Secure Separation Built on FlexPod Design Guide

Last Updated: October 20, 2011

Cisco Validated Design

Building Architectures to Solve Business Problems

About the Authors

Ramesh Issac

Ramesh Isaac, Technical Marketing Engineer, Systems Development Unit, Cisco

Ramesh Isaac has worked in data center and mixed-use lab settings since 1995. He started in information technology supporting UNIX environments and focused on designing and implementing multi-tenant virtualization solutions in Cisco labs over the last couple of years. Ramesh holds certifications from Cisco, VMware, and Red Hat.

Aeisha Duncan

Aeisha Duncan, Technical Marketing Engineer, Systems Development Unit (SDU), Cisco Systems

Aeisha Duncan, CCIE #13455, is a Technical Marketing Engineer for data center technologies in Cisco's Systems Development Unit. Prior to joining the SDU 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.

Alex Nadimi

Alex Nadimi, Solutions Architect, Systems Development Unit (SDU), Cisco Systems

Alex has been with Cisco for the past 15 years and is currently working as a Solutions Architect in Cisco's Systems Development Unit. 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.

Jon Benedict

Jon Benedict, Reference Architect, Infrastructure and Cloud Engineering, NetApp

Jon Benedict is a reference architect in the Infrastructure & Cloud Engineering team at NetApp. Jon is largely focused on designing, building, and evangelizing cloud and shared storage solutions based around NetApp for enterprise customers. Prior to NetApp, he spent many years as a consultant, integrator, and engineer with expertise in Unix and Linux. Jon holds many industry certifications including several from Red Hat.

About the Authors

Aleksandr Brezhnev

Aleksandr Brezhnev, Managing Principle Architect, Red Hat, Inc.

Aleksandr Brezhnev is a platform solutions architect at Red Hat. He is currently focused on partner enablement and solution development based on Red Hat virtualization and cloud products. Aleksandr is an expert in system tuning and database and application optimization on Red Hat platforms. He has been with Red Hat for more than 10 years and in prior roles he was a consulting engineer and technology development manager for strategic accounts in financial and healthcare verticals.

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 http://www.cisco.com/go/designzone.

Secure Separation Built on FlexPod Design Guide

Secure Separation Built on FlexPod Design Guide

Introduction

Goal of This Document

This document provides design guidance for implementing securely separated tenants in a non-virtualized (also known as Bare Metal) data center built on FlexPod™. Tenants may be a group of users, a group of applications, or any group which needs committed resources from the data center.

Cisco® and NetApp®, with Red Hat®, have jointly designed a best-in-breed non-virtualized FlexPod data center architecture with secure separation and have validated the design in a lab environment. The non-virtualized environment uses a native operating system (Red Hat Enterprise Linux® 6) on Cisco Unified Computing System™ (UCS™) blades, as opposed to a thin hypervisor.

The FlexPod data center solution is well suited for a variety of application workloads. Many customers are choosing to deploy some applications in a virtualized environment, while other applications remain dedicated to a server running a native operating system. FlexPod can run both virtualized and non-virtualized environments simultaneously. This design guide shows how to add non-virtualized workloads to the FlexPod data center solution and provide secure separation as needed.

This document describes the design challenges of implementing secure separation in a non-virtualized FlexPod environment and shares best practices for implementation. There are important design questions that must be answered prior to deployment since each situation is unique. This document guides the reader through the key considerations necessary for a successful deployment of secure separation in a non-virtualized data center built on FlexPod.

The secure separation validation test bed was configured as a simulated e-commerce environment. This e-commerce environment used non-virtualized systems hosted from common network and storage infrastructure that was securely separated within tenants and sub-segments within tenants. The tenants were broken up into two groups, simulated clients and the servers used to simulate the e-commerce workflow.

The Secure Separation built on FlexPod system design is built on:

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

- FlexPod Deployment Guide: http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/Virtualization/flexpod_deploy.html
- Red Hat Enterprise Linux built on FlexPod Deployment Guide: http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/Virtualization/flexpod_rhel.html
- The Enhanced Secure Multi-tenant Design Guide includes relevant design principles for virtualized environments on FlexPod: http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/Virtualization/securecldg_V2.html
- The Cisco Virtual Multi-Tenant Data Center Design Guide provides general design guidance for Data Center deployments: http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/VMDC/2.1/design_guide/vmdc21DesignGuide.html

Audience

The intended audience for this document includes, but is not limited to, sales engineers, field consultants, professional services personnel, IT managers, partner engineering personnel, and customers who want to deploy a non-virtualized FlexPod data center with secure separation.

