Enhanced Secure Multi-Tenancy Design Guide

Last Updated: October 8, 2010

Cisco Validated Design

Building Architectures to Solve Business Problems

About the Authors

Aeisha Bright

Aeisha Bright, Technical Marketing Engineer, Systems Architecture and Strategy, Cisco Systems

Aeisha Bright, CCIE #13455, is a Technical Marketing Engineer for data center technologies in Cisco's Systems Architecture and Strategy group. Prior to joining the SASU team, Aeisha spent 4 years as a Customer Support Engineer in Cisco's Technical Assistance Center where she supported LAN switching, VPN and Firewall technologies. She earned a B.S. in Computer Science from the University of Maryland at Baltimore County and an M.S. in Computer Networking from North Carolina State University.

Ramesh Isaac

Ramesh Isaac, Technical Marketing Engineer, Systems Architecture and Strategy, Cisco Systems

Ramesh has worked in data center and mixed use lab settings over the past 15 years. He started in information technology supporting Unix environments, with the last couple of years focused on designing and implementing multi-tenant virtualization solutions in Cisco labs. Ramesh holds certifications from Cisco, VMware, and Red Hat.

Alex Nadimi

Alex Nadimi, Solutions Architect, Systems Architecture and Strategy, Cisco Systems

Alex has been with Cisco for the past 15 years and is currently working as a Solutions Architect in Cisco Systems Architecture and Strategy group. Prior to this role, he worked as a Technical Marketing Engineer in the Cisco Central Marketing Organization. He has developed solutions and technical guidance on various technologies such as security, VPN networks , WAN transport technologies, data center solutions, and virtualization. Prior to Cisco, he has worked at Hughes LAN Systems and Northern Telecom . He holds a masters of science in electrical engineering from Louisiana State University.

Chris O'Brien

Chris O'Brien, Solutions Architect, Systems Architecture and Strategy, Cisco Systems

Chris O'Brien is a Solutions Architect for data center technologies in Cisco's Systems Architecture and Strategy group. He is currently focused on data center design validation and application optimization. Previously, O'Brien was an application developer and has been working in the IT industry for more than 15 years.

Chris Reno

Chris Reno, Reference Architect, Infrastructure and Cloud Enablement, NetApp

Chris Reno is a reference architect in NetApp's Infrastructure and Cloud Enablement group and is focused on creating, validating, supporting, and evangelizing solutions based on NetApp products. Before his current role, he worked with NetApp product engineers designing and developing innovative ways to do Q&A for NetApp products, including enablement of a large grid infrastructure using physical and virtualized compute resources. In these roles Reno has gained expertise in stateless computing, netboot architectures, and virtualization.

Henry Vail

Henry Vail, Reference Architect, Infrastructure and Cloud Enablement, NetApp

Henry Vail is a reference architect in NetApp's Infrastructure and Cloud Enablement team, developing strategic solutions that direct NetApp technologies toward realizing the potential of cloud architecture. In this role, he is focused on advancing cloud service automation and management frameworks, stateless computing, and ubiquitous workload deployment and mobility to deliver cohesive, elegant, and effective IT solutions. Prior to joining NetApp in 2008, Vail's career has included microelectronics research, control systems development, network systems engineering, software product engineering, content management integration, data protection, and enterprise systems architecture and management.


Mike Zimmerman

Mike Zimmerman, Reference Architect, Infrastructure and Cloud Enablement, NetApp

Mike Zimmerman is a reference architect in NetApp's Infrastructure and Cloud Enablement team. He focuses on the implementation, compatibility, and testing of various vendor technologies to develop innovative end-to-end cloud solutions for customers. Zimmerman started his career at NetApp as an architect and administrator of Kilo Client, NetApp's internal cloud infrastructure, where he gained extensive knowledge and experience building end-to-end shared architectures based upon server, network, and storage virtualization.


Serge Maskalik

Serge Maskalik, Senior Manager, vShield Engineering, VMware

Serge Maskalik is a Sr. Manager for the vShield Engineering team. He joined VMware through the Blue Lane acquisition. Serge started his career in 1998 in the Internet Service Provider arena, during which time he built large Service Provider data centers for providing Infrastructure as a Service. His focus at VMware now is working with product management, customers, and service providers to build the next generation security product portfolio for vCloud.


Wen Yu

Wen Yu, Senior Infrastructure Technologist, VMware

Wen Yu is a Sr. Infrastructure Technologist at VMware, with a focus on partner enablement and evangelism of virtualization solutions. Wen has been with VMware for six years during which time four years have been spent providing engineering level escalation support for customers. Wen specializes in virtualization products for continuous availability, backup recovery, disaster recovery, desktop, and vCloud. Wen Yu is VMware, Red Hat, and ITIL certified.

Enhanced Secure Multi-Tenancy Design Guide

Introduction

Goal of This Document

Cisco®, VMware®, and NetApp® have jointly designed a best-in-breed Enhanced Secure Multi-Tenancy (ESMT) Architecture and have validated this design in a lab environment. This document describes the design of and the rationale behind the Enhanced Secure Multi-Tenancy Architecture. The design includes many issues that must be addressed prior to deployment as no two environments are alike. This document also discusses the problems that this architecture solves and the four pillars of an Enhanced Secure Multi-Tenancy environment.

Audience

The target audience for this document includes, but is not limited to, sales engineers, field consultants, professional services, IT managers, partner engineering, and customers who wish to deploy an Enhanced Secure Multi-Tenancy (ESMT) environment consisting of best-of-breed products from Cisco, NetApp, and VMware.

Objectives

This document is intended to articulate the design considerations and validation efforts required to design, deploy, and backup Enhanced Secure Multi-Tenancy virtual IT-as-a-service.

Change Summary

New Features

- Virtualization of physical Unified Computing System™ (UCS) adapters with Cisco VIC adapters

Corporate Headquarters:
Cisco Systems, Inc., 170 West Tasman Drive, San Jose, CA 95134-1706 USA

- Enhanced environment security and visibility with Cisco Services
- Tenant backup and replication for data recovery and DR with NetApp SnapDrive® and SnapManager® products
- FCoE target capabilities with NetApp storage
- Pre-production/dev and test environment with VMware Cloud Director
- Open management framework for third-party orchestration

Foundational Components

Each implementation of an Enhanced Secure Multi-Tenancy (ESMT) environment will most likely be different due to the dynamic nature and flexibility provided within the architecture. For this reason this document should be viewed as a reference for designing a customized solution based on specific tenant requirements. The following outlines the foundational components that are required to configure a base Enhanced Secure Multi-Tenancy environment. Add additional features to these foundational components to build a customized environment designed for specific tenant requirements. It is important to note that this document not only outlines the design considerations around these foundational components, but also includes considerations for the additional features that can be leveraged in customizing an Enhanced Secure Multi-Tenancy environment.

The Enhanced Secure Multi-Tenancy foundational components include:

- Cisco Nexus® data center switches
- Cisco Unified Computing System
- Cisco Nexus 1000V Distributed Virtual Switch
- NetApp Data ONTAP®
- VMware vSphere™
- VMware vCenter™ Server
- VMware vShield™

Tenant Driven Requirements

The Enhanced Secure Multi-Tenancy environment provides an enterprise with the flexibility and agility to address a myriad of tenant requirements originating from the business, the application, or external regulations. The enterprise may find it necessary to provide tenants with additional services such as:

- Load balancing
- SSL offload
- Intrusion prevention and detection
- Network analysis

The design does not preclude the introduction of additional network or compute components to meet these specific objectives. This design recognizes that it may be necessary to use one or many intelligent network services to meet the security, scalability, or availability requirements of numerous tenant applications. The example applications in this design are based on previous efforts that focused on the use of network services to create:

- Architectural flexibility—Services are readily deployable and fulfill a role where the “mode” of deployment and model are themselves options
- Predictable traffic patterns—Steady and non-steady state should result in reliable flows
- Consistent convergence—Supporting the high availability requirements within the data center

- Client-to-server services—The introduction of network services for data center ingress and egress traffic

Introducing these services provides a comprehensive solution to meet the aforementioned requirements in a multi-tenant environment.

Problem Identification

Today's traditional IT model suffers from resources located in different, unrelated silos—leading to low utilization, gross inefficiency, and an inability to respond quickly to changing business needs. Enterprise servers reside in one area of the data center and network switches and storage arrays in another. In many cases, different business units own much of the same type of equipment, use it in much the same way, in the same data center row, yet require separate physical systems in order to separate their processes and data from other groups.

This separation often results in ineffectiveness as well as complicating the delivery of IT services and sacrificing alignment with business activity. As the IT landscape rapidly changes, cost reduction pressures, focus on time to market, and employee empowerment are compelling enterprises and IT providers to develop innovative strategies to address these challenges.

The current separation of servers, networks, and storage between different business units is commonly divided by physical server rack and a separate network. By deploying an Enhanced Secure Multi-Tenancy virtual IT-as-a-service, each business unit benefits from the transparency of the virtual environment as it still "looks and feels" the same as a traditional, all physical topology.

From the end customer viewpoint, each system is still separate with its own network and storage; however the divider is not a server rack, but an Enhanced Secure Multi-Tenancy environment. The servers, networks, and storage are all still securely separated, in some case much more so than a traditional environment.

And finally, when a business unit needs more servers, it simple requires an order to the IT team to "fire off" a few more virtual machines in the existing environment, instead of having to order new physical equipment for every deployment.

Design Overview

Cloud computing removes the traditional silos within the data center and introduces a new level of flexibility and scalability to the IT organization. This flexibility addresses challenges facing enterprises and IT service providers that include rapidly changing IT landscapes, cost reduction pressures, and focus on time to market. What is needed is a cloud architecture with the scalability, flexibility, and transparency to enable IT to provision new services quickly and cost effectively by using service level agreements (SLAs) to address IT requirements and policies, meet the demands of high utilization, and dynamically respond to change, in addition to providing security and high performance.

According to National Institute of Standards and Technology (NIST), cloud computing is defined as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of three service models and four deployment models.

There are a number of models for the kind of service offered by the cloud provider. The model embodied by this architecture is commonly referred to as Cloud Infrastructure as a Service (IaaS). In this environment, the capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the

underlying cloud infrastructure, but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls). Other models for cloud deployment include Cloud Software as a Service (SaaS), in which the consumer is given access to software running in the provider's environment, and Cloud Platform as a Service (PaaS), in which the consumer can use the provider's hardware to run applications created with the specific programming languages and tools supported by that provider. These other models can be built on top of the base IaaS environment described in this document.

Cloud environments may be separated into four categories based on the consumers they serve:

- Private cloud—The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.
- Community cloud—The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise.
- Public cloud—The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.
- Hybrid cloud—The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).

Many enterprises and IT service providers are developing cloud service offerings for public and private environments. Regardless of whether the focus is on public or private cloud services, these efforts share several objectives:

- Increase operational efficiency through cost-effective use of expensive infrastructure.
- Drive up economies of scale through shared resourcing.
- Provide rapid and agile deployment of customer environments or applications.
- Improve service quality and accelerate delivery through standardization.
- Promote green computing by maximizing efficient use of shared resources, lowering energy consumption.

Achieving these goals can have a profound, positive impact on profitability, productivity, and product quality. However, leveraging shared infrastructure and resources in a cloud-services architecture introduces additional challenges, hindering widespread adoption by IT service providers who demand securely isolated customer or application environments but require highly flexible management.

As enterprise IT environments have dramatically grown in scale, complexity, and diversity of services, they have typically deployed application and customer environments in silos of dedicated infrastructure. These silos are built around specific applications, customer environments, business organizations, operational requirements, and regulatory compliance (Sarbanes-Oxley, HIPAA, PCI) or to address specific proprietary data confidentiality. For example:

- Large enterprises need to isolate HR records, finance, customer credit card details, etc.
- Resources externally exposed for out-sourced projects require separation from internal corporate environments.
- Health care organizations must ensure patient record confidentiality.
- Universities need to partition student user services from business operations, student administrative systems, and commercial or sensitive research projects.

- Telcos and service providers must separate billing, CRM, payment systems, reseller portals, and hosted environments.
- Financial organizations need to securely isolate client records and investment, wholesale, and retail banking services.
- Government agencies must partition revenue records, judicial data, social services, operational systems, etc.

Enabling enterprises to migrate such environments to a cloud architecture demands the capability to provide secure isolation while still delivering the management and flexibility benefits of shared resources. Both private and public cloud providers must enable all customer data, communication, and application environments to be securely separated, protected, and isolated from other tenants. The separation must be so complete and secure that the tenants have no visibility of each other. Private cloud providers must deliver the secure separation required by their organizational structure, application requirements, or regulatory compliance.

However, lack of confidence that such secure isolation can be delivered with resilient resource management flexibility is a major obstacle to the widespread adoption of cloud service models. NetApp, Cisco, and VMware have collaborated to create a compelling infrastructure solution that incorporates comprehensive compute, network, and storage technologies that facilitate dynamic, shared resource management while maintaining a secured and isolated environment. VMware vSphere, VMware vShield, Cisco Unified Computing System, Cisco Nexus Switches, Cisco MDS Switches, and NetApp MultiStore® with NetApp Data Motion™ deliver a powerful solution to fulfill the demanding requirements for secure isolation and flexibility in cloud deployments of all scales.

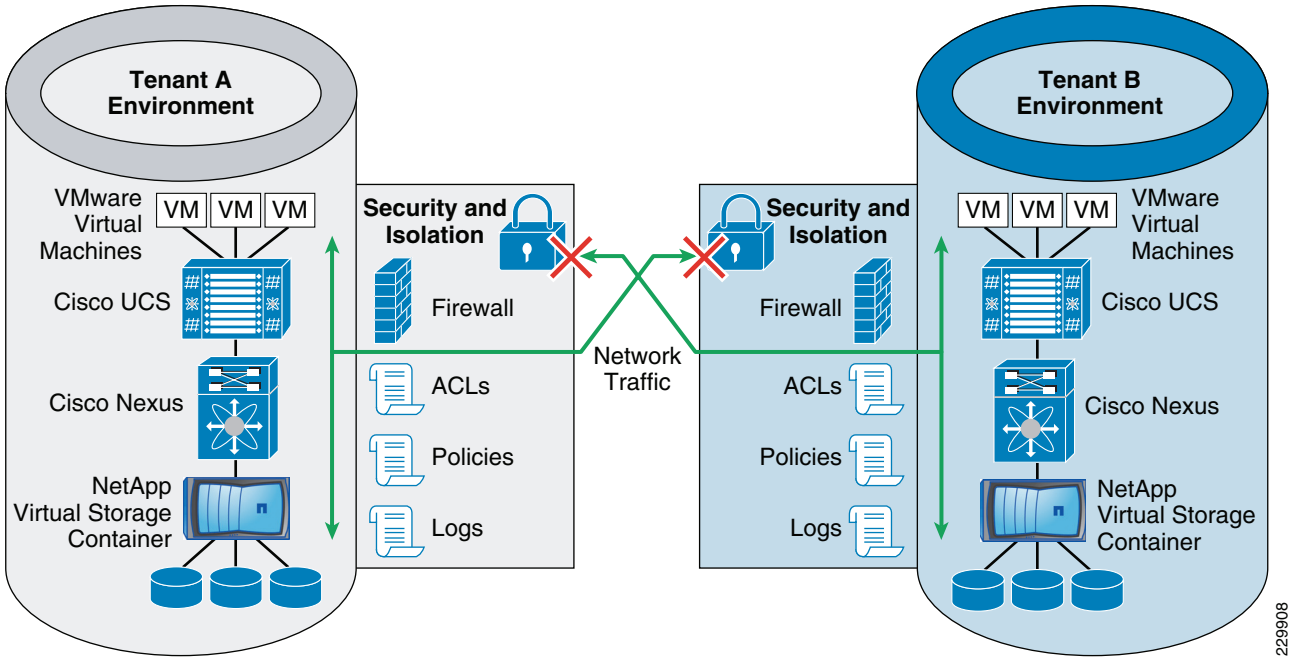
One of the main differences between traditional shared hosting (internal or external) and a typical IaaS cloud service is the level of control available to the user. Traditional hosting services provide users with general application or platform administrative control, whereas IaaS deployments typically provide the user with broader control over the compute resources. The secure cloud architecture further extends user control end-to-end throughout the environment: the compute platform, the network connectivity, storage resources, and data management. This architecture enables service providers and enterprises to securely offer their users unprecedented control over their entire application environment. Unique isolation technologies combined with extensive management flexibility deliver all benefits of cloud computing for IT providers to confidently provide high levels of security and service for multi-tenant customer and consolidated application environments.

Architecture Overview

One of the essential characteristics of a cloud architecture is the ability to pool resources. The provider's compute, network, and storage resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or data center). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

Each tenant subscribed to compute, network, and storage resources in a cloud is entitled to a given SLA. One tenant may have higher SLA requirements than another based on a business model or organizational hierarchy. For example, tenant A may have higher compute and network bandwidth requirements than tenant B, while tenant B may have a higher storage capacity requirement. The main design objective is to ensure that tenants within this environment properly receive their subscribed SLAs while their data, communication, and application environments are securely separated, protected, and isolated from other tenants.

Figure 1 Architecture Overview



Introducing the Four Pillars

The key to developing a robust design is clearly defining the requirements and applying a proven methodology and design principles. The following four requirements were defined as pillars for the Secure Cloud Architecture:

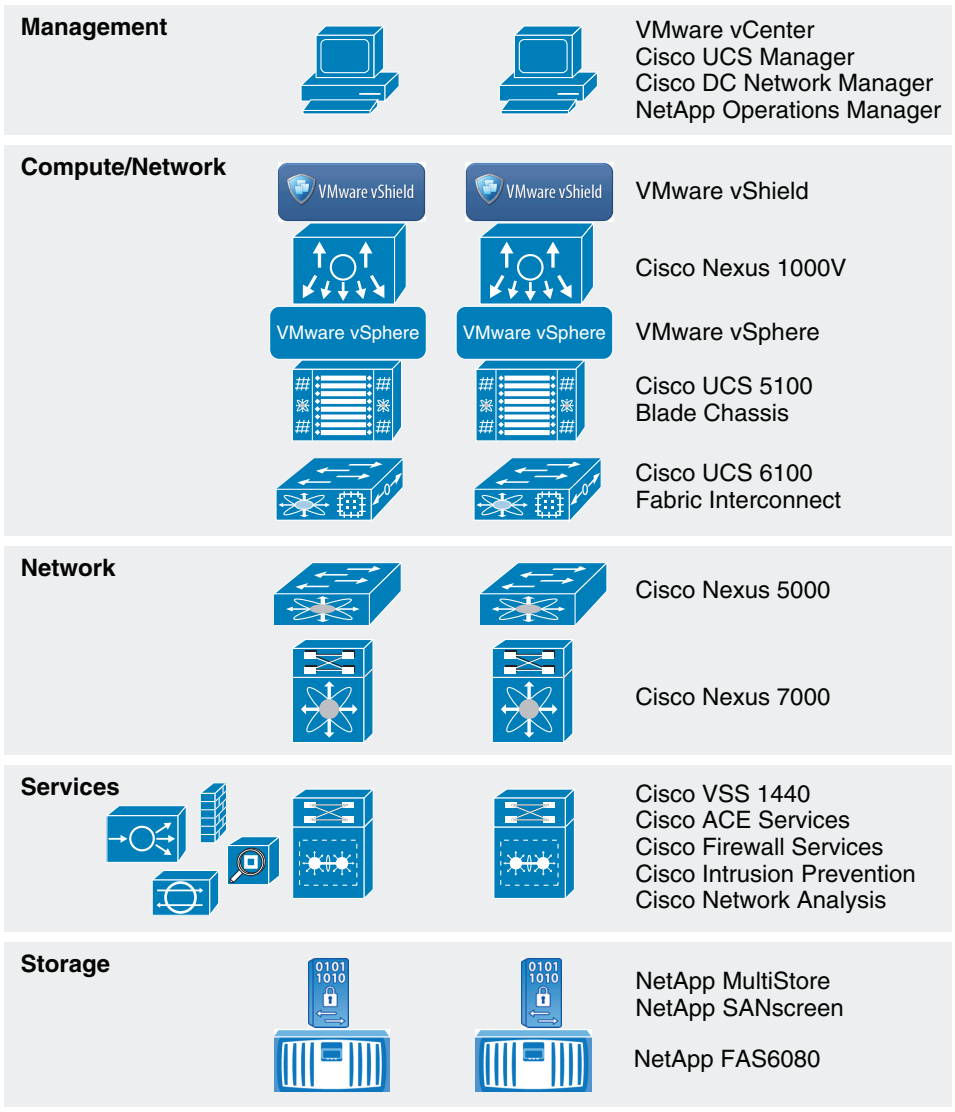
- **Availability** allows the infrastructure to meet the expectation of compute, network, and storage to always be available even in the event of failure. Like the Secure Separation pillar, each layer has its own manner of providing a high availability configuration that works seamlessly with adjacent layers. Security and availability are best deployed from a layered approach.
- **Secure Separation** ensures one tenant cannot disrupt other tenants' resources, such as virtual machine (VM), network bandwidth, tenant data, and storage. It also ensures protection against data loss, denial of service attacks, and unauthorized access. Each tenant must be securely separated using techniques such as access control, virtual storage controllers, VLAN segmentation, firewall rules, and intrusion protection. Secure seperation also implies a defense-in-depth approach with policy enforcement and protection at each layer.
- **Service Assurance** provides isolated compute, network, and storage performance during both steady state and non-steady state. For example, the network and the UCS blade architecture can provide each tenant with a certain bandwidth guarantee using Quality of Service (QoS); resource pools within VMware help balance and guarantee CPU and memory resources, while FlexShare® can balance resource contention across storage volumes.
- **Management** is required to rapidly provision and manage resources and view resource availability. Domain and element management provides comprehensive administration of the shared resources that compose the Secure Cloud architecture. The demarcation point for managing this design is defined by the interactive and programmable interfaces delivered by Cisco, NetApp, and VMware. The administrative interfaces and APIs in this portfolio address infrastructure components such as

vCenter, vCloud Director, UCS Manager, Data Center Network Manager, and the NetApp Manageability Suite. These element managers and their associated open APIs provide the foundation for delivering cohesive service lifecycle orchestration with solution partners.

Architecture Components

The architectural components of the Enhanced Secure Multi-Tenancy design are included to address the availability, secure separation, service assurance, and management requirements of its tenants. These design goals do not vary; however the business, compliance, and application requirements of any one tenant in the enterprise may. To address this reality, the ESMT design is flexible and extensible. The actual ESMT architectural components implemented in the enterprise to meet any one tenant's objectives may differ, but all are built starting with a specific set of foundational components. To clarify for the reader, this section is divided into two categories which identify the foundational components of ESMT and those driven by ESMT tenant requirements.

Figure 2 Architecture Components



Foundational Components

Compute

VMware vSphere and vCenter Server

VMware vSphere and vCenter Server offer the highest levels of availability and responsiveness for all applications and services with VMware vSphere, the industry's most reliable platform for data center virtualization. Optimize IT service delivery and deliver the highest levels of application service agreements with the lowest total cost per application workload by decoupling your business critical applications from the underlying hardware for unprecedented flexibility and reliability.

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, allowing IT administrators dramatically improved control over the virtual environment compared to other management platforms. VMware vCenter Server:

- Provides centralized control and visibility at every level of virtual infrastructure.
- Unlocks the power of vSphere through proactive management.
- Is a scalable and extensible management platform with a broad partner ecosystem.

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

VMware vShield

For organizations that want to leverage the benefits of cloud computing without sacrificing security, control, or compliance, the VMware vShield family of security solutions provides virtualization-aware protection for virtual data centers and cloud environments, enabling customers to strengthen application and data security, improve visibility and control, and accelerate IT compliance efforts across the organization. The vShield family of products includes:

- vShield Edge
- vShield App

vShield Edge

vShield Edge provides network edge security and services to isolate the virtual machines in a port group from the external network. The vShield Edge connects isolated, tenant stub networks to the shared (uplink) networks and provides common perimeter security services such as DHCP, VPN, and NAT. Common deployments of vShield Edge include at the DMZ, VPN Extranets, and multi-tenant Cloud environments where the Edge provides perimeter security for the Virtual Datacenters (VDCs). vShield Edge is compatible with Standard vSwitch, vNetwork Distributed Switch, and the Cisco Nexus 1000V. vShield Edge is also leveraged by VMware vCloud to isolate Organization and vShield App Networks.

vShield App

vShield App provides firewalling capability between virtual machines by placing a firewall filter on every virtual network adapter. The firewall filter operates transparently and does not require network changes or modification of IP addresses to create security zones. Rules can be written using vCenter groupings like Datacenter, Cluster, Resource Pools, and vApps or network objects like Port Group and VLAN to reduce the number of firewall rules and make the rules easier to track.

VMware vShield Zones is a centrally managed, stateful, distributed virtual firewall bundled with vSphere 4.1 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/.

Cisco UCS and UCSM

The Cisco Unified Computing System is a revolutionary new architecture for blade server computing. The Cisco UCS is a next-generation data center platform that unites compute, network, storage access, and virtualization into a cohesive system designed to reduce total cost of ownership (TCO) and increase business agility. The system integrates a low-latency, lossless 10 Gigabit Ethernet unified network fabric with enterprise-class, x86-architecture servers. The system is an integrated, scalable, multi-chassis platform in which all resources participate in a unified management domain. Managed as a single system whether it has one server or 320 servers with thousands of virtual machines, the Cisco UCS decouples scale from complexity. The Cisco UCS accelerates the delivery of new services simply, reliably, and securely through end-to-end provisioning and migration support for both virtualized and non-virtualized systems.

UCS Components

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.

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

Network

Cisco Nexus 7000

As Cisco's flagship switching platform, the Cisco Nexus 7000 Series is a modular switching system designed to deliver 10 Gigabit Ethernet and unified fabric in the data center. This new platform delivers exceptional scalability, continuous operation, and transport flexibility. It is primarily designed for the core and aggregation layers of the data center.

The Cisco Nexus 7000 Platform is powered by Cisco NX-OS (http://www.cisco.com/en/US/products/ps9372/index.html), a state-of-the-art operating system, and was specifically designed with the unique features and capabilities needed in the most mission-critical place in the network, the data center.

For more information, see: http://www.cisco.com/en/US/products/ps9402/index.html.

Cisco Nexus 5000

The Cisco Nexus 5000 Series (http://www.cisco.com/en/US/products/ps9670/index.html), part of the Cisco Nexus Family of data center class switches, delivers an innovative architecture that simplifies data center transformation. These switches deliver high performance, standards-based Ethernet and FCoE that enables the consolidation of LAN, SAN, and cluster network environments onto a single Unified Fabric. Backed by a broad group of industry-leading complementary technology vendors, the Cisco Nexus 5000 Series is designed to meet the challenges of next-generation data centers, including dense multisocket, multicore, virtual machine-optimized deployments, where infrastructure sprawl and increasingly demanding workloads are commonplace.

The Cisco Nexus 5000 Series is built around two custom components: a unified crossbar fabric and a unified port controller application-specific integrated circuit (ASIC). Each Cisco Nexus 5000 Series Switch contains a single unified crossbar fabric ASIC and multiple unified port controllers to support fixed ports and expansion modules within the switch.

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, nonblocking crossbar fabric capable of meshing all ports at wire speed. The unified crossbar fabric offers superior performance by implementing QoS-aware scheduling for unicast and multicast traffic. Moreover, the tight integration of the unified crossbar fabric with the unified port controllers helps ensure 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 switch is a software switch on a server that delivers Cisco VN-Link services to virtual machines hosted on that server. It takes advantage of the VMware vSphere framework to offer tight integration between server and network environments and help ensure consistent, policy-based network capabilities to all servers in the data center. It allows policy to move with a virtual machine during live migration, ensuring persistent network, security, and storage compliance, resulting in improved business continuance, performance management, and security compliance. Last but not least, it aligns management of the operational environment for virtual machines and physical server connectivity in the data center, reducing the total cost of ownership (TCO) by providing operational consistency and visibility throughout the network. It offers flexible collaboration between the server, network, security, and storage teams while supporting various organizational boundaries and individual team autonomy.

The Cisco Nexus 1000V supports VMware's vCloud Director. The vCloud director has three layers of networking available: provider, organizational, and vShield App. The Nexus 1000V supports all three of these network types via portgroup-backed network pools and their associated VLANs. The combination of vCloud and Nexus 1000V allows enterprises to offer self-service isolated network provisioning for multi-tenant environments.

Note The Nexus 1000V as of the release date of this document does not support VCNI, which is a new VMware technology available in vCloud director.

In addition to the security features offered by the Nexus 1000V, the Cisco virtual distributed switch supports VMware's vShield Edge technology. To achieve Enhanced Secure Multi-Tenancy, it is important to carve off one or more isolated Layer 2 adjacent segments on the Nexus 1000V for each tenant. These VLAN segments are further secured at the perimeter via vShield Edge which allows certain centralized services such as DNS or AD to be readily consumed by tenant virtual machines within the data center.

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

The Nexus 1010 Virtual Services Appliance hosts the Cisco Nexus 1000V Virtual Supervisor Module (VSM) and supports the Cisco Nexus 1000V Network Analysis Module (NAM) Virtual Service Blade to provide a comprehensive solution for virtual access switching. The Cisco Nexus 1010 provides dedicated hardware for the VSM, making the virtual access switch deployment much easier for the network administrator.

For more information on Nexus 1000V, see:
http://www.cisco.com/en/US/partner/products/ps10785/index.html.

For more information on Cisco VN-Link technologies see:
http://www.cisco.com/en/US/netsol/ns894/index.html.

Cisco Data Center Network Manager (DCNM)

DCNM is a management solution that maximizes overall data center infrastructure uptime and reliability, which improves business continuity. Focused on the management requirements of the data center network, DCNM provides a robust framework and rich feature set that fulfills the switching needs of present and future data centers. In particular, DCNM automates the provisioning process.

DCNM is a solution designed for Cisco NX-OS-enabled hardware platforms. Cisco NX-OS provides the foundation for the Cisco Nexus product family, including the Cisco Nexus 7000 Series.

For more information, see:
http://www.cisco.com/en/US/docs/switches/datacenter/sw/4_1/dcnm/fundamentals/configuration/guide/fund_overview.html.

Cisco Fabric Manager (FM)

Fabric Manager is a management solution for storage networks across all Cisco SANs and unified fabric. FM provides a robust centralized management station for SAN and unified fabric-enabled devices such as the MDS family of switches and the Nexus 5000. FM reduces TCO by being able to perform the tasks needed during a device's deployment cycle such as discovery, inventory, configuration, performance monitoring, and troubleshooting.

FM is designed for management of the MDS family of switches, the Nexus 5000 SAN features, and the UCS Fabric Interconnect with limited support.

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.

Storage

NetApp FAS Unified Storage

Each NetApp fabric-attached storage (FAS) controller shares a unified storage architecture based on the Data ONTAP 7G operating system (OS) and uses an integrated suite of application-aware manageability software. This provides an efficient consolidation of storage area network (SAN), network-attached storage (NAS), primary storage, and secondary storage on a single platform while allowing concurrent support for block and file protocols using Ethernet and Fibre Channel interfaces. These interfaces include Fibre Channel over Ethernet (FCoE), Network File System (NFS), Common Internet File System

protocol (CIFS), and iSCSI. This common architecture allows businesses to start with an entry-level storage platform and easily migrate to the higher-end platforms as storage requirements increase, without learning a new OS, management tools, or provisioning processes.

To provide resilient system operation and high data availability, Data ONTAP 7G is tightly integrated into the hardware systems. The FAS systems use redundant, hot-swappable components. Combined with the patented dual-parity RAID-DP® (high-performance RAID 6), the net result can be superior data protection with little or no performance loss. For a higher level of data availability, Data ONTAP provides optional mirroring, backup, and disaster recovery solutions. For more information, refer to: http://www.netapp.com/us/products/platform-os/data-ontap/.

NetApp Snapshot technology provides the added benefit of near-instantaneous file-level or full data set recovery, while using a very small amount of storage. The Snapshot technology creates up to 255 data-in-place, point-in-time images per volume. For more information, refer to: http://www.netapp.com/us/products/platform-os/snapshot.html.

Important applications require a quick response time, even during times of heavy loading. The FlexShare quality-of-service software is included as part of the Data ONTAP operating system to enable a fast response time when serving data for multiple applications. FlexShare allows storage administrators to set and dynamically adjust workload priorities. For more information, refer to: http://www.netapp.com/us/products/platform-os/flexshare.html.

While this solution focuses on specific hardware, including the FAS6080 and FAS3170, any of the FAS platforms, including the FAS6040 and FAS3140, are supported based on your sizing requirements and expansion needs with all of the same software functionality and features. Similarly, the quantity, size, and type of disks used within this environment may also vary depending on storage and performance needs. Additional add-on cards, such as the Flash Cache Modules (PAM II), can be used in this architecture to increase performance by adding additional system cache for fast data access. However, they are not required for functionality. For more information, refer to: http://www.netapp.com/us/products.

Ethernet Storage

Ethernet storage using NFS is one of the key technologies in this architecture. This technology is leveraged to provide tremendous efficiency and functional gains. Some of the key benefits of Ethernet-based storage are:

- Reduced hardware costs for implementation
- Reduced training costs for support personnel
- Simplified infrastructure supported by internal IT groups

The initial solution is to deploy a clustered pair of enterprise class NetApp storage controllers onto a dedicated virtual Ethernet storage network, which is hosted by a pair of core IP Cisco switches and an expandable number of edge switches. The virtual Ethernet storage network also extends to each host server through two fabric interconnects enabling direct IP storage access from within the compute layer. For more information, refer to: http://www.netapp.com/us/company/leadership/ethernet-storage/.

In addition, NetApp now supports the Cisco Discovery Protocol (CDP) which enables greater visibility into the Ethernet network from the storage perspective. CDP provides the storage administrator with information regarding the name of each Cisco switch that is attached and also specifically on which port or ports the storage system is attached. This feature helps to simplify initial Ethernet configuration as well as troubleshooting issues should they arise.

Stateless Computing Using Fibre Channel Boot Over Ethernet

Fibre Channel over Ethernet (FCoE) is an encapsulation of Fibre Channel frames transported over Ethernet networks with certain provisions to make sure that Fibre Channel standards are followed. FCoE architectures significantly reduce management complexity, required cable count, number of mezzanine cards for fabric segmentation, and power and cooling costs. FCoE is an inherent design element in the

Enhanced Secure Multi-Tenancy architecture because it is a core component available within the UCS in the form of virtualized server adapters and unified storage adapters on the NetApp controller. Fibre Channel and IP traffic from the UCS fabric interconnects are passed northbound to a Storage Protocols License enabled Nexus 5000 with attached NetApp IP and FCoE storage interfaces.

The deployment of an architecture consisting of FCoE booted physical resources provides great flexibility and resiliency to a multi-tenant infrastructure. An FCoE booted deployment consists of hosts in the environment with converged network adapters (CNAs) capable of translating SCSI commands. UCS hosts access their boot OS by using logical unit numbers (LUNs) on the NetApp unified storage controller.

FCoE booted hosts that use NetApp controllers have superior RAID protection and increased performance compared to traditional local disk arrays. Furthermore, FCoE booted resources can easily be recovered, are better utilized, and scale much faster than local disk installs. Operating systems and hypervisors provisioned by using NetApp controllers take advantage of storage efficiencies inherent in NetApp products. Another major benefit of FCoE booted architectures is that they can be deployed and recovered in minutes dependent on the operating system to be installed.

FCoE booted deployments effectively reduce provisioning time, increase utilization, and aid in the stateless nature of service profiles within UCS. An FCoE booted environment can be preconfigured and, through the use of NetApp technologies, can perform better, have greater data protection, and be easier to restore.

NetApp FilerView

NetApp FilerView® is the primary, element-level graphical management interface available on every NetApp storage system. FilerView is an intuitive, browser-based tool. This tool can be used to monitor and manage administrative tasks on individual NetApp storage systems. In addition to extensive configuration control over storage services, providers can leverage FilerView to assess storage resource capacity and utilization for the following:

- Physical disk aggregates
- FlexVol® logical volumes
- Quotas
- Block storage allocation for SAN attachments
- NFS/CIFS implementation for NAS attachments

FilerView provides control over administrative and user access to the NetApp storage system. Storage providers can use FilerView to inspect the health and status of NetApp storage systems and to configure notification and alerting services for resource monitoring.

NetApp MultiStore

NetApp MultiStore allows cloud providers to quickly and easily create separate and completely private logical partitions on a single NetApp storage system as discrete administrative domains called vFiler™ units. These vFiler units have the effect of making a single physical storage controller appear to be many logical controllers. Each vFiler unit can be individually managed with different sets of performance and policy characteristics. Providers can leverage MultiStore to enable multiple customers to share the same storage resources with minimal compromise in privacy or security. Administrative control of the virtual storage container can even be delegated directly to the customer. Up to 130 vFiler units can be created on most NetApp high-availability (HA) pairs using MultiStore technology. For more information, refer to: http://www.netapp.com/us/products/platform-os/multistore.html.

Business Driven Components

Compute

VMware vCloud Director

VMware vCloud Director (VCD) builds upon the VMware vSphere foundation and exposes virtualized shared infrastructure as multi-tenant virtual data centers that are completely decoupled from the underlying hardware. VMware vCloud Director also allows IT to expose virtual data centers to users through a Web-based portal and to define and expose a catalog of IT services that can be deployed within the virtual data center.

What this means is that using VMware technology, you can now deliver standardized IT services on shared infrastructure through a Web-based catalog. By standardizing service offerings, you can simplify many IT management tasks—from troubleshooting and patching to change management—and eliminate much of the administrative maintenance that burden your IT team today. You can also automate provisioning through policy-based workflows that empower validated users to deploy preconfigured services with the click of a button, translating into new opportunities for end users to procure IT resources exactly when they need them.

By standardizing processes, increasing automation and delivering IT as a service, you will drive additional savings beyond virtualization, while significantly reducing the amount of maintenance required per IT administrator.

A private cloud built on VMware technology will enable your IT department to transform itself into an efficient, agile, and user-friendly internal service provider. You can then deliver on the promise of IT as a Service, providing fully automated, catalog-based services to internal users through a Web-based portal.

VMware vCloud Director brings a new set of terminologies that are important for understanding the design principles used throughout this document.

- vSphere resources
  - vSphere resources are the vCenter™ Servers, ESXi hosts, resource pools, datastores, vNetwork Distributed Switches, and port groups that are used to provision cloud resources in VCD.
- Cloud abstraction layer
  - vCloud Director cells—Cloud cells are the Red Hat Enterprise Linux® 5 (RHEL5) servers that run the VCD software. Multiple cloud cells form the VCD cluster.
  - Provider Virual Data Center (vDC)—A provider vDC is a group of compute, memory, and storage resources from one vCenter. You can allocate portions of a provider vDC to your organizations using VCD.
  - External network—An external network uses a network in vSphere to connect to a network outside of your cloud. The network can be a public network such as the Internet or an external VPN network that connects to a given organization.
  - Organization—An organization is the fundamental grouping in VCD. An organization contains users, the vApps they create, and the resources the vApps use. An organization can be a department in your own company or an external customer to which you are providing cloud resources.
  - Organization vDC—An organization vDC provides an organization with the compute, memory, storage, and network resources required to create vApps.

- Network pool—A network pool is a collection of VM networks that are available to be consumed by vDCs to create vShield App networks and by organizations to create organization networks. Network traffic on each network in a pool is isolated at Layer 2 from all other networks.
- vApp—A vShield App is a virtual application that contains one or more VMs.
- Catalog—A catalog allows you to share vApp templates and media images with other users in your organization or with other organizations in VCD.

VMware vCenter Chargeback

vCenter Chargeback is an end-to-end cost reporting solution for virtual environments using vSphere. This Web-based application interacts with the vCenter Database to retrieve usage information, calculates the cost by using the defined chargeback formulas, and generates reports. Starting with version 1.5, vCenter Chargeback is integrated with VMware Cloud Director and vShield for resource usage statistics collection and reporting for the new abstraction layer.

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. SRM 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.

Network

Cisco Catalyst 6500 Virtual Switching System 1440

The Cisco Catalyst 6500 Series Virtual Switching System (VSS) 1440 allows for the merging of two physical Cisco Catalyst 6500 Series Switches together 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 Catalyst 6500 switches.

The VSS platform fully supports the use of Cisco integrated service modules such as the Cisco Application Control Engine (ACE), Firewall Services Module, and Network Analysis Module. In addition, the VSS platform is capable of supporting both gigabit and ten gigabit Ethernet devices allowing for network based services via a variety of appliance form factors.

Note The Catalyst VSS is used as a network-based services platform for the enhanced secure multi-tenant architecture.

Cisco Firewall Services

The Cisco Firewall Services Module (FWSM) is a stateful firewall residing within a Catalyst 6500 switching platform. The integrated module employs the power, cooling and space available in the chassis to provide data center security services. The FWSM module offers device level redundancy and scalability through multiple virtual security contexts. Each virtual security context may be transparently introduced at the Layer 2 network level or as a router “hop” 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 and application platforms perform server load balancing, network traffic control, service redundancy, resource management, encryption and security, and application acceleration and optimization, all in a single network device. The Cisco ACE technologies provide 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 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 Intrusion Prevention System (IPS) Appliances

The Cisco Intrusion Prevention System (IPS) appliances are network sensors that may be positioned throughout the data center as promiscuous network analysis devices or inline intrusion prevention systems. The Cisco IPS sensors protect the data center by detecting, classifying, and blocking network based threats via attack signatures associated with worms, viruses, and various application abuse scenarios. This process occurs on a per connection basis allowing legitimate traffic to flow unobstructed.

The Cisco IPS appliances support logical partitioning, allowing one physical sensor to become multiple virtual sensors. In this configuration, the virtual sensors may be deployed in any combination of promiscuous or inline modes. Each sensor, virtual or physical, may be finely tuned to inspect the application traffic pertinent to its network locale.

For more information, see:
http://www.cisco.com/en/US/products/hw/modules/ps2706/ps4452/index.html.

Cisco Network Analysis Module (NAM) Products

The Cisco Network Analysis Modules (NAM) comes in several form factors including:

- Integrated service module for the Catalyst 6500 switching platform
- Physical Appliance with multiple Gigabit or 10 Gigabit Ethernet support
- Virtual Service Blade for Cisco Nexus 1000v deployments

Regardless of the model, the NAM offers flow-based traffic analysis of applications, hosts, and conversations, performance-based measurements on application, server, and network latency, quality of experience metrics for network-based services and problem analysis using deep, insightful packet captures. The Cisco NAM includes an embedded, Web-based Traffic Analyzer GUI that provides quick access to the configuration menus and presents easy-to-read performance reports on Web for different types of services and traffic. The Cisco NAM line of products improves visibility into and monitors the performance of the many physical and virtual layers within the data center.

For more information, see:
http://www.cisco.com/en/US/products/ps5740/Products_Sub_Category_Home.html.

Cisco Application Network Manager (ANM)

The Cisco Application Network Manager is a client server application allowing administrators to provision, monitor, and maintain application network services in the data center. Employing RBAC, the Cisco ANM allows application owners or server administrators to create ACE-enforced application

policies within the network without impacting network configurations. In addition to supporting the Cisco ACE line of products, the Cisco ANM communicates with VMware's vCenter to allow seamless workflows across the virtualized data center as application polices and environments are deployed and evolve.

For more information, see: http://www.cisco.com/en/US/products/ps6904/index.html.

Cisco Security Manager

Cisco Security Manager is an enterprise-class management application designed to configure firewall, VPN, and intrusion prevention system (IPS) security services on Cisco network and security devices. Cisco Security Manager can be used in networks of all sizes—from small networks to large networks consisting of thousands of devices—by using policy-based management techniques.

Security Manager offers the following features and capabilities:

- Service-level and device-level provisioning of VPN, firewall, and intrusion prevention systems from one desktop
- Device configuration rollback
- Network visualization in the form of topology maps
- Workflow mode
- Predefined and user-defined service templates
- Integrated inventory, credentials, grouping, and shared policy objects
- Integrated monitoring of events generated by ASA and IPS devices—You can selectively monitor, view, and examine events from ASA and IPS devices by using the Event Viewer feature, introduced in Security Manager 4.0.

Storage

NetApp Virtual Storage Console (VSC)

Virtual Storage Console (VSC) provides integrated, comprehensive storage management for VMware infrastructures, including discovery, health monitoring, capacity management, provisioning, cloning, backup, restore, and disaster recovery. VMware administrators can access and execute all of these capabilities directly from VMware vCenter, enhancing both server and storage efficiencies without affecting the policies created by storage administrators. VSC operations may also be automated by using Java or XML APIs.

NetApp SnapManager for Virtual Infrastructure (SMVI)

NetApp SnapManager for Virtual Infrastructure (SMVI) works with VMware vCenter, automating and simplifying management of backup and restore operations. Administrators can now leverage an easy-to-use management tool to create application-consistent backups for their virtual machines. Additionally, they can instantly recover a datastore, VM, vmdk, or an individual file within a VM guest.

NetApp SnapDrive

NetApp SnapDrive enables the automation of storage-provisioning tasks and simplifies the process of taking error-free, host-consistent Snapshot copies of data. Server administrators use a wizard-based approach to map, manage, and migrate data between new and existing storage resources. SnapDrive can also eliminate the need to preallocate storage resources based on forecasted demand.

NetApp SnapManager Products for Microsoft Applications

SnapManager provides an integrated data management solution that enhances the availability, scalability, and reliability of application data. SnapManager provides rapid online backup and restoration of databases and application data, along with local or remote backup set mirroring for disaster recovery. SnapManager uses online Snapshot technology that is part of Data ONTAP and integrates application backup and restore APIs and Volume Shadow Copy Service (VSS). SnapManager provides the following data management capabilities:

- Migrating application databases and transaction logs to LUNs on storage systems
- Backing up application databases and transaction logs from LUNs on storage systems
- Verifying the backed-up application databases and transaction logs
- Managing backup sets
- Archiving backup sets
- Restoring application databases and transaction logs from previously created backup sets

NetApp SnapMirror

NetApp SnapMirror® is a Data ONTAP feature that provides data replication between two NetApp storage controllers or vFiler units. SnapMirror is typically deployed in disaster recovery scenarios in which business-critical data must be replicated offsite to protect against data loss and corruption if a failure or catastrophic event occurs at the primary site. Should such an event occur, the replicated data at the DR site can quickly and easily become writable, reducing RTO and ensuring business continuity.

NetApp Adapter for Site Recovery Manager (SRM)

The NetApp Disaster Recovery Adapter for Site Recovery Manager (SRM) is a storage-vendor-specific plug-in to VMware's SRM. This plug-in enables interaction between the SRM and the storage controller. The adapter interacts with the storage controller on behalf of the SRM to discover storage arrays and replicated datastores and to fail over or perform a failover test against the replicated datastores on which the virtual machines reside.

NetApp Data Motion

NetApp Data Motion software lets you easily and quickly migrate data across multiple storage systems while maintaining continuous user and client access to applications. Fully integrated with the Data ONTAP software platform, Data Motion integrates three proven NetApp software technologies—MultiStore, SnapMirror, and Provisioning Manager—to provide live data migration for your shared storage infrastructure. The result is that you can manage your physical or virtualized data center environment nondisruptively.

NetApp Provisioning Manager

NetApp Provisioning Manager allows service providers to streamline the deployment of cloud infrastructure and the delivery of tenant storage resources according to established policies. Provisioning Manager enables the cloud administrator to:

- Automate deployment of storage supporting the cloud compute infrastructure, the vFiler units, and the storage delivered to tenant environments.
- Make sure storage deployments conform to provisioning policies defined by the administrators or tenant SLAs.
- Provision multiprotocol storage with secure separation between tenant environments.

- Automate deduplication and thin provisioning of storage.
- Simplify data migration across the cloud storage infrastructure.
- Delegate control to tenant administrators.

Through the NetApp Management Console, Provisioning Manager delivers dashboard views that display a variety of metrics. These metrics can be leveraged to craft policies that further increase resource utilization, operational efficiency, and to make sure that storage provisions satisfy the desired levels of capacity, availability, and security. Provisioning policies can be defined within the context of resource pools that are aligned with administrative or tenant requirements. Cloud administrators can delegate Provisioning Manager access and control to tenant administrators within the confines of their separated storage environment, directly extending many of these benefits to their customers.

