed locations, so that data access can continue, uninterrupted, if one location is disabled.

By integrating virtualization in all aspects of a data center design, with capabilities of seamlessly moving and migrating services across geographically distributed sites, data centers can forgo disaster prevention and offer recovery the instant a failure occurs.

In additional to the Layer 2 DCI Extension Solution documented here, the following link provides additional best practices for disaster prevention and recovery.

http://www.cisco.com/en/US/netsol/ns749/networking_solutions_sub_program_home.html

Resiliency

Network resilience is paramount to data center design for customers who build redundant data centers. Redundant extended data centers should not be isolated due to a single link or device failure.

Achieving data center resiliency can be accomplished by deploying dual DCI Layer switches, with each switch having a link [part of the Multichassis EtherChannel (MEC)] going across the data center, forming a redundant data center interconnect.

A fully meshed DCI layer switch infrastructure can be achieved if adequate lamdas or optical gear are available.

The DCI Release 1.0 solution implements the topology shown in Figure 1-1 where each DCI Layer switch has one link as part of the DCI MEC.

VLAN Scalability

Taking a modular approach to data center design provides flexibility, and scalability in both network topology design, and utilization of physical resources.

VLANs are often divided into odd and even sections and forwarded via the resilient network to achieve HA, scalability and efficient cross-section Layer 2 link utilization.

Note It is a best practice to extend only those VLANs that must truly be extended to offer virtualization services.

As part of testing, the Catalyst 6500 was tested with up to 500 Layer 2 VLANs. Refer to Cisco DCI Solution Details & Testing Summary, page 2-1 for detailed platform scalability test results. The system is capable of scaling to a larger number of VLANs.

The Nexus 7000 was also tested for 500 VLANs. Under the current NX-OS release, and a system configured with vPC, this is a scaling limit to be enhanced in future releases.

Note A Nexus 7000 without vPC supports 4k VLANs on trunks. Once vPC is enabled, testing showed that up to 500 VLANs can be supported. Higher scale is underway in future releases.

Layers

Hierarchical network design has been commonly used in enterprise networking for many years. This model uses redundant switches at each layer of the network topology for device-level failover that creates a highly available transport between end nodes using the network. Data center networks often require additional services beyond basic packet forwarding, such as server load balancing, firewall, or intrusion prevention. These services might be introduced as modules populating a slot of switching nodes in the network, or as standalone appliance devices. Each of these service approaches supports the deployment of redundant hardware to preserve the high availability standards set by the network topology.

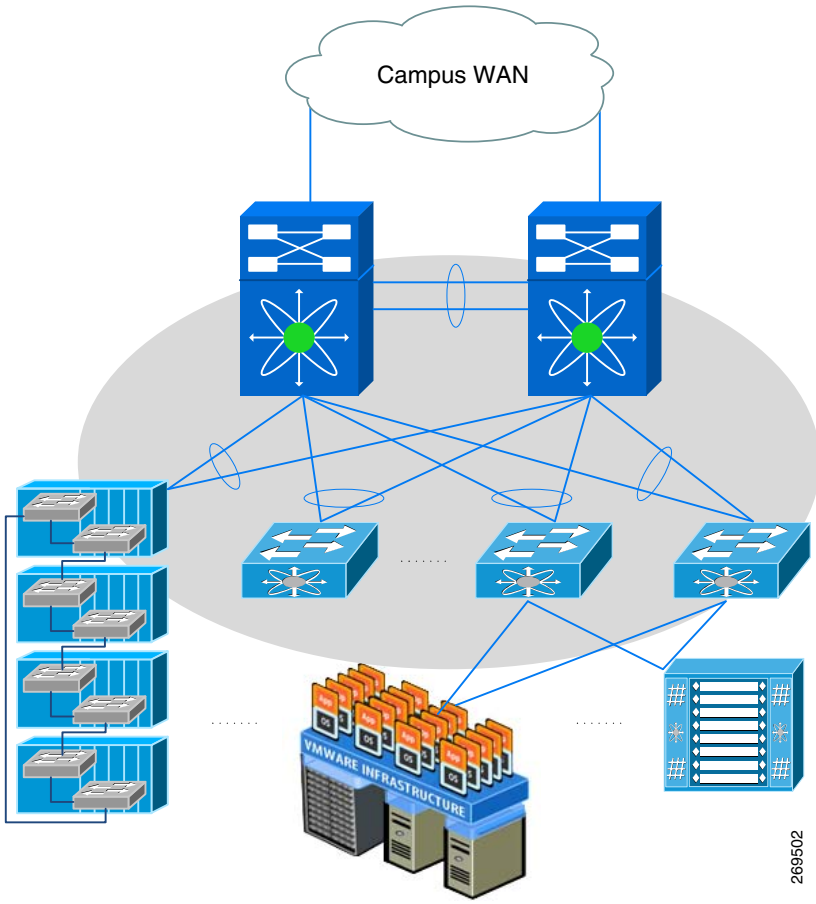
A structured data center environment uses a physical layout that corresponds closely with the network topology hierarchy. Decisions on cabling types, and the placement of patch panels and physical aggregation points must match interface types and densities of physical switches being deployed. In a new data center buildout, both can be designed simultaneously, taking into consideration power and cooling resource constraints. Investment concerns regarding the selection of switching platforms in existing data center facilities are strongly influenced by physical environment costs related to cabling, power, and cooling. Flexible networking requirements planning is vital when designing the physical data center environment. A modular approach to data center design ultimately provides flexibility, and scalability, in both network topology design and utilization of physical resources.

DC Core Layer

The Core Layer can be collapsed with a WAN Layer in a data center network. In some designs a customer may choose to collapse the data center Core/WAN and Aggregation Layers, which not only raises design issues but at times operational issues as well. The hierarchical network design adds stability and high availability characteristics by splitting out switching nodes based on their function, providing redundant switching units for each functional layer required. The core of a data center network is typically broken out into a pair of high performance, highly available, chassis-based switches. In larger, or geographically dispersed network environments, the core is sometimes extended to contain additional switches. The recommended approach is to scale the network core while continuing to use switches in redundant pairs. The primary function of the data center network core is to provide highly available, high performance, Layer 3 switching for IP traffic between other functional blocks in the network such as campus, Internet edge, WAN/branch, and data center. By configuring all links connecting to the network core as point-to-point Layer 3 connections, rapid convergence around any link failure is provided, and the control plane of the core switches is not exposed to broadcast traffic from end node devices, or required to participate in Spanning Tree Protocol (STP) for Layer 2 network loop prevention.

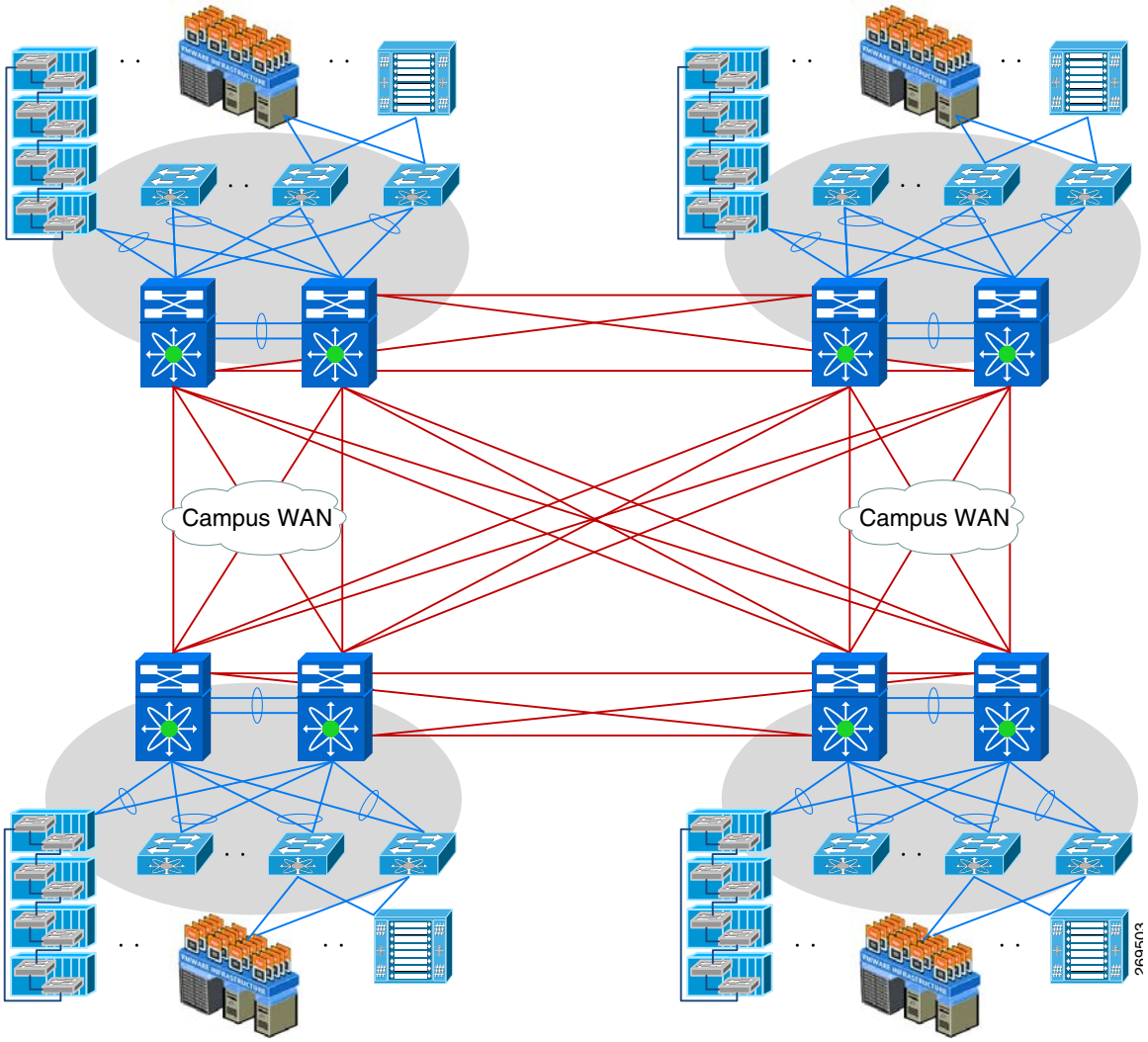
In small-to-medium enterprise environments, depending on scalability requirements, it is reasonable to connect a single data center aggregation block directly to the enterprise-switching core for Layer 3 transport to the rest of the enterprise network. Even the Core and Aggregation Layer could be collapsed as shown in Figure 1-2.

Figure 1-2 Small Simple DC Network



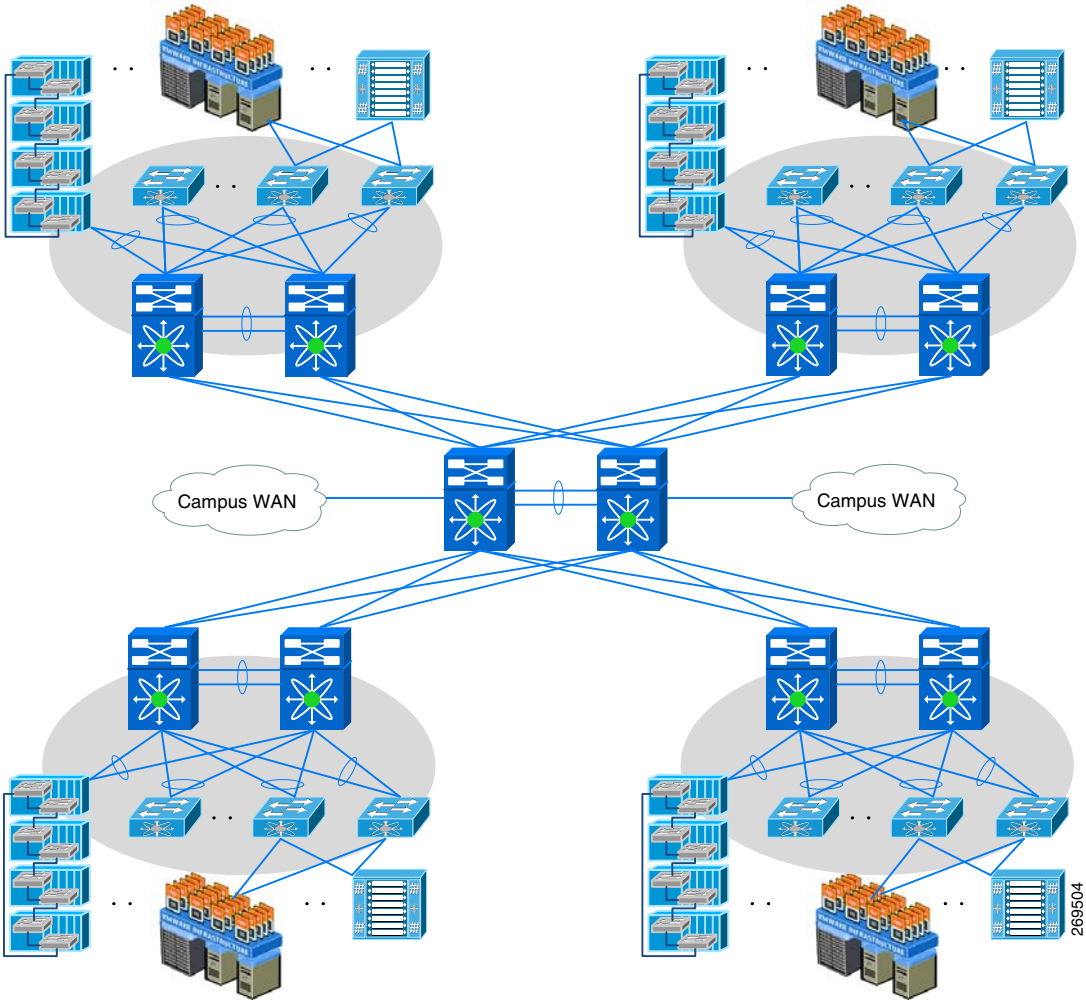
Provisioning a separate, dedicated pair of data center core switches provides additional insulation from the rest of the enterprise network for routing stability while providing scalability for future expansion of the data center topology. As the business requirements expand, and dictate two or more aggregation blocks serving separate pods or zones of the data center, a dedicated data center core network provides scaling expansion without requiring additional Layer 3 interface availability on the enterprise core as shown Figure 1-3.

Figure 1-3 Large DC Network with Collapsed Core Layer



Adding an additional Core Layer will simplify the design, create less operational issues in the network, and ease the link requirements as shown in Figure 1-4.

Figure 1-4 Large DC Network with Additional Core Layer



Additional reference regarding multi-tier architecture is available at:

http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DC_Infra2_5/DCInfra_4.html#wp1094957

DC Aggregation Layer

This document focuses on a redundant and efficient design. The Aggregation Layer is configured with Virtual Switching System (VSS), virtual Port Channel (vPC), or Non-VSS dual Catalyst 6500 Series switches in the DCI topology.

A VSS is two Catalyst 6500 (Cat6k) systems with VS-SUP720-10GE Supervisor modules connected together as a single virtual switch. A VSS system not only acts as a single port channel node but also unifies the control plane so the two physical switches appear as a single logical device (single router)

vPC in the Nexus 7000 offers a similar approach that allows building Layer 2 port channels that span across two Nexus 7000 chassis. From a control plane standpoint, the two Nexus 7000 chassis are separate entities.

The Aggregation Layer also acts as the root of the local STP domain. A key requirement of a DCI solution is the need to isolate spanning tree domains and keep them local to each data center. This is achieved by filtering spanning-tree Bridge Protocol Data Units (BPDU's) on the MEC link from local data centers going to the remote data center.

Layer 3 Routed Network

The Core switches are configured with BGP to interact with the WAN, and HSRP is configured to provide redundancy. Refer to Cisco DCI Solution Details & Testing Summary, page 2-1.

DC Access Layer

The Access Layer is connected to the Aggregation Layer switch using EtherChannel at the Aggregation Layer switches that may comprise a VSS or vPC system. The Access Layer can consist of fixed form or modular switches. The servers can be dual homed to a VSS system. The switches are configured to efficiently utilize all uplinks and minimize convergence times. Alternatively servers maybe be single homed.

DC Interconnect Layer

The DCI Layer provides the following options for the Layer 2 extension service:

- The DCI Layer could be part of the WAN router to be covered in Release 2.0 for MPLS deployments.
- In large-scale data centers where multiple aggregation blocks are built out, as shown in Figure 1-5, a separate set of redundantly configured devices form a separate DCI Layer, which can aggregate multiple pairs of Aggregation Layer switches, as was tested in the DCI Release 1.0.

  The purpose of this design is to:

  - Reduce the number of links that need to go across the WAN.
  - Allow Layer 2 connectivity not only across data centers but also within aggregation blocks in the same site.
- For smaller data centers that consist of a single set of aggregation switches, as shown in Figure 1-6, it might be hard to justify adding additional layer to extend the Layer 2 domain. In such scenarios, the Aggregation switches themselves could also be used to extend the Layer 2 domain and act as a DCI Layer switch (an assumption is made here that the Aggregation switches are VSS). The DCI Release 1.0 uses MEC to connect two data centers, ensuring both links are forwarding and optimally utilized under normal conditions.

