CHAPTER 1

Design Overview

This chapter provides an overview of the Virtualized Multiservice Data Center (VMDC) solution and contains the following topics:

Introduction

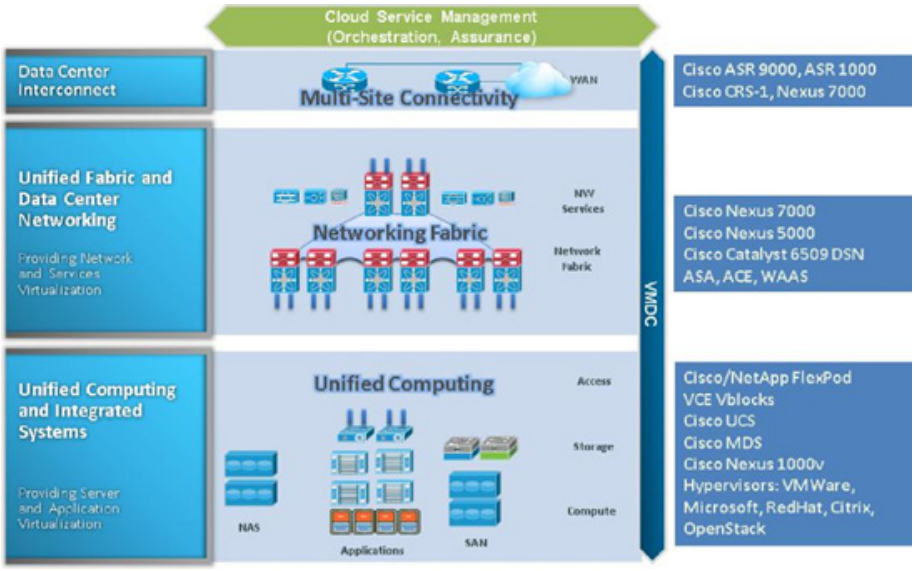
The cloud provides highly scalable, efficient, and elastic services accessed on-demand over the Internet or intranet. In the cloud, compute, storage, and network hardware are abstracted and delivered as a service. End users enjoy the functionality and value provided by the service without the need to manage or be aware of the underlying technology. A cloud deployment model differs from traditional deployments in its ability to treat the Data Center (DC) as a common fabric of resources. A portion of these resources can be dynamically allocated and deallocated when they are no longer in use.

The VMDC solution is the Cisco reference architecture for Infrastructure as a Service (IaaS) cloud deployments. This Cisco cloud architecture is designed around a set of modular DC components consisting of building blocks of resources called pods. These pods are comprised of shared resource pools of network, storage, and compute. Each of these components is virtualized and used by multiple tenants securely, so that each cloud tenant appears to have its own set of physical resources. Cloud service orchestration tools automate the resource provisioning workflow within the cloud DC.

The VMDC solution is targeted towards Enterprises building private clouds and Service Providers building public clouds. In the public cloud case, the tenant would typically be located remotely and have their own DC resources on site in addition to resources within the cloud. In the private case, the tenant could reside locally in another organizational unit logically separated from the IT DC or be located at another facility.

The VMDC system is built around the Cisco Unified Compute System (UCS), Nexus 1000V virtual switches, Multilayer Director Switch (MDS) storage switches, SAN and NAS storage arrays such as NetApp FAS arrays offering NetApp Unified Storage Architecture, Nexus 7000 Aggregation (switching and routing) and Nexus 5000 Access (switching) layers connecting into the Catalyst 6500 Data Center Service Node (DSN), Adaptive Security Appliance (ASA), Application Control Engine (ACE)-based Layer 4 (L4) - Layer 7 (L7) Services layer, and the ASR 9000 and 1000 WAN routers. Cloud orchestration is provided by the BMC Cloud Lifecycle Management (CLM) suite, and cloud assurance by the Zenoss Cloud Service Assurance (CSA) suite. Figure 1-1 shows the functional components of the VMDC solution:

Figure 1-1 VMDC System Overview



Note Data Center Interconnect (DCI), Cloud Orchestration, and Cloud Assurance are not covered in this document. Please refer to the following documentation for those aspects: Cisco VMDC Documentation on Cisco.com Design Zone.

There have been several iterations of the VMDC solution, with each phase encompassing new platforms, versions, and technologies. The previously released VMDC 2.2 and VMDC 3.0 solutions utilize end-to-end VRF-Lite for tenant segmentation within the cloud DC. The VMDC 2.2 solution utilizes a Virtual Port-channel (vPC)-based Layer 2 (L2) design in the Nexus DC fabric, whereas the VMDC 3.0 solution utilizes a FabricPath-based L2 design in the Nexus DC fabric.

Note For more information about previous versions of the VMDC solution, refer to the following documentation:
Cisco VMDC 2.2 Design Guide
Cisco VMDC 3.0 Design Guide

This document focuses on design considerations specific to aspects of the VMDC 2.3-based DC. The VMDC 2.3 solution forms the basis for the Service Provider Cloud Smart Solutions Standard-offer for the Cloud Ready Infrastructure (CRI). The VMDC 2.3 architecture is based on the prior VMDC 2.2 architecture, with some design changes. The key changes in the VMDC 2.3 solution, as compared to the VMDC 2.2 solution, are listed below.

- VMDC 2.2 is built for high server and VM density with up to 3000 servers and 70,000 VMs across six pods.
- VMDC 2.3 is built for small-to-medium server and VM density with up to 768 servers and 24,000 VMs across four pods.
- VMDC 2.3 design has been optimized from the VMDC 2.2 design to achieve higher tenant density with up to 2000 tenants across four pods.
- The Cisco ASR 1000 replaces the Cisco ASR 9000 as the DC WAN router (MPLS-PE).
- VMDC 2.3 does not use the DC Core Nexus 7000 layer.

- Instead of the Nexus 7018 with M1/M2 line cards, VMDC 2.3 uses the Nexus 7004 with F2 line cards as the Aggregation layer.
- VMDC 2.3 does not use the Catalyst 6500 DSN. Instead, services are provided by ASA and ACE appliances directly connecting to the Nexus 7000 Aggregation layer.
- VMDC 2.3 includes optimized tenancy models for Expanded Gold, Silver, and Bronze containers.
- VMDC 2.3 includes a new Copper container for Internet-based cloud access.

Cloud Data Center

The VMDC-based cloud DC consists of network, storage, and compute resources. The data centers are typically interconnected and provide access to the WAN, IP/Next Generation Network (NGN), or the public Internet. The DC provides multitenancy and multiservices, and also includes management elements for administrative functions, orchestration (cloud portals, Service Catalog, workflow automation), and assurance.