NetApp Protection Manager

Using NetApp Protection Manager, cloud and tenant administrators can group data with similar protection requirements and apply preset policies to automate data protection processes. Administrators can easily apply consistent data protection policies across the cloud storage infrastructure and within tenant environments designed to suit operational and service-level requirements. Protection Manager automatically correlates logical data sets and the underlying physical storage resources. This enables administrators to design and apply policies according to business-level or service-level requirements, alleviated from the details of the cloud storage infrastructure. Within the confines of established policies, secondary storage is dynamically allocated as primary storage grows. Protection Manager is integrated within the NetApp Management Console, providing a centralized facility for monitoring and managing all data protection operations. In addition, it allows cloud providers to appropriately grant control to tenant administrators. The integration of Provisioning Manager and Protection Manager within a single console allows cloud and tenant administrators to seamlessly provision and protect data through unified, policy-driven workflows.

NetApp Operations Manager

NetApp Operations Manager delivers centralized management, monitoring, and reporting tools. These tools enable cloud service providers to consolidate and streamline management of their NetApp storage infrastructure. With Operations Manager, cloud administrators can reduce costs by leveraging comprehensive dashboard views to optimize storage utilization and minimize the IT resources needed to manage their shared storage infrastructure. At the same time, administrators can improve the availability and quality of services delivered to their tenant customers. Cloud administrators can use Operations Manager to establish thresholds and alerts to monitor key indicators of storage system performance, enabling them to detect potential bottlenecks and manage resources proactively. Through the use of configuration templates and policy controls, Operations Manager enables administrators to achieve standardization and policy-based configuration management across their cloud storage infrastructure. This accelerates tenant deployments and mitigates operational risks. Operations Manager gives cloud providers comprehensive visibility into their storage infrastructure. Providers can continuously monitor storage resources, analyze utilization and capacity management, and gain insight into the growth trends and resource impact of their tenants. NetApp Operations Manager also addresses the business requirements of multi-tenant service providers, enabling chargeback accounting through customized reporting and workflow process interfaces.

NetApp SANscreen

NetApp SANscreen® enables cloud administrators to further improve the quality and efficiency of their service delivery with real-time, multiprotocol, service-level views of their cloud storage infrastructure. SANscreen is a suite of integrated products that delivers global, end-to-end visibility into the cloud service provider's entire networked storage infrastructure. SANscreen Service Insight offers providers a

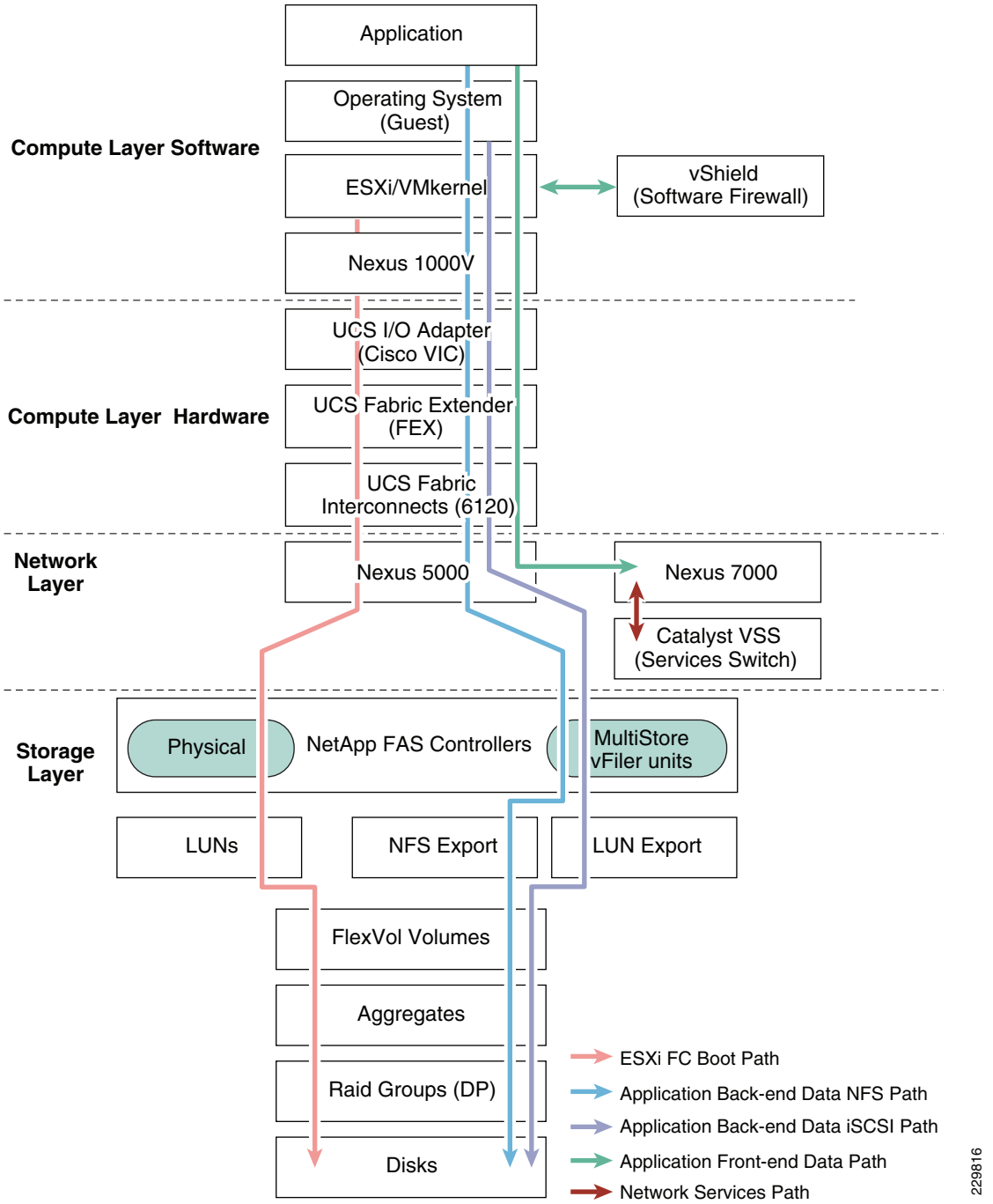
comprehensive view of their SAN and NAS environments, storage attachment paths, storage availability, and change management to closely monitor service level delivery to their customers. Service Insight provides the baseline framework for the SANscreen product suite and gives cloud providers a central repository for their inventory information. In addition, it provides reporting facilities that can be integrated into the provider's existing resource management systems and business processes for financial accounting and asset management. SANscreen Service Assurance applies policy-based management to the provider's networked storage infrastructure. This enables the cloud administrator to flexibly define best-practice policies to enforce storage network performance and availability requirements for each tenant environment. SANscreen Application Insight allows cloud service providers to discover near real-time performance data from their networked storage environment and map it to their tenant deployments. Administrators can then proactively load balance storage networks and systems to ensure customer service levels. SANscreen Capacity Manager provides real-time visibility into global storage resource allocations and a flexible report authoring solution. This product delivers decision support to the cloud service provider's capacity planning, storage tier analysis, storage service catalogs, trending and historical usage, audit, chargeback, and other business processes.

SANscreen VM Insight extends the administrator comprehensive networked storage visibility into the realm of virtual servers. The service path relationships between VMs and the networked storage infrastructure are correlated to enable the wealth of NetApp SANscreen service-oriented management for VM environments. Administrators can access SANscreen data from a unified console and also through VMware vCenter plug-in interfaces. From the virtual server environments to the shared storage allocations that resource a hosted tenant deployment, NetApp SANscreen delivers end-to-end visibility, flexible and proactive management, and service-level assurance for multi-tenant cloud service providers.

End-to-End Block Diagram

Understanding the flow from application to storage is key in building an Enhanced Secure Multi-Tenancy environment. Figure 3 provides an end-to-end path, such as ESXi FCoE boot starting from ESXi VMkernel at the compute layer to network layer to storage layer. In this model, there is a single Fiber Channel hop between the Cisco UCS 6100 and Nexus 5000 fabric.

Figure 3 End-to-End Block Diagram



Logical Topology

The tenant is the primary building block of the Secure Multi Tenant architecture. Therefore, the definition of a tenant is fundamental to the successful design and deployment of the ESMT architecture in any given enterprise. By definition a tenant is an occupant. In terms of this design the tenant resides

in or owns part of the shared data center infrastructure. Each enterprise defines a tenant by their own terms; there are no firm rules on what a tenant is or is not. Typically the enterprise defines the tenant along lines of business or organizational units defined by corporate structure, but one could use functional roles or individual application environments to construct a tenant.

Note In this design the tenant is a single application or workload within the enterprise.

The logical topology represents the underlying virtual components and their virtual connections that exist within the physical topology.

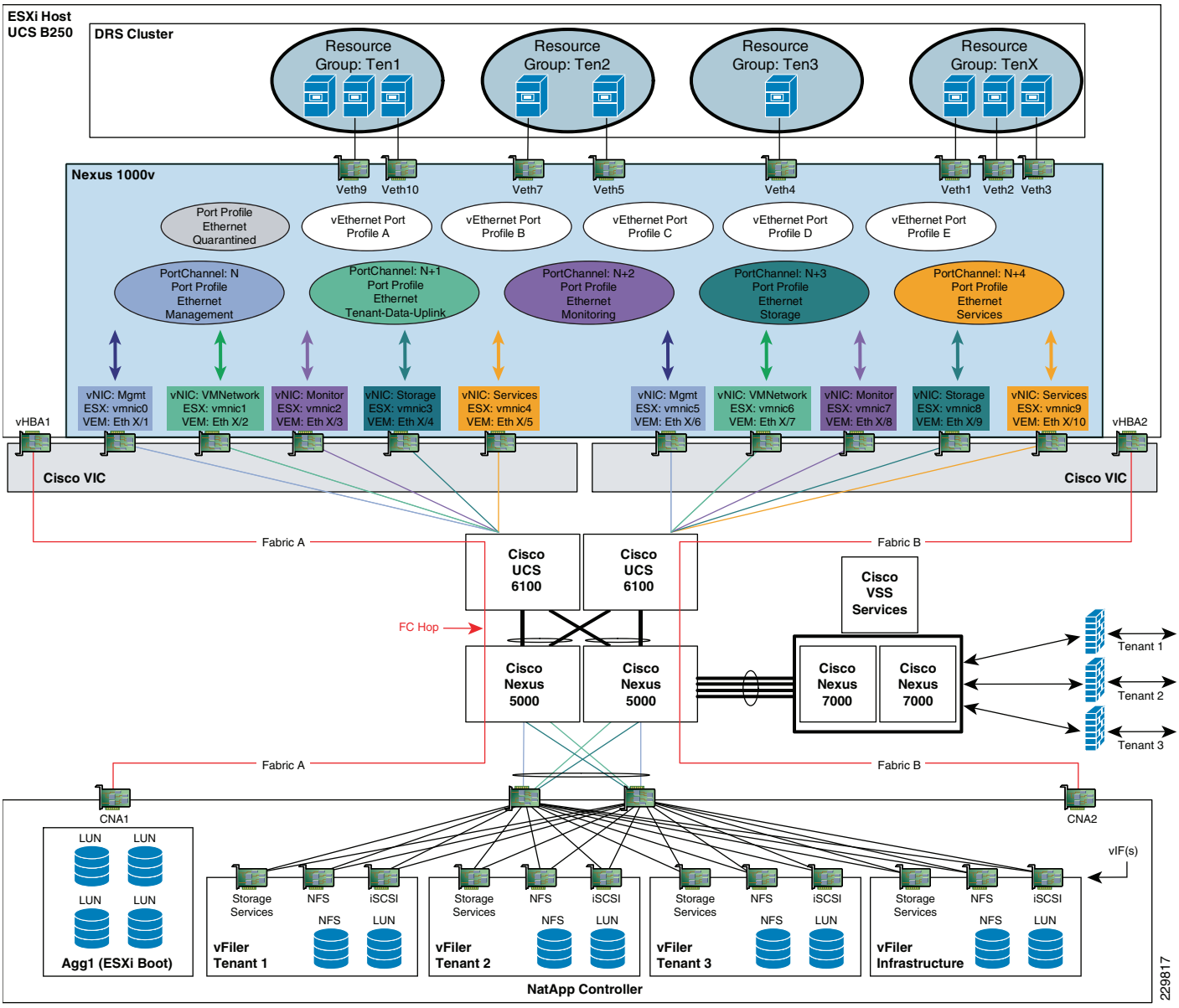
The logical architecture consists of many partitions falling into one of two categories, infrastructure and tenant. Infrastructure VMs are used in configuring and maintaining the environment, while tenant VMs are owned and leveraged by tenant applications and users. VM configuration and disk files for both infrastructure and tenant VMs are stored in distinct NetApp virtual storage controllers and are presented to each ESXi host's appropriate VMkernel interfaces as an NFS export or iSCSI target.

Each VMware virtual interface type, VMkernel, and individual VM interfaces connect directly to the Cisco Nexus 1000V software distributed virtual switch. At this layer, packets are tagged with the appropriate VLAN header and all outbound traffic is aggregated to the two Cisco 6100 Fabric Interconnects via the UCS 2104 fabric extenders in each chassis. The UCS B250 blade servers contain a pair of Cisco Virtual Interface Cards (VIC) Ethernet uplinks. Each Cisco VIC presents five virtual interfaces to the ESXi host, known as UCS vNICs, which allows for further traffic segmentation and categorization across all traffic types based on vNIC network policies. Using port aggregation between the Fabric "A" and "B" vNIC pairs enhances the availability and capacity of each traffic category. All inbound traffic is stripped of its VLAN header and switched to the appropriate destination virtual Ethernet interface. In addition, the Cisco VIC allows for the creation of multiple virtual Host Bus Adapters (vHBAs), permitting FC-enabled boot across the same physical infrastructure.

The two physical 10Gb Ethernet interfaces per physical NetApp storage controller are aggregated together into a single virtual interface. The virtual interface is further segmented into VLAN interfaces, with each VLAN interface corresponding to a specific VLAN ID throughout the topology. Each VLAN interface is administratively associated with a specific IP Space and vFiler unit. Each IP Space provides an individual IP routing table per vFiler unit. The association between a VLAN interface and a vFiler unit allows all outbound packets from the specific vFiler unit to be tagged with the appropriate VLAN ID specific to that VLAN interface. Accordingly, all inbound traffic with a specific VLAN ID is sent to the appropriate VLAN interface, effectively securing storage traffic, no matter what the Ethernet storage protocol, and allowing visibility to only the associated vFiler unit.

Note Infrastructure VMs may reside on a dedicated ESXi cluster (not shown in Figure 4) or share the same resources as the other tenant VMs. In either case, the partitioning of network, storage, and compute remains constant. Infrastructure VMs are considered a tenant.

Figure 4 Logical Topology Production Cluster



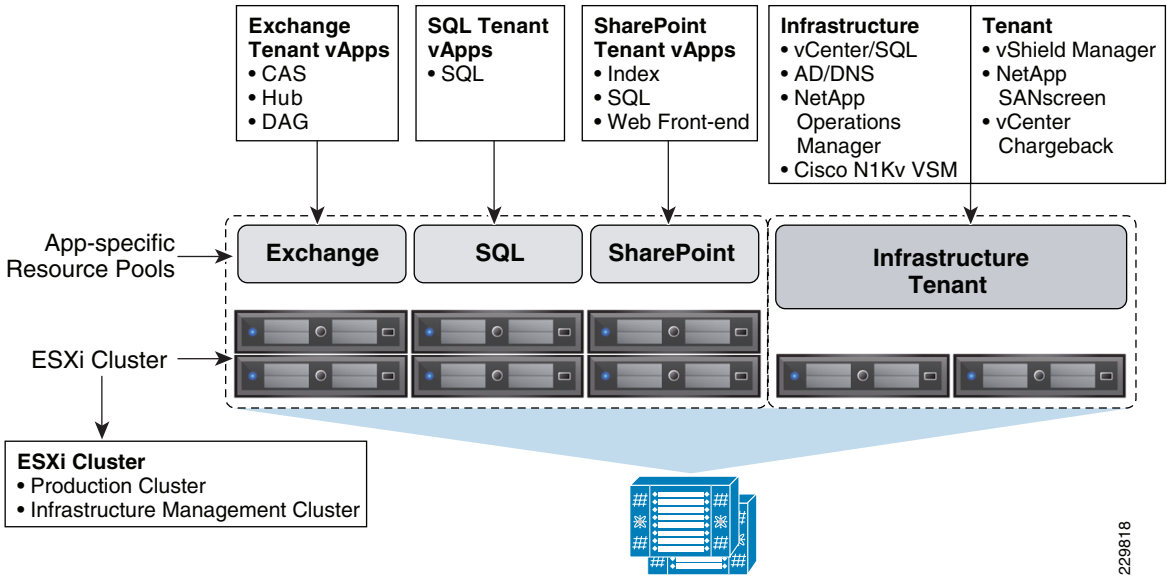
Primary Site Design

Production environments are designed to support the current generation of business critical applications leveraged by tenants within the enterprise. These essential applications may include Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Supply Chain Management (SCM), and sales force tracking. Production applications require varying levels of data center services to meet the security, availability, and scalability requirements of the business. The ESMT production environment topology addresses these objectives by providing each tenant the flexibility and control to meet their particular application needs. It is important to note that it is the application and the business

requirements of the application which drive the implementation of services. The production environment consists of shared compute, network, and storage resources which are virtualized to promote enterprise efficiency and agility.

Figure 5 depicts the primary data center VMware DRS cluster configuration. As shown, there are two dedicated clusters, Production and Infrastructure. The Production cluster supports three tenants, each with an independent enterprise application, namely Exchange 2010, SQL Server 2008, and SharePoint 2010. The Infrastructure cluster houses the Infrastructure tenant which contains virtual machines required to manage the ESMT architecture. Although the Infrastructure tenant could be deployed on the Production cluster, it is still considered a best practice to employ dedicated physical resources for operational efficiency, troubleshooting, and future growth of the management environment. The Production and Infrastructure tenant environments are detailed in the following section.

Figure 5 Primary Data Center Cluster Design



Cisco UCS supports multi-tenancy by allowing administrators to carve up the systems physical infrastructure by logical entities known as organizations. Organizations are logically isolated in the UCS fabric. UCS hardware and policies can be assigned to different organizations such that the desired customer or line of business has access to the right compute blade for a given workload. This rich set of policies in UCS can be applied per organization to ensure the right sets of attributes and I/O policies are assigned to the correct organization. Each organization can have their own pools of resources including:

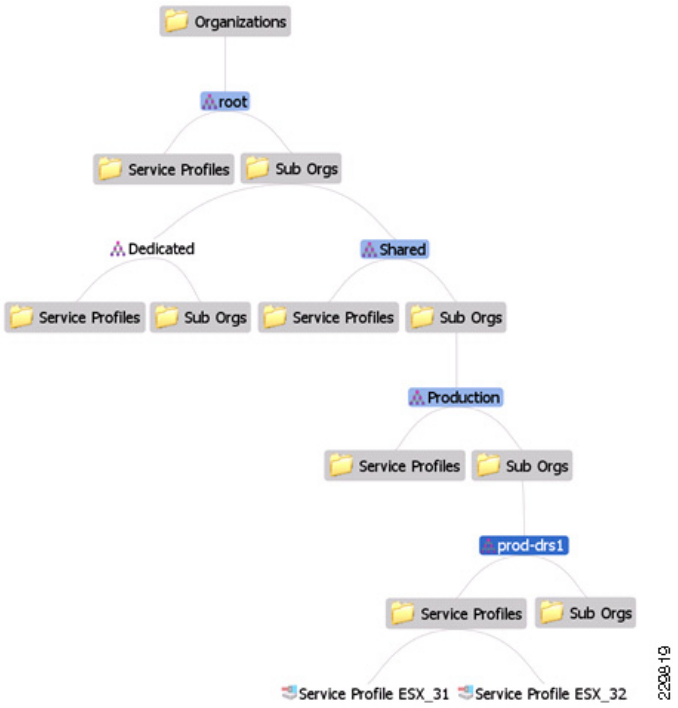
- Resource pools (Server, MAC, UUID, WWPN, etc.)
- Policies
- Service profiles
- Service profile templates

All organizations are hierarchical. The top-level organization is always root. The policies and pools that you create in root are system-wide and are available to all organizations in the system. However, any policies and pools created in other organizations are only available to organizations that are above it in the same hierarchy. For example, if a system has organizations named Finance and HR that are not in the same hierarchy, Finance cannot use any policies in the HR organization and HR cannot access any policies in the Finance organization. However, both Finance and HR can use policies and pools in the

root organization. Combine this functionality with the use of RBAC and UCS locales to assign or restrict user privileges and roles by organization, and the functional isolation provided by UCS is ideal for a multi-tenant environment.

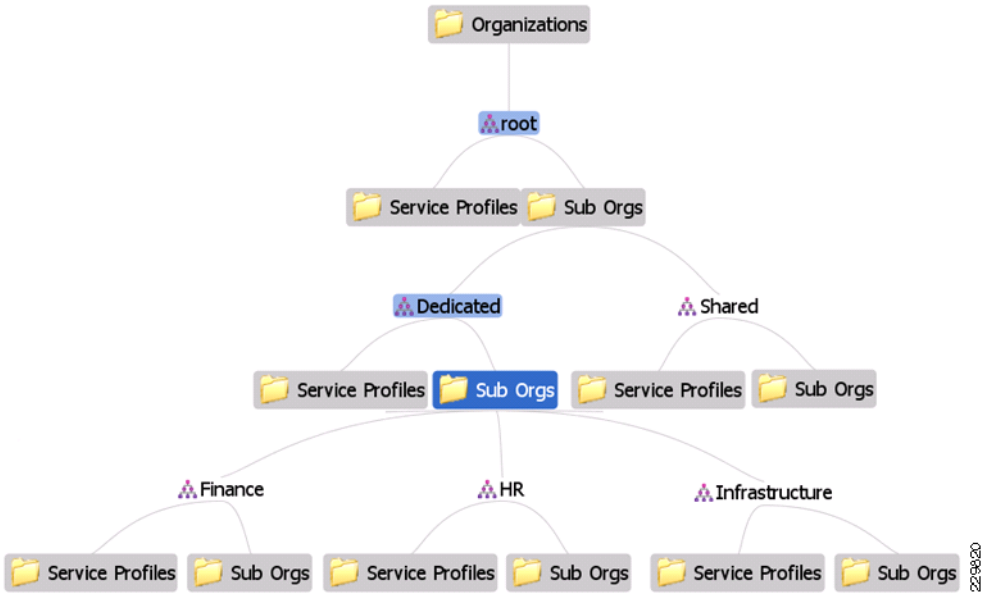
In this design, UCS Organizations were broken up into two primary sub organizations of the root organization, namely Shared and Dedicated. The name Shared was chosen to reflect that servers defined under this organization were a common resource for tenant consumption. Figure 6 details the UCS hierarchy for the Shared organization. The Shared organization within this design supports the Production sub-organization. The Production sub-organization defines another sub-organization named prod-drs1, which hosts the ESXi service profiles. This sub-organization delineates a set of service profiles (or server definitions) which directly aligns with the ESXi servers forming the VMware Distributed Resource Scheduler (DRS) cluster Production depicted in Figure 5. The Production sub-organization hierarchy can support multiple sub-organization definitions, such as prod-drs2 and prod-drs3, allowing for future placement of DRS clusters in the Shared organization.

Figure 6 Organizations in a Multi-Tenant Environment—Shared Organizations



The Dedicated organization isolates tenants at the compute layer via dedicated ESXi clusters or bare metal installations. The Dedicated organization speaks to tenants that may have business or application requirements that require committed resources. As show in Figure 5, the Infrastructure VMware DRS cluster has committed compute platforms to support the ESMT management applications. In Figure 7 those ESXi host resources are defined under the Infrastructure sub-organization. Figure 7 also indicates that the Finance and HR business units require dedicated server resources; this requirement is reflected under the “Dedicated” organizational unit within UCS.

Figure 7 Organizations in a Multi-Tenant Environment—Dedicated Organizations



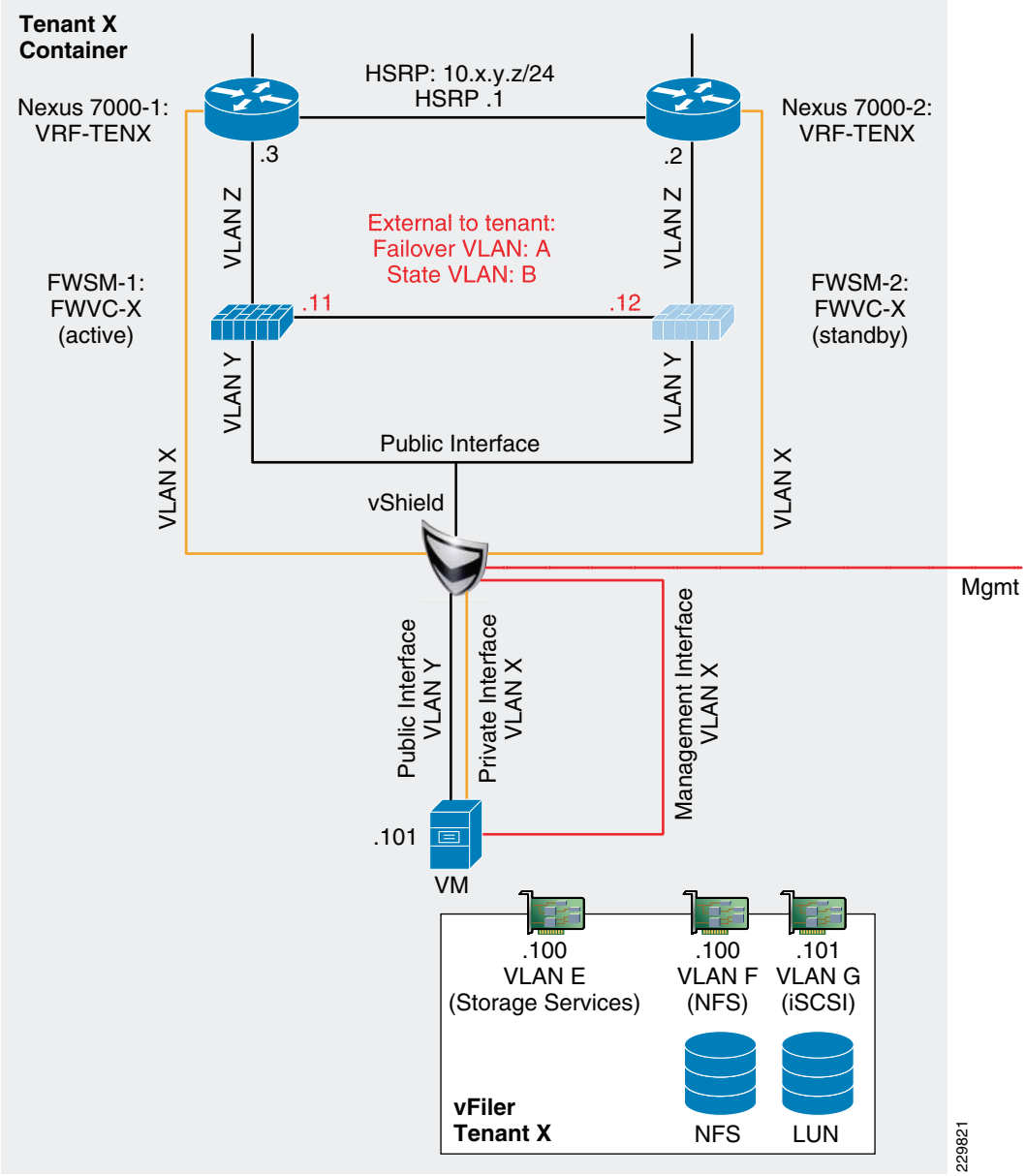
Production Tenant Model

Figure 8 illustrates the generic tenant model employed in the validation of this design where network, compute, and storage resources are allocated to support one or more applications contained on a virtual machine. The basic tenant leverages the following features and services:

- Layer 2 segmentation via VLANs
- Dedicated VRF for Layer 3 services
- Dedicated virtual firewall context to enforce stateful security services on ingress and egress data center tenant traffic
- vShield security services applied across the virtual compute layer to enforce inter-VM security policies
- VMware Resource Pools to impose compute resource allocation policy (not shown)
- Allocated vFiler units for storage services including file and block based

The tenant model is not restrictive to the number of VLANs, network services, or virtual machines employed. These are design considerations that must be revisited upon each tenant deployment in a data center invoking the ESMT design principals. To illustrate this point, Tenant Details highlights three enterprise workload examples and how the fundamental tenant structures were applied to address the particular needs of each application.

Figure 8 Generic Tenant Model



Tenant Details

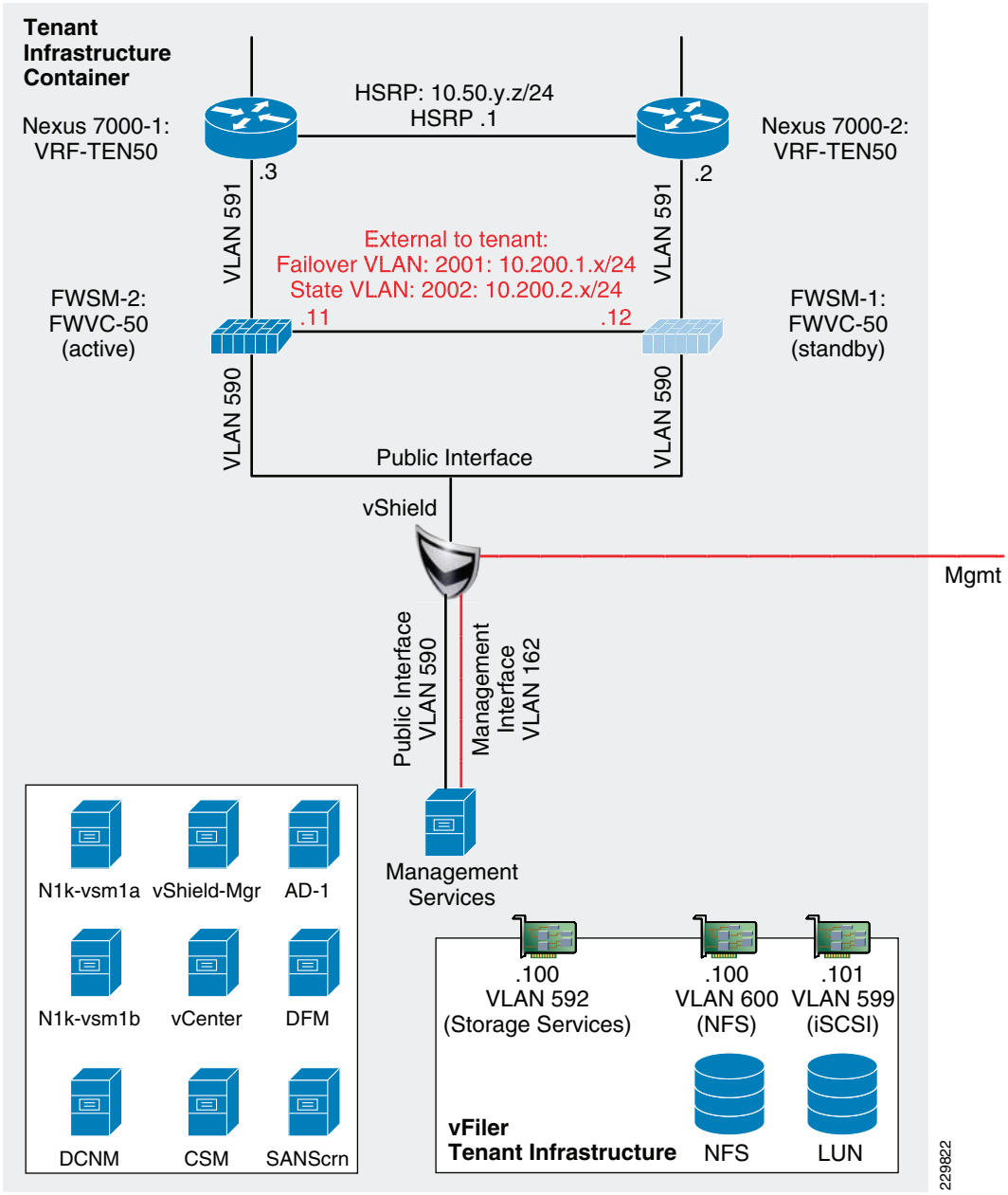
The following section details the tenant containers and associated applications housed within the enterprise Production environment. At a minimum, each of the tenants employ the generic tenant model described earlier and invoke other application services as necessary.

Tenant 0—Infrastructure

The Infrastructure tenant accommodates the management components of the virtualized data center. To this end, it is a “special” tenant, requiring the highest levels of service availability, scalability, security, and resource management. Ideally, all compute, network, and storage element managers leverage this tenant as a foundation. The various element managers employed in this solution are outlined in Architecture Components.

Figure 9 illustrates the Infrastructure tenants' logical topology and a sampling of the management applications services within the tenant container. In Figure 9, the Nexus 7000 aggregation switches dedicate a virtual routing instance as the first-hop IP router for the tenant. The HSRP virtual IP address 1 serves as the default gateway for all virtual machines in the container. The VRF is the Layer 3 tenant boundary; all other services are transparent from the server perspective.

Figure 9 Infrastructure Tenant



The Infrastructure tenant uses two types of firewall to secure the virtual management server farm, namely:

- A Firewall Service Module virtual context

- A vShield App (vApp) firewall virtual appliance
- A vShield Edge

In this deployment, the FWSM virtual context transparently bridges traffic between VLANs 591 and 590, providing stateful security services on data center ingress and egress tenant traffic. The virtual firewall context is dedicated to meet the security requirements of the Infrastructure tenant and therefore may be adjusted to meet the specific needs of the virtual machines and applications it front-ends.

The vShield 4.1 firewall virtual appliance (vApp) resides in the virtual access layer of the environment. The vShield App again provides another layer of security by controlling access between hosts on the same VLAN segment, essentially allow for prescriptive host isolation and in a larger sense VLAN consolidation within the data center as varied hosts may share the same VLAN without direct communication. Additionally, the vCenter container-based separation enables separation between the management and production clusters. vShield App can be enabled to allow only needed communication between the two clusters, creating clear separation between infrastructure and production tenants. It is important to note that a firewall is not only restrictive, but permissive, meaning that detailing the type of allowed traffic provides a new level of refinement and control within the virtual machine layer (e.g., allowing SQL traffic between vCenter and its database instance while implicitly denying all other communication between the nodes). The vShield App provides security between virtual machines within the tenant container.

The vShield Edge service allows administrators to securely pass traffic between tenant containers. The Infrastructure tenant virtual machines may securely communicate with other tenant virtual machines via vShield Edge. In this design, vShield Edge performs network address translation (NATing) between management services in the Infrastructure tenant and all tenant virtual machines in the Production DRS cluster. For example, AD services are centralized within the Infrastructure tenant but available to all tenants via vShield Edge. The granular rule set offered by the vShield Edge allows administrators to safely open communication between tenants.

Note In addition, to the security services of the FWSM virtual context and vShield virtual appliance, the cloud administrators may employ load balancing, network analysis, and intrusion prevention technologies described in later sections of this document if desired or required.

The virtual machines within the Infrastructure tenant are part of the Management VMware High Availability (HA) Cluster enabled with Distributed Resource Scheduler (DRS) and assigned to a resource pool. A resource pool allows cloud administrators to allocate appropriate compute resources (CPU and memory). The use of resource pools within DRS clusters provides hardware abstraction through compute aggregation, delivering both performance isolation and role-based access to the contents of individual pools. The Infrastructure tenant uses the Infrastructure ESX cluster's resource pool. Specific design considerations are highlighted in Tenant Availability.

Note The Infrastructure tenant is not required to reside on a distinct ESXi cluster; it may reside on the same ESXi cluster as other tenants with its own resource pool. However, a dedicated ESXi management cluster enables independent scalability of the infrastructure management environment and is the recommended approach.

The Infrastructure tenant uses a virtual storage controller, a NetApp vFiler unit, to store all relevant VMDKs and application datastores required by the management environment. This vFiler unit is assigned only to the Infrastructure tenant; it is isolated by the NetApp MultiStore solution through the implementation of VLAN segmentation, initiator groups (igroup), and IPSpaces. The NetApp vFiler unit may also be managed within a VM by the tenant administrator through the NetApp SnapDrive and "Storage Services" interface.

Tenant 1—Workload SharePoint

Figure 10 illustrates the logical topology of a tenant implementing a Microsoft SharePoint 2010 application environment. This tenant maintains the fundamental tenant structures of virtual routing instances, firewall instances, machines, and datastores discussed earlier. In addition, the SharePoint farm uses the application services available in the ACE virtual load balancer context and intrusion prevention technologies offered via multiple IPS virtual sensors. These devices provide application availability, scalability, and security enhancements to the tenant's application. Each of these features is introduced transparently at Layer 2 as "bumps in the wire."

Internal Zoning for SharePoint

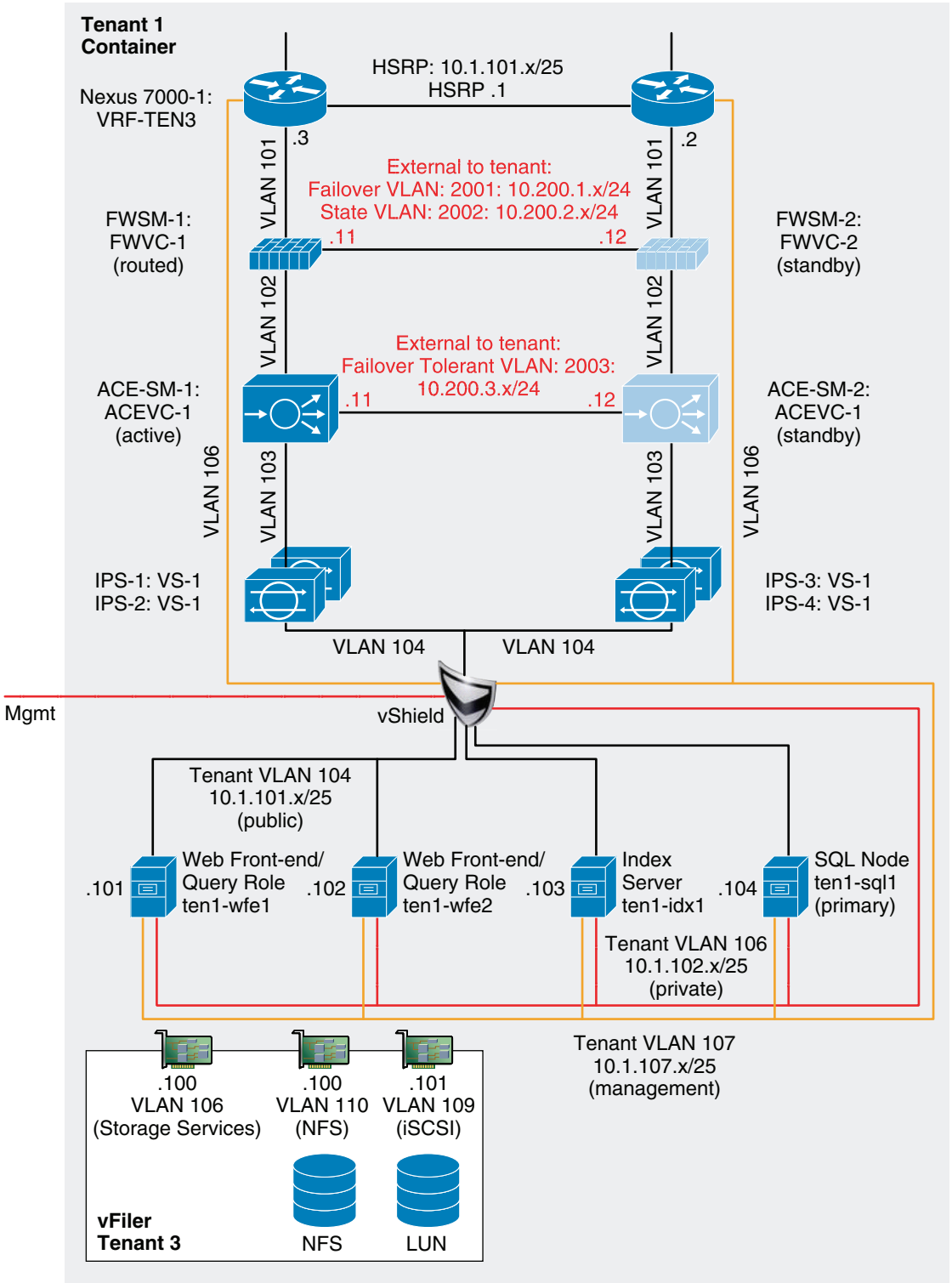
The SharePoint application consists of a database, Web tiers, and an index server. The vShield App vNIC-level virtual machine firewall allows us to carve up VLANs into multiple zones for the Web front ends, the index server, and the database backend server. To simplify the complexity of the virtual network and reduce the number of VLANs needed to create these three zones, the vShield App firewall rules are created using either custom vNIC Security Groups or via the use of vCenter containers. For example, it is possible to place the Web front end virtual machines into a vApp (collection of virtual machines) and create vApps for the remaining zones. Rules can be specified using the vApp names in the source and destination IP address fields. These are dynamic vCenter containers which would allow any of the tiers of the SharePoint application to be scaled by spinning up new virtual machines and they would automatically receive rules that belong to their tier. An example of a container-based rule would be from the SharePoint Web front end vApps to the database backend **vApp** allowing TCP/1433 (MS SQL DB port).

Note The vShield App functionality is being applied to all internal VLANs labeled "private" in Figure 10. These VLANs may support any number of services including storage services. The use of security services must be weighed against the risk and possible performance implications.

Edge Services for SharePoint

The SharePoint tenant environment uses the vShield Edge functionality to communicate with the Infrastructure tenant for AD and DNS services. The management interface of each virtual machine uses the vShield edge as the gateway to the Infrastructure tenants management services. This requires the host to implement a host-based route to the Infrastructure subnet. This host route is readily replicated via virtual machine template instantiation or cloning technologies.

Figure 10 Tenant 1 SharePoint

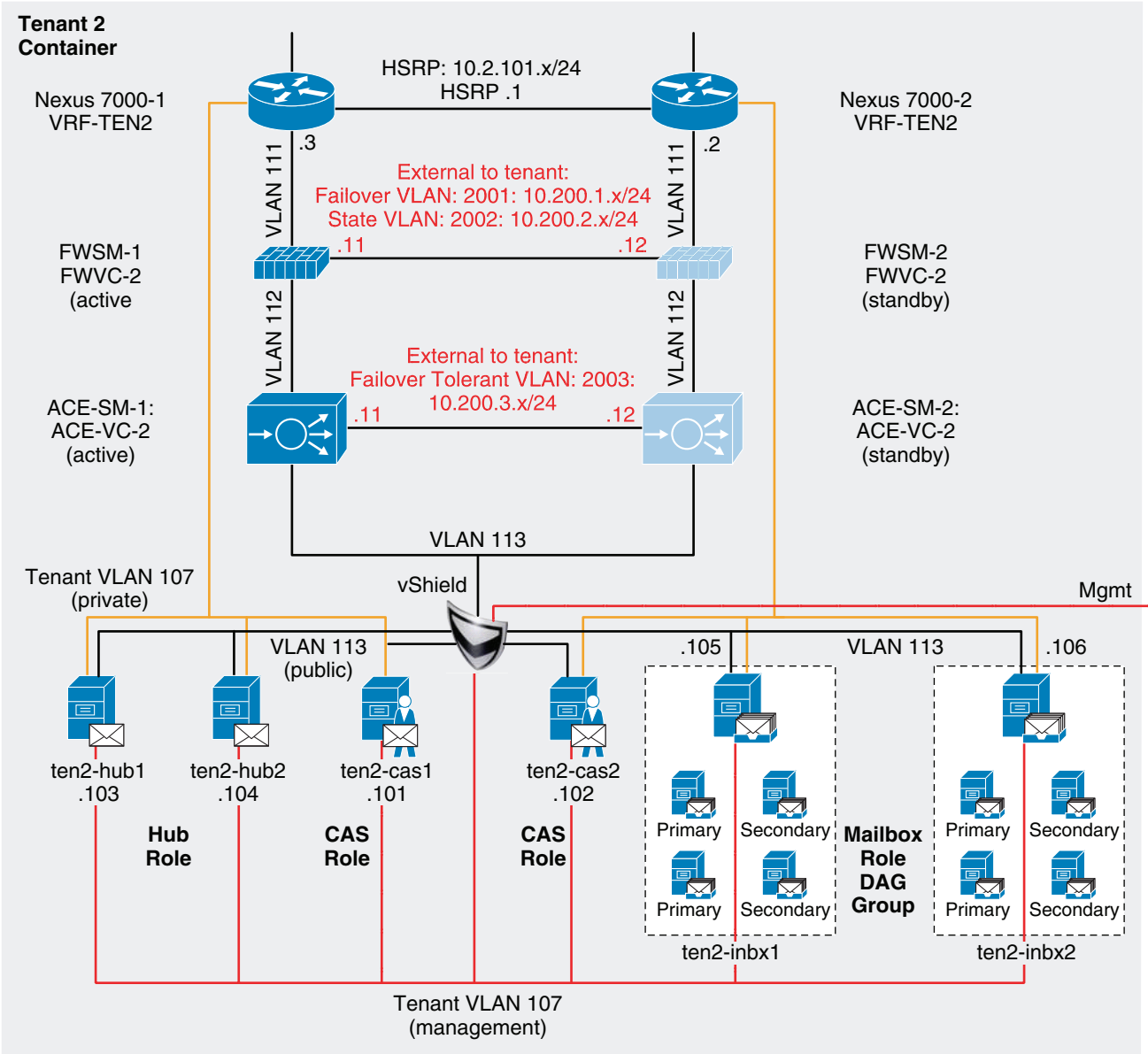


Tenant 2—Workload Exchange

The Microsoft Exchange 2010 application environment employs a subset of the services available to the SharePoint tenant environment. The Nexus 7000 VRF is the default gateway for the tenant servers. The use of Cisco firewall services and load balancers offers security and availability services from a client-server perspective, while VMware vShield App and Edge technologies secures inter-VM communication within and outside of the Exchange tenant application environment.

Notice the private VLAN segments within this environment do not employ the services of vShield. This design showcases the use of private VLANs (PVLANs) to construct a secure solution within a community or via complete isolation. This design also removes the performance implications of virtual appliance-based services.

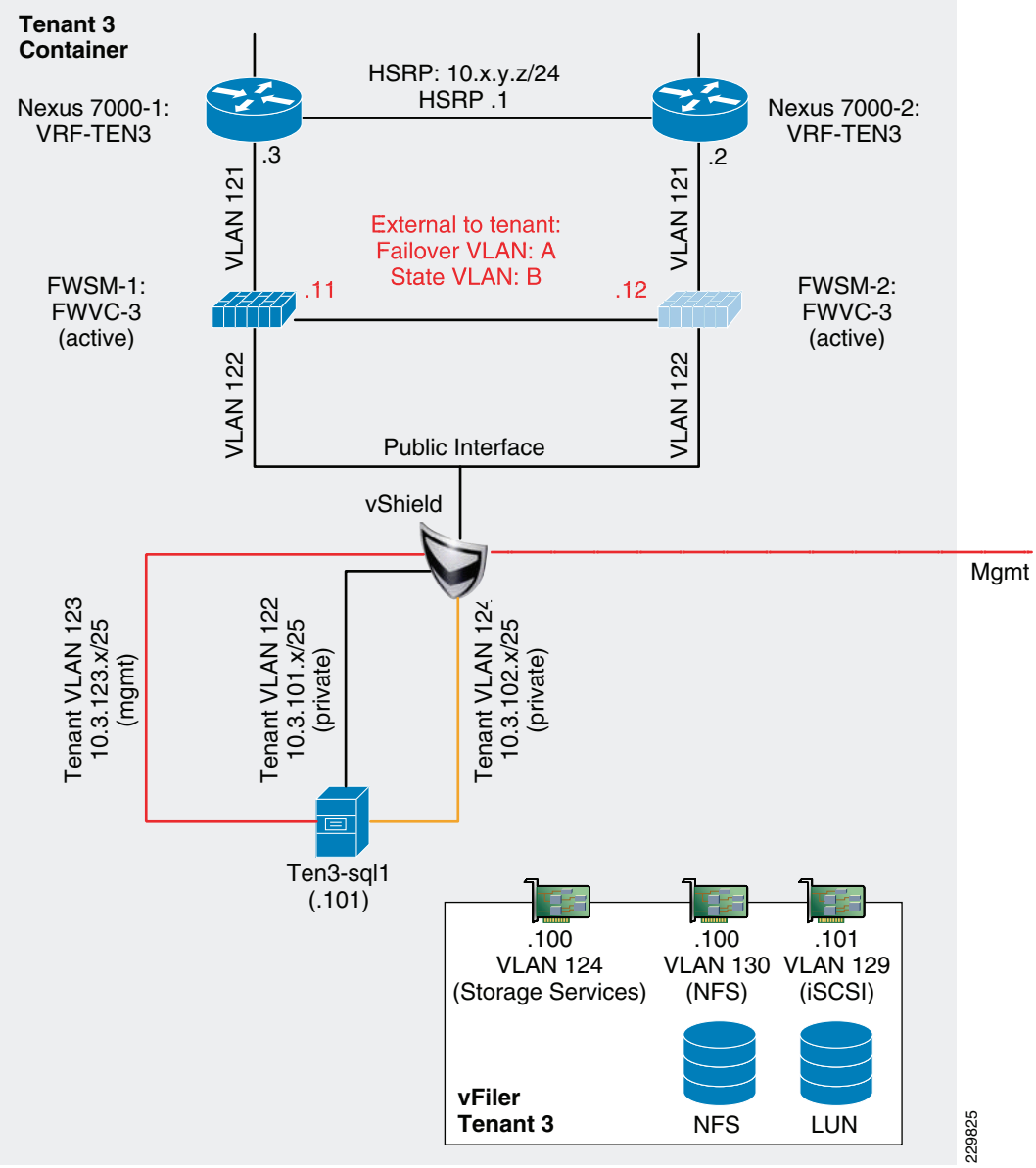
Figure 11 Tenant 2 Exchange



Tenant 3—Workload SQL

Tenant 3 supports a single SQL server instance within the data center. This instance of SQL may support application environments residing on other tenants within the enterprise depending on inter-tenant security policies. Figure 12 details the deployment model where firewall services are positioned to protect the database server. Again, the FWSM virtual context is in transparent mode and the virtual access layer uses a virtual application firewall. Inter-tenant traffic patterns are discussed in Secure Separation.

