



CHAPTER 1

Architecture Overview

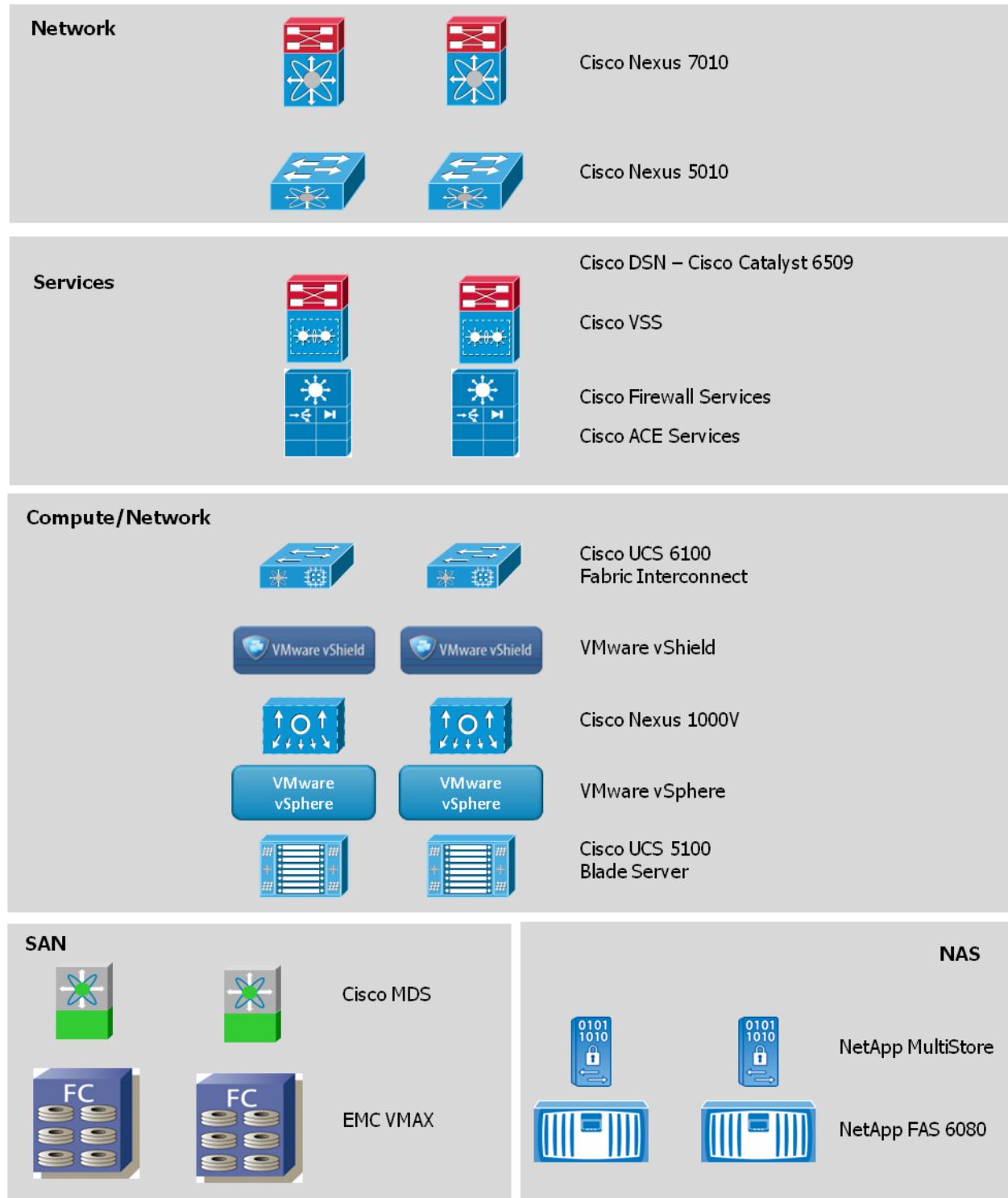
Revised: April 26, 2011

A cloud deployment model differs from traditional deployments in its ability to treat the data center as a common fabric of resources. A portion of these resources can be dynamically allocated and then de-allocated when they are no longer in use. Cisco VMDC leverages basic key building blocks:

- **Shared resource pools.** The resource pools consists of three main components: network, compute, and storage. Each of these components is virtualized so that each cloud tenant appears to have its own set of physical resources.
- **Service orchestration.** Service orchestration uses a set of tools and APIs to automate the provisioning process by using a predefined workflow. Service orchestration is presented as a web portal from which an end user can request specific resources from the cloud.
- **Business continuance.** Business continuity strives to ensure that essential functions can continue during and after a disaster. Business continuance planning seeks to prevent interruption of mission-critical services and to reestablish full functionality as swiftly and smoothly as possible.

Solution Key Components

Figure 1-1 Key Components of the Cisco VMDC 2.0 Solution, Compact Pod Design



Network

The following components were used in the network layer of the VMDC Compact Pod:

- [Cisco Nexus 7000](#), page 1-3
- [Cisco Datacenter Services Node \(DSN\)](#), page 1-3
- [Cisco Nexus 5000](#), page 1-4
- [Cisco Nexus 1000V](#), page 1-5

Cisco Nexus 7000

The Cisco Nexus 7000 Series is a modular switching system designed to deliver 10 Gigabit Ethernet and unified fabric. Designed for the core and aggregation layers of the data center, it delivers exceptional scalability, continuous operation, and transport flexibility.

It runs the Cisco NX-OS operating system (<http://www.cisco.com/en/US/products/ps9372/index.html>). For more information, see: <http://www.cisco.com/en/US/products/ps9402/index.html>.

Cisco Datacenter Services Node (DSN)

The DSN is an orderable option that includes a pair of Catalyst 6509-E chassis with the following modules:

- [Cisco Catalyst 6500 Virtual Switching System 1440](#), page 1-3
- [Cisco Firewall Services Module \(FWSM\)](#), page 1-4
- [Cisco Application Control Engine \(ACE\)](#), page 1-4

Cisco Catalyst 6500 Virtual Switching System 1440

The Cisco Catalyst 6500 Series Virtual Switching System (VSS) 1440 merges two physical Cisco Catalyst 6500 Series Switches into a single, logically managed entity. The key enabler of a VSS 1440 is the Virtual Switching Supervisor 720-10G. Once a VSS 1440 is created, it acts as a single virtual Catalyst switch delivering the following benefits:

- **Operational Manageability.** Two Catalyst 6500s share a single point of management, single gateway IP address, and single routing instance eliminating the dependence on First Hop Redundancy Protocols (FHRP) and Spanning Tree Protocols.
- **Availability.** Delivers deterministic, sub-200 millisecond Layer 2 link recovery through inter-chassis stateful failovers and the predictable resilience of EtherChannel.
- **Scalability.** Scales system bandwidth capacity to 1.4 Tbps by activating all available bandwidth across redundant switches.

The VSS platform supports Cisco integrated service modules, such as the Cisco Application Control Engine (ACE) and Firewall Services Module. It supports both 1- and 10-Gbps Ethernet devices allowing for network-based services.

Cisco Firewall Services Module (FWSM)

The Cisco Firewall Services Module (FWSM) is a stateful firewall residing within a Catalyst 6500 switching platform. The FWSM module supports device-level redundancy and scales through multiple virtual security contexts. A virtual security context can be transparent at Layer 2 or addressable at Layer 3. With either deployment model, the security policies associated with each virtual context are consistently applied to protect the related data center networks.

For more information, see:

<http://www.cisco.com/en/US/products/hw/modules/ps2706/ps4452/index.html>.

Cisco Application Control Engine (ACE)

The Cisco Application Control Engine (ACE) module performs server load balancing, network traffic control, service redundancy, resource management, encryption and security, and application acceleration and optimization. It provides device- and network service-level availability, scalability, and security features to the data center.

- The Cisco ACE offers the following device-level services:
- Physical redundancy with failover capabilities for high availability
- Scalability through virtualization allows ACE resources to be logically partitioned and assigned to meet specific tenant service requirements
- Security via access control lists and role-based access control

Network service levels support the following:

- Application availability through load balancing and health monitoring of the application environments
- Scalability of application load balancing, health monitoring, and session persistence policies as all are locally defined within each ACE virtual partition
- Security services, including ACLs and transport layer encryption (SSL/TLS) between the ACE virtual context, client population, and associated server farm

For more information, see:

http://www.cisco.com/en/US/products/ps5719/Products_Sub_Category_Home.html.

Cisco Nexus 5000

The Cisco Nexus 5000 Series (<http://www.cisco.com/en/US/products/ps9670/index.html>) switches deliver high performance, standards-based Ethernet and FCoE that enable the consolidation of LAN, SAN, and cluster network environments onto a single Unified Fabric. Each switch contains a single unified crossbar fabric ASIC and multiple unified port controllers to support fixed ports and expansion modules.

The unified port controller provides an interface between the unified crossbar fabric ASIC and the network media adapter and makes forwarding decisions for Ethernet, Fibre Channel, and FCoE frames. The ASIC supports the overall cut-through design of the switch by transmitting packets to the unified crossbar fabric before the entire payload has been received. The unified crossbar fabric ASIC is a single-stage, non-blocking crossbar fabric capable of meshing all ports at wire speed. The unified crossbar fabric implements QoS-aware scheduling for unicast and multicast traffic. Integration of the unified crossbar fabric with the unified port controllers ensures low latency lossless fabric for ingress interfaces requesting access to egress interfaces.

For more information, see: <http://www.cisco.com/en/US/products/ps9670/index.html>.

Cisco Nexus 1000V

The Nexus 1000V software switch delivers Cisco VN-Link services to virtual machines hosted on the server where the switch resides. Built on the VMware vSphere framework, it integrates server and network environments to ensure consistent, policy-based network capabilities to all servers in the data center. The Nexus 1000V aligns management of the operational environment for virtual machines and physical server connectivity in the data center, which enables a policy to follow a virtual machine during live migration, ensuring persistent network, security, and storage compliance.

For more information, see: <http://www.cisco.com/en/US/products/ps9902/index.html>.

For more information on Cisco VN-Link technologies see:

<http://www.cisco.com/en/US/netsol/ns894/index.html>.

Compute

The following components were used in the network layer of the VMDC Compact Pod:

- [Cisco UCS and UCSM, page 1-5](#)
- [VMware vSphere and vCenter Server, page 1-6](#)
- [VMware vShield Zones, page 1-6](#)

Cisco UCS and UCSM

The Cisco Unified Computing System (UCS) unites compute, network, storage access, and virtualization into a cohesive system. The system integrates a low-latency, lossless 10-Gigabit Ethernet unified network fabric with x86-architecture servers. All integrated resources participate in a unified management domain, whether it has one or 320 servers with thousands of virtual machines. The Cisco UCS accelerates the delivery of new services via end-to-end provisioning and migration support for both virtualized and non-virtualized systems.

The Cisco Unified Computing System is built from the following components:

- Cisco UCS 6100 Series Fabric Interconnects (<http://www.cisco.com/en/US/partner/products/ps10276/index.html>) is a family of line-rate, low-latency, lossless, 10-Gbps Ethernet and Fibre Channel over Ethernet interconnect switches.
- Cisco UCS 5100 Series Blade Server Chassis (<http://www.cisco.com/en/US/partner/products/ps10279/index.html>) supports up to eight blade servers and up to two fabric extenders in a six rack unit (RU) enclosure.
- Cisco UCS 2100 Series Fabric Extenders (<http://www.cisco.com/en/US/partner/products/ps10278/index.html>) bring unified fabric into the blade-server chassis, providing up to four 10-Gbps connections each between blade servers and the fabric interconnect.
- Cisco UCS B-Series Blade Servers (<http://www.cisco.com/en/US/partner/products/ps10280/index.html>) adapt to application demands, intelligently scale energy use, and offer best-in-class virtualization.
- Cisco UCS B-Series Network Adapters (<http://www.cisco.com/en/US/partner/products/ps10280/index.html>) offer a range of options, including adapters optimized for virtualization, compatibility with existing driver stacks, or efficient, high-performance Ethernet.

