Intel® Cloud Builders Guide: Cloud Design and Deployment on Intel® Platforms

Unified Networking with Cisco* Virtualized Multi-Tenant Data Center*

Intel® Xeon® Processor 5500 Series

Intel® Xeon® Processor 5600 Series

AUDIENCE AND PURPOSE

Cloud computing offers a path to greater scalability and lower costs for service providers, infrastructure hosting companies, and large enterprises. The establishment of an infrastructure that can provide such capabilities requires experience. Intel has teamed up with leading cloud vendors through the Intel® Cloud Builders program to help any customer design, deploy, and manage a cloud infrastructure.

The Intel Cloud Builders program provides a starting point as it supplies a basic hardware blueprint and available cloud software management solutions such as the Cisco* Virtualized Multi-Tenant Data Center (VMDC). You can use the use cases described in this paper as a baseline to build more complex usage and deployment models to suit specific customer needs.

We wrote this paper for cloud service providers, cloud hosters, and enterprise IT who want to realize revenue potential as they maximize their existing data center infrastructure to offer cloud computing services to their customers or internal users.

Table Of Contents

Executive Summary

The construction of a multi-tenant cloud by using a virtualized infrastructure implies that the infrastructure is effectively uniform throughout the data center. IT administrators use this design in large data centers not only to simplify the design and maintenance of the data center, but also to ensure consistent performance regardless of where the application is hosted in the data center.

Therefore, to the closest approximation, all servers are equal (from the perspective of the ability to host the workloads), all storage is uniformly accessible, and connectivity is uniform across the data center. The challenge in connectivity, of course, is to provide the connections to data and storage at low cost and to ensure delivery of the necessary bandwidth at sufficiently low latencies to meet the expectations of the workloads being hosted.

The emergence of 10 GbE fabrics has enabled the design of data center compute, storage, and connectivity to all use a common fabric. Data centers no longer require a separate fabric for storage and another for all other traffic. This convergence onto a single fabric has far reaching consequences, all of which yield simplification and delivery of the expected performance.

To make this transition approachable, this paper describes in some detail how to implement a multi-tenant data center through use of a single 10 GbE network fabric. Because this paper is only a starting point, it also includes "Things to Consider" that can inform on the options available to make the best choices in the design.

Introduction

The transition to a converged fabric is the result of the evolution of network technology to the point where a single fabric has sufficient throughput, low enough latency, and low cost to be the only means for connectivity in a data center. Cisco and Intel are uniquely positioned to deliver on the promise of improved efficiency and simplification for unified networking. Our years of collaboration to solve the most demanding requirements in the enterprise enabled this transition to 10 GbE as a converged fabric.

Now, as the benefits of cloud architectures continue to emerge, we have a huge demand for use of the convergence to deliver a flexible, scalable, and performant fabric for the data center. Imagine a data center where tens of thousands of servers host many tens of thousands of workloads. To achieve the low cost of service delivery expected by the customers, we require a high degree of automation. In this data center with tens of thousands of servers, the only practical way to deliver the required level of automation is when the compute, network, and storage are completely uniform. When any one server, storage, or network element requires special attention, then the automation has a special case. Every special case represents a point of risk as we maintain and expand the system over the years. Without the use of a converged fabric, we have one of these special cases: the storage and the connectivity require tight coordination through very different fabrics and management infrastructures.

The emergence of 10 GbE provides us with the opportunity to converge to one fabric. Therefore, a response to this converged networking solution stack is a critical next step for data center managers.

The inclusion of storage in the common fabric brings with it the opportunity to extend storage access through the use of industry standard protocols such as iSCSI (SCSI protocol over TCP/IP) and fiber channel over Ethernet (FCoE).

As the storage capacity requirements continue to expand, and as the availability and performance demands continue to increase, the ability to maximize industry standard platforms and protocols brings lower cost and simplification to the data center network design.

This paper focuses on the design of converged fabrics to meet the needs of the multi-tenant, cloud data center.

ISV Product Implementation Overview

The implementation describes a reference Infrastructure-as-a-service (IaaS) architecture that brings together core products and technologies from Cisco, Intel, and ecosystem partners to deliver a comprehensive, end-to-end cloud solution. Focused on IaaS cloud deployment, the Cisco Virtualized Multi-tenant Data Center (VMDC) solution provides customers with robust, scalable, and resilient options for cloud data center deployments.

This Cisco-driven, end-to-end architecture defines how to provision flexible, dynamic pools of virtualized resources that you can share efficiently and securely among different tenants, and provision quickly through process automation. Process automation reduces resource provisioning and improves time-to-market (TTM) for IaaS-based services. Shared resource pools consist of virtualized Cisco unified compute and virtualized storage area network (SAN) and network attached storage (NAS) storage platforms connected to the use of Cisco data center switches and routers.

This solution presents design and implementation guidance for a private or public IaaS cloud data center. The intended deployment uses a multi-tenant, differentiated service tier model.

When network architects design an IaaS architecture and the shared resource pools, they should consider the following design goals:

- **Secure separation**—Provides end-to-end tenant path isolation and security. Several security techniques at different layers of the network or infrastructure isolate tenants from each other. For example, we use virtual route forwarding instances (VRFs) at Layer 3 to stop communication between tenants at the layer 3 domain. Likewise, we use similar isolation features at compute and storage layers to provide complete isolation of tenants in a shared infrastructure.
- **Data center scalability**—A point of delivery (Pod)-based architecture provides network architects the ability to modularize the infrastructure into easily replicable units. These units are called Pods. Architects can plan for an initial Pod, which guarantees a certain scale and performance along with a scalable data center core network. This architecture provides a predictable and homogeneous method to add self-contained Pods as additional resources are needed.
- **High availability**—Availability ensures cloud resources accessibility even during a failure situation. We require availability to meet the expectations of service level agreements (SLAs) in a cloud deployment.

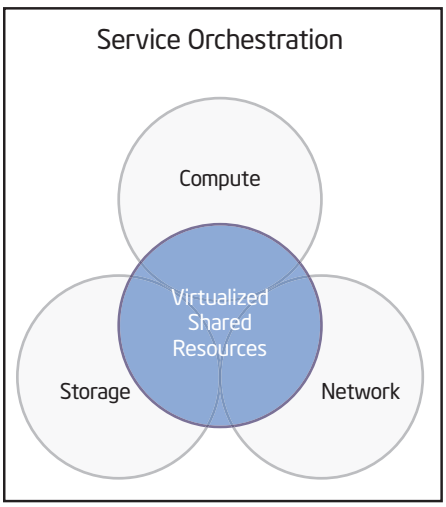


Figure 1. VMDC Building Blocks

- **Service assurance**—Provides mechanisms to define different service levels and defines how to adhere to them through the use of network quality of service (QoS) techniques during both steady and non-steady states. To differentiate IaaS service tiers, network architects can reserve and guarantee certain network bandwidths based on their subscription rules for the tier. For example, you can guarantee a Gold tenant with 1 Gb of bandwidth per virtual machine (VM) whereas a Silver tenant only gets 0.5 Gb per VM.

Test Bed Blueprint Overview

Hardware and Software Description

We list the hardware components within the Cisco VMDC in Table 1. They include networking (routing, switching), network services (firewall, load-balancing, intrusion detection), server computing, server virtualization,[1] SAN storage and system-level orchestration and management software.

Processor Description

Cisco UCS B-Series* blade servers support the following Intel® Xeon® processors[2]:

- Intel® Xeon® Processor 5500 Series* (2 or 4 cores per processor)
- Intel® Xeon® Processor 5600 Series* (4 or 6 cores per processor)
- Intel® Xeon® Processor 6500 Series* (4, 6, or 8 cores per processor)
- Intel® Xeon® Processor 7500 Series* (4, 6, or 8 cores per processor)

The following sections describe key Intel processor technologies that enable implementation of virtualized multi-tenant data centers:

Parallelism

Intel® enterprise servers include multiple processors and multiple computation engines to enable high-performance computing. The typical server model for most server applications includes two socket servers. Server workloads that require higher performance typically use four socket systems. If you need more processing power for the server workload, the number of processors per server can be increased.

Intel® Turbo Boost Technology

Intel® Turbo Boost Technology allows processors to deliver higher speed execution on demand by using available power to run at a higher frequency.

Intel® Hyper-Threading Technology

Many server applications lend themselves to parallel, multi-threaded execution. Intel® Hyper-Threading Technology enables simultaneous multi-threading within each processor core, up to two threads per core. Hyper-threading reduces computational latency, which makes optimal use of every clock cycle. For example, while one thread waits for a result or event, another thread executes in that core. In this way, hyper-threading maximizes the work from each clock cycle.

Intel® QuickPath Technology

Intel® QuickPath Technology is a scalable, shared memory architecture that delivers a high-memory bandwidth to enable top performance for bandwidth-intensive applications. It provides high–speed, point-to-point connections between processors, and between processors and the I/O hub. Each processor has its own dedicated memory, which it accesses directly through an integrated memory controller. In cases where a processor needs to access the dedicated memory of another processor, it can do so through a high-speed Intel® QuickPath Interconnect (Intel® QPI) that links all the processors.

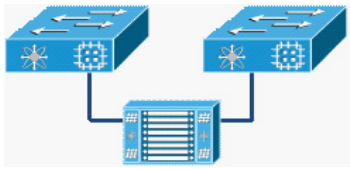
Features	Components	ICON
Network	Cisco Nexus 5010*	
	Cisco Nexus 7010*	
	Catalyst 6509*	
	Data center Services Node 6509-E (VSS – Virtual Switching System)*	
	Cisco Nexus 2148-T (FEX)*	
	Cisco Firewall Service Module* (FWSM) (for Catalyst 6500*)	
	Application Control Engine Module* for Catalyst 6500	
Compute	Cisco Unified Computing System* (UCS) UCS 5108 Blade Server Chassis* UCS Blade Server* UCS M61KR-I Intel Converged Network Adapter* (CNA) UCS M81KR Virtual Interface Card* (VIC) Cisco UCS 6120*, Cisco UCS 6140* fabric interconnect	

Table 1. Components of the VMDC Pod Resource Pool

Virtualization	VMware vSphere* VMware ESXi 4.0U1 Hypervisor* Intel® Virtualization Technology (Intel® VT-x, Intel® VT-d, Intel® VT-c)	
	Cisco Nexus 1000v* (virtual access switch)	
Security	Cisco Firewall Services Module* Cisco Application Control Engine* (ACE) VMware vShield* NetApp vFiler* and Virtual Service Domains* Cisco MDS* soft zoning and VSANs Cisco Nexus 1000V	
Storage Fabric	Cisco MDS 9506*, Cisco MDS 9513* (Director) and MDS 9148*, MDS 9134*	
Storage Array	EMC2 Symmetrix V-Max* with virtual provisioning	
Orchestration/Management	BMC Atrium Orchestrator* VMWare vCenter* Cisco UCS Manager* BMC BladeLogic for Server/Network* BMC Remedy IT Service Management Suite*	
Discrepancy Analysis	Determines if configuration deltas exist within the Cisco content services gateway (CSGs) defined in the server farm. Provides a report with found discrepancies; shows any divergent Cisco IOS* Software CLIs between Cisco CSGs	

Table 1. Components of the VMDC Pod Resource Pool

Intel® Intelligent Power Technology

Within a single server, Intel® Intelligent Power technology minimizes power consumption when server components are not fully utilized. Integrated power gates allow for reduction of individual idling cores to near-zero power independent of other operating cores. Automated low-power states automatically put processor and memory into the lowest available power states that will meet the requirements of the current workload. More CPU power states enhance processors, including a number of lower-power states, and the memory and I/O controllers have new power management features.

Intel® Virtualization Technology (Intel® VT)

Intel® Virtualization Technology (Intel® VT) enhances virtualization performance with new hardware assist capabilities across the following elements of the server:

- Processor: Intel® VT-x provides hardware-assisted page-table management, that allows guest operating systems more direct access to the hardware and that reduces compute-intensive software translation from the VM. Intel VT-x also includes Intel® VT FlexMigration and Intel® VT Flex-Priority, which are capabilities for flexible workload migration and performance optimization across the full range of 32-bit and 64-bit operating environments.
- Chipset: Intel® VT-d helps speed data movement and eliminates much of the performance overhead because it gives designated virtual machines their own dedicated I/O devices, which reduces the overhead of the virtual machine manager (VMM) required to manage I/O traffic.
- Network Adapter: Intel® VT-c enhances server I/O solutions because it integrates extensive hardware assists into the I/O devices that are used to connect servers to the data center network and storage infrastructure. Two technologies comprise Intel VT-c: Virtual Machine Device Queues, which accelerates throughput and reduces the load on the VMM and server processors, and Peripheral Component Interconnect Special Interest Group (PCI-SIG) single root I/O virtualization (SR-IOV), which delivers near-native throughput and provides dedicated, direct connectivity between VMs and hardware resources.

