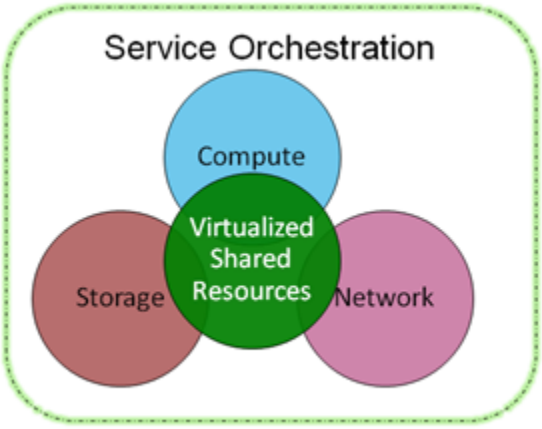
CHAPTER 2

Architecture Overview

A cloud deployment model is distinctive from traditional deployments in its ability to treat the data center as a common fabric of resources available in an on-demand basis. A portion of these pools are dynamically allocated to individual tenants and then deallocated when they are no longer in use. As depicted in Figure 2-1, VMDC defines two key building blocks:

- **Virtualized Shared Resource Pool**—The resource pool consists of network, compute, and storage components. These components are virtualized and used by multiple tenants securely.
- **Service Orchestration**—Service orchestration automates the resource provisioning workflow. It leverages a set of tools and APIs to dynamically provision cloud resources on demand. A tenant initiates the workflow process using a web portal to request specific resources.

Figure 2-1 VMDC Building Blocks



Note This document addresses design aspects of the shared resource pool that a customer must understand before implementing a cloud data center. It does not address service orchestration components and design. A separate module of the VMDC 2.0 document provides service orchestration design and implementation guidance.

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

- **Secure Separation**—Provides end-to-end tenant path isolation and security. Tenants are isolated from each other via several security techniques at different layers of the network or infrastructure. For example, virtual route forwarding instances (VRFs) are leveraged at the Layer 3 to stop

communication between tenants at Layer 3 domain. Likewise, similar isolation features are leveraged at compute and storage layers to provide complete isolation of tenants in a shared infrastructure.

- **Data Center Scalability**—A pod-based architecture provides network architects the ability to modularize the infrastructure into easily replicable units called pods. Architects can plan for an initial pod, which guarantees a certain scale and performance along with a scalable data center core network. This architecture provides a predictable and homogeneous method for adding self-contained pods as additional resources are needed.
- **High Availability**—Availability ensures that the cloud resources are accessible even during a failure situation. Availability is required to meet the expectations of service-level agreements (SLAs) in a cloud deployment.
- **Service Assurance**—Provides mechanisms to define different service levels and defines how to adhere to them using network QoS techniques during both steady and non-steady states. To differentiate IaaS service tiers, network architects can reserve and guarantee certain network bandwidths based on their subscription rules for the tier. For example, a Gold tenant could be guaranteed with 1 Gbs of bandwidth per VM whereas a Silver tenant only gets 0.5 Gbs per VM.

Table 2-1 presents example storage distinctions by service tier.

Table 2-1 Components of the Large Pod Resource Pool

Features	Components
Network	Cisco CRS Cisco Nexus 7010 Cisco Nexus 7018 Data Center Services Node 6509-E (VSS) Firewall Service Module Application Control Engine Module
Compute	Cisco Unified Computing System (UCS) • UCS5108 Blade Server Chassis • UCSB200-M1 Blade Server • UCS M71KR-E Converged Network adapter • UCS M81KR Virtual Interface card • Cisco UCS 6120, Cisco UCS 6140 fabric interconnect
Virtualization	VMware vSphere VMware ESXi 4.0U1 Hypervison Cisco Nexus 1000V (virtual access switch)
Security	Cisco Firewall Services Module (FWSM), ACE Application Control Engine VMware vShield NetApp vFiler and Virtual Service Domains MDS soft zoning and VSANs Cisco Nexus 1000V
Storage Fabric	Cisco MDS 9513

Table 2-1 Components of the Large Pod Resource Pool (continued)

Features	Components
Storage Array	EMC 2 Symmetrix VMAX with virtual provisioning NetApp FAS3170