- Cisco UCS Manager (<http://www.cisco.com/en/US/partner/products/ps10281/index.html>) provides centralized management capabilities for the Cisco Unified Computing System.

For more information, see: <http://www.cisco.com/en/US/partner/netsol/ns944/index.html>.

VMware vSphere and vCenter Server

VMware vSphere and vCenter Server provides centralized control and visibility at every level of the virtual infrastructure and provisions service delivery and application service agreements.

VMware vCenter Server provides a scalable and extensible platform that forms the foundation for virtualization management (<http://www.vmware.com/solutions/virtualization-management/>). VMware vCenter Server, formerly VMware VirtualCenter, centrally manages VMware vSphere (<http://www.vmware.com/products/vsphere/>) environments.

For more information, see <http://www.vmware.com/products/>.

VMware vShield Zones

VMware vShield Zones is a centrally managed, stateful, distributed virtual firewall bundled with vSphere 4.x, which takes advantage of ESXi host proximity and virtual network visibility to create security zones. By leveraging various VMware logical containers, it is possible to greatly reduce the number of rules required to secure a multi-tenant environment and therefore reduce the operational burden that accompanies the isolation and segmentation of tenants and applications. This new way of creating security policies closely ties to the VMware virtual machine objects and therefore follows the VMs during vMotion and is completely transparent to IP address changes and network re-numbering. Using vShield Zones within DRS (Distributed Resource Scheduler) clusters ensures secure compute load-balancing operations without performance compromise as the security policy follows the virtual machine.

In addition to being an endpoint and asset aware firewall, the vShield Zones contain microflow-level virtual network reporting that is critical to understanding and monitoring the virtual traffic flows and implement zoning policies based on rich information available to security and network administrators. This flow information is categorized into allowed and blocked sessions and can be sliced and diced by protocol, port and application, and direction and seen at any level of the inventory hierarchy. It can be further used to find rogue services, prohibited virtual machine communication, serve as a regulatory compliance visualization tool, and operationally to troubleshoot access and firewall rule configuration. Flexible user configuration allows role-based duty separation for network, security, and vSphere administrator duties.

The Flow Monitoring feature displays Allowed and Blocked network flows at application protocol granularity. This can be used to audit network traffic and as an operational troubleshooting tool.

For more information, see: <http://www.vmware.com/products/vshield-zones/>.

Storage

The following components were used in the storage layer of the VMDC Compact Pod:

- Cisco MDS 9513, page 1-7
- Cisco MDS 9134, page 1-7
- Cisco MDS 9148, page 1-7
- Cisco Management Interface, page 1-8

- [EMC Symmetrix VMAX, page 1-8](#)
- [EMC Powerpath VE, page 1-9](#)
- [NetApp FAS6080 Filer, page 1-9](#)
- [NetApp Snapshot, page 1-9](#)
- [NetApp Deduplication, page 1-9](#)
- [NetApp Rapid Clone Utility, page 1-9](#)

Cisco MDS 9513

The Cisco MDS 9513 Multilayer Director allows you to deploy high-performance SANs using a high-performance, protocol-independent switch fabric. It provides uncompromising high availability, security, scalability, ease of management, and transparent integration of new technologies for extremely flexible data center SAN solutions. The Cisco MDS 9513 is compatible with first-, second-, and third-generation Cisco MDS 9000 Family switching modules.

For more information, see: <http://www.cisco.com/en/US/products/hw/ps4159/index.html>.

Cisco MDS 9134

The Cisco MDS 9134 Multilayer Fabric Switch is a 1RU chassis that provides line-rate 4-Gbps and 10-Gbps ports. It expands from 24 to 32 ports in 8-port increments and optionally activates 2 10-Gbps ports. The Cisco MDS 9134 offers non-blocking architecture, with all 32 4-Gbps ports and the 2 10-Gbps ports operating at line rate concurrently.

The 10-Gbps ports support a range of optics for connection to the Cisco MDS 9000 family core using 10-Gbps Inter-Switch Link (ISL) connectivity. The Cisco MDS 9134 can also be stacked using copper CX4 X2 transceivers to cost effectively offer up to 64-port densities. The Cisco MDS 9134 supports quick configuration and task wizards that allow it to be deployed quickly and easily in networks of any size. Powered by Cisco MDS 9000 NX-OS/SAN-OS Software, it includes advanced storage networking features and functions and is compatible with Cisco MDS 9500 Series Multilayer Directors and Cisco MDS 9200 Series Multilayer Fabric Switches, providing transparent, end-to-end service delivery in core-edge deployments.

For more information, see: <http://www.cisco.com/en/US/products/hw/ps4159/index.html>.

Cisco MDS 9148

The Cisco MDS 9148 Multilayer Fabric Switch is a one rack unit (1RU) top-of-rack (ToR) chassis that provides 48 line-rate 8-Gbps ports for storage networking deployments. It can expand from 16 to 48 ports in 8-port increments. The Cisco MDS 9148 delivers a non-blocking architecture, with all 48 1/2/4/8-Gbps ports operating at line-rate concurrently.

The Cisco MDS 9148 supports the Cisco Device Manager Quick Configuration Wizard, which allows it to be deployed quickly and easily in networks of any size. Powered by Cisco MDS 9000 NX-OS Software, it includes advanced storage networking features and functions and is compatible with Cisco MDS 9500 Series Multilayer Directors and Cisco MDS 9200 and other 9100 Series Multilayer Fabric Switches, providing transparent, end-to-end service delivery in core-edge deployments.

For more information, see: <http://www.cisco.com/en/US/products/hw/ps4159/index.html>.

Cisco Management Interface

The following Cisco management interfaces were used in the storage layer of the VMDC Compact Pod:

- [Cisco Device Manager, page 1-8](#)
- [Cisco Fabric Manager, page 1-8](#)

Cisco Device Manager

Device Manager is a management solution for Cisco MDS 9000 Family switch chassis. It graphically depicts installed switching modules, the supervisor modules, the status of each port within each module, the power supplies, and the fan assemblies. Device Manager provides two views, Device View and Summary View. Use Summary View to monitor interfaces on the switch. Use Device View to perform the following switch-level configurations:

- Configure zones for multiple VSANs
- Manage ports, port channels, and trunking
- Manage SNMPv3 security access to switches
- Manage CLI security access to the switch
- Manage alarms, events, and notifications
- Save and copy configuration files and software image
- View hardware configuration
- View chassis, module, port status, and statistics

Cisco Fabric Manager

Fabric Manager is a management solution for the MDS family of switches, the Nexus 5000 SAN features, and the UCS Fabric Interconnect with limited support. It provides a robust centralized management station for SAN and unified fabric-enabled devices such as the MDS family of switches and the Nexus 5000. Using Fabric Manager, you can perform the tasks needed during a device's deployment cycle, such as discovery, inventory, configuration, performance monitoring, and troubleshooting.

The tables in the Fabric Manager Information pane correspond to dialog boxes in Device Manager. While Device Manager shows values for a single switch, Fabric Manager shows values for multiple switches. However, for verifying or troubleshooting device-specific configuration, Device Manager provides more detailed information than Fabric Manager.

For more information, see:

http://www.cisco.com/en/US/partner/docs/switches/datacenter/mds9000/sw/5_0/configuration/guides/fund/fm/fmfund_5_0_1.html.

EMC Symmetrix VMAX

EMC Symmetrix VMAX provides high-end storage for the virtual data center. It scales up to 2 petabyte (PB) of usable protected capacity and can be deployed with Flash Drives, Fibre Channel, and Serial Advanced Technology Attachment (SATA) drives, with tiering fully automated with FAST.

For more information, see: <http://www.emc.com/products/detail/hardware/symmetrix-vmax.htm>.

EMC Powerpath VE

With PowerPath/VE, you can standardize path management across heterogeneous physical and virtual environments. PowerPath/VE enables you to automate optimal server, storage, and path utilization in a dynamic virtual environment. This automation eliminates the need to manually load-balance hundreds or thousands of virtual machines and I/O-intensive applications in hyper-consolidated environments.

For more information, see: <http://www.emc.com/products/detail/software/powerpath-ve.htm>.

EMC Timefinder

EMC TimeFinder provides local storage replication for increased application availability and faster data recovery.

For more information, see: <http://www.emc.com/products/detail/software/timefinder.htm>

NetApp FAS6080 Filer

The NetApp FAS6080 provided Enterprise Class Network Attached Storage (NAS) Solution over fully redundant 10 Gigabit and Gigabit Ethernet LANs.

The NetApp FAS6080 Filer system is leveraged 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 datastore clones, array-based thin provisioning, automated policy-based datastore resizing, and direct access to array-based Snapshot copies.

For more information, see <http://www.netapp.com/us/products/storage-systems/fas6000/fas6000.html>.

NetApp Snapshot

Snapshot creates point-in-time copies of file systems, which you can use to protect data-from a single file to a complete disaster recovery solution. It supports up to 255 Snapshot copies per volume to create online backups for user-driven recovery.

For more information, see [h.http://www.netapp.com/us/products/platform-os/snapshot.html](http://www.netapp.com/us/products/platform-os/snapshot.html).

NetApp Deduplication

NetApp Deduplicaton can be leveraged to oversubscribe of real data storage.

For more information, see <http://www.netapp.com/us/products/platform-os/dedupe.html>.

NetApp Rapid Clone Utility

NetApp Rapid Clone is a plug-in for VMware vSphere and supports cloning and provisioning of virtual machines.

For more information, see <http://blogs.netapp.com/virtualization/2010/02/rcu-30-now-available.html>

Service Orchestration

Service orchestration is multi-domain configuration abstraction layer that manages the data center infrastructure. This abstraction layer enables a service catalog/portal-based configuration interface, in which the customer subscribing (application hosting community) to the infrastructure can pick from a limited number of customized service options, and host/place applications as virtual machines. Based upon these picks, configuration actions are executed across multiple domains, and to the device(s) within these domains, that together make up the service as represented within the customer facing portal.

Orchestration (integration across the domain tools) is fundamental as there is no single tool, within the Data Center that can configure the bundled services presented within the service catalog end-to-end. Orchestration coordinates the configuration requirements on top of the domain tools, and insures that all of the services defined within the service catalog/portal, are appropriately sequenced and correctly executed within each specific domain. Moreover, orchestration aggregates all of the individual service components within the service catalog, as a total services pool, and determines if there are sufficient resources across all of the components, to provide the service. The tools required for service orchestration include the following:

- Portal and service catalog (IT service management)
- Configuration management database
- Orchestration (runbook automation)
- Virtualized server provisioning and resource management
- Network provisioning and resource management
- Storage provisioning and resource management

In the the service orchestration layer of VMDC Compact Pod, the following components were used:

- [BMC Atrium Orchestrator, page 1-10](#)
- [BMC Remedy AR System, page 1-10](#)
- [BMC BladeLogic Network Automation, page 1-11](#)
- [BMC BladeLogic Server Automation Suite, page 1-11](#)

BMC Atrium Orchestrator

BMC Atrium Orchestrator automates manual tasks. Workflows based on ITIL standards can be built and adapted to match your processes, with components selected from a library of operator actions and workflow templates.