Intel® Advanced Encryption Standard (Intel® AES) Technology

Intel® Advanced Encryption Standard Technology enables robust encryption without the need for additional appliances and increased performance overhead. This technology improves CPU performance for encryption by as much as a 52 percent for secure Internet transactions and allows broader use of encryption throughout the data center.[3]

Intel® Trusted Execution Technology (Intel® TXT)

Intel® Trusted Execution Technology (Intel® TXT) addresses a critical security need for all server deployments, especially virtualized and cloud-based use models, because it helps to protect servers prior to OS launch or hypervisor launch. Intel® TXT complements other malware protections such as anti-virus and intrusion detection to help ensure that only trusted software is on the platform. Intel® TXT also helps protect VMs on trusted platforms so that you can more easily migrate them onto other trusted platforms or use them to create pools of platforms with trusted hypervisors.

Data Center Design with Pod Building Block

Cisco VMDC architecture, based on Pod, provides network architects the ability to modularize the infrastructure into easily replicable units called Pods. Architects can plan for an initial Pod, which guarantees a certain scale and performance along

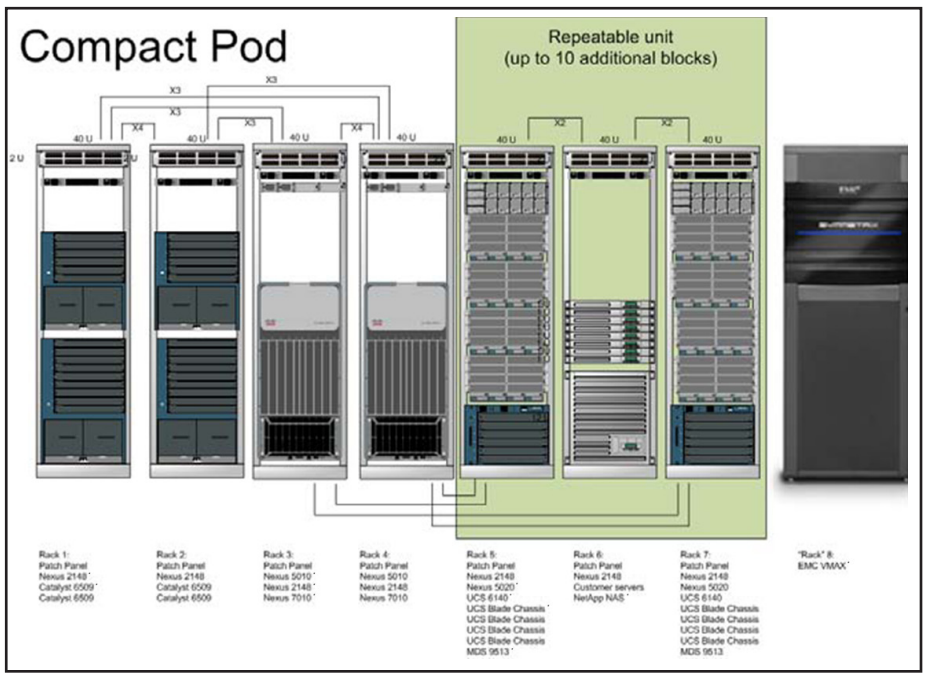


Figure 2. Physical Rack View of VMDC Pods

with a scalable data center core network. This architecture provides a predictable and homogeneous method to add self-contained Pods as you need additional resources. The test bed that is used for this reference architecture uses a data center built with the use of Pod.

All of the equipment defined in this document is in operation at Cisco Solution Labs in San Jose, CA and Research Triangle Park, NC. The equipment in the labs is typically in two configuration sizes: compact Pod and expanded Pod. The compact Pod allows Cisco to rapidly validate the solution as well as to deliver an entry-level pricepoint for customers, and the expanded Pod allows Cisco to test the system at a larger scale. Visibility to the solution will soon be available within Cisco's Customer Proof of Concept labs.

This view, in Figure 2, shows the physical rack layout of a Cisco VMDC Pod (network/compute/storage), as well as the aggregation and core network elements. NOTE: This is a typical rack layout, although specific customer layouts may vary based on environmental considerations within individual data centers.

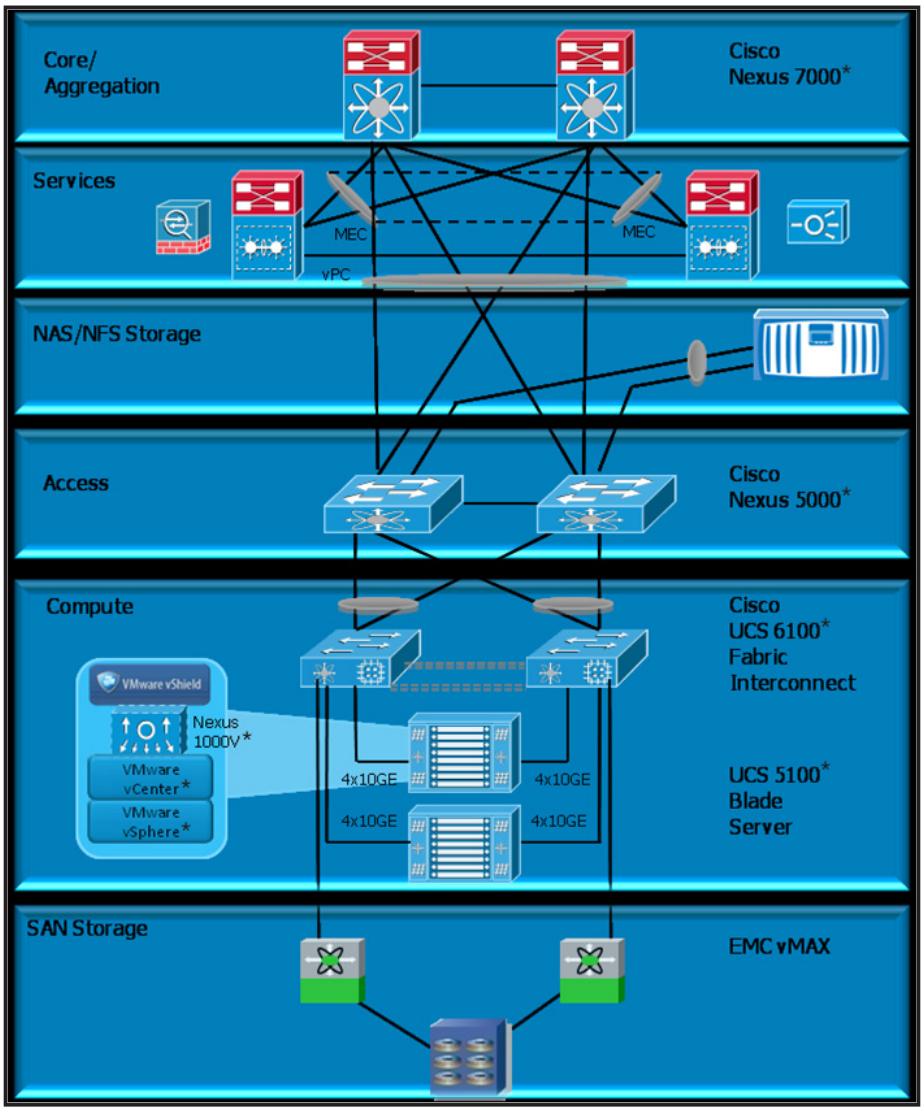


Figure 3. Cisco VMDC Pod - Logical Topology (includes layers of the network)

Racks 1 and 2 –Provide connectivity to non-Cisco UCS computing resources (through the Nexus 2148*) and aggregation-layer services (through the Catalyst 6500*). They provide high bandwidth interconnect (10 Gb scalable up to multiple 100 Gb links) and security, load-balancing, and segmentation services.

Racks 3 and 4 – Provide both core-level network services (through the Nexus 7000*), but also top-of-rack (ToR) aggregation services for Cisco UCS* and non-Cisco UCS computing services or other access-layer devices.

Racks 5, 6 and 7 –Provide unified computing services (through Cisco UCS) and access to both SAN and NAS storage.

Rack 8 – Provides SAN storage.

Figure 3 shows a diagram of the physical level view of the Cisco VMDC Pod. The Pod uses Cisco Data Center[4] best practices for core, aggregation and access-layer networking, security, isolation and QoS. The server computing, server virtualization, and SAN storage components are attached at the access layer. Each layer of the VMDC architecture supports the ability to segment and isolate traffic by user or class-of-service.

Core/aggregation layer – Provides core routing functionality over high-speed links, which brings together multiple aggregation layer areas within the data center.

Services layer – Provides security, load-balancing, traffic inspection, and routing services to multiple access layer users and devices. These services are run at the services layer on highly available platforms so that high-speed packet switching can move to the core layer.

NAS/ Network File System (NFS) storage layer – Provides IP connectivity to NAS/NFS storage services for virtual machines (NFS) or user-shared data (NAS/ common Internet file system (CIFS)).

This is typically considered similar to an access layer, with both 1 GbE and 10 GbE connectivity.

Access layer – Provides aggregation of computing devices and other end-points (wireless access points, printers, etc.) through 100 Mb/1 Gb/10 Gb links. The segmentation (through virtual local area network (VLANs)) that provides multi-tenancy capabilities across the network is overlaid on the access layer. A combination of both data (LAN) and storage (SAN/NAS) protocols is within the access layer.

Compute layer – Provides connectivity to Cisco UCS, which hosts virtualized (eg. VMware) and non-virtualized business applications. You can dedicate computing resources within the Cisco UCS to a specific application or shared by many applications. You can dedicate UCS resources to a specific user or share them with many business groups. Access to Cisco UCS is through unified 10 GbE links that can carry both LAN and SAN traffic.

SAN storage layer – Provides fiber channel (FC), iSCSI, or fiber channel over Ethernet (FCoE) connectivity to SAN storage. Uses zoning and logical unit number (LUN) masking to extend isolation capabilities to VMDC for multi-tenancy.

Figure 4 shows the physical 10 GbE connectivity between the layers within the Cisco VMDC. The figure highlights the VDC functionality that provides network-centric services to both the access and core layers through the aggregation layer.

Many customers will need to expand their cloud capacity, so it's important to understand how multiple Cisco VMDC Pods interconnect within the data center. Figure 5 shows a logical view of multiple Pods and how they interconnect into the broader network. Figure 5 highlights how you can deliver multiple classes of service to a shared infrastructure in a multi-tenant environment.

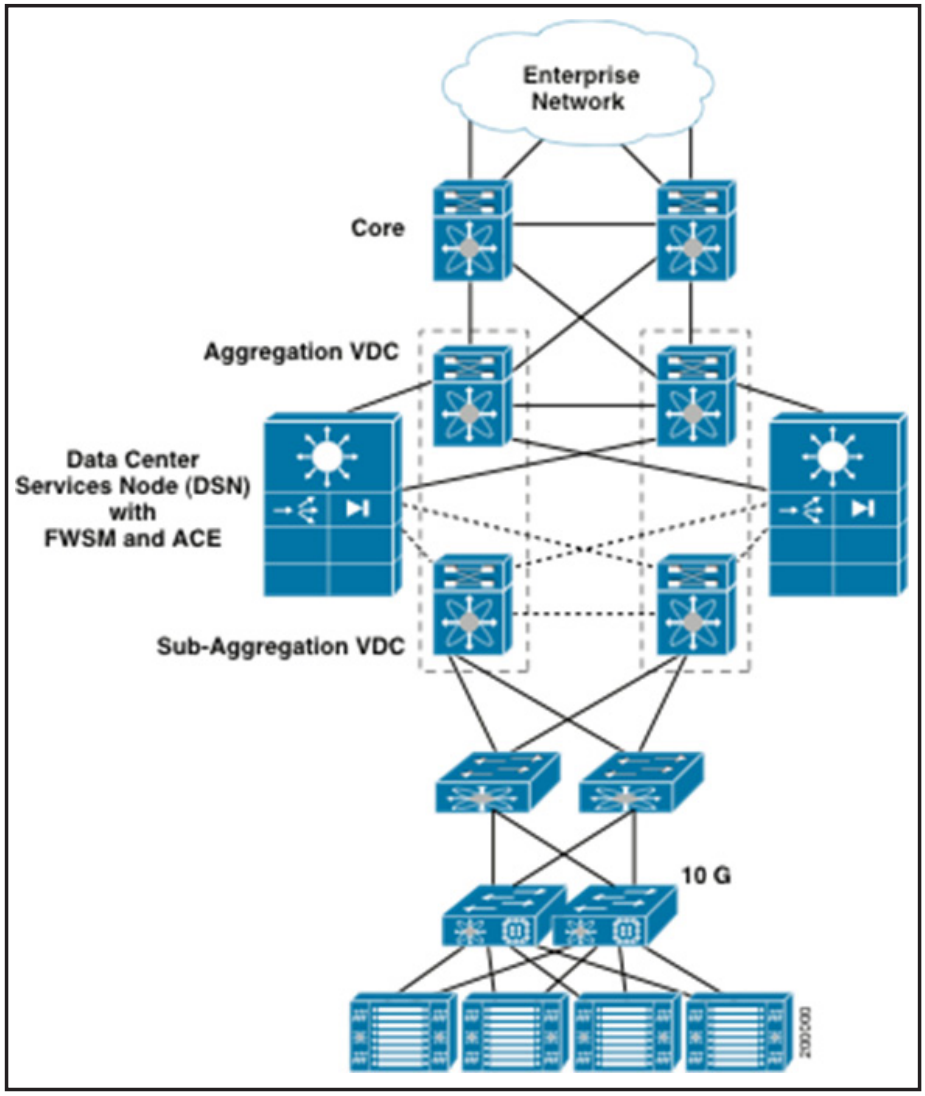


Figure 4. VMDC Pod, Physical Layout – 10 Gb Access Layer

Technical Review

Installation Overview

In general terms, an installation of Cisco VMDC will accomplish the following tasks:

- Prepare the physical environment. VMDC Pod is made up of equipment that will fill multiple standard equipment racks. We listed the equipment in Figure 2.
- Prepare the cabling between the equipment, including Cat5/6, Twinax and Fiber cabling to interconnect the servers, network, and storage.
- Prepare the system addressing and naming, including IP addressing, storage addressing, and DNS naming.
- Prepare the computing environment, including identification of application images for virtualization, operating systems, applications, and the associated storage (LUNs, file systems) and networking.
- Configure the elements within the VMDC architecture.
- Enable and configure the system management, network management and system orchestration.