This section discusses the following aspects of the cloud DC:

Hierarchical Network Architecture

Typical DC designs are based upon the classic, multilayer hierarchical network model. In general, such a model implements three layers of hierarchy:

1. A DC Core layer, characterized by a high degree of redundancy and bandwidth capacity, and thus, optimized for availability and performance. The Core layer connects to the DC WAN Edge router.
2. A DC Aggregation layer, characterized by a high degree of high-bandwidth port density capacity, and thus, optimized for traffic distribution and link fan-out capabilities to Access layer switches. Functionally, the nodes in the Aggregation layer typically serve as the L2/L3 boundary. Multiple Aggregation layers can connect to the Core layer, providing for increased density and east-west traffic within the DC.
3. A DC Access layer, serving to connect hosts to the infrastructure, and thus, providing network access, typically at L2 (VLANs).

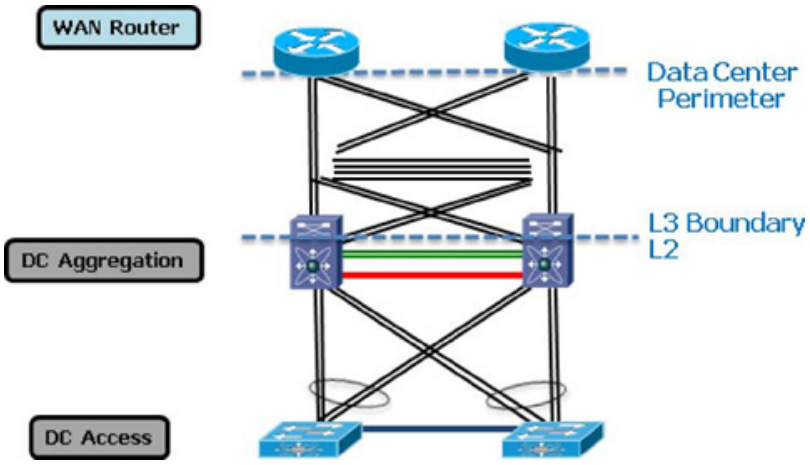
The previous VMDC 2.2 architecture utilized such a three-layer design within the DC, however, for the VMDC 2.3 architecture, the primary goals are to:

1. Increase tenancy scale
2. Reduce cost of the solution
3. Require fewer VMs per tenant (typically 1 VLAN and 4-5 VMs per tenant), targeting the Small, Medium Business (SMB) market for the public or virtual private cloud
4. Require less VM density and east-west bandwidth within a tenant

Based on these requirements, the VMDC 2.3 architecture has been optimized by eliminating the Core layer, as the Core layer adds to cost and reduces tenancy scale due to control plane (Border Gateway Protocol (BGP) and Virtual Routing and Forwarding (VRF)) limits on the platforms. Further, since most tenants are contained within the Aggregation layer (fewer VMs and VLANs per tenant), there is no need to have a Core layer that can provide for routing capabilities between multiple Aggregation layers. Instead, the WAN Edge router can provide the same functionality if needed.

Figure 1-2 illustrates these two layers of the hierarchical model.

Figure 1-2 VMDC 2.3 Two-Layer Hierarchical Model

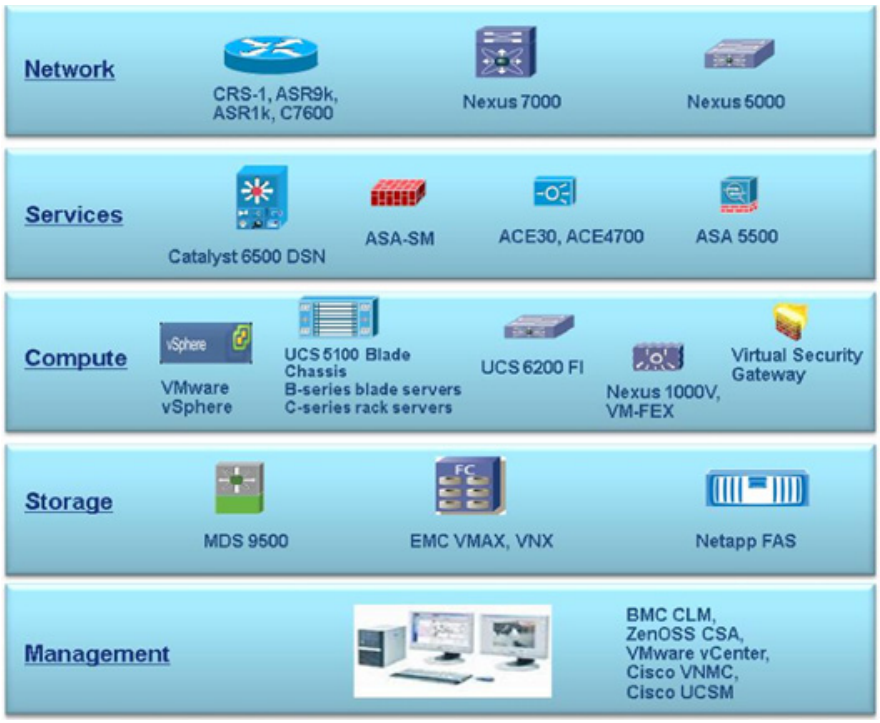


Benefits of such a hierarchical model include scalability, resilience, performance, maintainability, and manageability. The hierarchical design represents a structured approach to building the infrastructure, allowing for relatively easy expansion in modular increments. Redundant nodes and links at each level ensure no single point of failure, while link aggregation can be engineered for optimal bandwidth and performance through the Aggregation and Core layers. Devices within each layer perform the same functions, and this consistency simplifies troubleshooting and configuration. The effect is ease of maintenance at lower operational expense.

VMDC Layers

Figure 1-3 illustrates the functional layers within the VMDC architecture.

Figure 1-3 Functional Layers within the VMDC Data Center



The Network layer includes the WAN/PE router, which forms the DC perimeter to the Enterprise wide area or provider IP/NGN backbone and to the public Internet. These perimeter nodes may be dedicated to L3 routing functions or may be multiservice in nature, providing L2 interconnects between data centers, as well as L3 services. WAN/PE routers validated within the VMDC reference system architecture include the Cisco CRS-1, Cisco ASR 9000, ASR 1000, Cisco 7600, and Cisco Catalyst 6500 platforms. The Network layer also includes the aforementioned, two-layer hierarchy of switching nodes. Within the VMDC reference architecture, this portion of the infrastructure is comprised of Nexus 7000 systems serving as the aggregation nodes, and the Nexus 5000 systems as the access nodes. These systems allow for fine-tuning of port capacity and bandwidth to the level of aggregation or access density required to accommodate current and anticipated scale requirements. In the VMDC 2.3 architecture, the ASR 1000 is used as the WAN/PE router, the Nexus 7004 is used as the aggregation device, and the Nexus 5548 is used as the access device.