Figure 12 Tenant 3 SQL Server



Secondary Site Design

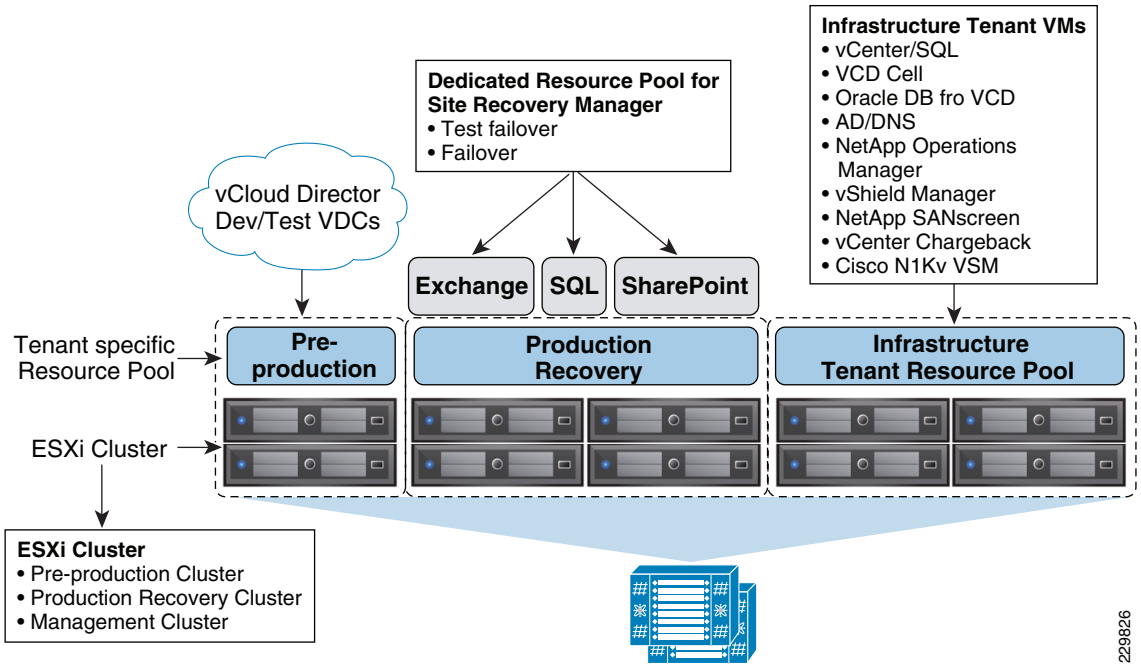
Due to the increase in data center expansion, complexity, and business needs, a secondary data center site is often required. Secondary data centers allow for business continuity and disaster recovery. These sites can be designed for load balancing, which allows for clustering and virtualization across the two sites, or for disaster prevention where the secondary site is a mirror of the primary, only to be used when resources in the primary site are unavailable. This design employs a mix of both methods by not only providing an environment for disaster recovery, but also providing an environment from which to run a Pre-production tenant during normal operations.

An investment in a secondary site for disaster recovery is important to many enterprises as it helps to protect against the negative business impacts of unavailable services and data. From a component perspective, secondary sites typically involve similar, if not the same, hardware and software configurations as the primary site data center. This helps to ensure that when tenant workloads are shifted from the primary to secondary site, they function as normal due to identical supporting environments at each site. In this design, the secondary data center is configured as an exact replica in both physical and logical aspects. Some parameters, such as IP address schemes, VLAN IDs, etc. may differ across sites. Certain infrastructure services and applications are also required to exist at each site and cannot be replicated between sites. Such services and applications include, but are not limited to, VMware vCenter, VMware Site Recovery Manager (SRM), Cisco Nexus 1000v Virtual Supervisor Module, etc.

One disadvantage to building a secondary site for disaster recovery is that this site is typically idle during normal operations, which leads to a very low return on investment. For this reason, this design includes a Pre-production tenant running at the secondary site during normal operations. This allows the enterprise to leverage the investment made in a secondary site, but still ensures primary site workloads (tenants) can be shifted to provide business continuity and disaster recovery.

Figure 13 depicts the data center VMware DRS cluster configuration. As shown, there are three dedicated clusters, Production, Infrastructure, and Pre-production. The Production cluster is a disaster recovery destination for the three tenants located on the primary site. The Infrastructure cluster houses another Infrastructure tenant which contains virtual machines required to manage the ESMT architecture at the secondary site. These machines are unique and are not recovered as part of the disaster recovery strategy. As with the primary site, the Infrastructure tenant could be deployed on the Production cluster, but it is still considered a best practice to employ dedicated physical resources. The remainder of this section documents the design of the Pre-production tenant.

Figure 13 Secondary Data Center Cluster Design



Pre-Production Tenant Model

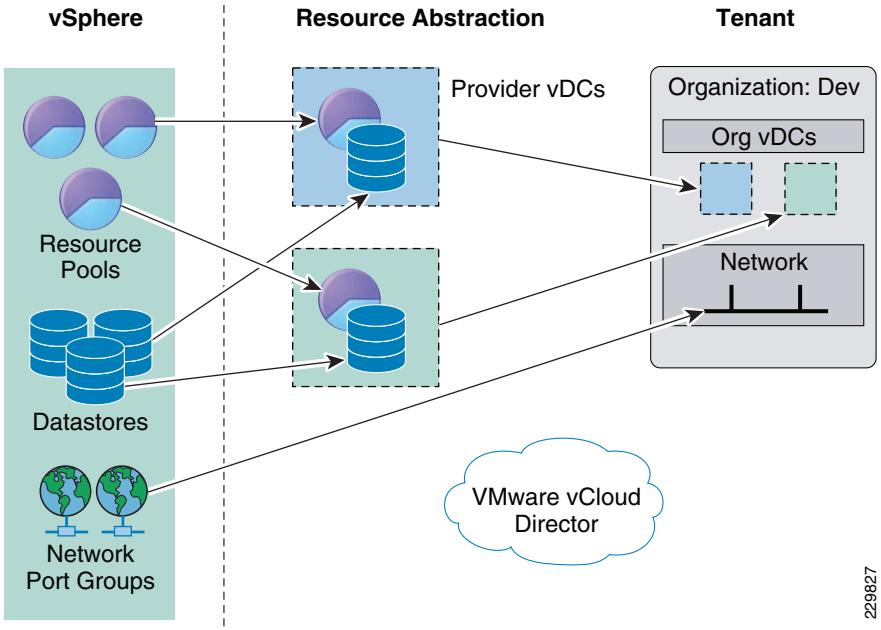
The fourth tenant in the ESMT architecture is the pre-production environment, typically development and test teams within an enterprise. The compute, network, and storage resource requirements for this particular tenant are provided by a self-service model, enabled by VMware vCloud Director. The entire environment is deployed in the recovery data center to enable an "active/active" model where both the Primary and Secondary sites are serving active workloads. The design principles of the four pillars are also applied in the pre-production vCloud environment in the secondary site.

Note VMware Cloud Director 1.0 release is not compatible with Site Recovery Manager. DR test/failover for the pre-production environment is out of the scope of this document.

Tenant 4—vCloud New Abstraction Layer

vCloud Director creates a new layer of abstraction for consumers of compute, network, and storage resources. Figure 14 illustrates the self-service dev/test tenant model enabled by vCloud Director. In this model, all underlying compute, network, and storage resources are abstracted by vCloud Director as a Virtual Data Center (VDC). Each tenant in a pre-production environment is identified by vCloud as an organization. Each organization is assigned a VDC (can be more than one); the VDC represents the compute and storage resources for end user consumption. Network connectivity is provisioned in the form of organization network, which can connect directly to the pre-production VLAN, or a private organization network that is self-contained or routed to the pre-production VLAN.

Figure 14 Self-Service Dev/Test Tenant Model Enabled by vCloud Director



The tenant model is not restrictive to the number of organization VDCs, network port groups assigned, and types of network connectivity to the external network. These are design considerations that must be revisited on each tenant deployment in a data center invoking the ESMT design principals. To illustrate this point, Tenant Details highlights a dev/test use case example and how the fundamental design principals were applied to address the particular needs of pre-production environments.

vCloud Availability

The overall compute availability design for tenants is enabled by the following features/product components:

- VMware HA
- vMotion
- VMware vCloud Director Stateless Cell

All design considerations for VMware HA, vMotion, and Storage vMotion highlighted in Availability are applicable for the pre-production vCloud tenant.

vCloud Director Stateless Cell

VMware vCloud Director utilizes stateless cells (vCD Cells) that provide all of the functionality required to provide the new layer of abstraction required for the self-service model through vCD Web portal. All cells interact with vCenter Server as well as directly with ESX/ESXi Hosts. Therefore, it is imperative to have more than one vCD Cell deployed in the pre-production ESXi cluster. As the environment scales to more ESXi Server hosts and/or vCenter Server instances, additional vCD Cell can be added. The cells are stateless by design, meaning in the event of cell VM outage, surviving cell(s) can take over Web portal access, vCenter/ESXi communication, and vCloud API calls.

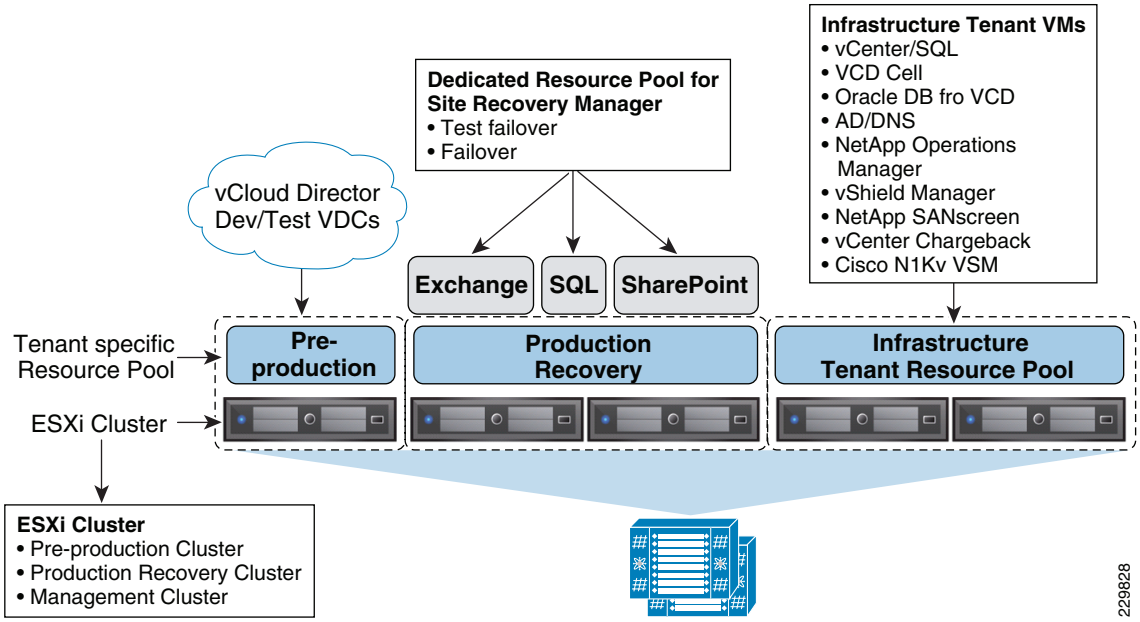
vCloud Secure Separation

Compute

Separation Between Production and Management

For the Recovery Data Center, the design principle of separation between compute resources for production, pre-production, and infrastructure management is applied. The UCS blade servers are separated into three separate ESXi clusters and dedicated resource pools within each ESXi cluster. Figure 15 illustrates the separation of compute cluster for production recovery, pre-production, and management.

Figure 15 High-Level Logical Topology of Resource Pool Allocation and Separation in Recovery Site



Cluster Design Considerations

- All clusters need to be enabled with VMware HA and DRS.
- Each cluster hosts its own dedicated resource pool/sub-resource pools.
- The Management cluster hosts the following infrastructure management/services virtual machines:
  - vCenter Server VMs
  - VMware Cloud Director Cell VMs
  - vCenter Chargeback Server VM with Chargeback and vCloud, VSM Data Collector
  - vShield Manager
  - Microsoft SQL DB VMs for vCenter Server
  - VMware Cloud Director Oracle DB VM
  - Nexus 1000V VSM
  - AD/DNS
  - Cisco DCNM
  - NetApp SANscreen
  - NetApp Operations Manager
- The Pre-production cluster hosts compute resources for vCloud Director organization consumption.

- The Production Recovery cluster hosts three separate resource pools, one for each of the production applications running in the primary site (Exchange, SQL, and SharePoint).

Separation Between Organizations Within vCloud Environment

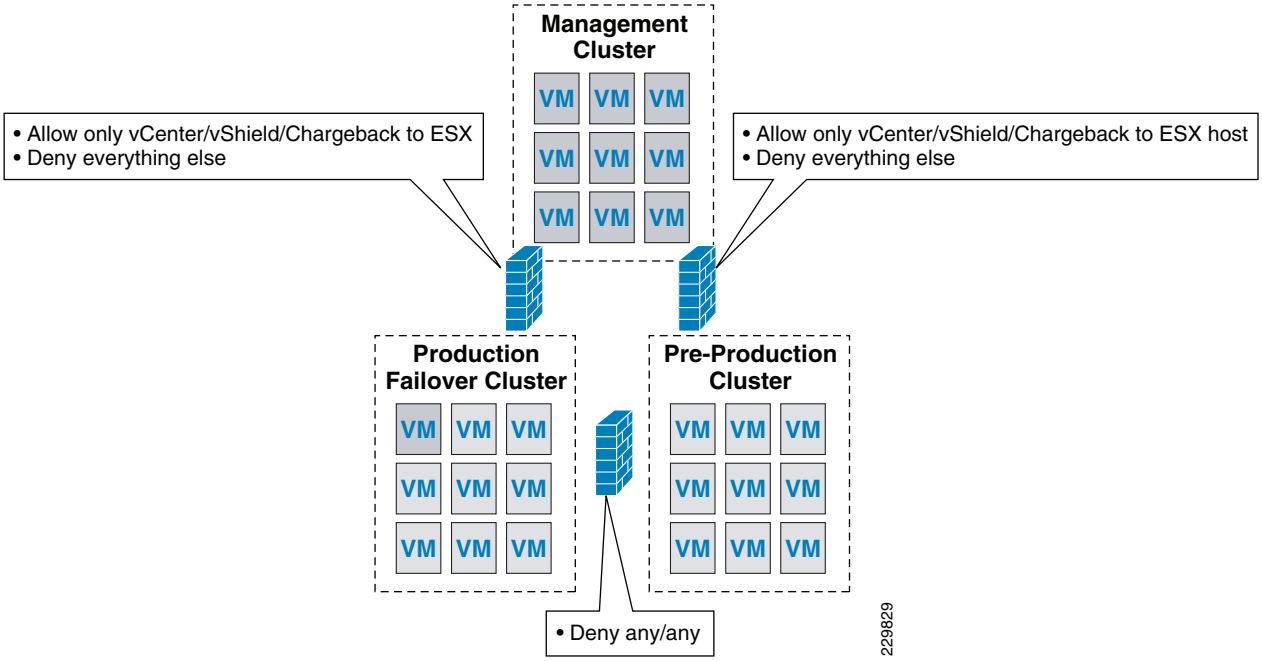
vCloud Director has built-in multi-tenancy for end tenants that consume the resources. Each organization is associated with its own organization VDC, completely isolated from other organizations. Though both organization VDCs consume resources from the same Provider VDC, vApps running in one organization VDC are completely isolated from other organization VDCs. Therefore, visibility and control of resources are restricted to each organization.

Network

Separation Between Production, Pre-Production, and Management

The pre-production environment has a dedicated dev/test VLAN. vShield App is used to provide policy-based separation at the individual cluster level. Given the compute separation based on ESXi cluster, policies can be defined with production, pre-production, and management clusters as the source and destination pairs, as illustrated in Figure 16.

Figure 16 Production, Pre-Production, and Management Clusters



It is best practice to allow all traffic between the management cluster and the other clusters at their initial deployment and leverage vShield Manager to capture all traffic types between the two. Once that is completed and analyzed, use deny any/any as the baseline and only allow needed traffic needed for the different clusters to communicate with the management cluster.

Separation Between Organizations Within vCloud Environment

It is imperative to understand the new network abstraction created by vCloud Director in order to properly design the environment for multi-tenancy.

External Network

An external network is a logical, differentiated network based on a vSphere port group/port profile created from the Cisco Nexus 1000V. In the context of a public service provider that uses VMware vCloud Director as the platform for offering services, the external network provides the interface to the

Internet for virtual machines connected to external organization networks. In the context of a private cloud deployment for pre-production environment within an enterprise ESMT architecture, an external network provides the interface to the dedicated VLAN for pre-production organizations.

vCloud Director is compatible with the Cisco Nexus 1000. With the Nexus 1000V, the network pool must be “backed by portgroup”. The “VLAN backed” and “VCD Network Isolation backed” are not used in this design. The “external” network for the pre-production environment is the pre-production VLAN, backed by the pre-production port group created on the Nexus 1000V. Both dev and test organizations share the same external network with isolation services provided by vShield Edge.

Note As each external network is differentiated, additional separation can be created by having a dedicated VLAN for dev and test, instead of sharing one. In this case, two external networks are created, one backed by a dev VLAN and another backed by test VLAN. vCloud Director provides the flexibility for administrators to design the architecture based on business requirements within the enterprise.

Network Pool

Each organization, dev and test in this case, is associated with their own network pool, each backed by a set of port groups.

- Dev Network Pool:
  - Two total port groups defined (one with routed VLAN and another with private, non-routable VLAN)
  - Three organization networks are to be created from the Dev Network Pool for vShield App network connectivity

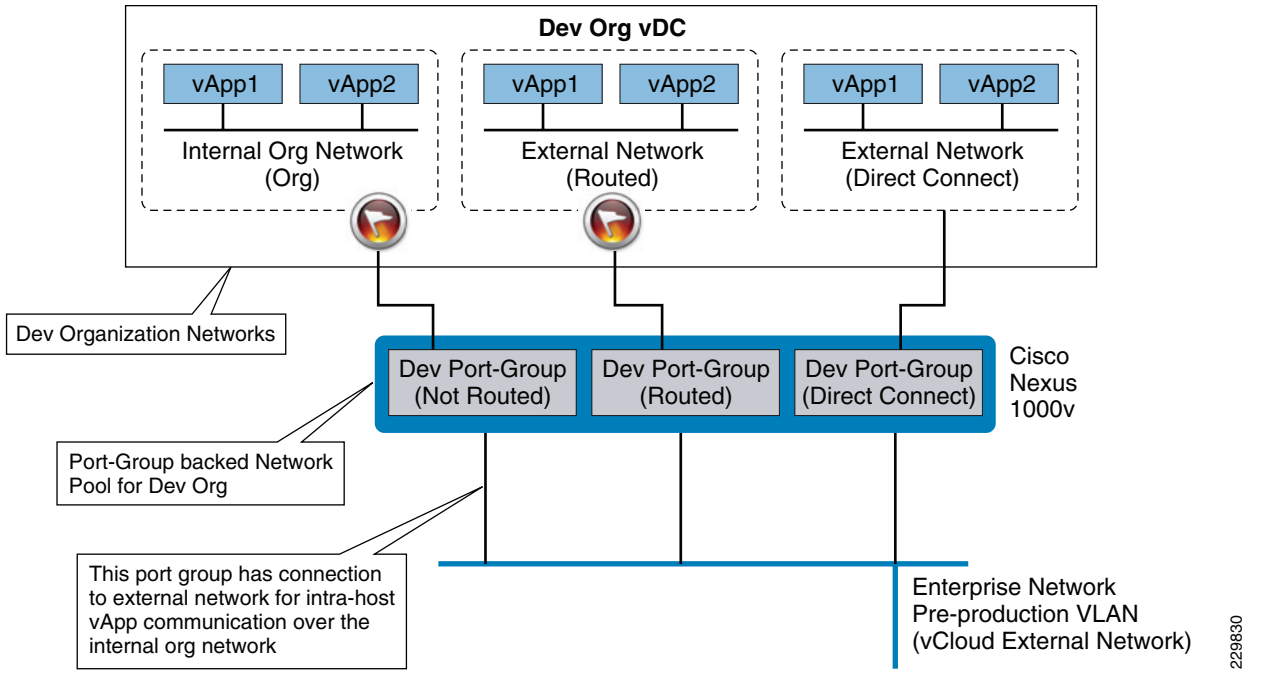
  Note Only two port groups need to be created to back this network pool as the “direct connect” organization network allows vApps to be directly connected to the External Network port group.

- Internal Organization Network—vApps connected to this network are only accessible by the dev organization. This is ideal for development teams that require an environment that is completely isolated from all other teams within the enterprise organization.

Note Access to virtual machines connected to this internal network can be done via the virtual machine remote console via the vCloud Web portal.

- External Organization Network—Direct Connection—vApps connected to this network are accessible by both dev and test organizations. This is useful for common access to a set of vApps shared between dev and test. One example use case would be a bug reporting application for dev and test teams to collaborate on development and test findings.
- External Organization Network—NAT-Routed Connection—vApps connected to this network can have access to the pre-production VLAN through address translation service provided by vShield Edge. Individual or groups of selected vApps can have private addresses translated by vShield Edge. Example use cases are:
  - The dev team finishes functional testing of new software code and the code is packaged into a vApp with a Web server front-end for the test team to access and initiate additional stress testing.
  - Collaboration (troubleshooting) between dev and test teams on certain dev setup.

Figure 17 vCloud Networking Options



Note Test Network Pool and organization networks design are identical to the above design for dev, aiming for simplicity and consistency.

Storage

Similar to the MultiStore design in the primary site, the pre-production vCloud tenant is assigned a dedicated vFiler unit for all of its storage needs. All organizations (in this example, dev and test) share volumes managed and provisioned by a vFiler unit dedicated for pre-production.

Further information on hardening security for vCloud Director implementation can be found in the vCloud Director Security Hardening Guide: http://www.vmware.com/files/pdf/techpaper/VMW_10Q3_WP_vCloud_Director_Security.pdf.

vCloud Service Assurance

Compute and Storage resources go hand-in-hand in the vCloud Director abstraction layer—Virtual Data Center. A provider vDC (PvDC) consists of a vSphere resource pool and datastore(s). Therefore, both are discussed in the context of service assurance for the pre-production environment.

The pre-production ESXi cluster can carve out up to three main resource pools, each representing a service level: Gold, Silver, and Bronze. The recommended values for resource pool CPU/memory reservations, limits, shares, and expandable reservation are the defaults (no CPU/memory reservation, enabled expandable reservation and unlimited). The rationale is that the resource allocation models in vCloud Director will define these specific settings at the sub-resource pool level to achieve the resource guarantee on a per organization vCD basis. Compute and storage service tiering can be achieved by leveraging the following built-in resource adjustment settings in vCloud Director and NetApp MultiStore:

- Organization Lease, Quota, Limits (compute)
- Organization vCD Allocation Model (compute)

- NetApp FlexShare (storage)

Note This is one way of designing a service level offering based on vSphere resource pool values. If there is a mixture of different model UCS blades in the environment, the service level offering can be based on the model of blade. For example, an ESXi Cluster with B250 blades can be considered a Gold level service offering and ESXi Cluster with B200 blades can be considered as the Silver level.

Design Considerations for Organization Lease, Quota, and Limits

With the PvDCs created based on SLA, resource allocation to dev and test organizations with VMware vCloud Director involves specifying leases. Leases provide a level of control over an organization's storage and compute resources by specifying the maximum amount of time that vApps can be running and that vApps and vApp templates can be stored. There are two types of leases, runtime lease and storage lease.

The goal of a runtime lease is to prevent inactive vApps from consuming compute resources. For example, if a user starts a vApp and goes on vacation without stopping it, the vApp continues to consume resources. A runtime lease begins when a user starts a vApp. When a runtime lease expires, vCloud Director stops the vApp.

The goal of a storage lease is to prevent unused vApps and vApp templates from consuming storage resources. A vApp storage lease begins when a user stops the vApp. Storage leases do not affect running vApps. A vApp template storage lease begins when a user adds the vApp template to a vApp, adds the vApp template to a workspace, downloads, copies, or moves the vApp template.

When a storage lease expires, Cloud Director marks the vApp or vApp template as expired or deletes the vApp or vApp template, depending on the organization policy that has been set.

Understand the needs for dev and test organization users prior to configuring the values for lease, quota, and limits and the corresponding actions to take for vApps. For best efficiency and preservation of storage space, it is best to configure maximums for each organization based on the service level agreement/objective instead of limitless maximums.

Leases:

- Run Time Lease—The maximum amount of time that vApps can be running in the organization virtual data center is determined by the lease value. A use case for test organization is to first determine the testing cycle required for each iteration of software testing and specify a value (in hours, days) with some buffer (i.e., extra day/few hours). That way, when test organization users start vApps to spin up testing, the vApps get powered off automatically after the specified lease value. If the nature of the testing is non-deterministic, then the conservative approach is to set lease value of "never expire".
- Storage Lease—This value determines how long powered off vApps are available before they are automatically cleaned up. Available cleanup actions are "move to expired items"or "permanently delete".

Quotas determine how many virtual machines each user in the organization can store and power on in the organization's virtual data centers. The quotas you specify act as the default for all new users added to the organization. This setting is useful to prevent the virtual machine sprawl issue as virtualization has made it very easy and efficient to provision new development/test workstation/servers. By limiting this on a per-user basis in a the dev and test organizations, end user behavior can be controlled to only provision what is needed and only what is needed.

Limits determine how many resource intensive operation per user and per organization. Such operations include move or copy vApps, upload vApps to the cloud, etc. It is good practice to limit these on a per-user and per-organization basis, to prevent any one user or organization from affecting one another,

by performing large number of such operations. Last but not least, the number of simultaneous virtual machine console connection can have an upper limit defined as well, to prevent such operation from occupying too many resources that could otherwise be used for true computing.

Resource Allocation to Organizations

When allocating compute resources to the individual organizations, three allocation models are available to accommodate different levels of service:

- Allocation Pool—Only a percentage of the resources are committed; this model allows for over commitment of resources.
- Pay-As-You-Go—Resources are committed only when vApps are powered on in the organization.
- Reservation Pool—Allocated resources are 100% committed to the organization.

Design Considerations for Pre-production Orgs Resource Allocation

The resource allocation model defined above determines the amount of resource guarantee for each organization virtual data center. The following grouping is used as a reference model for the pre-production environment in this architecture:

Dev Organization

- Gold-Dev vDC—Reservation pool with CPU/memory amount that needs to be guaranteed without impact to development schedule.
- Silver-Dev vDC—Allocation pool with CPU/memory amount that is 50% of the Gold Dev-vDC; additional utilization is not restricted, but can be metered and charged back to the organization for over-utilization of committed amount.

Test Organization

- Gold-Test vDC (reservation pool with CPU/memory amount that needs to be guaranteed without impact to testing schedule)
- Bronze-Test vDC (pay-as-you-go pool) consuming resources from Bronze Provider vDC (ideal for use cases of load gen testing or regression testing that only takes place after code freeze)

Design Considerations for NetApp FlexShare

The Gold, Silver, and Bronze pre-production resource pools are fully controlled by vCloud Director for Provider vDC creation. Given that VMware vCloud Director treats datastores as a generic set of storage resource pool—meaning one datastore is not treated differently from another in terms of vApps/virtual machines placement based on performance attributes—then the following approach is used in the design for a consistent end-to-end quality of service model:

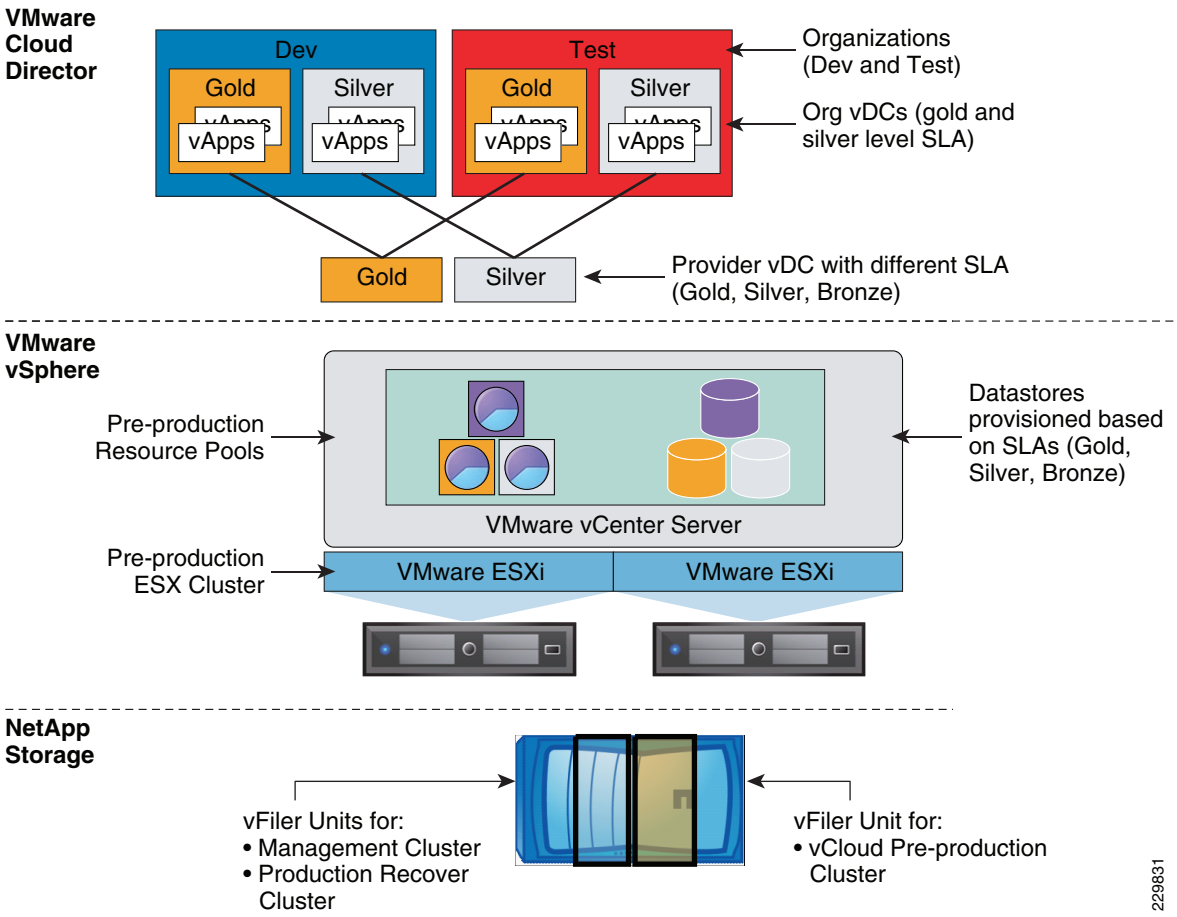
Define levels of service for storage volumes provisioned to the pre-production ESXi Server cluster (Gold, Silver, and Bronze)

The service level offerings can be based on the catalog defined by NetApp Operations Manager, which includes Snapshot copies and replication schedules or simply based on FlexShare values. In this design, FlexShare values are used as the baseline for service level offerings for simplicity:

- Gold datastore—Very high
- Silver datastore—Normal
- Bronze datastore—Low

The Gold datastore pairs up with the Gold Provider vDC during creation of the PvDC (the same applies for Silver and Bronze datastores and PvDCs). See Figure 18 for an SLA-based compute and storage resource consumption model by dev and test organizations.

Figure 18



vCloud Management

The new layer of abstraction introduced by vCloud Director eliminates the needs for dev and test organizations to access vCenter Server. All of the compute, storage, and network needs are addressed by accessing the self-service Web portal. The cloud admin typically carries out the following tasks to enable self-service for a given organization to consume resources from the vCloud environment:

- Create Organization
- Save portal URL for organization
- Allocate CPU/memory/storage resources to organization by creating org vDC
- Create and assign organization network
- Assign catalog to organization

For expansion of resources for a given organization, the following tasks are carried out by the cloud admin:

Expansion of Organization Resources

- If pre-existing defined limits have been set and can be extended, extend the limits
- If not:

- Create additional organization vDCs
- Create additional organization networks

Refer to Management for further information about available APIs for end-to-end orchestration, without any manual intervention to vCenter Server or vCloud Web portal.

Design Considerations—The Four Pillars

This section discusses design considerations for the four pillars:

- Availability
- Secure Separation
- Service Assurance
- Management

Availability

Availability is the first pillar and foundation for building an Enhanced Secure Multi-Tenancy environment. Eliminating planned downtime and preventing unplanned downtime are key aspects in the design of the Enhanced Secure Multi-Tenancy infrastructure. This section covers availability design considerations and best practices related to compute, network, and storage. See Table 1 for various methods of availability.

Table 1 Methods of Availability

Compute	Network	Storage
• UCS Dual Fabric Redundancy • vCenter Heartbeat • VMware HA • VMware Site Recovery Manager • vMotion • Storage vMotion • vShield Manager built-in backup	• EtherChannel • vPC • Device/Link Redundancy • MAC Learning • Active/Passive VSM • Fault Tolerant Service Integration • Application Health Monitoring	• NetApp RAID-DP • Virtual Interface (VIF) • NetApp HA • NetApp Snapshot • NetApp SnapDrive • NetApp SnapManager • NetApp SnapMirror • Adapter for VMware SRM

Highly Available Physical Topology

At the compute layer, Cisco UCS provides a unified compute environment with integrated management and networking to support compute resources. VMware vSphere, Cisco Nexus 1000V, and VMware vCloud Director run on top of the highly-available UCS platform and provide high availability at the logical virtualization layers. Refer to Design Considerations for Tenant Availability for design consideration for high availability in the virtualization layer. At the network layer, the architecture is enabled with Nexus 5000 as a unified access layer switch and Nexus 7000 as a virtualized aggregation

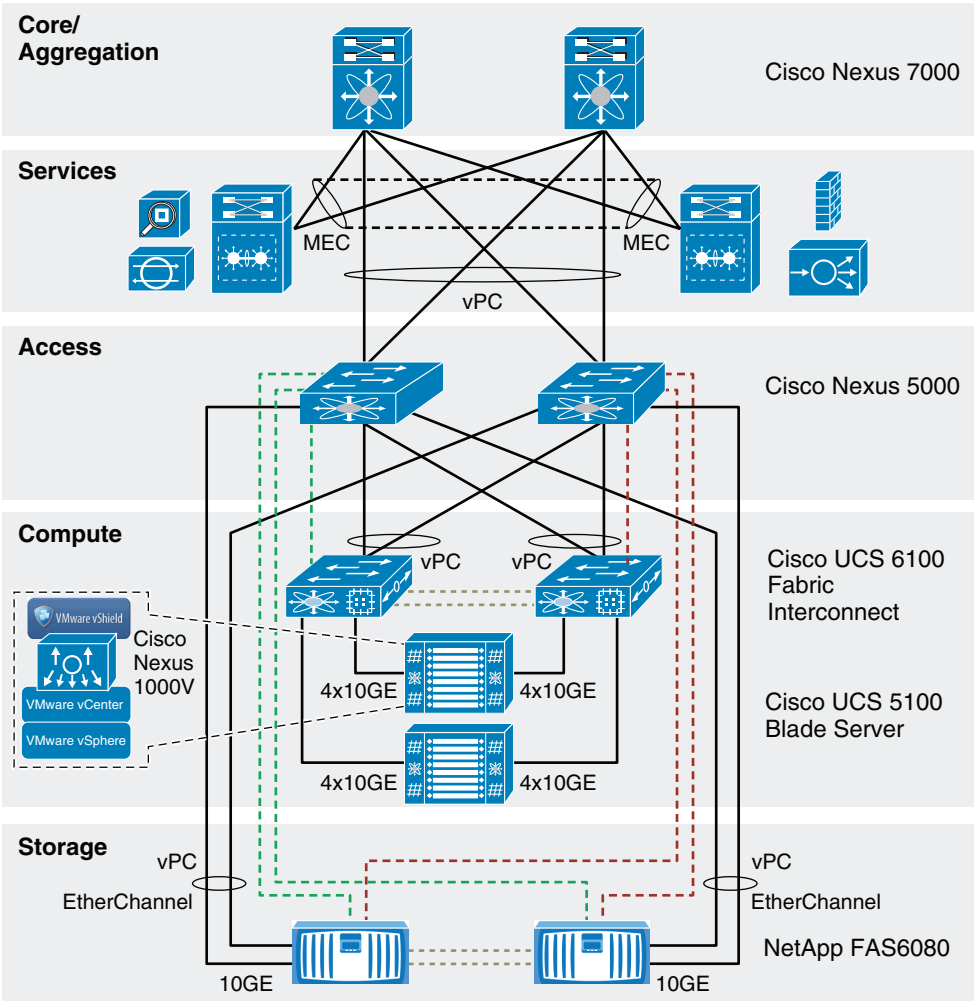
layer switch. The two UCS 6120 Fabric Interconnects provides a robust compute layer platform. Via vPC a topology with redundant chassis, card, and links with Nexus 5000 and Nexus 7000 provides a loopless topology.

Both the UCS 6120 Fabric Interconnects and NetApp FAS storage controllers are connected to the Nexus 5000 access switch via EtherChannel with dual-10 Gig Ethernet. The NetApp FAS controllers use redundant 10Gb NICs configured in a two-port Virtual Interface (VIF). Each port of the VIF is connected to one of the upstream switches, allowing multiple active paths by utilizing the Nexus vPC feature. This provides increased redundancy and bandwidth with a lower required port count.

The Cisco Nexus 5000 access layer switches provide dual-fabric SAN connectivity at the access layer and both UCS 6120 and NetApp FAS are connected to both fabrics via Fiber Channel (FC) for SANBoot. The UCS 6120 has FC links to each controller, each providing redundancy to the other. NetApp FAS is connected to the Nexus 5000 via dual-controller FCoE adapters in a full mesh topology.

The Nexus 7000 provides redundant paths to the Nexus 5000 access layer and VSS enabled service domain via vPC. vPC provides a logically loopless topology with convergence times based on Etherchannel and not spanning tree. The VSS-enabled service domain supports both integrated network service modules and port density for appliance-based services. The fault management and state traffic related to these network-based services are managed and contained within the VSS domain.

Figure 19 Physical Topology



Design Considerations for Compute Availability

Unified Computing System

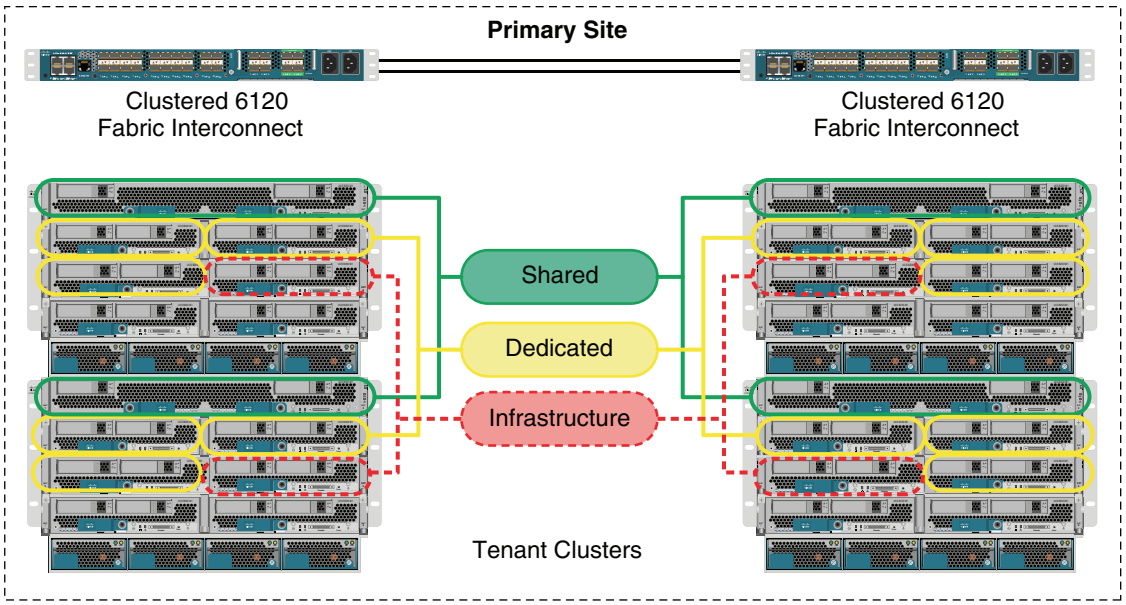
In the UCS, hardware can be presented in a stateless manner that is completely transparent to the OS and the applications that run on it. A Service Profile is made, creating a hardware overlay that contains specifics sensitive to the OS:

- MAC addresses
- WWN values
- UUID
- BIOS
- Firmware versions

The Service Profile boots from a LUN that is tied to the WWPN specified, allowing an installed OS instance to be locked with the Service Profile. The independence from server hardware allows installed systems to be re-deployed between blades. Through the use of pools and templates, UCS hardware can be deployed quickly to scale.

The Service Profiles form the backbone of the unified computing fabric within the data center. As shown in Figure 20, the UCS system is carved into "Shared", "Dedicated", or "Infrastructure" categories. Each of these represents a distinctive environment within the UCS fabric created by UCS policy. These servers are associated with different service profiles, server pools, and roles. The "Shared" category hosts VMware DRS clusters for tenants, the "Infrastructure" category consists of servers supporting the management tenant environment, and the "Dedicated" servers are platforms not employing virtualization or those virtualized and devoted to specific environments such as virtual desktop infrastructures (VDI). By using service profiles, the servers can be readily spread across the UCS fabric to minimize the impact of a single chassis failure.

Figure 20 UCS BladeTopology for Clusters



Note The implementation of uniformly configured sets of UCS 6120 Fabric Interconnects allows for an added degree of scalability within a single site. The VMware HA clusters can span differing UCS 5108 chassis and even UCS 6120 clusters creating greater opportunity to scale beyond the limits of a single fabric.

Design Considerations for Network Availability

Hierarchical Design

For more information on IP infrastructure high availability best practices, see: http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DC_3_0/DC-3_0_IPInfra.html.

This design guide addresses the components required to build a highly-available, multi-tenant, virtualized infrastructure. This document does not detail the end-to-end aspects of availability. The underlying assumption is that a highly-available infrastructure is the fundamental backbone of any

multi-tenant virtualization service. The key design attributes of this adaptation for Enhanced Secure Multi-Tenancy are described below, including newer design options based on Nexus 1000V and UCS Virtual Interface Card (VIC) capabilities.

The infrastructure design for multi-tenant is based on a three-tier core, aggregation, and access model as described in Figure 19.

Data center technologies are changing at a rapid pace. Cisco network platforms enable the consolidation of various functions at each layer and access technology, creating a single platform for optimized resources utilization. From a hierarchical perspective, there are two consolidation alternatives:

- **Aggregation layer**—Traditionally the aggregation layer is designed with a physical pair of hardware devices, enabling network connectivity at various speeds and functionality. With the Nexus 7000, the Virtual Device Context capability enables the consolidation of multiple aggregations topologies, with multiple distribution blocks represented as a logical entry in a single pair of Nexus 7000 devices. VDC-level separation is desirable because:
  - Compliance-level separation is required at the aggregation layer.
  - Explicit operational requirements, such HSRP control, active/active, site-specific topologies, and burn in address (BIA) requirements for specific access layer devices.
  - Separation of user space application separation against control (vMotion) and network management (SNMP, access to non-routed network, etc.).
  - Introduction of advanced network services such as Overlay Transport Virtualization (OTV) via VDC appliances.

This design uses a single VDC model where an active context supports the multiple tenants in the environment.

- **Access layer**—The second consolidation is at the access layer, which presents the most challenging integration requirements with a diverse set of devices. The access layer consists of server, storage, and network endpoints. The goal is to consolidate and unify the access layer with existing access layer topologies and connectivity types. The unification of the access layer needs to address the following diverse connectivity types:
  - Separate data access layer for class of network—Application, departmental segregation, functions (backup, dev-test)
  - Separate Fiber Channel (FC), Networked File System (NFS), and Tape Back storage topologies and access network
  - Edge layer networking—Nexus 1000V, VBS (Virtual Blade Servers), blade-systems, and stand-alone (multi-NIC) connectivity
  - 100 M, 1G, and 10 G speed diversity
  - Cabling plant—EOR (end of row) and TOR (top of rack)

This design mainly focuses on the consolidation of compute resources enabled via UCS and storage integration with NFS. Consolidation at the access layer requires a design consisting of these key attributes:

- Consolidation and integration of various data networks topologies
- Unified uplink—10Gbps infrastructure for aggregated compute function (more VMs pushing more data)
- Consolidation and integration of storage devices integrated with Ethernet-based topologies

Storage topology consolidation is one of the main drivers for customers to consider adopting unified access at the compute level. The consolidation of storage topologies into the existing Ethernet IP data infrastructure requires assurance and protection of storage traffic in term of response time as well as bandwidth. The remainder of this section describes the network availability and design attributes for the Enhanced Secure Multi-Tenancy architecture.

Core Availability

The data center core is meant to be a high-speed Layer 3 transport between the data center and any other outside entity. High availability at the core is an absolute requirement for in this design. Using the technologies available in the Nexus 7000, it can be achieved in the following ways:

- Device redundancy—The core is typically composed of two devices, each with a connection to outside of the data center and a connection back to the aggregation layer of the data center.
- Path redundancy—With the core comprised of Layer 3 links, this is done primarily using redundant routed paths. The core should have redundant paths to the campus and WAN as well as the aggregation layer.
- Supervisor redundancy—To account for hardware failures within the Nexus 7000, redundant supervisors can be installed.

This design employs all three of the aforementioned methods of high availability in the core layer of the data center.

Aggregation Layer Availability

To achieve high availability in the aggregation layer, many of the features used for availability in the core layer are utilized in addition to some key features available with the Nexus 7000:

- Virtual Port Channels (VPC)—Virtual port channels can be used to connect the aggregation layer to the access or services layer. VPC allows for redundant paths between entities while simultaneously removing the sub-optimal blocking architecture associated with traditional spanning tree designs. The details regarding the configuration and options to enable vPC can be found at: http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9670/configuration_guide_c07-543563.html.
- Virtual Route and Forwarding (VRF)—Redundant pairs of VRF instances provide Layer 3 services for their associated tenant VLAN segments.
- First hop redundancy—HSRP can be used to provide gateway redundancy for the edge devices in the data center. Each switch can become an HSRP peer in however many groups are required for the design. Ideally, each tenant would have an HSRP group.