For more information, see:

<http://www.bmc.com/products/product-listing/90902406-157022-1134.html>.

BMC Remedy AR System

BMC Remedy AR System enables you to automate a broad range of business solutions, from service desk call tracking to inventory management to integrated systems management without learning a programming language or complex development tools. It also acts as a single point of integration, including support for popular API types (such as Java and C), Web Services, ODBC, and utilities such as the BMC Atrium Integration Engine.

For more information see: <http://www.bmc.com/products/product-listing/22735072-106757-2391.html>.

BMC BladeLogic Network Automation

Using BMC BladeLogic Network Automation, you can implement policy-based automation for managing networks, combining configuration management with compliance assurance. Supported by a robust security model, this network automation solution enables organizations to dramatically reduce operational costs, improve operational quality, and achieve operational compliance. BMC BladeLogic Network Automation automates common tasks of device management, including the following:

- Quick, non-disruptive configuration changes
- Proactive assessment of changes and enforcement of configuration standards
- Rapid deployment of devices from predefined templates
- Simplify provisioning of Service Profiles with Network Containers
- Document planned, unplanned, and unauthorized network changes
- On-demand compliance and Key Performance Indicator reporting

For more information, see:

<http://www.bmc.com/products/product-listing/BMC-BladeLogic-Network-Automation.html>.

BMC BladeLogic Server Automation Suite

BMC BladeLogic Server Automation Suite enables customers to manage server and application lifecycle events-including Discovery, Inventory, Provisioning, Configuration, Change Control, and Continual Compliance. The BMC solution addresses three functional areas:

- **Configuration.** Configuration management tasks often make up the bulk of the activities performed in a data center-patching, configuring, updating, and reporting on servers, across multiple platforms. The BMC solution, by shielding users from underlying complexity, enables consistency in change and configuration management activities. At the same time, subject to security constraints, it exposes sufficient detail about servers under management to ensure effective and accurate administrative activities.
- **Compliance.** Most IT organizations are required to maintain their server configurations in compliance with some sort of policy-whether regulatory (such as SOX, PCI, or HIPAA), security, or operational. BMC BladeLogic Server Automation achieves and maintains compliance by defining and applying configuration policies. Then, it provides detailed reports on how well servers comply with these policies. If a server or application configuration deviates from policy, the remediation instructions are generated and packaged, and can be either automatically or manually deployed to the server. All operations performed on servers are constrained by the appropriate set of policies, ensuring that servers stay in compliance throughout configuration changes, software deployments, and patches.
- **Provisioning.** BMC BladeLogic Server Automation automates the OS installation and configuration for both physical and virtual servers-delivering rapid, consistent, and reliable server provisioning processes, and ensuring that all servers are set up in compliance with configuration policies.

For more information, see:

<http://www.bmc.com/products/product-listing/BMC-BladeLogic-Server-Automation-Suite.html>.

Business Continuity

VMware vCenter Site Recovery Manager (SRM) 4.0 provides business continuity and disaster recovery protection for virtual environments. Protection can extend from individual replicated datastores to an entire virtual site.

In a Site Recovery Manager environment, there are two sites involved—a protected site and a recovery site. Protection groups that contain protected virtual machines are configured on the protected site and can be recovered by executing the recovery plans on the recovery site.

Site Recovery Manager leverages array-based replication between a protected site and a recovery site. The workflow that is built into Site Recovery Manager automatically discovers datastores setup for replication between the protected and recovery sites. Site Recovery Manager provides protection for the operating systems and applications encapsulated by virtual machines running on a VMware ESX host. A Site Recovery Manager server must be installed both at the protected and recovery site. The protected and recovery sites must each be managed by their own vCenter Server.

Furthermore, VMware vCenter Site Recovery Manager 4.0 supports VMware vSphere, shared recovery site, and NFS.

The following components were used to provide business continuance for the VMDC Compact Pod:

- [VMware Site Recovery Manager, page 1-12](#)

VMware Site Recovery Manager

VMware vCenter Site Recovery Manager is a business continuity and disaster recovery solution that helps you plan, test, and execute a scheduled migration or emergency failover of vCenter inventory from one site to another. It provides the following features:

Disaster Recovery Management

- Discover and display virtual machines protected by storage replication using integrations certified by storage vendors
- Create and manage recovery plans directly from vCenter Server
- Extend recovery plans with custom scripts
- Monitor availability of the remote site and alert users of possible site failures
- Store, view, and export results of test and failover execution from vCenter Server
- Control access to recovery plans with granular role-based access controls

Non-Disruptive Testing

- Use storage snapshot capabilities to perform recovery tests without losing replicated data
- Connect virtual machines to an existing isolated network for testing purposes
- Automate execution of recovery plans
- Customize execution of recovery plans for testing scenarios
- Automate cleanup of testing environments after completing failover tests

Automated Failover

- Initiate recovery plan execution from vCenter Server with a single button
- Automate promotion of replicated datastores for use in recovery scenarios with adapters created by leading storage vendors for their replication platforms
- Execute user-defined scripts and halts during recovery
- Reconfigure virtual machines' IP addresses to match network configuration at failover site
- Manage and monitor execution of recovery plans within vCenter Server

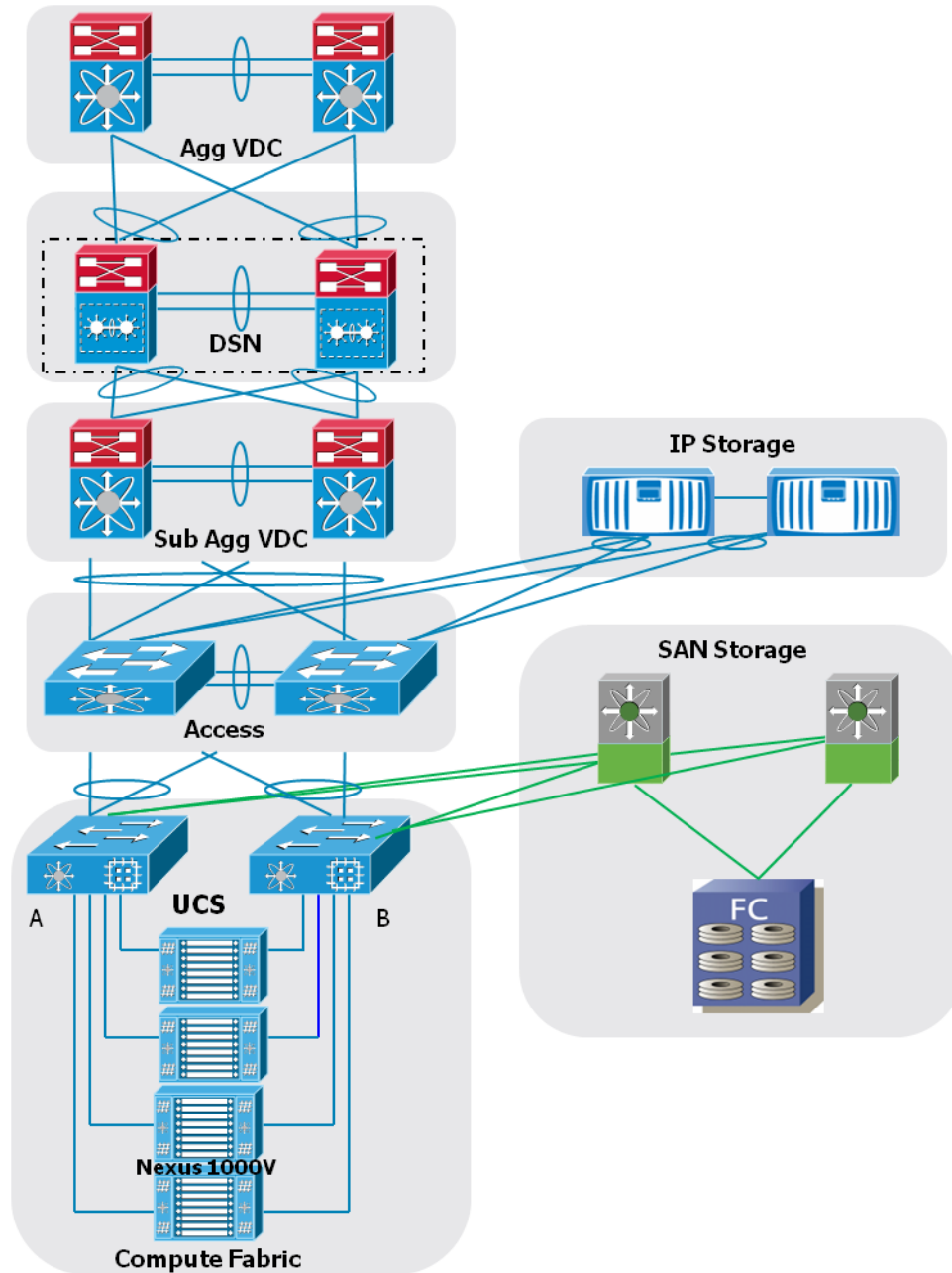
Modular Building Blocks

To scale the data center, Cisco VMDC defines two points of scale: the Point of Delivery and the integrated compute stack. Both repeatable building blocks provide for incremental growth to meet demands. This section defines how these building blocks relate to each other and the data center core and explains how they scale various resources. It contains the following topics:

- [Pod, page 1-13](#)
- [Integrated Compute Stack \(ICS\), page 1-15](#)