The Services layer comprises network and security services such as firewalls, server load balancers, SSL offload, intrusion prevention, network analysis, etc. A distinct difference arises between the conventional DC Services layer and cloud DC Services layer in that the solution set for the latter must support application of L4 - L7 services at a per-tenant level, through logical abstraction of the physical resources. Centralized services are most useful in applying policies that are broadly applicable across a range of tenants (or workgroups in the private case). This layer also serves as the termination point for remote access IPsec or SSL VPNs. Within the VMDC reference architecture, the Catalyst 6500 DSN can provide firewalling and server load-balancing services, in a service module form factor (i.e., the ACE30 and ASASM service modules); alternatively, these are available in appliance form-factors. In the VMDC 2.3 architecture, to keep to smaller footprint and cost, the ASA and ACE appliances serve as the Services layer. Specifically, the ASA 5585-X60 is utilized for firewall services, the ASA 5555X for IPsec/SSL VPN remote access, and the ACE 4710 for Server Load Balancing (SLB).

The Compute layer includes several sub-systems. The first is a virtual Access switching layer, which allows for extension of the L2 network across multiple physical compute systems. This virtual Access switching layer is of key importance in that it also logically extends the L2 network to individual VMs

within physical servers. The feature-rich Cisco Nexus 1000V fulfills this role within the architecture. Depending on the level of software functionality (i.e., Quality of Service (QoS) or security policy) or scale required, the Cisco UCS VM-FEX may be a hardware-based alternative to the Nexus 1000V. A second sub-system is that of virtual (i.e., vApp-based) services. These may include security, load balancing, and optimization services. Services implemented at this layer of the infrastructure will complement more centralized service application, with unique applicability directly to a specific tenant or workgroup and their applications or VMs. Specific vApp-based services validated within the VMDC 2.3 architecture include the Cisco Virtual Security Gateway (VSG), which provides security policy enforcement point within the tenant Virtual Data Center (vDC) or Virtual Private Data Center (VPDC). The third sub-system within the Compute layer is the computing resource. This includes physical servers, hypervisor software providing compute virtualization capabilities, and the VMs thus enabled. The UCS, featuring redundant 6200 Fabric Interconnects, UCS 5108 Blade Chassis, and B-Series Blade or C-Series RackMount servers, comprise the compute resources utilized within the VMDC reference architecture.

The Storage layer provides storage resources. Data stores will reside in SAN (block-based) or NAS (file-based) storage systems. SAN switching nodes (MDS) implement an additional level of resiliency, interconnecting multiple SAN storage arrays to the compute resources, via redundant Fibre Channel (FC) (or Fibre Channel over Ethernet (FCoE)) links. The VMDC architecture has been validated with both EMC and NetApp storage arrays and will also work with any other storage vendors.

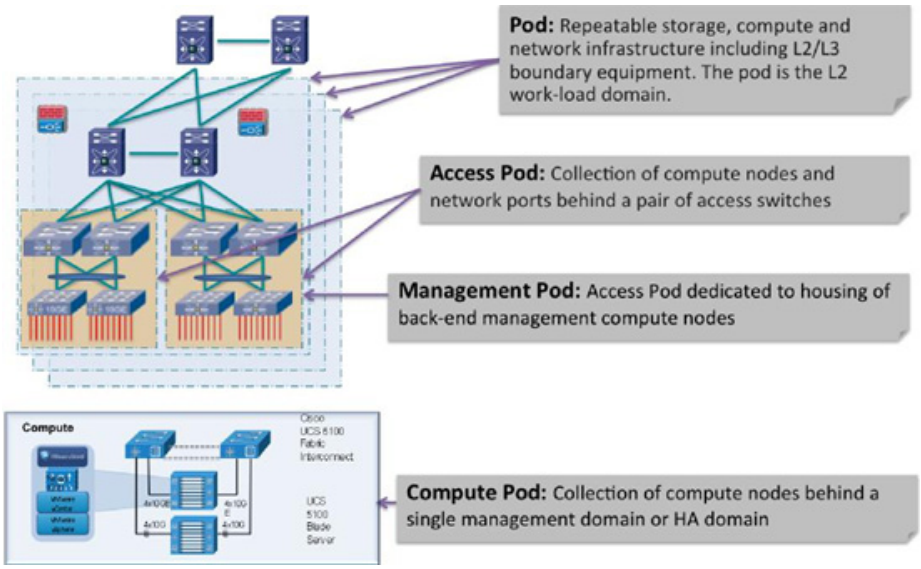
The Management layer consists of the hardware and software resources required to manage the multitenant infrastructure. These include domain element management systems, as well as higherlevel service orchestration systems. The domain management systems that have been validated within VMDC include the UCS Manager (UCSM), VMware vCenter and vCloud Director for compute resource allocation; EMC Unified Infrastructure Manager (UIM), NetApp OnCommand Unified Manager and OnCommand System Manager, NetApp VSC (Virtual Storage Console - a vCenter plugin that provides end-to-end virtual machine (VM) monitoring, provisioning, B&R and management for VMware vSphere environments running on NetApp storage) and Cisco Fabric Manager for storage administration; and the Cisco Virtual Supervisor Module (VSM) and Virtual Network Management Center (VNMC) for virtual access and virtual services management. Automated service provisioning, including cross-resource clooud service orchestration functions, are provided by BMC's Cloud Lifecycle Management (CLM) system; while cloud service assurance is provided by Zenoss Cloud Service Assurance (CSA). Typically, these management systems are hosted in a separate Management pod, so as to isolate the failure domains between the production and management systems.

Modular Building Blocks

The Pod

Previous iterations of the VMDC reference architecture defined resource containers called “pods” that serve as the basis for modularity within the cloud DC. As a homogenous modular unit of network, compute, and storage resources, the pod concept allows one to address environmental, physical, logical, and application-level requirements in a consistent way. The pod serves as a blueprint for incremental build-out of the cloud DC in a structured fashion. When resource utilization within a pod reaches a pre-determined threshold (i.e., 70-80%), the idea is that one simply deploys a new pod. From a service fulfillment and orchestration perspective, a pod represents a discrete resource management domain.