Figure 1-5 Large DC with Multiple Aggregation Blocks

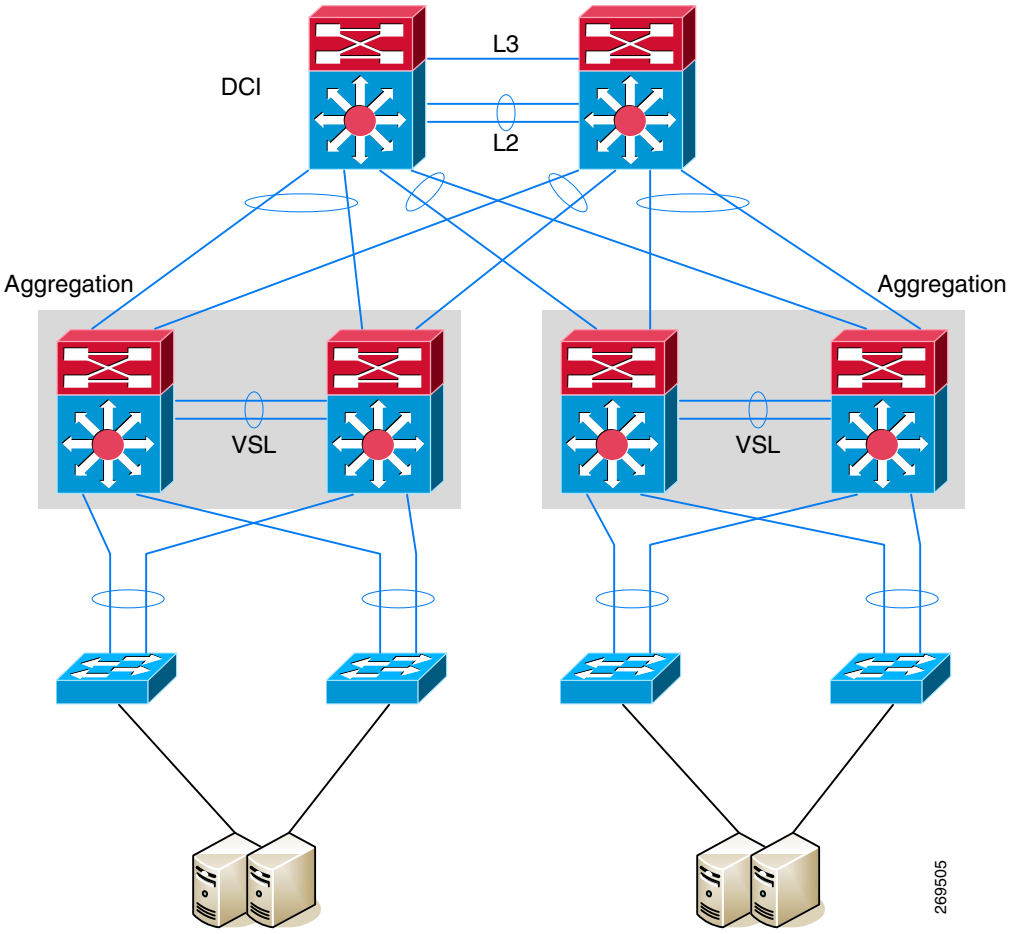
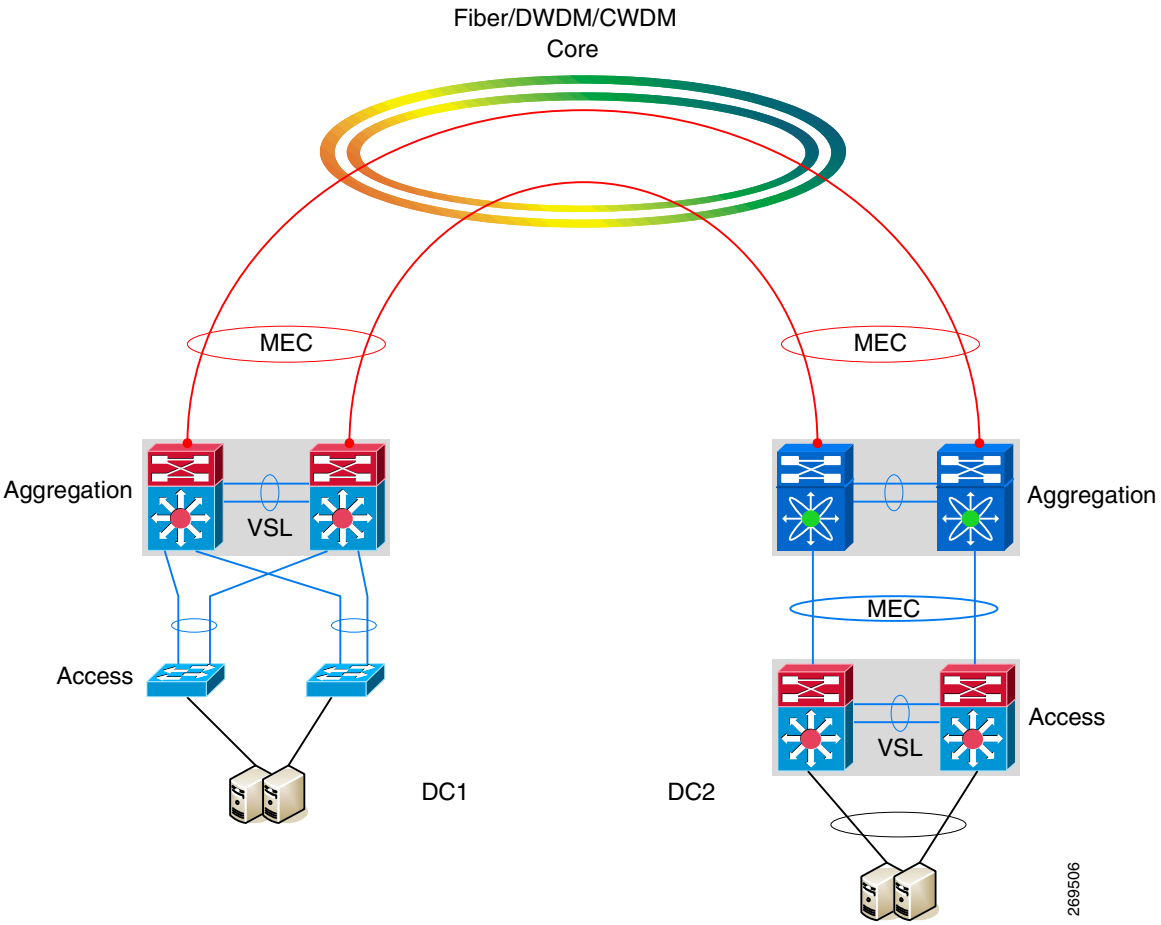


Figure 1-6 Small DC with Aggregation



The DCI Layer should also implement features such as CoPP and Storm Control. More details can be found in Storm Control and Flooding, page 1-16 and Protection of the Control Plane, page 1-17.

BPDU Filtering can also be enabled, selectively, on per-port basis. This feature disables participation in STP and keeps the port in forwarding mode. BPDU Filtering is used to assure STP isolation. This, however, also implies that the connection between the sites must be loop-free at all times. If a loop occurs on this connection, it might affect all data centers.

The reasoning behind isolating the STP domains is to:

- Assure that an STP loop event in one site doesn't affect the other sites.
- Assure an optimal placement for the STP primary and secondary root can be achieved in each data center.

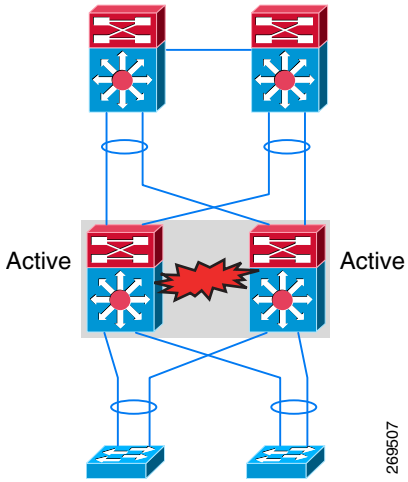
The VSS or vPC at the DCI Layer should be set up and configured with additional loop prevention and STP isolation during VSL or vPC peer link failures.

VSS is used for convergence and active/active forwarding through MEC. BPDUs are filtered at the DCI Layer to prevent extending a single spanning tree domain across multiple data centers.

Since BPDU's are filtered at the DCI Layer, additional measures need to be taken in addition to normal VSS and vPC configuration to avoid loops and avoid a “dual-active” state from occurring.

BPDU Filtering is enabled to completely isolate STP domains. When BPDU Filtering is enabled on a per-port basis it disables participation in STP and keeps the port in Forwarding mode. Refer to the information below on implementing features that can mitigate a dual-active state.

Figure 1-7 VSL Bundle Failure Causing Dual Active VSS



The following VSS dual-active detection mechanisms were tested and can be implemented:

- Enhancement to PAgP+ used in MEC when connecting Cisco switches, refer to Table 1-1
- L2 Fast-Hello Dual-Active Detection configuration on a directly connected link (besides VSL) between virtual switch members (supported starting with 12.2(33)SXI), the most recommended dual-active detection mechanism)

The following vPC dual-active mechanisms were tested:

- vPC pk-Link (Peer Keepalive Link)

WAN Layer 3 Hops

Routing protocol summarization is a common IP networking practice used to keep routing tables small for faster convergence and greater stability. In the data center hierarchical network, summarization may be performed at the data center Core or Aggregation Layers. Summarization is recommended at the data center Core if it is a dedicated layer that is separate from the enterprise Core. The objective is to keep the enterprise Core routing table as concise and stable as possible, so as to limit the impact of routing changes in other places of the network from impacting the data center, and vice versa. If a shared enterprise core is used, summarization is recommended at the data center Aggregation Layer. To enable summarization, proper IP address allocation must be used in the subnet assignment to allow them to be summarized into a smaller number of routes.

However, in active/active or active/standby data centers, where the same subnet is extended across multiple data centers summarization of all routes may not be possible and hence more specific routes even host routes can be used to avoid sub-optimal or asymmetric routing as it can cause severe issues with TCP congestion flow mechanisms.

Further detail will be addressed in subsequent releases of the DCI System Release.

Layer 2 Considerations

Building Layer 2 extensions for DCI is essential and critical in design. While extending Layer 2 is the primary focus of this section, routing plays an important role between data centers as well. Not all traffic has a requirement to be passed within the same VLAN (bridged). Most application or server-to-server traffic is able to traverse Layer 3 hops. To provide Layer 3 connectivity one additional VLAN is provisioned between the sites. This VLAN is a pure "transit VLAN" which connects all aggregation layer switches or core routers into a single subnet to establish a routing adjacency. This way traffic that needs to be routed between data centers can make use of the bandwidth available on the DCI link.

Adjacency Using VLANs

The requirement for clustered applications, legacy non-IP implementations as well as virtual machine mobility, dictate that some VLANs extend between data centers. It is important to understand that Layer 2 expansion should be carefully performed, and applied only to VLANs that necessitate such connectivity. The VLAN number space between data centers must be coordinated. If VLAN spaces are overlapping and can’t be reassigned or realigned, VLAN translation may be considered and is typically configured in the Aggregation or DCI Layers.

Loop Prevention

When building an interconnect between data centers, availability is of utmost concern. This often implies that there are redundant connections between sites, so that no single link failure can disrupt communication between them. Unless sites are geographically adjacent (within same campus), connections between those sites are costly. With that, it is of interest to utilize as much bandwidth as possible. More specifically, it is ideal that all links, rather than just one, shall be utilized.

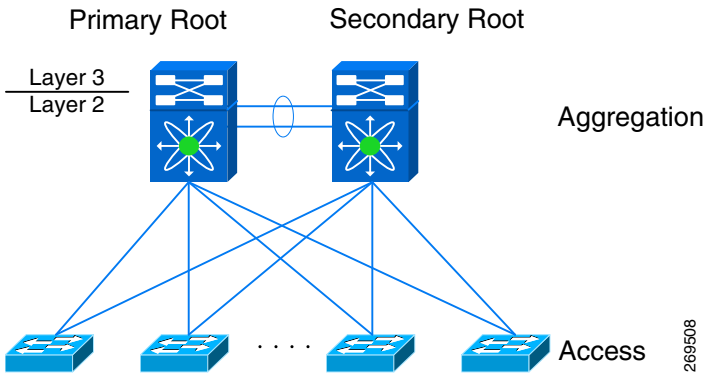
With Layer 2 extended, STP assures only one active path to prevent loops, which conflicts with the ideal of using all links. One way to resolve this conflict is to configure the spanning tree parameters so that even VLANs are forwarded over one link and odd VLANs are forwarded over the other. Another option is to present a single logical link to spanning tree by bundling links using EtherChannel. This release of testing focuses on the latter concept and utilizes it to extend VLANs between two or more data centers. Technologies to achieve this are VSS, which offers Multichassis EtherChannels (Catalyst 6500 series), or Virtual Port Channels (Nexus 7000 series).

Spanning Tree Protocol Design Considerations

Typically, an STP domain spans as far as a VLAN reaches. So if multiple data centers share a common VLAN, the STP domain extends to all those data centers.

Figure 1-8 shows that from a STP design standpoint, best practices place the primary and secondary root in an Aggregation Layer. The assumption is that the Aggregation Layer builds the top most level of a Layer 2 hierarchy (the VLAN spans from Access Layer switches to Aggregation Layer switches).

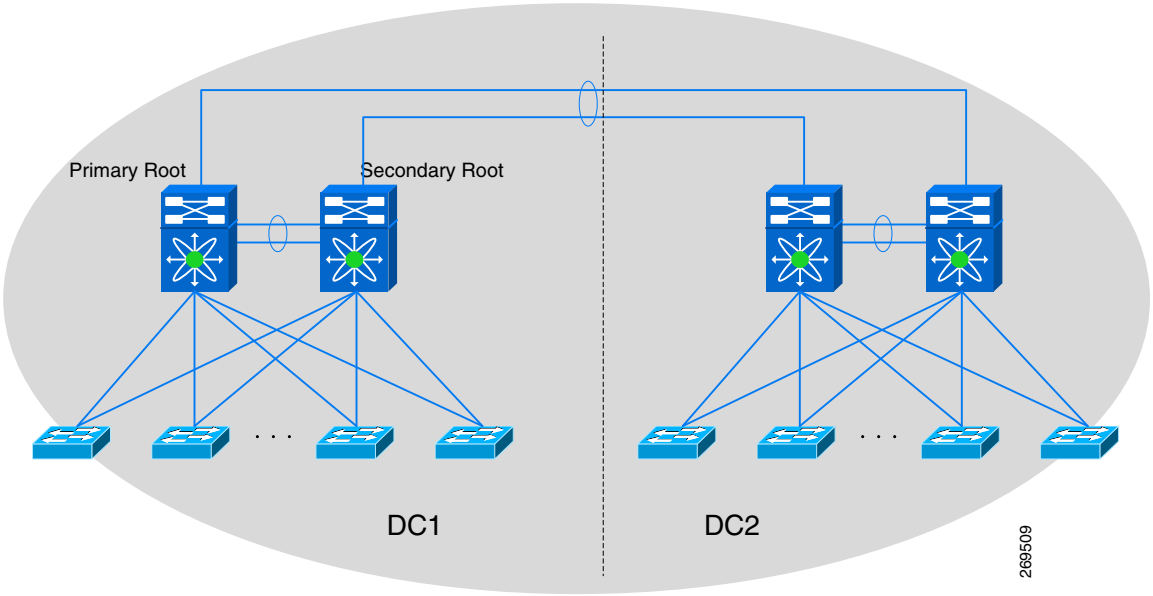
Figure 1-8 Aggregation Layer, Layer 2 Hierarchy



When a VLAN gets extended to another Aggregation Layer, the question arises where the primary and secondary STP root is placed.

In Figure 1-9 the grey oval indicates the extended STP domain. It becomes apparent in this illustration that there is no optimal placement of the STP root for data center 2. It is preferred that each data center has its own primary and secondary STP root.

Figure 1-9 VLAN Expansion to Another Aggregation Layer



Another STP design consideration regarding multiple data centers is the type of STP used. On Cisco Catalyst and Nexus switches the following STP modes are supported natively or compatibly:

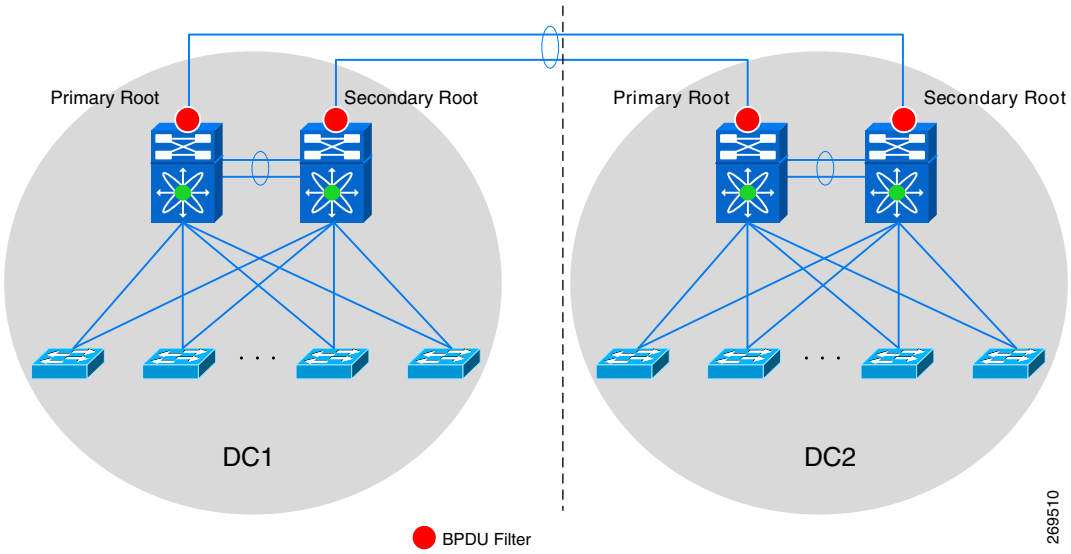
- IEEE 802.1D
- PVST+ (Per VLAN Spanning Tree)
- Rapid-PVST
- IEEE 802.1s MST (Multiple Spanning Tree)

As new data centers are built, the opportunity to move to a more scalable STP mode is given. If VLANs are extended between data centers, it is the existing data center STP mode that typically dictates the mode being used. It is preferable to have the freedom of moving to the STP mode that makes the most sense from a device support and scalability standpoint.

Ideally, STP events should be isolated to a single data center. Typically, any link state change of a port that participates in a VLAN, triggers a BPDU with the topology change notification (TCN) bit set. Ports that are configured as edge ports, or configured with Spanning Tree PortFast do not trigger TCN BPDUs. When the STP root receives a BPDU with the TCN bit set, it will send out a BPDU with the topology change (TC) bit set resulting in all bridges initiating a fast age out of MAC address table entries (802.1D or PVST+), or immediately flushing those MAC entries (Rapid-PVST or 802.1w). This may result in temporary flooding events in the respective VLANs, which is an undesired but essential side effect to ensure the MAC addresses re-learning.

These three design and implementation challenges (root placement, STP mode, BPDU isolation) can be addressed using BPDU Filtering. Figure 1-10 shows the impact of applying BPDU filters.

Figure 1-10 BPDU Filtering Impact



By applying BPDU filters, the STP domain for a VLAN, or a set of VLANs, is divided (indicated by the blue ovals). Consequently, the STP root is localized to each data center (the STP root design is now optimal). Given that there is no interaction between the data centers, from an STP standpoint, each data center can make use of an independent STP mode. An existing data center might run Rapid-PVST while the new data center can run MST. Also, BPDU TCN messages are local to the data center and will not initiate any fast aging of MAC address table entries in other connected data centers.

Note Use caution when applying BPDU filters. A Layer 2 loop can be introduced if filtering is applied in parts of the network where multiple paths exist. In Figure 1-10, the link between the data centers represents a portchannel and logically appears as a single STP link.

Storm Control and Flooding

For large Layer 2 domains, the following traffic types are flooded to all ports that participate in a VLAN:

- Broadcast
- Layer 2 multicast (non-IP-can not be pruned using IGMP)
- Traffic to unknown destinations (unknown unicast)

Some protocols, essential in IP networking, rely on broadcast. The most common is the Address Resolution Protocol (ARP), but depending on the network environment, there may be others. A hardware fault or application level error can cause massive levels of broadcast or unknown unicast traffic. In extreme cases it can happen that a whole data center is affected and effectively no “good” traffic can be passed anymore. To stop a spread of such traffic across multiple data centers, the Catalyst and Nexus families of switches allow suppression of these kinds of traffic types.

It is important to baseline broadcast, Layer 2 non-IP multicasts, and unknown unicast traffic levels before applying suppression configurations which should be applied with headroom and constantly monitored. The introduction of new, or deprecation of legacy, applications could change traffic levels.

When applied properly, even a STP loop in one data center may only cause a limited amount of traffic to be sent to other data center(s).

For Storm Control details on specific platforms refer to:

Catalyst 6500

http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SXF/native/configuration/guide/storm.html

Nexus 7000

http://www.cisco.com/en/US/docs/switches/datacenter/sw/4_1/nx-os/security/configuration/guide/sec_storm.html

Security & Encryption

Security for data centers is a primary concern for data center architects. Security considerations range from the physical location over physical access control to the data center, all the way to traffic flow to and from, as well as within and between data centers. Encryption of critical or confidential data is also often vital.

Note This section focuses on security and encryption as it relates to Ethernet & IP networking. Technologies like storage media encryption provided by the Cisco MDS 9500 Fibre Channel director switches is outside the scope of this document, as are physical security aspects of a data center.

Securing and Hardening Network Infrastructure

Running a secure and hardened infrastructure is the foundation for continued operations of a data center. Access via SSH, or console/terminal servers to the infrastructure has to be secured and controlled. Requirements here are no different from what has been known from the past.

AAA Infrastructure

Access control via Authentication, Authorization and Accounting (AAA) using protocols such as TACACS+ or RADIUS must be used to prevent unauthorized access to data center devices. Also, access to the management interface should be allowed from specific subnets only. Client/Server subnets should not be able to access the management interface/network. The AAA infrastructure (Cisco ACS) should be built as fault tolerant since it plays a critical role in access and management devices. The infrastructure is often built using an out-of-band (OOB) management network, providing complete separation of management and client/server traffic.

Protection of the Control Plane

State of the art switching platforms perform forwarding in hardware using Application Specific Integrated Circuits (ASICs). Only traffic destined to the switch, typically management traffic or protocols that run among network devices, uses the switch's CPU. Given that switches aggregate large amounts of bandwidth, the CPU can potentially be stressed beyond normal levels through incorrect configurations and/or malicious traffic. To protect the CPU from such attacks, rate limiters and QoS policies can be applied. This is called Control Plane Policing (CoPP). Various Catalyst as well as Nexus platforms offer hardware based CoPP capabilities.

Each network environment is different. Protocol scaling, number of servers as well as traffic levels dictate the correct CoPP values. The Nexus 7000 has CoPP configured by default. The network administrator can choose from multiple pre-defined profiles (strict, moderate, lenient, none) and adjust these profiles to meet specific needs.

To find the optimal CoPP configuration, perform a baseline that includes low and high network activity conditions, whereupon an optimal configuration can be found (i.e. baseline plus margin). Continuous monitoring should be in place to observe CPU trends that may trigger an adjustment of CoPP policies.