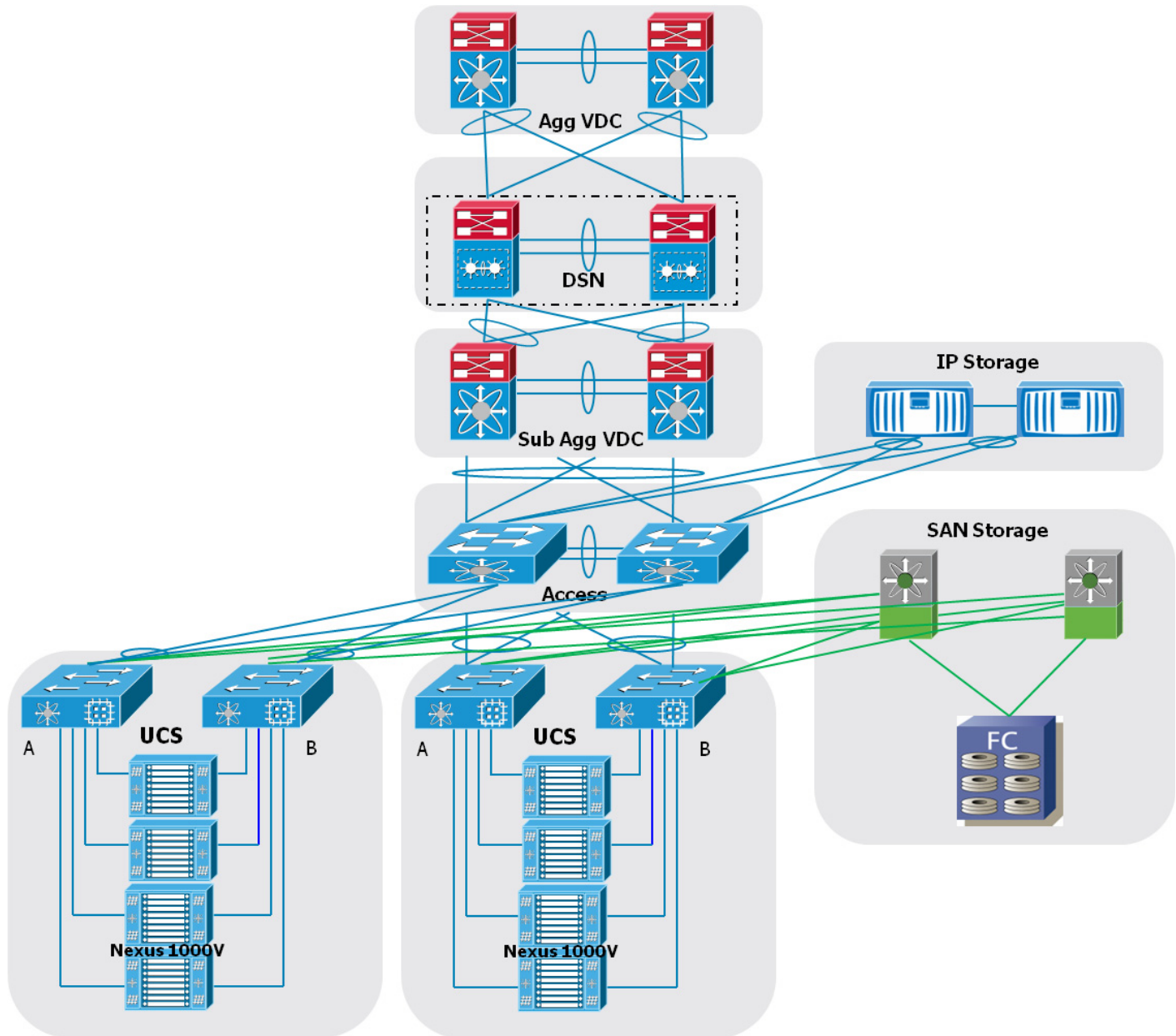
Pod

A pod identifies modular unit of data center components. This modular architecture provides a predictable set of resource characteristics (network, compute, and storage resource pools, power, and space consumption) per unit that is added repeatedly as needed. In this discussion, the aggregation layer switch pair, services layer nodes, and one or more integrated compute stacks are contained within a pod (see [Figure 1-2](#)).

Figure 1-2 Pod Components

To scale a pod, customers add additional integrated compute stacks (see [Figure 1-3](#)). You can continue to scale in this manner until the pod resources are exceeded.

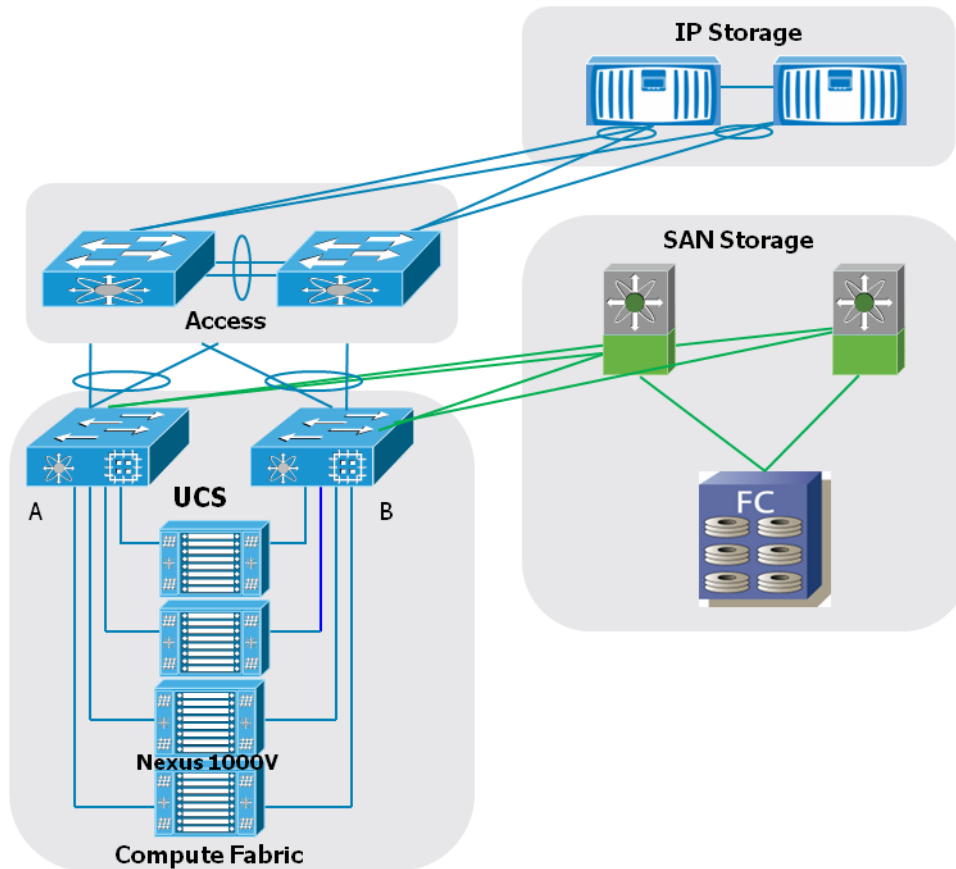
Figure 1-3 Expanding a Pod with Integrated Compute Stacks



Integrated Compute Stack (ICS)

An integrated compute stack can include network, compute, and storage resources in a repeatable unit (see [Figure 1-4](#)). In this discussion, the access layer switch pair, storage, and compute resources are contained within an integrated compute stack.

Figure 1-4 Integrated Compute Stack Components



Multi-Tenant Concepts

Multi-tenancy refers to the logical division of a shared pool of network, compute, and storage resources among multiple groups. Cisco VMDC relies on key concepts to deliver a solution that meets the requirements of these groups. This section explains the specific interpretation of multi-tenancy in the VMDC solution with the following topics:

- [Tenant Defined, page 1-16](#)
- [Differentiated Services, page 1-17](#)

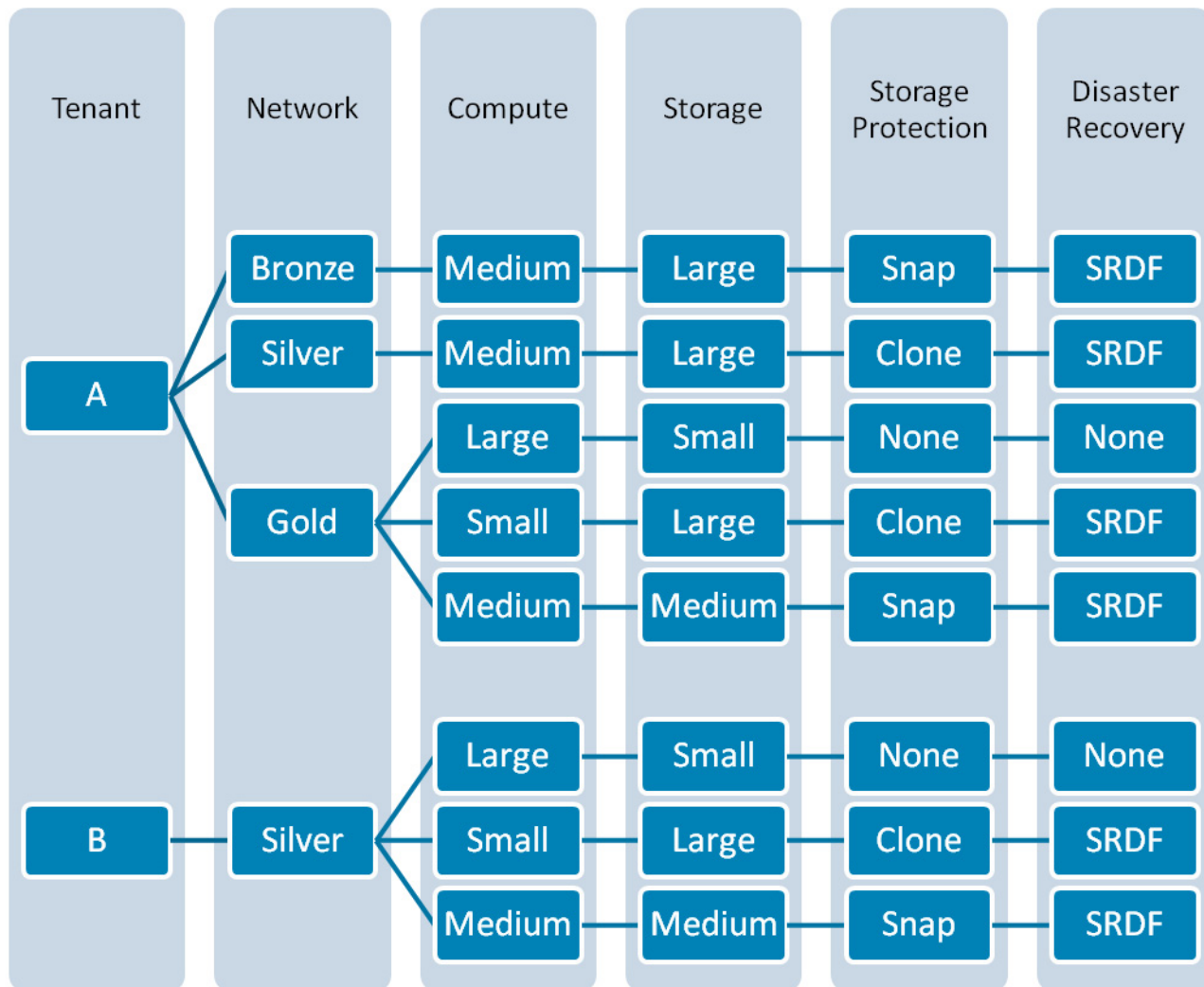
Tenant Defined

In the enterprise private cloud deployment model, the tenant is referenced as a department or business unit, such as engineering or human resources. In the public cloud deployment model, a tenant is an individual consumer, an organization within an enterprise, or an enterprise subscribing to the public cloud services. In either model, each tenant must be securely separated from other tenants because they share the virtualized resource pool.

When a tenant deploys an application or adds a new application, they first select a network container, then a VM size, a storage allocation, a storage protection, and a disaster recovery tier that meets the requirements of the application.

As shown in [Figure 1-5](#), a tenant can select among multiple attributes to define the performance and operation of a virtual server.

Figure 1-5 *Tenants and Virtual Servers*



Differentiated Services

The cloud is the 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 manage the underlying technology. Cloud services are differentiated at three layers in the VMDC solution: network, compute, and storage.

Network Layer

- **Application tiers.** Service tiers can provide differentiated support for application hosting. In some instances, applications may require several application tiers of VMs. For example, a Gold profile could have three application tiers to host web, application, and database (DB) services on different VMs and VLANs. Each tier could provide multiple VMs each for redundancy and provide load balancing. A Silver profile could also have three tiers for web, application, and DB services, but each tier might have multiple VMs on the same VLAN for redundancy and load balancing. A Bronze profile could have three tiers but with the web, application, and DB services residing on the same VM and VLAN.
- **Stateful services.** Customer or employee workloads can also be differentiated by the services applied to each tier. These services can be firewalls, encryption, load balancers, protocol optimization, application firewalls, WAN optimization, advanced routing, redundancy, disaster recovery, and so on. Within a service like firewalls, you can further differentiate among tiers as with inter-VLAN, intra-VLAN, or intra-host inspections. For example, a Gold tier might include firewall inspection, SSL off loading, IPSec encryption, server load balancing, and WAN optimization. A Silver tier might offer only firewall inspection and server load balancing.
- **Quality of Service agreements.** Bandwidth control during periods of network congestion can be key to managing application response time. QoS policies can prioritize bandwidth by service tier. Traffic classification, prioritization, and queuing and scheduling mechanisms can identify and offer minimum bandwidth guarantees to tenant traffic flows during periods of congestion. For example, a Gold service tier might be given the highest priority and a minimum network bandwidth guarantee of 50%. A Bronze service tier might receive best-effort treatment only and no minimum bandwidth guarantees.