Use Case Details

This paper focuses on unified networking and multi-tenancy use cases.

Unified Networking:

Enterprise data center networks are complex systems. We typically connect servers in the data center to several different networks for diverse functions such as production traffic, backups, storage, management, and VM migration. While Ethernet is the predominant networking technology in use, many data centers utilize FC technology for their storage network. If the traffic load demands it, individual network ports are aggregated to provide more capacity for a particular function. These practices result in many network connections for each server and require the corresponding cables and switch ports to support them. Up to 6-12 network connections are common for a virtualized server today, which impacts the data center CapEx and OpEx due to the overall complexity of this arrangement.

The transition to 10 GbE allows consolidation of multiple separate Ethernet ports into fewer 10 GbE ports which greatly simplifies the data center network, reduces costs, and simultaneously provides greater overall platform networking bandwidth capability. If you have used FC in the past, you can use this additional Ethernet capacity effectively to consolidate separate storage network traffic onto a common 10 GbE unified networking infrastructure which drives an even simpler and more cost effective data center. 10 GbE consolidation and unified networking can greatly simplify network management.

Multi-tenancy:

Multi-tenancy refers to the capability of the data center to host multiple customers such that the resources for each customer (network, storage, and compute) are logically separate from other customers' resources, with security separating them. This is a critical attribute of any cloud computing deployment, as it is one of the key items that differentiates cloud computing and IaaS from collocation or dedicated infrastructure for each application. It is relevant in public clouds that host multiple customers with the same or different service-level requirements, and in private clouds in which multiple departments or organizations share the same cloud infrastructure. You must support different degrees of multi-tenancy throughout the data center. The data center architecture should balance logical and physical segmentation. You assign unique resources to each tenant in a modular data center. These resources include different policies, pools, and QoS definitions. Virtualization at different layers of a network allows the infrastructure to provide logical isolation without the dedication of physical resources to each customer.

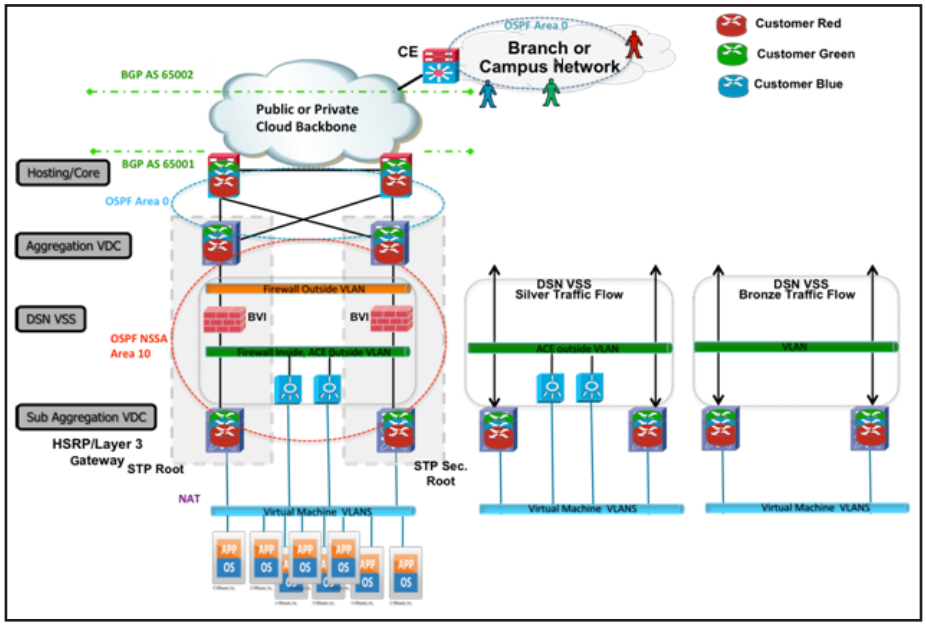


Figure 5. VMDC - End-to-End Logical Topology (Bronze, Silver, Gold)

A scalable data center that supports multi-tenancy architecture should include:

- Modular, multi-tenant data center design with data center infrastructure modules optimized for different scale and cost points
- Service orchestration for on-demand provisioning of resources
- Service-tier based design to allow for differentiated services
- Workload mobility and disaster recovery capability for business continuance
- Security at each layer of the data center

Execution and Results

Once you build the physical and logical environment within the data center, the next set of tasks focuses on validation of the architecture. Since the Cisco VMDC brings together dozens of technology elements, the breadth of use cases is very large. Table 2 highlights the main areas of attention that cloud architects and IT operations should focus on to validate the architecture.

Within each area, you should follow defined test cases to validate that each function works properly and performs to the expected levels. In this document we will not detail all functional areas, but rather highlight the segmentation/ isolation (multi-tenancy) function.

Network Segmentation and Isolation

Within the core of the network, routing protocols allow us to logically segment the network infrastructure. You could use each segment for a class of service (Gold, Silver, Bronze) or dedicate it to a set of users (a tenant). Beyond the use of routing protocols, the core switches use Virtual Data Context (VDC) functionality to virtualize the hardware into logical switches, which further isolates segments of the infrastructure from other segments.

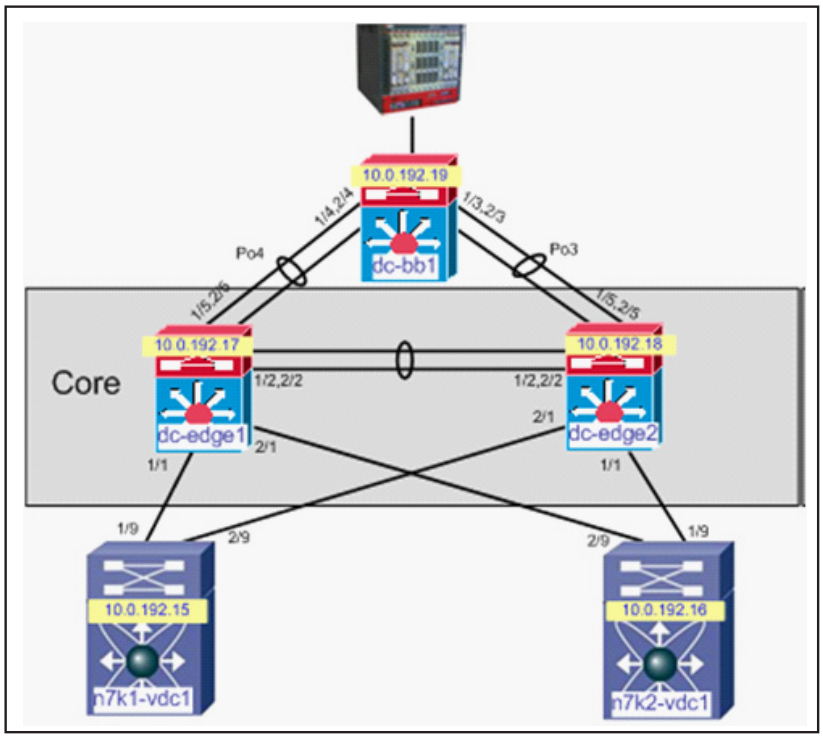


Figure 6. VMDC - Core Network Topology

Features	Technologies
Data center end-to-end functionality validation for SAN and NAS	End-to-end feature/integration testing, including QoS for all datacenter network layers from access to wide area network (WAN) edge on all platforms; ESX/VM provisioning, boot up, and maintenance; and SAN/NAS storage design verification
Disaster recovery scenario validation	Transparent movement of data center workloads for business continuance (active backup scenario)
Automation validation	Validation of service orchestration, portal, service catalog validation with element manager, and integration for compute and network
Data center services functionality validation	Validation of service tier offerings with data center services node (firewall and load balancing), multi-tenancy through VLANs, zoning, LUN masking
Failover scenario validation	Validation of redundancy designs (with baseline steady state traffic) – routing, virtual PC (vPC)/MEC, equal cost multipath (ECMP), virtual switching system (VSS), hot standby router protocol (HSRP), active-active service modules, clustering
Security validation	End-to-end security validation on various components
Scalability verification	Multi-dimensional scalability (VLAN, media access control (MAC), HSRP, routes, contexts, VM) within scope of architecture

Table 2

Requirement Matrix	Component Features that Address Requirements
End-to-end VRF Lite	Individual VRFs for all customers and all departments from BC-BB1 though VDC2 in the sub-aggregation layer to provide Layer 3 segregation.
IGP (interior gateway protocol) (open shortest path first (OSPF))	OSPF is configured from the edge routers to VDC2 in the sub-aggregation layer.
Border gateway protocol (BGP)	BGP is configured between the edge routers and DC-BB1.

Table 3

We use VRF lite to configure each router. There are 32 VRFs. We configure the VRFs in DC BB1 and extend them to the sub-aggregation layer. We configure BGP between DC-BB1, DC-Edge1, and DC Edge2. We advertise client networks on DC-BB1 over BGP to the edge routers, DC-Edge1, and DC Edge2. We configure the edge routers with BGP and OSPF. The edge routers are OSPF internal routers (IR) and are only in OSPF backbone area 0.

Through the use of route maps, we redistribute BGP client routes into OSPF at edge routers to the aggregation layer. Server subnets advertised into OSPF by the aggregation layer to the edge routers are redistributed into BGP at the edge routers to DC-BB1. We redistribute the server subnet routes into BGP through the use of the metric option in the redistribute command so that server subnet routes for all odd VRFs are sent to DC-BB1 from DC-Edge1 and server subnet routes for all even VRFs are sent to DC-BB1 from DC-Edge2. This configuration allows for the control of client-to-server (north to south) traffic to load balance from north to south. ECMP automatically takes care of this configuration for server to client traffic (south to north).