Orchestration/Management	Domain Management:
	• UCS Manager
	• Nexus 1000V VSM
	• VMware vCenter
	• Fabric Manager
	BMC Cloud Lifecycle Management (CLM):
	• BMC BladeLogic Server Automation
	• BMC BladeLogic Network Automation
	• BMC Remedy Action Request Suite
	  • Service Request Manager (SRM)
	  • Atrium Core
	  • Atrium Orchestrator
	  • Remedy AR System Server
	  • Cloud Extension Pack
	• BMC Remedy Change Management

Discrepancy analysis	Determines if configuration deltas exist within the Cisco CSGs defined in the server farm Provides a report with found discrepancies; shows any diverging Cisco IOS ® Software CLIs between Cisco CSGs

End-to-End Topologies

Figure 2-2 shows the end-to-end logical topology for Gold, Silver, and Bronze service classes.

Figure 2-2 End-to-End Logical Topology (Gold, Silver, Bronze)

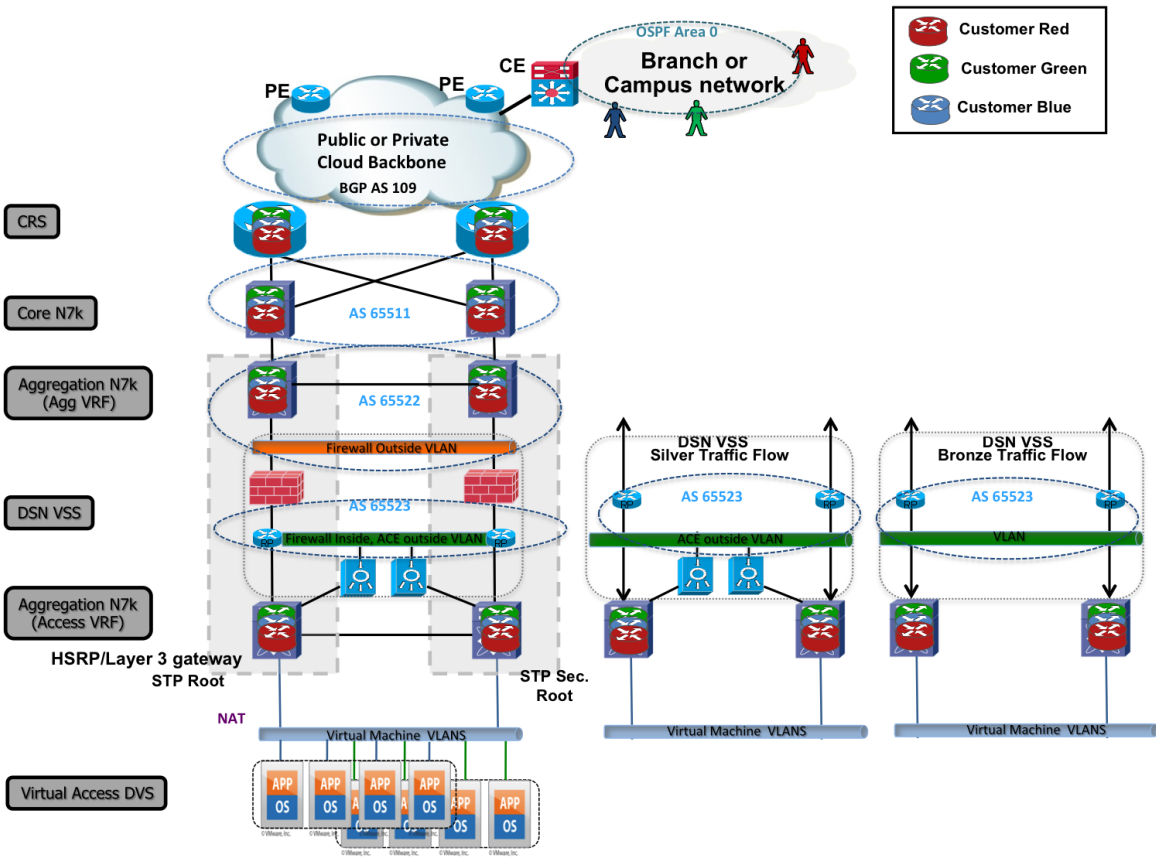


Figure 2-3 the end-to-end physical topology.

Figure 2-3 End-to-End Physical Topology (Gold, Silver, Bronze)