Compute Layer

- **Virtual servers.** Typically, cloud providers want to offer multiple service tiers and provide different service level agreements (SLAs). Cloud services can be differentiated into predefined service tiers by varying resource allocation: virtual machine resources. Service profiles can vary based on the size of specific virtual machine (VM) attributes, such as CPU, memory, and storage capacity. Service profiles can also be associated with VMware Distributed Resource Scheduling (DRS) profiles to prioritize specific classes of VMs. For example, a Gold service can consist of VMs with dual core 3-GHz virtual CPU (vCPU), 8 GB of memory, and 500 GB of storage. A Bronze service can consist of VMs with a single core 1.5 GHz vCPU, 2 GB of memory, and 100 GB of storage.

Storage Layer

- **Storage allocation.** Applications require various amounts of disk space to operate. The ability to tune that allocation ensures that applications are not over or under provisioned, which uses resources more intelligently.
- **Storage protection and disaster recovery.** To meet datastore protection, recovery point, or recovery time objectives, service tiers can vary based on provided storage features, such as RAID levels, disk types and speeds, and backup and snapshot capabilities. For example, a Gold service could offer three tiers of RAID-10 storage using 15K rpm Fibre Channel (FC), 10K rpm FC, and SATA drives. While a Bronze service might offer a single RAID-5 storage tier using SATA drives.

The VMDC solution defines options for differentiating levels of resource allocation within the cloud. In this reference architecture, the different levels of services are described as services tiers. Each service tier includes a different set of resources from each of the groups: compute, storage, and network.

Tiered Service Models

The Cisco VMDC architecture allows providers to build service-level agreements (SLAs) that support their tenant or application requirements. The following example is not meant to be a strict definition resource allocation scheme, but to simply demonstrate how differentiated service tiers could be built.

In the Cisco VMDC Compact Pod design we define three service tiers: Gold, Silver, Bronze.

Each service tier is a container that defines different network, compute and storage service levels (see [Table 1-1](#)).

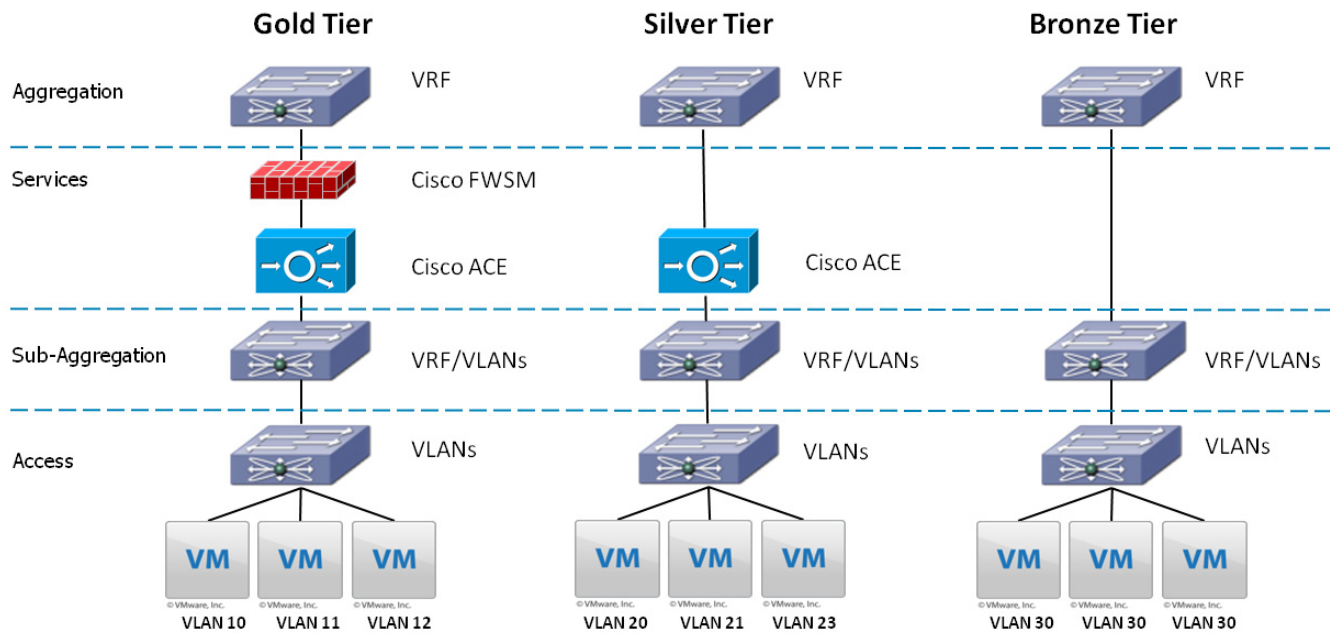
Table 1-1 **Example Network and Data Differentiations by Service Tier**

	Gold	Silver	Bronze
Services	Firewall and Load Balancing Services	Load Balancing Services	No additional services
Bandwidth	40%	30%	20%
Segmentation	Single VRF	Single VRF	Single VRF
VLAN	Multiple VLANs per client	Multiple VLANs per client	Single VLAN per client
Data Protection	Clone - Mirror copy (local site)	Snap - Virtual copy (local site)	none
Disaster Recovery	Remote replication (any-point in-time recovery)	Remote replication (With specific RPO/RTO)	none

The following sections identify how the resources differ among the three tiers.

Network Resources

The Cisco VMDC solution leverages Multi-VRF, VLAN, and virtualized services, such as firewall and load balancing contexts, to extend end-to-end network virtualization into the data center. [Figure 1-6](#) depicts the network components assigned to Bronze, Silver, and Gold service tiers in the VMDC solution.

Figure 1-6 Network Resources by Service Tier

Each tenant leverages a number of service tiers to provide a dedicated virtual network (or virtual private data center). Depending upon the tenant application requirements, multiple application tiers can be created within a Gold or Silver container, allowing each separate tier of the application to reside in a separate VLAN within the VRF.

For each Gold and Silver container, a unique VRF and three VLANs are provisioned. The Gold tenant container includes a dedicated virtual firewall and load-balancing instance. The Silver container includes only a virtual load-balancing instance. The Bronze container is assigned a unique VRF and a single VLAN, and no firewall or load-balancing services are provided.

These service tier definitions form a baseline to which additional services may be added for enhanced security, PCI compliance, data store protection, business continuity, or disaster recovery.

Compute Resources

Server virtualization entails running multiple virtual servers on a single physical blade server. The two primary characteristics to consider are vCPU and RAM.

The number of virtual machines (VMs) that can be enabled depends on the workload type being deployed and the CPU and memory capacity of the blade server. Cisco UCS B-series blade servers are two-socket blades based on the Intel Xeon series processor. Each socket has four cores for a total of 8 cores, or 8 vCPUs, per blade.

At the compute layer, service tier differentiation defines three compute workload sizes called Large, Medium, and Small. As [Table 1-2](#) shows, we enabled 32 Small VMs per blade server by allocating 0.25 vCPU for each virtual machine. A Large VM has a dedicated vCPU limiting the total number of Large workloads to 8 per blade server.

[Table 1-2](#) lists the workload options and compute resource sizes.

Table 1-2 *Compute Resources by Virtual Server Size*

	Virtual Server Options		
	Large	Medium	Small
vCPUs per VM	1 vCPU	0.5 vCPU	0.25 vCPU
Cores per CPU	4	4	4
VM per CPU	4 VM	16 VMs	32 VMs
VM per vCPU Oversubscription Ratio	1:1 (1)	2:1 (0.5)	4:1 (0.25)
RAM allocated per VM	16 GB dedicated	8 GB dedicated	4 GB from shared pool

Storage Resources

The Cisco VMDC architecture defines three static storage allocation sizes called Large, Medium, and Small. These storage arrays are highly available and reliable.

[Table 1-3](#) lists the storage resources sizes.

Table 1-3 *Virtual Server Storage Size Allocation*

	Storage Resource Options		
	Large	Medium	Small
Base storage (GB)	300	150	50
Storage growth increment (GB)	50	50	50

You can further refine the service tiers by differentiating the backup and recovery options. Snap and Clone techniques can create point-in-time consistent copies of tenant volumes. To provide support for disaster recovery, Snap volumes can be replicated to multiple locations. [Table 1-4](#) presents example storage distinctions by service tier.

Table 1-4 *Service Tier Distinctions for Storage Backup and Recovery*

	Gold	Silver	Bronze
Backup (retention length options)	1 mo., 6 mo., or 1yr.	1 mo., 6 mo., or 1yr.	1 mo., 6 mo., or 1yr.
Data protection	Clone – Mirror copy (local site) – SNAP copies every 4 hrs.; 36 hr. retention	Snap – Virtual copy (local site) SNAP copies every 8 hrs.; 36 hr. retention	None
Disaster recovery	Remote replication SRDF	Remote replication Symmetrix Remote Data Facility (SRDF)	None

Compact Pod Network Topology

Data center networking technology is currently an area of rapid change. Higher-performance end nodes and the migration to 10-Gigabit Ethernet for edge connectivity are changing design standards, while virtualization capabilities are expanding the tools available to the network architect. When designing the

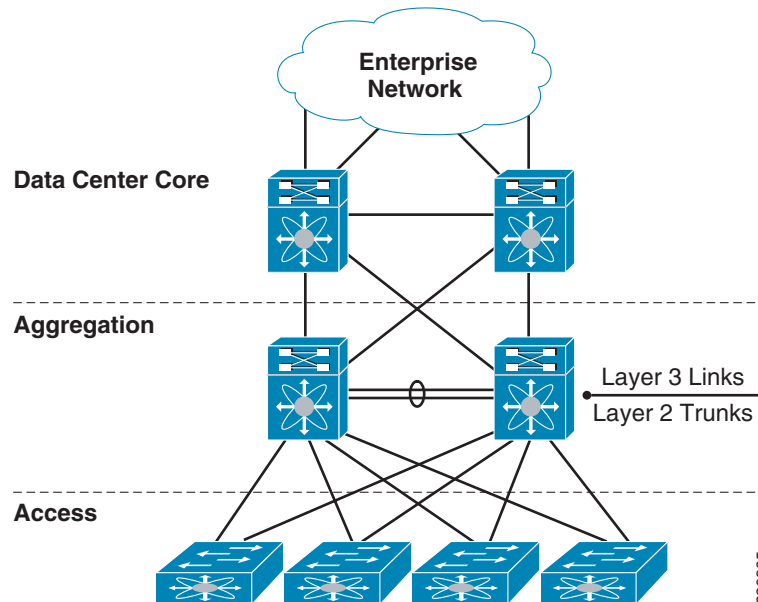
data center network, the experienced architect relies on a solid hierarchical foundation for high availability and continued scalability. This foundation also provides the flexibility to create different logical topologies utilizing device virtualization, the insertion of service devices, as well as traditional Layer-3 and Layer-2 network configurations. The following section describes the hierarchical network design reference model as applied to meet the requirements and constraints commonly found in today's data centers. As a reference model, this topology is flexible and extensible, and may need to be extended or modified to meet the requirements of a specific enterprise data center network.