This output shows the isolated routing table (per-VRF) for router 1

```
DC-Edge1#sh ip route vrf Dept-1-Bronze-1
Routing Table: Dept-1-Bronze-1
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
E1 - OSPF external type 1, E2 - OSPF external type 2
i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
ia - IS-IS inter area, * - candidate default, U - per-user static route
o - ODR, P - periodic downloaded static route
Gateway of last resort is not set
1.0.0.0/32 is subnetted, 4 subnets
O IA 1.1.1.11 [110/4] via 121.2.1.16, 10:31:16, TenGigabitEthernet2/1.1901
[110/4] via 121.1.1.15, 10:31:16, TenGigabitEthernet1/1.1801
O IA 1.1.1.12 [110/4] via 121.2.1.16, 10:31:16, TenGigabitEthernet2/1.1901
[110/4] via 121.1.1.15, 10:31:16, TenGigabitEthernet1/1.1801
O 1.1.1.15 [110/2] via 121.1.1.15, 10:31:16, TenGigabitEthernet1/1.1801
O 1.1.1.16 [110/2] via 121.2.1.16, 10:31:16, TenGigabitEthernet2/1.1901
101.0.0.0/16 is subnetted, 1 subnets
O IA 101.1.0.0
[110/43] via 121.2.1.16, 10:31:16, TenGigabitEthernet2/1.1901
[110/43] via 121.1.1.15, 10:31:16, TenGigabitEthernet1/1.1801
99.0.0.0/8 is variably subnetted, 4 subnets, 2 masks
C 99.10.1.0/24 is directly connected, Port-channel4.1701
B 99.15.1.0/24 [20/0] via 99.10.1.19, 3w6d
C 99.1.1.17/32 is directly connected, Loopback1
B 99.1.1.19/32 [20/0] via 99.10.1.19, 3w6d
111.0.0.0/24 is subnetted, 1 subnets
O IA 111.1.9.0 [110/3] via 121.2.1.16, 10:31:16, TenGigabitEthernet2/1.1901
[110/3] via 121.1.1.15, 10:31:16, TenGigabitEthernet1/1.1801
121.0.0.0/24 is subnetted, 5 subnets
```

```
C 121.1.1.0 is directly connected, TenGigabitEthernet1/1.1801
C 121.2.1.0 is directly connected, TenGigabitEthernet2/1.1901
C 121.3.1.0 is directly connected, Port-channel2.1601
O 121.4.1.0 [110/2] via 121.3.1.18, 10:31:16, Port-channel2.1601
O 121.5.1.0 [110/2] via 121.3.1.18, 10:31:16, Port-channel2.1601
```

This output shows the isolated routing table (per-VRF) for router 2

```
DC-Edge2#sh ip route vrf Dept-1-Bronze-1
Routing Table: Dept-1-Bronze-1
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external
i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
ia - IS-IS inter area, * - candidate default, U - per-user static route
o - ODR, P - periodic downloaded static route

Gateway of last resort is not set
1.0.0.0/32 is subnetted, 4 subnets
O IA 1.1.1.11 [110/4] via 121.5.1.15, 10:31:19, TenGigabitEthernet2/1.1821
[110/4] via 121.4.1.16, 10:31:19, TenGigabitEthernet1/1.1921
O IA 1.1.1.12 [110/4] via 121.5.1.15, 10:31:19, TenGigabitEthernet2/1.1821
[110/4] via 121.4.1.16, 10:31:19, TenGigabitEthernet1/1.1921
O 1.1.1.15 [110/2] via 121.5.1.15, 10:31:19, TenGigabitEthernet2/1.1821
O 1.1.1.16 [110/2] via 121.4.1.16, 10:31:19, TenGigabitEthernet1/1.1921
101.0.0.0/16 is subnetted, 1 subnets
O IA 101.1.0.0
[110/43] via 121.5.1.15, 10:31:19, TenGigabitEthernet2/1.1821
[110/43] via 121.4.1.16, 10:31:19, TenGigabitEthernet1/1.1921
99.0.0.0/8 is variably subnetted, 4 subnets, 2 masks
C 99.11.1.0/24 is directly connected, Port-channel3.1721
B 99.15.1.0/24 [20/0] via 99.11.1.19, 3w6d
C 99.1.1.18/32 is directly connected, Loopback1
B 99.1.1.19/32 [20/0] via 99.11.1.19, 3w6d
111.0.0.0/24 is subnetted, 1 subnets
O IA 111.1.9.0 [110/3] via 121.5.1.15, 10:31:19, TenGigabitEthernet2/1.1821
[110/3] via 121.4.1.16, 10:31:19, TenGigabitEthernet1/1.1921
121.0.0.0/24 is subnetted, 5 subnets
O 121.1.1.0 [110/2] via 121.3.1.17, 10:31:19, Port-channel2.1601
O 121.2.1.0 [110/2] via 121.3.1.17, 10:31:19, Port-channel2.1601
C 121.3.1.0 is directly connected, Port-channel2.1601
C 121.4.1.0 is directly connected, TenGigabitEthernet1/1.1921
C 121.5.1.0 is directly connected, TenGigabitEthernet2/1.1821
```

Aggregation and Sub-Aggregation Layers (Services VDC Sandwich Design)

Data center service insertion requirements may include server load balancing devices, security devices such as firewall and intrusion prevention, and others. There are multiple approaches for the integration of these services into the data flow. Design decisions include the use of modules in external services chassis, the use of appliances, and whether to run the service devices in a transparent or routed mode.

One design approach is to configure all services in transparent mode, but to insert an additional layer of routing instances between the server farm subnets and the services devices. We have shown this approach previously in design guidance with the use of VRFs, and the deployment of multiple VRFs also provides the capability to direct traffic independently through multiple virtual contexts on the service devices, through the virtualization of both the routing functions and the services devices in the design.

Another design approach includes the configuration of the firewall services modules in transparent mode and the Cisco Application Control Engine* (ACE) modules in routed mode through the use of Source NAT (SNAT) in each context of the ACE. This approach enables the accessibility of client-to-server traffic destined to the virtual IP (VIP) addresses that are not configured on the same IP subnet as the outside IP address configured on the VLAN interface within each ACE context. We configure static routes to each VIP host address or subnet in the aggregation layer. We configure SNAT in each ACE context such that all client source IP addresses are translated to an IP address on the same IP subnet configured on the inside VLAN interfaces. The servers also reside on this same IP subnet. For return traffic from the server to the client, we configure default routes in each ACE context that point to the HSRP address configured in each VRF in the aggregation layer. We implemented the latter approach in the design.

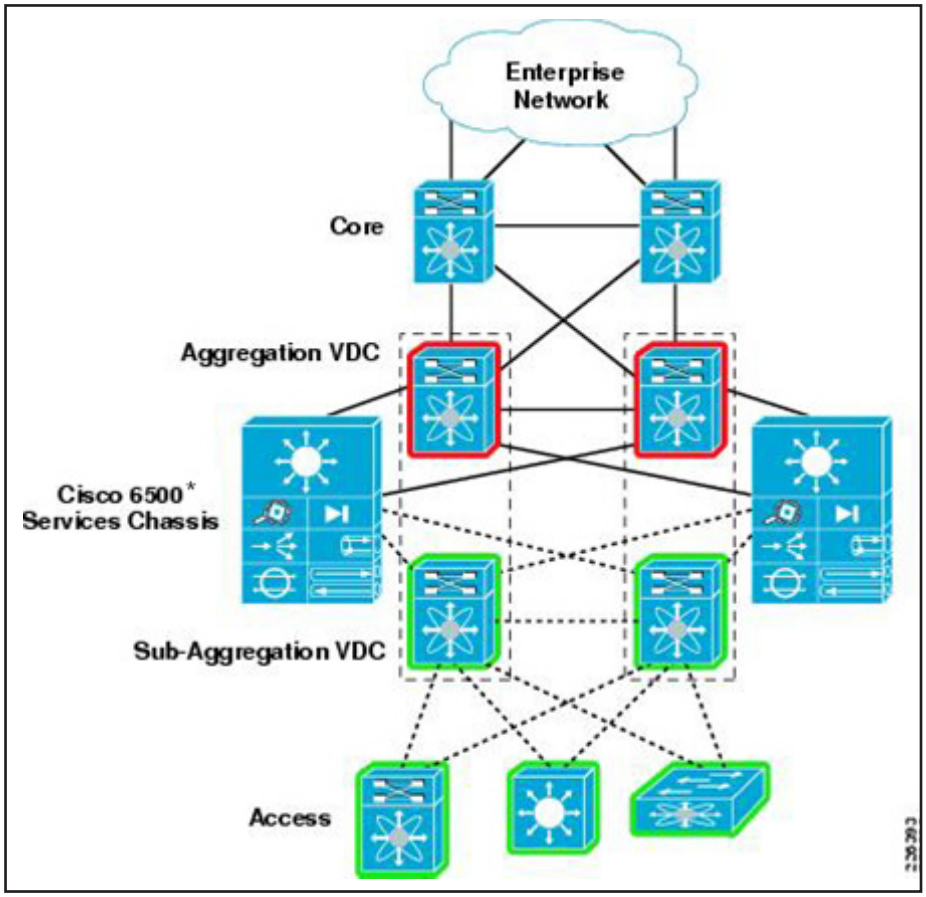


Figure 7. VMDC – Aggregation Network (and VDC) Topology

The VDC capability of the Nexus 7000 Series* enables the network architect to make use of another type of virtualization in the design, to improve ease of configuration, supportability, and security. We can create the sub-aggregation, a secondary virtual switching layer, through the use of VDCs located between the services devices and the access switches. We refer to this topology as a services VDC sandwich.

All of the access layer switches shown in Figure 7 attach only to the sub-aggregation VDCs. We could also attach different classes of servers to access-layer switches that connect directly to the main aggregation layer above the services chassis, if they either do not require services or are serviced by a different group of services devices. Additional design considerations for this type of topology include the following:

- Similar designs have been deployed through the use of a single pair of switches with separate VLANs and VRFs to provide the routing instance below the services chassis. The insertion of a separate set of VDCs into the design still represents the use of a single physical pair of switches to perform these functions but provides better isolation between the routing environments above and below the services chassis. This conceptually provides for easier support and configuration without the increase of the impact of a single-switch failure due to the introduction of a second set of VDCs.

- The security model is more robust since the operating environment of the sub-aggregation VDCs is completely separate from the primary aggregation layer. Instead of only separate VLANs and VRFs on the same switch, there are separate virtual switches with completely different sets of processes and physical ports.
- You may require additional interfaces for the VDC sandwich topology as compared to a VRF sandwich topology. The services chassis must have separate physical connections into both sets of VDCs as opposed to VLANs that share the same trunks. You must also provision additional interface count to support the inter-switch link between the two sub-aggregation VDCs.
- Cisco validated this model through the use of FWSM and ACE modules that run in transparent mode, where the two layers of VDCs are direct IP routing peers. Layer 3 control plane load on the VDC below the services may be limited by the use of static routes that point to an HSRP address shared between the primary aggregation VDCs to support IP unicast traffic flows. IP multicast traffic is not supported over a combination of static routes and HSRP addresses. If you require IP multicast, then you may use an IGP such as OSPF or enhanced interior gateway routing protocol (EIGRP).
- VDCs provide the distinction between the routing instances of the aggregation and the sub-aggregation layers; however, you may use multiple VRFs in the sub-aggregation layer to support additional virtualization capabilities. You may map distinct VRFs in the sub-aggregation layer with the use of VLANs to separate contexts within the virtualized service devices such as the FWSM and ACE, which allows the split of active contexts between both services chassis. If you require services between layers of a multi-tier application architecture, placement of these tiers in subnets that belong to separate VRFs will allow for powerful, multi-context service insertion between tiers.
- A services VDC sandwich that uses external services chassis provides independent connectivity between the services and both aggregation switches. If the aggregation switch on the left side of the topology fails, then the services on the left side have dual connectivity and can maintain a primary role. Service appliances run in transparent mode that only support single connections to carry a given VLAN (such as the Adaptive Security Appliance (ASA) 5580*) will not be dual-homed if they are attached directly to the aggregation. You can still deploy these appliances in a highly available manner through the use of redundant appliances.

Storage Layer

Depending on your storage requirements, you can implement a SAN or NAS solution. The following sections describe how we implemented each type of solution in Cisco VMDC.

- Storage component
- SAN design
- NAS design

Storage Component

As an essential piece to every data center, the storage component of the SAN provides several capabilities to the data center, including the ability to remotely boot from SAN and virtual (thin) provisioning for increased performance and data protection. These segregate capabilities ultimately alleviate responsibilities from the host and SAN.

Scenarios to Address

Secure Separation

- Device mapping
- LUN masking
- VSANs
- Zoning

NOTE: In these examples, the storage array in use is an EMC VMAX*, with FC SAN attached storage. We used Symmetrix Management Console (SMC) as the interface to configure the EMC VMAX array. GUI displays are from Cisco Fabric Manager.

LUN Masking (via SMC)

1. Configure the mask view.
   a. Right-click the Symmetrix Masking folder and select **Device Masking and Mapping > Masking Views Maintenance > Create Masking View**.
   b. Name the new masking view.
   c. Select an existing storage group, or create a new group following the steps in "Configure Storage Groups."
   d. Select an existing port group, or create a new group following the steps in "Configure Port Groups."
   e. Select an existing initiator group, or create a new group following the steps in "Configure Initiator Groups."
   f. *Optional* Click **Set Dynamic LUN Addresses** to manually configure the LUN addresses for each device.
   g. Click **OK** to confirm the new masking view.
2. Configure the storage groups.
   a. Right-click the Symmetrix Masking folder and select **Device Masking and Mapping > Storage Groups Maintenance > Create Storage Group**.
   b. Name the new storage group.
   c. Select the device source type.
   d. Add available devices to the group members column.
   e. Click **OK** to confirm the new storage group.