Services Availability

High availability for the services layer can be achieved whether using appliances or the service chassis design. Appliances can be directly attached to the aggregation switches or to a dedicated services switch, usually a Cisco 6500 series switch. The service chassis design involves using modules designed for the 6500 series chassis.

If using appliances, high availability can be achieved by logically pairing two physical service devices together. The pairs can be used in an active/standby model or an active/active model for load balancing, depending on the capabilities of each appliance pair. Certain service appliances, such as the ASA and ACE, can load balance by dividing their load among virtual contexts that allow the appliance to act as multiple appliances. This can also be achieved with their service module counterparts. This can be particularly valuable in a tenant environment where it is desirable to present each tenant with their own appliance to ensure separation.

The same can be achieved using the service chassis design, but HA would also need to be implemented at the service chassis level. An ideal way to implement this would be to use the Virtual Switching System (VSS). With VSS, redundant modules can be paired across different chassis, but management can be simplified by presenting the two chassis as one to the administrator. Details regarding VSS can be found at: http://www.cisco.com/en/US/products/ps9336/products_tech_note09186a0080a7c837.shtml.

This design uses a services chassis design with VSS implemented. Service modules that support virtual contexts, such as the Firewall Services module and the Application Control Engine module, are implemented in an active/active failover pair. The Intrusion Prevention System (IPS) appliance is also used. A pair of IPS appliances is attached to both switches in the VSS pair. Etherchannel load balancing is configured on the 6500 switch and IPS appliances to insure link resiliency. The Cisco IPS appliances support virtualization allowing for operational independence and more granular traffic inspection rules per device context.

Access Layer Availability

Access layer is designed with the following key design attributes in Nexus 5000:

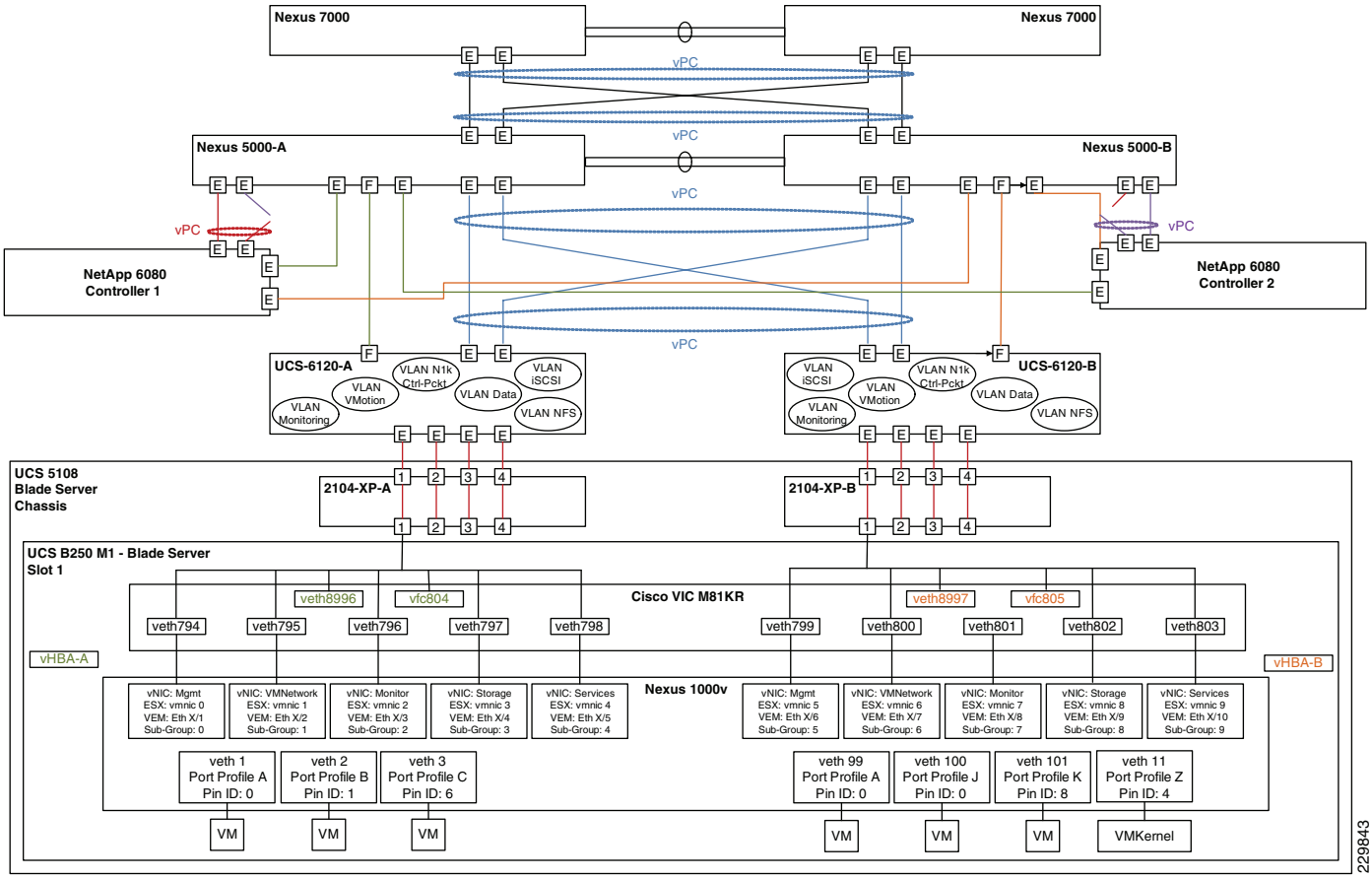
- Enables loop-less topology via Virtual Port-Channel (vPC) technology. The two-tier vPC design is enabled such that all paths from end-to-end are available for forwarding (see Figure 19).
- Nexus 7000 to Nexus 5000 is connected via a single vPC between redundant devices and links. In this design four 10Gbps links are used, however for scalability one can add up to eight vPC members in the current Nexus software release.
- The design recommendation is that any edge layer devices should be connected to Nexus 5000 with port-channel configuration.
- RPVST+ is used as spanning tree protocol. MST option can be considered based on multi-tenant scalability requirements. Redundant Nexus 7000 is the primary and secondary root for all VLANs with matching redundant default gateway priority.

Edge Device Layer Availability

The edge device connectivity consists of all devices that connect to the access layer. This includes the Cisco UCS, Nexus 1000V, and NetApp FAS 6080. The network availability for NetApp FAS 6080 is covered in Design Considerations for SAN Availability and FCoE and hence is not addressed here. Depending on the hardware and software capability, many design choices are possible with UCS and Nexus 1000V.

Most of the designs options and some of the best practices are described in the white paper at: http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9902/white_paper_c11-558242_ns944_Networking_Solutions_White_Paper.html. This white paper provides a foundation for understanding UCS and Nexus 1000V design choices; however this design guide covers newer options available in Nexus 1000V software and Cisco UCS hardware. This document may supersede or suggest a design change based on the requirements of the Enhanced Secure Multi-Tenancy design. Figure 21 depicts the UCS and Nexus 1000V connectivity for a multi-tenant environment and the design attributes that follow.

Figure 21 Edge Layer Connectivity



Unified Computing System:

- Fabric Availability—The UCS provides two completely independent fabric paths A and B. The fabric failover is handled at the Nexus 1000V level and thus not used in this design.
- Control Plane Availability—The UCS 6100 is enabled in active/standby mode for the control plane (UCS Manager) managing the entire UCS system.
- Forwarding Path Availability—Each fabric interconnect (UCS 6100) is recommended to be configured in end-host mode. Two uplinks from each UCS 6100 are connected as port-channel with LACP “active-active” mode to Nexus 5000.
- Blade Server Path Availability—Each blade server is enabled with M81KR VIC (Virtual Interface Card) providing 20Gbps connectivity to each fabric, capable of supporting up to 58 PCIe virtual NIC or HBA interfaces.

Nexus 1000V:

- Supervisor Availability—The Virtual Supervisor Module (VSM) is a virtual machine which can be deployed in variety of ways. In this design guide, it is deployed under UCS blade along with Virtual Ethernet Module (VEM). The Nexus 1000V and the Nexus 1010 Virtual Services Appliance supports redundant VSMs. The active and standby are recommended to be configured under separate UCS blade servers with the anti-affinity rule under vCenter such that the VSMs can never be operating under the same blade server.

- Forwarding Path Availability—Each ESXi host runs a VEM, which is typically configured with two uplinks connected to 10Gbp interface of the blade server. When connected to vCenter, the port-profile designated for uplinks automatically creates port-channel interface for each ESXi host.
- With the Cisco M81KR VIC, these uplinks are refined to multiple virtual interfaces that are provisioned through UCSM to provide up to 58 virtual interfaces that can be configured to Fibre Channel or Ethernet devices. These virtual uplinks can be used to create an added layer of traffic isolation and shaping. In this design, the VEM uses ten vNICs exposed as Ethernet interfaces to the hypervisor.
- The VSM creates Ethernet port-profile uplinks to map to the defined Ethernet M81KR virtual interfaces. These create port channels that do not run LACP and are not treated as host vPCs. This feature creates the source-mac based pinning to one of the uplinks and silently drops packet on other links for any packet with source MAC on that link. As a reminder the Nexus 1000V does not run spanning tree protocol and thus a technique is needed to make MAC address available via single path.
- The VLANs carried in through the Ethernet port-profile uplinks are broken up into vEthernet port-profiles specified as VMware port-groups. These vEthernet profiles are presented as distributed virtual port groups in the vNetwork Distributed Switch the VSM provides to vCenter.
- VLANs that need to be brought up before the VEM contacts the VSM are specified as system VLANs. Specifically, this includes the Control/Packet VLANs on the appropriate uplink(s), which is required for the VSM connectivity. It also applies to the ESXi management VLAN on the uplink if the management port is on the Nexus 1000V. These VLANs come up on the specified ports first, to establish vCenter connectivity and receive switch configuration data. On the ESXi host where the VSM is running, if the VSM is running on the VEM, the storage VLAN also needs to be a system VLAN on the NFS VMkernel port.
- Virtual Machine Network Availability—The port-profile capability of Nexus 1000V enables seamless network connectivity across the UCS domain and ESXi cluster. In this design guide, each virtual machine is enabled with three virtual interfaces, each inheriting a separate profile associated with a port channel. The profiles are designed with connectivity requirements and secure separation principles discussed in Secure Network Isolation. The front-end, back-end, and VM/application management traffic flows are separated with distinct traffic profiles. Using the sub-group ID associated with mac pinning and QoS service polices, tenant service-level strategies may be applied at the virtual port level.

Design Considerations for SAN Availability and FCoE

FCoE is the transport mechanism for Fibre Channel. Architectures leveraging FCoE must still adhere to best practices for Fibre Channel SAN design. Some issues to consider when designing an FCoE booted fabric include, but are not limited to, virtual SANs (VSANs), zone configurations, n-port virtualization, fan in/fan out ratios, high-availability (HA), and topology size. If they are configured incorrectly, each of these components can lead to a fabric that is not highly available due to the Fibre Channel requiring a loss-less nature. In this Enhanced Secure Multi-Tenancy architecture, an improperly configured SAN impacts the boot OS and in turn the tenant VMs and data sets. A basic understanding of Fibre Channel SANs is required for the design of the FCoE booted environment.

Cisco VSANs are a form of logically partitioning a physical switch to segment traffic based on design needs. By deploying VSANs, an administrator can separate primary boot traffic from secondary traffic, ensuring reliability and redundancy. Additionally, as deployments grow, subsequent resources can be placed in additional VSANs to further aide in any segmentation needs from a boot or data access perspective. For instance, as a multi-tenant environment grows beyond the capacity of a single UCSM, additional FCoE booted hosts can be added without impacting existing compute blades or deploying new switches dependent upon port counts.

Zoning within a fabric is used to prevent extraneous interactions between hosts and storage ports, which can lead to an abundance of initiator cross-talk. Through the creation of zones, which exist in a given VSAN, a single initiator port can be grouped with the desired storage port to increase security, improve performance, and help troubleshoot the fabric. A typical SAN booted architecture consists of redundant fabrics with primary and secondary boot paths constructed by using zones in each fabric.

As SANs grow, the amount of switches required to accommodate the port count increases. This is particularly true in legacy bladecenter environments because each Fibre Channel I/O module constitutes another switch to manage with its own security implications. From a performance perspective, this is a concern because each switch or VSAN within an environment has its own domain ID, adding another layer of translation. N-port ID Virtualization (NPIV) is a capability of the Fibre Channel protocol that allows multiple N-ports to share a single physical port. NPIV is particularly powerful in large SAN environments because hosts that log into an NPIV-enabled device are actually presented directly to the north-bound fabric switch. This improves performance and ease of management. NPIV is a component requirement of the northbound Nexus 5000. The UCS Fabric Interconnect operates in N-Port Virtualization mode (PV) meaning it does not operate as a FC switch or consume domain ID resources. The Fabric Interconnect allows the hosts to leverage the fabric without increasing its logical size.

The fan-in characteristics of a fabric are defined as the ratio of host ports that connect to a single target port. Fan-out is the ratio of target ports or LUNs that are mapped to a given host. Both are performance indicators, with the former relating to host traffic load per storage port and the latter relating to storage load per host port. The optimum ratios for fan-in and fan-out are dependent on the switch, storage array, HBA vendor, and the performance characteristics of the IO workload. High-availability within a FC fabric is easily attainable by using a configuration of redundant paths and switches. A given host is deployed with a primary and redundant initiator port, which is connected to the corresponding fabric. With a UCS deployment, a dual-port mezzanine card is installed in each blade server and matching vHBAs. In addition, boot policies are set up providing primary and redundant access to the target device. These ports access the fabric interconnect as N-ports, which are passed along to the northbound Nexus switch. Zoning within the redundant FC switches is configured so that if one link fails then the other handles the data access. Multipathing software is installed dependent on the OS to maintain LUN availability and load balancing. Through the use of UCS service profile templates, primary and secondary FCoE paths for each host can easily be configured to make sure that no given target port is over or underutilized.

When designing SAN booted architectures, considerations are made regarding the overall size and number of hops that an initiator would take before it is able to access its provisioned storage. The performance of a given fabric improves as the amount of hops and devices connected across a given inter-switch link (ISL) decreases. A common target ratio of hosts across a given switch link could be between 7:1 and 10:1, while an acceptable ratio may be as high as 25:1. This ratio can vary significantly depending on the size of the architecture and the performance required.

SAN and FCoE connectivity should involve or include:

- The use of redundant VSANs and associated zones
- The use of redundant ISLs where appropriate
- The use of redundant FCoE target ports
- The use of redundant fabrics with failover capability for Fiber Channel SAN booted infrastructure
- The use of NPIV enabled Nexus switches with the addition of the storage protocol’s license

Design Considerations for Storage Availability

Data Availability with RAID Groups and Aggregates

RAID groups are the fundamental building blocks when constructing resilient storage arrays containing any type of application data sets or virtual machine deployments. A variety of protection and cost levels are associated with different types of RAID groups. Selecting a storage controller that offers superior protection is an important decision to make when designing a multi-tenant environment because hypervisor boot LUNs, guest virtual machines, and application data sets are all deployed on a shared storage infrastructure. Furthermore, the impact of multiple drive failures is magnified as more data is housed on a given disk and disk size increases. Deploying a NetApp storage system with RAID-DP offers superior protection coupled with an optimal price point.

RAID-DP is a standard Data ONTAP feature that safeguards data from double-disk failure by means of using two parity disks. With traditional single-parity arrays, adequate protection is provided against a single failure event such as a disk failure or bit error during a read. In either case, data is recreated by using parity and data remaining on the unaffected disks. With a read error, the correction happens almost instantaneously, and often the data remains online. With a drive failure, the data on the corresponding disk must be recreated, which leaves the array in a vulnerable state until all data has been reconstructed onto a spare disk. With a NetApp array deploying RAID-DP, a single event or second event failure is survived with little performance impact, because a second parity drive exists. NetApp controllers offer superior availability while requiring fewer physical disks.

Aggregates are concatenations of one or more RAID groups that are then partitioned into one or more flexible volumes. Volumes are made available as file level (NFS or CIFS) mount points or they are partitioned into LUNs for block-level access (iSCSI, FCP, FCoE). With NetApp inherent storage virtualization, all data sets or virtual machines housed within a shared storage infrastructure leverage the benefits of RAID-DP. For example, with a maximum UCS deployment, 640 local disks (two per blade) could be configured in 320 independent RAID-1 arrays all housing the separate hypervisor OS. Conversely, using a NetApp array deploying RAID-DP, these OSs could be located within one large aggregate to take advantage of pooled resources from a performance and availability perspective.

Highly Available Storage Configurations

Inferior RAID configurations are detrimental to data availability. Similarly, the overall failure of the storage controller that serves data can be catastrophic. NetApp controllers deployed with RAID-DP and high-availability (HA) pairs provide continuous data availability for multi-tenant solutions. The deployment of an HA pair of NetApp controllers results in the environment being available both in the event of an unforeseen failure and when system upgrades are needed.

Storage controllers in an HA pair have the capability to seamlessly take over their partner's roles and activities in the event of a system failure. These include controller personalities, IP addresses, SAN information, and access to the data being served. This is accomplished through simple administrative setup, providing redundant paths to the storage from each controller and configuring the cluster interconnections. In the event of an unplanned outage, a node assumes the identity of its partner with no reconfiguration required by any attached hosts. HA pairs also allow non-disruptive upgrades for software installation and hardware upgrades. A simple command is issued to take over and give back controller identity.

Consider the following when deploying an HA pair:

- Best practices should be deployed to make sure that any node can adequately handle the total system workload.
- Storage controllers communicate heartbeat information using a cluster interconnect cable.
- The takeover process takes seconds.

- TCP sessions to client hosts are reestablished following a timeout period.
- Some parameters must be configured identically on each controller in the HA pair.

For additional information regarding NetApp HA pairs, refer to: http://media.netapp.com/documents/clustered.pdf.

Storage Network Connectivity (VIFs) Using LACP

A Virtual Interface (VIF) is a mechanism that allows the aggregation of a network interface into one logical unit. Combining links aids in network availability and bandwidth. NetApp provides three types of VIFs for network port aggregation and redundancy:

- SingleMode
- Static MultiMode
- Dynamic MultiMode

The secure, multi-tenant architecture leverages Dynamic MultiMode VIFs due to the increased reliability and error reporting and is also compatible with Cisco Virtual Port Channels. A Dynamic MultiMode VIF uses Link Aggregation Control Protocol (LACP) to group multiple interfaces together to act as a single logical link. This provides intelligent communication between the storage controller and the Cisco Nexus and enables load balancing across physical interfaces as well as failover capabilities.

Design Considerations for Tenant Availability

Tenant Availability

The overall tenant availability design is enabled by the following features/product components:

- VMware HA
- vCenter Heartbeat
- vMotion
- Storage vMotion

VMware HA

All ESXi Clusters (production, pre-production, and production failover reserve) are enabled with VMware HA for automatic restart of virtual machines in the event of host or virtual machine guest operating system failures, thereby reducing the impact of unplanned outages. The following are key design considerations for VMware HA in this architecture:

- **Primary HA Nodes**—The first five ESXi hosts added to the VMware HA cluster are primary nodes; subsequent hosts added are secondary nodes. Primary nodes are responsible for performing failover of virtual machines in the event of host failure. In large environments where the ESXi cluster spans multiple UCS blade chassis, ensure the first five ESXi servers are added in a staggered fashion across all available blade chassis. Doing so will prevent a single blade chassis failure, causing loss of HA functionality where virtual machines from the failed blade chassis do not get restarted automatically.
- **HA Admission Control Policy**—Three admission control policies dictate the amount of compute resource reservation set aside to accommodate host failures:

  - Host Failures Cluster Tolerates—This policy works well for environments that have virtual machines with identical set of CPU and memory resource requirements. In most enterprise environments, virtual machines often times require varying resource requirements, therefore this is the ideal admission control policy for a multi-tenant environment.
  - Percentage of Cluster Resource Reserved—With the Percentage of Cluster Resources Reserved admission control policy, VMware HA ensures that a specified percentage of aggregate cluster resources is reserved for failover. For the ESMT architecture, all production virtual machines (Exchange CAS, Hub and DAG VMs, SQL VMs and SharePoint DB, Index VMs) are configured with reserved CPU and memory resources. The key design consideration is to ensure the amount of CPU and memory resource reservation is greater than or equal to the total amount of "reserved" CPU and memory resources for all critical production VMs. For pre-production virtual machines or non-critical virtual machines with no resource reservation, the default amount of memory reserved is 0 MB and CPU reserved is 256MHZ.
  - Specify Failover Host—This policy puts a dedicated host to standby, as the failover host. To ensure that spare capacity is available on the failover host, HA prevents any virtual machines from being powered on or vMotion to the standby host. Additionally, DRS does not use the standby host for balancing. In a multi-tenant environment where virtual machines serve both production and non-production workloads, this policy is not ideal as the standby compute resources do not get utilized.

- **VM Restart Priority**—Ensure all production virtual machines have restart priority set to "High".
- **Host Isolation Response Setting**—This setting dictates whether HA leaves the virtual machines powered on or off in the event of host isolation, where the management network is unavailable. With the fully redundancy from Nexus 1010V to the Fabric Interconnect to access and aggregation layer, there is a very slim chance that the management network will become unavailable. However, for protection against a case where only the management network is unavailable while the production network is unaffected, change the default setting of "Shutdown" to "Leave powered on" for all critical/production virtual machines.
- **VM Monitoring**—Ensure all critical/production virtual machines have VM monitoring sensitivity level set to "High" or a custom value. The default values for VM monitoring are shown in Table 2.

Table 2 VM Monitoring Default Values

Setting	Failure Interval (seconds)	Reset Period
High	30	1 hour
Medium	60	24 hours
Low	120	7 days

- **Maximums**—It is important to stay within the boundary of maximum number of host/virtual machines per ESXi Cluster. Exceeding the boundary will result in virtual machines not getting restarted automatically by HA. vSphere 4.1 release has increased the scalability limits of VMware HA tremendously. The new maximums are the following:
  - 32 ESXi hosts
  - 160 virtual machines per ESXi host
  - Up to 3200 virtual machines per cluster

vCenter Heartbeat

vCenter servers in both data centers (Protected and Recovery) are configured with vCenter Heartbeat to ensure vCenter is protected against hardware and application failures. The following are key design considerations for vCenter Server availability in this architecture:

- **Separation of vCenter Server and SQL database**—Ensure each vCenter Server database runs on a separate virtual machine than vCenter Server; this is important for the scalability of the environment and also problem isolation.
- **Protection for all vCenter instances**—Ensure each vCenter Server and MS-SQL database instances in Protected and Recovery site are installed with vCenter Server Heartbeat.
- **Server anti-affinity**—vCenter Heartbeat has built-in protection against server failure. To guarantee protection, the primary and secondary vCenter and MS-SQL virtual machines must have anti-affinity rule established at the ESX cluster level. The anti-affinity rule will restrict both virtual machines from being powered on or migrated to the same ESX Server host.
- **Network Separation**—vCenter Heartbeat architecture has requirements for two separate networks, one for Principal (Public) interface where end users access vCenter Server and a second, the Channel interface, which is responsible for the heartbeat communication between the primary and secondary virtual machines. To fulfill this requirement, for each instance of vCenter Server virtual machine (both primary and secondary), the vNIC for the "Public Network" needs to be connected to the routable Management Port Profile/management VLAN and the vNIC for the "Physical Channel"/heartbeat needs to be connected to the non-routable Services Port Profile/Heartbeat VLAN.
- **Storage**—Each vCenter Server, SQL Server virtual machine instance requires its own virtual disk. All virtual disks assigned to vCenter Server and MS-SQL virtual machines need to reside on a datastore managed by the Management vFiler.
- **Split-brain avoidance**—Split-brain Avoidance ensures that only one server becomes active if the VMware Channel connection is lost, but both servers remain connected to the Principal (Public) network. Split-brain Avoidance works by pinging from the passive server to the active server across the Principal (Public) network. If the active server responds, the passive server does not failover, even if the VMware Channel connection is lost. With the Cisco UCS Palo adapter and active/active dual fabric connectivity, it is unlikely that the channel interface will lose complete communication, however it is still a good practice to enable Split-brain Avoidance feature.

vMotion

vMotion is a well-adopted feature in vSphere for elimination of application downtime. From an availability standpoint, ensure all infrastructure services virtual machines and tenant virtual machines are capable of getting migrated with vMotion between ESXi Servers. With the Cisco Nexus 1010V and Palo adapter, this is enabled easily by simply configuring the VMkernel vMotion interface to connect to the Services port profile/vMotion VLAN.

Storage vMotion

Storage vMotion is complementary to NetApp Data Motion capability. Storage vMotion addresses the need to live migrate tenant or infrastructure virtual machines between different tiers of storage or to migrate workload off datastores that are overly used. Data Motion addresses the need to migrate live individual vFiler units from one physical NetApp storage controller to another (i.e., equipment refresh or controller load balance). A key consideration for these two technologies is to ensure that Storage vMotion and Data Motion operations do not occur simultaneously.

vShield Manager Backup

vShield Manager has built-in backup capability to ensure firewall rules configured can easily be restored. As a best practice, perform a manual backup of vShield Manager after all firewall rules have been defined for the environment and after new rules have been added after new tenant instantiation. Regular scheduled backup should also be configured.

Tenant Backup, Recovery, and Replication

The Enhanced Secure Multi-Tenancy architecture has two parts for the storage of a tenant's virtual machines and applications. First, each tenant's guest OS and application binaries (that is, the virtual machine's C: drive) are installed within the vmdk virtual disk files on a tenant-specific NFS datastore. Each tenant's NFS datastore within ESXi is an NFS exported flexible volume on storage that is owned and managed by that tenant's vFiler unit. Second, any application-specific data (for example, databases, transaction logs, and so forth) is stored on iSCSI RDMs. These RDMs are also exported from that particular tenant's vFiler unit as iSCSI LUNs and are stored within separate, flexible volumes. These functions are split to provide support for data backup and replication methods for certain applications and protocols.

NetApp SnapManager for Virtual Infrastructure (SMVI) is used to back up, restore, and replicate tenant guest OSs and application binaries on NFS datastores. SMVI integrates with the VMware VSS requestor to automate the creation of crash-consistent Snapshot copies. SMVI also provides a method for the restore of virtual machines, virtual disk files, and individual files on virtual disks. For this reason, VMware tools must be installed on each virtual machine in the environment. Remote replication is achieved using NetApp SnapMirror, which is tightly integrated with NetApp SMVI. This makes it possible to invoke local Snapshot copies (backups) and replication of those Snapshot copies from a single user interface, using a single process.

SnapDrive and SnapManager are used to back up, restore, and replicate application-specific data stored on iSCSI RDMs. These applications are installed locally on the virtual machine running the application and are designed to automate the tasks of creating storage and protecting specific application data. SnapDrive performs the underlying storage operations, such as creating iSCSI LUNs, mapping iSCSI LUNs to virtual machines as iSCSI RDMs, creating Snapshot copies on the storage system, and so on. This requires the ability to talk to both the VMware vCenter Server and the NetApp vFiler unit to function properly. When configuring VMware vShield and other firewall or security measures, specific rules must be created to allow this type of communication. Multiple versions of SnapManager are available, each corresponding to a particular application:

- **NetApp SnapManager for Microsoft Exchange**—Creates on-disk backups of Exchange databases and triggers SnapMirror replication of RDM devices containing application data.
- **NetApp SnapManager for Microsoft SQL**—Creates on-disk backups of SQL databases and triggers SnapMirror replication of RDM devices containing application data.
- **NetApp SnapManager for Microsoft SharePoint Server**—Creates on-disk backups of SharePoint databases and triggers SnapMirror replication of RDM devices containing application data.

In an RDM-based solution such as this, application-consistent backups of data in the Exchange, SQL Server, and SharePoint VMs are created using NetApp SnapManager for Exchange, SQL, and SharePoint, respectively. These applications perform scheduled backups of the transaction logs and databases and, if configured, also initiate SnapMirror updates. The SnapManager products also provide granular recovery points for these Microsoft applications from within the same user interface.

Figure 22 Backup and Replication Summary

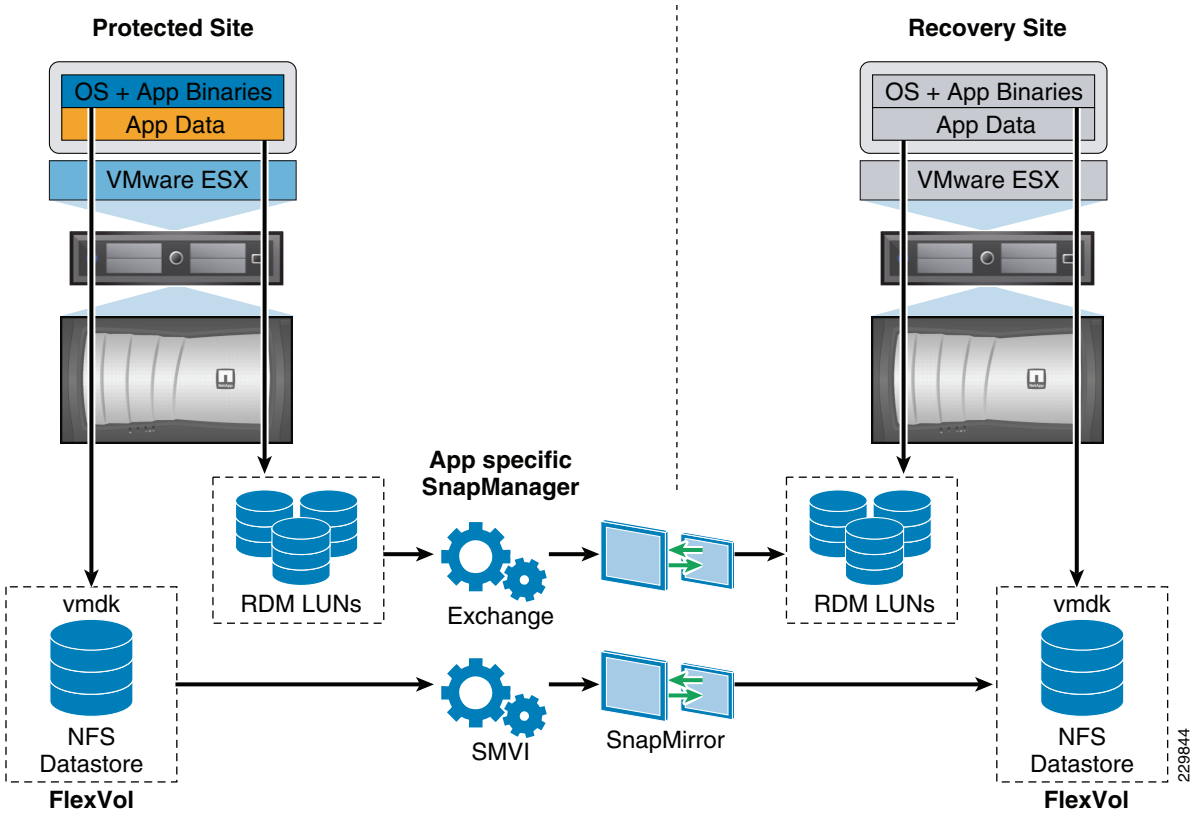
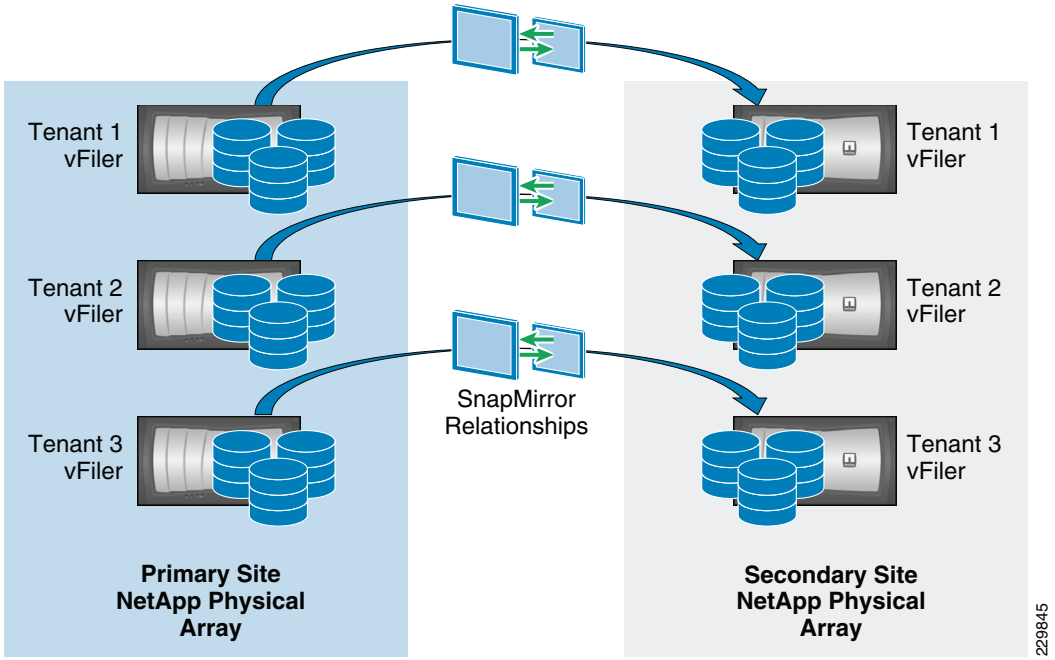


Figure 23 Tenant SnapMirror Relationships



Disaster Recovery

After data replication is initiated between the primary and secondary sites, VMware Site Recovery manager automates the processes involved in failing over to the secondary site should a disaster event occur. Leveraging the NetApp Adapter for Site Recovery Manager, SRM detects existing SnapMirror relationships within each vFiler unit and determines which virtual machines are associated within each relationship. Virtual machines can then be placed into Protection Groups that determine the order in which they are powered on in relation to other virtual machines. Based on these and other factors, SRM builds a customized Recovery Plan that outlines the necessary processes and workflows required to move from the primary site operation to the secondary site operation. Upon the execution of an SRM Recovery Plan, SRM will:

- Quiesce and break the NetApp SnapMirror relationships.
- Connect the replicated datastores to the ESX hosts at the DR site.
- If desired, power off the virtual machines, such as in testing and development instances, at the DR site, freeing compute resources.
- Reconfigure the virtual machines as defined for the network at the DR site.
- Power on the virtual machines in the order defined in the recovery plan.
- Execute any custom commands that have been stored in the recovery plan.

Figure 24 Pre-Failover Summary

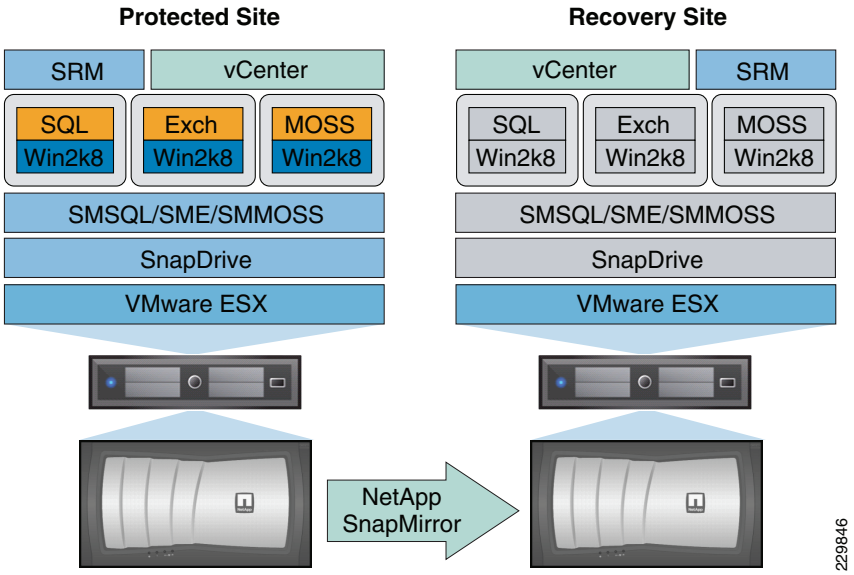
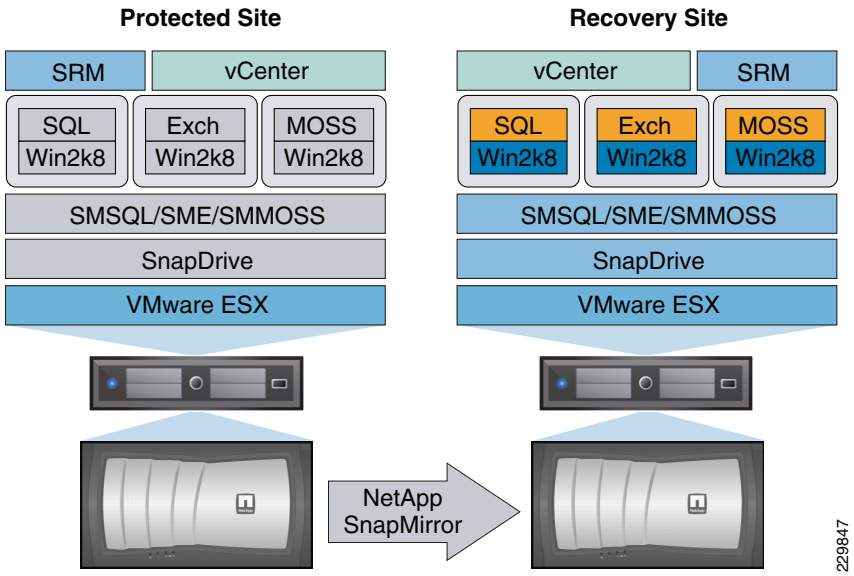


Figure 25 Post-Failover Summary



For more detailed information on this and other subjects, refer to the following documents and the NetApp Technical Library at http://www.netapp.com/us/library.

- TR-3822: Disaster Recovery of Microsoft Exchange, SQL Server, and SharePoint Server VMware vCenter Site Recovery Manager, NetApp SnapManager and SnapMirror, and Cisco Nexus Unified Fabric http://www.netapp.com/us/library/technical-reports/tr-3822.html
- TR-3785: Microsoft Exchange Server, SQL Server, and SharePoint Server Mixed Workload on VMware vSphere 4, NetApp Unified Storage (FC, iSCSI, and NFS), and Cisco Nexus Unified Fabric http://media.netapp.com/documents/tr-3785.pdf
- TR-3671: VMware vCenter Site Recovery Manager in a NetApp Environment http://media.netapp.com/documents/tr-3671.pdf

Infrastructure Tenant

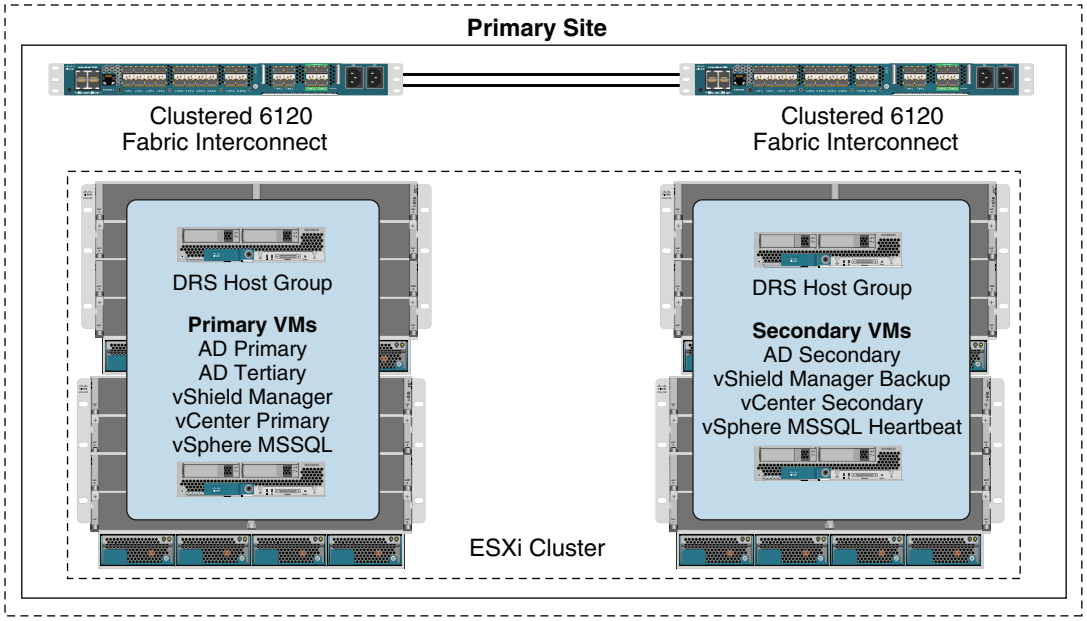
The infrastructure tenant is the foundation of the entire architecture because it serves as the basis for enabling the enterprise to offer IT as a Service to the different consumer of resources within an enterprise. Therefore, it is crucial to have a highly-available infrastructure for hosting all management services-related virtual machines. To achieve complete isolation between resources dedicated to infrastructure management and tenants, two groupings of compute, network, and storage resources are defined, Management Cluster and Resource Groups. Management Cluster is basically the Infrastructure tenant in this architecture, a dedicated ESXi cluster hosting service virtual machines that provide provisioning and management capabilities to the administrators. The Resource Groups consist of ESXi Servers configured as an ESXi Cluster, with resource pools carved out specifically for tenant application VMs, in this case, tenants 1, 2, and 3 (Exchange, MSSQL, and SharePoint) for the primary site. The same design principle is applied on the secondary site, where there is a dedicated ESXi Server cluster for management, production failover, and pre-production vCloud environments.

Management Cluster

Management components can be configured as a tenant within a production tenant or exist within a dedicated tenant cluster. In the design, management services were configured as clustered or as adjacent peer services. These clustered and peer service VMs are isolated to differing sets of hosts across the adjacent fabrics with DRS Host Groups to provide increased resiliency.

DRS Host Groups are a new feature with vCenter 4.1. Host Groups set an affinity for a listing of VMs to run across a defined list of hosts. By using Host groups, VMs are tied to a group of hosts that have a degree of isolation from each other, either through the chassis level or managing fabrics, creating an automatic anti-affinity between the VMs associated to these differing VM DRS groups. The breakdown of management resources across the cluster are shown in Figure 26.

Figure 26 Management Resources Across the Cluster



Network isolation and firewalling is provided as it is to other tenants. Additional Layer 2 isolation for cluster services and database transactions are implemented through vShield.

Infrastructure Tenant Recovery

The infrastructure tenant service virtual machines need not be replicated to the Secondary site. In the ESMT design, all service virtual machines are deployed in the same manner as the Primary site for consistency. For vShield Manager in particular, the recommended approach is to obtain all configured rules from the Primary site and re-apply them to the vShield Manager instance in the Secondary site. This can be done easily with the REST API call to get all the firewall rules on Primary site; invoke another REST API call to vShield Manager to apply the rules that have been obtained and saved. Doing so will ensure test failover or actual failover of production workload adheres to the same security policies that have been defined.

Secure Separation

Secure Separation ensures one tenant does not have access to another tenant's resources, such as virtual machine (VM), network bandwidth, and storage. Each tenant must be securely separated from every other tenant and from other dangers to the data center.

Key Threats in the Data Center

The threats that IT security administrators face today have grown from relatively trivial attempts to wreak havoc on networks to sophisticated attacks aimed at profit and theft of sensitive corporate data. Implementation of robust data center security capabilities within a multi-tenant environment to safeguard sensitive mission-critical applications and data is a cornerstone in the effort to secure enterprise networks.

The multi-tenant data center is exposed to threats from outside and from other tenants. Secure threats from other tenants are an additional security risk that requires mitigation.

Attack vectors have moved higher in the stack to subvert network protection and aim directly at applications. HTTP-, XML-, and SQL-based attacks are useful efforts for most attackers because these protocols are usually allowed to flow through the enterprise network and enter the intranet data center.

The following are some of the threat vectors affecting the multi-tenant data center:

- Unauthorized access
- Interruption of service
- Data loss
- Data modification

Unauthorized access can include unauthorized device access and unauthorized data access. Interruption of service, data loss, and data modification can be the result of targeted attacks. A single threat can target one or more of these areas. Specific threats can include privilege escalation, malware, spyware, botnets, denial-of-service (DoS), traversal attacks (including directory, URL), cross-site scripting attacks, SQL attacks, malformed packets, viruses, worms, and man-in-the-middle. In addition to these threats, many new threats are entering the enterprise network through legitimate applications, such as E-mail or through the Web. Viruses, spam, and malware are examples of such threats. These threats can significantly decrease user productivity, lead to loss of data, and cause sensitive information to be compromised.

Table 3 summarizes the threats in the data center and the network components and services that can be leveraged to mitigate those threats. The remainder of this section discusses these services in a multi-tenant design.

Table 3 Data Center Threats and Mitigation Resources

Network Services	Botnets	DOS	Spyware/ Malware	Data Leakage	Visibility	Network Abuse	Control
Policy Enforcement					Yes	Yes	Yes
Application Control Engine (ACE)		Yes		Yes	Yes		Yes
IPS integration	Yes		Yes		Yes	Yes	Yes

Table 3 Data Center Threats and Mitigation Resources

Network Services	Botnets	DOS	Spyware/ Malware	Data Leakage	Visibility	Network Abuse	Control
Switching Security		Yes		Yes		Yes	
Secure Device Access					Yes	Yes	Yes
Telemetry	Yes	Yes			Yes	Yes	
Firewall Services	Yes	Yes			Yes	Yes	Yes

Security Architectural Framework

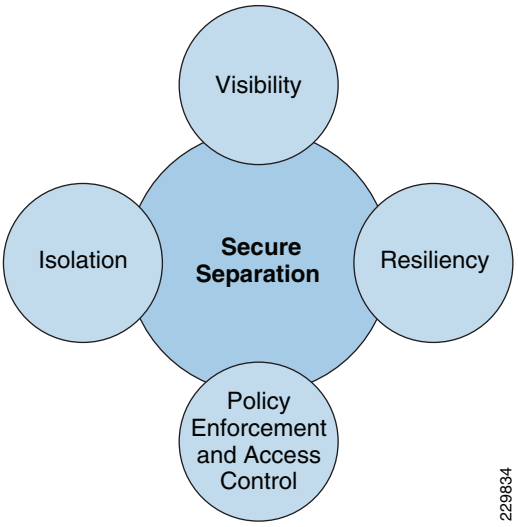
Introduction to Principals

Secure separation is the partition that prevents one tenant from having access to another's environment and also prevents a tenant from having access to the administrative features of the cloud infrastructure. The following briefly describes the main security principals that are implemented in this architecture.

- Isolation—Isolation can provide the foundation for security for the multi-tenant data center and server farm. Depending on the goals of the design, it can be achieved through the use of firewalls, access lists, VLANs, virtualization, storage, and physical separation. A combination of these can provide the appropriate level of security enforcement to the server applications and services within different tenants.
- Policy Enforcement and Access Control—Within a multi-tenant environment, the issue of Access Control and Policy Enforcement looms large and requires careful consideration. Capabilities of devices and appliances within each layer of the architecture can be leveraged to create complex policies and secure access control that can enhance secure separation within each tenant.
- Visibility—Data centers are becoming very fluid in the way they scale to accommodate new virtual machines and services. Server virtualization and technologies, such as VMotion, allow new servers to be deployed and to move from one physical location to another with little manual intervention. When these machines move and traffic patterns change, it can create a challenge for security administrators to actively monitor threats within the infrastructure. This architecture leverages the threat detection and mitigation capabilities that are available at each layer of the network to gather alarm, data, and event information and dynamically analyze and correlate the information to identify the source of threats, visualize the attack paths, and suggest and optionally enforce response actions.
- Resiliency—Resiliency implies that end-points, infrastructure, and applications within the multi-tenant environment are protected and can withstand attacks that can cause service disruption, data enclosure, and unauthorized access. Proper infrastructure hardening, providing application redundancy, and implementing firewalls are some of the steps needed in order to achieve the desired level of resiliency.

Figure 27 shows the security architecture framework implemented in this design.

Figure 27 Security Architecture Framework



The multi-tenant design outlined in this document implements the above described services into the multi-tenant end-to-end architecture and incorporates the security based design practices as outlined in the Cisco SAFE security reference architecture (http://www.cisco.com/en/US/docs/solutions/Enterprise/Security/SAFE_RG/SAFE_rg.html).