For CoPP details on specific platforms refer to:

Catalyst 6500

http://www.cisco.com/en/US/prod/collateral/switches/ps5718/ps708/prod_white_paper0900aecd802ca5d6.html

Nexus 7000

http://www.cisco.com/en/US/docs/switches/datacenter/sw/4_1/nx-os/security/configuration/guide/sec_cppolicing.html

Other Security Mechanisms

Further security mechanisms include:

- **NetFlow:** Allows monitoring the overall amount of traffic and number of flows and that can trigger an alarm in case levels go beyond pre-defined thresholds.
- **Address Spoofing Filters:** Incorrectly configured end stations might inject traffic which could impact overall network performance.
- **Native VLAN:** Generally it is a best practice to not allow user traffic on the native VLAN. If there is still a requirement to have user traffic on the native VLAN it is recommended to configure the switch to tag traffic for the native VLAN.

- **Port Security:** Port Security allows for specifying how many MAC addresses should be allowed on a physical port. Faulty network interface cards (NICs) have shown in the past that random traffic may be sent to the network. If the source MAC address is “randomized,” this can result in either creating unnecessary MAC moves and/or CAM tables filling up. The latter can result in unnecessary flooding of traffic. Caution should be used for servers hosting multiple virtual machines since those could be hosting tens of MAC addresses.

Further Best Practices can be found here:

http://preview.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SX/best/practices/recommendations.html

Encryption

Enterprises and Service Providers store data in data centers that contain information which needs to be protected as intellectual property, sensitive information like patient and financial data, and regulatory information. Depending on the type of corporation there might be regulatory requirements that have to be followed by the respective organization. Examples of such data protection requirements are HIPAA, PCI, Sarbanes-Oxley (SOX), Basel II, and FDA.

Data that is passed between locations has to be encrypted. This requirement holds true for data base synchronization traffic, applications exchanging state information as well as users accessing applications. Generally speaking as soon as traffic leaves the physical data center premises, data needs to be encrypted. This not only assures data privacy but also data integrity. No one should be able to modify data while it is in transit.

While some applications encrypt data at the presentation or application layer, it is typical that hundreds if not thousands of applications are active within an enterprise corporation. Changing or enhancing all corporate applications to communicate over a secure channel can be very time consuming and costly. Network-based encryption, however, is able to encrypt all traffic between two or more locations since it acts at the lower layers in the 7 layer OSI model and therefore does not require any changes to the application.

One of the most popular network-based encryption protocols is IPSec which encrypts IP packets and sends them across an encrypted tunnel. Encryption performance and the operational aspects of setting up IPSec tunnels has been greatly simplified over the years with automated tunnel set up mechanisms (DMVPN). Based on the way the IPSec tunnel technology is defined today, it is that traffic can not be bridged into a tunnel but must be routed into a tunnel.This effectively means that all traffic passed via an IPSec tunnel is routed. In other words, IPSec natively doesn't allow bridged traffic to be encrypted.

As mentioned, data centers often not only require Layer 3 but Layer 2 connectivity between each other. This either requires the use of an additional Layer 2-in-Layer 3 tunneling protocol (such as L2TPv3) that, in turn, could be encrypted using IPSec or other protocol. In 2006 the IEEE ratified the 802.1AE standard, also known as MAC security standard (MACsec). MACsec encrypts all Ethernet frames, irrespective of the upper layer protocol. With MACsec, not only routed IP packets but also IP packets where the source and destination is in the same subnet or even non-IP traffic are encrypted.

DCI Network Components

Other components of DCI that were monitored but not actively tested in release 1.0 include Microsoft Clusters. The Microsoft Cluster was configured in multiple data centers and heartbeat on the cluster was used to monitor server availability all convergence numbers were within limits.

Table 1-1 lists Cisco Data Center Interconnect network components.

Table 1-1 DCI Network Components

Component	Make/Model	SW Version	Notes
Server Cluster	Generic	Windows Server 2003	Each site will have a cluster of 2 servers
DHCP Server	Generic	4.2.0.30P2 (Solaris 10)	Required for providing IP Addresses to Microsoft Cluster
DCI Layer and Aggregation Layer	Nexus 7000 - N7K-SUP1 - N7K-C7010-FAB - N7K-M132XP - SFP-10G-SR	NX-OS 4.2(1)	
DCI, Core, Aggregation and Access Layers	WS-C6509-E VS-S720-10G-3CXL [1] VS-S720-10G-3C [2] WS-SUP720-3BXL X2-10GB-SR WS-X6748-GE-TX WS-X6716-10GE WS-X6708-3C WS-X6704-3BXL WS-CAC-6000W	12.2(33)SXI1	VSS + ISSU is available in 12.2(33)SXI
DWDM	ONS-15454 15454-OPT-AMP-17 15454-40-WSS-C 15454-40-DMX-C 15454-MS-ISC-100T 15454-OSCM 15454-OSC-CSM 15454-TCC2P-K9 15454-OPT-PRE 15454-OTU2-XP	9.0	

1. VSS capable Supervisor
2. VSS capable Supervisor

Customer Benefits

The Cisco DCI Release 1.0 provides the following customer benefits:

- **Customer Revenue Impact:** Gives customers end-to-end DCI options, so they can enable new functionality with minimal network upgrade and flexible network design (Layer 2 or Layer 3 data center). This provides a better return on investment for new services.
- **Installation & Deployment Impact:** The DCI System Release enables Cisco's existing customers to enable their end-to-end DCI strategies. Since Cisco is a known leader in industry solutions and services, customers will find it easy to successfully deploy this solution.
- **Maintenance and Ongoing Operational Impact**: Customers using Cisco platforms will find it easy to upgrade their existing infrastructure to offer DCI services.
- **Customer's Competitive Position:** Gives the customer options to use their existing Cisco infrastructure to rollout new services quickly. From a management perspective, customers can customize existing management tools.
- **Regulation or Standard Compliance:** This solution enables business continuity and disaster prevention, elements which are requirements for critical data security, separation and backup and compliance for regulations such as HIPAA.

Networking Technology

This testing focuses on switches connected directly to each other via a Gigabit Optical Fiber Network (Dark Fiber or DWDM). MEC or vPC PortChannels run on these fiber networks to create Inter-DC connectivity. This deployment is protocol agnostic.

Optical Transport (CWDM, DWDM)

CWDM can match the basic switching capabilities of dense wavelength-division multiplexing (DWDM), but with the inherent trade off of lower capacity for lower cost. Consequently, there is no intrinsic reason why CWDM should not be a viable technology in terms of performance and price for the coming generation of reconfigurable optical metro networks based on DWDM wavelength switching.

DCI Release 1.0 assumes the availability of Dark Fiber as a transport to connect multiple data centers. The Dark Fiber Optical Layer provides connectivity between different sites. The maximum and minimum distances tested between 2 data centers are 100km and 50km respectively. The Optical Layer also provides connectivity for storage transport between data centers.

Note SR and LR optics were used to connect the Catalyst 6500 and Nexus 7000 to local ONS systems.

Cisco ONS 15454

The Cisco ONS 15454 Multiservice Transport Platform (MSTP) is the most deployed metropolitan-area (metro) and regional DWDM solution in the world. The platform features two- through eight-degree reconfigurable optical add/drop multiplexer (ROADM) technology that enables wavelength provisioning across entire networks and eliminates the need for optical-to-electrical-to-optical (OEO)

transponder conversions. The ONS 15454 MSTP interconnects with Layer 2, Layer 3 and storage area network (SAN) devices at speeds of 40 Gbps. It delivers any service type to any network location and supports all DWDM topologies.

The platform offers reconfigurable optical add/drop multiplexing (ROADM) functions that allow zero to 40 channels of pass-through or add/drop in both the C-band and the L-band, A-Z wavelength provisioning, and full real-time power monitoring of each individual wavelength.

Layer 1 DCI extension in a 3-site topology included 3 different 2-degree ROADM nodes. An MSTP ring is formed by connecting 1 degree at each ROADM site and each span is capable of carrying 40 channels. Each ROADM site has a main node controller shelf and a subtending 454 chassis to accommodate the OTU2-XP Xponder used in the DCI test. Each span was designed to be a 40-wavelength span capable of supporting up to 40 x 10 GigE clients ports from transponders.

The OTU2-XP was tested as a dual 10 GigE transponder or 2 x 10GE transponder mode and in optical fiber switched or splitter mode. This dual transponder mode. In this mode, the OTU2-XP can accept 2 x 10 GigE client signals (from Nexus 7000 or Catalyst 6500) and map Client Port 1 (LAN PHY) to trunk port 3 (DWDM trunk) and client port 2 to trunk port 4. In a subtending shelf of the MSTP as tested, this can provide up to 24 Client 10 GigE's in a single 454 chassis.

Additional documentation can be found at:

http://wwwin.cisco.com/sptg/cmtsotbu/optical/products/mstp/salestools.shtml

ONS 15454 MSTP

http://www.cisco.com/go/optical

ONS 15454 OTU2-XP XPonder

http://www.cisco.com/en/US/prod/collateral/optical/ps5724/ps2006/data_sheet_c78-500937.html

Release 9.0 ONS 15454 DWDM Procedure Guide

http://www.cisco.com/en/US/docs/optical/15000r9_0/dwdm/procedure/guide/454d90_provisiontmxpcards.html#wp554539

Virtual Switch System (VSS)

The Virtual Switch System (VSS) is available on the Catalyst 6500 switches, providing the following benefits:

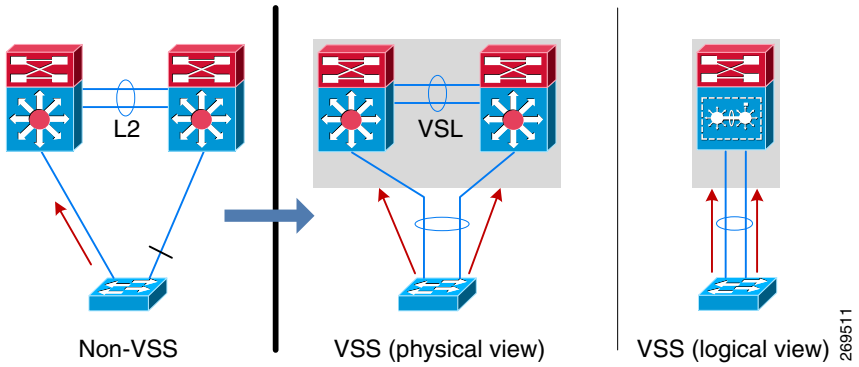
- Consolidation of control and management planes across multiple chassis
- Active-Active data plane
- Stateful switchover (SSO) across chassis

VSS also provides Multichassis EtherChannel (MEC) as a means to connect upstream and downstream devices to each of the Catalyst 6500 switches in the VSS domain. MEC provides the following benefits:

- Removes dependence on STP for link recovery
- Doubles effective bandwidth by utilizing all MEC links
- Reduce the number of Layer 3 routing neighbors since the VSS domain is seen as a single entity

Figure 1-11 shows a VSS system with Catalyst 6500 switches.

Figure 1-11 VSS with Catalyst 6500 Series Switches



The Virtual Switch Link (VSL) is at the heart of healthy VSS functionality. Thus, protecting the VSL bundle is of upmost importance. There are several best practices to a accomplish this.

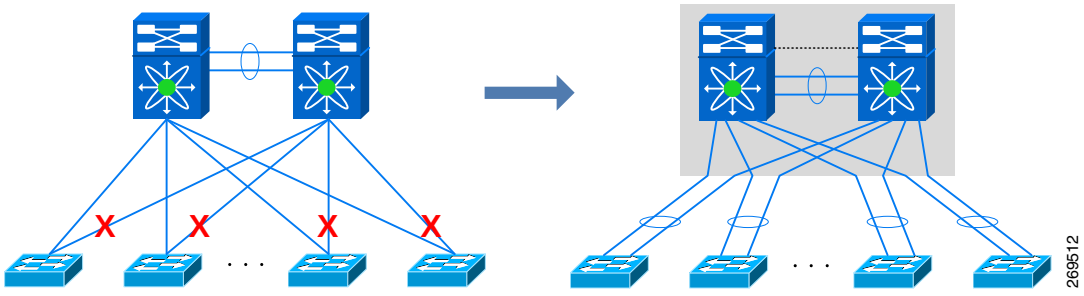
- Using one port from the Supervisor and another from the 67XX line cards to form the VSL bundle
- Over provisioning bandwidth to the VSL link
- Using diverse fiber paths for each of the VSL links
- Managing traffic forwarded over the VSL link by avoiding single-homed devices

Virtual Port Channel (vPC)

Similar to VSS on the Catalyst 6500, Virtual Port Channel (vPC) offers port channel distribution across two devices, allowing redundant yet loop-free topologies. Currently, vPC as a technology is offered on the Nexus 7000 and Nexus 5000 platforms.

Compared to traditional STP-based environments, vPC allows redundant paths between a downstream device and its two upstream neighbors. With STP, the port channel is a single logical link that allows for building Layer 2 topologies which offer redundant paths without STP blocking redundant links.

Figure 1-12 vPC Physical Topology

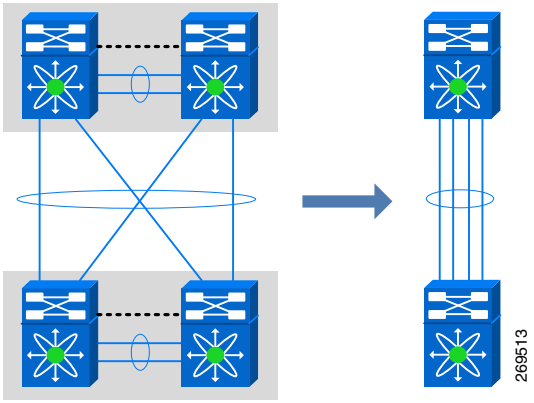


In contrast to VSS, vPC does not unify the control plane of the two peers. Each vPC peer is configured separately and runs its own independent instance of the operating system. The only interaction between the two chassis is facilitated using the Cisco Fabric Service (CFS) protocol, which assures that relevant configuration pieces and MAC address tables of the two peers are in synch.

A downstream device sees the vPC domain as a single LACP peer since it uses a single LACP ID. Therefore the downstream device does not need to support anything beyond IEEE 802.3ad LACP. If the downstream device doesn't support 802.3ad LACP, a port channel can be statically configured (**channel-group** *group* **mode on**). Currently, NX-OS does not support PAgP which typically does not pose a problem given LACP standardization acceptability and longevity.

vPC can be configured back-to-back. This configuration is one of the key building blocks used for the testing performed in this validation effort. Figure 1-12 and Figure 1-13 shows the physical view and how it translates to the STP and LACP logical view. There are no loops so spanning tree sees the topology as a single Layer 2 link. While a link or vPC peer failure results in reduced bandwidth, there is no spanning tree re-convergence needed. The only convergence necessary at a port channel layer.

Figure 1-13 vPC Physical Topology

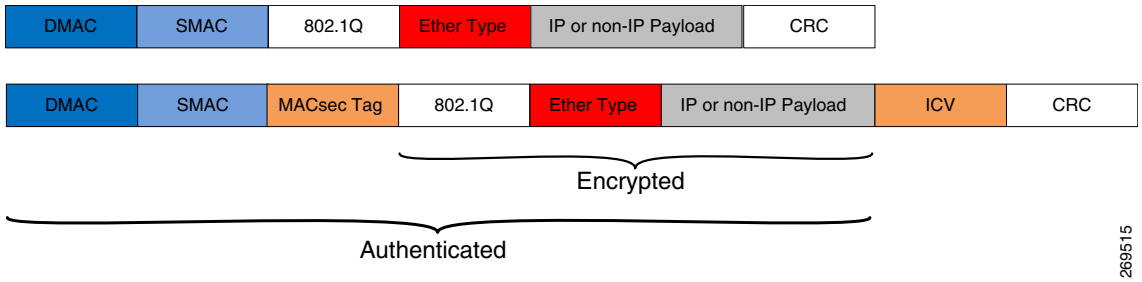


IEEE 802.1AE MACsec with Nexus 7000

As part of the Cisco TrustSec infrastructure, the Nexus 7000 supports IEEE 802.1AE standards based encryption (also known as MACsec) on all interfaces. This hardware level encryption is where each port has its dedicated crypto engine that can be turned on in NX-OS configuration. Since each port has dedicated resources, there is no tax on the supervisor or line card CPU. All data encryption and decryption is performed at the port level. Apart from the additional MACsec header, there is no overhead or performance impact when turning on port level encryption. From a deployment standpoint, the main difference between MACsec and IPSec is that MACsec is performed at a link layer (meaning hop-by-hop basis), while as IPSec is fundamentally building a tunnel that goes from one IP address to another over potentially many hops.

802.1AE not only protects data from being read by others sniffing the link, it assures message integrity. Data tampering is prevented by authenticating relevant portions of the frame. Figure 1-14 shows how a regular Layer 2 frame is encrypted.

Figure 1-14 Encrypted Layer 2 Frame



The MACsec Tag plus the Integrity Check Value (ICV) make up 32 Bytes (no 802.1AE metadata is considered). While the 802.1Q header, the ether type and payload are encrypted, destination and source MAC are not. As for the integrity check, the whole frame, with the exception of the ICV and the CRC, is considered. This assures that not even a source or destination address of the frame could be manipulated.

CHAPTER 2

Cisco DCI Solution Details & Testing Summary

Cisco Data Center Interconnect (DCI) solution system releases extend LAN and SAN connectivity across geographically dispersed data centers. These solutions allow organizations to provide high-performance, non-stop access to business-critical applications and information. They support application clustering, as well as Virtual Machine (VM) mobility between data centers, to optimize computing efficiency and business continuance.

These solutions offer flexible connectivity options to address LAN, Layer 3 and SAN extension needs across optical (Dark Fiber/DWDM using VSS, vPC), Multiprotocol Label Switching (MPLS), or IP infrastructures. They also provide:

- Transparency to the WAN network architecture
- Support for multi-site data center architectures
- LAN and SAN extension, with per-data center fault isolation
- A resilient architecture that protects against link, node, and path failure
- Optional payload encryption
- SAN Write Acceleration and compression, to increase replication options

Cisco DCI solutions offer a scalable, flexible platform that satisfies requirements for server and application mobility and business continuance.

Release 1.0 of the solution focuses on the Layer 2 extension of data centers across geographically dispersed data centers using the Dark Fiber/DWDM approach and uses the VSS and vPC capabilities on the Catalyst 6500 and Nexus 7000, respectively.

Interoperability of VSS and vPC

VSS with MEC and vPC are two fundamentally different features. VSS allows the Catalyst 6500 system consisting of two switches to appear as a single network entity, while vPC maintains the separate control planes of the two devices which still allows the capability of link aggregation from physical links coming from two different Nexus 7000s. These two technologies, consequently, have their respective place in the overall architecture as a way to eliminate spanning tree and enable the EtherChannel initiation and termination, to and from the two physically separate chassis.

DCI Topologies

This DCI design takes into account multiple scalability and resiliency factors and provides a solution based on them.