3. Configure the port groups.
   a. Right-click the Symmetrix Masking folder and select **Device Masking and Mapping > Port Groups Maintenance > Create Port Group**.
   b. Name the new port group.
   c. Add the available ports to the group members columns.
   NOTE: These ports are front-end FA ports only.
   d. Click **OK** to confirm the new port group.
4. Configure the initiator groups.
   a. Right-click the Symmetrix Masking folder and Select **Device Masking and Mapping > Initiator Groups Maintenance > Create Initiator Group.**
   b. Name the new initiator group.
   c. Add the available initiators to the selected initiators column.
   d. *Optional* If the host is not in the available initiators list, the **New Initiator** option allows you to manually add the pWWN to the selected initiators column.
   e. Click **Set HBA Flags** to open a new window to modify flags associated with this group.
   f. Click **OK** to confirm the new initiator group.

Software Configuration

To ensure data separation, scalability, and future expansion, as well as high availability and redundancy at key points of failure, we enabled the following software features:

- VSANs—General data separation
- Zone/Zoneset—Granular data separation
- NPV/NPIV—Host end scalability
- PortChannel—Redundancy/failover protection between edge and core switches

Design Considerations

Configure each component of a SAN layer redundantly such that upon a failure, the standby peer becomes active immediately with no service disruption or data loss. Also, the SAN should support expansion and scale to accommodate a company's needs for additional hosts and storage.

Configuration Purpose

This section details the purposes for configuration of the following five features:

- VSANs
- Zoning
- N-Port ID Virtualization (NPIV)
- N-Port Virtualizer (NPV)
- PortChannel

VSANs

To achieve the same level of isolation as physically separate fabrics at a lower cost, a VSAN creates separate virtual fabrics on a common physical infrastructure. Membership in a VSAN is based on physical port, and no physical port can belong to more than one VSAN; a node connected to a physical port belongs to that port's VSAN. VSANs provide strict hardware isolation and replicate the FC services created for each new VSAN. When you create a new VSAN, you create and enable a separate set of services that includes name server, zone server, domain controller, alias server, and login server, across those switches that you configured to carry the new VSAN. This services replica enables the isolated environments to satisfy high-availability requirements over a shared physical infrastructure. VSANs also interconnect isolated SAN fabrics in remote data centers over a common long-haul infrastructure. Because frame tagging is performed in hardware, you can multiplex the traffic from several VSANs across a single fiber pair and transport it a greater distance, all while it remains completely isolated. VSANs scale over a redundant physical infrastructure to provide flexible isolated SAN fabrics that achieve high-availability goals.

Zoning

Within each VSAN, the active zone set contains one or more zones. Each zone has one or more members that are allowed to communicate among each other.

On Cisco MDS* switches, there is an option to do basic or enhanced zoning. Enhanced zoning advantages include the prevention of parallel configuration attempts which ensures consistency within the fabric, the distribution of zonesets without activation which avoid hardware changes for hard zoning on the switch, and the enhancement of error reporting to simplify the troubleshooting process.

N-Port ID Virtualization (NPIV)

NPIV allows an FC host connection or N-Port to be assigned multiple N-Port IDs or FC IDs over a single link. You can manage all FCIDs on an FC fabric as unique entities on the same physical host. You can use different applications in conjunction with NPIV. In a virtual machine environment where many host operating systems or applications run on a physical host, you can now manage each virtual machine independently from zone, alias, and security perspectives. In a Cisco MDS 9000* family environment, each host connection can log in as a single virtual SAN (VSAN).

N-Port Virtualizer (NPV)

An extension to NPIV, the N-Port Virtualizer (NPV) feature allows the blade switch or top-of-rack fabric device to behave as an NPIV-based host bus adapter (HBA) to the core FC director. The device aggregates the locally connected host ports or N-Ports into one or more uplinks (pseudo-interswitch links) to the core switches. The only requirement of the core director is that it supports the NPIV.

PortChannel

You can configure a PortChannel without restrictions to logically bundle physical links from any port on any Cisco MDS 9000 Family Fibre Channel Switching Modules*. This feature allows you to deploy highly available solutions with great flexibility. During a port, ASIC, or module failure, the stability of the network is not affected because the logical PortChannel remains active even though the overall bandwidth is reduced. The MDS PortChannel solution scales to support up to 16 ISLs per PortChannel and aggregates 1-, 2-, 4-, 8-, or 10-Gbps FC links. This feature aggregates up to 20,400 MB of application data throughput per PortChannel for exceptional scalability. The MDS PortChannel solution neither degrades performance over long distances nor requires specific cabling. The MDS PortChannel uses flow-based load balancing to deliver predictable and robust performance independent of the distance covered.

Summary of Device Configurations

VSAN

```
vsan 100 information
name:10g-topo state:active
interoperability mode:default
loadbalancing:src-id/dst-id/oxid
operational state:up
```

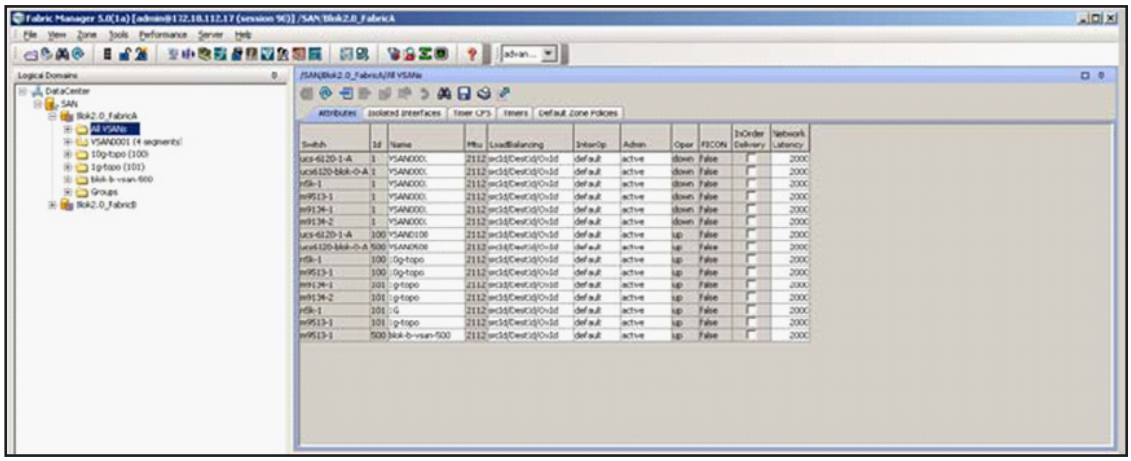


Figure 8. Cisco Fabric Manager (MDS 9000 SAN Switch) – Viewing Switch Inventory (Ports)

Zone/Zoneset

```
zoneset name 10g-fab-a vsan 100
zone name Host-1_to_V-Max vsan 100
* fcid 0x01002a [pwwn 20:00:00:25:b5:01:00:0f] [host-1]
* fcid 0x010015 [pwwn 50:00:09:72:08:1f:3d:64] [vmax-10Fa]
* fcid 0x010012 [pwwn 50:00:09:72:08:1f:3d:58] [vmax-7Fa]
* fcid 0x010013 [pwwn 50:00:09:72:08:1f:3d:5c] [vmax-8Fa]
* fcid 0x01000f [pwwn 50:00:09:72:08:1f:3d:60] [vmax-9Fa]
```

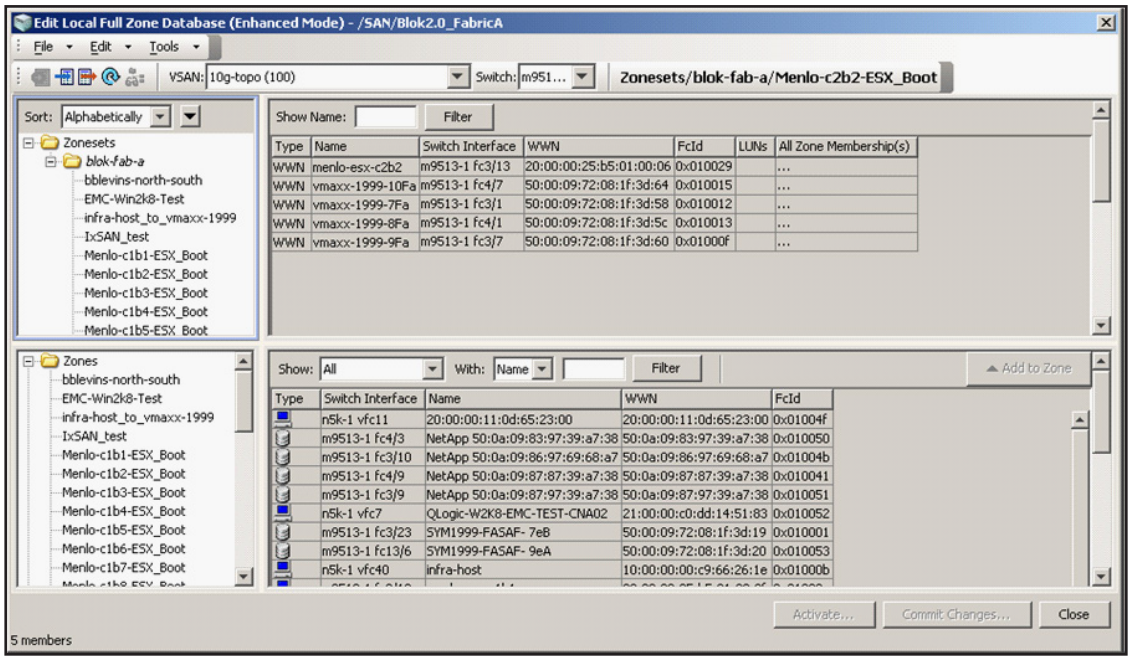


Figure 9. Cisco Fabric Manager (MDS 9000 SAN Switch) – Viewing Logical Port Associations

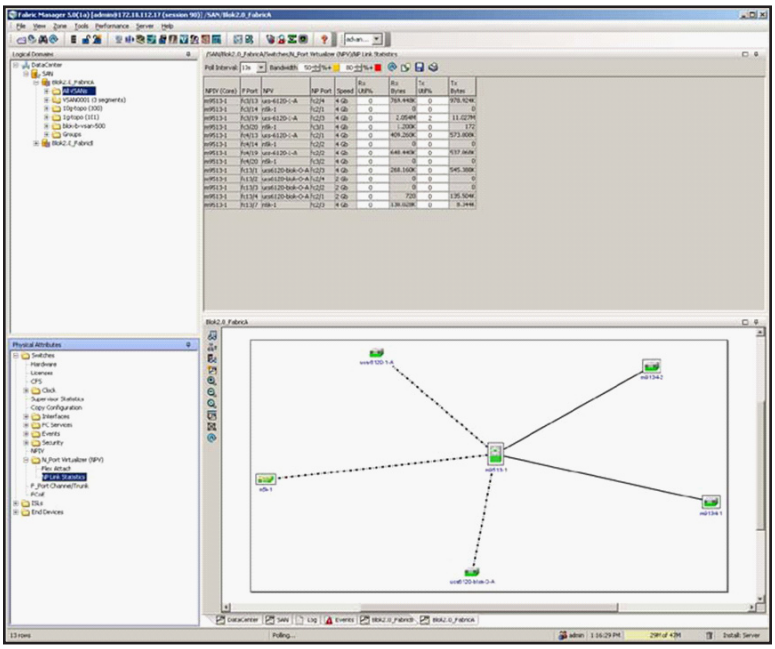


Figure 10. Cisco Fabric Manager (MDS 9000 SAN Switch) – Viewing Port Status

PortChannel

```
interface port-channel 1
switchport mode E
switchport rate-mode dedicated
switchport trunk mode off
port-channel 1 is up
Hardware is FC
Port WWN is 24:01:00:0d:ec:3b:b6:40
Admin port mode is E, trunk mode is off
snmp link state traps are enabled
Port mode is E
Port vsan is 101
Speed is 20 Gbps
…
Member[1] : fc1/1
```

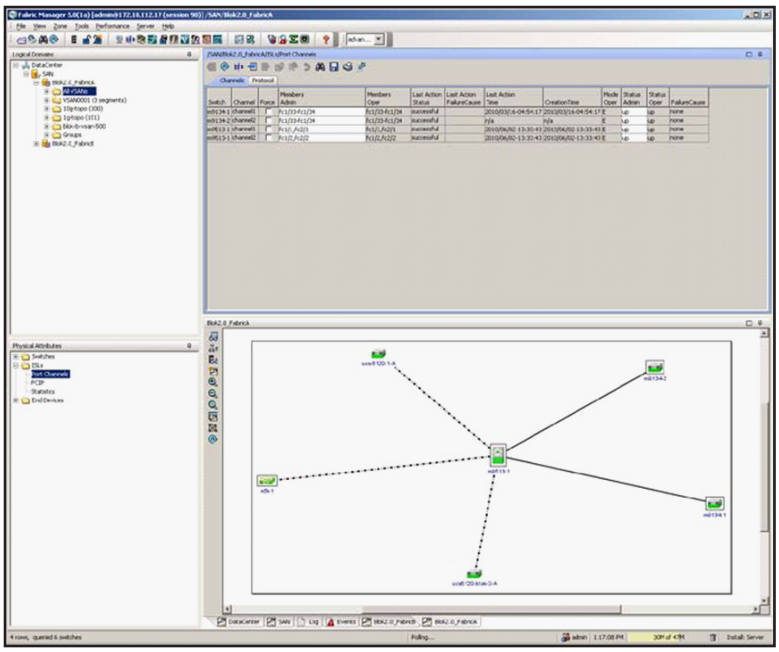


Figure 11. Cisco Fabric Manager (MDS 9000 SAN Switch) – Viewing Logical Storage Mappings