Mapping of Security Principals to Features

Secure separation is one of the four main pillars within a multi-tenant environment. The network, compute, storage, and management components within this architecture provide features and capabilities which together form the security framework and ensure secure separation within tenants. The following section provide a mapping of features of the network, storage, compute, and associated management elements needed to implement the corresponding security principals within the multi-tenant architecture.

Table 4 Mapping of Security Principals to Features

Principal	Network	Compute	Storage
Isolation	ASA/FWSM firewall services Virtual Firewall Nexus 1000V security Features ACE virtual contexts	vSphere/vCenter vShield Edge and App	NetApp Data ONTAP, MultiStore, IPSpaces and VLAN interfaces
Visibility	Netflow, ERSPAN Syslog ASA/FWSM/IPS/ACE Event Notifications ASA/FWSM/IPS/ACE Telemetry Data Export	vShield Manager Logging and monitoring capabilities Virtual firewall Logging	NetApp Data ONTAP audit logs

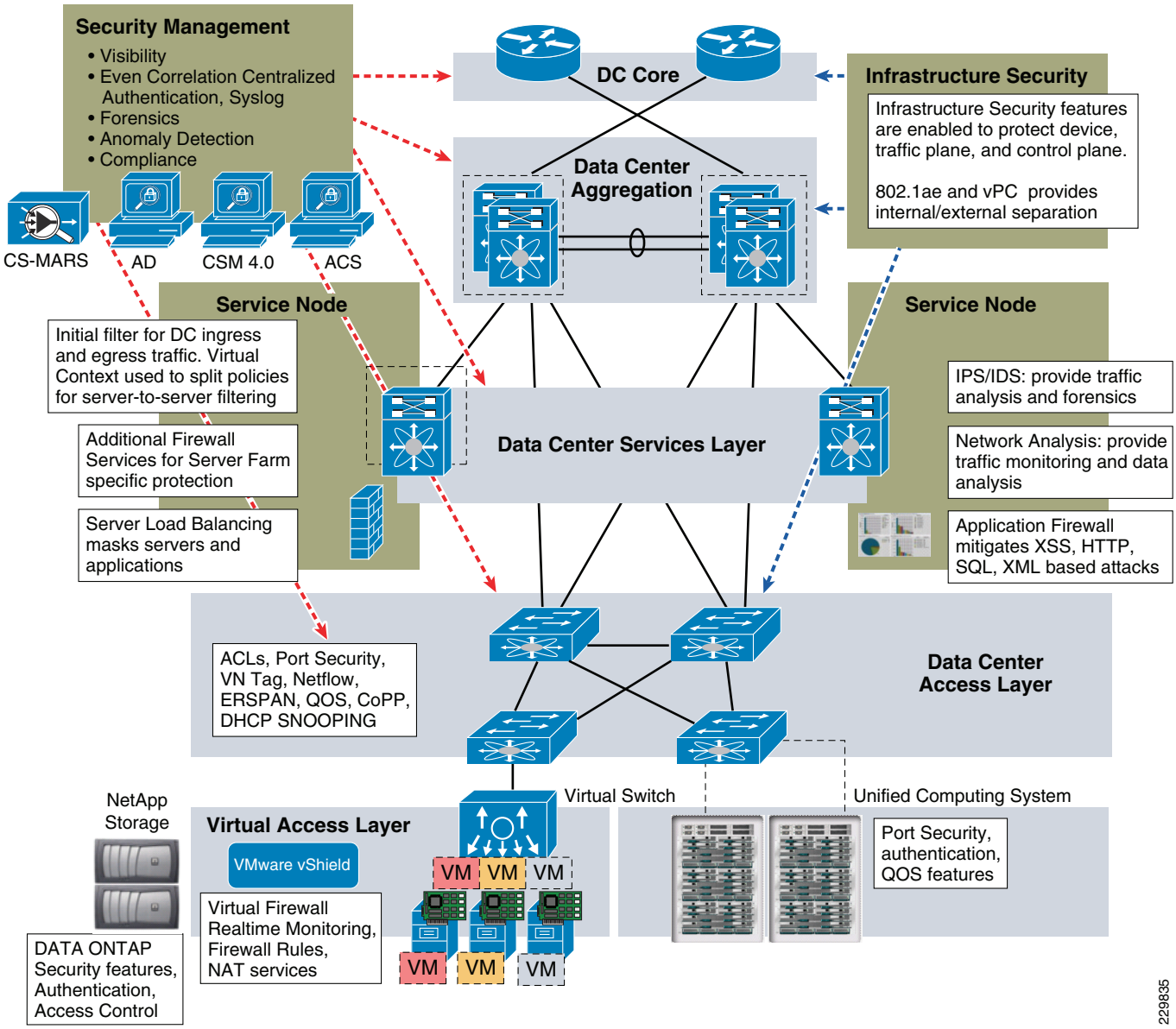
Table 4 Mapping of Security Principals to Features

Policy Enforcement /Access Control	Active Directory Services Cisco ACS device management services RBAC features on UCS	vSphere/vCenter RBAC and Access Control capabilities	NetApp Data ONTAP, LDAP, and Microsoft Active Directory support with RBAC
Resiliency	Device Hardening ACE Loadbalancing and Offload Services	vShield Edge and App Firewall Rules	NetApp Data ONTAP advanced settings and options

Architectural Framework

Figure 28 illustrates the security architectural framework used in this design. Figure 28 highlights the functional areas of the solution, its components, and their corresponding security features end-to-end.

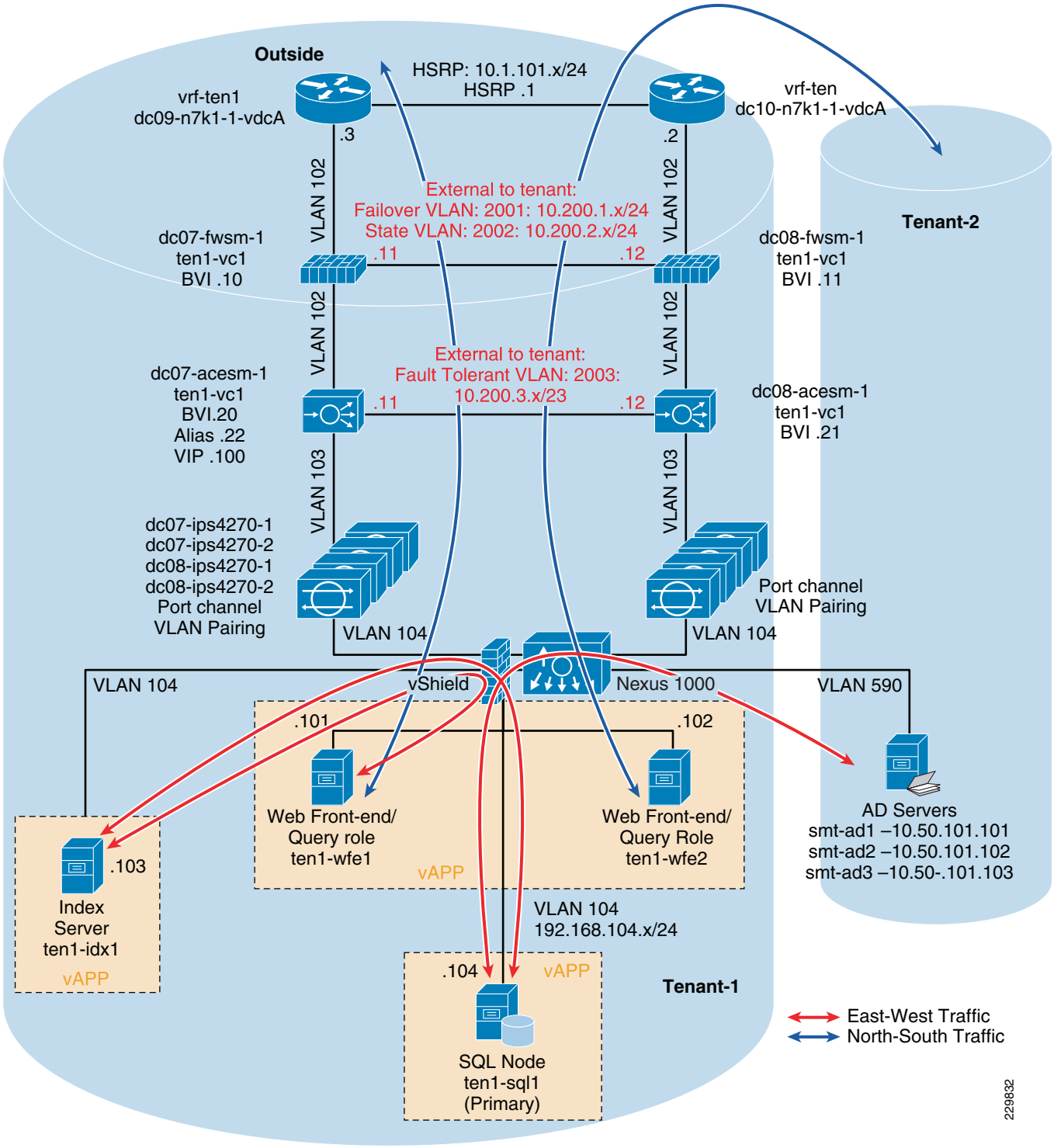
Figure 28 Enhanced Secure Multi-Tenancy Architectural Framework



Traffic Flow Models

To create an Enhanced Secure Multi-Tenancy infrastructure one needs to address the different traffic patterns existing within the environment. Recognition of these flows leads to the development and deployment of a comprehensive set of security policies. The traffic flows within this shared multi tenant infrastructure can be divided into two distinct categories, north-south and east-west.

Figure 29 *Sample Traffic Flows within a Multi-Tenant Architecture*



Client-to-Server—North-South Flows

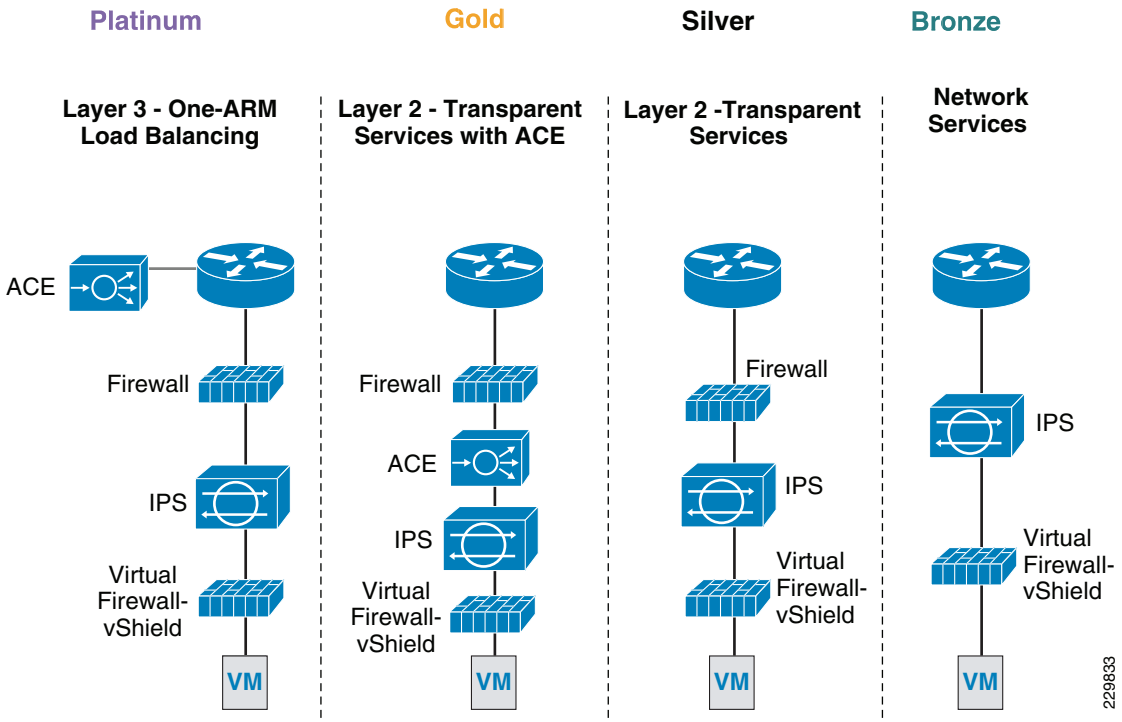
North-south traffic flows are either ingress or egress in relation to the data center and are commonly understood as client-to-server in nature. This traffic traverses the data center and is readily exposed to any number of services in its path including firewalling, load balancing, intrusion detection, and network analysis devices. In the multi-tenant environment, traffic between tenants may also be forced through the data center network services. This functionality is ideal as each tenant policy is uniformly applied between each other. The exposure of ingress-egress traffic flows to security services in the data center is dependent upon application specific requirements and the overall security policies of the enterprise.

Figure 29 highlights the client-to-server traffic flows and various services that may be applied to each flow in and out of the data center or between tenants. In this example, client-to-server traffic traverses the Nexus 7000 aggregation layer and a select number of network-based services. As shown in Figure 29, inter-tenant traffic may also cross the Nexus 7000 and in this case is categorized as a north-south communication flow.

In a multi-tenant environment, different tenants may require different levels of security protection. Tenants with very stringent security requirements will require a host of virtual and physical appliances to satisfy their needs, while other tenants may require basic network protection. For example, the human resource department would require a more stringent security infrastructure than the development-test organization. The capability to provide different levels of security protections as an add-on for different tenants needs to be planned up front and the design needs to be tightly integrated with the overall network architecture. This architecture leverages Cisco FWSM, Cisco ACE, Cisco IPS, and VMware vShield to augment solution security.

It is important to note the flexible nature of this architecture, where the security architect can use any combination of security appliances to create his/her own security offerings. The arrangements in Figure 30 illustrate some possible examples of the different tiers of security service offerings available and their corresponding network components.

Figure 30 North-South Service Model Examples



In all of the examples shown in Figure 30, the tenants have virtual firewall services and intrusion protection services. The virtual firewall provides protection and isolation at the access layer for virtual machines and the Intrusion Protection System (IPS) provides protection from known network worms and viruses, DoS traffic, and directed hacking attacks. This functionality is highly beneficial in an Enhanced Secure Multi-Tenancy environment by quickly identifying attacks and providing forensic information so that attacks can be cleaned up before substantial damage is done to network assets. IPS is designed to monitor and permit all traffic that is not malicious and can be deployed within the environment to cover all of the network segments. By deploying IPS at the edge of the network one can thwart attacks and unwanted traffic is stopped before entering the network.

To achieve a higher level of security and protection, one can deploy physical firewalls at the edge as well as the Cisco Application Control Engine (ACE). The ACE service module can be used to scale Web applications and servers. It has security benefits as well, where the ACE can protect against DoS attacks and protect Web servers by masking the servers' real IP addresses and providing a single IP address for clients to connect over a single or multiple protocols such as HTTP, HTTPS, and FTP. In this architecture, the firewalls and the IPS are deployed in transparent mode and the ACE can be deployed in transparent mode or in one-Arm configuration mode.

Note Figure 30 is an example and not a prescription for network-based services. There are numerous implementation models that may be used in the data center to apply network services. For more information on service integration in the data center, review Data Center Service Integration: Service Chassis Design Guide at http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/dc_servchas/service-chassis_design.html#wp58871.

Server-to-Server—East-West Flows

East-west traffic refers to the communication between servers within the data center access layer; it is commonly referred to as server-to-server traffic. Securing inter-server communication can be an application-based requirement or an enterprise-based requirement. Typically, enterprise-class applications require more availability, scalability, and/or processing power than a single server instance can provide. To address these issues, application developers use dedicated server roles. Each role is specialized and dependent on other servers to complete their function. Virtual machines fully support this application model. In the shared infrastructure of secure multi tenancy, server-to-server flows between virtual machines may occur within a single tenant container or between tenants.

To optimize east-west traffic patterns within a virtualized data center, it is recommended to use a virtual firewall appliance such as vShield to provide secure connectivity between virtual machines. This service may securely support intra- or inter-tenant communication. For example, a virtual firewall can provide secure connectivity for tenant virtual machines which need access to infrastructure services such as Active-Directory residing in a common infrastructure tenant. Figure 29 illustrates secure communication between servers in a single tenant and between tenants.

There are three primary ways to implement a virtual firewall:

- The first method is to use traditional virtual firewall model, where the virtual firewall can either block traffic between tenants and virtual machines residing in different VLANs or the firewall can be configured to allow specific pre-determined traffic flows. The traffic flows that are allowed across VLANs still need to be routed at the aggregation layer by the Nexus 7000. This functionality can be implemented by using virtual firewall rules in vShield Apps.

- Another way to implement east-west traffic flows is to use a virtual firewall to perform NAT translation between two network segments while simultaneously using the firewall capability to only allow certain traffic flows. This eliminates the need to use routing in the aggregation layer. This functionality can be implemented by using more sophisticated virtual firewalls like the vShield Edge.
- It may be necessary to create firewall rules between virtual machines within the same broadcast domains. The firewall rules in this case will be applied to specific virtual machines, rather than networks. The advantage of this method is that one can reduce the number of VLANs within the system and the traffic flow is restricted within the Layer 2 switching infrastructure. This capability can be achieved by configuring vShield Apps.

Design Considerations for Tenants

The following sections discuss the security features within the compute layer to augment the security discussed earlier and implemented within the network domain of the data center. The security implemented at this layer considers the same traffic models north-south and east-west.

Perimeter Security for Tenants

A typical multi-tier application topology is shown in Figure 34. The vShield Edge provides network address translation service between the Production tenants and the Infrastructure tenant VMs, such as DNS and Active directory services. This allows for secure separation of tenant while allowing the operational efficiency of centralized authentication. One can use the vShield realtime flow monitoring capability to ascertain the ports that are active between those two VLANs and create firewall rules to lock down the traffic accordingly. Vshield Edge can also be used to monitor and restrict traffic between the external interface and internal segments. This monitoring would complement the monitoring and firewall capabilities by the physical FWSM/ASA and IPS appliances.

Internal Zoning for Tenants

The ability to enforce security policies within a single tenant environment is critical to properly manage and control traffic between server roles.

The example n-tier application consists of a database, Web tiers and an index server. vShield App vNIC-level virtual machine firewall allows us to carve up VLAN 106 into multiple zones for the Web Front ends, the Index server, and the Database backend server. The advantage of vShield app is that one can create firewall rules between virtual machines in the same VLAN. This would simplify the complexity of the virtual network and reduce the number of VLANs needed to create these three zones. The vShield App firewall rules are created using either custom vNIC Security Groups or via the use of vCenter containers. For example, it is possible to place the Web front end virtual machines into a vApp (collection of virtual machines) and create vApps for the remaining zones. Rules can be specified using the vApp names in the source and destination IP address fields. These are dynamic vCenter containers which would allow any of the tiers of the SharePoint application to be scaled by spinning up new virtual machines and they would automatically receive rules that belong to their tier. An example of a container-based rule would be from SharePoint Web front end vApp to Database backend vApp allow TCP/1433 (MS SQL DB port).

Visibility

In an Enhanced Secure Multi-Tenancy architecture, an accurate view of the infrastructure is critical. Visibility implies access to an infrastructure-wide intelligence that provides an accurate vision of network topologies, attack paths, and the extent of the damage. Traditional point security solutions are incapable of meeting the needs of an increasingly sophisticated multi-tenant environment. An Enhanced Secure Multi-Tenancy environment requires a higher degree of visibility that is only attainable with infrastructure-wide security intelligence and collaboration. This implies the collection of various forms of network telemetry present on networking equipment, security appliances, storage clusters, and virtual compute and user endpoints to obtain a consistent and accurate view of the activity within each tenant.

As part of the event monitoring, analysis, and correlation, logging and event information generated by routers, switches, firewalls, storage arrays, intrusion prevention systems, and virtual protection appliances are collected, trended, and correlated. The architecture also uses the collaborative nature between security platforms such as intrusion prevention systems, firewalls, and virtual firewalls, switches, and appliances. Visibility can effectively be implemented by performing the following:

- Identify threats—Collecting, trending, correlating, and logging event information help identify the presence of security threats, compromises, and data leak.
- Confirm compromises—By being able to track an attack as it transits the network, and by having visibility on the endpoints, the architecture can confirm the success or failure of an attack.
- Reduce false positives—Endpoint and system visibility help identify whether a target is in fact vulnerable to a given attack.
- Reduce volume of event information—Event correlation dramatically reduces the number of events, saving security operators precious time and allowing them to focus on what is most important.
- Determine the severity of an incident—Enhanced endpoint and network visibility allows the architecture to dynamically increase or reduce the severity level of an incident based on the degree of vulnerability of the target and the context of the attack.
- Reduce response times—Having visibility over the entire network makes it possible to determine attack paths and identify the best places to enforce mitigation actions.

Network Visibility

Monitoring the network infrastructure is the most critical part of achieving optimum visibility. Server virtualization brings new challenges for visibility into what is occurring at the virtual network level. Traffic flows can now occur within the server—between virtual machines—without needing to traverse a physical access switch. If a virtual machine is infected or compromised, it might be more difficult for administrators to spot without the traffic forwarding through security appliances. The network infrastructure implemented in this guide leverages features and capabilities at all layers of the network to provide a consistent and accurate view of the state of the network.

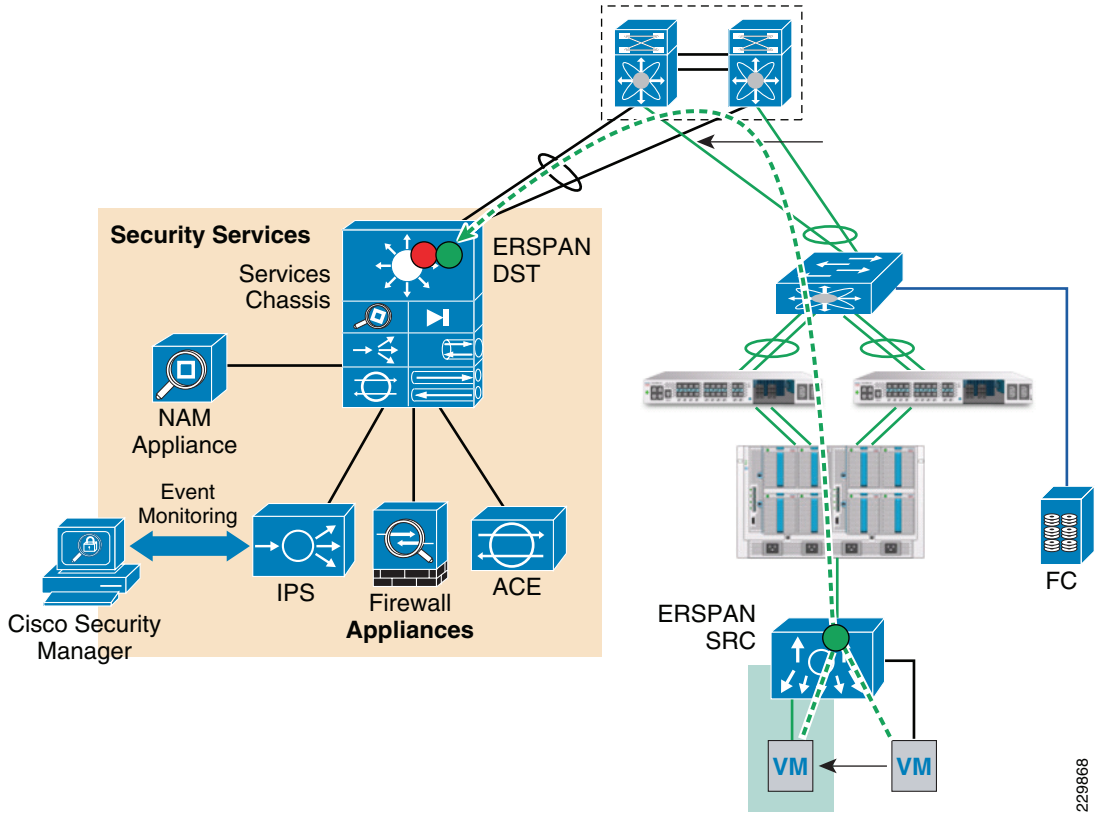
The following is a list of features and network components that are used to provide visibility and monitoring capability within the Enhanced Secure Multi-Tenancy architecture.

- Network Analysis Module/Appliance
- Encapsulated Remote Switched Port Analyzer (ERSPAN) is a very useful tool for gaining visibility into network traffic flows. This feature is supported on the Cisco Nexus 1000V. ERSPAN can be enabled on the Cisco Nexus 1000V and traffic flows can be exported from the server to external devices.
- Cisco Intrusion Prevention System (IPS) provides deep packet and anomaly inspection to protect against both common and complex embedded attacks. The IPS devices used in this design are Cisco IPS 4270s with 10-Gigabit Ethernet modules.
- Cisco Security Manager 4.0 (CSM)—Although Cisco Security manager is mainly a configuration tool, the CSM 4.0 software release provides event management capabilities that enhance network visibility. In this design the following event monitoring capabilities of CSM are leveraged.
  - Receipt of syslog messages from Cisco ASA appliances and Security Device Event Exchange (SDEE) messages from Cisco IPS sensors
  - Real-time and historical event viewing
  - Cross-linkages to firewall access rules and IPS signatures for quick navigation to the source policies
  - Pre-bundled set of views for firewall and IPS monitoring

- Customizable views for monitoring select devices or a select time range
- Administrative options to turn event collection on or off for select security devices

Figure 31 illustrates the capability to use ERSPAN and Cisco Security Manager to monitor events and traffic flows within the network.

Figure 31 Nexus 1000V Network Visibility



ERSPAN uses GRE tunnels to route traffic to the appropriate destination. The Nexus 1000V supports ERSPAN, allowing network administrators to observe the traffic associated with the following:

- The individual vNIC of a virtual machine connected to a VEM
- The physical ports associated with the ESX host
- Any port channels defined on the VEM
- Any VLAN that is defined on the VEM

In this design Nexus 1000V is configured with the capability to forward monitored traffic from any VLAN or groups of VLANS to the NAM appliance, via the service chassis, using ERSPAN. This approach provides the flexibility to monitor individual VLANs or groups of VLANS.

Vshield Manager

vShield Manager provides a great level of visibility into both the system operational states (captured by system events), configuration changes performed by all users (audit logs), and virtualization layer intra-host and inter-host networking communications (traffic flow). Each of these is made visible in the vShield Manager user interface and remote syslog server. It is recommended to configure vShield to send event notifications to a remote syslog server to keep trails of all information.

System Events

System events are events that are related to vShield App operation. System events are raised to detail every operational action, such as a vShield App reboot or a break in communication between a vShield App and the vShield Manager. A report is available at a per-vShield App level and the level of event severity can be filtered in the vShield Manager UI. By default, all levels of events are captured (both informational and critical); it is recommended to stay with default settings to ensure all events are captured for audit and troubleshooting.

Audit Logs

The Audit Logs capture all actions performed by vShield Manager users. This information is readily available in the vShield Manager UI. A log of firewall operation based on matching firewall rules against traffic is also available for each vShield App instance. The logs can be downloaded via the vShield Manager interface for each vShield App.

Traffic Flow

Traffic Flow Monitoring is a traffic analysis tool that provides a detailed view of the traffic on a virtual machine network that passed through a vShield App. The Flow Monitoring output defines which machines are exchanging data and over which applications. This data includes the number of sessions, packets, and bytes transmitted per session. Session details include sources, destinations, direction of sessions, applications, and ports being used. These session details can be used to create App Firewall allow or deny rules. This level of visibility for the virtual machine east- and west-bound traffic enables the cloud administrator to ensure 100% separation based on policies. Any unwanted virtual machine communication detected could be corrected on the fly by adding deny rules from the established session.

Storage Visibility

NetApp Data ONTAP provides audit log capabilities that capture tasks performed using administrative access from either the console or remote shell sessions. While virtual storage controllers do not provide direct console or remote shell sessions, any tasks performed on a virtual storage controller are also logged in the centralized audit log. Should corporate policies dictate centralized audit log collection, NetApp also supports the transfer of audit logs to a remote syslog host. Because a single audit log exists between the physical and virtual storage controllers, third-party software can easily be integrated to parse the single audit log into separate logs for the cloud administrator and individual tenants.

Resiliency

Resiliency is the ability to cope, adapt, and overcome challenges. Resiliency is a characteristic of a well-designed data center where security threats or high stress loads challenge the availability of enterprise applications. This section of the document addresses the ability of the secure multi tenant environment to withstand these dangers and adverse conditions.

Network Layer Resiliency

Effective multi-tenant security demands an integrated defense-in-depth approach that includes the hardening of network infrastructure against security threats and Denial of Service (DOS) attacks. Network resiliency implies the implementation of best practices to preserve the resiliency and survivability of routers and switches, helping the network maintain availability even during the execution of an attack. This can be achieved by implementing the following:

- Routing Protocol Hardening—Routing is one of the most important parts of the infrastructure that keeps a network running and, as such, it is absolutely critical to take the necessary measures to secure it. Attacks may target the router devices, the peering sessions, and/or the routing information. Fortunately, protocols like OSPF, EIGRP, and RIPv2 provide a set of tools that help secure the routing infrastructure. The routing infrastructure can be protected by implementing neighbor authentication, route peer definition, and route filtering:

- iACLs—Infrastructure protection access control lists (iACLs) are an access control technique that shields the network infrastructure from internal and external attacks. In a nutshell, iACLs are extended ACLs designed to explicitly permit authorized control and management traffic bound to the infrastructure equipment, such as routers and switches, while denying any other traffic directed to the infrastructure address space. In this design, iACls are best utilized when implemented at the core routers at the edge of the network infrastructure. By only allowing authorized control and management traffic, iACLs help protect routers from unauthorized access and DoS attacks based on unauthorized protocols and sources.
- Control Plane Policing (CoPP) and Control Plane Protection—Control Plane Policing and Control Plane Protection are security infrastructure features that allow the configuration of QoS policies that rate limit the traffic sent to the route processor in Cisco devices. Both features help protect routers from unauthorized access and DoS attacks, even when they originate from valid sources and for valid protocols. Both features also help protect routing sessions by preventing the establishment of unauthorized sessions and by reducing the chances for session reset attacks.

More complete design considerations for network device resiliency can be found at: http://www.cisco.com/en/US/docs/solutions/Enterprise/Security/Baseline_Security/securbase.pdf.

Compute Layer Resiliency

Infrastructure hardening is equally important in the virtualization layer as the physical layer. The following are common ones from the VMware vSphere Hardening Guide, vShield Administration Guide, and vCloud Director Administration Guide:

Virtual Machine:

- Disable Copy/Paste to Remote Console—When VMware Tools runs in a virtual machine, by default you can copy and paste between the guest operating system and the computer where the remote console is running. As soon as the console window gains focus, nonprivileged users and processes running in the virtual machine can access the clipboard for the virtual machine console. It is recommended that you disable copy and paste operations for the guest operating system.
- Ensure Unauthorized Devices are Not Connected—Ensure that no device is connected to a virtual machine if it does not need to be there. For example, serial and parallel ports are rarely used for virtual machines in a data center environment and CD/DVD drives are usually connected only temporarily during software installation. Try to limit new virtual machine deployment to be based on templates, to ensure consistency and less chance of human error.
- Limit VM log file size and number—Apply virtual machine settings to limit the total size and number of log files. Uncontrolled logging could lead to datastore getting filled up and result in denial of service. Normally a new log file is created only when a host is rebooted, so the file can grow to be quite large, but you can ensure new log files are created more frequently by limiting the maximum size of the log files. If you want to restrict the total size of logging data, VMware recommends saving 10 log files, each one limited to 1000KB. Datastores are likely to be formatted with a block size of 2MB or 4MB, so a size limit too far below this size would result in unnecessary storage utilization. Each time an entry is written to the log, the size of the log is checked, and if it is over the limit, the next entry is written to a new log. If the maximum number of log files already exists, when a new one is created the oldest log file is deleted. A denial of service attack that avoids these limits could be attempted by writing an enormous log entry, but each log entry is limited to 4KB, so no log files are ever more than 4KB larger than the configured limit. A second option is to disable logging for the virtual machine.
- Prevent Virtual Machines from Taking Over Resources—By default, all virtual machines on an ESX/ESXi host share the resources equally. By using the resource management capabilities of ESX/ESXi, such as shares and limits, you can control the server resources that a virtual machine consumes.

ESX/vCenter:

- Configure remote syslog—Remote logging to a central host provides a way to greatly increase administration capabilities. By gathering log files onto a central host, you can easily monitor all hosts with a single tool as well as do aggregate analysis and searching to look for such things as coordinated attacks on multiple hosts. Logging to a secure, centralized log server can help prevent log tampering and provides a long-term audit record.
- Configure NTP time synchronization—By ensuring that all systems use the same relative time source (including the relevant localization offset) and that the relative time source can be correlated to an agreed-upon time standard (such as Coordinated Universal Time-UTC), you can make it simpler to track and correlate an intruder's actions when reviewing the relevant log files.
- Ensure vSphere management traffic is on a restricted network.
- Ensure VMotion Traffic is isolated—The security issue with VMotion migrations is that information is transmitted in plaintext and anyone with access to the network over which this information flows may view it. Ensure that VMotion traffic is separate from production traffic on an isolated network. This network should be non-routable (no Layer 3 router spanning this and other networks), which will prevent any outside access to the network.
- Ensure IP- Based Storage Traffic is isolated—This type of configuration may expose IP Based Storage traffic to unauthorized virtual machine users. To restrict unauthorized users from viewing the IP Based Storage traffic, the IP Based Storage network should be logically separated from the production traffic. Configuring the IP Based Storage adapters on separate VLANs or network segments from the VMkernel management and service console network will limit unauthorized users from viewing the traffic.
- Ensure that port groups are not configured to the value of the native VLAN.
- Avoid user login to the vCenter Server system.
- Avoid user login to the ESXi Server directly (except for Cloud Administrator).
- Restrict usage of vSphere Administrator Privileges—By default, vCenter Server grants full administrative rights to the local administrators account, which can be accessed by domain administrators.

vShield Manager:

It is recommended to generate or import an SSL certificate into the vShield Manager to authenticate the identity of the vShield Manager Web service and encrypt information sent to the vShield Manager Web server. As a security best practice, you should use the generate certificate option to generate a private key and public key, where the private key is saved to the vShield Manager.

For additional security hardening guidelines, refer to the following documents:

- http://www.vmware.com/files/pdf/techpaper/VMware_vSphere_HardeningGuide_May10_EN.pdf
- http://www.vmware.com/pdf/vshield_41_admin.pdf
- http://www.vmware.com/pdf/vcd_10_admin_guide.pdf

Storage Layer Resiliency

A NetApp FAS controller in an Enhanced Secure Multi-Tenancy environment, similar to most security procedures and goals, focuses on restricting access to systems and data. This straightforward goal is important and requires following proper security best practices. These practices include a definitive set of policies, procedures, and significant planning. This section does not attempt to cover security in that level of detail. Rather, this section illustrates the security items pertinent to hardening a NetApp FAS controller.

Tenant users should not have direct access to the NetApp FAS controller. In other words, logins should not be mapped to, or created on, the storage meant for non-administrative users. Direct login access should be strictly provided only to the cloud administrator and tenant administrators roles. Even then, the cloud administrator is the only role with complete access to the NetApp FAS controller. Tenant administrators only need to see and affect change on their own vFiler units. Additionally, cloud and tenant administrators should be granted access based on a group policy as opposed to a shared login that many people use. This enables tracking the specific actions of specific users. Do not create a single "administrator" account to which multiple people have access.

By default, administrators can login from any network and any host. This needs to be secured as soon as the administrator accounts are created. Plan a limited set of management networks and/or administrative hosts, such as one per vFiler unit and one for the physical NetApp FAS controller. Allow administrative access only from those networks. Finally, the administrative access should use LDAP over SSL for centralized authentication and authorization in favor of NIS.

After establishing the tenant administrators, the next design consideration concerns the use and planning of VLANs. VLANs are one of the primary means of segregating traffic at the network layer in the Enhanced Secure Multi-Tenancy environment and the underlying storage needs to follow suit. Proper planning is required before setting up the vFiler units to make sure that the proper VLANs are extended from the network to the storage. Separate VLANs are required for management traffic, NFS traffic, and iSCSI traffic. This requirement is on a per-tenant basis, so if more than one tenant uses iSCSI, then each tenant should have its own iSCSI VLAN. NFS and management traffic should be treated to the same level of separation.

The next step is to disable any unnecessary or insecure services. For example, remote access should be done by way of SSH. Because RSH, FTP, and telnet are inherently insecure, they should be disabled as soon as the configuration of SSH is verified. Additionally, SSH2 should be used instead of SSH1. Also, if other services, such as TFTP or CIFS, are not being used, they should also be disabled. If the use of the Web console is enabled, force the use of HTTPS instead of HTTP.

In addition to the use of VLANs, access to data shares should be strictly confined to the specific hosts that need access. For example, access to an NFS share should be granted on a per-host basis as opposed to a per-subnet basis. Each individual IP address should be listed instead of simply using a blanket "network/mask" access policy. NFS traffic should be restricted to specific interfaces, rather than allowing NFS access on all interfaces.

Access to the iSCSI LUNs requires entering the exact Initiator Name when creating the igroup. From there, the network interfaces that accept iSCSI traffic should be limited as well. An additional layer of security can be added to the iSCSI LUN by enforcing a CHAP login/password challenge. For FCP LUNs, NetApp and VMware highly recommend that customers implement a "single-initiator, multiple storage target" zoning policy. Data access should not go through the management device. Furthermore, the management console should be on its own VLAN, completely separated from all other traffic.

For additional security recommendations, refer to TR-3649: Best Practices for Secure Configuration of Data ONTAP 7G.

Policy Enforcement and Access Control

The level of complexity shared virtualized data centers present in comparison to legacy data centers is exponential. Creating common policies and authentication measures across the environment is imperative in minimizing operational complexities and maximizing security. The Enhanced Secure Multi-Tenancy Architecture provides policy enforcement and access control methods in a unified approach across all layers of the solution in order to address both complexity and security concerns.

Centralized Authentication

Centralized authentication is an important feature within the Enhanced Secure Multi-Tenancy environment that prescribes users and groups be created globally across the entire solution rather than locally within each layer or component. In this design, the Cloud Administrator leverages Microsoft Active Directory to establish a hierarchy of users and groups, ensuring that each device in the environment adheres to the same global authentication design. Without the use of centralized authentication, users and groups are defined locally on each device, thus creating an increase in operational complexity and risk for mis-configuration, which can potentially lead to an insecure environment. In order for centralized authentication to serve its purpose, each device must include some form of Active Directory integration and, more specifically, support the type of naming service or services being leveraged. Some examples of name services are LDAP and TACACS+ with Cisco ACS, which are used as name services for the different devices throughout this particular design. No matter whether the device is a Cisco UCS Manager, a NetApp storage controller, or a VMware vCenter Server, each is integrated with Active Directory using some form of name service.

Users and Groups

As aforementioned, global users and groups are created within the environment in order to simplify tasks associated with configuring and maintaining the authentication design. Careful planning should be done prior to initial configuration to ensure users and groups are created with the thought of organizational structure in mind. Table 5 outlines example Active Directory users and groups created globally within an Enhanced Secure Multi-Tenancy environment. This table is meant to serve as an example only; users and groups will vary depending on the environment.

Table 5 Active Directory Users and Groups in an ESMT Environment

Group Name	User(s)	Privileges
Cloud Admins	Cloud Admin	Full access and privileges throughout the environment
	Infrastructure Admin	Access and privileges limited to physical infrastructure components only
	UCS Admin	Access and privileges limited to Cisco UCS Manager and UCS components only
	Server Virtualization Admin	Access and privileges limited to VMware vSphere environment
	Network Admin	Access and privileges limited to Cisco Nexus components only, including the Nexus 1000V
	Storage Admin	Access and privileges limited to NetApp storage controllers only
Tenant A Admins	Tenant A Admin	Access and privileges limited to Tenant A's virtual environment including virtual machines and virtual storage controller only
	Tenant A Server Admin	Access and privileges limited to Tenant A's virtual machines only
	Tenant A Application Admin	Access and privileges limited to specific Tenant A applications only
	Tenant A Services User	Access and privileges limited to only running services within Tenant A's virtual environment, i.e. SnapDrive for Windows

Table 5 Active Directory Users and Groups in an ESMT Environment

Tenant A Users	Tenant A User A	Access limited to necessary applications
	Tenant A User B	Access limited to necessary applications
	Tenant A User N…	Access limited to necessary applications

Privileges and Roles

Once users and groups have been established globally within Active Directory, roles and privileges must be assigned to specific users and groups to grant access to or allow tasks to be performed on objects within the environment. This is typically referred to as Role Based Access Control (RBAC). Privileges are defined as the individual tasks that can be performed on objects, while roles are a collection of such privileges that are grouped together based on a user or group's job description or role in the environment. Roles are typically created to define a common set of privileges in order to easily assign multiple privileges to the necessary users and groups. Each device in the environment owns different objects and assigns privileges and roles in a different fashion. For this reason, privileges and roles are assigned to global users and groups individually within each device. Care should be taken when assigning roles and privileges to groups rather than individual users. It can be convenient to assign roles and privileges to groups, but ensure only the appropriate users that require such roles and privileges are members of that particular group.

Building on Table 5, Table 6 provides an example of the privileges that may be assigned to users within the environment. Table 6 should only be used as a guide in creating customized authentication policies for a particular environment.

Table 6 Privileges Assigned to Users

Group Name	User(s)	Privileges
Cloud Admins	Cloud Admin	Full access and privileges throughout the environment
	Infrastructure Admin	Access and privileges limited to physical infrastructure components only
	UCS Admin	Access and privileges limited to Cisco UCS Manager and UCS components only
	Server Virtualization Admin	Access and privileges limited to VMware vSphere environment
	Network Admin	Access and privileges limited to Cisco Nexus components only, including the Nexus 1000V
	Storage Admin	Access and privileges limited to NetApp Storage controllers only

Table 6 Privileges Assigned to Users

Group Name	User(s)	Privileges
Tenant A Admins	Tenant A Admin	Access and privileges limited to Tenant A's virtual environment including virtual machines and virtual storage controller only
	Tenant A Server Admin	Access and privileges limited to Tenant A's virtual machines only
	Tenant A Application Admin	Access and privileges limited to specific Tenant A applications only
	Tenant A Services User	Access and privileges limited to only running services within Tenant A's virtual environment, i.e. SnapDrive for Windows
Tenant A Users	Tenant A User A	Access limited to necessary applications
	Tenant A User B	Access limited to necessary applications
	Tenant A User N…	Access limited to necessary applications

Secure Isolation

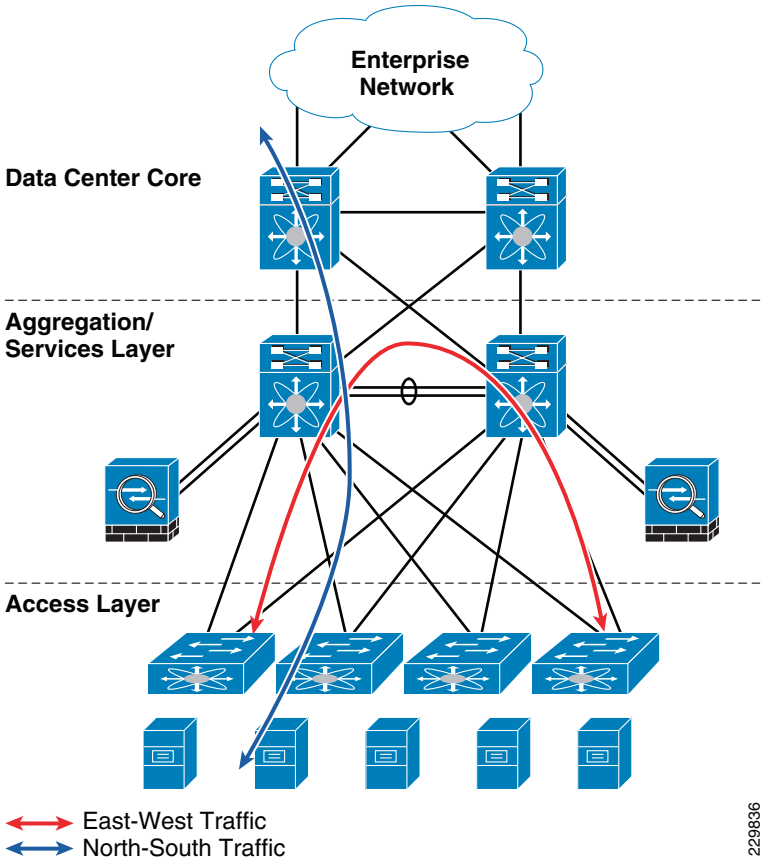
Isolation can provide the foundation of security for the multi-tenant data center and server farm. Depending on the goals of the design it can be achieved through the use of firewalls, access lists, VLANs, virtualization, storage, and physical separation. A combination of these can provide the appropriate level of security enforcement to the server applications and services within the different tenants.

Secure Network Isolation

The network infrastructure consists of the virtual access layer, services layer, and the core layer. The topological view of the network infrastructure outlining the different layers is illustrated in Figure 32.

The best practices and design considerations for each layer are described in the following sections.

Figure 32 Topological View of Network Infrastructure



Secure Isolation at the Virtual Access Layer

Secure separation at the access layer (Layer 2) is an essential requirement of multi-tenant network design. The separation is critical since it defines operational manageability and access. This separation provides compliance with regulatory requirements:

- Degree of confidentiality
- Performance
- Span of administrative control and accountability

Virtual Switches

The access layer of the network provides connectivity for serverfarm end nodes residing in the data center. The networking technology in the access layer has evolved greatly by the introduction of virtual switches such as the Nexus 1000V Virtual Distributed Switch. Nexus 1000V provides virtual access layer switching within a virtualized environment. Nexus 1000V is the glue between VMs interfacing the UCS 6100 fabric connection and the network infrastructure. The introduction of software switching at the access layer has greatly enhanced security capabilities within the data center. Virtual switches provide a seamless and consistent mechanism to apply security policies to virtual machines irrespective of where the virtual machine resides. In addition to virtual switches, the UCS provides additional security features that can be leveraged.

Port-Profile capability in Nexus 1000V is the primary mechanism by which network and security policy are defined and applied to virtual machines and is a dynamic way of defining connectivity with a consistent set of configurations applied to many VM interfaces at once. Essentially, port-profiles extend VLAN separation to VMs in flexible ways to enable security and QoS policies and are one of the

fundamental ways a VM administrator associates the VM to the proper VLANs or a subnet and hence provides network isolation. Port-profiles can be used to define configuration options as well as security and service level characteristics to virtual machines. Using this functionality, the security policies applied to each virtual machine do not change when the virtual machine is migrated within the data center. In this design, the Port-Profile capability of the Nexus 1000V is seamlessly integrated with virtual firewalls such as vShield to secure multiple tenants from threats within and from outside. In addition to leveraging the capabilities of the virtual firewall, this implementation also incorporates other security capabilities of the Nexus 1000V. Nexus 1000V security features include:

DHCP Protection

DHCP protection is critical to ensure that a client on an access edge port is not able to spoof or accidentally bring up a DHCP server nor exhaust the entire DHCP address space by using a sophisticated DHCP starvation attack. Both of these attacks are addressed with the Cisco IOS DHCP snooping feature that performs two key functions:

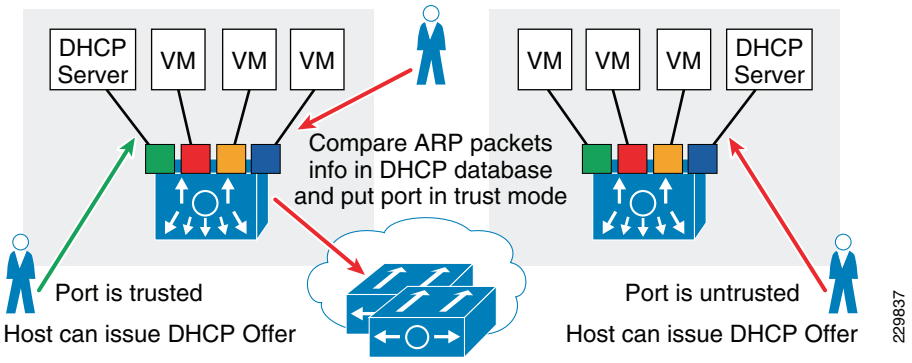
- Rogue DHCP Server Protection—If reserved DHCP server responses (DHCPOFFER, DHCPACK, and DHCPNAK) are received on an untrusted port (such as an access port), the interface is shut down.
- DHCP Starvation Protection—Validates that the source MAC address in the DHCP payload on an untrusted (access) interface matches the source MAC address registered on that interface.