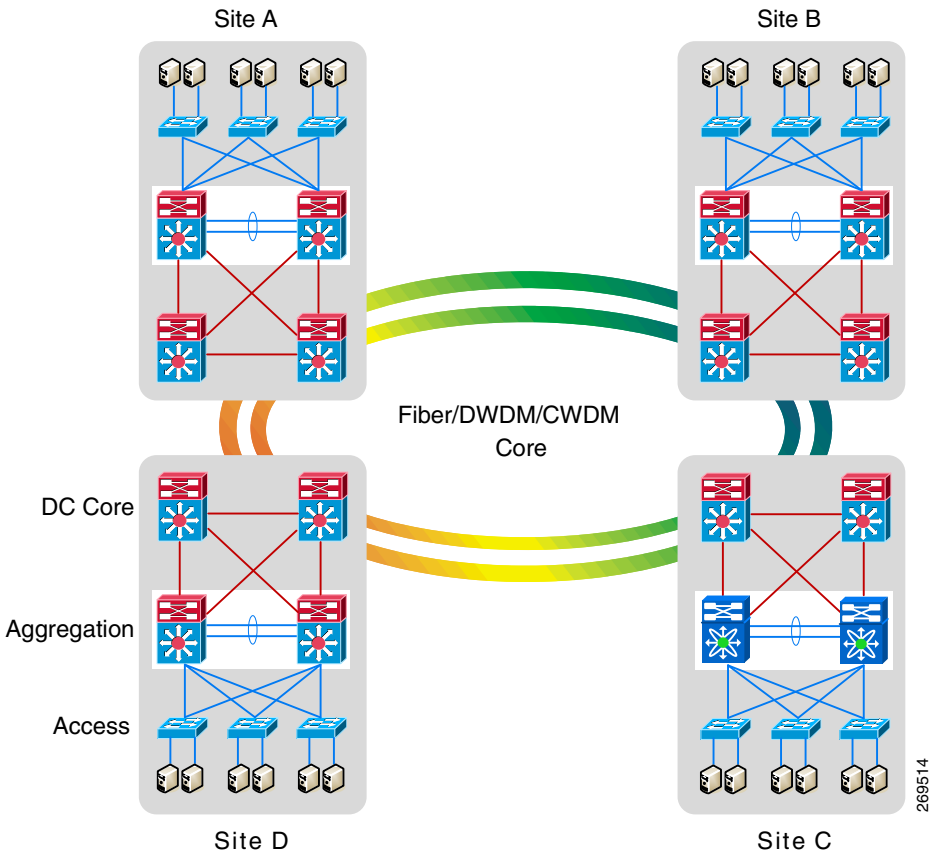
Consider the following factors before choosing a appropriate DCI design:

- Number of VLANs that need to be extended across data centers
- Number of data centers that need to interconnected
- Total amount of intra-data center bandwidth required
- Convergence and recovery times in failure scenarios
- Number of servers (MAC address scalability)
- Platform capable of providing existing data center features
- Possibility of leveraging existing network equipment

This document presents multiple data center interconnect topology options. Select a design option based on customer specific requirements and platform positioning.

Figure 2-1 shows a customer use case topology implementing DCI System Release 1.0 testing.

Figure 2-1 DCI Fiber, DWDM, and CWDM Core Topology



This method assumes the following physical connectivity:

- Fiber/DWDM between sites
- Optionally leverage the DWDM gear (e.g. ONS 15454) on the Nexus 7000 and the Catalyst 6500
- Cisco ONS 15454 to enable Multiplexing of wavelengths

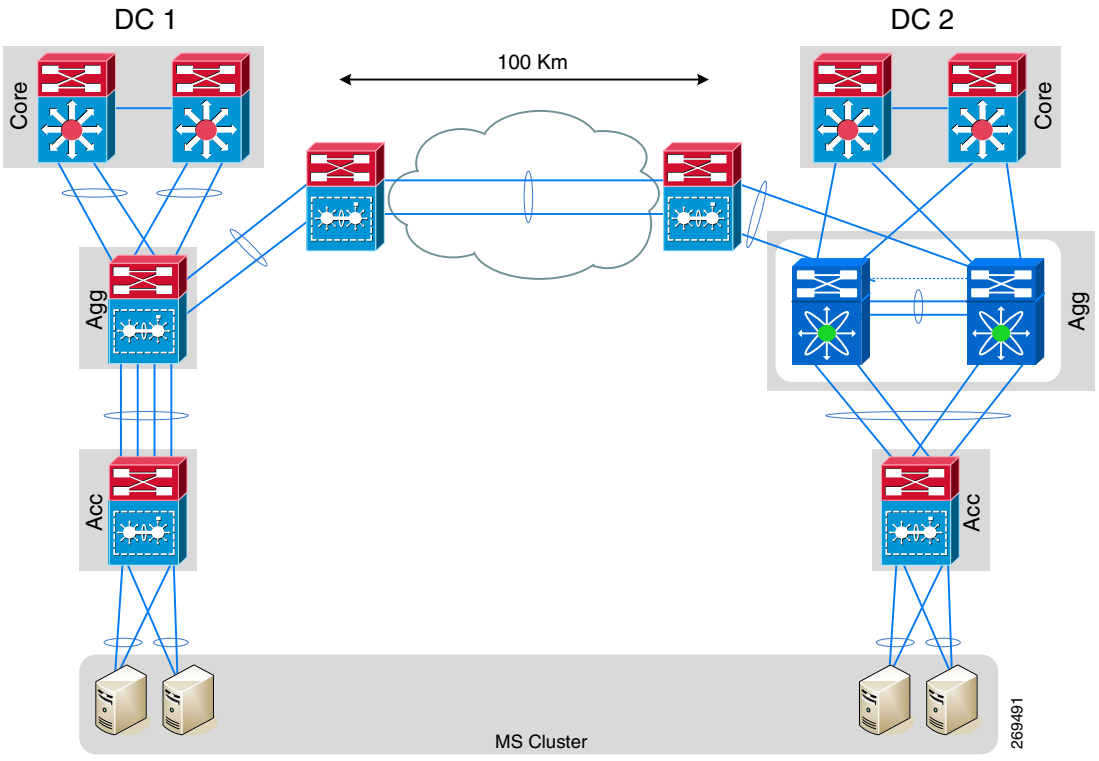
2 Sites VSS-VSS Case Study

The first deployment case study as illustrated in Figure 2-2 uses MEC to connect a pair of Catalyst 6500 switches running VSS at the DCI Layer of either data center.

Likely Customers

This is an ideal deployment scenario for customers requiring high bandwidth capacity between two data centers only who can also leverage existing VSS capable aggregation layer switches. The data centers could be deployed in an active/active or active/standby state. Refer to Table 2-1 for hardware details. Refer to Test Findings and Recommendations, page 2-43 for test case considerations.

Figure 2-2 VSS-VSS 2-Site Test Topology



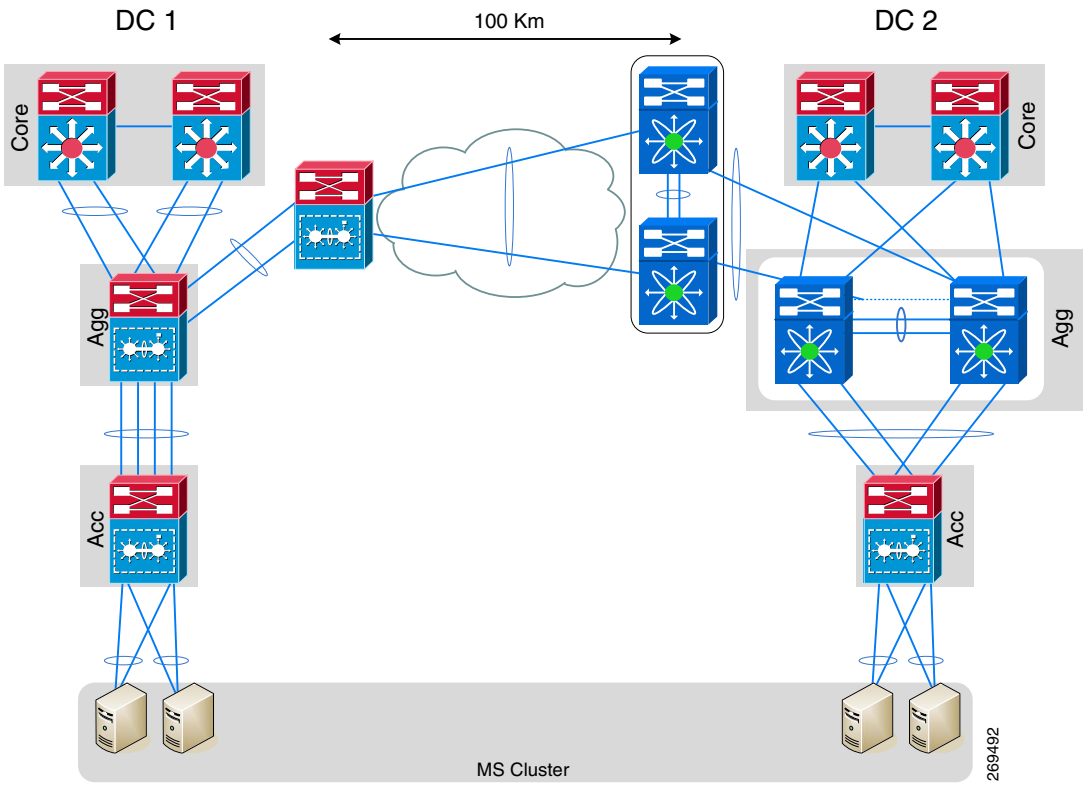
2 Sites VSS-vPC Case Study

The second deployment case study as illustrated in Figure 2-3 uses MEC to connect a pair of Catalyst 6500 and Nexus 7000 switches at the DCI Layer of either data center.

Likely Customers

This is an ideal deployment scenario for customers requiring high bandwidth capacity between two data centers only who have an existing Catalyst 6500 based data center and are building a new data center with the Nexus 7000. Refer to Table 2-1 for hardware details. Refer to Test Findings and Recommendations, page 2-43 for test case considerations.

Figure 2-3 VSS-vPC 2-Site Test Topology



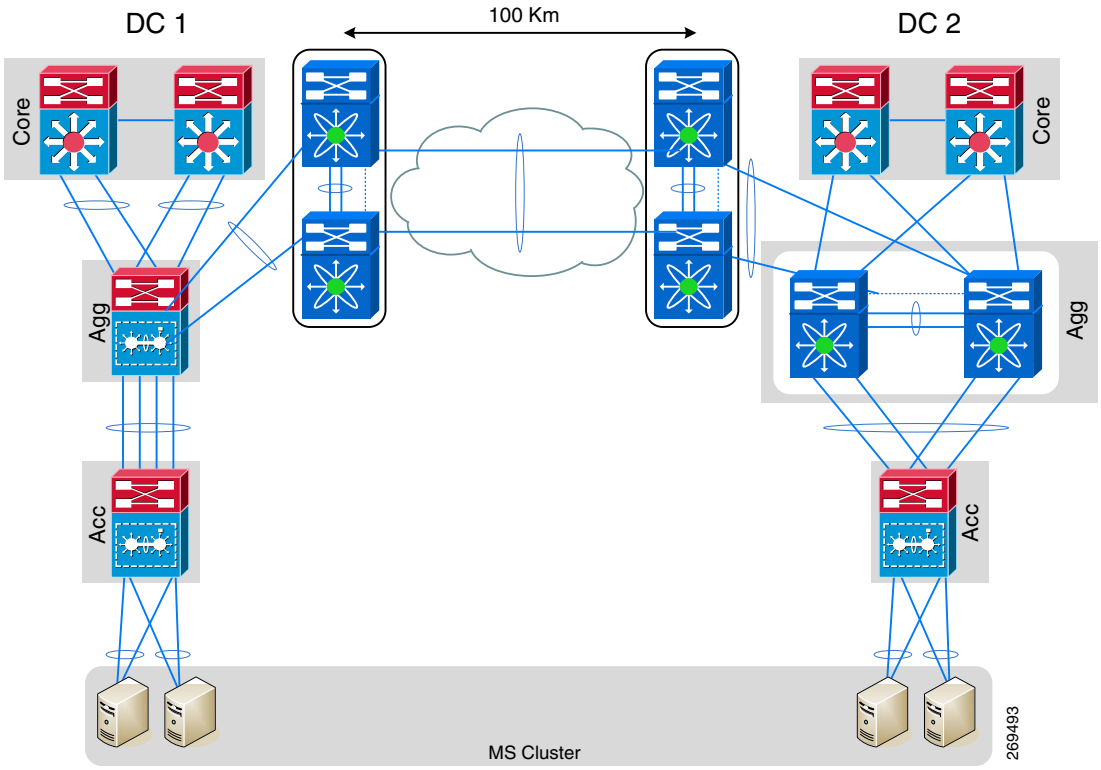
2 Sites vPC-vPC Case Study

The third deployment case study as illustrated in Figure 2-4 uses vPC to connect a pair of Nexus 7000 switches at the DCI Layer of either data center.

Likely Customers

This is an ideal deployment scenario for customers requiring high bandwidth capacity between two data centers only with a large number of aggregation uplinks. Refer to Table 2-1 for hardware details. Refer to Test Findings and Recommendations, page 2-43 for test case considerations.

Figure 2-4 vPC-vPC 2-Site Test Topology



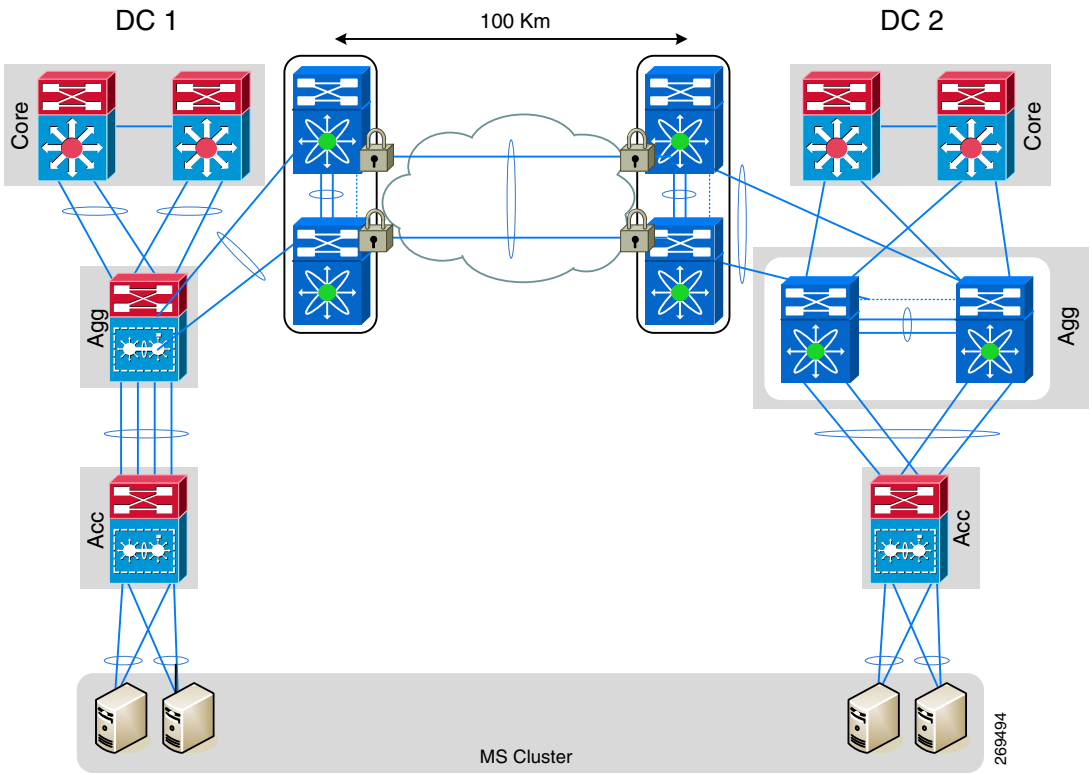
2 Sites vPC-to-vPC with 802.1AE Encryption Case Study

The fourth deployment case study as illustrated in Figure 2-5 uses vPC to connect a pair of Nexus 7000 switches at the DCI Layer of either data center.

Likely Customers

This is ideal for new deployment scenario customers requiring high bandwidth capacity between two data centers only and a large number of aggregation uplinks with encryption. Encryption is provided using 802.1AE standard. Refer to Table 2-1 for hardware details. Refer to Test Findings and Recommendations, page 2-43 for test case considerations.

Figure 2-5 vPC-vPC (with 802.1AE) 2-Site Test Topology



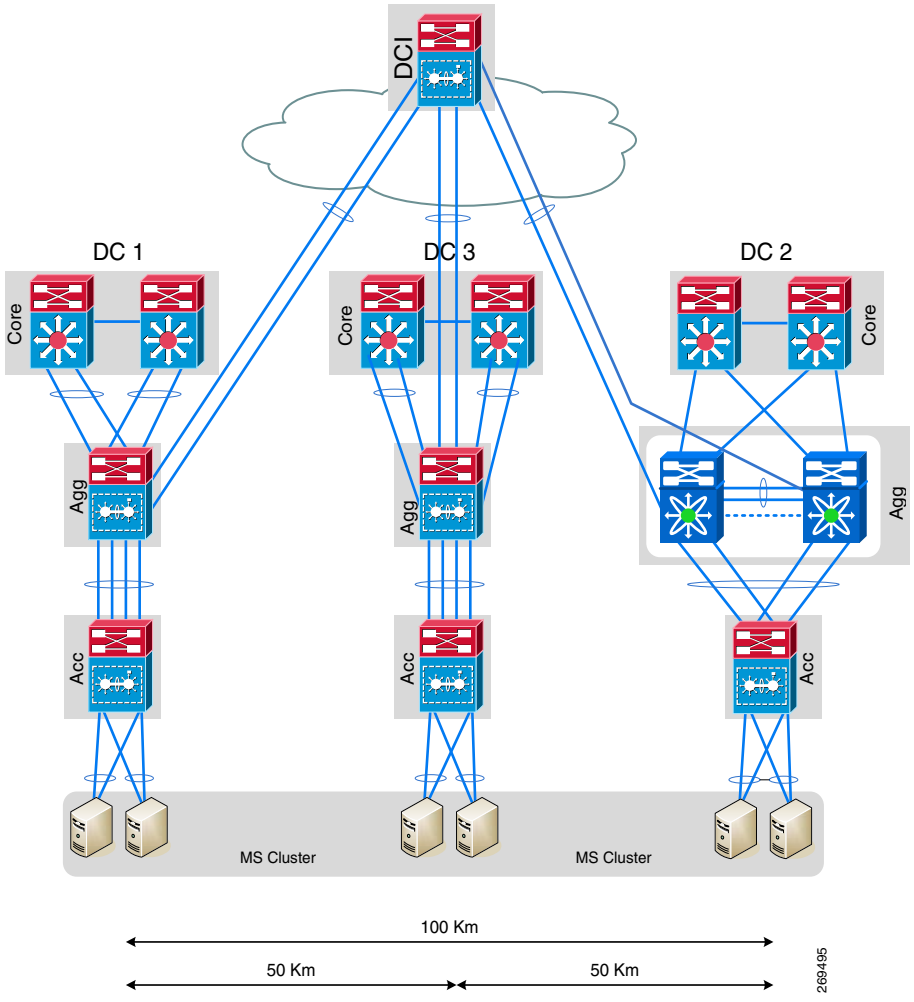
3 Sites VSS at DCI Case Study

The fifth deployment case study as illustrated in Figure 2-6 uses MEC to connect multiple data centers to Catalyst 6500 switches using VSS which form a DCI VSS (core) Layer. The individual switches are strategically places in the two main data centers. The local aggregation switches connect using the SR Optics, and the remote DCI VSS Layer switch is connected using the DWDM network.

Likely Customers

This is an ideal deployment scenario for customers requiring high bandwidth capacity between more than two data centers who want to leverage existing VSS capable aggregation layer switches. This deployment could have data centers made up of VSS or vPC based systems at the aggregation layer. Refer to Table 2-1 for hardware details. Refer to Test Findings and Recommendations, page 2-43 for test case considerations.