Device Alias

```
device-alias name host-1 pwwn 20:00:00:25:b5:01:00:0f
device-alias name vmax-7Fa pwwn 50:00:09:72:08:1f:3d:58
device-alias name vmax-8Fa pwwn 50:00:09:72:08:1f:3d:5c
device-alias name vmax-9Fa pwwn 50:00:09:72:08:1f:3d:60
device-alias name vmax-10Fa pwwn 50:00:09:72:08:1f:3d:64
```

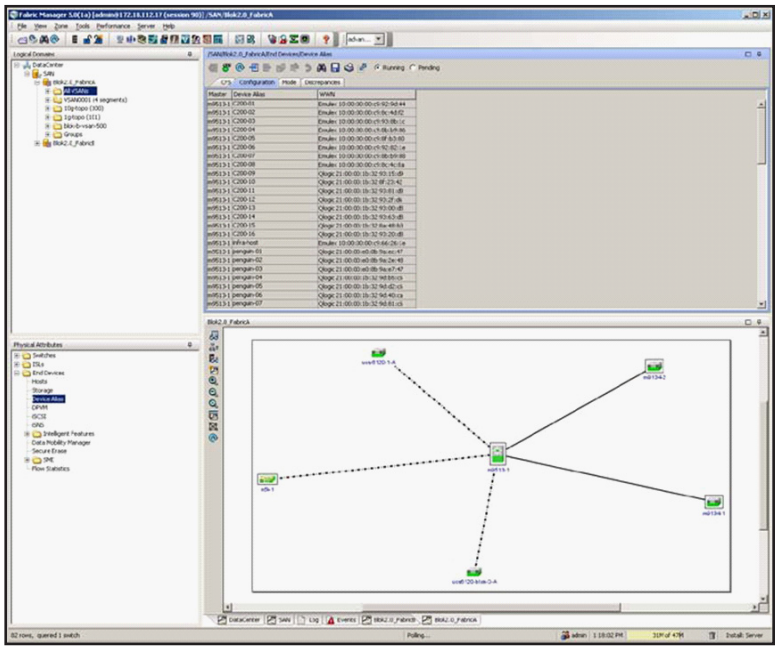


Figure 12. Cisco Fabric Manager (View of the Storage SAN)

Storage Layer - NAS

The NAS technologies, particularly NFS version 3, provide:

- No immediate cost for new infrastructure equipment
- Simplified storage provisioning
- Default support for VMware "thin provisioning"
- No disruptions during increases and decreases in space allocation
- Default support for large frame sizes

We used the NetApp* FAS6080* filer system for this solution. Through NFS, customers receive an integration of VMware* virtualization technologies with WAFL*, NetApp's advanced data management and storage virtualization engine. This integration provides transparent access to VM-level storage virtualization offerings, such as production-use data deduplication, immediate zero-cost VM and data store clones, array-based thin provisioning, automated policy based data store resizing, and direct access to array-based Snapshot* copies. NetApp provides integrated tools such as Site Recovery Manager*, SnapManager for Virtual Infrastructure*, the Rapid Cloning Utility*, and the Virtual Storage Console*.

Architecture Overview

When compared to FC and iSCSI, NFS provides higher performance and lower per-port storage costs. Both FC and iSCSI require the purchase of expensive adapters and even complete separate infrastructures. NFS requires only an additional file server (filer). We use NFS data stores in shared pools for virtual machines. The Cisco VMDC solution uses the NetApp FAS6080 server to provide support for NFS and VMware.

For vendor-specific guidance, see "NetApp and VMware vSphere Storage Best Practices" at the following URL: http://media.netapp.com/documets/tr-3749.pdf.

[NOTE: GUI screenshots are from the NetApp storage array, FilerView* and System Manager* tools, as well as from VMware ESX* client].

NFS Data Stores on NetApp

The deployment of VMware with NetApp's advanced NFS results in a high-performance, easy-to-manage implementation that provides VM-to-data-store ratios that you cannot accomplish through the use of other storage protocols, such as FC. This architecture can result in a ten-fold increase in data store density with a correlating reduction in the number of data stores. When you deploy NFS, the virtual infrastructure receives operational savings, as there are fewer storage pools to provision, manage, back up, and replicate.

For more information, see NetApp Document ID TR- 3749 Version 2.0 "NetApp and VMware vSphere Storage Best Practices."

Scenarios to Address

- Secure separation (end-to-end tenant security)
  - Logical datastore access separation
- Scalability
  - High-speed network connections
- High Availability
  - Hardware redundancy
  - Logical traffic path redundancy
- Service assurance
  - QoS
  - Distributed resource scheduling (DRS)
  - Reliable transmission
  - Efficient data delivery

Software Configuration

- Secure separation (end-to-end tenant security)
  - Logical data store access separation (see virtual access)
  - Cisco VLAN 802.1Q Trunking (See Network Design)
  - Cisco UCS VLAN support (See Compute_10G)
  - VMware support for 802.1Q trunking (See Virtual Access)
  - VMware vShield* zones (See Virtual Access)
  - VMware vSphere* support for NetApp NFS verson 3: each VMware ESXi* host, acting as NFS clients, must be configured with the correct export path.
  - NetApp NFS verson 3 support: filers, acting as NFS servers, must be configured to allow the client machines (ESXi hosts) access to the entire storage system.
  - NetApp 802.1Q* trunking support: filers need to be configured with VLAN-virtual interfaces that match the VLAN interfaces of the NFS clients.
  - NetApp Virtual Filer* (vFiler*): vFiler is not used in this solution; however, it allows users to access their own "virtual" filer that is under their control as they see fit. This "virtual filer" uses the physical resources of a single NetApp FAS6080.

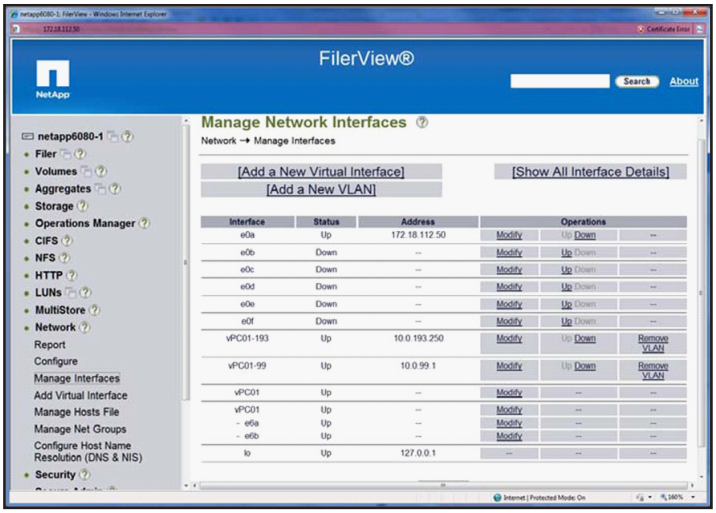


Figure 13. NetApp FilerView – Viewing Network Interfaces

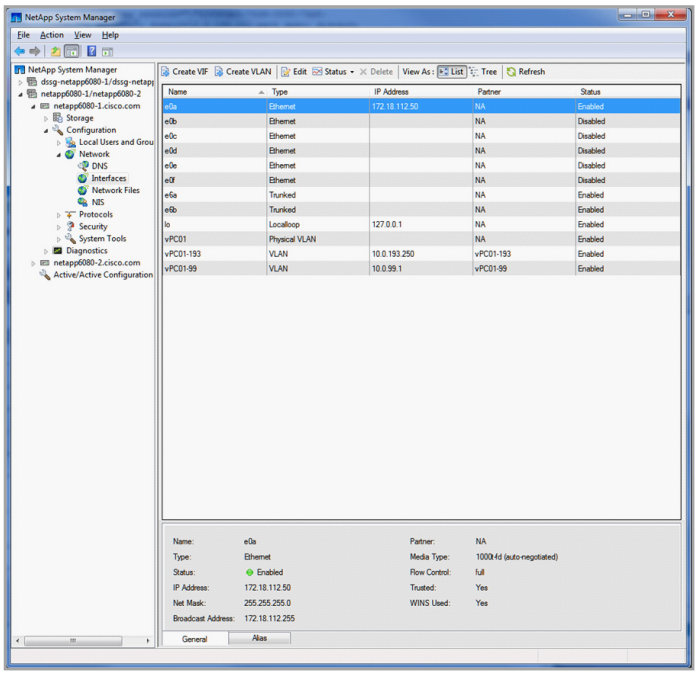


Figure 14. NetApp System Manager – Viewing Interface Associations

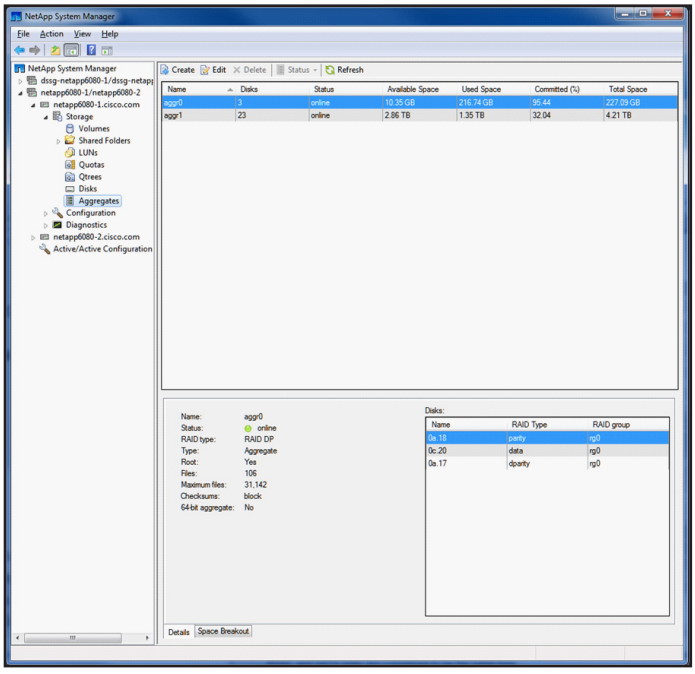


Figure 15. NetApp System Manager – Viewing Storage Aggregates

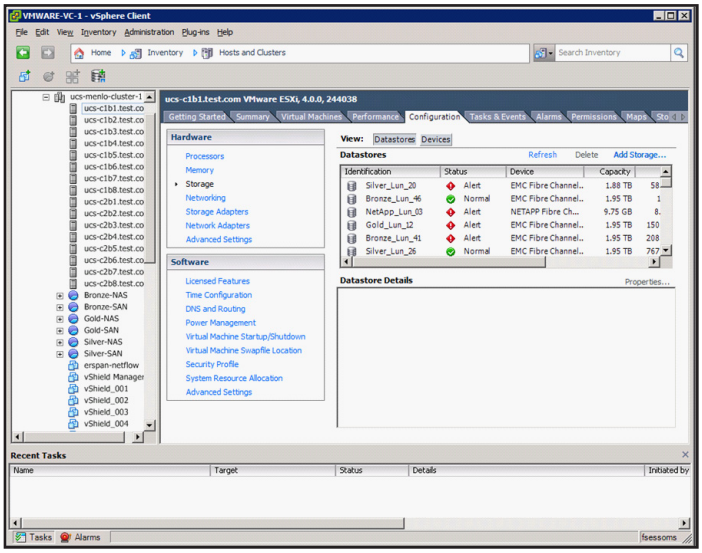


Figure 16. VMware vSphere ESX Client – Viewing VM to Server to Storage Mappings

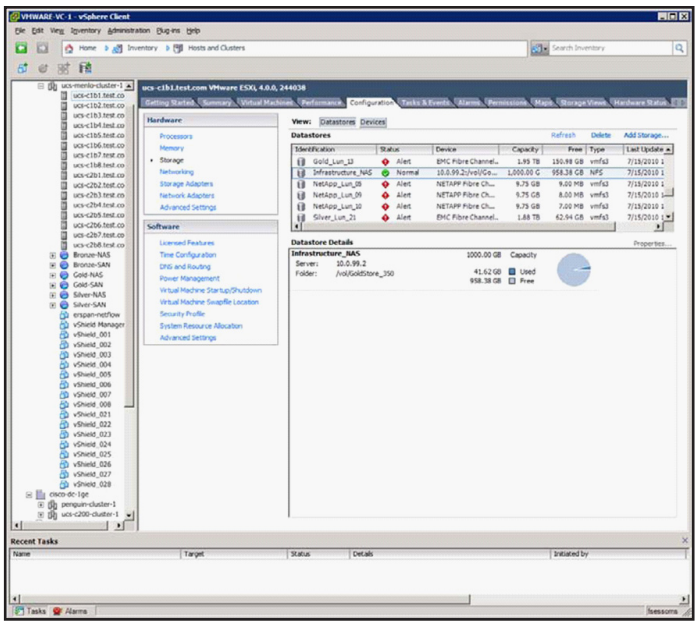


Figure 17. VMware vSphere ESX client – Viewing Storage Utilization

Next Steps

The current VMDC design provides a broad framework to deploy cloud computing today, but Cisco constantly seeks to innovate in ways that move the architecture forward and provide new capabilities for customers. This paper now examines several key elements that allow customers to look at their data center functionality in very new ways.