ARP Spoofing Protection

ARP spoofing protection ensures that a client on an access edge port is not able to perform a Man-in-the-Middle (MITM) attack by sending a gratuitous ARP that presents its MAC address as that associated with a different IP address, such as that of the default gateway. This attack is addressed with the Cisco IOS Dynamic ARP Inspection (DAI) feature that validates that the source MAC and IP address in an ARP packet received on an untrusted interface matches the source MAC and IP address registered on that interface.

Figure 33 illustrates DHCP snooping functionality.

Figure 33 DHCP Snooping Functionality



IP Spoofing Protection

IP spoofing protection is provided by implementing IP Source Guard on Port-Profiles.

Port Security

The port security feature is used to restrict input to an interface by limiting and identifying MAC addresses of the Virtual Machine that are allowed to access the port. In multi-tenant environments the port security feature is used to lock down a critical server to a specific port. The Nexus 1000V provides three modes of port security:

- Static—MAC address of the Virtual Machine has been statically configured to be secure on that port. Static MAC address can only be configured on a virtual Ethernet (veth) interface.
- MAC address of the Virtual Machine is dynamically learned when the VM starts to send traffic. The MAC address configuration is not persistent after a VSM reboot.
- MAC address of the Virtual Machine is dynamically learned when the VM starts to send traffic. The MAC address configuration is persistent after a VSM reboot.

Once a MAC address has been secured, The Nexus 1000V can be configured to work in two modes to ether limit the number of Mac addresses that can be supported on a single interface or prevent a secure MAC to be seen on another port.

Virtual Firewalls

A virtual firewall, such as VMware vShield Edge and Zones, is a centrally managed, stateful, distributed virtual firewall bundled and tightly integrated with vSphere 4.1 and Nexus 100V which takes advantage of its proximity to the ESX host and its networking components to create security zones within the multi-tenant environment. The vShield Zones integrates into the VMware vCenter and leverages virtual inventory information, such as vNICs, port groups, clusters, and VLANs, to simplify firewall rule management and trust zone provisioning.

Within this multi-tenant environment virtual firewalls can be used to isolate network segments and lock down data flows to specific ports, right at the access layer. Depending on how the applications are deployed over the network, different kinds of policy-based separation options are available:

- Option 1—Multi-tier application deployed on separate Layer 2 subnets for each application component (i.e., Web front end, database and application server backend). The best option for this type of deployment is to use the port group as the container for separation policy—each port group (VLAN) is used as a container for firewall rule creation and enforcement. Scalability is built in for this approach as more virtual machines are added to their respective VLAN as the policy is defined at the VLAN level, not individual IP addresses. This functionality can be implemented using vShield App-Zones virtual firewall.
- Option 2—Multi-tier application deployed on the same Layer 2 subnet. The best option for this type of deployment is to use resource pool or vShield App container-based separation. Each application component can be packaged as a vShield App and each instance of vShield App is used as either a source or destination container for the separation policy. This model also has scalability built-in as virtual machines are added to their respective vShield App containers and the container-based separation policies are enforced automatically.
- Option 3—Multi-home virtual machines in DMZ or virtual machines that are complaint with specific industry security compliance such as HIPPA or PCI. In a physical world with virtual firewall, these virtual machines would reside on dedicated vSphere ESXi servers. The best option for this type of deployment is to use vNIC level separation. vNIC level separation defines distinct security zones between the DMZ and backend application infrastructure. vNIC separation allows only a specific NIC on the DMZ or compliant virtual machine to communicate to and from the backend corporate network. This controls access across the shared vSphere infrastructure while simultaneously supporting regulations such as HIPPA or PCI by isolating compliant from non-compliant virtual machines.
- Option 4—A dev/test environment that undergoes constant changes (development environment spin up, code testing) that could be detrimental to production environment if network separation is not implemented properly. For this type of deployment, the best option is to use vShield Edge to provide NATng and DHCP functionalities. All virtual machines can reside on the same Layer 2 subnet, with vShield Edge to provide separation for virtual machines in the same functional area, such as development compute farm and testing compute farm. Further firewall rules can be defined within the private network behind vShield Edge based on source and destination IP address pairs.

In this architecture virtual firewalls are used for the following:

- Intra-tenant virtual machine separation—For the SharePoint application, each of the component virtual machines is packaged into a vShield App container for each functional area: front-end Web server vApp, index server vApp, and backend SQL vShield App. For the Exchange application, CAS, Hub, and DAG virtual machines are packaged into their respective functional areas' vShield App. For the SQL server application, vShiedl App container creates firewall rules based on the individual communication needs of that one server and prevents unauthorized acess to all other traffic**.** For isolation between virtual machines, a virtual firewall such as vShield App can be used to isolate different hosts. In this design, the SharePoint components, such as the frontend machine, index server, and the SQL database reside in the same VLAN and vShield App firewall is used to implement firewall rules between them.
- For inter-tenant virtual machine separation—Given each tenant gets assigned a dedicated vSphere resource pool, the resource pool itself can be used as the container for separation. Firewall rules are created to restrict cross tenant access between Exchange, SQL, and SharePoint tenants and also isolate the vCloud pre-production tenant virtual machines.
- For VLAN reduction of both production and pre-production environments—For the production SharePoint and Exchange environments, only one VLAN is needed for each as opposed to three. For the pre-production environment, only one VLAN is designated for the dev and test virtual machines. Overall management overhead is reduced due to a fewer number of VLANs/PVLANs to provision, control, and manage.
- To provide secure connectivity between different tenants and shared infrastructure and shared resources, without the need to connect through the aggregation layer—In this implementation vShield Edge was used to connect tenants to Active Directory by using the NAT functionality between the different segments.
- To provide additional firewall functionality for north-south traffic—North-south traffic is protected by the appliances within the services layer and a virtual firewall can provide further protection and monitoring capability for inter-tenant and client-server traffic.
- Hide and isolate the network segments where shared resources reside from other parts of the enterprise network—This can be particularly useful if these resources are leveraged by other parts of the enterprise infrastructure that are public facing, such as the Internet edge and the DMZ.

Secure Isolation at the Services Layer and Aggregation Layer

The aggregation layer and the services layer provide an excellent filtering point for multi-layer protection of the data center. The aggregation layer enables virtualized data and control plane path for end-to-end Layer 3 network connectivity. Examples of techniques used for Layer 3 virtualization include Virtual Route Forwarder (VRF), MPLS, etc. This design methodology is commonly referred to as network virtualization. The devices in this layer of the architecture are Nexus 7000 Series switches. The Nexus 7000 provides more than sufficient slot and port density to support the surrounding core, services, and the access layer devices within the topology. In addition, the Nexus devices offer a rich set of Layer 2, Layer 3, and virtualization features permitting a new level of segmentation and control within a single aggregation device in the data center.

This services layer provides a building block for deploying firewall services for ingress and egress filtering of data center bound traffic. In this architecture the recommendations for all the tiers of service provide a symmetric traffic patterns to support stateful packet filtering and inspection. The services layer provides intrusion protection, server load balancing, and firewall services. The Services Layer consists of VSS-enabled Catalyst 6500 series switches using service modules and, in case of a high-powered user, a dedicated appliance platforms can be used. The appliance services may attach directly to the Nexus 7000 aggregation layer or use the port density available on the services chassis themselves. From a

security perspective, the services layer design used for this solution is based upon previous efforts surrounding services chassis design, but expands upon this foundation of load balancing and firewalling to include intrusion detection, intrusion prevention, and monitoring capabilities.

As outlined previously, a high-powered tenant which uses the Platinum Security Service offering will use a Layer 3 design within the services layer using separate firewall and ACE appliances. Other tenants that use the other security service offerings will use a Layer 2 design using transparent contexts within the firewall and ACE modules. In either case the network isolation is provided by the following:

- VLANs are used to separate tenants and resources or components within each tenant.
- The Nexus 7000 provides the aggregation functionality for the data center infrastructure and provides each tenant with separate VRF instances. The Nexus 7000 performs routing and provides connectivity outside of the data center.
- Inter-tenant network isolation is provided by a physical firewall, as well as a virtual firewall such as vShield. Inter-tenant traffic flow flows through the aggregation firewall and Nexus 7000.
- In service tiers that use a Layer 2-based design in the services layer, the firewalls and the ACE module are configured in transparent mode. The firewall modules have been configured for multiple contexts using the virtual context feature. This virtualization feature allows the firewall to be divided into multiple logical firewalls each supporting different interfaces and policies. The use of distinct contexts allows for sharing of network resources while at the same time providing seamless separation of tenant traffic flows within the multi-tenant environment.
- The ACE module or appliance can be used to terminate HTTP, HTTPS, Secure Socket Layer (SSL), and other protocols. By using virtual contexts or Layer 3 routing, one can achieve robust separation between server clusters at different tenants. The ACE provides the additional benefit of storing server certificates locally. This allows the ACE to act as a proxy SSL connection for client requests and forward the client request in clear text to the server. This functionality makes it possible for the Cisco IPS devices to inspect the traffic in clear text.

More information on the implementation details for network virtualization is available at: http://www.cisco.com/en/US/solutions/ns340/ns414/ns742/ns815/landing_cNet_virtualization.html.

VLAN Design Considerations

VLANs are the primary means of network separation in multi-tenant design. VLAN design also requires planning and awareness of the administrative roles of compute, application, storage, and network resources. One of the challenges of a multi-tenant environment is to accommodate varied types and functional uses of VLANs with appropriate security while providing operational stability and control. Each major resource space (compute, storage, and network) requires the various control and manageability of the devices within its scope. Typical functional uses of VLANs in traditional design are covered below for reference and are a basis for further design considerations on how to consolidate and classify each function based on its criticality in multi-tenant design.

VLAN Planning and Design Scope

VLAN design enables cloud administrators to provide differentiated services by separating the critical operational requirements of managing cloud components. In addition to properly securing the VLANs, it is important to map administrative scope and the access requirements to each category of VLANs.

- Management VLANs

  These VLANs are used for managing VMware ESXi hosts, NetApp storage controller, and networking devices such as Nexus 1000V, Nexus 5000, and Nexus 7000. This category also includes VLANs for monitoring, SNMP management, and network level monitoring (RSPAN).

- Monitoring VLANs

  This category includes VLANs for monitoring the infrastructure and for network-level visibility via SPAN, RSPAN, and ERSPAN, as well as netflow.

- Storage VLANs

  These VLANs are used for tenant datastores and iSCSI connectors to meet application requirements. VMkernel interfaces are designated within these VLANs for NFS and iSCSI traffic types.

- Service VLANs

  These VLANs provide non-routed segments for vMotion as well as Nexus 1000V control and packet traffic.

- Data VLANs

  The data VLANs consist of any VLANs used in servicing the customer application. Typically this includes front-end VLANs for application and back-end VLANs for application access to databases or storage fabric. Additionally each tenant requires VLANs for traffic monitoring and backup VLANs for data management.

Naming Design

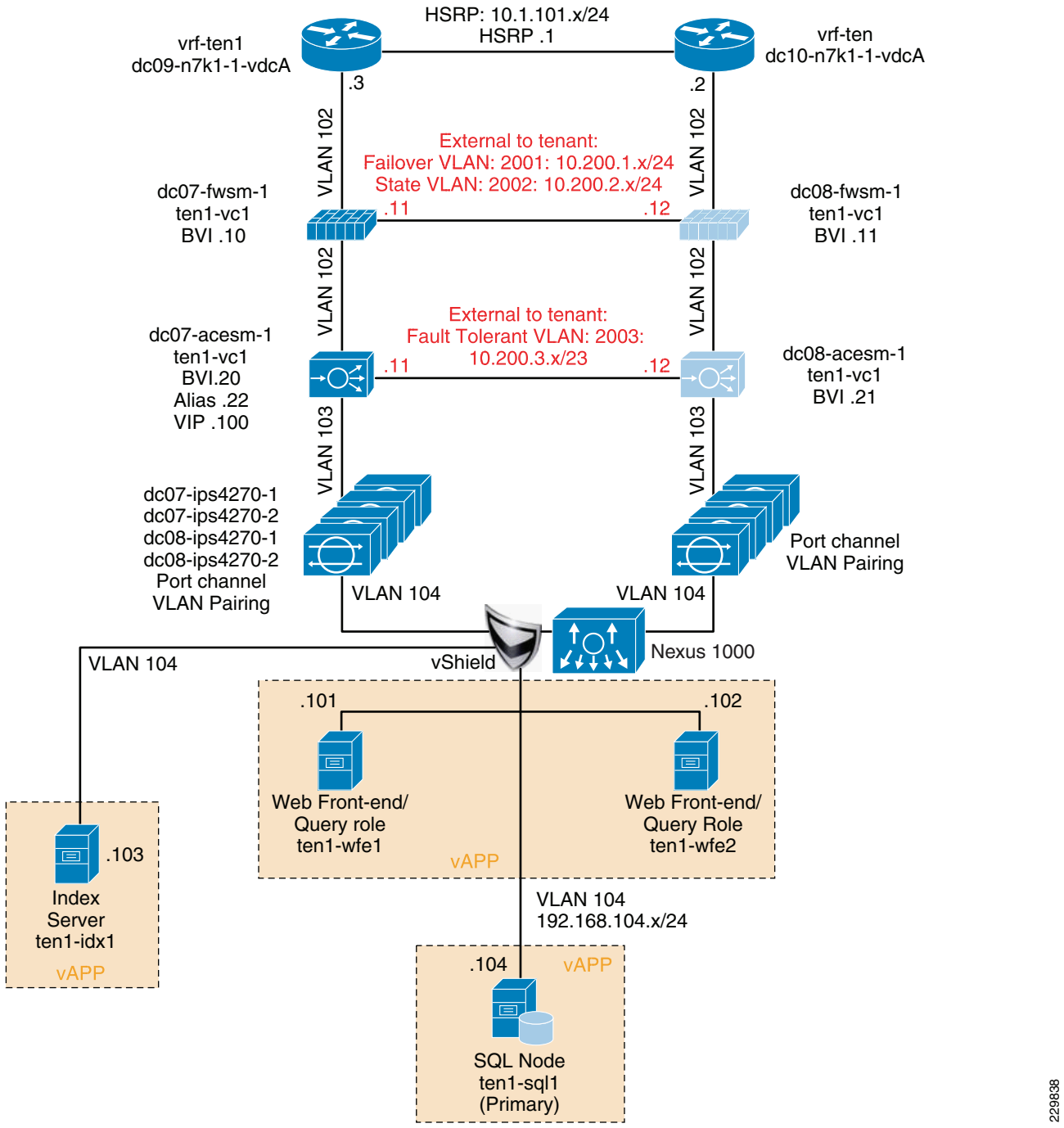
Multi-tenant design requires end-to-end operational consistency for managing, provisioning, and troubleshooting various resources. The VLAN classification group described above should further be enhanced with a consistent naming convention. The VLAN naming convention is crucial for interworking a diverse set of groups in a multi-tenant environment. The following guideline can be used in naming VLANs:

- Identify services levels for each multi-tenant entity. The services level can be categorized as Platinum, Gold, Silver, Bronze, and Default class.
- Identify tenants and appliances utilizing a particular VLAN, e.g., Sales, Marketing, HR, vShield, and Management.
- Identify types of applications or functions each VLAN provides, e.g., Transactional, Bulk, NFS datastore, and Application IO.
- Define subnet identification for each VLAN so that VM administrators and networking staff can identify subnet to port-profile to VLAN association.

The common workflow models start at the application group requesting a VM. Three separate working groups (compute, storage, and network) provide resources to enable the requested services. By providing a consistent naming for VLANs, VMs, and Nexus 1000V port-profiles, the workflow can be significantly streamlined. This may seem trivial, however the validation experience shows that it is extremely hard to associate a VM to its proper port-profile, which can then be associated with the correct VLAN in order to optimize operational efficiency.

The Nexus 1000V is the glue between VMs interfacing the UCS 6100 fabric connection. The port-profile is a dynamic way of defining connectivity with a consistent set of configurations applied to many VM interfaces at once. Essentially, port-profiles extend VLAN separation to VMs in flexible ways to enable security and QoS separation. The port-profile is one of the fundamental ways a VM administrator associates the VM to the proper VLANs or subnet. The port-profile name should also follow the same VLAN naming convention. By matching VLAN names and policy to port-profile name, both compute and network administrators can refer to the same object and provisioning and troubleshooting becomes much easier.

Figure 34 Design Considerations and Best Practices for Network Isolation

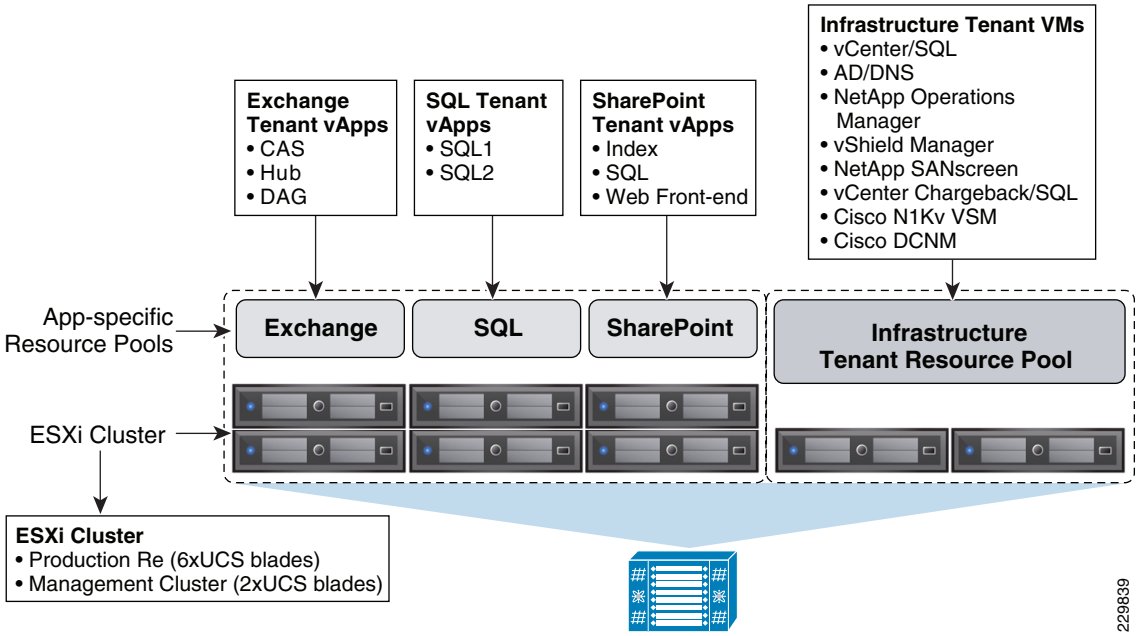


Secure Compute Isolation

Resource Pool separation at vSphere layer

vSphere provides the ability to aggregate all physical compute resources into a resource pool, thereby abstracting the notion of a physical server. It is imperative to design resource pools with separation in mind—separation in the context of infrastructure management services and raw compute resources for tenant consumption. This model of separation enables better security and easier problem isolation. Figure 35 is a reference model that satisfies both requirements.

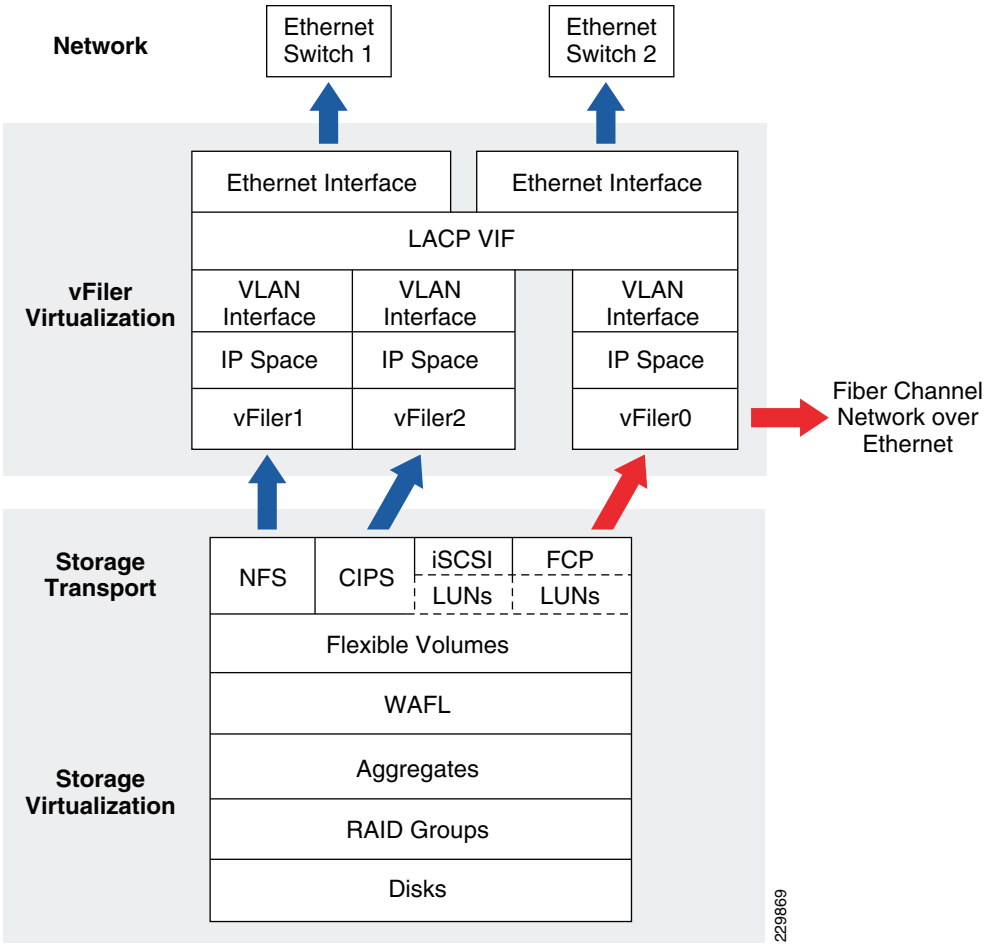
Figure 35 Production Site Resource Pool Isolation/Separation or Recovery Site Resource Pool Isolation/Separation



Secure Storage Isolation

This section examines how the storage layer provided by NetApp secures the separation of tenant data. The technologies involved in storage virtualization are also demonstrated.

Figure 36 Technologies Involved in Storage Virtualization



As previously mentioned, physical disks are pooled into RAID groups, which are further joined into abstract aggregates. To maximize parallel I/O, we configure the largest aggregate possible, which is then logically separated into flexible volumes. Each flexible volume within an aggregate draws on the same storage pool, but has a unique logical capacity. These volumes can be thin-provisioned and the logical capacity can be resized as needed by the storage administrator.

In this architecture, MultiStore is used to deploy multiple vFiler units to manage one or more volumes. vFiler units are isolated virtual instances of a storage controller and have their own independent configurations. These virtual storage controllers have virtual network interfaces and, in this architecture, each interface is associated with a VLAN and IP space. IP spaces provide a unique and independent routing table to a virtual storage controller and prevent problems in the event that two VLANs have overlapping address spaces.

The physical storage controller is accessed through vFiler0, which administers the other vFiler units and is the only vFiler unit providing Fibre Channel services. All Ethernet storage protocols (that is, NFS, CIFS, and iSCSI) are served by unprivileged vFiler units. This includes both infrastructure data (for example, NFS datastores for VMware ESXi served from the infrastructure vFiler unit) and tenant data (for example, a database LUN accessed by using iSCSI from a tenant vFiler unit).

vFiler units serve as the basis for the secure separation of storage. Each vFiler unit encapsulates both the data and administrative functions for a given tenant and is restricted to the VLANs associated with that tenant. Therefore, even the tenant administrator (who has root privileges on his or her vFiler unit) cannot

connect to another tenant's vFiler unit, let alone access the data managed by it. Furthermore, the Ethernet storage network implements strict access control to block any IP traffic other than the defined storage, backup, and administration protocols.

Cloud Administrator Perspective

Every IT organization requires certain administrative infrastructure components to provide the necessary services and resources to end users. These components within the secure cloud architecture include various physical and virtual objects, including storage containers and storage controllers. These objects play an important role in maintaining overall cloud operations. However, from a security aspect, they are treated exactly the same as tenant resources and are isolated from other tenants as such. The infrastructure Ethernet storage (that is, NFS for ESXi datastore, iSCSI for the vCenter database, and so on) is separated onto its own non-routed VLAN. The Fibre Channel over Ethernet SAN used to boot the ESXi hosts is isolated because tenants do not have access to Fibre Channel initiators—the only initiators present are the HBAs presented by the Cisco Virtual Interface Card (VIC) within each UCS blade. Even within a VLAN, all management traffic is configured to use only secure protocols (HTTPS, SSH, and so on), and local firewalls and port restrictions are enabled where appropriate.

The cloud administrator configures all storage containers, both physical (aggregates) and virtual (flexible volumes), and then assigns virtual storage containers to individual tenant vFiler units in the form of flexible volumes. After a flexible volume has been assigned to a tenant vFiler unit, the tenant can either export the flexible volume directly by using NAS protocols or further redistribute storage by using block-base LUNs or lower-level directories called qtrees. Because the cloud administrator owns all storage allocations, tenants can only use the storage directly allocated to their vFiler unit. If additional storage is needed, the cloud administrator may resize the currently allocated flexible volume(s) for that tenant or assign an additional flexible volume. Tenants cannot use more than their allocated storage. Only the cloud administrator, who can responsibly manage storage resources among the tenants, has the ability to allocate storage capacity.

Tenant Perspective

Each tenant possesses his or her own authentication measures for both administrative and data access to their vFiler unit and its underlying storage resources. Tenant administrators can choose the necessary export method and security exports between the application and the storage. As an example, a tenant administrator can create custom NFS export permissions for his or her assigned storage resources or export storage by using LUNs and leverage iSCSI with CHAP between the application VMs and the storage. The method by which application and/or user data is accessed from a tenant's vFiler unit is customizable by the tenant administrator. This creates a clean separation between storage provisioning (undertaken by the cloud administrator) and storage deployment (managed by the tenant administrator).

Service Assurance

Service Assurance is the third pillar that provides isolated compute, network, and storage performance during both steady state and non-steady state. For example, the network can provide each tenant with a certain bandwidth guarantee using QoS, resource pools within VMware help balance and guarantee CPU and memory resources, while FlexShare can balance resource contention across storage volumes.

Table 7 Methods of Service Assurance

Compute	Network	Storage
• UCS QoS System Classes for Resource Reservation and Limit • UCS Cisco VIC Rate Limiting • vSphere Resource Pool • VM/vApp Resource Reservation • Distributed Resource Scheduler (DRS)	• QoS—Queuing • QoS—Bandwidth control • QoS—Rate Limiting	• FlexShare • Storage Reservations • Thin Provisioning

Design Considerations for Compute Service Assurance

To support the service assurance requirement for a multi-tenant environment with differing levels of service, the following features are used:

- vSphere Resource Pool settings
- Distributed Resource Scheduling (DRS)

VMware vSphere Resource Pool Settings

Compute resource assurance for production and pre-production tenants can be achieved by setting the following attributes built in for resource pools:

- Reservation—Affects guaranteed CPU or memory allocation for the tenant's resource pool. A nonzero reservation is subtracted from the unreserved resources of the parent (host or resource pool). The resources are considered reserved, regardless of whether virtual machines are associated with the resource pool.
- Limit—Defines the maximum amount of CPU and/or memory resource a given tenant can utilize.
- Shares—Set to "High, normal, or low" on a per-tenant resource pool level; under transient (non-steady state) conditions with CPU and/or memory resource contention, tenants with "high" shares or larger number of shares configured have priority in terms of resource consumption.
- Expandable Reservation—Indicates whether expandable reservations are considered during admission control. With this option enabled for a tenant, if the tenant powers on a virtual machine in their respective resource pool and the reservations of the virtual machines combined are larger than the reservation of the resource pool, the resource pool can use resources from its parent or ancestors.

Management and Production Resource Pool Settings

Note The following guidelines are for references only; actual values/settings may vary depending on resource requirements/service.

Table 8 Resource Poll Settings Guidelines

Resource Pool Type	Reservation (CPU/Memory)	Limits	Shares	Expandable Reservation
Management	Reserved (non-zero); amount should be at least the minimum resource required for all management/service VMs	Unlimited	N/A as the management cluster is dedicated and separated from Production and Pre-production	Enabled
Production: Exchange	Reserved (non-zero); amount should be at least the minimum resources required for CAS, Hub and DAG VMs	Unlimited	NA as all three production applications are treated equal (this value can be set if specific production apps have special resource needs	Enabled
Production: SharePoint	Reserved (non-zero); amount should be at least the minimum resources required for front-end web server, index and SQL DB VMs	Unlimited	Same as Exchange	Enabled
Production: MS-SQL	Reserved (non-zero); amount should be at least the minimum resource required for active and standby SQL server nodes	Unlimited	Same as Exchange	Enabled

Virtual Machine Resource Settings

The following settings are configurable for individual virtual machines running in production and pre-production clusters, all of which are defined the same way as resource pool settings, except they are applied on a per virtual machine basis:

- CPU/memory reservation
- CPU/Memory shares

Note Applications such as Exchange, SQL, and SharePoint have strict CPU/memory resource requirements. Ensure reserved resource values are in compliance with Microsoft best practice recommendations. For other types of applications that the VMs may host, use the "Share" value to provide preferential treatment for more critical VMs to accommodate non-steady state condition where VMs are competing for resources in the same ESXi host.

VMware DRS Load Balancing

VMware Distributed Resource Scheduler (DRS) can be set to be fully automated at the cluster level so the management, production, and pre-production ESXi clusters are load balanced during VM/vApp power on, steady state, and non-steady state conditions where the cluster experiences ESXi host failure. Ensure all clusters are enabled with DRS with migration threshold set to normal.

Design Considerations for Network Service Assurance

QoS-Based Classification

Classification based on application and tenant services level is the primary requirement for resource pooling in a multi-tenant environment. Traffic classification is the foundation to protect resources from oversubscriptions and provide service level assurance to tenants. The QoS classification tools identify traffic flows so that specific QoS actions can be applied to the desired flows. Once identified, the traffic is marked to set the priority based on pre-defined criteria. The marking establishes the trust boundary in which any further action on the traffic flow can be taken without re-classifying the traffic at each node in the network. Once the packet is classified, a variety of action can be taken on the traffic depending upon the requirements of the tenant.

This design guide provides classification and service levels to the infrastructure as well as the tenant level. The design that follows is illustrated in Figure 38. To provide such services levels, the network layer must be able to differentiate the bidirectional traffic flows from application to storage and application to user access for each tenant. In addition, resilient operation of control plane functions (such as console management of devices, NFS datastore, control and packet traffic for Nexus 1000V, and many more) is critical for the stability of the entire environment. This service level assurance or dynamic management of traffic flows is a key to multi-tenant design. The first step in meeting this goal is to adopt the following classification principles at various hierarchical layers of the network:

- Classification Capability of Layer 2 Network
- Identify the Traffic Types and Requirements for Multi-Tenant Network
- Classify the Packet Near to the Source of Origin

Each of the above principles and ensuing design decisions are described in the following sections.

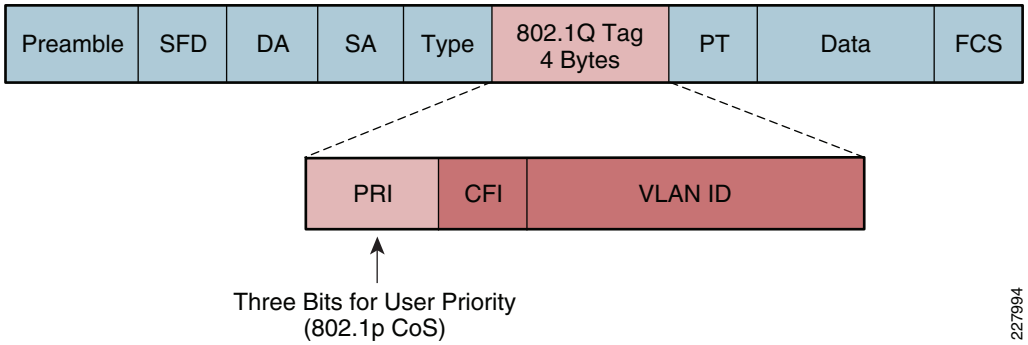
Classification Capability of Layer 2 Network

The industry standard classification model is largely based on RFC 2474, RFC 2597, RFC 3246, as well as informational RFC 4594. The data center QoS capability is rapidly evolving to adopt newer standards under the umbrella standard of DCB (Data Center Bridging). For more information on these standards, see:

- Data Center Bridging Task Group: http://www.ieee802.org/1/pages/dcbridges.html
- Priority-based Flow Control: http://www.ieee802.org/1/pages/802.1bb.html

This design guide uses 802.1Q/p Class of Service (CoS) bits for the classification as shown in Figure 37. The three bits gives eight possible services type, out of which CoS 7 is reserved in many networking devices; thus this design consists of a class of service model based on the remaining six CoS fields.

Figure 37 802.1Q/p CoS Bits



In addition, the number of classes that can be applied in a given network depends on how many queuing classes are available in the entire network. The queuing class determines which packet gets a priority or drop criteria based on the oversubscription in the network. If all devices have a similar number of classes available, then one can maintain end-to-end symmetry of queuing classification. This design guide uses five queuing classes, excluding FCoE since the minimum number of queuing classes supported under UCS is five excluding FCoE class. These five classes are shown in Table 9.

Table 9 Services Class Mapping with CoS and UCS

CoS Class	UCS Class	Network Class Queue
5	Platinum	Priority
6	Gold	Queue-1
4	Silver	Queue-2
3	FCoE	Reserved, unused
2	Bronze	Queue-3
0 and 1	Best-effort	Default Class

In UCS all QoS is based on 802.1p CoS values only. IP ToS and DSCP have no effect on UCS internal QoS and thus cannot be used to copy to internal 802.1p CoS, however DSCP/ToS set in IP header is not altered by UCS. CoS markings are only meaningful when CoS classes are enabled. One cannot assign more than one CoS value to a given class. If all the devices do not have the same number of queues/capability to classify the traffic, it is generally a good idea to only utilize the minimum number of classes offered, otherwise application response may become inconsistent.

Identify the Traffic Types and Requirements for Multi-Tenant Network

This is the most critical design decision in developing services level models in multi-tenant design. The VLAN separation decision and methods discussed previously also overlap this map of classification. The traffic profiling and requirements can vary from tenant to tenant. The cloud service network administration should develop a method to identify customer traffic types and application response requirements. Methods to identify application and traffic patterns are beyond the scope of this design guide. However, the following best practices can be used to classify traffic based on its importance and characteristics:

Infrastructure Type of Traffic—This is a global category that includes all traffic types except tenant data application traffic. There are three major types of traffic flow under the infrastructure category:

- Control Plane Traffic—This traffic type includes the essential signaling protocols and data traffic required to run the ESX-host as well VMs operating system. The operational integrity of these is of the highest priority since any disruption on this class of service can have multiple impacts ranging from slow response from ESXi host to guest VM operating systems shutting down. The type of traffic that falls in this category includes ESX-host control interfaces (VMkernel) connected to NFS datastore and Cisco Nexus 1000V control and packet traffic. The traffic profile for this class can range anywhere from several MB/second to bursting to GB/second. The traffic of these characteristics are classified with CoS of 5 and mapped to a “priority” queue and Platinum class where appropriate. The priority queue available in networking devices offers the capability to serve this type of traffic since the priority queue is always served first without any bandwidth restrictions as long as the traffic is present.
- Management Traffic—This traffic type includes the communication for managing the multi-tenant resources. This includes ESXi management access, storage and network device management, and per-tenant traffic (application and VM administration). The traffic requirement of this type of traffic

may not be high during steady state, however access to the critical infrastructure component is crucial during failure or congestions. Traffic with these characteristics is e classified with CoS of 6 and mapped to a queue and Gold class where appropriate.

- vMotion Traffic—vMotion is used for migrating VMs behind the scenes. This traffic originates from ESXi hosts during the VM move, either via automation or user-initiated action. This type of traffic does not require higher priority, but may require variable bandwidth as memory size and page copy can vary from VM to VM in addition to the number of VMs requiring vMotion simultaneously. Delaying vMotion traffic simply slows the background copy and makes the operation take longer. Traffic with these characteristics is classified with CoS of 4 and mapped to a queue or Silver class where appropriate.

Tenant Data Plane Traffic—This traffic category comprises two major traffic groups. The first one consists of back-end traffic, which includes storage traffic and back-end VM-to-VM traffic for multi-tier applications. The second group consists of user access traffic (generically called front-end application traffic). Each of these traffic groups would require some form of protection based on each tenant's application requirements. Each class also requires some form of service differentiation based on enterprise policy. For this reason each of these traffic groups are further divided into three levels of service, Platinum, Gold, and Silver. The mapping of the services class to CoS/Queue/UCS-class is show in Table 10. Identifying each user tenant application and user requirement and developing a service model that intersects the various requirements of each tenant is beyond the scope of this design guide. For this reason, in this design guide the services level classification is maintained at the tenant level. In other words, all tenant traffic is treated with a single service level and no further differentiation is provided. However the design methodology is extensible to provide a more granular differentiation model.

- Back-end User Data Traffic—This traffic type includes any traffic that an application requires to communicate within a data center. This can be application to application traffic, application to database, and application to each tenant storage space. The traffic bandwidth and response time requirements vary based on each tenant's requirements. In this design three levels of services are proposed for back-end user data; each service is classified in separate CoS classes based on the requirements. The services level classification helps differentiating various IO requirements per tenant. Table 10 explains and maps the services class based on IO requirements of the application. Each IO requirement class is mapped to CoS type, queue type, and equivalent UCS bandwidth class.

Note In this design guide, CoS 6 is used for data traffic, which is a departure from traditional QoS framework.

Table 10 Services Levels for Back-End User Data Traffic

Services Class	IO Requirements	Cos/Queue/UCS-Class	Rational
Platinum	Low latency, Bandwidth Guarantee	5/Priority-Q/Platinum class	Real-time IO, no rate limiting, no BW limit, First to serve
Gold	Medium latency, No Drop	6/queue-1/Gold class	Less than real-time, however traffic is buffered
Silver	High latency, Drop/Retransmit	4/queue-2/Silver class	Low bandwidth guarantee, Remarking and policing allowed, drop and retransmit handled at the NFS/TCP level

- Front-end User Data Plane Traffic—This class of traffic includes the front-end VM data traffic for each tenant accessed by user. The front-end user traffic can be further sub-divided into three distinct classes of traffic. Each of these subclasses has unique requirements in term of bandwidth and response-time. Each traffic subclass is described below with the classification rationale.
- Transactional and Low-Latency Data—This service class is intended for interactive, time-sensitive data applications which requires immediate response from the application in either direction (example of such could be Web shopping, terminal services, time-based update, etc.). Excessive latency in response times of foreground applications directly impacts user productivity. However not all transactional application or users require equal bandwidth and response time requirements. Just like back-end user traffic classification, this subclass offers three levels of services, Platinum, Gold, and Silver and related mappings to CoS/Queue/UCS-Class, as shown in Table 11.

Table 11 Services Levels for Transactional User Data Traffic

Services Class	Transactional Requirements	Cos/Queue/UCS-Class	Rational
Platinum	Low latency, Bandwidth Guarantee	5/Priority-Q/Platinum class	Real-time IO, no rate limiting, no BW limit, First to serve
Gold	Medium latency, No Drop	6/queue-1/Gold class	Less than real-time, however traffic is buffered, policing is allowed
Silver	High latency, Drop/Retransmit	4/queue-2/Silver class	Low bandwidth guarantee, drop and retransmit permitted, policing or remarking allowed.

- Bulk Data and High-Throughput Data—This service class is intended for non-interactive data applications. In most cases this type of traffic does not impact user response and thus productivity. However this class may require high bandwidth for critical business operations and may be subject to policing and re-marking. Examples of such traffic include E-mail replication, FTP/SFTP transfers, warehousing application depending on large inventory updates, etc. This traffic falls into the Bronze services class with CoS of 2, as shown in Table 12.

Table 12 Services Levels for Bulk User Data Traffic

Services Class	Transactional Requirements	Cos/Queue/UCS-Class	Rational
Bronze	Bulk Application and High Throughput	2/queue-3/Bronze class	

- Best Effort—This service class falls into the default class. Any application that is not classified in the services classes already described is assigned a default class. In many enterprise networks, a vast majority of applications default to best effort service class; as such, this default class should be adequately provisioned (a minimum bandwidth recommendation for this class is 25%). Traffic in this class is marked with CoS 0.

- Scavenger and Low-Priority Data—The scavenger class is intended for applications that are not critical to the business. These applications are permitted on enterprise networks, as long as resources are always available for business-critical applications. However, as soon as the network experiences congestion, this class is the first to be penalized and aggressively dropped. Furthermore, the scavenger class can be utilized as part of an effective strategy for DoS (denial of service) and worm attack mitigation. Traditionally in enterprise campus and WAN network this class is assigned a CoS of 1 (DSCP 9).

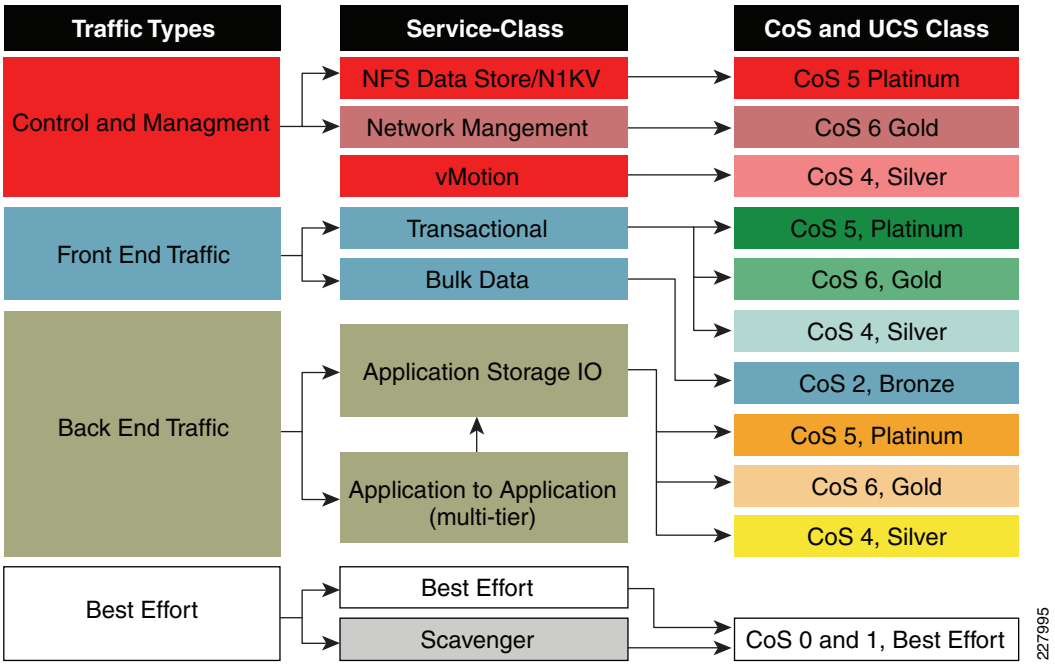
In this design guide the best effort and scavenger classes are merged into one class called "best-effort" in UCS 6100 and "class-default" in Nexus 5000.

Table 13 Services Levels for Best Effort User Data Traffic

Services Class	Transactional Requirements	Cos/Queue/UCS-Class	Rationale
Default (best-effort, class-default, scavenger class)	Any application that is not classified in above categories or not matched with given classification rules or marked with very high probability of drop	0 &1/default-queue/Best-effort class	Default class as well as class that is less then default (scavenger class for deploying DoS services)

Figure 38 summarizes the type of traffic, services class, and associated CoS mapping that would be marked as per the services model proposed in this design.

Figure 38 Enhanced Secure Multi-Tenancy Service Class Model



Classify the Packet Near to the Source of Origin

By keeping the classification near the source of the traffic, network devices can perform queuing based on a mark-once, queue-many principle. In multi-tenant design there are three places that require marking:

- Classification for the Traffic Originating from ESXi Hosts and VM
- Classification for the Traffic Originating from External to the Data Center
- Classification for the Traffic Originating from Networked Attached Devices

Classification for the Traffic Originating from ESXi Hosts and VM

When incorporating UCS in your data center design, there are two options available to classify traffic originating at the ESXi host or VM:

- Cisco M81KR VIC QoS marking—All traffic leaving a vNIC can be classified as a singular CoS value. For more information on configuration, see the configuration guide at: http://www.cisco.com/en/US/partner/docs/unified_computing/ucs/sw/gui/config/guide/1.3.1/UCSM_GUI_Configuration_Guide_1_3_1_chapter18.html#task_9463F54D9EC0498FB977AD8D1A8D2096.
- Use the Nexus 1000V as a classification boundary—Traffic can be classified and marked using the Modular QoS CLI.

In this design the Nexus 1000V is used as classification boundary and thus the traffic originating from a VM is treated as un-trusted. Any traffic from an ESXi host, guest VM, or appliance VM is categorized based on the above services levels. The classification and marking of traffic follows the Modular QoS CLI (MQC) model in which multiple criteria can be used to classify traffic via use of ACLs, DSCP, etc. The class-map identifies the traffic classes in a group, while the policy map provides the capability to change the QoS parameter for a given service class. All packets that are leaving an uplink port of the Virtual Ethernet Module (VEM) on each server blade are marked based on policy map. The services policy is then attached to the port-profile. This QoS consistency is maintained for stateless computing where VMs can move to any blade server with a consistent set of access and classification criteria.

Classification for the Traffic Originating from External to the Data Center

The Nexus 7000 is a natural boundary for classifying traffic entering or leaving the data center. The traffic originating from outside the data center boundary may have either DSCP-based classification or no classification at all. The traffic originating from the data center towards a tenant user can be either re-mapped to DSCP scope defined by a larger enterprise-wide QoS service framework or simply trusted based on the CoS classification defined in the above section. If the traffic is marked with the proper QoS classification in either direction, no further action is required as the Nexus 7000 by default treats all the ports in a trusted mode. DSCP to CoS translation is done via three higher order bits in the DSCP field and similarly for CoS to DSCP translation.

For more information on Nexus 7000 QoS, see:
https://www.cisco.com/en/US/docs/switches/datacenter/sw/4_2/nx-os/qos/configuration/guide/qos_nx-os_book.html

Classification for the Traffic Originating from Networked Attached Devices

In this design the Nexus 5000 is used as the classification boundary at the network access layer. It can set trusted or un-trusted boundaries or both, depending on requirements of the multi-tenant design. The following functionality is required:

- If a type of device connected to the network cannot set the CoS value, then that device is treated as un-trusted and both classification and setting the CoS value is required.

  If the traffic is known to come from a trusted boundary (which implies that it has already been marked with the proper CoS), then only classification based on match criteria is required; otherwise even though the packet has a CoS value, the value itself is untrusted and must be overridden to enforce the commonly-defined CoS values defined by the administrator.

- In this design guide the traffic from the UCS-6100 and Nexus 7000 is always trusted as they represent a trusted boundary. However the traffic originating from the storage controller (NetApp FAS) is not trusted and thus requires classification and marking with the proper CoS. The Nexus 5000 QoS model consists of a separate class and policy map for each type of QoS functions. QoS is an umbrella framework of many functionalities and the QoS functionality in Nexus 5000 is divided into three groups:
- "QoS" is used for classification of traffic inbound or outbound at the global (system) as well as the interface level.
- "Network-qos" is used for setting QoS related parameters for given flows at the global (system) level.
- "Queuing" is used for scheduling how much bandwidth each class can use and which queue can schedule the packet for delivery. In this design queuing is applied as an output policy.

All three types of QoS follow the MQC model.