Figure 2-6 3 Site Test Topology with VSS Core



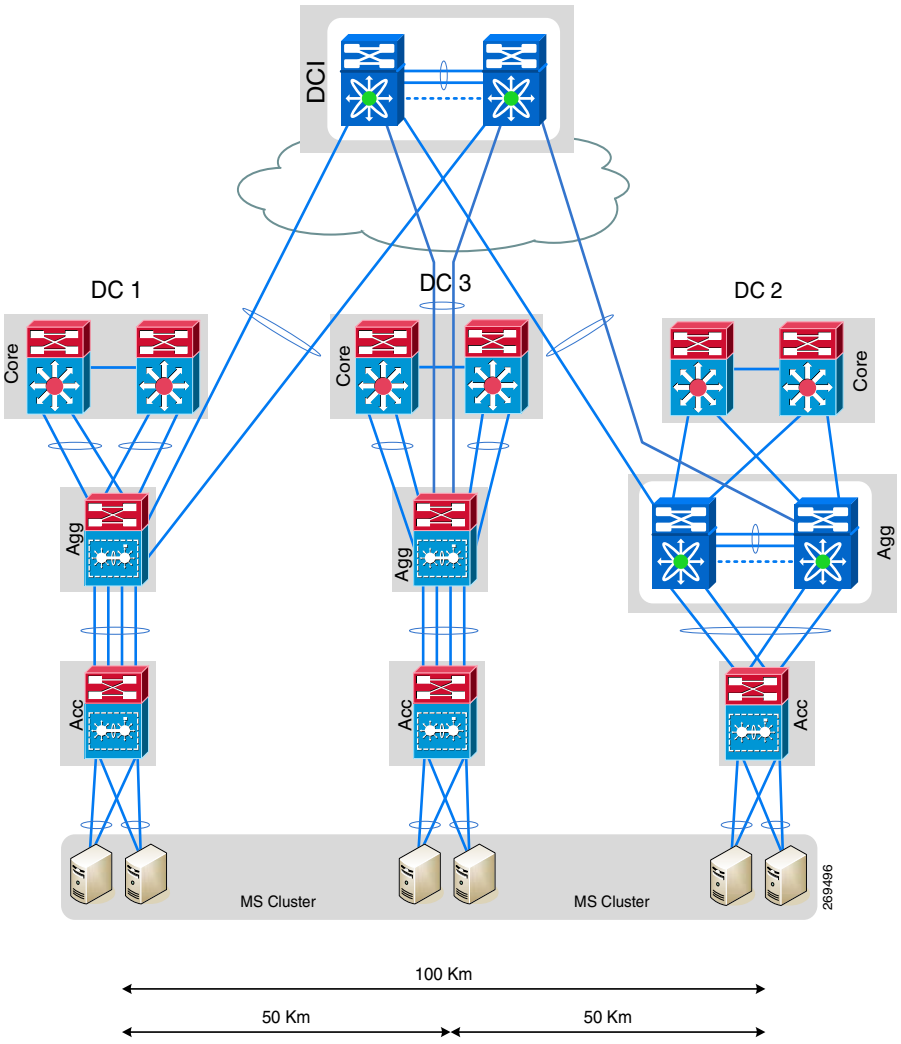
3 Sites vPC at DCI Case Study

The sixth deployment case study as illustrated in Figure 2-7 uses vPC to connect multiple data centers to Nexus 7000 switches which form a DCI vPC (core) Layer. The individual switches are strategically places in the two main data centers. The local aggregation switches connect using the SR or LR Optics and the remote DCI vPC Layer switch is connected using the DWDM network.

Likely Customers

This is an ideal deployment scenario for customers requiring high bandwidth capacity between more than two data centers. This deployment could have data centers made up of VSS or vPC based systems at the aggregation layer. Refer to Table 2-1 for hardware details. Refer to Test Findings and Recommendations, page 2-43 for test case considerations.

Figure 2-7 3 Site Test Topology with vPC Core



Testing Overview

The material provided below is intended to illustrate, through practice, the concepts and technologies that have been discussed above in this document. In order to achieve this, multiple test topologies were built and tested so that the large bulk of requirements for successful data center interconnect deployments were met.

Two key technologies were leveraged to achieve data center interconnect and Layer 2 extension: VSS on the Catalyst 6500 and vPC on the Nexus 7000. Given these two technologies alone, there are several options for deployment that can be explored. For the purposes of this testing and this paper, six distinct options, or use cases, became the focus of testing. These six use cases are listed below. Details of these use cases and how they were built and tested in the lab are provided in later sections.

- VSS-to-VSS 2-site connectivity
- VSS-to-vPC 2-site connectivity
- vPC-to-vPC 2-site connectivity
- vPC-to-vPC 2-site connectivity with 802.1AE encryption
- Multi-Site connectivity using VSS core
- Multi-Site connectivity using vPC core

For each of these use cases, as they were tested, certain levels of scale were built into the topologies in order to get closer to real-world deployment scenarios. The following list provides the scale numbers that may be of interest:

- 500 Layer 2 VLANs (see note below)
- 100 VLAN SVIs
- 10,000 client-to-server flows
- 20 Gbps traffic flows between data centers

Note 500 Layer 2 VLANs was noted during testing as the Cisco NX-OS 4.2(1) scalability limit for the Nexus 7000 using vPC.

Table 2-1 provides a quick summary of the hardware and software utilized in the testing for this solution. Details of the hardware configuration for each of the use cases are provided in further detail in later sections.

Table 2-1 DCI Hardware/Software Coverage

Place in Network	Platform	Hardware Used	Software Used
Core	Catalyst 6500	WS-SUP720-3BXL WS-X6704-10G	12.2(33)SXI1 IOS Software Modularity
Aggregation	Catalyst 6500	VS-S720-10G-3C WS-X6708-10GE-3C WS-X6716-10GE-3C	12.2(33)SXI1 IOS Software Modularity
	Nexus 7000	N7K-SUP1 N7K-M132XP-12	NX-OS 4.2(1)

Table 2-1 DCI Hardware/Software Coverage (continued)

Place in Network	Platform	Hardware Used	Software Used
Access	Catalyst 6500	VS-S720-10G-3C WS-X6708-10GE-3C WS-X6748-GE-TX	12.2(33)SXI1 IOS Software Modularity
DCI	Catalyst 6500	VS-S720-10G-3C WS-X6708-10GE-3C	12.2(33)SXI1 IOS Software Modularity
	Nexus 7000	N7K-SUP1 N7K-M132XP-12	NX-OS 4.2(1)
Optical Transport	ONS 15454	15454-OPT-AMP-17 15454-40-WSS-C 15454-40-DMX-C 15454-MS-ISC-100T 15454-OSCM 15454-OSC-CSM 15454-TCC2P-K9 15454-OPT-PRE 15454-OTU2-XP	9.0
Server		32-bit Generic	Windows 2003 Microsoft Clustering

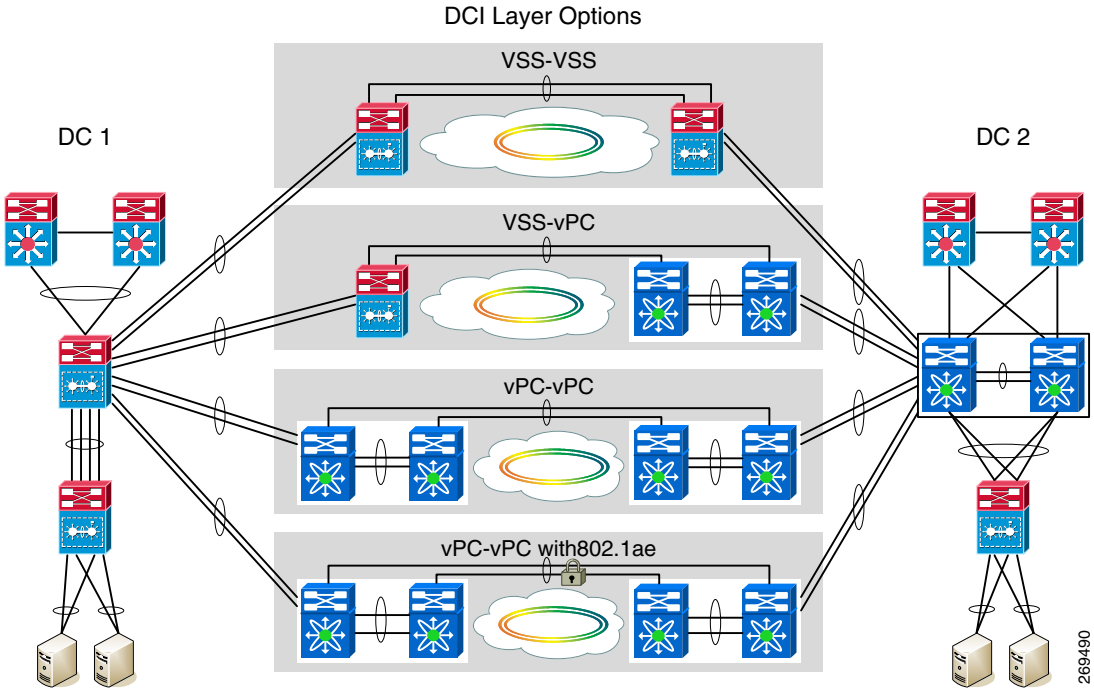
Note While the following content provides a summary of the test results as observed, detailed results are available in a separate document. Full configurations are also available in a separate document.

Dual-Site Test Topologies

As indicated above, there were four use cases covered with two data centers connected at Layer 2. As in all use cases, these two sites are connected by DWDM running over dark fiber. The ONS-15454 platform was leveraged to provide the optical connections between the data centers. Two 10 Gbps protected links were deployed to provide a 20 Gbps aggregate bandwidth between the data centers. For the most part, the optical transport was a “black box” as far as the testing was concerned, with only a small handful of tests taking place in that domain. Fiber spools were deployed in the test topology to provide a real fiber distance of 100 kilometers between the two data centers.

DCI Layer options for dual-site deployments are called out in the Figure 2-8 and correspond respectively to the first four use cases listed above. Each of these deployment options is covered in more detail in the material below.

Figure 2-8 Dual-Site Use Cases Summary View



Each of the two data centers leverages the logical layer model, where Core, Aggregation and Access are deployed separately in order to serve their discrete functions. The hardware and software details for these three layers, as deployed in testing, are common to all four use cases, and so described directly below. The Data Center Interconnect (DCI) Layer is where these four use cases differentiate from one another.

This differentiation is explained as these four use cases are explored in more depth below.

Core Layer—Hardware

In each data center, the Core Layer is composed of a pair of Catalyst 6500s using the following hardware (per switch):

- (1) WS-SUP720-3BXL
- (2) WS-X6704-10GE

Supervisor 720-3BXL was used to accommodate the larger routing tables that are commonly seen at this position in the network. The links connecting the Core Layer to the Aggregation Layer are 10 Gigabit Ethernet (GE) Layer 3 links. A single link of the same speed is connecting the Core switches to each other. There are 10 Gigabit Ethernet links connecting the Core switches to a traffic generator, also. The traffic profiles are discussed in more detail in another section below.

As illustrated in the diagram, there are two 10 GE links connecting each Core Layer switch to the Aggregation Layer. With VSS being employed at the Aggregation Layer in Data Center 1, this pair of switches is presented, logically, as a single switch. The two links from each Core switch, therefore, are bundled into a single logical link, using EtherChannel technologies.

The Core switches in Data Center 2 are connected to the Nexus 7000s in the Aggregation Layer in a similar fashion as described above for Data Center 1. The difference is that the vPC technology in Nexus 7000 can only be used to build port channels for Layer 2 links. Thus, the links connecting the Core Layer Catalyst 6500s to the Aggregation Layer Nexus 7000s in Data Center 2 were individual links, using Equal Cost MultiPathing (ECMP) to facilitate the load balancing.

Core Layer—Software

The complete software configurations for the solution as tested are available in a separate document. The following few paragraphs address, at a high level, the configurations that were used in the Core Layer of each data center in each tested topology.

All of the links emanating from the Core Layer are Layer 3 links. The links going north towards the Campus core (not pictured) are running BGP. The IXIA test tool, connected to a black box switch north of the Core Layer in each data center, injected 1000 BGP routes into the BGP domain for testing. BGP-to-OSPF redistribution of those routes was done in these Core Layer switches. 5000 OSPF routes were also injected into the Core Layer switches using the IXIA test tool.

The links both between the Core Layer switches and from them to the Aggregation Layer switches were running OSPF, and all links were in Area 0. The OSPF hello and dead timers for all links in all topologies were configured as 10 seconds and 40 seconds (default), respectively. OSPF SPF timers were set to default on all routers in all topologies. For BGP, the default timers were used.

BGP peer sessions were configured and established without authentication or encryption. OSPF authentication and encryption was configured on all OSPF-enabled devices in all topologies.

Aggregation Layer—Hardware

As mentioned above, the Aggregation Layers in the respective data centers, while serving the same function, are different in that they are built on different switching platforms. The Aggregation Layer in Data Center 1 is built on the Catalyst 6500 platform using the VSS technology to present the pair of physical switches as a single logical entity.

The hardware needed for this, and deployed in these switches, includes (per switch):

- (1) VS-S720-10G-3C
- (2) WS-X6708-10G-3C

Each of the inter-switch links (ISLs) connecting the Aggregation Layer switches to neighbor switches are 10 GE. The Multichassis EtherChannel (MEC) technology is used to bundle the links coming from the different physical chassis as a single logical EtherChannel link.

Not shown in the topology diagram is the Virtual Switch Link (VSL) necessary for VSS to operate. The VSL is composed of two 10 GE links bundled together using EtherChannel technology. Two links are used for two main reasons:

1. To provide redundancy for the critical VSS heartbeat messages and control plane messages sent between the two physical chassis.
2. To provide an aggregate bandwidth (20 Gbps here) large enough to accommodate data plane traffic that may need to pass between one chassis and the other, on its way to the destination. To provide further redundancy, the VSL links are split between the supervisor 10 GE uplink and one of the 10 GE line cards in the chassis.

Note Having one of the VSL links on the supervisor is critical for two reasons. First, it removes identical hardware as a single point of failure. For instance, even if the links were split between two WS-X6708-10G-3C line cards, in the event of an issue compromising that particular type of line card, the entire VSL would be compromised, leading to the dangerous dual active condition. The second reason to put one of the VSL links on the supervisor has to do with the timing of events following a single chassis reload. If both VSL links are on line cards, the rebooting chassis may see itself as VSS Active before the line cards and all links come fully online. By having one of the VSL links on the supervisor, the VSL is sure to come online first so that VSS roles can be set before other links come online.

In Data Center 2, the hardware used in each Nexus 7000 chassis was the same:

- (2) N7K-SUP1
- (2) N7K-M132XP-12

All the ISLs connecting these Aggregation Layer switches to other devices were 10 GE from the 32-port 10GE line cards. As with VSS and the VSL, there is a EtherChannel connecting the two Nexus 7000 chassis in the vPC pair called the Peer Link. For redundancy, the two links in this bundle are split across multiple line cards. For further protection from an accidental dual active state, there is another link between the two chassis called the Peer Keepalive Link (PKL). This link does not directly attach the two chassis. Rather, it is a separate Layer 3 connection.

Note While the out-of-band management network was used to build the PKL in the test topology, the best practice is that a separate Layer 3 link be used to build the PKL, unless only a single supervisor is available.

Aggregation Layer—Software

The complete software configurations for the solution as tested are available in a separate document. The following few paragraphs address, at a high level, the configurations that were used in the Aggregation Layer of each data center in each tested topology.

In traditional Cisco data center designs, the Aggregation Layer demarcates the boundary between Layer 2 and Layer 3. In the tested topologies, the Layer 3 links, running OSPF as the IGP, connect up to the Core Layer. The remaining links in the Aggregation Layer are Layer 2.

In those data centers where the Aggregation Layer was made up of two Catalyst 6500 switches, the Layer 2 link between these switches formed the VSL that helped to maintain the integrity of the VSS system. In the data center where the Aggregation Layer was made up of two Nexus 7000 switches, the Layer 2 link between these two switches formed the vPC Peer Link necessary for the vPC feature. For the VSL link that connected the two Catalyst 6500 switches in the VSS system, Port Aggregation Protocol (PAgP), in Desirable mode, was used to bundle the two links into a single port channel. For the Peer Keepalive link connecting the Nexus 7000 switches, Link Aggregation Control Protocol (LACP) was used, with the ports in Active mode.

Due to the virtual nature of either two Catalyst 6500 or two Nexus 7000 switches combined using either VSS or vPC, the single links that connect the Aggregation Layer to the Access Layer are bundled into one logical Multichassis EtherChannel (MEC). This is configured just as a normal EtherChannel, using either PAgP with the member interfaces in desirable mode (in the case of two Catalyst 6500s being connected) or Link Aggregation Control Protocol (LACP) with the members in active mode (in the case of a Catalyst 6500 connecting to a Nexus 7000).

The same is true of the MEC links connecting the Aggregation Layer to whatever DCI Layer option was being tested at the time.

Aside from OSPF, the other Layer 3 protocol running at the Aggregation Layer was Hot Standby Router Protocol (HSRP) which provided gateway redundancy to hosts in a given subnet. Of the 500 Layer 2 VLANs that were configured in the Layer 2 domain that was extended between sites in the 2-site or 3-site test topologies, 100 were mapped to Layer 3 interfaces, or SVIs. These SVIs were all configured with HSRP, sharing a common group across all data centers in a given topology.

Since multiple devices were participating in a single HSRP group, the configured HSRP priorities were manipulated such that the active HSRP router was the Aggregation Layer VSS device in DC 1. One of the vPC devices in the Aggregation Layer of DC 2 was the standby HSRP router for each HSRP group while the second Nexus 7000 remains in HSRP Listening state. The configuration is the same with the third site present in the topology, with the VSS device in that third site also waiting in Listening state.

The Layer 2 VLANs are trunked from the Aggregation Layer switches to the Access Layer switches using 802.1q VLAN tagging. The trunking mode is set to on. For the entire Layer 2 domain, the VTP mode is configured as transparent.

Rapid spanning-tree protocol (rPVST+) is configured on all Layer 2 devices, including at the Aggregation Layer. Since the Aggregation, Access and DCI Layers are all either VSS or vPC, though, spanning-tree is not blocking any ports as there are no Layer 2 loops to prevent. The STP Root switch was configured as the VSS switch in DC 1, while the Secondary Root switch was the first vPC switch in DC 2.

Fast Hello was selected as the preferred VSS Dual Active detection mechanism. This is because Fast Hello has been identified as the most reliable and robust Dual Active detection mechanism. VSS chassis and VSL link failure events with the Fast Hello mechanism enabled yielded sub-second convergence times for failure and full recovery. Refer to Table 2-2 for test result details.

Access Layer—Hardware

The Access Layer in each of the two data centers consists of a VSS pair of Catalyst 6500 switches. The hardware configuration of each of these Access Layer switches is as follows.

- (1) VS-S720-10G-3C
- (2) WS-X6748-GE-TX
- (2) WS-X6708-10GE-3C

The 10 Gigabit Ethernet line cards provide the uplinks to the Aggregation Layer switches, as well as one of the two links bundled into the VSL. The supervisor uplinks provide the other link for the VSL bundle. The Gigabit Ethernet ports on the 48-port line cards provide server access connectivity.

Access Layer—Software

The Access Layer was not really an integral part of release 1.0 testing, and details on the configurations used at this layer can be found in the separate results document. In brief, the Access Layer devices were strictly Layer 2 as used in testing. The spanning-tree, Layer 2 trunk and EtherChannel configurations on the Catalyst 6500 devices in this layer were similar to those same features as deployed at the Aggregation Layer.

VSS-to-VSS DC Interconnect

The first use case tested and deployed is illustrated above in Figure 2-2. In this use case, data center inter connectivity is accomplished by a pair of Catalyst 6500 switches using VSS at the DCI Layer of either data center. The MEC between them is built across the optical transport (DWDM in this case) so that a full 20 Gbps of bandwidth exists to connect the two data centers.

DCI Layer—Hardware

The Catalyst 6500s deployed at the DCI Layer in each of the two data centers have similar hardware deployments, as listed below:

- VS-S720-10G-3C
- WS-X6708-10G-3C

As mentioned, the links between the VSS pairs in each data center are 10 Gigabit Ethernet, and bundled into a MEC. Likewise, the links from the DCI Layer VSS switches to the Aggregation Layer VSS switches are also 10 GE and bundled into a MEC. As per best practices, the respective VSLs are built with dual-10 GE links distributed between the supervisor and one line card.

Figure 2-2 illustrates clearly that with VSS at the Access, Aggregation and DCI Layers in each data center, spanning-tree is essentially eliminated, as is unused bandwidth. These concepts of STP elimination and full link utilization are discussed in earlier sections of this document.

DCI Layer—Software

In the VSS-to-VSS test topology, the VSS devices in either data center were strictly Layer 2. These devices connected to each other across the DWDM ring in order to extend the Layer 2 domain from one data center to the other. From the perspective of most Layer 2 protocols, then, the configuration at the DCI Layer was the same as at either the Aggregation or Access Layers.

There are some important configurations to point out at the DCI Layer, though, and these will apply to all of the 2-site topologies that were tested.

Outside of the test cases used to validate the system fault isolation capabilities, the spanning-tree domain extended between the two data centers, along with the 500 VLANs. That means that there was a single STP root for both data centers (it was DC2-AGG-7K-1 in all 2-site use cases) and a single STP secondary root (was DC1-AGG-6K).

For the fault isolation test cases, the larger STP domain was split into two through the use of BPDU Filtering. For this test case, a Layer 2 loop, and subsequent broadcast storm, was introduced in one data center while the other data center was monitored to see if it was impacted from the storm.