Figure 1-4 Pod Concept



In general practice, the pod concept may serve simply as a framework, with designers defining their own variants tuned to specific environmental or performance characteristics. As Figure 1-4 illustrates, a pod can be defined at different levels of modularity, supporting growth in differing increments. Within the VMDC reference architecture, however, a general purpose utility compute pod extends from the Compute and Storage layers to the L2 ports on the aggregation nodes serving as the L2/L3 boundary, and up to and including components within the network Services layer. The port and MAC address capacity of the aggregation nodes are thus key factors in determining how many pods a single pair of aggregation nodes will support within the cloud DC.

Special Purpose Pods

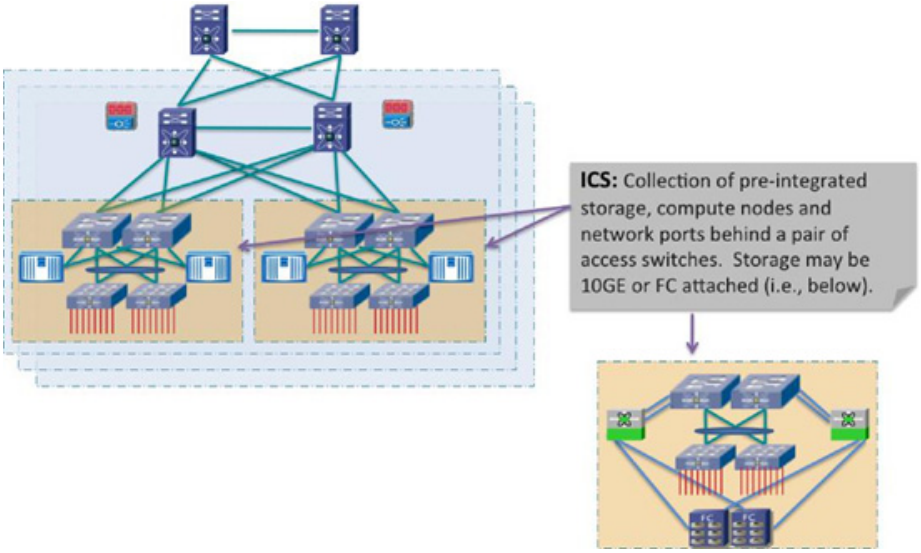
A major premise behind building general purpose homogeneous compute pods and applying logical segmentation overlays to meet business or security policy requirements is that this maximizes utilization of resources, however, in some cases there may be a unique requirement - for ease of operation, special performance tuning, or to meet special security objectives - to physically separate some of the compute nodes out from a general purpose pod and place them in a dedicated, perhaps application-specific pod. The VMDC architecture provides the flexibility to build special purpose pods, and such is the case with the management pod concept.

Back-end management compute nodes may be placed within a general purpose compute pod, and logically isolated and firewalled from production hosts. For smaller, less complex or more streamlined environments, this is an excellent option, however, in larger environments, a separate pod dedicated to back-end management servers (i.e., bare metal and virtualized) is recommended. In the various VMDC 2.X releases, the as-tested systems have in fact included a separate access pod in which servers are dedicated to back-end infrastructure management functions. The benefits of this option include creation of a more discrete troubleshooting domain in the event of instability or failures. The architecture flexibility allows for logical isolation and firewalling or for dedicated firewalls (physical or in vApp form) to be placed on the perimeter of the management container. In practice, Role-based Access Controls (RBAC) tied to directory services would be applied to categorize and limit user access and change control authority as per their functional roles within the organization.

The Integrated Compute and Storage Stack

An Integrated Compute and Storage (ICS) stack represents another potential unit of modularity within the VMDC cloud DC, representing a sub-component within the pod. An ICS is a pre-integrated collection of storage, compute, and network resources, up to and including L2 ports on a pair of access switching nodes. Figure 1-5 illustrates the location of the ICS within a pod. Multiples instances of an ICS are deployed like building blocks to fill the capacity of a pod.

Figure 1-5 ICS Concept



Working with eco-system partners, Cisco currently supports two ICS stack options, a Vblock and a FlexPod. A Vblock comprises UCS and EMC storage systems, offered in several combinations to meet price, performance, and scale requirements. Similarly, a FlexPod unit combines UCS compute and storage resources, however in this case, NetApp storage systems are used. The VMDC reference architecture will accommodate more generic units of compute and storage, including storage from other third-party vendors, however, the business advantage of an ICS stack is that pre-integration takes the guesswork out of balancing compute processing power with storage Input/Output Operations Per Second (IOPS) to meet application performance requirements.

The Cisco/Netapp FlexPod units are offered in a range of sizes designed to achieve specific workload requirements. The FlexPod architecture is highly modular or "podlike'. Each of the component families of the FlexPod can be scaled both up (adding resources to a FlexPod unit) for greater performance and capacity, and out (adding more FlexPod units) for environments that require consistent, multiple deployments while supporting the same feature-sets and functionality as the base FlexPod.

Some of the key benefits of FlexPod with clustered Data ONTAP are:

- **Non-Disruptive Operations**—Customers never have to deal with a storage outage again. Storage services are always available, even while systems or software are being upgraded or replaced. Immortal data storage infrastructure is now a reality.
- **On-demand Flexibility—**Businesses can scale their storage (both up and out and for performance and capacity), compute, and network resources almost without limits to keep up with today's monumental data growth, all without an interruption in service.

- **Operational Efficiency and Multi-Tenancy**—Organizations can operate more efficiently and become agile by managing multiple systems as a single entity. One storage pool supports a diverse set of applications that companies use to run the business. Storage-related services required to support and protect the business are all automated by using storage service catalogs.