QoS-Based Assurance

QoS classification differentiates between the application flows and storage IO requirements upon which the services assurance model is built. Service assurance provides the resilient framework for developing the services level in a multi-tenant design. The services assurance at the network layer addresses two distinct design requirements for both the control function and tenant user data plane of the multi-tenant infrastructure:

- Network Resources Performance Protection (Steady State)
- Network Resources Performance Protection (Non-Steady State)

Network Resources Performance Protection (Steady State)

This functionality addresses how to protect the service level for each traffic type and services class in steady state. In a normal operation, the networking resources should be shared and divided to meet the stated goal of the service or protection. Once the traffic is classified based on services level, the shared bandwidth offered at the network layer must be segmented to reflect the services priority defined by the CoS field. There are two distinct methods to provide steady state performance protection:

- **Queuing**—Queuing allows the networking devices to schedule a packet delivery based on the classification criteria. The end effect of the ability to differentiate which packet can get a preferential delivery is to provide the differentiation in terms of response time for applications when oversubscription occurs. The oversubscription is a general term used for defining resources congestion that can occur for a variety of reasons in various spaces of a multi-tenant environment. Some examples that can trigger a change in resources map (oversubscription) are failure of multi-tenant components (compute, storage, or network), unplanned application deployment causing high bandwidth usage, or aggregation layer in the network supporting multiple unified fabrics. It is important to be aware that the queuing only takes effect when a given bandwidth availability is fully utilized by all the services classes. As described in Architecture Overview, the compute layer (UCS) usually offers a 1:1 subscription ratio and the storage controller offers enough bandwidth that queuing may not be occurring all the time. However, it is critically important to address the functional requirement of multi-tenant design that one cannot always be sure about

overuse of resources. The congestion always occurs in the end-to-end system, whether hidden inside application structure, VM NIC, CPU, or at the network layer. The oversubscription is elastic and thus the choke points move at various levels in the end-to-end systems. It is the ability of the queuing capability in each networking device to handle such dynamic events that determines the quality of service level.

This queuing capability is available at all layers of the network, albeit with some differences in how it functions in each device. The capability of each devices and its design recommendation are addressed below.

- **Bandwidth Control**—As discussed above, queuing allows managing the application response time by matching the order in which queues gets serviced, however it does not control the bandwidth management per queuing (service) class. Bandwidth control allows network devices an appropriate amount of buffers per queue such that certain classes of traffic do not over utilize the bandwidth, allowing other queues to have a fair chance to serve the needs of the rest of the services classes. Bandwidth control goes hand in hand with queuing, as queuing provides the preference on which packet are delivered first, while bandwidth provides how much data can be sent per queue. The Cisco VIC and the Nexus 1000V have the capability to provide bandwidth control at the VM level. With the Cisco VIC, each vNIC can be policed to a desired bandwidth. The same can be achieved using Modular QoS on the Nexus 1000V.

QoS-Based Classification describes the types of traffic and services classification based on the service level requirements. In that section each services class is mapped a queue with appropriate CoS mapping. Once the traffic flow is mapped to the proper queue, the bandwidth control is applied per queue. The queue mapping shown in Table 14 is developed based on the minimum queuing class available based on end-to-end network capability. The design principles applied to selecting queuing and bandwidth control requires the capability of following attributes in a multi-tenant design.

- **Topology**—The multi-tenant design goal is to offer scalability and flexibility in services offering. The three-tier model selected in this design has a flexibility and management point of selectively choosing the technique at each layer. The new paradigm of unification of access layer technologies (storage and data) and topologies (Fiber Channel, Ethernet, and FCoE) requires careful treatment of application and IO requirements. In a multi-tenant environment this translates into enabling necessary control and service level assurance of application traffic flows per tenant. If the aggregation and access layer is collapsed the QoS functionality and bandwidth control gets difficult since the traffic flow characteristic changes with a two-tier model. For example, the traffic from VM to VM may have to flow through the access layer switch since the dual-fabric design requires the traffic to exit the fabric and be redirected via an access layer switch that has a knowledge of the MAC address reachability. With a two tier design, the Layer 3 and Layer 2 functionality may get merged and thus one has to manage Layer 3 to Layer 2 flows, marking/classification between Layer 3 and Layer 2; aggregation-to-aggregation flows are now mixed with access layer flows (VM to VM). Thus traffic management and bandwidth control can be complex as the environment grows with diverse VM-to-VM or aggregation-to-aggregation flows supporting diverse communication connectivity. The unified access layer with Nexus 5000 allows control of the bandwidth and queuing, specifically targeting the traffic flow behavior at the edge for compute, storage, and networking.

Oversubscription Model (Congestion Management Point)

In this design UCS represents unified edge resources where the consolidation of storage IO and IP data communicates via 10G interfaces available at each blade. UCS is designed with a dual-fabric model in which each data path from individual blades eliminates the network bandwidth level oversubscription all the way up to UCS 6100 fiber interconnects. However, when UCS is used in multi-tenant environment the need for service level for each tenant(s) (which are sharing the resources in a homogeneous way)

requires management of the bandwidth within the UCS as well as at the aggregation point (Fiber Interconnects) where multiple UCSs can be connected. There are many oversubscription models depending upon the tiered structure and the access-to-aggregation topologies.

In this design working from compute layer up to Layer 3, the major boundaries where oversubscriptions can occur are:

- **VM to UCS Blades**—The density of VM and application driving VM network activity can oversubscribe 10G interface. Notice that the Nexus 1000V switch providing virtual Ethernet connectivity is not a gated interface; in other words, it is an abstraction of physical Ethernet and thus offers no signaling level limit that exists in physical Ethernet. Major communication flow that can occur is between VM to VM either within the blade or residing on a different blade; the later flow behavior overlaps the fiber interconnect boundary (described below) since those VM-to-VM communications must flow through the 6100 to an access layer switch.

Note The Cisco Virtual Interface Card presents a virtual abstraction of a physical Ethernet interface providing bandwidth controls equivalent to a physical adapter. The Cisco VIC is fully supported by the Cisco Nexus 1000V virtual switching platform and can assist in developing an oversubscription model where VM-to-VM traffic between blades is successfully managed.

- **Fiber Interconnect to access layer**—The uplinks from the UCS 6100 determine the oversubscription ratio purely from the total bandwidth offered to UCS systems, since each UCS systems can offer up to 80 GB/second of traffic to the UCS 6100. The maximum number of 10GBps links that can be provisioned from a UCS 6100 (from each fabric) is eight; the resulting oversubscription could be 2:1 or 4:1 depending on the number of UCS systems connected. In the future uplink capacity may rise to sixteen 10GBps links. The fiber interconnect manages the application flows (both directions) for two major categories of traffic:
  - Back-end user data traffic—VM to VM (either VM residing on a separate blade or to other UCS systems in a domain)
  - VM to storage (NFS datastore and Application IO per tenant)—Front end user data traffic-VM to users in each tenant

The UCS 6100 upstream (towards users and storage) traffic queuing and bandwidth control is designed based on services classes defined in QoS-Based Classification. The UCS QoS class capability and CoS mapping based on traffic classes is shown in Table 14. The queuing capability of UCS 6100 is integrated with the QoS services classes it offers. In other words, the QoS systems class is mapped to CoS mapping; e.g., Platinum class when assigned CoS value of 5, the CoS-5 is treated as priority class and is given a first chance to deliver the packet. Notice also that the Gold class is designated as “no-drop” to differentiate the IO and transactional services class based on tenant requirements. The no-drop designated class buffers as much as it can and does not drop the traffic; the resulting behavior is higher latency but bandwidth is guaranteed.

Bandwidth control becomes an important design attribute in managing services levels with the unified fabric. Bandwidth control in terms of weights applied to each class is also shown in Table 14. Notice that the weight multiplier can range from 1 to 10. The multiplier automatically adjusts the total bandwidth percentage to 100%. Table 14 does not reflect the bandwidth control applicable to a multi-tenant design, as effective values are highly dependent on application and user tenant requirements. However, Platinum class requires a careful bandwidth allocation since the traffic in this class is treated with higher priority and unlimited bandwidth (NFS datastore and Platinum tenant application IO).

The weight of 1 is referred as best-effort, however that does not mean the traffic in the respective class is treated as best-effort.

Table 14 shows the weight of one (1) is applied to all classes; the effective bandwidth is divided in equal multiples of five (total classes) (essentially a ratio of a weight of the class to total of weight presented as percentage of bandwidth as a whole number).

Table 14 UCS-Queuing and Bandwidth Mapping

QoS System Class	CoS Mapping	Drop Criteria	Weight (1-10)	Effective BW%
Platinum	5	Tail Drop	1 (best-effort)	20
Gold	6	No Drop	1 (best-effort)	20
Silver	4	Tail Drop	1 (best-effort)	20
Bronze	2	Tail Drop	1 (best-effort)	20
FCoE	3	Not Used	Not Used	
Default	0,1	Tail Drop	1 (best-effort)	20

For additional information on UCS 6100 QoS, see: http://www.cisco.com/en/US/docs/unified_computing/ucs/sw/gui/config/guide/GUI_Config_Guide_chapter16.html.

Within the access layer—The oversubscription at this boundary is purely a function of how many access layer devices are connected and how much inter-devices traffic management is required. Two major categories of application flow that require management:

- Back-end traffic (storage IO traffic)—In this design the storage controller (NetApp FAS) is connected to the Nexus 5000 with two 10GB/second links forming a single EtherChannel. The NFS datastore traffic flow is composed of ESXi host and guest VM operation, which is the most critical flow for the integrity of the entire multi-tenant environment. Per-tenant application traffic flow to a storage controller requires the management based on services levels described in QoS-Based Classification. This design guide assumes that each tenant vFiler unit is distributed over dual-controller and thus offers up to 40 GB/second bandwidth (the FAS6080 can have up to a maximum of five dual-port 10Gb adapters, thus ten 10Gb/second ports per controller and supports up to eight active interfaces per LACP group) and thus oversubscription possibility for managing the traffic from storage is reduced. However, the traffic flow upstream to the VM (read IO) is managed at the Nexus 5000 with bandwidth control.
- Front-end user traffic—In this design application flows from VM to user tenant are classified with a per tenant services class. The front-end user traffic requires bandwidth control on upstream as well as downstream. Upstream (to the user) bandwidth control should reflect the total aggregate bandwidth from all networked devices (in this design primarily UCS systems). The downstream (to the VM) bandwidth control can be managed per class at either the Nexus 7000 or Nexus 5000. In this design the Nexus 5000 is used as the bandwidth control point.

The Nexus 5000 QoS components are described in QoS-Based Classification. The queuing and bandwidth capability reflecting the above requirements are shown in Table 15. In Nexus 5000, queuing can be applied globally or at the interface level. In general it is a good design practice to keep the queuing policy global, as it allows for the same type of queuing and bandwidth for all classes to all interfaces in both directions. If the asymmetric QoS services requirement exists, then multiple levels of policy can be applied (interface and global). Each Ethernet interface supports up to six queues, one for each system class. The queuing policy is tied to via QoS group, which is defined when the classification policy is defined.

The bandwidth allocation limit applies to all traffic on the interface including any FCoE traffic. By default class is assigned 50% bandwidth and thus requires modification of both bandwidth and queue-limit to distribute the buffers over the required classes. For the port-channel interface the

bandwidth calculation applies as a sum of all the links in a given LACP group. The queues are served based on s WRR (weighted round robin) schedule. For more information on Nexus 5000 QoS configuration guidelines and restrictions, see: http://www.cisco.com/en/US/partner/docs/switches/datacenter/nexus5000/sw/qos/Cisco_Nexus_5000_Series_NX-OS_Quality_of_Service_Configuration_Guide_chapter3.html.

Table 15 shows the mapping of CoS to queue and bandwidth allocation. Table 15 does not reflect the bandwidth control applicable to a multi-tenant design, as effective values are highly dependent on application and user tenant requirements.

Table 15 Nexus 5000-Queuing and Bandwidth Mapping

QoS System Class	CoS Mapping	Queue	BW Allocation (%)	Drop Criteria
Platinum	5	Priority	20	Interface bandwidth
Gold	6	Queue-1	20	WRR
Silver	4	Queue-2	20	WRR
Bronze	2	Queue-3	20	WRR
FCoE	3	Not Used		Not Used
Default	0,1	Queue-4	20	WRR

Caution This design utilizes the VPC technology to enable loop-less design. The VPC configuration mandates that both Nexus 5000s be configured with a consistent set of global configuration. It is recommended to enable QoS polices at the systems level before the VPC is enabled. If the QoS configuration is applied after the VPC configuration, both Nexus 5000s must enable the QoS simultaneously. Failure to follow this practice would disable all the VLANs belonging to VPC topology.

Network Resources Performance Protection (Non-Steady State)

This functionality addresses how to protect the services level for each traffic type and services class in a non-steady state. Non-steady state is defined by any change in the resources pool, such as a failure of any component of a multi-tenant environment. vMotion or new VM provisioning can affect the existing resources commitment or protection of application flows. Non-steady state performance is often identified as a set of events that triggers the misbehavior of or over commitment of resources.

In a multi-tenant environment, tenants must be protected from each other. In practice, a tenant may require resources in which application and IO traffic may drastically vary from normal usage. In other cases a tenant environment may have been exposed to a virus, generating abnormal amounts of traffic. In either case, a set of policy controls can be enabled such that any un-predictable change in traffic patterns can be either treated softly by allowing applications to burst/violate for some time above the service commitment or by a hard policy to drop the excess or cap the rate of transmission. This capability can also be used to define service level such that non-critical services can be kept at a certain traffic level or the lowest service level traffic can be capped such that it cannot influence the higher-end tenant services. Policing as well as rate-limiting is used to define such services or protection levels. These tools are applied as close to the edge of the network as possible, since it is intuitive to stop the traffic from entering the network. In this design the Nexus 1000V is used in a policing and rate-limiting function for three types of traffic:

- **vMotion**—vMotion is used for live VM migration. The vMotion traffic requirement can vary for each environment as well as VM configuration. VMware traditionally recommends a dedicated Gigabit interface for vMotion traffic. In this design the vMotion traffic has been dedicated with a non-routable VMkernel port. The traffic for the vMotion from each blade-sever is kept at 1 GBps to

reflect the traditional environment. This limit can either be raised or lowered based on requirements, however the design should consider that vMotion is run to a completion event (thus it may take longer with lower bandwidth, but will complete in time) and should not be configured such that the resulting traffic rate impacts critical traffic, such as NFS datastore traffic.

- **Differentiated transactional and storage services**—In a multi-tenant design, various methods are employed to generate a differentiated services. For example, the "priority" queue is used for the most critical services and "no-drop" is used for traffic that cannot be dropped, but can sustain some delay. Rate-limiting is used for services that are often called fixed-rate services, in which each application class or service is capped at a certain level, beyond which the traffic is either dropped or marked with high probability drop CoS. In this design Silver traffic is designated as fixed-rate services and rate-limited at the edge of the network.
- **Management**—In this design guide the ESXi management interface (VLAN) is merged into a common management VLAN designated for all the resources. Traditionally the ESXi management interface is provided a dedicated 1GBps interface, however in practical deployments the bandwidth requirement for managements function is well below 1GBps. UCS enables non-blocking 10Gps access bandwidth and so offers bandwidth in excess of 1GBps. However, the management VLAN should be enabled with rate-limiting to cap the traffic at 1GBps.

For Nexus 1000V rate-limiting configuration and restriction guidelines, see: http://www.cisco.com/en/US/docs/switches/datacenter/nexus1000/sw/4_0/qos/configuration/guide/qos_4policing.html.

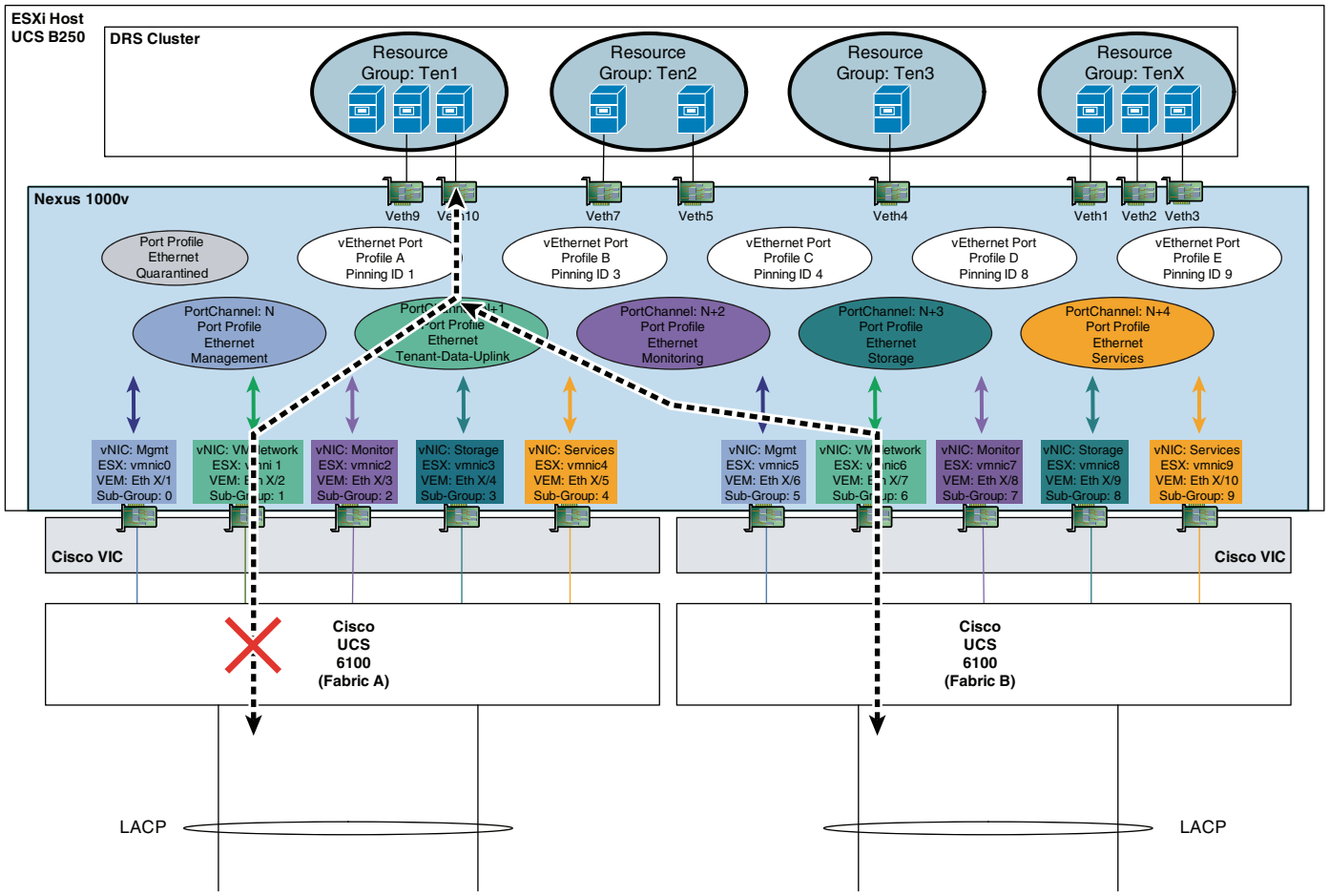
Traffic Engineering with End-to-End Service Assurance

QoS-Based Classification and Design Considerations for Network Service Assurance developed a design model for a multi-tenant environment. In this design, multiple types of traffic use the same class of service, which in turn uses the same queue/interface. With UCS redundant fabric (A and B) capability and the virtual port-channel host mode (vPC-HM) mac-pinning feature available in the 4.0(4)SV1(2) release for Nexus 1000V, it is possible to add much diversity in managing the traffic in a steady state condition. Dual fabric in UCS in conjunction with mac-pinning enables the possibility of traffic engineering such that a service class can split its traffic types, effectively doubling the classification buckets available. Of course during the failure of a fabric all traffic types will use the same path and resources. One of the big advantages of static pinning is to migrate the active/standby design that customers have been deploying with VMware vSwitches.

The mac-pinning is configured under each vEthernet port-profile such that when a port-profile is attached to the VM interfaces, any traffic from that interface mac-address is pinned to a given fabric ID via the port channel sub-group ID. In vPC-HM mac-pinning mode, a distinct sub-group ID is automatically assigned to each Ethernet interface on the VEM. Combine this functionality with the Cisco Virtual Interface Card's (VIC) capacity to present multiple virtual adapters to the hypervisor and the ability to direct and rate limit traffic becomes even more granular. The use of UCS service profiles ensures that the boot order and identity of these virtual interfaces is uniform across the cluster, providing consistency to the mac-pinning configuration. In the case of a failure of any component along the path, the Nexus 1000V uplink automatically select the available fabric for recovering the traffic.

Figure 39 explains the mac-pinning capability available in the latest Nexus OS release 4.0(4)SV1(3). The Cisco VIC virtual adapters are presented as Ethernet interfaces to the Cisco Nexus 1000V VEM, and as such are dynamically assigned a distinct vPC-HM sub-group ID. In the event of a failure on fabric A, the traffic assigned to the "Tenant-Data-Uplink" port channel sub-group ID "1" is reassigned to the remaining fabric available on vmnic6, sub-group ID "6".

Figure 39 MAC-Pinning and Failover



Port Channel	Sub-Group IDs
Management	0, 5
Tenant Data	1, 6
Monitoring	2, 7
Storage	3, 8
Services	4, 9

The multi-tenant model for all types of traffic and associated services classes along with their proposed differentiated services level is described in Table 16. Notice that the rationale for assigning traffic to various resources (fabric, UCS-class, queue) can vary based on each customer's preference. Each tenant can be configured with a unique CoS based on the needs of the applications that reside on that tenant. In this design, each tenant corresponds to a single application and, therefore, will be assigned a specific CoS for all data coming from that particular tenant. A more complicated design could involve assigning different levels of service to various applications residing on a single tenant. The key design point is that with UCS (dual-fabric) and Nexus 1000V (mac-based pinning), the customer can traffic engineer the multi-tenant user services level requirement with sufficient diversity.

Table 16 End-to-End Traffic Engineering Service Class Map

Traffic Type	Classification Category	CoS	Traffic Engineering Fabric/Class	Rational
NFS Data Store	VMkernel/Control	5	Fab-A/Platinum	Live ESX/VM OS Data
Nexus 1000V Control	System/Control	5	Fab-A/Platinum	Nexus 1000 Operation
Nexus 1000V Packet	System/Network-Control	5	Fab-A/Platinum	Nexus 1000 Operation
Platinum IO Low Latency, BW Guarantee	Tenant Data	5	Fab-B/Platinum	Load-share Fab-B wrt CoS 5 since NFS is in Fab-A
Platinum Transactional	Tenant Data	6	Fab-A/Platinum	Time Sensitive Traffic
Nexus 1000V Management	System/Control	6	Fab-B/Gold	Split Nexus 1000 control from Fab-A getting all
ESXi Service Console	vswif/Control	6	Fab-B/Gold	Same as above
Gold IO Med Latency, No Drop	Tenant Data	6	Fab-A/Gold DCE to buffer	Load-share Fab-A, since Platinum-IO is on Fab-A
Gold Transactional	Tenant Data	6	Fab-B/Gold	Time Sensitive Traffic
vMotion	VMkernel/Control	4	Fab-A/Silver	Rate Limited/not often, run to completion
Silver Transactional	Tenant Data	4	Fab-A/Silver	Competing with vMotion only when vMotion occurs
Silver IO High Latency, Drop/Retransmit	Tenant Data	4	Fab-B/Silver	Fab-A has vMotion
Bulk	Tenant Data	2	Fab-A/Bronze Fab-B/Bronze	Bulk and High Throughput Transaction

Note For more information on QoS in the secure multi-tenant architecture, see: http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/Virtualization/securecldg.html.

Design Considerations for Storage I/O Assurance

NetApp FlexShare

NetApp FlexShare allows the storage administrator to prioritize workloads, which increases control over how the storage system resources are used. Data access tasks that are executed against a NetApp controller are translated into individual read or write requests, which are processed by WAFL® (Write Anywhere File Layout) within the storage controller's operating system, Data ONTAP. As WAFL processes these transactions, requests are completed based on a defined order versus the order in which they are received. When the storage controller is under load, the FlexShare defined policies prioritize resources including system memory, CPU, NVRAM, and disk I/O based on business requirements.

With FlexShare enabled, priorities are assigned to either volumes containing application data sets or operations executed against a NetApp controller. FlexShare uses these defined priorities to logically arrange the order in which tasks are processed within the system. All WAFL requests are processed at the same speed regardless of importance, but FlexShare prioritizes these operations. For example, the data for a tenant that has a Platinum service level is given preferential treatment because it has a higher priority compared to tenants with Gold, Silver, or Bronze service levels.

Operations performed against a NetApp controller are defined as either user or system operations, providing yet another layer of prioritization. The ones that originate from a data access request, such as NFS, CIFS, iSCSI, or FCP, are defined as user operations and all other tasks are defined as system operations. An administrator can define policies in which data access is processed prior to tasks such as restores and replication to make sure that service levels are honored as other work is executed.

When designing a multi-tenant architecture, it is important to understand the different workloads on the storage controller and the impact of setting priorities on the system. Improperly configured priority settings can impact performance, adversely affecting tenant data access. Adhere to the following guidelines when implementing FlexShare on a storage controller:

- Enable FlexShare on all storage controllers.
- Make sure that both nodes in a cluster have the same priority configuration.
- Set priority levels on all volumes within an aggregate.
- Set volume cache usage appropriately.
- Tune for replication and backup operations.

For more information, refer to: http://www.netapp.com/us/products/platform-os/flexshare.html and to the FlexShare Design and Implementation Guide.

Storage Reservation and Thin Provisioning Features

Thin provisioning with NetApp is a method of storage virtualization that allows administrators to address and oversubscribe the available raw capacity. A common practice within the storage industry is to allocate the projected capacity from the pool of available resources as applications or virtual machines are deployed. However, storage is often underutilized before the actual capacity used matches the projected requirements. Thin provisioning allows enterprises to purchase storage as required without the need to reconfigure parameters on the hosts that attach to the array. This saves organizations valuable money and time with respect to the initial purchase and subsequent administration overhead for the life of the storage controllers. Thin provisioning provides a level of “storage on demand” as raw capacity is treated as a shared resource pool and is only consumed as needed.

When deploying thin-provisioned resources, administrators should also configure associated management policies on the thinly provisioned volumes within the environment. These policies include volume auto-grow, Snapshot auto-delete, and fractional reserve. Volume auto-grow is a space management feature that allows a volume to grow in defined increments up to a predefined threshold. Snapshot auto-delete is a policy related to the retention of Snapshot copies, protected instances of data, providing an automated method to delete the oldest Snapshot copies when a volume is nearly full. Fractional reserve is a policy that allows the percentage of space reservation to be modified based on the importance of the associated data. When these features are used concurrently, Platinum-level tenants have priority to upgrade their space requirements. In effect, a Platinum tenant would be allowed to grow its volume as needed, and the space would be reserved from the shared pool. Conversely, lower-level tenants would require additional administrator intervention to accommodate requests for additional storage.

The use of thin provisioning features within a multi-tenant environment provides outstanding ROI as new tenants are deployed and grow, which requires more storage. Environments can be designed to improve storage utilization without having to reconfigure the UCS and virtualization layer. Using management policies can distinguish resource allocation afforded to tenants of varying service levels.

For additional details regarding thin provisioning and the latest best practices, refer to the following technical reports:

- NetApp Thin Provisioning: Improving Storage Utilization and Reducing TCO (http://media.netapp.com/documents/tr-3563.pdf)
- Thin Provisioning in a NetApp SAN or IP SAN Enterprise Environment (http://media.netapp.com/documents/tr-3483.pdf)

Management

Domain and Element Managers

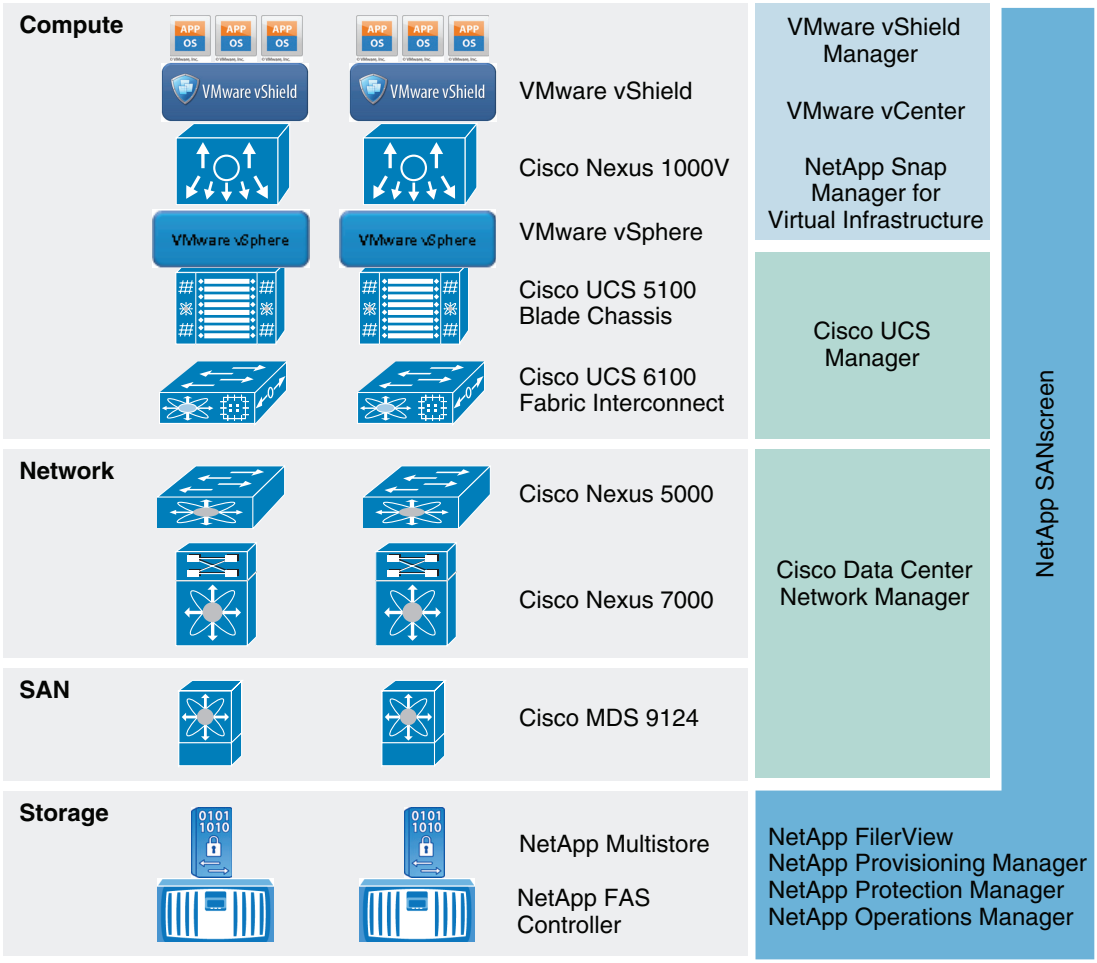
As the storage demands of multi-tenant environments grow, so do the challenges of managing them. Multi-tenant service providers require comprehensive control and extensive visibility of their shared infrastructure to effectively enable the appropriate separation and service levels for their customers. The varied and dynamic requirements of accommodating multiple customers on a shared infrastructure drive service providers toward storage management solutions that are more responsive and comprehensive while minimizing operational complexity.

In its current form, components within each layer are managed by:

- vCenter (http://www.vmware.com/products/vcenter/)
- vCenter Chargeback (http://www.vmware.com/products/vcenter-chargeback/)
- vShield Manager (http://www.vmware.com/products/vshield/)
- vCloud Director (see vCloud Management)
- UCS Manager (http://www.cisco.com/en/US/partner/products/ps10281/index.html)
- Data Center Network Manager (http://www.cisco.com/en/US/partner/products/ps9369/index.html)
- NetApp FilerView (http://www.netapp.com/us/products/platform-os/filerview.html)
- Provisioning Manager (http://www.netapp.com/us/products/management-software/provisioning.html)
- Protection Manager (http://www.netapp.com/us/products/management-software/protection.html)
- SnapManager for Virtual Infrastructure (http://www.netapp.com/us/products/management-software/snapmanager-virtual.html)
- NetApp Virtual Storage Console for vSphere (http://www.netapp.com/us/products/management-software/vsc/virtual-storage-console.html)
- Operations Manager (http://www.netapp.com/us/products/management-software/operations-manager.html)
- SANscreen (http://www.netapp.com/us/products/management-software/sanscreen/sanscreen.html)

These are illustrated in Figure 40. This section discusses the options available to cloud and tenant administrators for managing across compute, network, and storage.

Figure 40 Management Components



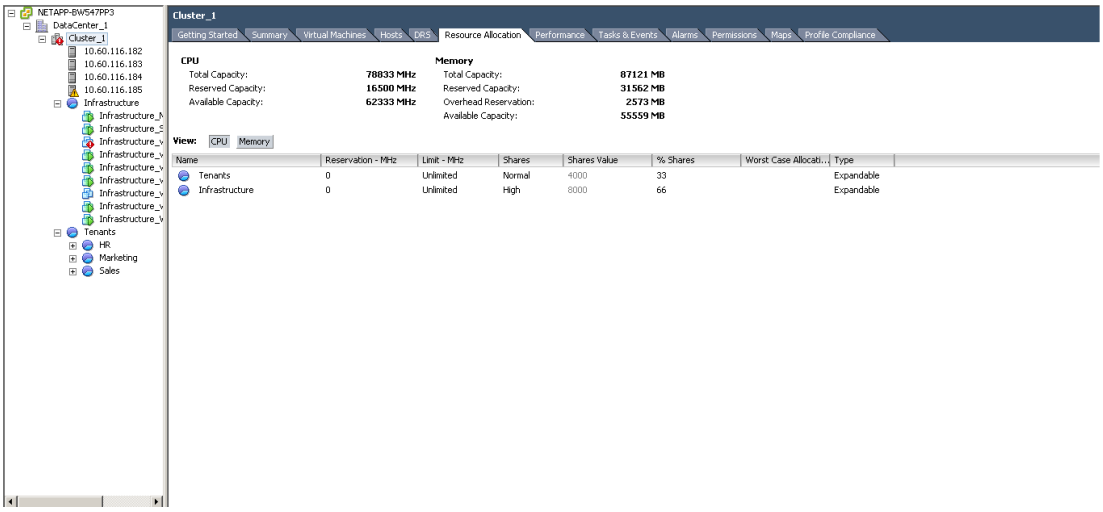
Compute Domain

VMware vSphere Resource, Capacity, and Health Management

VMware vCenter simplifies resource and capacity management for both Cloud and Tenant Administrators. Here are a few main points about vSphere management features used in multi-tenant environments:

- The Resource Allocation Tab in vCenter Server (Figure 41) displays detailed CPU and memory allocation at individual resource pool and virtual machine levels. A Cloud Administrator can use information provided at the cluster level to get an overview of CPU and memory resources allocated to infrastructure virtual machines and individual tenants.

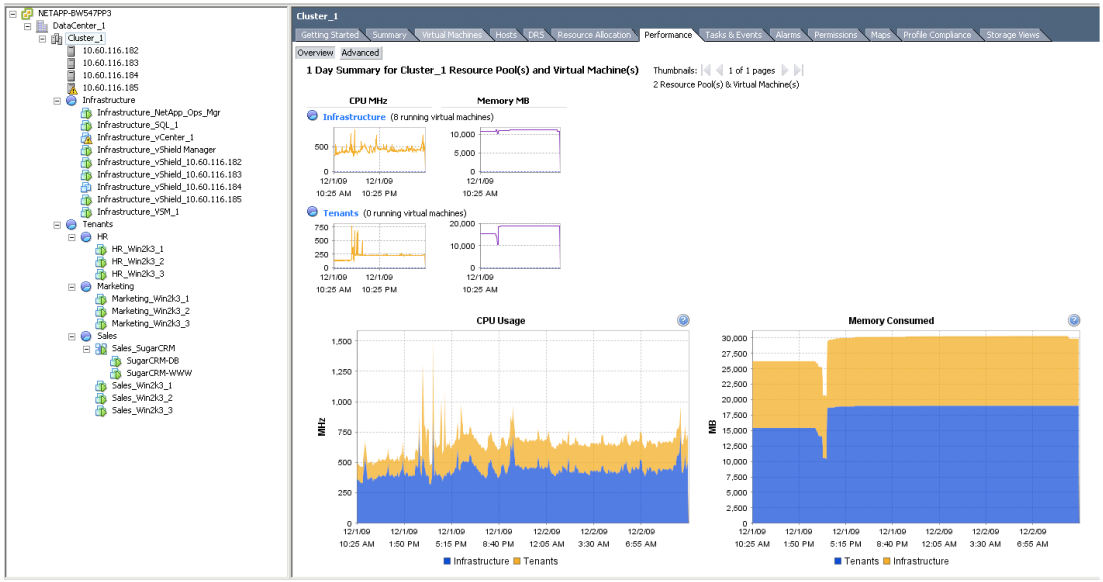
Figure 41 Resource Allocation Tab in vCenter Server



227997

- Tenant Administrators can use information provided at the resource pool level to get an overview of CPU and memory resource allocated to the virtual machines or vShield Apps.
- The performance charts in vCenter Server (Figure 42) provide a single view of all performance metrics at both the data center and individual resource pool level. Information such as CPU, memory, disk, and network is displayed without requiring you to navigate through multiple charts. In addition, the performance charts include the following views:
  - Aggregated charts show high-level summaries of resource distribution, which helps Cloud and Tenant Administrators identify the top consumers.
  - Thumbnail views of virtual machines, hosts, resource pools, clusters, and datastores allow easy navigation to the individual charts.

Figure 42 Performance Charts in vCenter Server

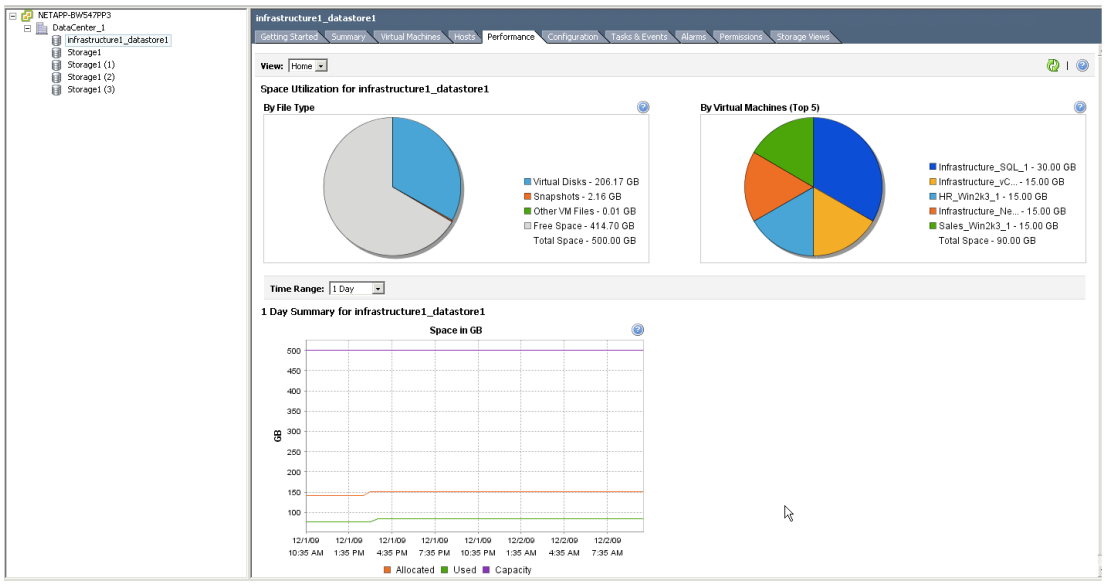


227998

- The vCenter Storage plug-in provides detailed utilization information for all datastores dedicated for infrastructure and tenant virtual machines. The following information is available for the Cloud Administrator for each datastore (NFS, iSCSI, or FCP):
  - Storage utilization by file type (virtual disks, snapshots, configuration files)
  - Individual virtual machine storage utilization overview
  - Available space

Note NetApp MultiStore provides the flexibility to dedicate specific NFS or iSCSI volumes to individual tenants. In that situation, the vCenter Storage plug-in can be used by the Tenant Administrator to monitor their respective datastore. To do that, however, permission also needs to be explicitly assigned to the group of Tenant Administrator to enable secure and isolated access.

Figure 43 vCenter Datastore Utilization Chart



- NetApp Virtual Storage Console (a vCenter plugin) is complementary to the vCenter Storage plugin. The Storage Console allows the Cloud Administrator to get a holistic view of storage utilization from the vSphere datastore level, to volume, LUN, and aggregate levels in the NetApp FAS storage controller.
- Events and Alarms in vCenter Server provide better monitoring of infrastructure resources. Low level hardware and host events are now displayed in vSphere Client to quickly identify and isolate faults. Alarms can now be set to trigger on events and notify the Cloud Administrator when critical error conditions occur. In addition, alarms are triggered only when they also satisfy certain time interval conditions to minimize the number of false triggers. vCenter Server provides a number of default alarm configurations to simplify proactive infrastructure management for Cloud Administrators.

Recommended alarms to configure in a multi-tenant shared services infrastructure are:

- At the Cluster level:
  - Network uplink redundancy loss (default alarm to monitor loss of network uplink redundancy on a virtual switch)

  - Network connectivity loss (default alarm to monitor network connectivity on a virtual switch)
  - Migration error (default alarm to monitor if a virtual machine cannot migrate, relocate, or is orphaned)
  - Cluster high availability error (default alarm to monitor high availability errors on a cluster)
- At the Datastore level:
  - Datastore usage on disk (default alarm to monitor datastore disk usage)
  - Datastore state for all hosts (this is not a default alarm; it needs to be defined to monitor the datastore access state for all hosts)

For Tenant Administrators, the following alarms are recommended to monitor individual tenant virtual machine resource utilization

- At the resource pool level:
  - Virtual machine cpu usage
  - Virtual machine memory usage
  - Virtual machine total disk latency

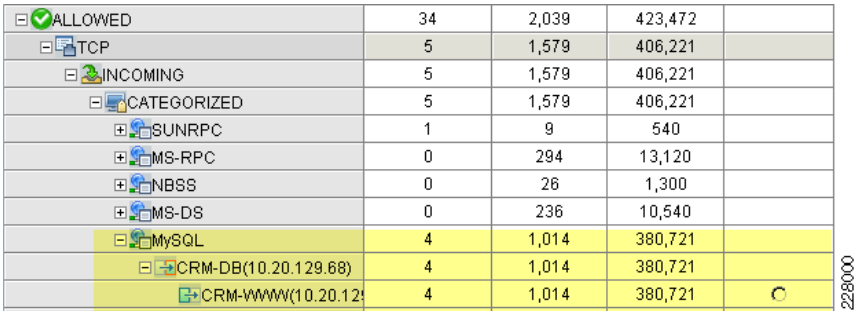
VMware vShield Resource Management

vShield provides visibility into the virtual network and maps out how network services are accessed between virtual and physical systems by displaying microflow level reports. Each network conversation is recorded with statistics such as source and destination IP addresses mapped to virtual machine names, TCP/UDP ports, and protocol types across all layers of the OSI model. Each microflow is mapped to virtual machine, cluster, virtual data center, or network containers such as vSphere portgroup or the Nexus VLAN. This allows the user to browse the flows at top levels, such as virtual data center or have a granular report at each virtual machine level. Each microflow is also marked as allowed or blocked to track the firewall activity and it is possible to create firewall rules right from the flow reports to immediately stop malicious activity or modify existing firewall rules. The following are the common use cases for these VMflow reports:

- During the initial installation of new applications or entire tenant environments, it is important to audit the virtual network to discover which protocols and ports are needed to be open on the firewall to achieve a positive security model where only needed protocols are admitted.
- Historical usage information in bytes or sessions sliced by protocol, application, virtual machine, or even entire data center, allows for capacity planning or tracking application growth.
- Troubleshooting of firewall policies without the need to require users to repeat the operation which is failing. This is no longer needed since all history is kept and blocked flows are visible at virtual machine levels.

Figure 44 depicts the logging capability of vShield, where traffic can be analyzed at the protocol level.

Figure 44 Logging Capability of vShield



In addition to virtual network flow information, firewall management requires the administrator to understand which operating system network services and applications are listening on virtual machines. Not all default network services need to be accessible and should be locked down to avoid exposure of various vulnerabilities—vShield provides such inventory on per virtual machine basis. The combination of service and open port inventory per virtual machine plus the network flow visibility eases the job of the administrator in setting up and managing virtual zones and access to tenant resources.

VMware Chargeback

vCenter Chargeback enables the administrators to associate resource utilization with a cost. Though most enterprises do not actually charge for the internal service offerings, the reports generated by Chargeback can help bringing awareness to all consumers of IT resources. This typically helps reducing the problems of wasteful resource requests and allocations.

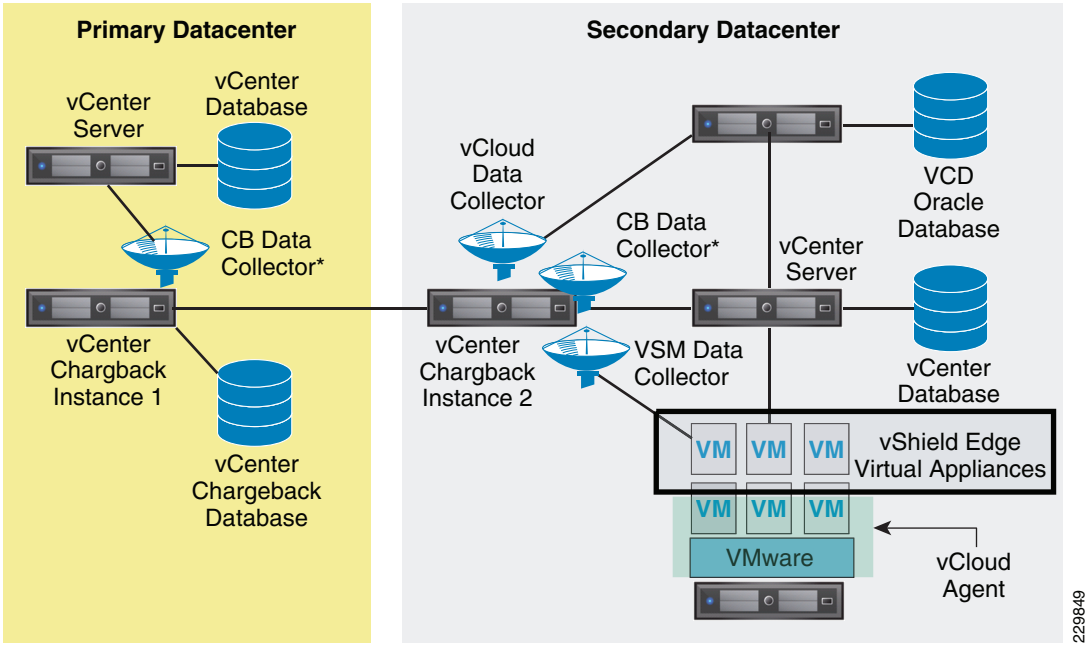
It is important to understand the architectural components of vCenter Chargeback prior to fitting it into an Enhanced Secure Multi-Tenancy architecture. vCenter Chargeback consists of the following core components:

- Chargeback application
- Data Collector:
  - Chargeback data collector—Responsible for vCenter Server data collection
  - vCloud and vShield Manager (VSM) data collector—Responsible for collection of utilization/allocation on new abstraction layer created by vCloud Director (i.e., virtual data center, vApp, media, templates, external network bandwidth, network count)
- Load Balancer (only needs to be installed once for the Chargeback cluster)
- Chargeback database

When the first instance of vCenter Chargeback is installed, the data collector (Chargeback data collector is installed by default with vCloud Director and VSM data collectors being optional) and load balancer are installed along with the application. Additional instances can be installed and added to the original instance to form a cluster for scalability. In a multi-site Enhanced Secure Multi-Tenancy architecture deployment, the following can be used as a reference design:

- Primary Datacenter—First instance of Chargeback with Chargeback data collector and load balancer installed
- Secondary Datacenter—Second instance of Chargeback with Chargeback, vCloud Director and VSM data collectors installed

Figure 45 VMware Chargeback Multi-Site Topology



- Given the vCloud Director environment is serving the pre-production dev/test environments, there is only one instance of the vCloud Director and the VSM Data Collector installed. If high availability is desired for charge report generation for both production and pre-production environments, then both primary and secondary Chargeback instances need to have all three data collectors installed.
- In current release of vCenter Chargeback (version 1.5), there is no site locality awareness for the data collectors. Chargeback data collector on primary data center may communicate with vCenter Server from either primary or secondary site; the same goes for the Chargeback data collector on the secondary site. As for the vCloud and VSM data collectors, they only communicate with vCloud environment in the secondary datacenter as that is the only deployment available for communication.
- The current release of vCenter Heartbeat (version 6.3) does not have built-in support for Chargeback failover. However, Chargeback can co-exist with vCenter Servers protected by vCenter Heartbeat.