Hierarchical Network Design Reference Model

Hierarchical network design has been commonly used in 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 one of the switching nodes in the network or as standalone appliance devices. Each service approach also 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 correlates tightly to the hierarchy of the network topology. Decisions on cabling types and the placement of patch panels and physical aggregation points must match the interface types and densities of the physical switches being deployed. In a new data center build-out, the two can be designed simultaneously, also taking into consideration the constraints of power and cooling resources. When seeking to avoid significant new investment within an existing data center facility, an architect must consider the pre-existing physical environment of cabling, power, and cooling when selecting switching platforms. Careful planning in conjunction with networking requirements and an eye toward flexibility for the future is critical when designing the physical data center environment. Taking a modular approach to data center design provides flexibility and scalability in both network topology design and utilization of physical resources.

Figure 1-7 illustrates the primary network switching layers of the hierarchical network design reference model for the data center environment. The overall hierarchical model is similar to the reference topology for enterprise campus design, but the term *aggregation layer* replaces the term *distribution layer*. The data center network is less concerned with distributing network access across multiple geographically disparate wiring closets and is focused aggregating server resources and providing an insertion point for shared data center services.

Figure 1-7 Hierarchical Network Design Reference Model

The reference model in Figure 1-7 shows the boundary between Layer-3 routed networking and Layer-2 Ethernet broadcast domains at the aggregation layer. Larger Layer-2 domains increase the physical flexibility of the data center—providing the capability to manually or virtually relocate a server to a different physical rack location with less chance of requiring a change of IP addressing to map to a specific subnet. This physical flexibility comes with a tradeoff. Segregating the network into smaller broadcast domains results in smaller spanning tree domains and failure domains, which improve network stability, reduce convergence times and simplify troubleshooting. When determining how to scale Layer-2 domains, the network architect must consider many factors including the access switching model in the use and nature of the underlying applications being serviced. Cisco has introduced features such as bridge assurance and dispute mechanism into switching products to allow greater scalability of Layer-2 domains with increased stability of the STP.

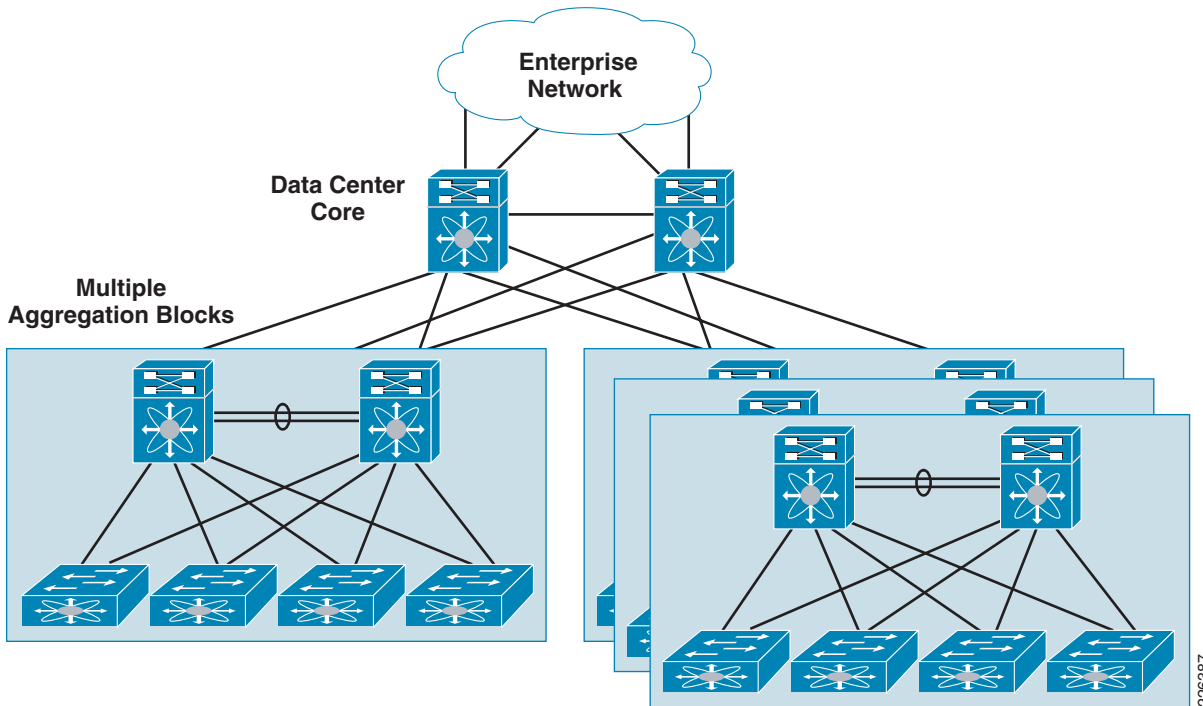
Core Layer

The hierarchical network design model gains much of its stability and high availability characteristics by splitting out switching nodes based on their function, and 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 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 among the other functional blocks of the network, such as campus, Internet edge and WAN. 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 STP for Layer-2 network loop prevention.

In small-to-medium enterprise environments, it is reasonable to connect a single data center aggregation block, or pod, directly to the enterprise switching core for Layer-3 transport to the rest of the enterprise network. Provisioning a separate, dedicated pair of data center core switches provides additional insulation from the rest of the enterprise network for routing stability and also provides a point of 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 for scale expansion without requiring additional Layer-3 interfaces to be available on the enterprise core. An illustration of scaling the data center topology with a dedicated core and multiple aggregation blocks is provided in Figure 1-8.

Figure 1-8 **Scaling the Data Center with a Dedicated Core**



Cisco's premier switching platform for the data center core is the Nexus 7000 Series switch. The Nexus 7000 Series has been designed from the ground up to support the stringent uptime requirements of the data center. The Nexus 7000 Series switches are optimized for support of high density 10-Gigabit Ethernet, providing scalability in the 18-slot chassis up to 128 wire rate 10-Gigabit Ethernet interfaces when ports are configured in a dedicated mode using the N7K-M132XP-12 I/O Module. The Nexus 7000 Series hardware is coupled with Cisco NX-OS, a modular operating system also designed specifically for the requirements of today's data center networks. NX-OS is built on the industry-proven SAN-OS software-adding virtualization, Layer-2, and Layer-3 features and protocols required in the data center environment. NX-OS includes high availability features, such as granular process modularity, In-Service Software Upgrade (ISSU), and stateful process restart, that are specifically targeted at the service-level requirements of the enterprise or service provider data center.

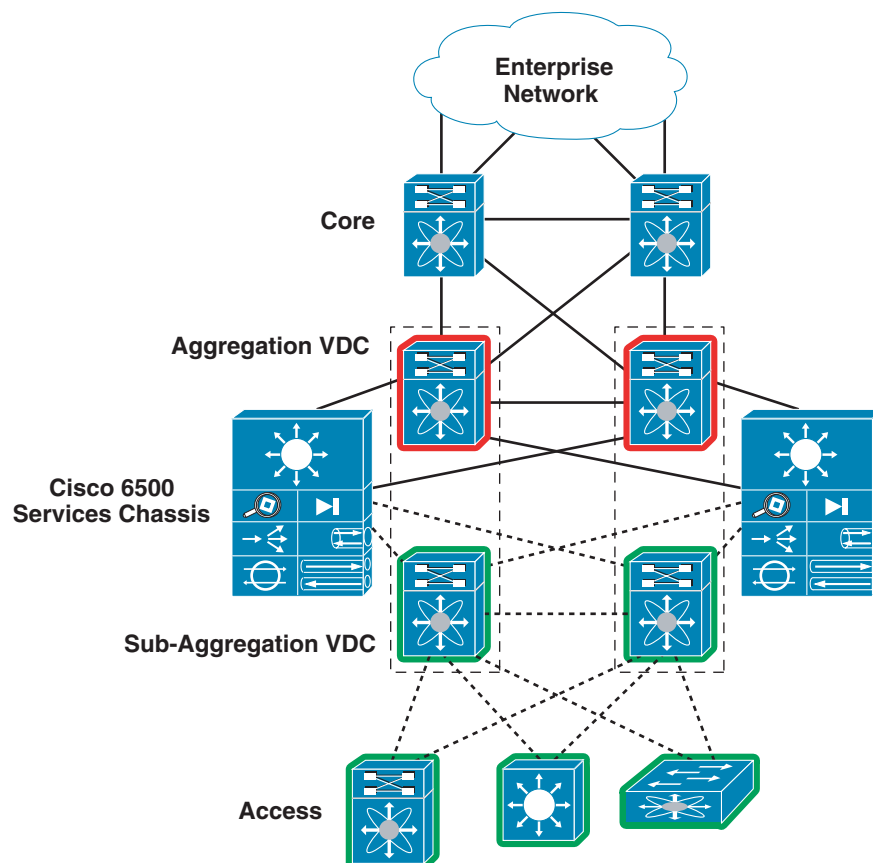
When choosing switching platforms to provision layers of the data center network topology, the network architect must be aware of specific features and interface types required by the network design. The Nexus 7000 Series offers unique virtualization features such as Virtual Device Contexts (VDCs) and Virtual Port Channels (vPCs). The Nexus 7000 Series switches also have excellent high availability features, throughput, and 10-Gigabit Ethernet port densities; however, NX-OS does not support some of the features found in Cisco IOS-based switching platforms. Another Cisco switching platform commonly found in the core of today's data centers is the Cisco Catalyst 6500. The Catalyst 6500 offers software features such as support of Multi Protocol Label Switching (MPLS), VLAN Mapping, and Q-in-Q multiple-level VLAN tagging that may be required in specific designs. The Cisco Catalyst 6500 also offers greater diversity of physical interface types and support for services modules directly installed within the chassis.

Services VDC Sandwich Model

Data center service insertion requirements may include server load-balancing devices, security devices such as firewall and intrusion prevention, and others. Multiple approaches exist for the integration of these services into the data flow. Design decisions include using modules in external Services Chassis, using appliances, and whether to run the service devices in a transparent or routed mode. One very flexible design approach is to use all services in transparent mode, but to insert an additional layer of routing instances between the server farm subnets and the services devices. This approach has been shown in design guidance using VRFs, and the deployment of multiple VRFs also provides the capability to direct traffic independently through multiple virtual contexts on the service devices, leveraging the virtualization of both the routing functions and the services devices in the design.

The VDC capability of the Nexus 7000 Series enables the network architect to leverage another type of virtualization in the design, to improve ease of configuration, supportability, and security. A secondary virtual switching layer called the *sub-aggregation* can be created using VDCs, located between the services devices and the access switches. This topology is referred to as a services VDC sandwich and is the topology used in the VMDC 2.0 Compact Pod architecture. An example of this topology using services modules located in external Catalyst 6500 chassis is shown in Figure 1-9.

Figure 1-9 Services Sandwiched Between VDCs



All the access layer switches shown in [Figure 1-9](#) attach only to the sub-aggregation VDCs. Different classes of servers could also be attached 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 considerations when designing this type of topology include the following:

- Similar designs have been deployed only using 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 using 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 increasing 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 being only separate VLANs and VRFs on the same switch, they are separate virtual switches with completely different sets of processes and physical ports.
- Additional interfaces may be required for the VDC sandwich topology as compared with a VRF sandwich topology. The Services Chassis must have separate physical connections into both sets of VDCs as opposed to VLANs sharing the same trunks. Additional interface count must also be provisioned to support the inter-switch link between the two sub-aggregation VDCs.
- This model has been validated by Cisco using Firewall Services Module (FWSM) running in transparent mode and Application Control Engine (ACE) modules running in routed 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 using static routes pointing 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 IP multicast is a requirement, then an IGP such as OSPF or EIGRP may be used.
- VDCs provide the distinction between the routing instances of the aggregation and the sub-aggregation layers; however, the use of multiple VRFs in the sub-aggregation layer may be utilized to support additional virtualization capabilities. Distinct VRFs in the sub-aggregation layer may be mapped using VLANs to separate contexts within the virtualized service devices such as the FWSM and ACE, allowing active contexts to be split between both Services Chassis. If services are required between layers of a multi-tier application architecture, placing these tiers in subnets belonging to separate VRFs will allow for powerful, multi-context service insertion between tiers.
- A services VDC sandwich using 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, the services on the left side have dual connectivity and can maintain a primary role. Service appliances run in transparent mode, such as the Adaptive Security Appliance (ASA) 5580, that only support single connections to carry a given VLAN. In transparent mode, such an appliance will not be dual-homed if attached directly to the aggregation, but it can still be deployed in a highly available manner by using redundant appliances.