Objectives

This document articulates the design principles and implementation best practices required to design and deploy secure separation built on FlexPod in a non-virtualized environment. The design principles and best practices are derived from lab validation activities.

Architecture Overview

Data center resources are shared by applications and users. In both virtualized and non-virtualized environments, the administrator's goal is to use data center resources as efficiently as possible, while providing the levels of service demanded by applications and users. The applications and users within the data center need to coexist and interact or remain isolated from each other, as dictated by organizational requirements.

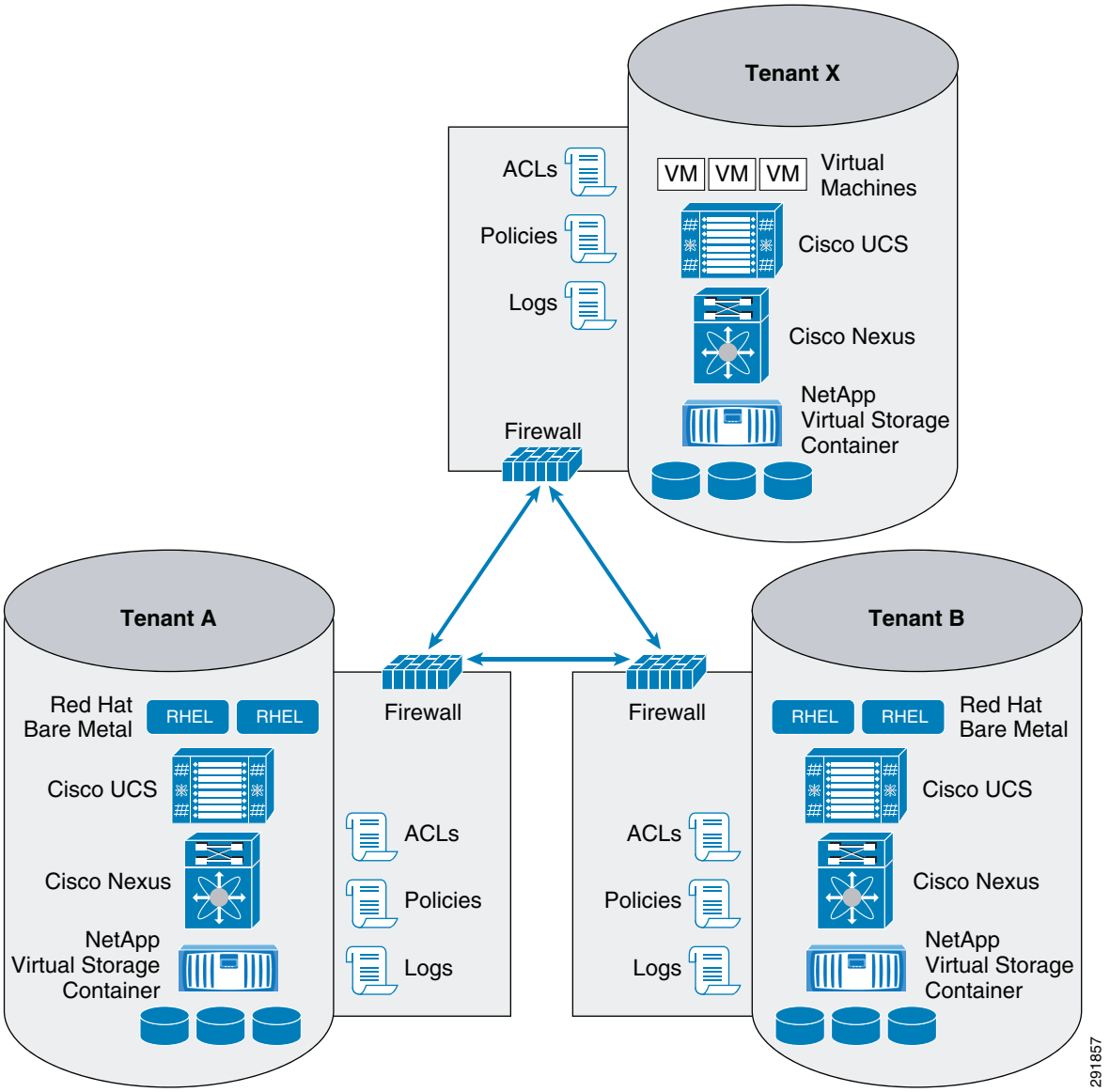
Multi-tenant data center design leverages common resources to cost effectively provide the tenant with what appears to be a dedicated environment. Shared resources must ensure one tenant cannot impede the access of another tenant and must be built in a manner that allows continued functionality in the event of component failure. To achieve these goals, Cisco and NetApp have worked with Red Hat to implement securely separated non-virtualized tenants in a data center built on FlexPod.