Hierarchy Management for Chargeback

vCenter Chargeback interacts with the vCenter Server to determine the utilization of the computing resources by various virtual machines that are created in the vCenter Server hierarchy. vCenter Chargeback enables administrators to create multiple chargeback hierarchies, which can be different from the vCenter Server hierarchies. Hierarchies created in vCenter Chargeback can be logical parents of any of the vCenter server inventory objects such as ESX host, virtual machine folder, and resource pools; such logical parents can represent a department or business unit within an enterprise.

Design Considerations

If the vCenter hierarchy already matches up with the way organizations, business units, or departments are mapped within an enterprise, then a Chargeback hierarchy can be created for chargeback metering and reporting; otherwise, custom hierarchy can be created by placing vCenter entities (resource pool, host folder, VM folder) into it. In this ESMT architecture, vCenter Server hierarchy is organized by production, pre-production, and management clusters as unit of separation, therefore no custom

Chargeback hierarchy needs to be created. To accommodate differences in enterprise practices, vCenter Chargeback provides the flexibility to accommodate it. This design is a reference with simplicity in mind; if desirable, custom hierarchy model can be used.

To successfully generate a cost report, it is imperative to understand the chargeback cost elements defined by vCenter Chargeback. Refer to vCenter Chargeback User Guide (http://www.vmware.com/pdf/vCenterChargeback_v_1_5_Users_Guide.pdf) to fully understand the following cost -related elements prior to any report generation:

- Chargeable computing resource
- Base rate
- Rate factor
- Fixed cost
- Billing policy
- Cost model
- Cost template

Cisco UCS Manager (UCSM)

The UCS platform is managed by a HTTP-based GUI interface. UCS manager provides a single point of management for the UCS system and manages all devices within UCS as a single logical entity. All configuration tasks, operational management, and troubleshooting can be performed through the UCS management interface. The following is a summary of the basic functionality provided by the UCS manager:

- It manages the fabric interconnect, the chassis, the servers, the fabric extender, and the adapters.
- It provides hardware management, such as chassis and server discovery, firmware management, and backup and configuration restore functionality.
- It manages system wide pools that can be shared by all the servers within the chassis. These pools include MAC pools, World Wide Node Name (WWNN) pools, World Wide Port Name (WWPN) pools, and the Universally Unique Identifier Suffix (UUID) pools.
- It can be used to define service-profiles, which are logical representations of a physical servers that include connectivity as well as identity information.
- Create an organizational hierarchy as well as role-based-access control.
- Create configurational and operational policies that determine how the system behaves under specific circumstances. Some of the policies that can be set include:
  - Boot policy—Determines the location from which the server boots
  - QOS definition policy—Determines the outgoing QoS parameters for a vNIC or vHBA
  - Server discovery policy—Determines how system reacts when a new server is discovered
  - Server pool policy—Qualifies servers based on parameters such as memory and processor power
  - Firmware policy—Determines a firmware version that will be applied to a server
  - Port, adapter, blade, and chassis policy—Defines intervals for collection and reporting of statistics for ports, adapter, blades, and chassis respectively

Network Domain

Cisco Network and UCS Infrastructure Management

The Data Center Network Manager

The Cisco Data Center Network Manager (DCNM) provides an effective tool to manage the data center infrastructure and actively monitor the storage area network (SAN) and local area network (LAN). In this design one can manage the Cisco Nexus 5000 and 7000 switches. Cisco DCNM provides the capability to configure and monitor the following features of Nexus-OS.

- Ethernet switching
  - Physical ports and port channels
  - VLANS and private VLANs
  - SPT protocol
  - Loopback and management interfaces
- Network security
  - Access control lists and role-based access control
  - Authentication, authorization and accounting services
  - ARP inspection and DHCP snooping-storm control, and port security
- Fibre-Channel
  - Discovery and configuration of zones
  - Troubleshooting, monitoring, and configuring of fibre-channel interfaces
- General
  - Virtual Device Context and SPAN analyzer
  - Gateway Load Balancing Protocol

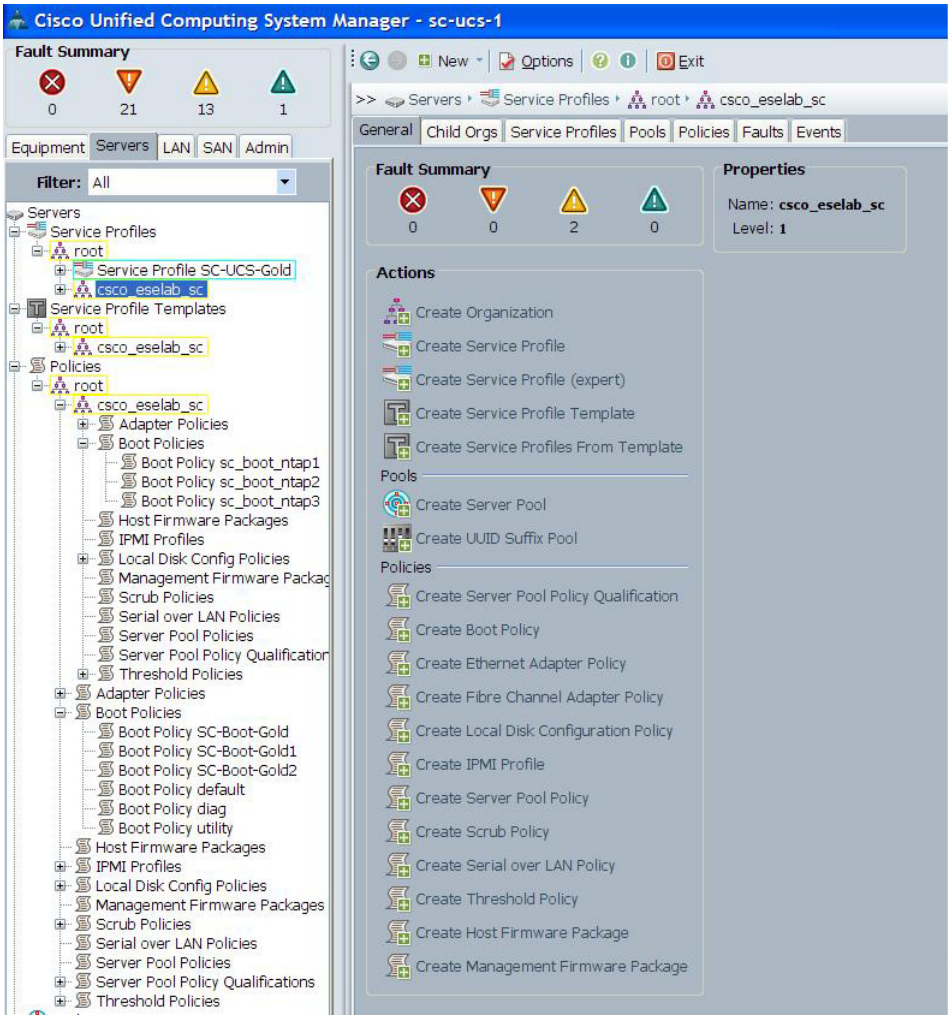
DCNM also provides the capability for hardware inventory and event browser and is embedded with a topology viewer that performs device discovery, statistical data collection, and client logging.

Deploying DCNM and Using UCS Manager in a Secure Cloud

A stateless computing model dictates that booting from remote storage is transparent and logically independent from the physical device. The UCS platform provides the capability to define SAN boot parameters that can be moved from one physical blade to the other without resorting to additional configuration of boot parameters and network parameters within the chassis and the network. This would allow a logical server to be booted from a remote storage on different blades at different times.

The UCS manager can be used to implement policies that correspond to each of the tenant's requirements. For example, server pools can be used to reserve server platforms with specific hardware requirements and QoS parameters can be used to set system-wide policies to address tenant SLAs. Figure 46 depicts the management interface that is used to set various policies within a UCS chassis.

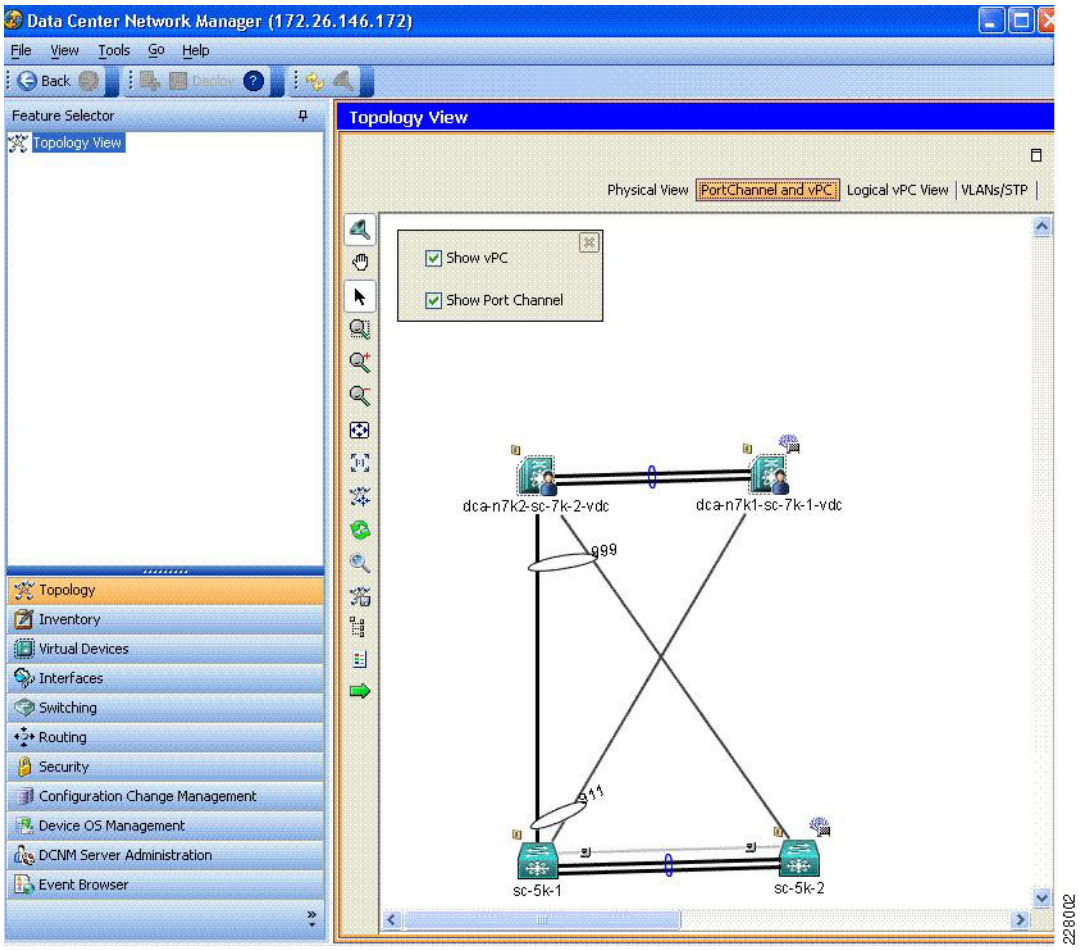
Figure 46 Management Interface for Setting Policies Within a UCS Chassis



DCNM in conjunction with UCS manager can also be used to implement and monitor VPC connectivity between the 6100 fabric interconnect Nexus 5000 and between Nexus 5000 and 7000. VPC functionality provides a redundant topology that is loop-less, which implies fast convergence time after a failover.

Operationally, with DCNM one can view configurations of Nexus 5000 and Nexus 7000 switches and take snapshots of configuration changes. In addition the managed devices can send their logging information to the DCNM. DCNM also provides one with a graphical view of the network infrastructure. Figure 47 depicts how DCNM an be used to get a topological view of the network infrastructure.

Figure 47 Topological View of Network Infrastructure



Cisco Network Service Delivery Management

A secure multi-tenant environment compels administrators to observe and to control the myriad of tenant traffic patterns occurring within the data center. To meet these objectives, network service devices may be readily introduced into or accessed within the shared environment. These network service management devices include:

- Cisco Application Network Manager (ANM)
- Cisco Security Manager (CSM)
- Cisco Network Analysis Modules and Appliances (NAM)

The Cisco ANM allows for centralized configuration, operation, and monitoring of Cisco data center networking equipment and services. Cisco ANM provides this management capabilities for the following Cisco products:

- Cisco ACE modules and 4710 appliances
- Cisco Content Services Switch (CSS)
- Cisco Content Switching Module (CSM)
- Cisco Content Switching Module with SSL (CSM-S)

- Cisco ACE Global Site Selector (GSS)

The Cisco ANM integrates with VMware vSphere environments, providing continuity between the application server and network operator. Cisco ANM can add, delete, activate, and suspend traffic and change load-balancing weights for virtual machines benefiting from Cisco ACE load-balancing services. From within VMware vCenter, administrators have access to Cisco ANM's real-server monitoring graphs with real time application performance information. The ANM expedites implementation by allowing discovery tools to automate importation and mapping of virtual machines to existing Cisco ACE server objects. All of these functions are subject to role based authentication and authorization.

The Cisco Security Manager (CSM) allows one to manage both Cisco firewall and intrusion prevention devices—devices that may be virtualized to secure specific tenant and therefore application environments. The CSM provides a comprehensive view of the security policies being applied and events occurring on the devices it manages. The CSM is typically managed by the security administrators within the enterprise requiring authenticated access.

The Cisco network analysis devices provide administrators visibility into data center traffic. The NAM devices provide performance monitoring and troubleshooting capabilities. These devices are available as service modules, physical appliances, or as Virtual Service Blades as part of the Nexus 1010 platform. Combine the any of the NAM capabilities with the Nexus 1000V ERSPAN and Netflow functionality and administrators have a virtual NIC view of the data center and its applications.

The Cisco Secure Access Control Server (ACS) enforces the access control policy for network or service devices within the secure multi-tenant data center. The Cisco ACS supports AAA protocols such as TACACS+ and RADIUS as well as directory databases such as LDAP and Active Directory. Centralizing authentication across the infrastructure is ideal. The ACS server allows one to choose a central repository for user information and the flexibility to offer that data as an authentication service to multiple platforms using the protocol of their choice.

Storage Domain

NetApp Storage Infrastructure and Service Delivery Management

Multi-tenant service providers and enterprises require comprehensive control and extensive visibility of their shared infrastructure to effectively provide the appropriate separation and service levels for their customers or application environments. NetApp provides cohesive management solutions that enable providers to achieve dramatically improved efficiency, utilization, and availability. NetApp offers a holistic approach focused on simplifying data management that effectively addresses the particular operational challenges of these environments.

User interfaces such as FilerView and the Data ONTAP® command line are instrumental in the provider's initial build-out of the cloud service architecture. However, for the subsequent processes involved with the routine service operations, the use of these interactive tools should be discouraged in favor of the standards-oriented, policy-driven facilities of the NetApp management software portfolio. The following sections introduce data management structures and concepts to consider when designing a storage management solution for Enhanced Secure Multi-Tenancy architectures.

Resource Groups

Effective data management in a multi-tenant, shared infrastructure solution begins with grouping storage objects according to desired relationships. NetApp Operations Manager allows the flexible creation of resource groups that can address common characteristics such as the underlying storage systems (version, capability, configuration, and so on), geographical location, application environments, business units, or departments. Groups and subgroups can define scope around the operational management of storage infrastructure, providing granular controls to delegate storage administration that intuitively aligns with organizational models. Grouping can help organize multi-tenant resource

management with alerting, reporting, and role-based access control. For example, tenant organizations and sub-organizations can be defined into groups to apply access controls or to direct reports and alerts to the appropriate administrators.

Resource Pools

While resource groups organize logical relationships across storage objects, resource pools describe the physical organization of storage systems and aggregates to be sourced for provisioning operations. Resource pools can be used to group storage according to common attributes such as size, cost, performance, and availability. Create resource pools according to how storage will be used or preferentially treated; for example, by separating primary from secondary data across varying performance configurations or by grouping homogeneous storage types. Within an Enhanced Secure Multi-Tenancy architecture, resource pools can be used to design and prescribe how storage is sourced from the shared storage infrastructure to accommodate various tenant service requirements.

Data Sets

A data set is a collection of data objects (volumes, qtrees, LUNs) that are managed as a single unit, following the same provisioning and protection requirements. A data set includes primary data objects as well as all replicas of those objects derived from their data protection configuration. For example, a data set can include the primary application data within a tenant environment, any associated replicated data on secondary storage, and cascaded replication targets at a remote site. Storage administrators should design data sets to segregate particular application requirements, configuration standards for data objects, or common protection requirements.

Provisioning and Protection Policies

NetApp Provisioning Manager and Protection Manager provide a cohesive solution to automate storage provisioning, data protection, and data migration. Storage administrators create provisioning policies, which standardize and govern how data is provisioned and organized, to deliver the desired structure, placement, performance, efficiency, isolation, and availability. Likewise, protection policies can fully automate the provisioning and configuration of the desired data protection design and processes for primary data and secondary, tertiary, and further replication. Provisioning and protection policies prescribe how storage objects are allocated from underlying resource pools into data sets that are logically associated with resource groups.

Resource Labels

Resource labels provide a way to refine the scope of resources used within a provisioning or protection policy. Labels can be associated with resource pools or specific resource pool members. When a policy-driven provisioning or protection operation is initiated, labels can be used to restrict the resources available to satisfy that request. Resource labels provide a valuable filtering mechanism for policy design and data management within a hierarchical multi-tenant environment. For example, a storage administrator can apply granular control over which storage resources are used to satisfy requests for specific storage service offerings or for particular tenant organizations or applications.

vFiler Unit Templates

A vFiler unit template is a baseline configuration to help standardize vFiler unit deployment. vFiler unit templates include configuration settings such as CIFS, DNS, NIS, and host specifications that may apply to multiple vFiler units. For example, a cloud administrator could design a vFiler unit template for each tenant organization or sub-organization to define default configuration values to standardize all vFiler units deployed on behalf of a tenant.

Storage Services

Storage administrators can combine provisioning and protection policies with resource pool assignments and vFiler unit templates to create a consolidated storage service. Unique storage services can be designed according to desired service-level objectives, such as Platinum, Gold, Silver, or whatever context is desired. Provisioning Manager includes a Storage Service Catalog facility to publish these

storage services to administrators and automation frameworks. By selecting the desired storage service from this catalog, new data sets can automatically be created to comply with desired provisioning and protection policies in a single workflow. In an Enhanced Secure Multi-Tenancy architecture, storage services can be designed to deliver tenant data objects that comply with baseline storage offerings defined by the cloud storage administrator. Storage services can also be further tailored to meet the specific needs of a particular tenant, organization, or application.

NetApp SANscreen

NetApp SANscreen® provides administrators with end-to-end service path assurance and insight into the shared infrastructure assigned to support the resources deployed to tenant environments. SANscreen discovers and correlates resources, configuration, and utilization data collected from data sources (the devices and element managers that comprise the Enhanced Secure Multi-Tenancy infrastructure). Administrators can configure service path policies and associate them throughout the infrastructure according to availability or service-level objectives such as path redundancy, resource isolation, and capacity requirements. SANscreen then monitors those service paths and reports any policy violations. Administrators can classify resources across many annotation types such as organizational unit, application, geographical location, infrastructure segmentation, service tier, or other custom criteria. Annotation rules can be defined to automatically classify resources upon discovery into prescribed annotation types according to discovered configuration properties. For example, a storage quality service tier may be automatically classified as Platinum, Gold, or Silver according to volume type, configuration, or performance characteristics. Correlating these annotations across the resource and utilization data provides a rich, contextual view into tenant environments and their application resources. This allows SANscreen to provide granular monitoring, reporting, and trending analysis to cloud and tenant administrators. This includes chargeback data when cost metrics are applied.

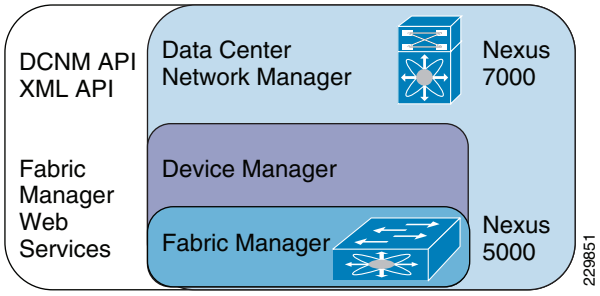
Open Framework Approach

The components in this design have a rich distribution of element managers and open APIs. These APIs allow for customization of configuration, monitoring, and process automation through third-party or independently-created interfaces.

Cisco Network and Storage Managers

Cisco network and storage elements are managed by DCNM and Fabric Manager, which are reachable through XML APIs. NX-OS uses NETCONF (http://tools.ietf.org/html/rfc4741) over SSH for remote XML management. You can configure NX-OS devices by using the Cisco DCNM web services API on the DCNM server. You create XML-based API requests by using the SOAP protocol.

Figure 48 Nexus 7000 and Nexus 5000



The DCNM server configures devices using the XML management interface. For more information on establishing the XML management interface on Cisco NX-OS devices, see: http://www.cisco.com/en/US/docs/switches/datacenter/sw/nx-os/xml/user/guide/xml_mgmt_interface.pdf.

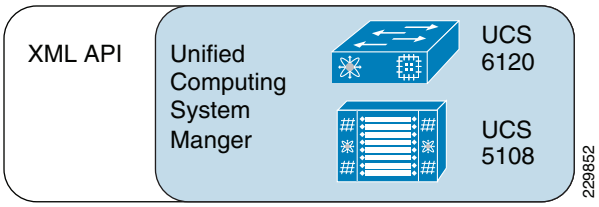
Nexus 7000 also leverages methods available in the DCNM Web Services API to initiate changes to Ethernet interfaces, VRFs, VPCs and VLANs.

Nexus 5000 can be natively set up to initiate XML-based call home messages for event notifications. Through the Fabric Manager API, extended visibility into connectivity can be reported to an external application. The DCNM API allows Ethernet configurations for the Nexus 5000, but zone changes are currently out of scope for the DCNM and Fabric Manager APIs.

For more information on the DCNM and Fabric Manager APIs, see:

- DCNM API: http://www.cisco.com/en/US/partner/docs/switches/datacenter/sw/4_2/dcnm/web_services/api/guide/overview.html
- Fabric Manager API: http://www.cisco.com/en/US/docs/switches/datacenter/mds9000/sw/4_1/configuration/guides/fm_4_1/wsrv.html#wp2185299

Figure 49 Unified Computing System

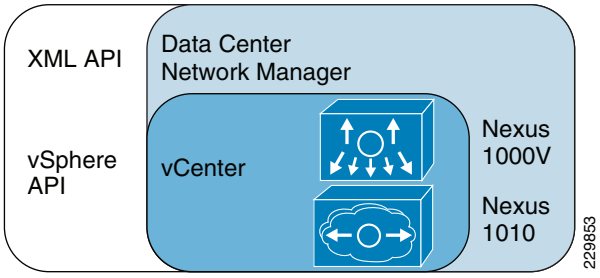


The UCS Manager API follows an object model-based approach, which distinguishes it from a traditional function call-based API. Instead of using a unique API function for each exposed task, the UCS responds to state changes made through the API (XML documents).

This object model approach gives complete coverage of the functionality of UCSM within the API. Every object referenced in the UCSM GUI or CLI is available within the XML API, because both the UCSM GUI and CLI are built from the XML API.

For more information on the UCSM XML API see: http://www.cisco.com/en/US/docs/unified_computing/ucs/sw/api/ucs_api.pdf

Figure 50 Nexus 1000V and Nexus 1010



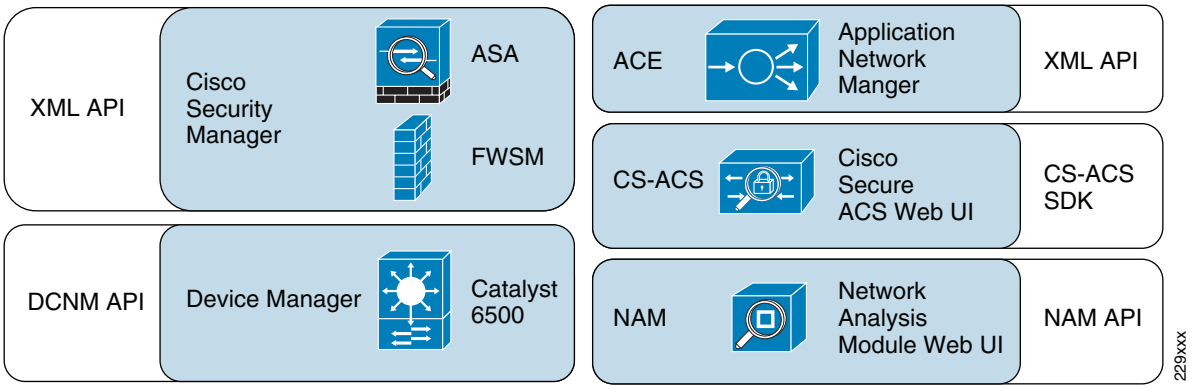
The Nexus 1000V and Nexus 1010 Virtual Services Appliance can both utilize NETCONF over SSH for remote XML management. Standard operations covering interface, port-profile, VLAN, and QoS configurations are implemented through this XML API. The complete list of the supported CLI commands in XML can be referenced from the XML schema definitions (XSD) which are included in the Nexus 1000V download. In addition to the direct XML API, the Nexus 1000V can be managed and configured by the DCNM API.

For more information on the Nexus 1000V XML API see: http://www.cisco.com/en/US/docs/switches/datacenter/nexus1000/sw/4_0_4_s_v_1_3/xml_api/programming/guide/n1000v_xml_api_1oview.html.

Addition of hosts to the Nexus 1000V VSM or Nexus 1010 VSM Distributed Virtual Switch (DVS) can be automated through vSphere API with either VMware's vSphere Power CLI, vSphere Perl SDK, or vSphere Web Services SDK.

For more information on the vSphere API see: http://www.vmware.com/support/pubs/sdk_pubs.html.

Figure 51 Cisco Appliances and Service Modules



Cisco Application Performance, Security, and Monitoring appliances and service modules in this design have their own managers, APIs, and SDKs to expand orchestration possibilities. For more information about the APIs and SDKs of these products see:

- Cisco Security Manager: http://www.cisco.com/en/US/products/ps6498/ (The Cisco Security Manager does not currently have an API, but the ASA and FWSM are both configurable via CLI).
- Catalyst 6500: http://www.cisco.com/en/US/docs/switches/datacenter/sw/4_2/dcnm/web_services/api/guide/web_services.pdf
- Application Control Engine (ACE): http://www.cisco.com/en/US/docs/interfaces_modules/services_modules/ace/vA2_3_0/configuration/administration/guide/xml.html
- Cisco Secure Access Control Server(CS-ACS): http://www.cisco.com/en/US/docs/net_mgmt/cisco_secure_access_control_system/5.1/sdk/overview.html
- Network Analysis Module: http://developer.cisco.com/web/nam/resources

NetApp Manageability SDK

The NetApp Manageability SDK includes API library bindings in C/C++, Java, Perl, and .Net for all industry-leading operating systems and platforms. Detailed API documentation, design guides, and examples are provided for Data ONTAP, Operations Manager, Provisioning Manager, and Protection

Manager. The SDK provides WSDL and support for Web services over HTTP and HTTPS that can be used with various SOAP toolkits. Orchestration frameworks can leverage the NetApp Manageability SDK to simplify the delivery of data resources by using best-practice, policy-based storage services.

NetApp Virtual Storage Console SDK

The NetApp Virtual Storage Console SDK includes both Java and Web archive formats to facilitate SOAP development. Virtual Storage Console automation simplifies the best-practice provisioning and de-provisioning of VMware datastores and virtual machines within Enhanced Secure Multi-Tenancy architectures.

NetApp SANscreen Connect API

The NetApp SANscreen Connect API is based on SOAP and Web services, providing orchestration frameworks with access to the SANscreen change and service repository to report on devices, connectivity, configurations, service paths, policies, and violations. Enhanced Secure Multi-Tenancy providers can integrate SANscreen data into centralized IT dashboards, operations centers, and business processes.

VMware Manageability SDK

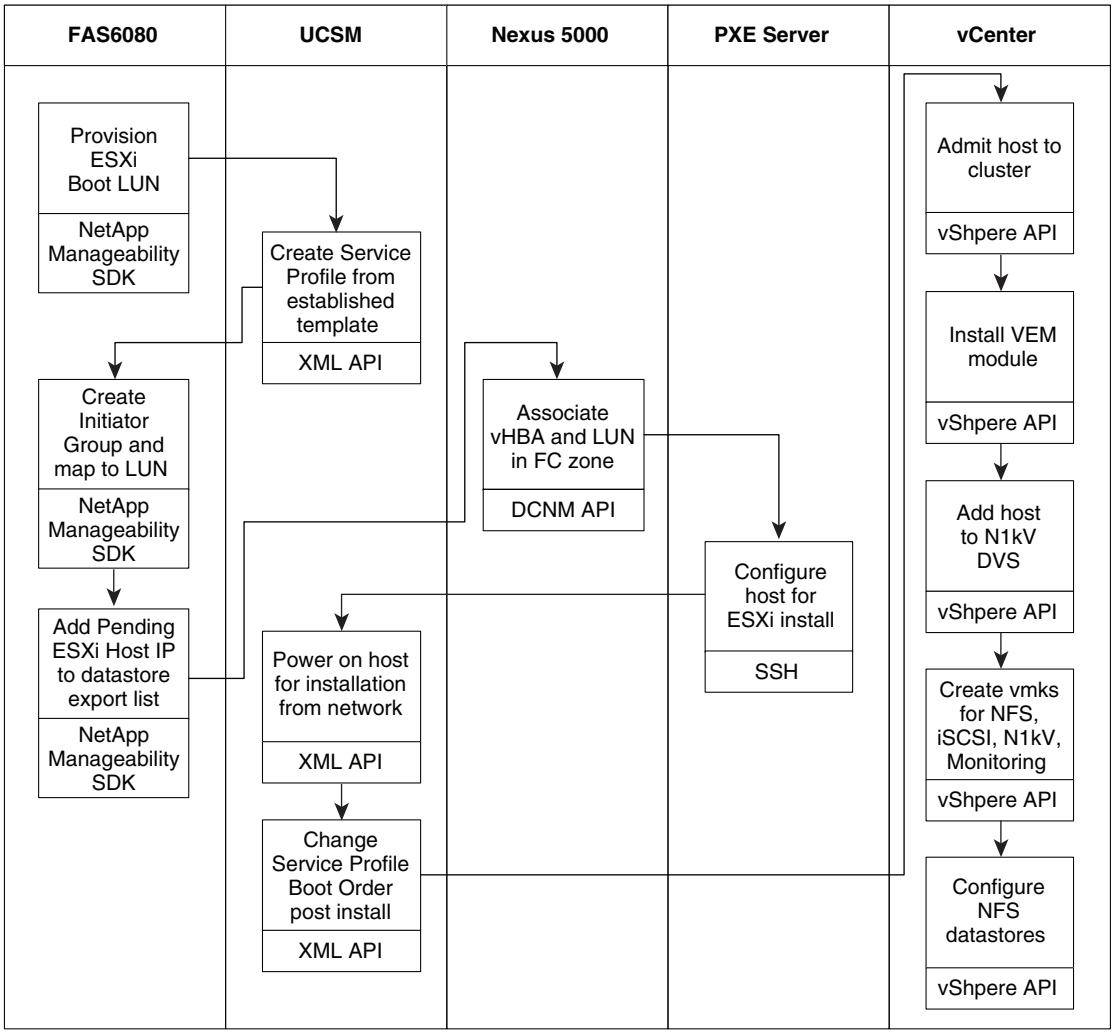
All VMware components referenced in this architecture has rich set of open APIs for custom scripting or third-party management solutions integration. Visit the following site for available APIs and documentation for the vSphere platform, vCloud Director, vShield, and vCenter Chargeback: http://communities.vmware.com/community/developer/

Example Use Cases/Flows

Workflow—Expand Tenant Compute Cluster

The workflow in Figure 52 presents a production placement with tenants active in the environment. A UCS B250 M2 is being added to an available slot in the production chassis.

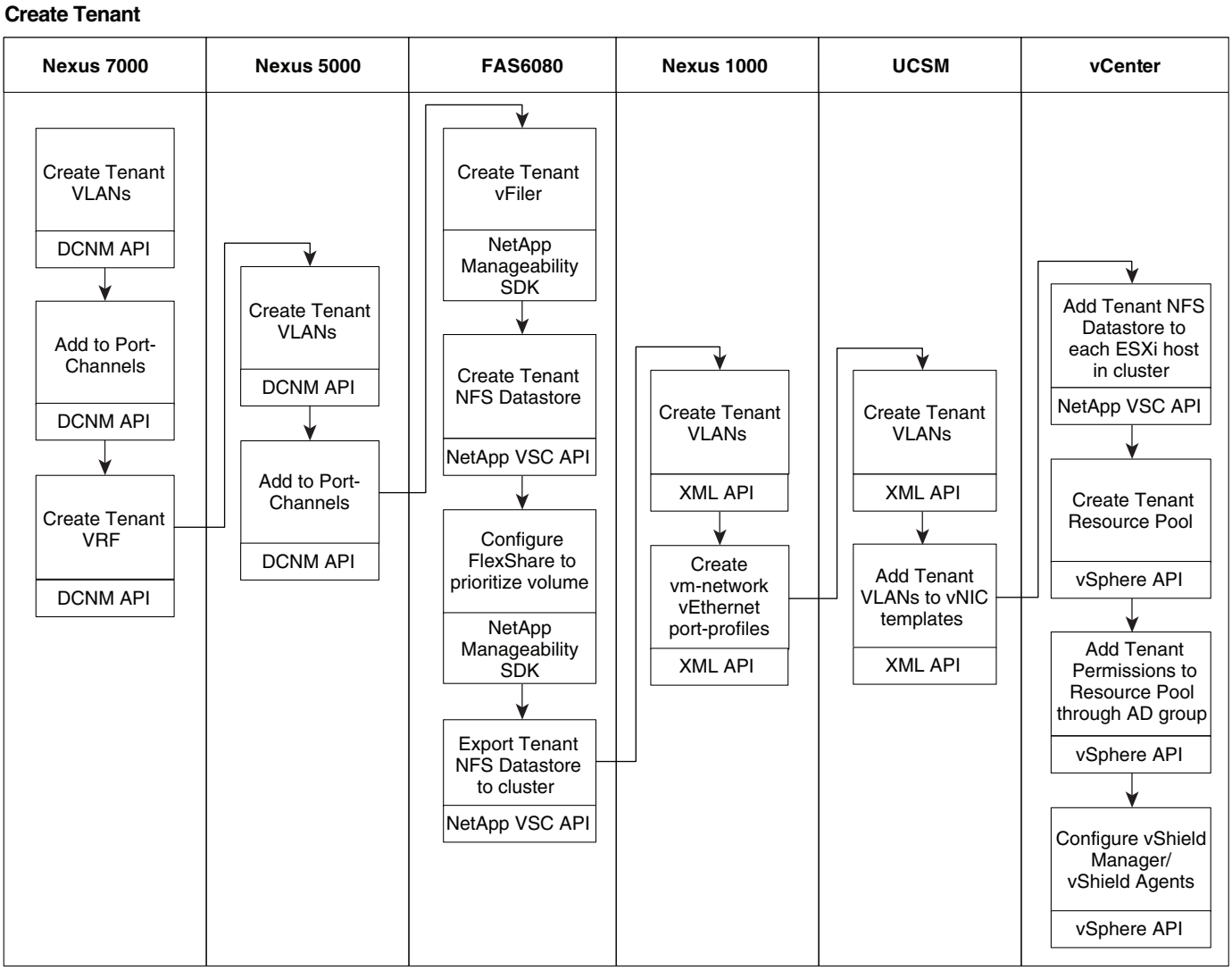
Figure 52 Expand Tenant Compute Cluster



Workflow—Create Tenant

The workflow in Figure 53 presents a production cluster with established tenants that are having a new tenant added.

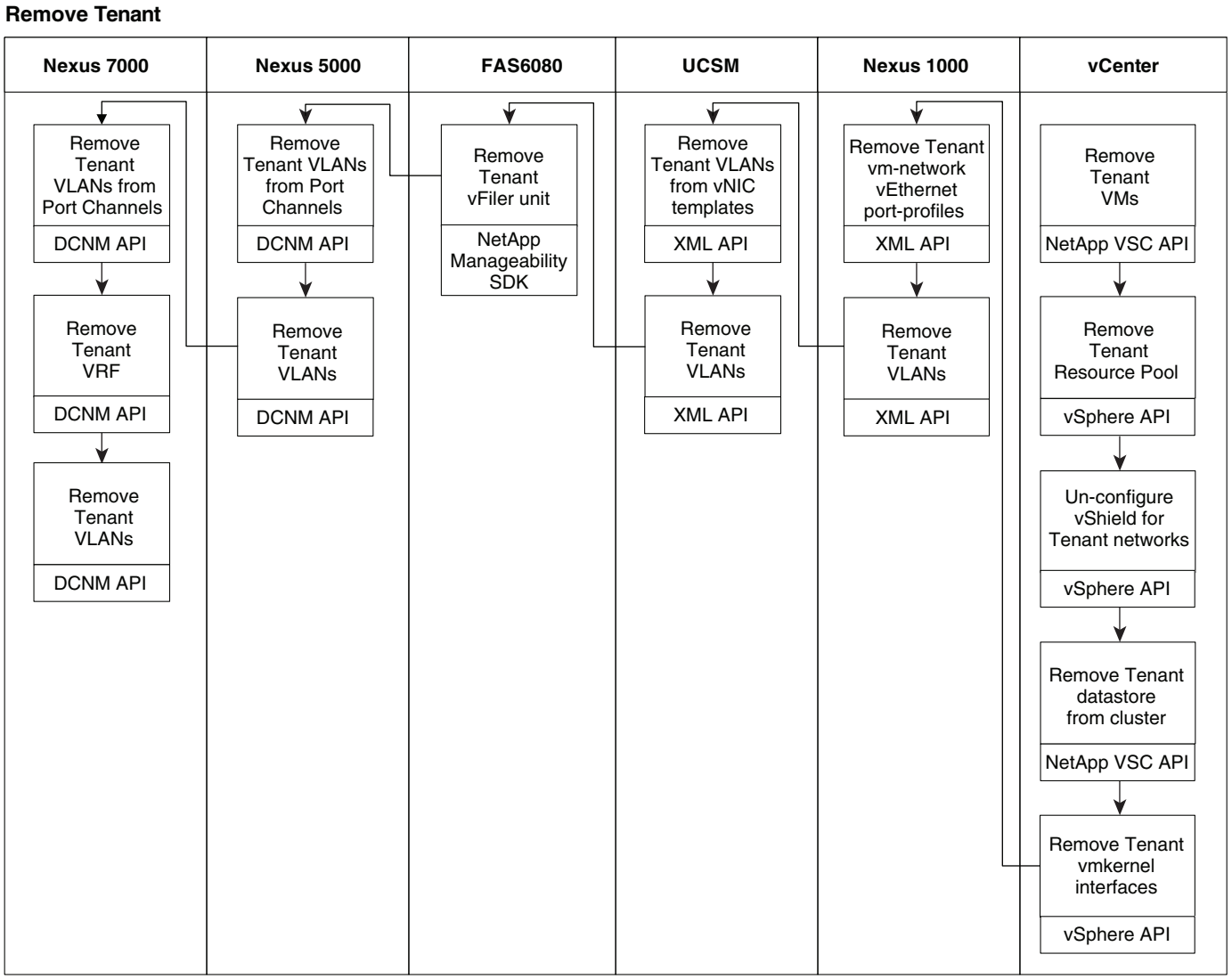
Figure 53 Create Tenant



Workflow—Remove Tenant

The workflow in Figure 54 presents a production cluster that is removing a tenant that is no longer in use.

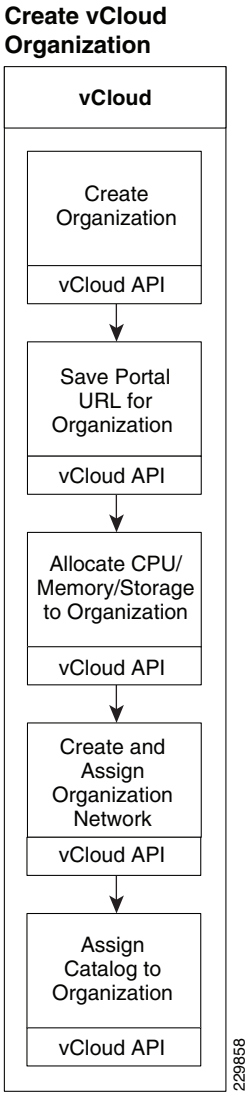
Figure 54 Remove Tenant



The workflows in Figure 55 and Figure 56 have the pre-requisite of the existing environment being configured for vCloud Director (i.e., vCenter Server(s) and vShield Manager attached to vCloud Director, with resource pools, network port groups, and datastores already configured).

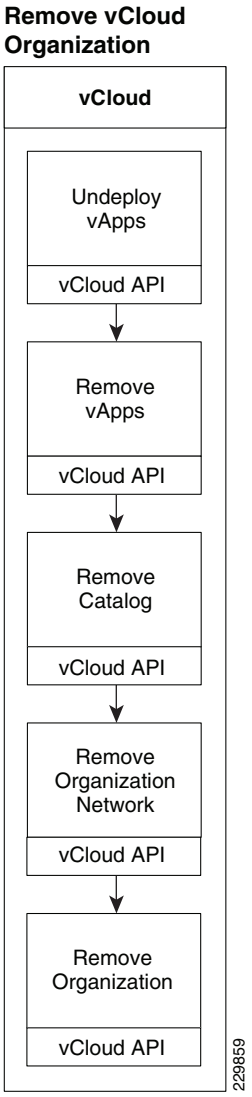
Creation of Pre-production Organization/Team/Department

Figure 55 Create vCloud Organization



Removal of a vCloud Organization

Figure 56 Removal of vCloud Organization



Appendix A—PCI Design Considerations

Regulatory requirements, such as the PCI data security standard, is often a mandatory business requirement for a vast majority of enterprise networks implementing s multi-tenant infrastructure. The PCI Solution for Retail includes a set of configured and audited architectures that incorporate technology from Cisco and Cisco partners to help retailers meet the requirements of the Payment Card Industry (PCI) Data Security Standard. The architectural components in this design and the best practices outlined in this document provides the foundation for meeting today's regulatory compliance requirements, such as PCI. The PCI requirements are tightly integrated within the architecture outlined in this document and the ultimate goal is to ensure that this Secure-Multi-Tenant architecture complies with the new PCI 1.3 requirements due at Fall of 2010.

Business Benefits

- Build a foundation for compliance - with a network infrastructure that helps enterprise address many of the PCI requirements and helps optimize security for sensitive information.
- This architectures help companies build a network that securely and reliably protects their brand images and assets while mitigating the financial risk of noncompliance fines and penalties.
- Enable secure new business initiatives - by eliminating the need to re-design the network to add capabilities.

Table 17 summarizes the twelve PCI 1.2 requirements.

Table 17 PCI Data Security Standard Requirements

Build and Maintain a Secure Network	1. Install and maintain a firewall configuation to protect data. 2. Do not use vendor-supplied defaults for system passwords and other security requirements.
Protect Cardholder Data	3. Protect stored cardholder data. 4. Encrypt transmission of cardholder data across open, public networks.
Maintain a Vulnerability Management Program	5. Use and regularly update anti-virsus software or programs. 6. Develop and maintain secure systems and applications.
Implement Strong Access Control Measures	7. Restrict access to cardholder data by business need-to-know. 8. Assign a unique ID to each person with computer access. 9. Restrict physical access to cardholder data.
Regularly Monitor and Test Networks	10. Track and monitor all access to network resources and cardholder data. 11. Regularly test security systems and processes.
Maintain an Information Security Policy	12. Maintain a policy that addresses information security.

A self assessment for the Enhanced Secure Multi-Tenancy design was performed to map the requirements to the security components within the architecture. Table 18 outlines the results of the self-assessment.

Table 18 Self-Assesment of ESMT Security Components

Requirement	Component
Install and maintain a firewall configuration to protect data.	ASA/Virtual Firewall/Vcloud
Do not use vendor-supplied defaults for system passwords and other security requirements.	SAFE/Network Foundation Protection Best Practices
Protect stored data.	NetApp Vfiers, management and data protection features
Encrypt transmission of cardholder data and sensitive information across public networks.	IPSec features and protocols across Internet and IPSec termination in DMZ
Use and regularly update anti-virsus software.	UCS Service Profiles allow for easy software upgrades/IPS
Develop and maintain secure systems and applications.	Network Foundation Protection, virtual firewall, IPS
Regularly test security systems and processes.	Management, NAM
Restrict access to data by business need-to-know.	Tenant Separation, Firewall Rules
Restrict physical access to cardholder data.	Restricting Access to Storage Appliances
Assign a unique ID to each person with computer access.	Cisco ACS and Active Directory
Maintain a policy that addresses information security.	Management, RBAC
Track and monitor all access to network resources and cardholder data.	IPS/NAM/NETFLOW

Note The security best practices provided in this document outline the architectural framework for complying with the PCI standards. The deployment guide based on this design will go through a formal process of validating this architecture against the next version of PCI standard—PCI 2.0.

About NetApp Products and Solutions

NetApp provides no representations or warranties regarding the accuracy, reliability, or serviceability of any information or recommendations provided in this publication, or with respect to any results that may be obtained by the use of the information or observance of any recommendations provided herein. The information in this document is distributed AS IS, and the use of this information or the implementation of any recommendations or techniques herein is a customer's responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. This document and the information contained herein must be used solely in connection with the NetApp products discussed in this document.

NetApp, the NetApp logo, Go further, faster, Data ONTAP, FilerView, FlexClone, FlexShare, FlexVol, MultiStore, NearStore, NetApp Data Motion, RAID-DP, SANscreen, SnapDrive, SnapManager, SnapMirror, SnapRestore, Snapshot, SnapVault, vFiler, and WAFL are trademarks or registered trademarks of NetApp, Inc. in the United States and/or other countries. All other brands or products are trademarks or registered trademarks of their respective holders and should be treated as such.

About Cisco Validated Design (CVD) Program

The CVD program consists of systems and solutions designed, tested, and documented to facilitate faster, more reliable, and more predictable customer deployments. For more information visit www.cisco.com/go/designzone.

ALL DESIGNS, SPECIFICATIONS, STATEMENTS, INFORMATION, AND RECOMMENDATIONS (COLLECTIVELY, "DESIGNS") IN THIS MANUAL ARE PRESENTED "AS IS," WITH ALL FAULTS. CISCO AND ITS SUPPLIERS DISCLAIM ALL WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OR ARISING FROM A COURSE OF DEALING, USAGE, OR TRADE PRACTICE. IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THE DESIGNS, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

THE DESIGNS ARE SUBJECT TO CHANGE WITHOUT NOTICE. USERS ARE SOLELY RESPONSIBLE FOR THEIR APPLICATION OF THE DESIGNS. THE DESIGNS DO NOT CONSTITUTE THE TECHNICAL OR OTHER PROFESSIONAL ADVICE OF CISCO, ITS SUPPLIERS OR PARTNERS. USERS SHOULD CONSULT THEIR OWN TECHNICAL ADVISORS BEFORE IMPLEMENTING THE DESIGNS. RESULTS MAY VARY DEPENDING ON FACTORS NOT TESTED BY CISCO.

CCDE, CCENT, Cisco Eos, Cisco Lumin, Cisco Nexus, Cisco StadiumVision, Cisco TelePresence, Cisco WebEx, the Cisco logo, DCE, and Welcome to the Human Network are trademarks; Changing the Way We Work, Live, Play, and Learn and Cisco Store are service marks; and Access Registrar, Aironet, AsyncOS, Bringing the Meeting To You, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, CCSP, CCVP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Collaboration Without Limitation, EtherFast, EtherSwitch, Event Center, Fast Step, Follow Me Browsing, FormShare, GigaDrive, HomeLink, Internet Quotient, IOS, iPhone, iQuick Study, IronPort, the IronPort logo, LightStream, Linksys, MediaTone, MeetingPlace, MeetingPlace Chime Sound, MGX, Networkers, Networking Academy, Network Registrar, PCNow, PIX, PowerPanels, ProConnect, ScriptShare, SenderBase, SMARTnet, Spectrum Expert, StackWise, The Fastest Way to Increase Your Internet Quotient, TransPath, WebEx, and the WebEx logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0809R)