1 – Unified Network Services – While virtualization adds tremendous flexibility and efficiency to the data center, it also creates visibility challenges for IT. In the past, what was visible to network or security teams was often hidden as VMs communicated within a single server or as capabilities (switching, firewall, load-balancing) became services within the hypervisor. Cisco began to address this challenge in 2009 when it introduced the Nexus 1000v* virtual switch. This switch provided VM visibility to the network and security teams within the ESX hosts as well as vMotion* mobility. Cisco now extends the framework to move network services into a virtualized form factor as it adds firewall (Virtual Security Gateway*) and Application Acceleration (vWAAS*) capabilities, all managed through a single integration management framework (Cisco Virtual Network Management Center*). Over time, Cisco will expand the list of network services that can be virtualized, which will provide customers with enhanced visibility and control over their network in both physical and virtual data centers as well as in both software and hardware.

2 – Application Mobility – As customers begin to virtualize their server infrastructure, they understand the value of the ability to dynamically move a VM from one server to another (for example, vMotion* or Live Migration*). This mobility offers new ways to look at high availability and resource utilization. In the past, application mobility was limited to a single geographic location. But as customers begin to better understand the power of this functionality, they demand the ability to maximize this between geographic locations. Cisco has worked with ecosystem partners to create innovations aimed at solutions to this problem. Within the network, Cisco introduced Overlay Transport Virtualization* (OTV) which allows customers to seamlessly extend a Layer 2 domain across geographic locations. OTV allows customers to build these extended networks without the sacrifice of the control they expect within a single location. Beyond the network, Cisco has worked closely with ecosystem partners EMC and NetApp to extend this functionality across the storage layer and to allow mobility of the associated data.

Things to Consider

In any large-scale implementation, there are hundreds of design elements to consider. Cisco describes these in detail in Cisco Validated Designs (Cisco Design Zone - Data Center), but this paper will highlight some of the major considerations:

- Modular design—The use of a modular design allows new technologies to be added but will severely impact other portions of the data center.
- Hierarchical design—The use of a design that allocates certain functions (such as routing, switching, and security) to different layers will allow new services to be added in a consistent manner.
- Unified storage—The use of a design that supports multiple storage protocols (including FC, FCoE, iSCSI, NFS, CIFS) over a unified network fabric will allow the greatest flexibility to deliver multiple application services to users.
- Network-level intelligence—The use of a design that includes intelligent services within the network will allow for next-generation capabilities required by highly-virtualized environments (for application mobility and disaster avoidance) as well as mobile and video services.
- Unified infrastructure to unified IT—Virtualization technology blurs the lines between traditional IT groups such as applications, compute, storage, and networking. The utilization of technology designed to unify this migration will simplify the method to create IT processes that best manage the next-generation data center.

Design Considerations

The architects of a scalable cloud data center should consider the following requirements:

- Unified network
- Cloud tenants and service tiers
- Secure tenant separation
- Data center scalability through modular infrastructure
- High availability
- Service assurance

The following sections describe how the Cisco VMDC solution supports each of the requirements above.

Unified Network

The Cisco Unified Network* allows architects to create a single network infrastructure that is able to deliver data, storage, video, and voice traffic with predictable QoS. Unified networking begins with a 10 GbE high-speed transport. This transport maximizes Cisco QoS technology to deliver multiple classes of services, from deterministic to best-effort.

As data and storage traffic aggregate on the 10 GbE links, FCoE technology is able to guarantee the service levels and management that storage administrators have come to expect with FC. FCoE within the Cisco Unified Computing System* (UCS) and extended to the SAN Storage arrays allows LAN and SAN traffic to co-exist over a single cabling infrastructure. This simplified cabling reduces costs for cabling, patch-panels, cable-trays and NIC/HBAs (now called Converged Network Adapters (CNAs)). This reduction of cabling

also eliminates demands for power and cooling within the computing and storage systems, which reduces overall data center costs.

Beyond unification of the underlying 10 GbE transport, the Cisco VMDC makes use of virtualized services within the network. Called Unified Network Services*, these are delivered as VM appliances within the Cisco Unified Computing System. These include the Nexus 1000v vSwitch, the VSG firewall services and vWAAS. These virtualized services provide control and visibility to the network administrator that is often lost when virtualizing applications.

Cloud Tenants and Service Tiers

A tenant is an entity that subscribes to cloud services. In the enterprise private cloud deployment model, that entity is a department or sub-organization, such as development, test, research and development, or human resources. As shown in Figure 18, multiple users in the same department belong to the same tenancy. Within the tenancy, you can implement multiple workloads by different users who belong to the same department.

You must securely separate each tenant from other tenants who share the common virtualized resource pool. However, workloads owned by one tenant will be visible to others unless you configure firewalls to block communications among different applications.

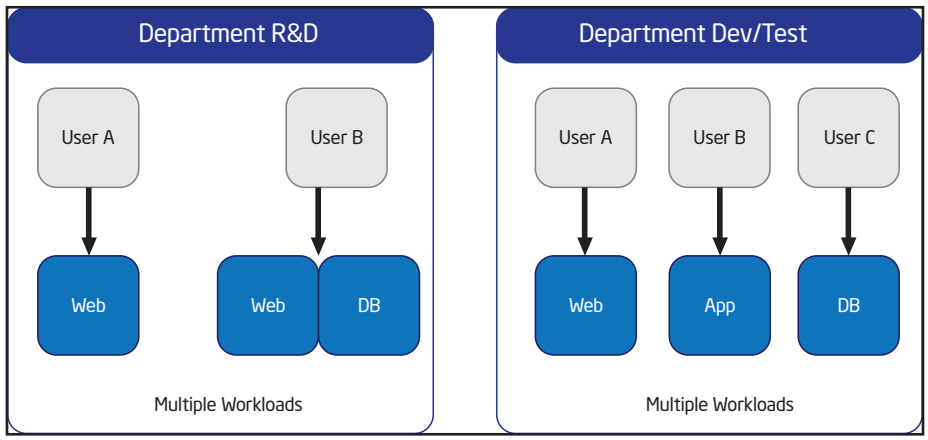


Figure 18. Tenants and Workloads

In the public cloud deployment model, a tenant is an individual consumer or group within an enterprise subscribing to the virtual private cloud services hosted by a service provider.

Cloud providers, whether service providers or enterprises, want an IaaS offering with multiple feature tiers and pricing levels. The cloud is a source of highly scalable, efficient, and elastic services accessed on-demand over the Internet or intranet. In the cloud, compute, storage, and network hardware are abstracted and delivered as a service. End users only consider the functionality and value provided by the service; they do not need to understand or manage the underlying technology.

	Bronze	Silver	Gold
Services	No additional services	Firewall services	Firewall and load balancing services
Bandwidth	20 percent	30 percent	40 percent
Segmentation	One VLAN per client, single VRF	Multiple VLANs per client, Single VRF	Multiple VLANs per client, single VRF
Data Protection	None	Snap – virtual copy (local site)	Clone - mirror copy (local site)
Disaster Recovery	None	Remote replication (With specific RPO/ RTO)	Remote replication (any-point in-time recovery)

Table 4. Example Network and Data Differentiations by Service Tier

To tailor workload or application requirements to specific customer needs, the cloud provider can differentiate services with a multi-tiered service infrastructure and QoS settings. Customers can use and purchase such services under a variable pricing model. Administrators can design infrastructure and resource pools so that end users can add or expand services when they request additional compute, storage, or network capacity. This elasticity allows the provider to maximize the user experience by with the offer of a custom, private data center in virtual form.

Typically, cloud providers want to offer three, four, or five different service tiers and provide different service level agreements (SLAs). You can differentiate IaaS cloud services into pre-defined service tiers as they vary support of the following features:

- Virtual machine resources
- Storage features
- Application tiers
- Stateful services
- Quality of service agreements

The VMDC solution defines options to differentiate IT cloud services. In this reference architecture, we call these cloud services services tiers. Typically when we talk about service tiers, we look at the server CPU and storage options.

But if a web application is being hosted in the cloud model, load balancing and firewall inspection are also required. To achieve secure separation of tenant data Layer 2 and Layer 3, you must enable features such as virtual routing and forwarding (VRF) and VLANs. With this virtual network separation configured, service tiers contain virtual compute, storage, and network resources.

This Cisco VMDC solution qualifies a three-tier model of Bronze, Silver, and Gold tiers that comprise IaaS services (See Table 4). These tiers define service levels for compute, storage, and network performance.

With the use of this tiered model, you can do the following:

- Offer service tiers with differing abilities
- Support customer segmentation based on desired service levels and functionality
- Support clients based on their requirements
- Allow for differentiated application support based on service tiers

Secure Tenant Separation

The following design considerations provide secure tenant separation[5] and path isolation:

- Network separation
- Compute separation
- Storage separation

Network Separation

End-to-end virtualization of the network requires separation at each network layer in the architecture. The VMDC design uses the following technologies to virtualize the network:

- Implementation of network Layer 3 (core/aggregation) separation is through the use of Cisco VRF Lite
- Implementation of network Layer 2 (access) separation is through the use of VLANs
- Implementation of network services (firewall and load balancing services) separation is through the use of the Cisco FWSM and the Cisco ACE service module
- Implementation of the tenant path isolation is through the use of virtual private networks (VPNs) using VRF Lite technology
- Implementation of client-server traffic separation is through the use of the following technologies:
  - Dedicated virtual firewall context on the firewall module that belongs to a particular tenant is used to provide traffic inspection
  - DOS attack prevention
  - L4-7 protocol inspection
  - ACLs to control what comes though the firewall
  - The VMDC firewall model employs a tiered model. The service tiers are mapped to secured access mechanisms, which include secure sockets layer (SSL), multiprotocol label switching (MPLS), and Internet protocol security (IPSec) VPNs.

Compute Separation

Virtualization introduces new security challenges and concerns. Traditionally, security policies were applied at the physical server level. However, as physical hosts can now contain multiple logical servers, policy must be applied at the VM level. Also, new technologies, such as vMotion, introduced VM mobility within a cluster where policies follow VMs as they are moved across switch ports and among hosts.

Finally, virtual computing continues to aggregate higher densities of VMs. This high-density model forces us to reconsider firewall scale requirements at the aggregation layer of the network. The result is that high-density compute architectures may require the distribution of security policies to the access layer.

To address some of these new security challenges and concerns, we recommend the deployment of virtual firewalls at the access layer of the data center infrastructure to create intra-tenant zones. You should also use per-VLAN firewalls at the aggregation layer. Like a firewall at the aggregation layer, Layer 2 firewalling can enforce security among the tiers of an application.

Storage Separation

Cisco MDS SAN networks offer segmentation mechanisms similar to VLANs in Ethernet. Cisco calls these mechanisms VSANs and they work in conjunction with fibre channel (FC) zones.

Data Center Scalability

Capacity at scale, or elasticity, is an essential attribute of clouds. Elasticity is the ability to scale resources up or down in minutes in accordance with service level agreements (SLAs). Elasticity increases resources on-demand and scales resource utilization as needed. This design uses a concept called Pod to achieve elasticity and to simplify capacity planning without the disruption of the existing environment. A Pod identifies a discrete, homogenous, modular unit of data center components. Because they are homogenous and modular, Pods support templates for incremental build-out of the data center that address environmental, physical, logical, and application requirements. This modular architecture provides a predictable set of resource characteristics per unit that is added repeatedly as needed.

Initially, a customer implements the data center through the use of a base Pod and expands the data center with the addition of more Pods. In the data center, Pod-based architectures provide predictable resource pools, power, and space consumption. As shown in Figure 19, the core layer is common to multiple Pods; as additional Pods are needed, the customer connects them to the network through the core layer.