The following design concepts need to be addressed in any data center to achieve these goals:

- Security
- Service assurance
- Availability
- Management

These design concepts apply to both virtualized and non-virtualized environments and are key elements of this securely separated non-virtualized data center design, as well as the enhanced secure multi-tenant architecture (virtualized environment) referenced earlier. Whether the requirements call for 1 or N tenants, implementing the data center by addressing these design concepts ensures a secure, manageable, available, and scalable design which will meet future organizational needs.

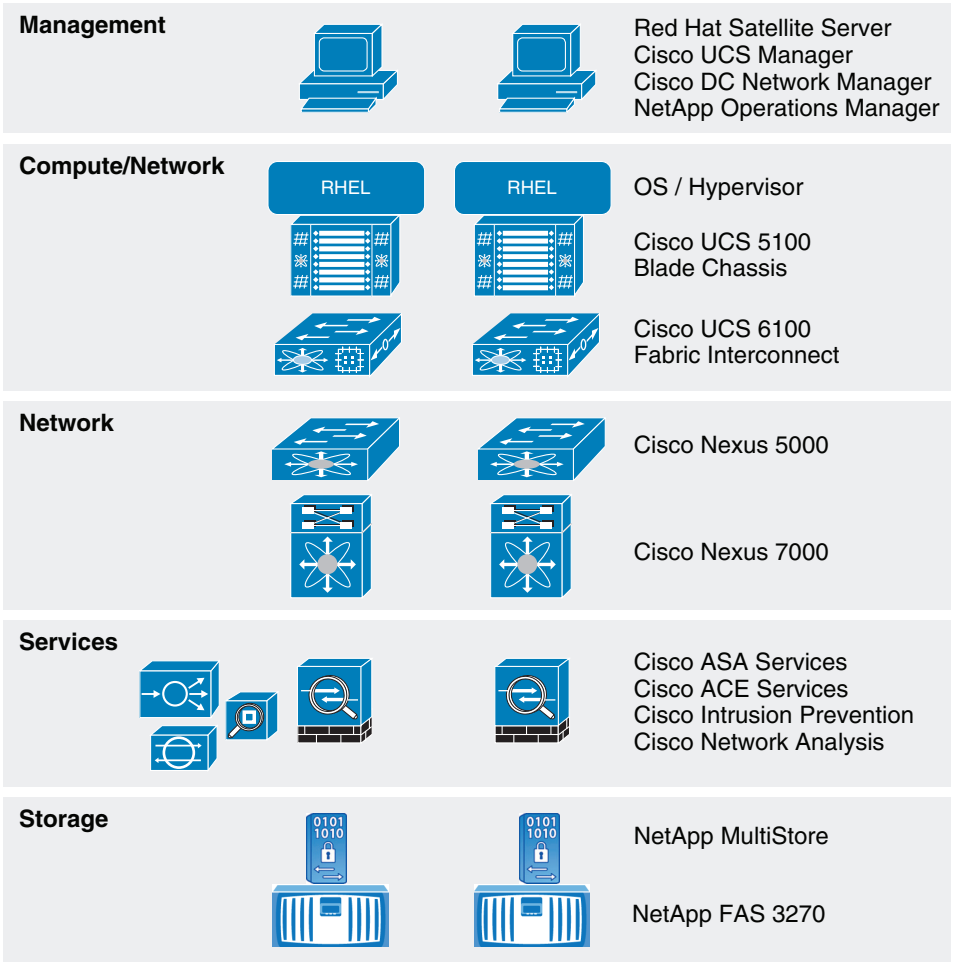
Figure 1 Architecture Overview



Organizational needs dictate the requirements for each tenant and between tenants and may be driven by business, compliance, or application requirements. The design is built on foundational components which create the environment for providing securely separated services. The specific requirements for each tenant determine how the foundational components are implemented to deliver the expected services.

Architecture Components

Figure 2 Architecture Components



Compute

Cisco Unified Computing System

The Cisco UCS 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.