Note

For more detail on the VDC services sandwich architecture with virtualized services, refer to the *Data Center Service Patterns* document at the following URL:

http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DC_3_0/dc_serv_pat.html

Aggregation VDC

The primary function of the aggregation VDC is to provide highly available, high performance Layer-3 switching IP traffic from the services layer to the other functional blocks of the network, such as data center core (if deployed). In smaller enterprise environments, it is reasonable to collapse the core functionality into the aggregation VDC and provide connections to the campus, Internet edge, and WAN at this layer.

Services Layer - Datacenter Services Node (DSN)

The Cisco VMDC 2.0 DSN design case virtual switching system (VSS) and Cisco FWSM and Cisco ACE virtualization.

Virtual Switching System (VSS)

VSS combines two physical Cisco Catalyst 6500 Series Switches into one virtualized switch. This arrangement enables a unified control plane and also allows both data planes to forward simultaneously. With VSS, multi-chassis EtherChannel (MEC) is introduced, which allows a port channel to be formed across two physical switches. vPC and VSS both provide enhanced system availability through redundant systems, eliminate reliance on Spanning Tree Protocol, achieve faster convergence times, and enable full system availability at all times. For the Cisco DSN use cases, the sub-aggregation layer switches can run in vPC mode and interconnect to the Cisco DSN through MEC, which will be running in the VSS. An additional benefit of integrating VSS with Cisco DSN is that this integration increases the number of supported service modules per chassis from four to eight in a single VSS domain, enabling an active-active highly available service chassis.

Transparent FWSM

A transparent firewall requires less configuration than a routed firewall, since there is no routing protocol to configure or list of static routes to maintain. It requires only a single IP subnet on the bridge-group interface, and forwards BPDUs between bridging devices that live on attached segments. In that way, it is truly transparent, and not a bridge itself. The VLANs on the different interfaces of the transparent FWSM carry different VLAN numbers, so a transparent device is often said to be “stitching” or “chaining” VLANs together.

Routed ACE

The active/standby Services Chassis design leverages the ACE context in a routed server load-balancing mode. The ACE has a single logical interface (VLAN), which exists between itself and the Aggregation VDC. The ACE has a second interface which resides on the Access Layer VLAN where the real servers sit. The ACE routes traffic destined to Enterprise Cloud to an HSRP address on the Aggregation Layer VDC.

The advantage of this deployment model is that the ACE virtual context is exposed to only those flows requiring its services. Non-load-balanced flows traverse in and out of the server farm without being processed by the ACE; while load-balanced flows benefit from dedicated ACE services positioned for optimal performance.

To force traffic back to the ACE context, it is necessary to configure Source NAT on the ACE. Source NAT on the ACE is simple to deploy and enforces symmetric traffic flow by using IP address pools dedicated to the ACE virtual context and basic Layer 2 switching. From a traffic flow perspective, source NAT readily fits into many existing data center designs by introducing the load balancer as a new separate Layer 2 adjacency in the network. To accommodate application-logging requirements, the network administrator may use the ACE's HTTP header manipulation feature by inserting the original source IP address of an HTTP traffic flow into the HTTP header. The disadvantage to this technique is the loss of source IP logging on the server for non-HTTP applications.

Sub-Aggregation VDC

The sub-aggregation layer of the data center provides a consolidation point where access layer switches are connected providing connectivity between servers for multi-tier applications, as well as connectivity across the services, aggregation, and core layers of the data center network to clients residing within the campus, WAN, or Internet. The sub-aggregation layer typically provides the boundary between Layer-3 routed links and Layer-2 Ethernet broadcast domains in the data center. The access switches are connected to the sub-aggregation layer using 802.1Q VLAN trunks to provide the capability of connecting servers belonging to different VLANs and IP subnets to the same physical access switch.

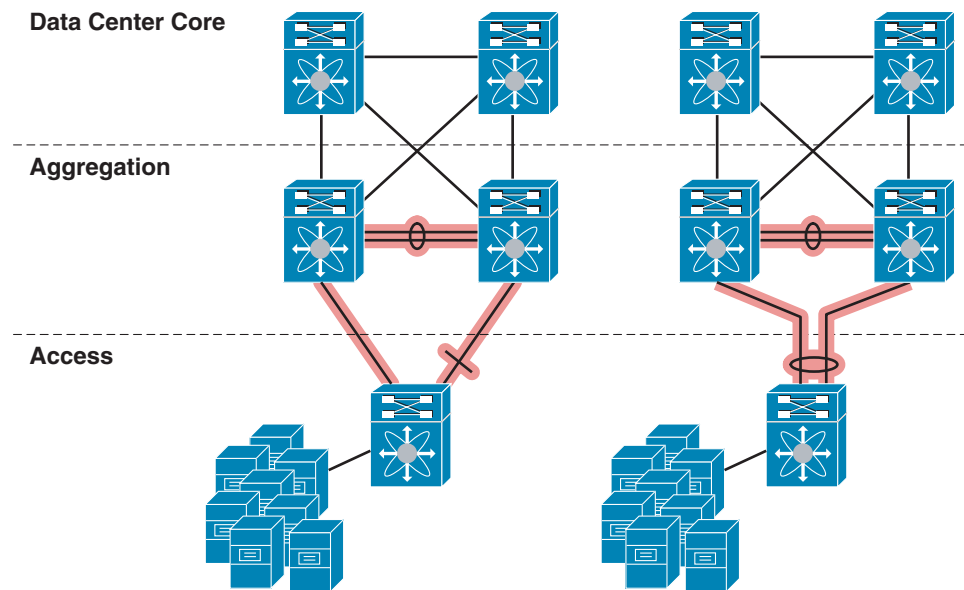
Traditional models of access-layer connectivity include links from each access-layer switch into both switches forming the sub-aggregation-layer redundant switch pair. This approach provides network resiliency in the event of a single link or interface failover or failure of one of the sub-aggregation switches. The inter-switch link between the two sub-aggregation switches is also an 802.1Q trunk that carries all VLANs in use in the server farm. The STP is active independently for each VLAN instance using the Rapid Per VLAN Spanning Tree Plus (RPVST+) model, which blocks redundant ports when they are not needed to avoid network loops. Features such as Virtual Port Channels (vPC) on the Cisco Nexus 7000 Series and Virtual Switching System (VSS) on the Catalyst 6500 series have been introduced to allow both switches in the aggregation pair to act as a single switching unit from a STP and port channel perspective. This approach allows all links between an access switch and the sub-aggregation layer to be active as a single port channel instead of having STP blocking a redundant path.

Loop Free Layer 2 Design

A data center is generally made of similar simple building blocks, replicated at will to achieve the desired level of scalability. The solution provides for redundancy, which means that devices in a particular position are at least duplicated. The traditional network design is a typical example of that: a pair of aggregation switches to which as many access switches as necessary are connected in a redundant way. The two main drawbacks of this solution are as follows:

- There is no Layer-2 multipathing for a particular VLAN, and the per-VLAN load balancing that allows using both uplinks of an access switch needs user configuration. There is no way of escaping this constraint as it dictated by the way bridging requires a spanning tree in the data plane.
- The dependency on STP for the creation of a spanning tree topology in the data plane, introducing delay in the convergence and potential risks.

Port channel technology is solving those remaining issues for the very specific case of the interconnection of two switches (see [Port Channels/EtherChannels](#)). Alone, link aggregation cannot be used to create a fully redundant data center, as it does not protect against the failure of a single switch. Cisco has recently introduced two technologies that lift this latter limitation. Both VSS (on the Catalyst 6000) and vPC (on the Nexus 7000) allow creating a Layer-2 port channel interface distributed across two different physical switches. This limited step-up in the channeling capability is enough to provide the simple building block required to build an entire data center with no dependency on the spanning tree model. [Figure 1-10](#) shows a high level use of the solution.

Figure 1-10 Loop-Free Network

The left part of [Figure 1-10](#) illustrates the current model, where the redundancy is handled by STP. The right part of [Figure 1-10](#) represents the solution introduced by distributing the end of a channel across the two aggregation switches.

The logical view shows that the redundancy has been hidden from STP. As far as the “[Rules for STP Network Stability](#)” are concerned, the right side of [Figure 1-10](#) shows the best solution, where the following are depicted:

- The number of blocked ports has been eliminated
- The freedom of STP has been also entirely removed, as it cannot open a loop even if it wanted to.

However, the recommendation is to keep STP on as a backup mechanism. Even if the redundancy has been hidden to STP, it is still there, at a lower layer. It is just handled by a different mechanism. STP helps protect against a configuration error that breaks a channel into individual links, for example.

Access Layer

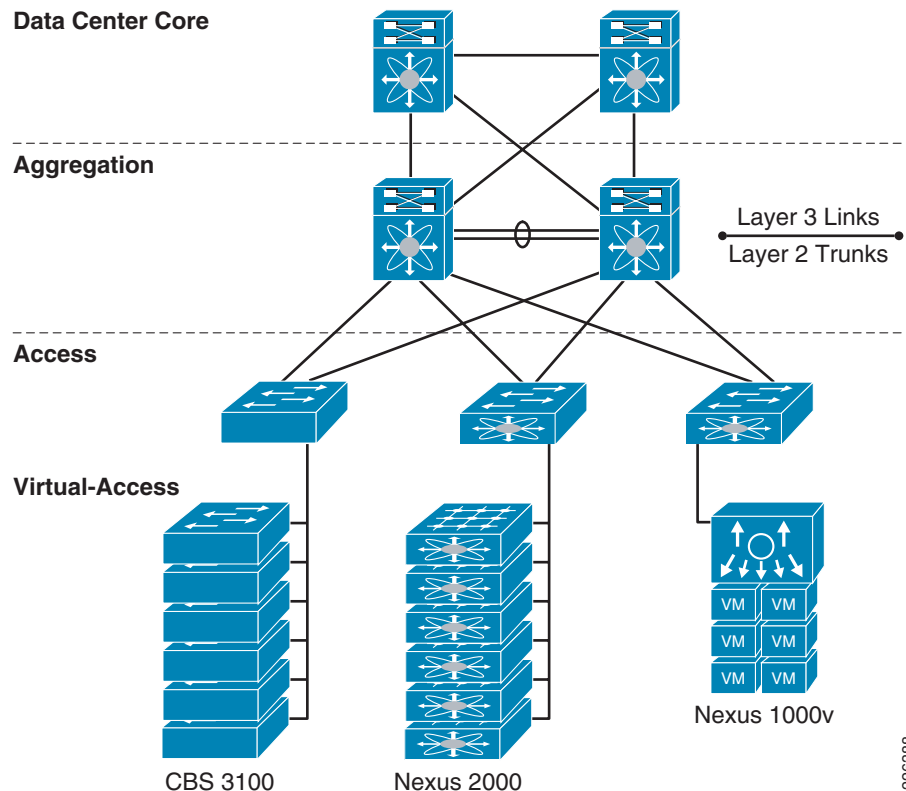
Virtual Access Evolution

The evolution of networking technology in the data center is most evident at the access layer of the network and within the server farm. Several options for building the data center access layer introduce switch virtualization that allows the function of the logical Layer-2 access layer to span multiple physical devices. For example:

- Cisco Nexus 5000 Series switches work in conjunction with the Cisco Nexus 2000 Series Fabric Extenders to act as a single virtual access switch while providing ToR connectivity for servers in multiple racks.
- The software-based switching implementation in the Cisco Nexus 1000V Virtual Distributed Switch also provides virtual access layer switching capabilities designed to operate in server virtualization environments.