SAN and NAS Architecture

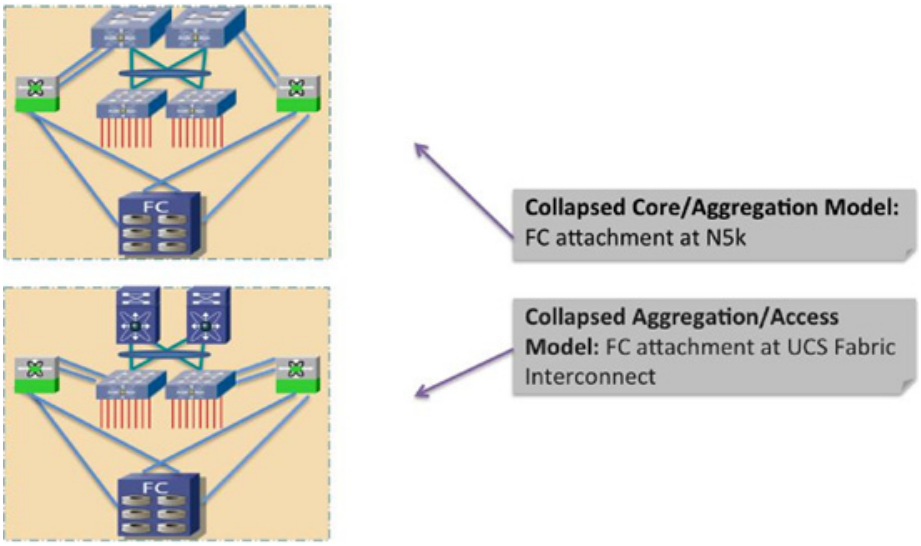
SAN Architecture

The VMDC 2.3 SAN architecture remains unchanged from previous (2.x) designs. It follows current best practice guidelines for scalability, high availability, and traffic isolation. Key design aspects of the architecture include:

- Leverage of Cisco Data Center Unified Fabric to optimize and reduce LAN and SAN cabling costs
- High availability through multi-level redundancy (link, port, fabric, Director, Redundant Array of Independent Disks (RAID))
- Risk mitigation through fabric isolation (multiple fabrics, Virtual SANs (VSANs))
- Data store isolation through N-Port Virtualization (NPV)/N-Port Identifier Virtualization (NPIV) virtualization techniques, combined with zoning and Logical Unit Number (LUN) masking

The hierarchical, pod-based infrastructure model described in this document lends itself to two possible attachment points for storage, within the pod and/or at the aggregation nodes - i.e., distributed or centralized. In practice, which option is most suitable for a particular deployment will depend on application characteristics and anticipated traffic patterns for interactions involving data store access. Companies often employ both options in order to satisfy specific application requirements and usage patterns. In terms of the VMDC validation work, the focus to date has been on consideration of storage as a distributed, pod-based resource. This is based on the premise that in a hierarchical, cloud-type DC model, it is more efficient in terms of performance and traffic flow optimization to locate data store resources as close to the tenant hosts and vApps as possible. In this context, there are two methods of attaching FC storage components into the infrastructure. The first method follows the ICS model of attachment via the Nexus 5000, and the second method provides for attachment at the UCS Fabric Interconnect. Both methods are illustrated in Figure 1-6.

Figure 1-6 FC SAN Attachment Options



In both scenarios, Cisco's unified fabric capabilities are leveraged with Converged Network Adapters (CNAs) providing "SAN-ready" servers, and N-Port Virtualizer on the UCS Fabric Interconnect or Nexus 5000 Top-of-Rack (ToR) switches enabling each aggregated host to be uniquely identified and managed through the fabric and over uplinks to the SAN systems. In order to match the current maximum processing capability of the SAN system, and thus, eliminate lack of bandwidth between the SAN components and their point of attachment to the network infrastructure as a potential bottleneck, multiple FC links are used from each (redundant) Nexus 5000 or UCS Fabric Interconnect to the MDS SAN switches.

In the FlexPod aligned VMDC 2.3 validation, the NetApp FAS arrays were FC attached to the Nexus 5000 access switch in the FlexPod. For more information on the FlexPod architecture, refer to http://www.netapp.com/us/media/tr-4036.pdf. From a storage/FlexPod automation perspective, OnCommand Workflow Automation (WFA), NetApp's storage automation product, makes common storage management processes simple and easy. Storage experts can easily define common storage management processes like provisioning, setup, migration, and decommissioning, and make them available for execution by approved users. WFA can leverage the current automation policies to demonstrate the value of a "Storage Service Catalog" and can also integrate with the existing orchestration systems. More details can be found at https://communities.netapp.com/community/products_and_solutions/storage_management_software/workflow-automation

Although Figure 1-6 shows a very simplistic SAN switching topology, it is important to note that if greater SAN port switching capacity is required, the architecture supports (and has been validated with) more complex, two-tier core-edge SAN topologies, as documented in the VMDC 2.0 Compact Pod Implementation Guide, and more generally in the Cisco SAN switching best practice guides, available at http://www.cisco.com/en/US/prod/collateral/ps4159/ps6409/ps5990/white_paper_C11-515630.html.

NAS Architecture

The VMDC 2.3 NAS architecture follows current best practice guidelines for scalability, high availability, and traffic isolation. Key design aspects of the FlexPod architecture include:

- Infrastructure resiliency through multi-level redundancy of FRU components, multipath HA controller configurations, RAID-DP, and software enhancements that help with failures from a software perspective and a hardware perspective.

- Risk mitigation through fabric isolation and multi-level redundancy of connections (multiple fabrics, vPCs or port-channels, interface groups at the storage layer).
- Cisco virtual Port Channels (vPC) address aggregate bandwidth, link, and device resiliency. The Cisco UCS fabric interconnects and NetApp FAS controllers benefit from the Cisco Nexus vPC abstraction, gaining link and device resiliency as well as full utilization of a nonblocking Ethernet fabric. From a storage perspective, both standard LACP and the Cisco vPC link aggregation technologies play an important role in the FlexPod design.
- Network redundancy in clustered Data ONTAP is supported by both the interconnect and the switching fabric, permitting cluster and data and management network interfaces to fail over to different nodes in the cluster, which extends beyond the HA pair.

Compute Architecture

The VMDC compute architecture is based upon the premise of a high degree of server virtualization, driven by DC consolidation, the dynamic resource allocation requirements fundamental to a cloud model, and the need to maximize operational efficiencies while reducing Capital Expense (CAPEX). The architecture is based upon three key elements:

1. **Hypervisor-based virtualization—**In this release as well as previous system releases, VMware's vSphere plays a key role, enabling the creation of VMs on physical servers by logically abstracting the server environment in terms of CPU, memory, and network touch points into multiple, virtual software containers.
2. **Unified Computing System (UCS)—**Unifying network, server and I/O resources into a single, converged system, the CS provides a highly resilient, low-latency unified fabric for the integration of lossless 10-Gigabit Ethernet and FCoE functions with x-86 server architectures. The UCS provides a stateless compute environment that abstracts I/O resources and server personality, configuration and connectivity, facilitating dynamic programmability. Hardware state abstraction makes it easier to move applications and operating systems across server hardware.
3. **Nexus 1000V—**The Nexus 1000V provides a feature-rich alternative to VMware's Distributed Virtual Switch (DVS), incorporating software-based VN-link technology to extend network visibility, QoS, and security policy to the VM level of granularity.