By configuring BPDU Filtering on those DCI Layer interfaces that connected the switches to the DWDM ring, the spanning-tree domain was separated between the two data centers. In this case, the Aggregation Layer VSS switch was the primary STP root in DC 1. One of the Nexus 7000 Aggregation Layer vPC switches was the primary root in DC 2 while the other Nexus 7000 was the secondary STP root.

To prevent the induced broadcast storm from propagating to the other data center in the Layer 2 domain, storm control was enabled on the DCI Layer interfaces connecting these switches to the Aggregation Layer.

VSS-to-vPC DC Interconnect

The second use case tested and deployed is illustrated in Figure 2-3. Here, one data center uses a pair of Catalyst 6500 switches running VSS at the DCI Layer, while the other data center leverages the vPC functionality on the Nexus 7000 platform. While VSS and vPC are different in their ability to combine multiple physical switches into a single logical unit, both support the concept of Multichassis EtherChannel, and thus can be used to connect the two sites using this commonality. Again, the MEC between the two data centers is built across the optical transport so that a full 20 Gbps of bandwidth is available.

DCI Layer—Hardware

The Catalyst 6500s deployed at the DCI Layer of Data Center 1 is the same as in the previously-covered VSS-to-VSS use case, as listed below:

- (1) VS-S720-10G-3C
- (2) WS-X6708-10G-3C

The Nexus 7000s deployed at the DCI Layer of DC 2 are configured with the following hardware:

- (2) N7K-SUP1
- (2) N7K-M132XP-12

While the platforms are different in each data center, the connectivity is fairly similar, with 10 GE links being used everywhere. The exception is the management link that is used to create the Layer 3 Peer Keepalive Link between the two Nexus 7000 systems. Though the technology being used at the DCI Layer in DC 2 is vPC on the Nexus 7000 platform, it delivers on the same goals of STP elimination and full link utilization through its support of the MEC technology.

DCI Layer—Software

The software features configured at the DCI Layer in this use case were identical to those outlined in the VSS-VSS 2-site test topology. Please refer to that section for more information.

vPC-to-vPC DC Interconnect

The third use case tested and deployed is illustrated above in Figure 2-4. In this use case, both data centers use a pair of Nexus 7000s and the vPC technology at the DCI Layer. 20 Gbps is facilitated between the data centers across the dark fiber using the MEC capabilities of the Nexus 7000 in either data center.

DCI Layer—Hardware

The Nexus 7000s deployed at the DCI Layer of DC 2 are configured with the following hardware:

- (2) N7K-SUP1
- (2) N7K-M132XP-12

Similar to other Nexus 7000 vPC configurations covered already in the testing discussion, 10 GE links are used everywhere, including the Peer Link. The Peer Keepalive Link is, again, provided through a separate Layer 3 link.

DCI Layer—Software

The software features configured at the DCI Layer in this use case were identical to those outlined in the VSS-VSS 2-site test topology. Please refer to that section for more information.

vPC-to-vPC DC Interconnect with 802.1AE Encryption

The final 2-site use case tested and deployed is illustrated in Figure 2-5. The only difference between this use case deployment and the vPC-to-vPC deployment discussed previously is that the MEC links that connect the two DCI Layer vPC pairs are encrypted using 802.1AE (MACsec) encryption, as facilitated by Cisco TrustSec on the Nexus 7000.

Note As part of the testing effort it was discovered that for a particular traffic pattern, drops could be observed that resulted in 802.1AE re-keying to fail. Therefore test results are not reported in this version of this document and will be reported as soon as the issue is root caused. This is being tracked under CSCtb34740.

DCI Layer—Hardware

There are no changes to the hardware configuration in this topology from that of the vPC-to-vPC topology. The components used are listed.

- (2) N7K-SUP1
- (2) N7K-M132XP-12

DCI Layer—Software

The software features configured at the DCI Layer in this use case were identical to those outlined in the vPC-vPC 2-site test topology. Please refer to that section for more information on those common Layer 2 features deployed.

In this use case there is a key difference. Here, 802.1AE encryption is configured on the MEC links connecting the two data centers. This means that any data traversing the Layer 2 path between the two data centers was encrypted.

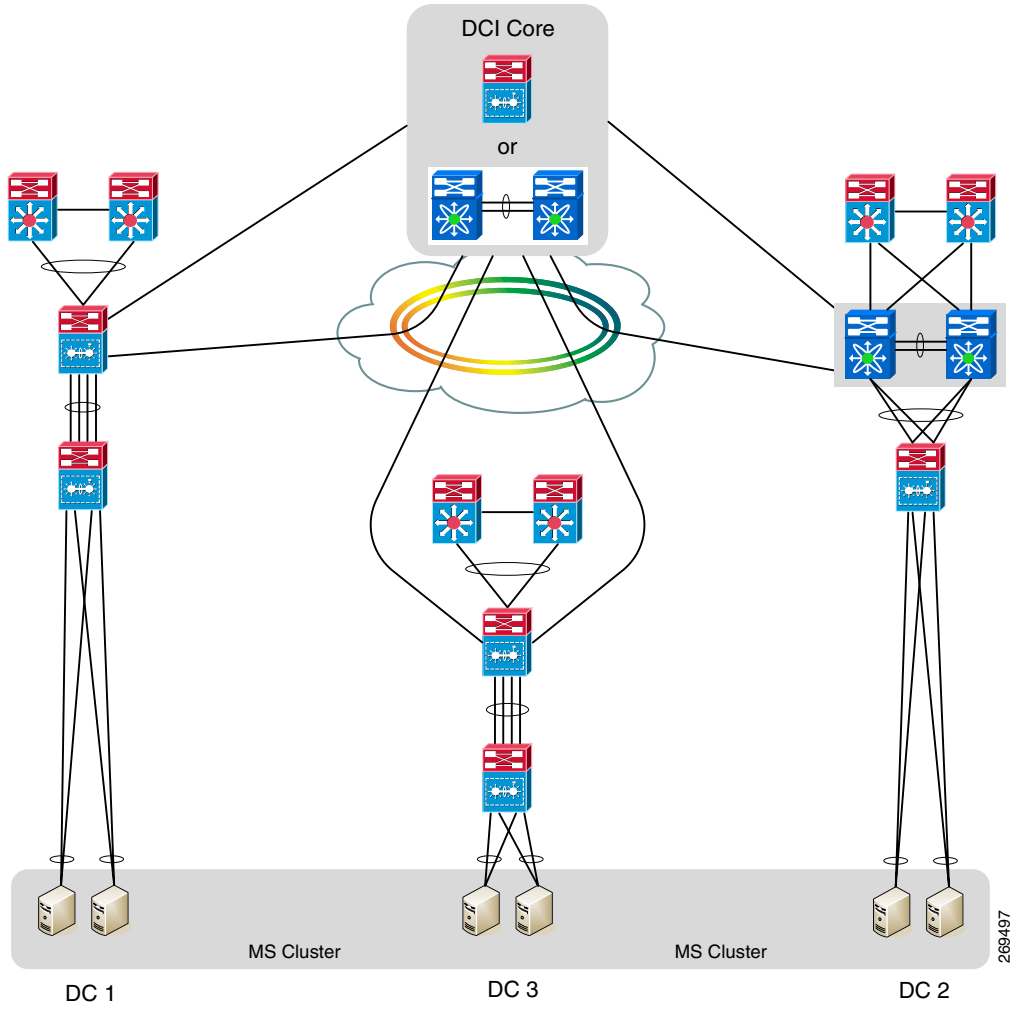
Multi-Site Test Topologies

The final two use cases were focused on connecting three or more data centers (three were validated through testing, but the results should be extensible to more) are connected for purposes of extending the Layer 2 domain between them.

To accomplish this, VSS and vPC technologies on the Catalyst 6500 and Nexus 7000 were leveraged again. This time, they were used to create logical star topologies, with either a VSS or vPC pair at the center of the star and the data centers forming the arms of the star.

Figure 2-9 illustrates the summary view of the topology used to test the two multi-site use cases. Within the DCI Layer, the option to use either a VSS or vPC core switch pair is given.

Figure 2-9 Multi-Site Use Cases Summary View



Of course, in order to have either a VSS or vPC switch pair at the center, or "core," of the star topology, a VSL or a Peer Link must be built to enable communication between the two switches. In theory, the two switches could reside anywhere. One option is to have both of them located physically at a single site. Another option is to have both of them located at some non-data center site (the least practical). The best option, though, is to have the location of the two switches split between two of the sites. By not

having them co-located, the single point of failure is eliminated and disaster recovery abilities are enhanced. Having them located at different sites, though, requires that the VSL or Peer Link be built between the two sites, at a distance, using the dark fiber.

In the testing that was done in support of this data center interconnect solution, the two sites that held the physical DCI Layer VSS or vPC switches were 100 kilometers apart (the distance generated using fiber spools). These two data centers were each 50 kilometers from the third data center.

The connections from the logically central switch pair to the respective data center Aggregation Layers are made through use of the MEC technology. Once again, since VSS or vPC is used in this multi-site topology at all Aggregation, Access and DCI layers, spanning-tree is functionally eliminated and we have achieved the goal of utilizing all links.

The links from the Aggregation Layer switches to the VSS or vPC "core" switch pair may be either directly connected, via short-range SR optics or connected over a great distance through the ONS-based optical ring, depending on the proximity of the Aggregation switch to the physical DCI Layer switch.

Other than the switches comprising the DCI "core" switch, the topologies used to validate the two multi-site use cases were exactly the same. The sections below delve deeper into the configurations of these two topologies.

Figure 2-10 shows the multi-site test topology with a VSS core. It can be referenced to illustrate the hardware at each of the three data center sites.

At the Core, Aggregation and Access Layers, DC 1 and DC 2 are the exact same topologies as used in the 2-site testing. They are built with the following hardware.

DC 1 & 2 Core Layer (per switch)

- (1) WS-SUP720-3BXL
- (2) WS-X6704-10GE

DC 1 & 2 Aggregation Layer (per switch)

- (1) VS-S720-10G-3C
- (2) WS-X6708-10G-3C (or)
- (2) WS-X6716-10G-3C

DC 1 & 2 Access Layer (per switch)

- (1) VS-S720-10G-3C
- (2) WS-X6748-GE-TX
- (2) WS-X6708-10GE-3C

As in the dual site test topology, all inter-switch links are 10 Gigabit Ethernet. All VSLs and the vPC Peer Link are 10 GE as well. The vPC also has the 1 GE Layer 3 Peer Keepalive Link.

The differences between the two use cases are called out in the sections below.

DC Interconnect with VSS Core

In this use case, the three data centers connect into the pair of Catalyst 6500s that form the DCI VSS "core" via multiple 10 GE links between the DCI switches and the Aggregation Layer. Figure 2-6, shown above, provides a view of this test topology.

DCI Layer—Hardware

The hardware at the DCI Layer in the multi-site with VSS core topology reflects what is used for other VSS pairs in this testing. One of the VSS switches lives in DC 1 and the other in DC 2. Each of the VSS switches has the following hardware.

- (1) VS-S720-10G-3C
- (2) WS-X6708-10GE-3C

The two switches are connected to each other (with VSL) across the dark fiber DWDM ring. Each switch has a single 10 GE connection into a single Aggregation Layer switch in each data center, meaning there are two 10 GE links from each DC Aggregation Layer to this VSS core. Since the switch pairs at the Aggregation Layer are either VSS or vPC pairs, the MEC connecting them to the VSS core is treated as a single logical link.

DCI Layer—Software

Though the physical and logical topologies between this 3-site use case and the 2-site use cases explored earlier differ, they are functionally the same at the DCI Layer. In each case, the DCI Layer exists to facilitate extending the Layer 2 domain between data centers. As such, the software features that were in use in the 2-site test topologies are also in use here.

In the extended Layer 2 domain, outside of the fault isolation test cases, the STP root is shared between the multiple data centers. The primary and secondary root switches are the same as in the 2-site topologies, with one of the vPC Nexus 7000 switches in DC 2 being the primary and the Aggregation Layer VSS switch in DC 1 being secondary.

In the fault isolation test cases, BPDU filtering was configured on the DCI Layer interfaces that faced the Aggregation Layer devices. This broke the single STP domain into three (four, including the VSL link), each with its own STP primary and secondary root switches. Storm control was used at the Aggregation Layer DCI-facing interfaces to prevent broadcast propagation between data centers, as in the 2-site test topologies.

DC Interconnect with vPC Core

In this use case, the three data centers connect into a pair of Nexus 7000 switches that form the DCI vPC “core” via multiple 10 GE links between the DCI switches and the Aggregation Layer. Figure 2-7, shown above, provides a view of this test topology.

DCI Layer—Hardware

The hardware at the DCI Layer in the multi-site with vPC core topology was consistent to what was used for the other vPC pairs in this testing. As with the multi-site VSS core topology, one of the vPC switches lives in DC 1 and the other in DC 2. Each of the vPC switches has the following hardware.

- (2) N7K-SUP1
- (2) N7K-M132XP-12

The two switches are connected to each other (via Peer Link) across the dark fiber DWDM ring. Each switch has a single 10 GE connection into a single Aggregation Layer switch in each data center, meaning there are two 10 GE links from each DC Aggregation Layer to this vPC core. Since the switch pairs at the Aggregation Layer are either VSS or vPC pairs, the MEC connecting them to the vPC core is treated as a single logical link.

DCI Layer—Software

Even with a pair of Nexus 7000 switches at the core of this 3-site test topology, there is no difference in configuration, functionally, from the VSS-core 3-site topology described above. BPDU Filtering and storm control are applied similarly, when and where needed to achieve the goals of fault isolation.

Testing Methodology

The testing performed to validate the six use cases was systems-level testing, meaning that it took into account the behavior of the entire system (the entire test topology) as a given test was being executed. This includes monitoring traffic for correct transmission through the network and monitoring the resources of the individual switches in the test topology for errors.

The testing performed was also focused on the entire system in that it assumed that discrete features such as EtherChannel or HSRP or OSPF already worked as designed. Instead of validating these individual features, the testing aimed to validate the stability and resiliency of the system as a whole.

The goal of the test setup was to achieve, as best possible within given constraints, an environment closely resembling a real-world deployment. This was accomplished through the use of selective configuration and traffic scaling, a mixture of real-world traffic profiles and realistic network device deployments.

Each test case was begun with specific pass or fail criteria already defined. For the component failure test cases, the criteria included a convergence time target for both the failure and recovery events. For the fault isolation test cases, the criteria for a successful run was that the fault would be isolated to a particular data center, and not compromise the other data centers connected on the same Layer 2 domain.

Test Scope

As mentioned above, all tests used to validate the six use cases were negative in nature, meaning they introduced some non-normal event into the system and measured the system's recovery from that event.

The majority of these negative tests were based on various component failures within the system. There were link failures, line card failures and full chassis/supervisor failures. Some of these component failures were less intrusive than others, due to various degrees of redundancy built into the network system. The system converged rather easily, for example, around single link failures, compared to full line card failures in which many links were impacted at one moment. Other events were more destructive. Some tests compromised the entire VSL link in a VSS pair causing what is called a dual-active condition in which each of the Catalyst 6500s to believe itself to be the VSS Active switch. Such a condition can wreak havoc on a network system unless proper safeguards are built in.

Another type of negative test case purposely caused a Layer 2 loop in a given data center. Such a loop would cause a broadcast storm that would, in turn, consume a bulk of forwarding resources on the switches in that data center. With each of the multiple data centers a part of the same Layer 2 domain, there is potential for the broadcast storm to propagate to the other data center(s). Such a fault could "take out" every data center in a customer's extended system and make critical business services unavailable. Therefore, this type of test would aim to validate the safeguards put in place to keep such a fault isolated to a single data center.

Test Tools and Traffic Profile

For each of the use cases, whether dual site or multi-site, there were two basic types of traffic that were injected into the test topology. For traffic flows that went from data center to data center (as if server-to-server communication), an EMIX (Enterprise Mix) traffic profile was used. For traffic flows that stayed within a single data center, traversing between Core Layer and Access Layer (as if campus client-to-server communication), an IMIX (Internet Mix) traffic profile was used. A high-level view of this traffic setup is provided in Figure 2-10 and Figure 2-11.

Figure 2-10 High-Level View of 2 Site Traffic Profile

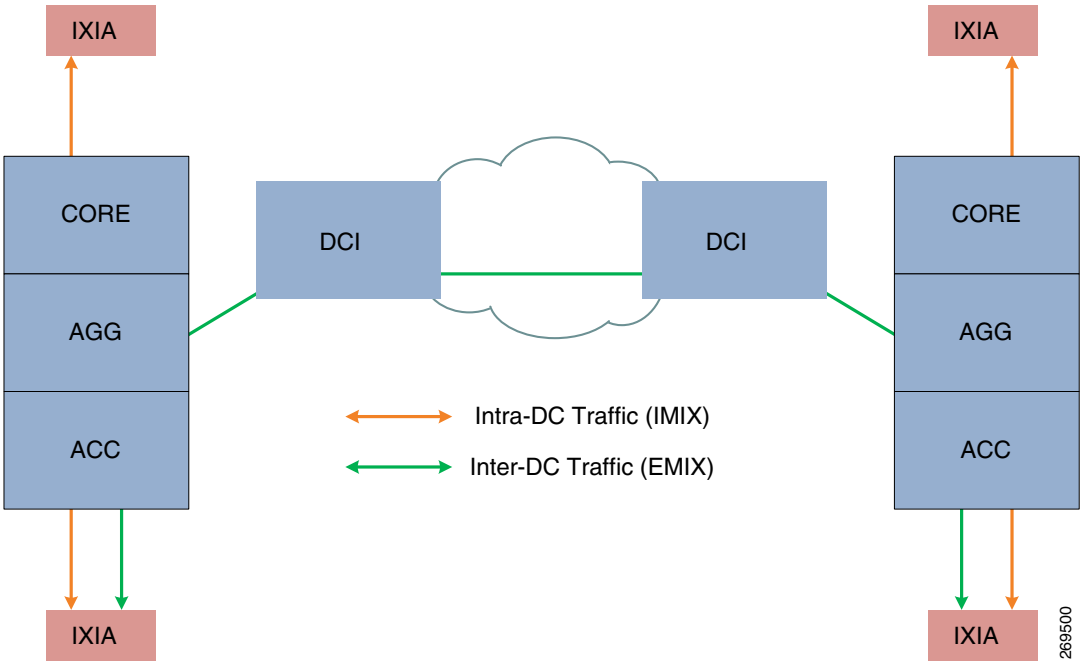
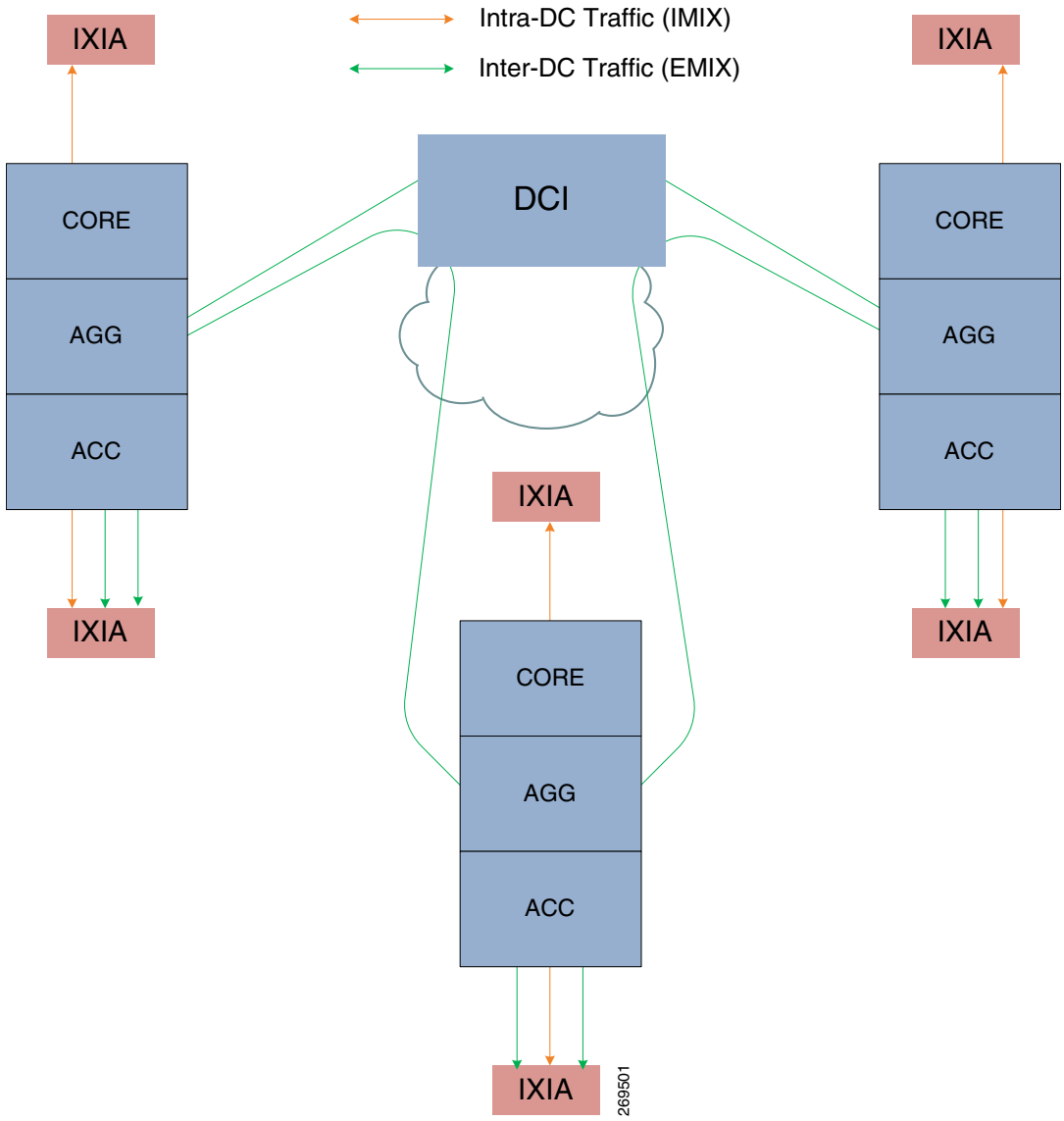


Figure 2-11 High-Level View of 3 Site Traffic Profile



The IXIA test tool was used to generate all of the traffic used in this testing. All of the traffic was Layer 2 and Layer 3 stateless traffic.