Figure 1-11 illustrates these examples of access-layer virtualization in the data center network. The virtual-access sublayer does not represent an additional level of Layer-2 switching; it conceptually exists as virtual I/O modules or line cards extended from a centralized management and control plane. This approach offers many of the benefits of EoR switching, such as reduced aggregation switch port density requirements and fewer points of management, while providing cable-management benefits similar to a ToR model.

Figure 1-11 Data Center Virtual-Access Evolution



The Cisco Nexus 5000 Series switches provide high-density 10-Gigabit Ethernet connectivity and innovative storage integration capabilities for the support of FCoE. With a Layer-2 capable implementation of NX-OS, the Nexus 5000 is optimized for the evolving data center access layer. For customers requiring a density of 1-Gigabit Ethernet server connectivity, the Nexus 2000 Fabric Extenders may be deployed in conjunction with a Nexus 5000 Series switch and treated as a single virtual chassis in the access layer of the data center topology. This approach may be used to provide ToR switching to multiple racks of servers, with all management functions for the Nexus 2000 Fabric Extenders centralized into their associated Nexus 5000 Series switch. The Nexus 5000 Series can also be placed middle-of-row (MoR) to provide 10-Gigabit Ethernet interfaces to nearby servers.

Implementations of hypervisor-based server virtualization systems include software-based logical switching capabilities within the server. The Nexus 1000V virtual distributed switch allows the network architect to provide a consistent networking feature set across both physical servers and virtualized servers. The Nexus 1000V operates as a virtualized chassis switch, with Virtual Ethernet Modules (VEMs) resident on the individual virtualized servers managed by a central Virtual Supervisor Module (VSM) that controls the multiple VEMs as one logical modular switch. The VSM provides a centralized point of configuration and policy management for the entire virtual distributed switch. Both the Cisco Nexus 2000 Fabric Extenders and the Cisco Nexus 1000V represent variations on the evolving capabilities of the data center virtual-access sub-layer.

Storage Integration

Another important factor changing the landscape of the data center access layer is the convergence of storage and IP data traffic onto a common physical infrastructure, referred to as a unified fabric. The unified fabric architecture offers cost savings in multiple areas including server adapters, rack space, power, cooling, and cabling. The Cisco Nexus family of switches, particularly the Nexus 5000 Series is spearheading this convergence of storage and data traffic through support of Fibre Channel over Ethernet (FCoE) switching in conjunction with high-density 10-Gigabit Ethernet interfaces. Server nodes may be deployed with converged network adapters (CNAs) supporting both IP data and FCoE storage traffic, allowing the server to use a single set of cabling and a common network interface. The Cisco Nexus 5000 Series also offers native Fibre Channel interfaces to allow these CNA attached servers to communicate with traditional Storage Area Network (SAN) equipment.

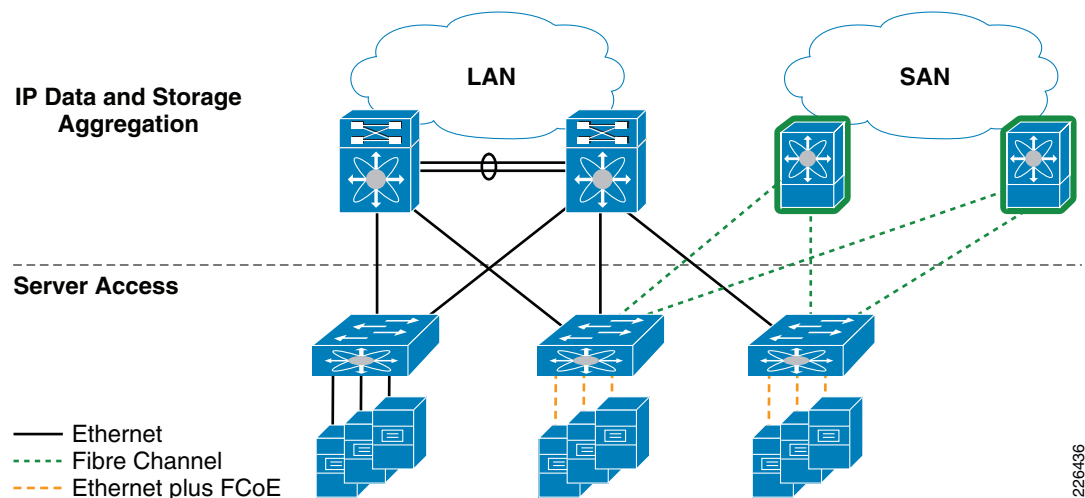
At its initial product release, the Cisco Nexus 5000 supports a unified fabric switching approach only at the edge of the data center topology. Over time, the Cisco Nexus family will allow further consolidation of FCoE-based storage traffic into the aggregation layer of the data center. Choosing Cisco Nexus switching platforms for new data center investment today positions the network architect to take advantage of additional I/O consolidation features as they are released across the product family. [Figure 1-12](#) illustrates a topology with CNA-attached servers running both FCoE traffic and IP data traffic over a common interface to a Nexus 5000 switch. The Nexus 5000 splits out the FCoE traffic and provides native Fibre Channel interface connections back into Fibre Channel switches to connect to the shared SAN.



Note

In the VMDC design, the UCS 6120 was used to extend the FCoE access layer so that FibreChannel traffic no longer has to pass through the Nexus 5000 to reach the SAN fabric.

Figure 1-12 Access Layer Storage Convergence with Nexus 5000



SAN Design Reference Model

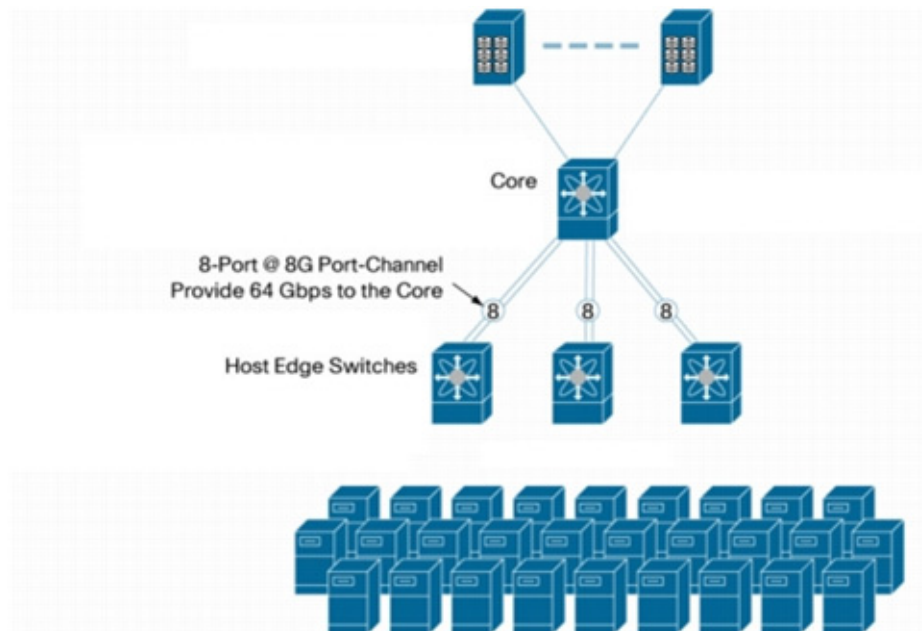
Core/Edge Design Model

It is common practice in SAN environments to build two separate, redundant physical fabrics (Fabric A and Fabric B) in case a single physical fabric fails. Large SAN architectures are classified as one of the following topologies in a physical fabric:

- Two-tier: Core-edge design
- Three-tier: Edge-core-edge design

Within the two-tier design, servers connect to the edge switches, and storage devices connect to one or more core switches (Figure 1-13). This allows the core switch to provide storage services to one or more edge switches, thus servicing more servers in the fabric. The interswitch links (ISLs) will have to be designed so that the overall fabric maintains both the fan-out ratio of servers to storage and the overall end-to-end oversubscription ratio.

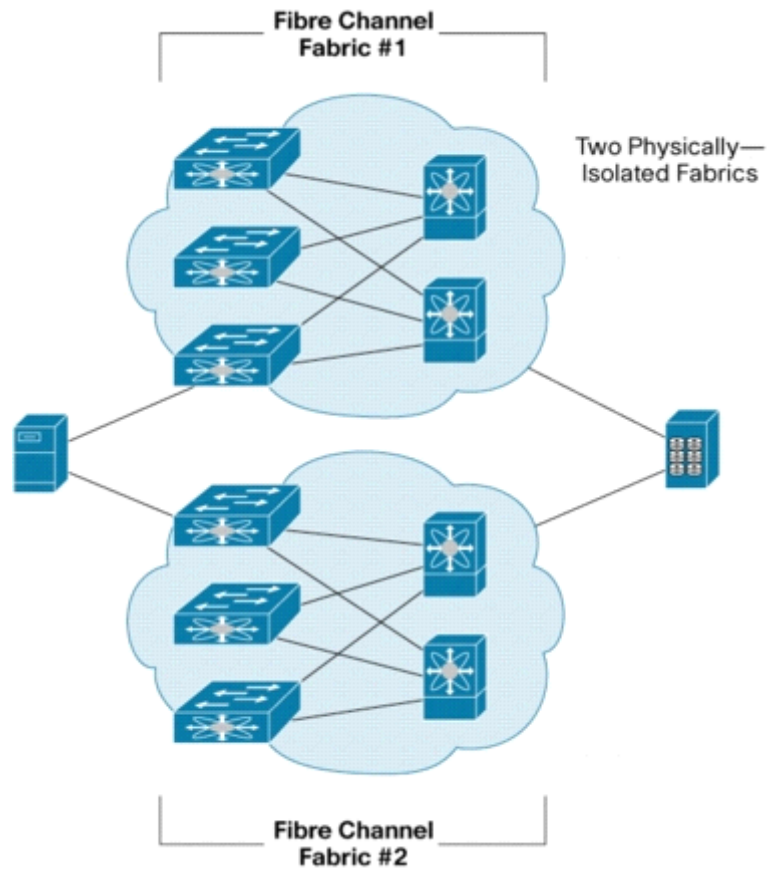
Figure 1-13 Sample Core-Edge Design



Fabric Redundancy

Another area that requires attention in a Fibre Channel SAN is the fabric itself. Each device connected to the same physical infrastructure is in the same Fibre Channel fabric. This opens up the SAN to fabric-level events that could disrupt all devices on the network. Changes such as adding switches or changing zoning configurations could ripple through the entire connected fabric. Therefore, designing with separate connected fabrics helps to isolate the scope of any such events. The Cisco Systems Virtual SAN (VSAN) capability offers a way to replicate this environment, namely, the isolation of events, using the same physical infrastructure. (See Figure 1-14)

Figure 1-14 *Designing SANs with Isolated Fabrics*



Interswitch links (ISLs)

The connectivity between switches is important as the SAN grows. Relying on a single physical link between switches reduces overall redundancy in the design. Redundant ISLs provide failover capacity if a link fails.