This system release utilizes VMware's vSphere 5.0 as the compute virtualization operating system. A complete list of new enhancements available with vSphere 5.0 is available online. Key baseline vSphere functionality leveraged by the system includes ESXi boot from SAN, Auto Deploy, VMware High Availability (VMware HA), and Distributed Resource Scheduler (DRS).

Fundamental to the virtualized compute architecture is the notion of clusters. A cluster consists of two or more hosts with their associated resource pools, virtual machines, and data stores. Working in conjunction with vCenter as a compute domain manager, vSphere's more advanced functionality, such as HA and DRS, is built around the management of cluster resources. vSphere supports cluster sizes of up to 32 servers when HA and/or DRS features are utilized. In general practice, however, the larger the scale of the compute environment and the higher the virtualization (VM, network interface, and port) requirement, the more advisable it is to use smaller cluster sizes in order to optimize performance and virtual interface port scale. Therefore, in large VMDC deployments, cluster sizes are limited to eight servers; in smaller deployments, cluster sizes of 16 or 32 can be utilized. As in the VMDC 2.2 release, three compute profiles (Gold, Silver, and Bronze) are created to represent Large, Medium, and Small workload types. Gold has 1 vCPU/core and 16G RAM, Silver has .5 vCPU/core and 8G RAM, and Bronze has .25 vCPU/core and 4G of RAM. The above sizing characteristics are provided as generic reference points in the VMDC architecture and have no bearing to Application or IOPS requirements.

While the VMDC 2.3 architecture works with Vblocks and FlexPods, the system has been validated with FlexPod ICS, which has the following characteristics:

- The architecture comprises multiple UCS 5100 series chassis (5108s), each populated with eight (half-width) server blades.
- Each server has dual 10GigE attachment, i.e., to redundant A and B sides of the internal UCS fabric.
- The UCS is a fully redundant system, with two 2208XP Series Fabric Extenders per chassis connecting up to a pair of UCS 6248UP Fabric Interconnects.
- Internally, four uplinks per Fabric Extender feed into dual Fabric Interconnects to provide the maximum bandwidth possible per server. This means that for server-to-server traffic within the UCS fabric, each server will have 10GigE bandwidth.
- Each UCS 6248 Fabric Interconnect aggregates via redundant 10GigE EtherChannel connections into the access switch (Nexus 5548UP). The number of uplinks provisioned will depend upon traffic engineering requirements. For example, in order to provide an eight-chassis system with an 8:1 oversubscription ratio for internal fabric bandwidth to aggregation bandwidth, a total of 80G (8x10G) of uplink bandwidth capacity must be provided per UCS system.
- For FC connectivity, eight ports on the Nexus 5548 provide 8Gig FC direct connectivity to the NetApp FAS storage arrays. In order to maximize IOPS, the aggregate link bandwidth from the Nexus 5548 to the FAS should match the processing capability of the storage controllers.
- The Nexus 1000V functions as the virtual Access switching layer, providing per-VM policy and policy mobility.

The current version of the CVD is "VMWare vSphere 5.1 on FlexPod Clustered Data ONTAP" which can be found at FlexPod Animal release.

For NAS connectivity, the FlexPod architecture leverages both the Unified Target Adapter (UTA) and the traditional 10GbE Ethernet adapter for storage connectivity. The UTA provides the greatest flexibility when migrating to an end-to-end FCoE design, however, a standard 10GbE can be used for IP-based storage designs. The vPC links between the Nexus 5548 switches and NetApp storage controllers' UTA are converged, supporting both FCoE and traditional Ethernet traffic at 10Gb providing a robust connection between initiator and target. The UTAs installed in each NetApp storage controller use FCoE to send and receive FC traffic to and from the Cisco Nexus switches over 10GbE. The Cisco UCS system also uses FCoE to send and receive FC traffic to and from the various Cisco UCS components (for example, the Cisco UCS B-Series blade servers and Cisco UCS C-Series servers). The FlexPod Animal topology is the first to leverage true end-to-end FCoE, which significantly simplifies the network design and therefore reduces application time to market.

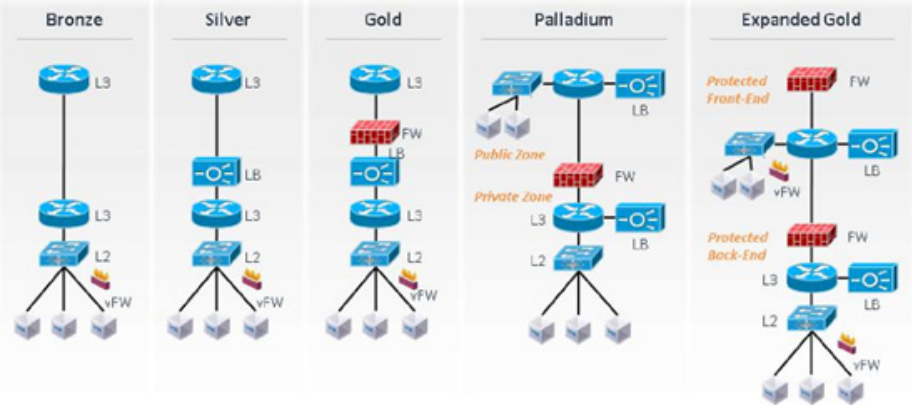
Multitenancy Architecture

Virtualization of compute and storage resources enables sharing across an organizational entity. In contrast, virtualized multitenancy, a concept at the heart of the VMDC reference architecture, refers to the logical isolation of shared virtual compute, storage, and network resources. In essence, this is “bounded” or compartmentalized sharing. A tenant is a user community with some level of shared affinity. For example, within an Enterprise, a tenant may be a business unit, department, or workgroup. Depending upon business requirements or regulatory policies, a tenant “compartment” may stretch across physical boundaries, organizational boundaries, and even between corporations. A tenant container may reside wholly within their private cloud or may extend from the tenant's Enterprise to the provider's facilities within a public cloud. The VMDC architecture addresses all of these tenancy use cases through a combination of secured datapath isolation and a tiered security model, which leverages classical security best practices and updates them for the virtualized multitenant environment.

Tenancy Models

Earlier VMDC releases presented five tenancy models or containers. High-level, logical depictions of these models are illustrated in Figure 1-7.

Figure 1-7 Existing VMDC Tenancy Models



The first three models provide a baseline, simple set of tenant containers, which were combined with different levels of network services in a tiered fashion, hence the Bronze, Silver, and Gold nomenclature. The two most interesting containers from this set are Bronze and Gold. Bronze is seemingly the most basic, but simplicity broadens its applicability. One tends to think of these containers as single tenant in nature, but in practice, a Bronze container may be used to support multiple tenants, with homogenous requirements, i.e., similar workload profiles, QoS, or security policies, or perhaps this is a community of interest using the same application set.