Note A more complete and detailed view of the test traffic profile can be seen in the complete test results document.

Test Consistencies

For each test case executed, a series of steps preceded and followed. At the beginning of each test, two scripts were executed. One script would monitor and log the CPU and memory utilization statistics of each network device in the entire test topology using SNMP. The other script would take a snapshot of the inband health of each device, looking at the Ethernet Out-of-Band Channel (EOBC) statistics and the stats for specific intra-chassis communication protocols. Upon completion of the test procedure, the SNMP script collecting the CPU and memory statistics would be stopped and graphs would be generated illustrating CPU and memory health of each device for the duration of the test. Also, the inband health script would take a second snapshot and compare it to the results of the first run of that script. Any unexpected increase in error counters or other indication of compromised system health would be flagged for review.

Test Convergence Results

Table 2-2 through Table 2-6 provide detailed summaries of the Data Center Interconnect System Release 1.0 test convergence results for dual site and multi-site testing. Each table provides results for a different use case and they are broken down further by test type. The following dual-site and multi-site use cases are listed below:

Dual-Site

1. Dual-Site VSS-VSS Test Results, page 2-26
   - VSS-VSS Hardware Failure
   - VSS-VSS Link Failure
2. Dual-Site VSS-vPC Test Results, page 2-27
   - VSS-vPC Fault Isolation
   - VSS-vPC Hardware Failure
   - VSS-vPC Link Failure
3. Dual-Site vPC-vPC Test Results, page 2-30
   - vPC-vPC Hardware Failure
   - vPC-vPC Link Failure

Multi-Site

1. Multi-Site with VSS Core Test Results, page 2-35
   - DCI Release 1.0 Multi-Site with VSS Core-HW Failure
   - DCI Release 1.0 Multi-Site with VSS Core-Link Failure
   - DCI Release 1.0 Multi-Site with VSS Core-Fault Isolation
2. Multi-Site with vPC Core Test Results, page 2-40
   - DCI Release 1.0 Multi-Site with vPC Core-HW Failure
   - DCI Release 1.0 Multi-Site with vPC Core-Link Failure
   - DCI Release 1.0 Multi-Site with vPC Core-Fault Violation

Figure 2-12, Figure 2-13, and Figure 2-14 can be used as references when reading these results. These figures provide device name and interface numbers that correlate to the specific failure activity outlined in the Test Details column.

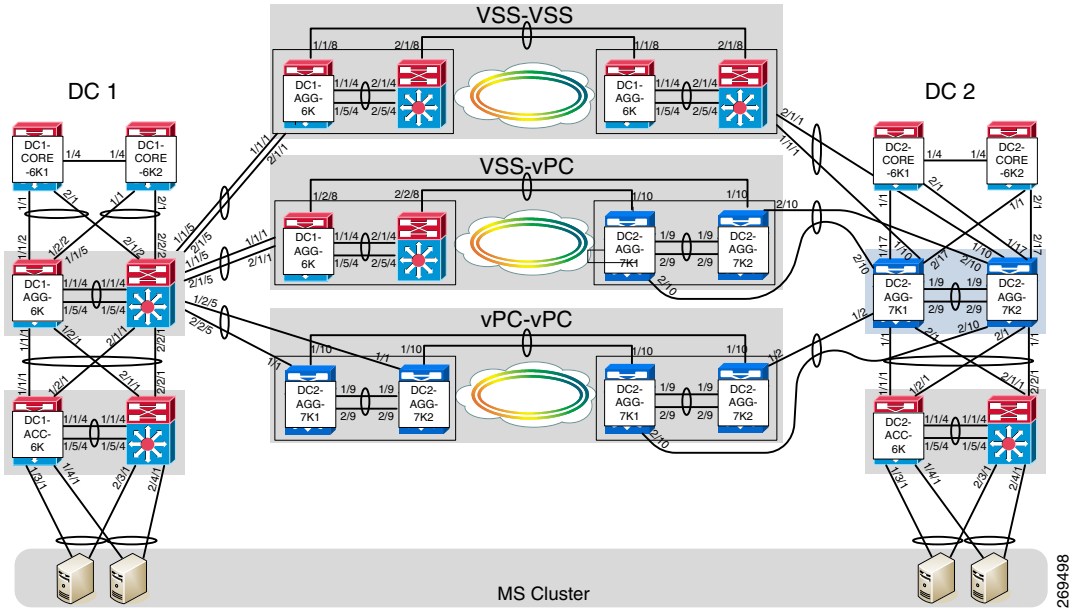
Note There are footnotes below the results table that point out specific issues encountered during testing. These footnotes speak to unresolved issues only.

For complete test results, including test procedures and a list of all issues (resolved and unresolved) encountered during testing, refer to the Cisco Data Center Interconnect Test Results System Release 1.0 document.

Dual-Site Testing

When reviewing results given in Table 2-2, Table 2-3, and Table 2-4, refer to Figure 2-12.

Figure 2-12 Dual-Site Test Topology Details



Dual-Site VSS-VSS Test Results

Table 2-2 provides a detailed test results summary for DCI dual-site VSS-to-VSS testing by test type.

Table 2-2 2-Site VSS-to-VSS Test Results

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
VSS-VSS Hardware Failure						
DC1 DCI 6k Active Chassis Failure	Powered down VSS active chassis for failure. Powered up VSS active chassis for restoration.	0.693s	0.968s	0.440s	2.791s	Pass
DC1-DCI-6K Linecard Failure WS-X6708-10GE-3c slot 1	Physically ejected DC1-DCI-6K1's WS-X6708-10GE-3c in slot 1 for failure. Module reinserted for restoration.	0.179 s	0.322 s	0.037 s	0.077 s	Pass
DC2-DCI-6K Active Chassis Failure	Powered down VSS active chassis for failure. Powered up VSS active chassis for restoration.	1.010s	1.026s	1.116s	2.310s	Pass
DC2-DCI-6K Linecard Failure WS-X6708-10GE-3c slot 2	Physically ejected DC2-DCI-6K1's WS-X6708-10GE-3c in slot 2 for failure. Module reinserted for restoration.	1.727s	2.339s	0.226s	2.084s	Pass
VSS-VSS Link Failure						
DC1-DCI-6K Link Failure 2/1/4 of DC1-DCI-6k	Physically disconnected the cable connected to DC1-DCI-6K1's interface TenGig2/1/4 for failure. Link reconnected for restoration.	0.0s	0.016s	1.764s	0.0s	Pass
DC2-DCI-6K Link Failure 1/1/4 of DC2-DCI-6k	Physically disconnected the cable connected to DC2-DCI-6K1's interface TenGig1/1/4 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass

Table 2-2 2-Site VSS-to-VSS Test Results (continued)

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
DC2-DCI-6k Whole VSL Link Failure 1/1/4 and 2/1/4 of DC2-DCI-6k1.	Physically disconnected both of DC2-DCI-6k1's VSL links, TenGig 1/1/4 and 2/1/4 for failure. Both links reconnected for restoration.	1.360s	1.198s	0.790s	1.133s	Pass
DWDM Failure-Fiber Cut b/t DC	The active link between ONS 1 and ONS 2 was physically disconnected for failure, and reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass

Dual-Site VSS-vPC Test Results

Table 2-3 provides a detailed test results summary for DCI dual-site VSS-to-vPC testing by test type.

Table 2-3 2-Site VSS-to-vPC Test Results

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
VSS-vPC Fault Isolation						
DC1 Loop and Verify Fault Isolation	A spanning tree loop was created on DC1-ACC-6k by connecting an ethernet cable to interfaces Gig 1/3/48 and 2/3/48 with BPDU filtering enabled on each port. Storm control was configured on DC1-DCI-6k's Po10 physical interfaces to mitigate and constrain the event's impact to Datacenter 1.	N/A	N/A	N/A	N/A	Pass

Table 2-3 2-Site VSS-to-vPC Test Results (continued)

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
DC2 Loop and Verify Fault Isolation	A spanning tree loop was created on DC2-ACC-6k by connecting an ethernet cable to interfaces Gig 1/3/48 and 2/3/48 with BPDU filtering enabled on each port. Storm control was configured on DC1-DCI-6k's Po10 physical interfaces to mitigate and constrain the event's impact to Datacenter 2.	N/A	N/A	N/A	N/A	Pass
VSS-vPC Hardware Failure						
DC2-DCI-7K1 Line Card Failure slot 1 of DC2-DCI-7k1	Physically ejected DC2-DCI-7K1's N7K-M132XP-12 in slot 1 for failure. Module reinserted for restoration.	0.893s	1.009s	1.230s	1.170s	Pass
DC2-DCI-7K2 Line Card Failure slot 2 of DC1-DCI-7k2.	Physically ejected DC2-DCI-7K2's N7K-M132XP-12 in slot 2 for failure. Module reinserted for restoration.	0.148s	0.264s	0.034s	0.290s	Pass
DC1 DCI 6k Active Chassis Failure	Powered down VSS active chassis for failure. Powered up VSS active chassis for restoration.	0.574s	0.956s	1.176s	2.60s	Pass
DC1 DCI 6k Linecard Failure 6708 slot 1 of DC1-DCI-6k	Physically ejected DC2-DCI-7K1's N7K-M132XP-12 in slot 1 for failure. Module reinserted for restoration.	1.262s	1.766s	0.160s	0.218s	Pass
DC2-DCI-7K1 Chassis Failure	Powered down DC2-DCI-7K1 for failure, powered back up for restoration.	0.497s	0.791s	1.968s	1.777s	Pass

Table 2-3 2-Site VSS-to-vPC Test Results (continued)

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
DC2-DCI-7K2 Chassis Failure	Powered down DC2-DCI-7K2 for failure, powered back up for restoration.	0.507s	0.817s	1.456s	1.969s	Pass
VSS-vPC Link Failure						
DC2-DCI-7K vPC Complete Peer Link Failure DC2-DCI-7k1: e 1/9, e2/9	Physically disconnected both of DC2-DCI-7k1f's vPC peer links, Ethernet 1/9 and 2/9 for failure. Both links reconnected for restoration.	1.233s	1.516s	1.477s	1.443s	Pass
DC2-DCI-7K vPC Peer Keepalive Link Failure dc2-dci-7k1 shut down mgmt 0	Physically disconnected the cable connected to DC2-DCI-7K1's interface Management 0 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass
DC2-DCI-7K vPC MEC Single Link Failure DC2-DCI-7K1 shut e 1/10	Physically disconnected the cable connected to DC2-DCI-7K1's interface Ethernet 1/10 for failure. Link reconnected for restoration.	0.473s	0.987s	0.295s	N/A [1]	Pass
DC2-DCI-7K1 Link Failure 1/9 of DC2-DCI-7k1.	Physically disconnected the cable connected to DC2-DCI-7K1's interface Ethernet 1/9 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass
DC2-DCI-7K2 Link Failure e 2/9 of DC2-DCI-7k2.	Physically disconnected the cable connected to DC2-DCI-7K2's interface Ethernet 2/9 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass

Table 2-3 2-Site VSS-to-vPC Test Results (continued)

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
DC1-DCI-6k Link Failure 2/1/4 of DC1-DCI-6k1.	Physically disconnected the cable connected to DC1-DCI-6k1's interface TenGig 2/1/4 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass
DC1-DCI-6K Whole VSL Link Failure 1/1/4 and 2/1/4 of DC1-DCI-6k1.	Physically disconnected both of DC1-DCI-6k1's VSL links, TenGig 1/1/4 and 2/1/4, for failure. Both links reconnected for restoration.	0.526s	1.034s	0.692s	0.619s	Pass
DWDM Failure-Fiber Cut b/t DC	The active link between ONS 1 and ONS 2 was physically disconnected for failure, and reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass

1. These results were not available.

Dual-Site vPC-vPC Test Results

Table 2-4 provides a detailed test results summary for DCI dual-site vPC-to-vPC testing by test type.

Table 2-4 2-Site vPC-to-vPC Test Results

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
vPC-vPC Hardware Failure						
DC1-DCI-7K1 Chassis Failure	Powered down DC1-DCI-7K1 for failure, powered back up for restoration.	1.528s	1.546s	0.238s	0.222s	Pass
DC1-DCI-7K1 Linecard Failure slot 1 of DC1-DCI-7k1	Physically ejected DC1-DCI-7k1's N7K-M132XP-12 in slot 1 for failure. Module reinserted for restoration.	0.635s	0.808	2.176s	2.373s	Pass

Table 2-4 2-Site vPC-to-vPC Test Results (continued)

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
DC1-DCI-7K2 Chassis Failure	Powered down DC1-DCI-7K2 for failure, powered back up for restoration.	0.537s	0.828s	2.542s	1.327s	Pass
DC1-DCI-7K2 Linecard Failure slot 2 of DC1-DCI-7k2.	Physically ejected DC1-DCI-7k2's N7K-M132XP-12 in slot 2 for failure. Module reinserted for restoration.	0.677 s	0.966 s	2.80s	2.537s	Pass
DC2-Agg-7k1 Chassis Failure	Powered down DC2-Agg-7k1 for failure, powered back up for restoration.	0.504s	0.140s	0.378s	0.323s	Pass
DC2-Agg-7k2 Chassis Failure	Powered down DC2-Agg-7k2 for failure, powered back up for restoration.	0.138s	0.436s	0.373s	0.161s	Pass
DC2-DCI-7K1 Chassis_Failure	Powered down DC2-DCI-7K1 for failure, powered back up for restoration.	1.528s	1.546s	0.238s	0.222s	Pass
DC1-DCI-7K1 Linecard_Failure slot 1 of DC2-DCI-7k1.	Physically ejected DC1-DCI-7k1's N7K-M132XP-12 in slot 1 for failure. Module reinserted for restoration.	0.635s	0.808s	2.176s	2.373s	Pass
DC1-DCI-7K2 Chassis_Failure	Powered down DC1-DCI-7K2 for failure, powered back up for restoration.	0.381s	0.192s	0.381s	0.192s	Pass
DC2-DCI-7K2 Linecard_Failure slot 2 of DC2-DCI-7k2	Physically ejected DC2-DCI-7K2's N7K-M132XP-12 in slot 2 for failure. Module reinserted for restoration.	0.491s	0.507s	0.023s	0.013s	Pass
vPC-vPC Link Failure						

Table 2-4 2-Site vPC-to-vPC Test Results (continued)

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
DC1-Agg-6k Single VSL Link Failure 2/1/4 of DC1-Agg-6k1	Physically disconnected the cable connected to DC1-Agg-6k1's interface TenGig2/1/4 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass
DC1-DCI-7K vPC Peer Keepalive Link Failure	Physically disconnected the cable connected to DC1-DCI-7K1's interface Management 0 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass
DC1-DCI-7K vPC Peer Keepalive Link and Peer Link Failure	Physically disconnected the cables connected to DC1-DCI-7K1's peer keepalive interface, Management 0, and entire vPC Peer Link, Ethernet 1/9 and 2/9 for failure. Links reconnected for restoration.	sustained loss	sustained loss	N/A	N/A	Fail [1]
DC1-DCI-7K-vPC Single Peer Link Failure 1/9 of DC1-DCI-7k1.	Physically disconnected the cable connected to DC1-DCI-7K1's interface Ethernet 1/9 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass
DC1-DCI-7K1 Link Failure 1/9 of DC1-DCI-7k1.	Physically disconnected the cable connected to DC1-DCI-7K1's interface Ethernet 1/9 for failure. Link reconnected for restoration.	1.310s	0.0s	0.0s	0.0s	Pass

Table 2-4 2-Site vPC-to-vPC Test Results (continued)

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
DC1-DCI-7K2 Link Failure 2/9 of DC1-DCI-7k2.	Physically disconnected the cable connected to DC1-DCI-7K2's interface Ethernet 2/9 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass
DC2-DCI-7K vPC Peer Keepalive Link Failure dc2-dci-7k disconnect mgmt 0	Physically disconnected the cable connected to DC2-DCI-7K1's interface Management 0 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass
DC2-DCI-7K vPC Peer Keepalive Link and Peer Link Failure	Physically disconnected the cables connected to DC2-DCI-7K1's peer keepalive interface, Management 0, and entire vPC Peer Link, Ethernet 1/9 and 2/9 for failure. Links reconnected for restoration.	sustained loss	sustained loss	N/A	N/A	Fail [2]
DC2-DCI-7K vPC Single Peer Link Failure DC2-DCI-7K1 shut 1/9.	Shut down DC2-DCI-7K1's interface Ethernet 1/9 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass
DC2-DCI-7K1 Link Failure DC2-DCI-7K1 shut 1/9.	Physically disconnected the cable connected to DC2-DCI-7K1's interface Ethernet 1/9 for failure. Link reconnected for restoration.	0.066s	0.012s	0.0s	0.0s	Pass

Table 2-4 2-Site vPC-to-vPC Test Results (continued)

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
DC2-DCI-7K2 Link Failure dc2-dci-7k2 e2/9	Physically disconnected the cable connected to DC2-DCI-7K2's interface Ethernet 2/9 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass
DWDM Failure-Fiber Cut b/t DC	The active link between ONS 1 and ONS 2 was physically disconnected for failure, and reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass

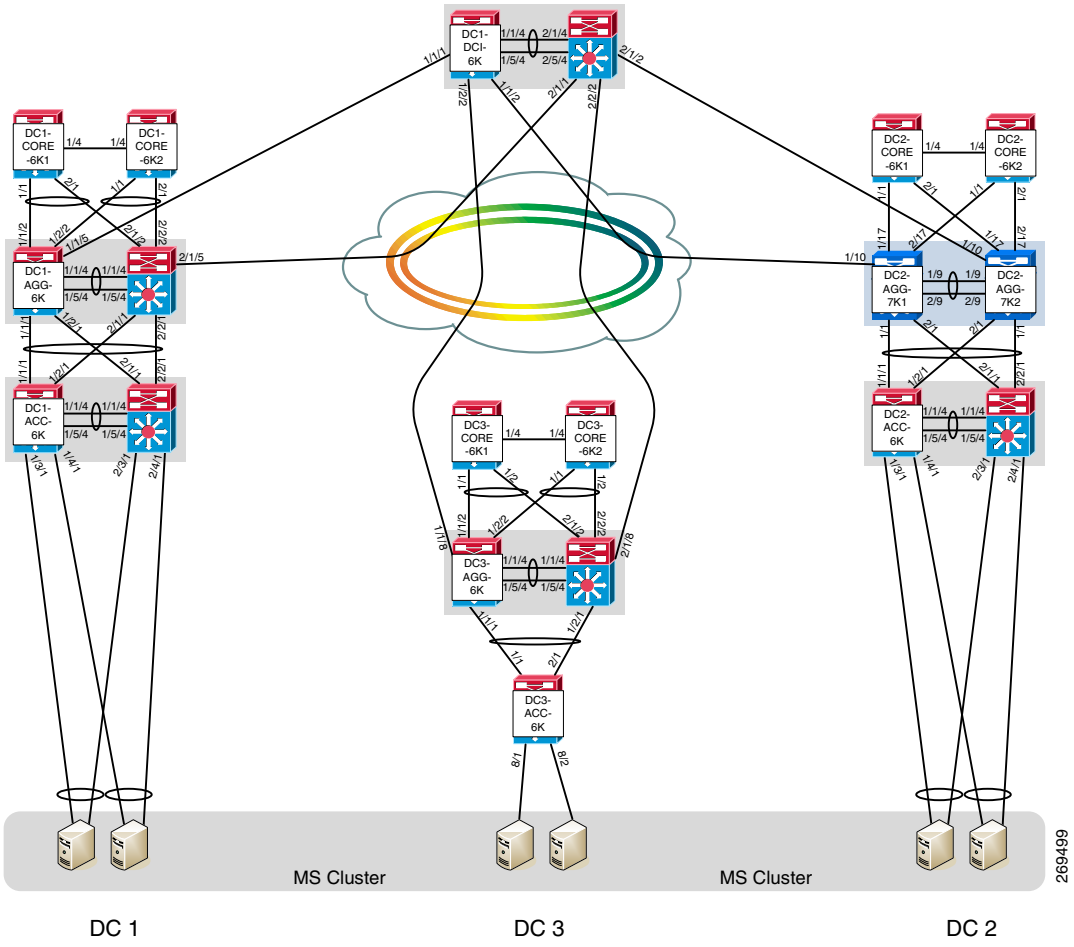
1. By failing all vPC peer link members plus the vPC peer keepalive link simultaneously, a vPC dual active condition is forced. If this scenario happens on a vPC domain that does not contain the STP root, the STP dispute mechanism inadvertently blocks links that cause intermittent traffic drop. While this is considered an unlikely scenario (triple failure within <3 seconds), NX-OS 4.2(3) will provide a configuration option to disable the STP dispute mechanism (CSCtb31482).
2. Same footnote 1.