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

- Cisco UCS 6100 Series Fabric Interconnects (http://www.cisco.com/en/US/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/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/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/products/ps10280/index.html) adapt to application demands, intelligently scale energy use, and offer best-in-class computing.
- Cisco UCS B-Series Network Adapters (http://www.cisco.com/en/US/products/ps10277/prod_module_series_home.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/products/ps10281/index.html) provides centralized management capabilities for the Cisco Unified Computing System.

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

Cisco UCS Manager

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

Red Hat Enterprise Linux

Red Hat Enterprise Linux (http://www.redhat.com/promo/Red_Hat_Enterprise_Linux6/) supports today's flexible, varied enterprise architectures. Red Hat Enterprise Linux 6 provides a rock-solid foundation for every deployment. By incorporating software technologies developed by Red Hat, its partners, and the open source community, the latest release boasts an extensive list of features—some new, some improved—including:

- Scalability and performance on the latest hardware and hypervisors
- Efficiency and resource management designed to minimize and manage your datacenter power requirements
- Consistency across physical, virtual, and cloud to make your applications flexible

Available for almost a decade, the world-class Red Hat Enterprise Linux platform has a reputation as a reliable, high-performance, high-value alternative to proprietary UNIX systems.

Red Hat Network Satellite

Red Hat Network (RHN) Satellite is a systems management platform that makes Linux deployable, scalable, manageable, and consistent. RHN Satellite provides administrators with the tools to efficiently manage their systems, lowering per-system deployment and management costs. RHN Satellite offers superior security by having a single centralized tool, secure connection policies for remote administration, and secure content. Use RHN Satellite to ensure security fixes and configuration files are applied across your environment consistently.

- One click software updates in an easy to use interface
- Role based administration
- Flexible delivery architectures—Satellite, Proxy, hosted
- Group systems together for easier administration
- Automate formerly manual tasks
- Manage the complete life cycle of your Linux infrastructure
- Track the performance of your Linux systems

For more information, see: http://www.redhat.com/red_hat_network/.

Network

Cisco Nexus 7000

The Cisco Nexus® 7000 Series is a modular switching system designed to deliver 10 Gigabit Ethernet and unified fabric in the data center. This 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 Series Supervisor modules are based on a dual core processor that scales the control plane beyond 15 terabits per second (Tbps). The supervisors control the Layer 2 and Layer 3 services, redundancy capabilities, configuration management, status monitoring, and power and environmental management.

Virtual Device Contexts (VDCs) on the Cisco Nexus 7000 enable complete separation of the control and data plane functionality, creating two or more logical devices with no feature interaction between them. Virtual Routing and Forwarding (VRF) instances can be used to virtualize Layer 3 forwarding and routing tables.

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, which 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 multi-socket, multi-core, 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, non-blocking 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 Adaptive Security Appliance

The Cisco Adaptive Security Appliance (ASA) provides advanced stateful firewall and VPN concentrator functionality in one device, and for some models, an integrated intrusion prevention system (IPS) module or an integrated content security and control (CSC) module. The ASA includes many advanced features, such as multiple security contexts (similar to virtualized firewalls), transparent (Layer 2) firewall or routed (Layer 3) firewall operation, advanced inspection engines, IPsec VPN, SSL VPN, and clientless SSL VPN support, and many more features.

For more information, see: http://www.cisco.com/en/US/products/ps6120/index.html.

Cisco Application Control Engine

The Cisco Application Control Engine (ACE) module and appliance 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

The Cisco Intrusion Prevention System (IPS) 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/docs/security/asa/asa84/asdm64/configuration_guide/modules_ips.html.

Cisco Network Analysis Module 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 Data Center Network Manager

Cisco Data Center Network Manager (DCNM) is a management system for the Cisco Unified Fabric. It enables you to provision, monitor, and troubleshoot the data center network infrastructure. It provides visibility and control of the unified data center so that you can optimize for the quality of service (QoS) required to meet service-level agreements.

Cisco DCNM increases overall data center infrastructure uptime and reliability, thereby improving business continuity. It provides a robust framework and comprehensive feature set that meets the routing, switching, and storage administration needs of data centers. Cisco DCNM streamlines the provisioning for the unified fabric and monitors the SAN and LAN components. Cisco DCNM provides a high level of visibility and control through a single web-based management console for Cisco Nexus, Cisco MDS, and Cisco Unified Computing System products.

For more information, see: http://www.cisco.com/en/US/docs/switches/datacenter/mds9000/sw/5_2/configuration/guides/fund/DCNM-SAN-LAN_5_2/DCNM_Intro.html.