High Availability

Cloud data centers offer IaaS to either internal enterprise customers or to external customers of service providers. Customers control the services through the use of SLAs, which can be stricter in service provider deployments than in an enterprise. A highly available data center infrastructure is the foundation of SLA guarantees and successful cloud deployment.

An end-to-end, highly available network infrastructure design provides predictable operational continuity. Because organizations must satisfy SLAs made for business application uptime, they cannot lose connectivity due to equipment downtime. Therefore, the data center design must ensure that a single hardware failure in the network does not affect the cloud subscribers' service. The Cisco VMDC utilizes the following techniques to create a highly available and resilient end-to-end infrastructure.

- Physical redundancy
  - Node redundancy
  - Hardware redundancy with the node
  - Link redundancy
- Access layer technologies such as system posture tokens (SPTs)
- Compute layer technologies
  - UCS end-host mode
  - Cisco Nexus 1000V and MAC-pinning
  - Redundant VSMs in active-standby mode
  - High availability within the cluster
  - Automated disaster recovery plans

Service Assurance

Today, IaaS SLAs often emphasize service availability. Differentiated service levels requirements exist because specific applications or traffic may require preferential treatment within the cloud. Some applications are mission critical, and some are interactive, while others are bulk or utilized simply for dev-test purposes. This differentiated treatment ensures that in the event of congestion or failure conditions, critical traffic is provided a sufficient amount of bandwidth to meet throughput requirements. Traditionally, an SLA framework includes consideration of bandwidth, delay, jitter, and loss per service class.

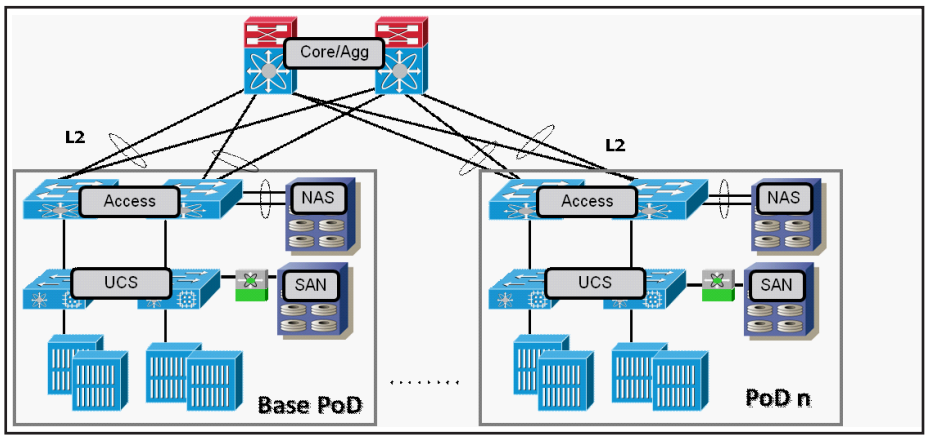


Figure 19. Pod Expansion Concept

Glossary

ACE: The Cisco ACE service module provides server load balancing and source NAT (SNAT).

DSN: The Cisco Data Center Services Node (DSN) is a Cisco Catalyst 6500 Series Switch with FWSM and ACE service modules dedicated to security and server load balancing functions.

FCoE: Fibre Channel over Ethernet Initialization Protocol offers the capability to transport Fibre Channel payloads on top of an Ethernet network.

FWSM: The Cisco Firewall Services Module (FWSM) provides Layer 2 and Layer 3 firewall inspection, protocol inspection, and network address translation (NAT).

IaaS: Cloud infrastructure services, also known as "Infrastructure as a Service (IaaS)", delivers computer infrastructure - typically a platform virtualization environment - as a service. Rather than purchasing servers, software, data center space, or network equipment, clients instead buy those resources as a fully outsourced service. Suppliers typically bill such services on a utility computing basis and the amount of resources consumed (and therefore the cost) will typically reflect the level of activity. From: http://en.wikipedia.org/wiki/Cloud_computing.

Intel AES: Intel® Advanced Encryption Standard (AES) Technology enables robust encryption without the need for additional appliances and increased performance overhead. This technology improves CPU performance for encryption by as much as a 52 percent for secure Internet transactions and allows broader use of encryption throughout the data center.

Intel QPI: Intel® QuickPath Technology is a scalable, shared memory architecture that delivers a high memory bandwidth to enable top performance for bandwidth-intensive applications. It provides high speed point-to-point connections between processors, and between processors and the I/O hub. Each processor has its own dedicated memory that it accesses directly through an Integrated Memory Controller. In cases where a processor needs to access the dedicated memory of another processor, it can do so through a high-speed Intel® QuickPath Interconnect (Intel QPI) that links all the processors.

Intel® TXT: Intel® Trusted Execution Technology (Intel® TXT) addresses a critical security need for all server deployments, especially virtualized and cloud-based use models as it helps to protect servers prior to OS launch or hypervisor launch. Intel® TXT complements other malware protections such as anti-virus and intrusion detection to help ensure that only trusted software is on the platform. Intel® TXT protects VMs on trusted platforms, so you can easily migrate them onto other trusted platforms or created pools of platforms with trusted hypervisors.

Intel VT-c: Intel® Virtualization Technology for Connectivity (Intel VT-c) enhances server I/O solutions by integrating extensive hardware assists into the I/O devices that are used to connect servers to the data center network and storage infrastructure. Two technologies comprise Intel VT-c: Virtual Machine Device Queues, which accelerates throughput and reduces the load on the VMM and server processors, and PCI-SIG SR-IOV, which delivers near-native throughput and provides dedicated, direct connectivity between VMs and hardware resources.

Intel VT-d: Intel® Virtualization Technology for Directed I/O (Intel VT-d) helps speed data movement and eliminates much of the performance overhead as it gives designated virtual machines their own dedicated I/O devices, which reduces the overhead of the VMM to manage I/O traffic.

Intel VT-x: Intel® Virtualization Technology (Intel VT-x) provides hardware assisted page-table management, which allows guest OS more direct access to the hardware and reduces compute-intensive software translation from the VMM. Intel VT-x also includes Intel VT FlexMigration and Intel VT Flex-Priority, which are capabilities for flexible workload migration and performance optimization across the full range of 32-bit and 64-bit operating environments.

NAS: Network Attached Storage is a storage server or appliance that uses file-based protocols such as NFS or CIFS to enable clients (typically servers and PCs) to access files over a TCP/IP network. From: http://en.wikipedia.org/wiki/Network-attached_storage.

NIC: A network interface card is hardware that enables a server to interface to an Ethernet or TCP/IP local area network (LAN). An NIC is not necessarily a card in the server; it can be integrated as LAN on a server motherboard (LOM).

PaaS: Cloud platform services or "Platform as a Service (PaaS)" deliver a computing platform and/or solution stack as a service, which often consumes cloud infrastructure and sustains cloud applications. It facilitates deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layer. From: http://en.wikipedia.org/wiki/Cloud_computing.

Pod: Cisco VMDC Pod-based architecture provides network architects the ability to modularize the infrastructure into easily replicable units called a point of delivery (Pod). Architects can plan for an initial Pod, which guarantees a certain scale and performance along with a scalable data center core network. This architecture provides a predictable and homogeneous method to add self-contained Pods as additional resources are necessary.

QoS: Quality of Service (QoS) is a mechanism to define cloud service quality. To tailor workload or application requirements to specific customer needs, the cloud provider can differentiate services with a multi-tiered service infrastructure and quality of service (QoS) settings. Customers can use and purchase such services under a variable pricing model.

SaaS: Software as a Service (SaaS) delivers software as a service over the Internet, which eliminates the need to install and run the application on the customer's own computers and simplifies maintenance and support. From: http://en.wikipedia.org/wiki/Cloud_computing

SAN: A storage area network is a storage server or appliance that uses block-based protocols typically based on SCSI to access files over a Fibre Channel or TCP/IP network. From: http://en.wikipedia.org/wiki/Storage_area_network.

Service Tier: Cisco VMDC reference architecture defines a three-tier model of Bronze, Silver, and Gold tiers that compromises IaaS services. These tiers define service levels for compute, storage, and network performance.

SLA: Service level agreements (SLAs) define the support levels in cloud services. Typically, cloud providers want to offer three, four, or five different service tiers and provide different service level agreements (SLAs). IaaS cloud services can be differentiated into pre-defined service tiers when they vary support of the following features: virtual machine resources, storage features, application tiers, stateful services, and network bandwidth.

Tenant: A tenant is an entity that subscribes to cloud services. Each tenant must be securely separated from other tenants who share the common virtualized resource pool. However, workloads owned by one tenant will be visible to others unless firewalls are configured to block communications among different applications.

UCS: Cisco Unified Computing System(UCS) is a data center platform that unites computing, networking, storage access, and virtualization into a cohesive system, and integrates a low-latency, lossless 10 Gigabit Ethernet unified network fabric with enterprise-class, x86-architecture servers.

vCPU: Virtual CPU (vCPU) is an entity that corresponds to a physical CPU in a guest VM. If the system has "n" cores, then the maximum number of vCPUs that can be allocated to a guest is "n."

VLAN: A virtual LAN, commonly known as a VLAN, is a group of hosts with a common set of requirements that communicate as if they were attached to the same broadcast domain, regardless of their physical location. A VLAN has the same attributes as a physical LAN, but it allows for end stations to be grouped together even if they are not located on the same network switch. You can achieve network reconfiguration through software instead of by the physical relocation of devices. From: http://en.wikipedia.org/wiki/VlanVRF.

VM: A virtual machine (VM) is a software implementation of a machine (i.e. a computer) that executes instructions (not programs) like a physical machine. From: http http://en.wikipedia.org/wiki/Virtual_machine.

vMotion: vMotion technology enables VM mobility within a cluster, where policies follow VMs as they move across switch ports and among hosts.

VMDC: Cisco Virtualized Multi-tenant Data Center (VMDC) solution is a reference IaaS architecture that brings together core products and technologies from Cisco, NetApp, EMC, BMC, and VMware to deliver a comprehensive end-to-end cloud solution. Focused on IaaS cloud deployment, the Cisco Virtualized Multi-tenant Data Center (VMDC) solution, version 2.0, provides customers with robust, scalable, and resilient options for cloud data center deployments. This Cisco-driven, end-to-end architecture defines how to provision flexible, dynamic pools of virtualized resources that you can share efficiently and securely among different tenants and provision quickly through process automation. Process automation reduces resource provisioning and improves time-to-market (TTM) for IaaS-based services. Shared resource pools consist of virtualized Cisco unified compute and virtualized SAN and NAS storage platforms connected through the use of Cisco data center switches and routers.

VRF: In IP-based computer networks, Virtual Routing and Forwarding (VRF) is a technology that allows multiple instances of a routing table to co-exist within the same router at the same time. Because the routing instances are independent, you can use the same or overlapping IP addresses without conflicting with each other. The simplest form of VRF implementation is VRF Lite. In this implementation, each router within the network participates in the virtual routing environment in a peer-based fashion. From: http://en.wikipedia.org/wiki/VRF.

VPN: A virtual private network (VPN) is a computer network that uses a public telecommunication infrastructure such as the Internet to provide remote offices or individual users with secure access to their organization's network. It aims to avoid an expensive system of owned or leased lines accessible to only one organization. From: http://en.wikipedia.org/wiki/Virtual_private_network.

VSG: Cisco Virtual Security Gateway* (VSG*) for Cisco Nexus 1000V Series Switches* is a virtual appliance that secures and provides trusted access to virtualized data centers in enterprise and cloud provider environments and at the same time meets the requirements of dynamic policy-driven operations, mobility-transparent enforcement, and scale-out deployment for dense multi-tenancy.

vWAAS: Cisco Virtual Wide Area Application Services (vWAAS) is a cloud-ready WAN optimization solution that accelerates applications delivered from private and virtual private cloud infrastructure, through the use of policy-based on-demand orchestration.

Endnotes

1. Security and Virtualization in the Data Center http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DC_3_0/dc_sec_design.html

2. Intel Xeon processors: http://www. intel.com/xeon

3. Intel Internal measurements using a web banking workload running PHP and Windows Server 2008 R2, comparing number of banking sessions (users) for an Intel® Xeon® processor X5689 (3.33 GHz) vs. Intel® Xeon® processor X5570 (2.93 GHz).

4. Data Center Design—IP Network Infrastructure http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DC_3_0/DC-3_0_IPInfra.html#wp1043848 Data Center Service Patterns http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DC_3_0/dc_serv_pat.html

5. Designing Secure Multi-Tenancy into Virtualized Data Centers http://www.cisco.com/en/US/netsol/ns1050/networking_solutions_sub_program_home.html

To learn more about deployment of cloud solutions,
visit www.intel.com/cloudbuilders

Disclaimers

Printed in USA 1010/SK/PRW/PDF Please Recycle 324446-001 US