Note As part of the testing effort it was discovered that for a particular traffic pattern, drops could be observed that resulted in 802.1AE re-keying to fail. Therefore test results are not reported in this version of this document and will be reported as soon as the issue is root caused. This is being tracked under CSCtb34740.

Multi-Site Testing

When reviewing results given in Table 2-5, refer to Figure 2-13. When reviewing results given in Table 2-6, refer to Figure 2-14.

Figure 2-13 Multi-Site Test Topology Details VSS Core



Multi-Site with VSS Core Test Results

Table 2-5 provides a detailed test results summary for DCI multi-site VSS core testing by test type.

Table 2-5 3-Site VSS Core Test Results

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
DCI Release 1.0 Multi-Site with VSS Core-HW Failure						
DC1-DCI-6K Chassis Failure	Powered down VSS active chassis for failure. Powered up VSS active chassis for restoration.	0.732s	0.082s	0.165s	0.128s	Pass
DC1-DCI-6K Line Card Failure	Physically ejected DC1-DCI-6K1's WS-X6708-10GE-3c in slot 1 for failure. Module reinserted for restoration.	1.420s	1.202s	.838s	0.488s	Pass
DC1-DCI-6K(2) Line Card Failure	Physically ejected DC1-DCI-6K2's WS-X6708-10GE-3c in slot 1 for failure. Module reinserted for restoration.	0.624	3.00s	1.148s	1.927s	Pass
DC2-DCI-6K Chassis Failure	Powered down VSS active chassis for failure. Powered up VSS active chassis for restoration.	0.221s	0.165s	0.877s	0.121s	Pass
DCI Release 1.0 Multi-Site with VSS Core-Link Failure						
MultiSite_VSS-DC1-Agg-6k Whole VSL Link Failure	Physically disconnected both of DC1-Agg-6k1's VSL links, TenGig 1/1/4 and 1/5/4 for failure. Both links reconnected for restoration.	5.020s (0.920s)	1.153s (1.977s)	0.404s (0.999s)	2.373s (1.104s)	Fail [1]
DC1-DCI-6K Link Failure	Physically disconnected the cable connected to DC1-DCI-6k1's interface TenGig 1/1/4 for failure. Link reconnected for restoration.	0.170s	0.0s	0.0s	0.0s	Pass

Table 2-5 3-Site VSS Core Test Results (continued)

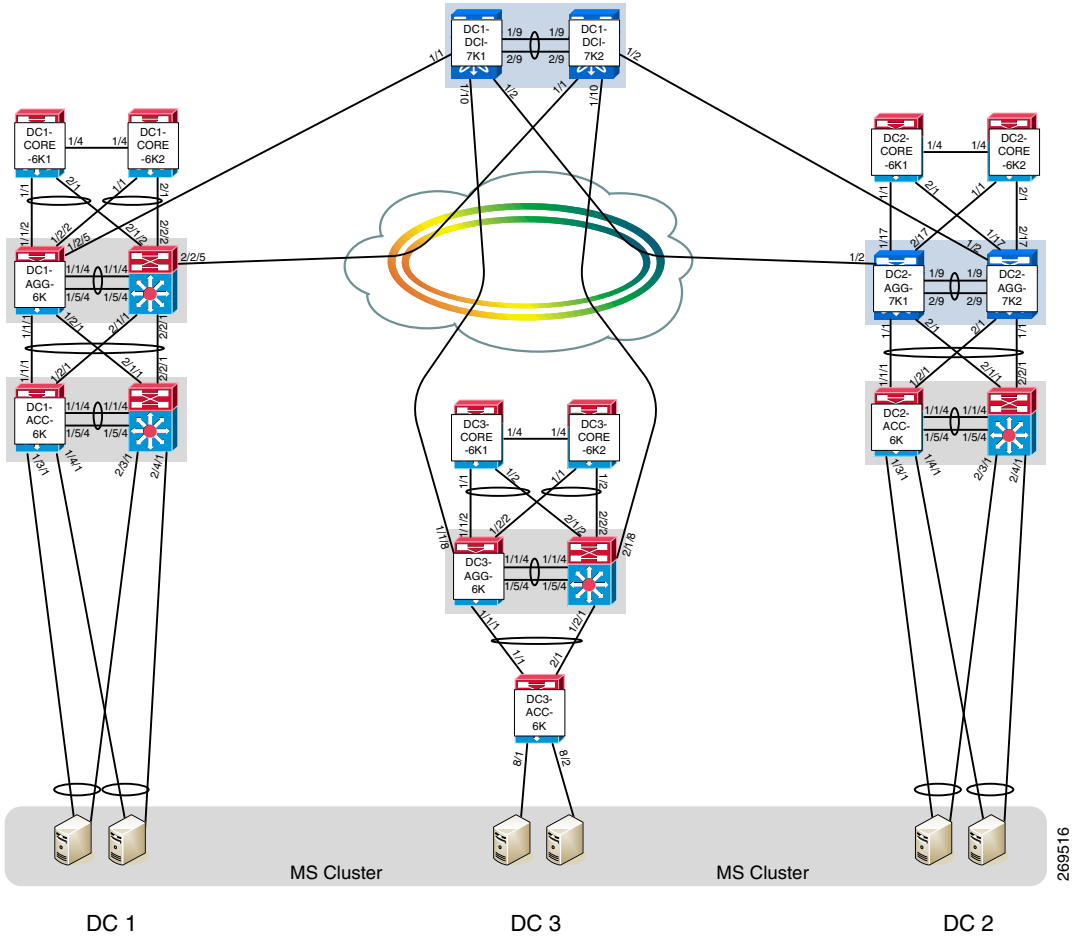
Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
DC2-Agg-7k1 vPC Peer Link and Keepalive Link Failure	Simulate total vpc link failure by shutting DC2-Agg-7k1's interfaces ethernet 1/9, ethernet 2/9 and management 0. No shut interfaces for restoration.	Sustained Loss	Sustained Loss	N/A	N/A	Fail [2]
DC1-DCI-6K Whole VSL Link Failure	Physically disconnected both of DC1-DCI-6K1's VSL links, TenGig 1/1/4 and 2/1/4 for failure. Both links reconnected for restoration.	2.020s (0.949s)	2.702s (2.574s)	2.036s (0.969s)	0.921s (1.788s)	Fail [3]
DCI Release 1.0 Multi-Site with VSS Core-Fault Isolation						
MultiSite_VSS-Create Loop in DC1 and Verify Fault Isolation	A spanning tree loop was created on DC1-ACC-6k by connecting an ethernet cable to interfaces Gig 1/3/48 and 2/3/48 with BPDU filtering enabled on each port. Storm control was configured on DC1-DCI-6k's Po10 physical interfaces to mitigate and constrain the event's impact to Datacenter 1.	N/A	N/A	N/A	N/A	Pass

Table 2-5 3-Site VSS Core Test Results (continued)

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
MultiSite_VSS-Create Loop in DC2 and Verify Fault Isolation	A spanning tree loop was created on DC2-ACC-6k by connecting an ethernet cable to interfaces Gig 1/3/48 and 2/3/48 with BPDU filtering enabled on each port. Storm control was configured on DC1-DCI-6k's Po10 physical interfaces to mitigate and constrain the event's impact to Datacenter 2.	N/A	N/A	N/A	N/A	Pass
MultiSite_VSS-Create Loop in DC3 and Verify Fault Isolation	A spanning tree loop was created on DC3-ACC-6k by connecting an ethernet cable to interfaces Gig 1/13 and 2/13 with BPDU filtering enabled on each port. Storm control was configured on DC1-DCI-6k's Po70 physical interfaces to mitigate and constrain the event's impact to Datacenter 3.	N/A	N/A	N/A	N/A	Pass

1. With the 12.2(33)SXI1 release, long convergence times were observed for the tests that resulted in VSS dual active condition. These convergence times were reduced significantly when tested with the 12.2(33)SXI2a release. For transparency, the convergence times observed with 12.2(33)SXI2a are included in the table, in parentheses.
2. By failing all vPC peer link members plus the vPC peer keepalive link simultaneously, a vPC dual active condition is forced. If this scenario happens on a vPC domain that does not contain the STP root, the STP dispute mechanism inadvertently blocks links that cause intermittent traffic drop. While this is considered an unlikely scenario (triple failure within <3 seconds), NX-OS 4.2(3) will provide a configuration option to disable the STP dispute mechanism (CSCtb31482).
3. Same as footnote 1.

Figure 2-14 Multi-Site Test Topology Details vPC Core



Multi-Site with vPC Core Test Results

Table 2-6 provides a detailed test results summary for DCI multi-site vPC core testing by test type.

Table 2-6 3-Site vPC Core Test Results

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
DCI Release 1.0 Multi-Site with vPC Core-HW Failure						
Multi-Site_vPC-DC1-DCI-7K Chassis Failure	Powered down DC1-DCI-7K1 for failure, powered back up for restoration.	1.493s	1.785s	1.228s	1.952s	Pass
MultiSite_vPC-DC1-DCI-7K Line Card Failure	Physically ejected DC1-DCI-7k1's N7K-M132XP-12 in slot 1 for failure. Module reinserted for restoration.	0.893s	2.424s	1.484s	1.436s	Pass
MultiSite_vPC-DC1-DCI-7K2 Chassis Failure	Powered down DC1-DCI-7K1 for failure, powered back up for restoration.	0.533s	1.815s	1.840s	2.180s	Pass
MultiSite_vPC-DC1-DCI-7K2 Linecard Failure	Physically ejected DC1-DCI-7k2's N7K-M132XP-12 in slot 2 for failure. Module reinserted for restoration.	0.115s	0.002s	0.765s	0.002s	Pass
vPC_vPC-DC1-Agg-6K Chassis Failure	Powered down VSS active chassis for failure. Powered up VSS active chassis for restoration.	.620s	.1836s	0.0s	0.020s	Pass
DCI Release 1.0 Multi-Site with vPC Core-Link Failure						
MultiSite_vPC-DC1-Agg-6k Whole VSL Link Failure	Physically disconnected both of DC2-DCI-6k1's VSL links, TenGig 1/1/4 and 1/5/4 for failure. Both links reconnected for restoration.	0.680ss	1.113s	0.800s	1.161s	Pass

Table 2-6 3-Site vPC Core Test Results (continued)

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
MultiSite_vPC-DC1-DCI-7K Link Failure	Physically disconnected the cable connected to DC1-DCI-7k1's interface Ethernet 1/9 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass
MultiSite_vPC-DC1-DCI-7K Peer Keepalive Link Failure	Physically disconnected the cable connected to interface Management 0 for failure. Link reconnected for restoration.	0.0s	0.0s	0.0s	0.0s	Pass
MultiSite_vPC-DC2-Agg-7k 1 vPC Peer Link and Keepalive Link Failure	Physically disconnected the cables connected to DC2-Agg-7k1's peer keepalive interface, Management 0, and entire vPC Peer Link, Ethernet 1/9 and 2/9 for failure. Links reconnected for restoration.	sustained loss	sustained loss	N/A	N/A	Fail [1]
DCI Release 1.0 Multi-Site with vPC Core-Fault Violation						
MultiSite_vPC-Create Loop in DC1 and Verify Fault Isolation	A spanning tree loop was created on DC1-ACC-6k by connecting an ethernet cable to interfaces Gig 1/3/48 and 2/3/48 with BPDU filtering enabled on each port. Storm control was configured on DC1-DCI-6k's Po10 physical interfaces to mitigate and constrain the event's impact to Datacenter 1.	N/A	N/A	N/A	N/A	Pass

Table 2-6 3-Site vPC Core Test Results (continued)

Test Case	Test Details	Failure Ucast	Failure Mcast	Restore Ucast	Restore Mcast	Result
MultiSite_vPC-Create Loop in DC2 and Verify Fault Isolation	A spanning tree loop was created on DC2-ACC-6k by connecting an ethernet cable to interfaces Gig 1/3/48 and 2/3/48 with BPDU filtering enabled on each port. Storm control was configured on DC1-DCI-6k's Po10 physical interfaces to mitigate and constrain the event's impact to Datacenter 2.	N/A	N/A	N/A	N/A	Pass
MultiSite_vPC-Create Loop in DC3 and Verify Fault Isolation	A spanning tree loop was created on DC3-ACC-6k by connecting an ethernet cable to interfaces Gig 1/13 and 2/13 with BPDU filtering enabled on each port. Storm control was configured on DC1-DCI-6k's Po70 physical interfaces to mitigate and constrain the event's impact to Datacenter 3.	N/A	N/A	N/A	N/A	Pass

1. By failing all vPC peer link members plus the vPC peer keepalive link simultaneously, a vPC dual active condition is forced. If this scenario happens on a vPC domain that does not contain the STP root, the STP dispute mechanism inadvertently blocks links that cause intermittent traffic drop. While this is considered an unlikely scenario (triple failure within <3 seconds), NX-OS 4.2(3) will provide a configuration option to disable the STP dispute mechanism (CSCtb31482).

Test Findings and Recommendations

Test cases resulted in the following findings and recommendations.

- One VSL ports on a supervisor, due to the fact that the sup will boot before the linecards, and at least one link is needed before the VSL can establish.
- Each VSS node should be dual-homed, to increase topology stability.
- Configuring Port Load Share Deferral on the peer switch to 60 seconds decreased convergence numbers.
- Of the 3 dual active mechanism ePagP and fast hello delivered similarly the best performance numbers. BFD was also tested but offered slightly higher convergence numbers than the others.
- Fast Hello was selected as the preferred VSS dual active detection mechanism because Fast Hello has been identified as the most reliable and robust dual active detection mechanism. VSS chassis and VSL link failure events with the Fast Hello mechanism enabled yielded sub-second convergence times for failure and full recovery.
- HSRP hello timers were changed to hello = 1 second and dead = 3.
- HSRP preempt delay minimum timer was changed to 60 due to the amount of VLANs we had, therefore, it allowed for a more graceful HSRP recovery.
- OSPF was configured with 'passive interface default' to minimize unnecessary control traffic (OSPF hellos) and reduce device overhead associated with higher numbers of peer relationships. Md5 authentication was implemented as a best practice to eliminate the possibility of rogue OSPF routers interfering with routing. Non Stop Forwarding (NSF) for OSPF was enabled to assist in keeping traffic and route disruption effects at a minimum. OSPF router ID's were configured referencing the routers' loopback addresses to provide predictable router ID's and consequentially routing protocol roles. Default timers
  -interface timers default are 10 hello, 40 dead.
- All Nexus 7k Devices Under Test (DUT's) were configured to use their management 0 interface as their peer keepalive link source. The Cisco NX-OS software uses the peer-keepalive link between the vPC peers to transmit periodic, configurable keepalive messages. You must have Layer 3 connectivity between the peer devices to transmit these messages. The system cannot bring up the vPC peer link unless the peer-keepalive link is already up and running.
- Throughout all spanning tree domains, VTP transparent mode was selected because the Nexus 7000 platform does not support Server or Client VTP mode.
- Rapid Pvst+ mode spanning tree was selected due to it's superior link failure recovery abilities, and pervasive adoption.
- Etherchannel mode desirable, and active, were implemented where possible. The desirable (PagP) and active (LACP) modes offer channel negotiation that prevent traffic forwarding problems resulting from a misconfiguration.
- Redundancy mode SSO is the default dual supervisor redundancy mechanism, and is necessary for VSS and NSF to function.
- All interfaces were configured with MTU 9216 to accommodate jumbo frames generated in some traffic profile streams.
- BPDU Filtering (prevents sending or processing received BPDU's) was configured on all ONS ring facing interfaces providing spanning tree domain separation between all data centers. DC1-AGG-6k was the spanning tree root of all VLANs in Datacenter 1. DC2-AGG-7k1 was the primary root of all VLANs for Datacenter 2, with DC2-AGG-7k2 acting as the secondary root. DC3-AGG-6k was the primary root for all VLANs in the Datacenter 3 domain.

Summary

Cisco is the leading global supplier of internetworking solutions for corporate intranets and the global Internet. In light of disaster recovery responsibilities and business continuum, regulatory compliance has emerged as a most challenging issue facing business and enterprise. From Securities and Exchange (SEC) Rule 17a to Sarbanes-Oxley and HIPAA, many legislative requirements now dictate how electronic data is stored, retrieved, and recovered. Organizations failing to meet new mandates face significant penalties and incalculable risk to corporate position and reputation.

Cisco Data Center Interconnect does not compromise off-site protection of data and electronic communications, ensuring regulatory compliance in keeping information safe and accessible, as required by the United States Securities and Exchange Act Rules 17a-3, 4 (17 CFR 240.17a-3, 4).

Cisco Data Center Interconnect compatible platforms supersede competitor portfolio gaps for DCI profiling by meeting and supporting Layer 2 transport mechanisms, high availability, spanning tree isolation, loop avoidance, multipath load balancing, swift end-to-end convergence and aggregation across WANs, queuing, and interconnect traffic encryption.

To this end, the Cisco Data Center Interconnect (DCI) system solution deploys state-of-the-art technology on robust proprietary platforms for strategic customers to extend subnets beyond Layer 3 boundaries of single site data centers, stretching clustered Layer 2 connected node functionality to promote resilient routing flexibility, and virtualization, while offsetting STP loops and broadcast storms, and identifying active IP addresses or subnets.

DCI System Release 1.0 tested two data center and multiple data center scenarios which were 100kms and 50kms apart using a DWDM network. The tests covered the most common to highly unlikely failure scenarios under constant traffic loads. In most scenarios, convergence times were observed as sub-second, highlighting rare failure scenarios that caused a couple seconds of outage. All testing was performed with a mix of EMIX traffic and the system was loaded to verify line rate scalability.

This release discovered a wide range of issues during testing. All show stopper and critical issues were resolved and verified within the scope of testing, and unresolved issues are documented as such.