A Gold container, with both firewall and server load balancer services applied, assumes a higher degree of security and availability. As in the Silver container, multiple VLANs support logical segmentation for N-tiered applications. The idea is that one could combine these tenant containers together in various combinations to support more complex scenarios if desired.

The fourth container type (Palladium) demonstrates a further incremental evolution of tenancy models from simple multisegment containers toward logical approximations of a vDC overlay on the physical shared infrastructure. With the notion of a separate front-end and back-end set of zones, each of which may have a different set of network services applied, the Palladium container begins to more closely align with traditional zoning models in use in physical IT deployments.

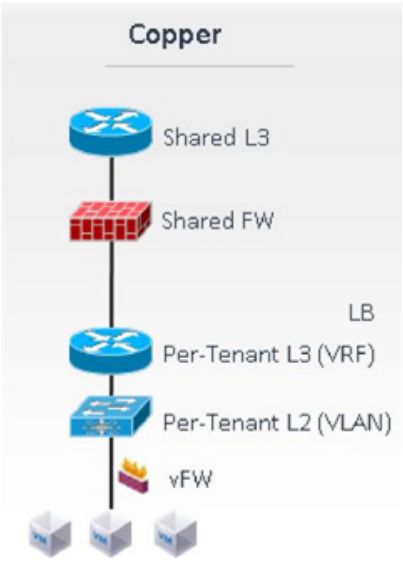
The fifth container type (Expanded Gold container) incrementally evolves the vDC concept, providing more expansion of protected front-end and back-end zones while furthering the notion of separate public (i.e., Internet or Demilitarized Zone (DMZ)) or shared (i.e., campus/inter-organizational) access from private access. It also includes secured remote IPsec or SSL VPN access. In this case, the term “private” can mean that the vDC is routed over the private Enterprise WAN or through the public cloud provider's IP/NGN via a private MPLS VPN. In the public cloud scenario, this type of virtual DC linked to the tenant Enterprise via an L2 or L3 MPLS VPN, is commonly termed a VPDC. MPLS VPNs are often used by public cloud providers as transport for hybrid managed cloud services. Such services may include IP addressing, security (i.e., firewalling, managed DMZ, zoning, secured remote VPN access), and server resiliency solutions.

New Tenancy Model Introduced in VMDC 2.3

A new tenancy model, the Copper container, is introduced in VMDC 2.3. This tenancy model has been designed to provide higher tenancy scale in VMDC cloud deployments, and is suitable for Internet-based access to cloud resources. The Copper container is relevant to SMBs who require one VLAN and a handful of VMs in the cloud. Such customers require isolation and security, but typically do not wish to pay higher fees for utilizing their own virtual firewall context in the cloud. The Copper container solves this need by utilizing a common firewall shared across such tenants, with each tenant getting their own VLAN and VRF instance for isolation behind the shared firewall.

Figure 1-8 shows the new Copper tenancy model for VMDC 2.3.

Figure 1-8 New VMDC 2.3 Copper Tenancy Model



Storage Multitenancy

From a storage perspective, secure multitenancy is the use of secure virtual partitions within a shared physical storage environment for the purpose of sharing the physical environment among multiple distinct tenants. For instance, a storage SP might configure a storage array in such a way that each of three different customers is provisioned a certain portion of the array's disk capacity and network resources. In a secure multi-tenant environment, each customer would have access only to the resources explicitly provisioned to that customer. The customer would not have access to other customers' data, and not even be aware of the existence of the other customers or the fact that they share a common physical array.

The VMDC 2.3 system has been validated with the FlexPod ICS, with NetApp FAS storage arrays. NetApp FAS arrays can be utilized in clustered Data ONTAP or Data ONTAP 7-Mode configurations. Clustered Data ONTAP is an inherently multi-tenant storage operating system and is architected in such a way that all data access is done through secure virtual storage partitions. It is possible to have a single partition that represents the resources of the entire cluster or multiple partitions that are assigned specific subsets of cluster resources. These secure virtual storage partitions are known as Storage Virtual Machines (SVM). A SVM is effectively isolated from other SVMs that share the same physical hardware. Because it is a secure entity, a SVM is only aware of the resources that have been assigned to

it and has no knowledge of other SVMs and their respective resources. Each SVM operates as a separate and distinct entity with its own security domain. Tenants may manage the resources allocated to them through a delegated SVM administration account.

Cloud Services

Another concept at the heart of the VMDC reference architecture is the notion of differentiated service tiering. Simply put, tenants may have unique requirements in terms of network throughput, compute processing, storage performance, or data store privacy characteristics, and a successful multitenant deployment must be able to address these needs.

Differentiated Services

By definition, in a cloud-based model, compute, storage, and network infrastructure are abstracted and delivered "as a service." In order to tailor workload characteristics or application performance to specific needs, the cloud administrator has various methods at hand for providing differentiated service tiers and ensuring that tenant privacy and Service Level Agreement (SLA) objectives are met:

- **Tiered Workload Definitions—**The secret to building a cloud-ready infrastructure is in categorizing the set of applications that must be supported and distilling these into their basic workload characteristics. Once these are reasonably understood, they can in most cases be addressed by a set of standard service profiles. For example, characteristics which apply to the ICS include VM attributes (CPU ratio, memory and associated storage capacity), storage attributes (RAID levels, disk types and speeds, and protection mechanisms), and support for various degrees of application tiering.
  - In the context of FlexPod and Netapp FAS storage arrays, the NetApp Virtual Storage Tier is a self-managing data-driven service layer for storage infrastructure that lends itself very well to hosting tiered workloads. VST is natively built into the Data ONTAP operating system and works by leveraging block-sharing technologies such as NetApp primary storage deduplication and file/volume FlexClone to reduce the amount of cache required and eliminate duplicate disk reads. This is extended with Flash Cache and Flash Pool technology, which provides intelligent caching enabling real-time assessment of workloadbased priorities, and enables I/O data requests to be optimized for cost and performance without requiring complex data classification.
- **Availability Mechanisms—**Availability mechanisms may be applied at various layers of the infrastructure to ensure that communication requirements are met. For example, within a vSphere cluster, DRS and vMotion or Fault Tolerance (FT) may be used to provide optimal resource allocation, even in the event of server failure. Similarly, within the SAN, data protection mechanisms such as snapshots, cloning, and backup archiving help to ensure that data store integrity is preserved through various types of failure scenarios. Network services, such as SLB, encryption, advanced routing and redundancy, can further help to achieve availability targets. The larger the shared domain (ICS, pod, or entire DC level), the broader the impact of the availability mechanisms utilized at that particular layer of the hierarchy. As these typically do not come without added cost, the goal would be to ensure that broadly scoped availability methods meet minimum targeted requirements for the entire tenant community.
- **Secure Isolation—**In a multitenant environment, the ability to securely contain and isolate tenant traffic is a fundamental requirement, protecting tenant resources and providing risk mitigation in the event that a specific tenant's privacy is breached. Like availability, isolation mechanisms are applied in a multilayered fashion in order to implement the requisite infrastructure protection and security

zoning policies on a per-tenant basis. In practice, techniques fall into two categories of physical and logical isolation mechanisms, however, VMDC analysis focuses mainly on logical mechanisms. These include various L2 and L3 mechanisms, such as multiple vNICs (i.e., for specific control or data traffic), 802.1q VLANs, MPLS VRF instances, VSANs, combined with access control mechanisms (i.e., RBAC and directory services, IPsec or SSL VPNs), and packet filtering and firewall policies.

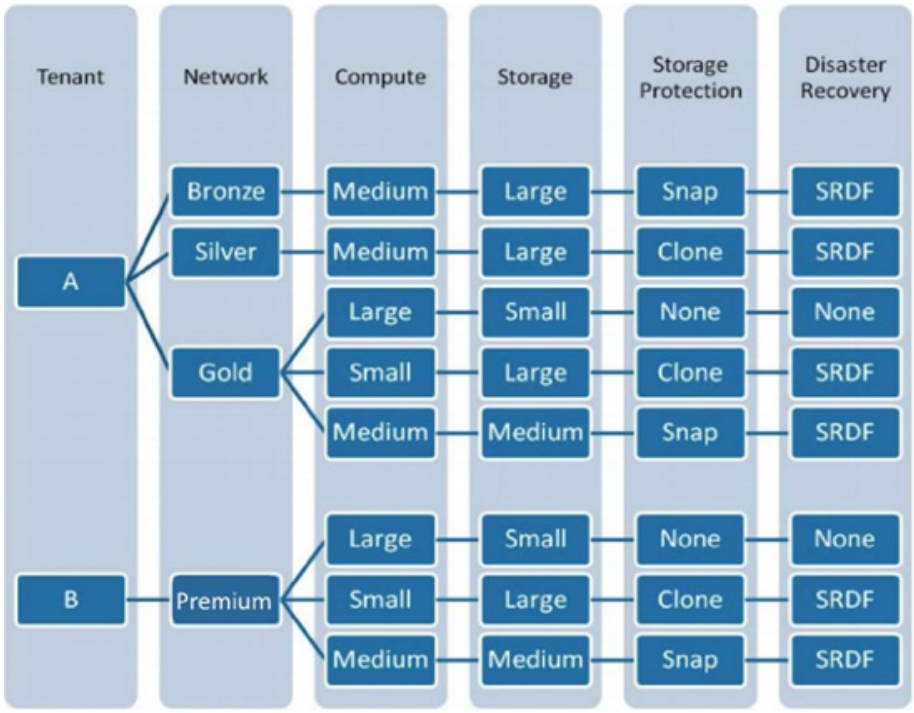
  - From a FlexPod and NetApp storage perspective, clustered Data ONTAP provides a logical mechanism for secure isolation of tenants using Storage Virtual machines (SVMs). Secure multi-tenancy using SVM allows businesses to consolidate tenants into shared resources, and assures that tenants will not have access to resources that are not explicitly assigned to them. Tenants sharing the same physical hardware can operate independently and with the expectation that no single tenant will consume resources unfairly. For example, production and dev/test can run on the same system without the risk of dev/test affecting production workloads. Because it is a secure entity, an SVM is only aware of the resources that have been assigned to it and has no knowledge of other SVMs and their respective resources. Each SVM operates as a separate and distinct entity with its own security domain. Tenants may manage the resources allocated to them through a delegated SVM administration account, and each SVM may connect to unique authentication zones such as Active Directory, LDAP, or NIS.

- **Service Assurance Mechanisms—**Service assurance is a function of availability and QoS policies. The implementation of QoS policies allows for differentiated classification and treatment of traffic flows per tenant, per service tier during periods of congestion.
- **Management—**The ability to abstractly represent per-tenant resources and services in the form of a Service Catalog is a prerequisite for automated service fulfillment and service assurance functions, i.e., the “Day 1” and “Day 2” management tasks which are so essential to operating under an IaaS model. The Service Catalog is effectively the highest level of abstraction for the underlying cloud resources. Accurate representations of these resources as policy-based tenancy models to the service catalog rely on interactions directly with domain element managers or middleware Management layers via standardized interfaces (i.e., APIs, MIBS, etc.). The more intelligent the middleware layer, the less work has to be done at higher levels in the management framework to understand the tenancy models and commission or decommission resources on a per-tenant basis.

Service Tiering

Previous VMDC releases were modeled based on three baseline categories of tenant network services tiers - Bronze, Silver, and Gold - represented in terms of firewalling, server load balancing, SSL offload, and QoS policy (i.e., three data classes of service), combined with three workload models, each with specific compute attributes, associated storage characteristics, and business continuance services. Figure 1-9 is a high-level conceptual illustration of these models, demonstrating a variety of ways in which these resources and services can be applied in combination to meet business or application requirements in a tiered fashion.

Figure 1-9 VMDC Service Tiers



In VMDC 2.3, these definitions are augmented, with the expansion of the Gold service tier to create a premium "Expanded Gold" tier. This premium tier is enabled through a QoS framework which adds SLA support for low latency Voice over IP (VoIP) and multimedia (i.e., video) traffic, in addition to control and differentiated data traffic classes. VMDC 2.3 also introduces an SMB "Copper" tier.

In the context of FlexPod and NetApp storage arrays, refer to the following links for more information on tiering, replication, backup, and DR technologies:

- **Virtual Storage Tiering**: http://www.netapp.com/in/technology/virtual-storage-tier/index.aspx
- **SnapMirror Datasheet**: http://www.netapp.com/in/products/protection-software/snapmirror.aspx
- **SnapMirror Best Practices**: http://www.netapp.com/us/media/tr-4015.pdf
- **SnapVault Datasheet**: http://www.netapp.com/in/products/protection-software/snapvault.aspx
- **SnapVault Best Practices**: http://www.netapp.com/us/media/tr-4183.pdf