Cisco Application Network Manager

The Cisco Application Network Manager (ANM) is a client-server application allowing administrators to provision, monitor, and maintain application network services in the data center. Employing role-based access control (RBAC), the Cisco ANM allows application owners or server administrators to create ACE-enforced application policies within the network without impacting network configurations.

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 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.

For more information, see:

http://www.cisco.com/en/US/products/ps6498/index.html.

Storage

NetApp FAS Unified Storage

Each NetApp® fabric-attached storage (FAS) controller shares a unified storage architecture based on the Data ONTAP® 8 (7-Mode) 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.

Data ONTAP 8 is tightly integrated into the hardware systems to provide resilient system operation and high data availability. The FAS systems use redundant, hot-swappable components and offers patented double-parity RAID-DP®. For a higher level of data availability, Data ONTAP provides optional mirroring, backup, and disaster recovery solutions. For more information, see: 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. Snapshot technology creates up to 255 data-in-place, point-in-time images per volume. For more information, see: http://www.netapp.com/us/products/platform-os/snapshot.html.

FlexShare® quality-of-service software is included as part of the Data ONTAP operating system to enable fast response when serving data for multiple applications. FlexShare allows storage administrators to set and dynamically adjust workload priorities. For more information, see: http://www.netapp.com/us/products/platform-os/flexshare.html.

The quantity, size, and type of disks used vary depending on storage and performance needs. Add-on cards, such as the Flash Cache (PAM II) modules, can be used in this architecture to increase performance by increasing system cache for fast data access. For more information, see: http://www.netapp.com/us/products.

Ethernet Storage

Ethernet storage using NFS 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

NetApp 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.

Stateless Computing Using Fibre Channel Boot over Ethernet

Fibre Channel over Ethernet is an encapsulation of Fibre Channel frames transported over Ethernet networks with provisions to ensure the Fibre Channel standards are followed. FCoE architectures reduce management complexity, required cable count, number of mezzanine cards for fabric segmentation, and power and cooling costs. FCoE is a design element in the FlexPod architecture, which is a core capability of the UCS (implemented in virtualized server adapters) and the NetApp storage controller (implemented in unified storage adapters). Fibre Channel and IP traffic from the UCS fabric interconnects travels through the Cisco Nexus 5000 switch to the NetApp storage IP and FCoE interfaces.

Leveraging UCS service profiles, deployed by using FCoE-booted physical resources, provides flexibility and resiliency to a multi-tenant infrastructure.

FCoE-booted hosts using NetApp controllers have superior RAID protection and increased performance compared to traditional local disk arrays and FCoE-booted resources can easily be recovered, are better utilized, and scale much more quickly than local disk installs. Another major benefit is that they can be deployed and recovered in minutes. FCoE-booted deployments effectively reduce storage provisioning time, increase storage utilization, and aid in the stateless nature of service profiles within UCS. Furthermore, UCS service profiles ensure hosts are auto-assigned to specific VLANs and VSANs and leverage preconfigured pools of MAC addresses and WWPNs, which could be incorporated into access control lists (ACLs) or port security settings.

NetApp MultiStore and IP Spaces

Storage controllers consist of pools of resources and the methodology to access data—whether it is CIFS, NFS, iSCSI, or FCP. NetApp MultiStore® allows users 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 serving Ethernet-based storage. Each vFiler unit can be individually managed with different sets of performance and policy characteristics.

When considering how an application or user accesses the data contained within a vFiler unit, it is important to understand how virtual interfaces (VIFs), VLAN interfaces, and the IP space feature work together to effect secure separation of the various tenants housed on a physical controller. Virtual

interfaces are configured on the storage system and allow for network redundancy and increased bandwidth of the physical network ports. VLAN interfaces, which are layered on a given VIF, enable VLAN tagging across the bonded interface. VLAN interfaces are assigned to an IP space, which is a distinct IP address space with private routing tables. IP spaces are created when vFiler units need to have their own secure storage, administration, and routing leading to secure multi-tenancy. When vFiler units are used in conjunction with IP spaces, multiple tenants share the same storage resources while maintaining privacy and security.

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

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 off-site to protect against data loss and corruption should a failure or catastrophic event occur 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 providing for business continuity.

NetApp DataMotion

NetApp DataMotion™ 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, DataMotion integrates three proven NetApp software technologies—MultiStore, SnapMirror, and Provisioning Manager—to provide live data migration for shared storage infrastructure.

NetApp Provisioning Manager

NetApp Provisioning Manager streamlines the deployment of tenant storage resources according to established policies. Provisioning Manager enables the administrator to:

- Automate deployment of storage supporting the 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 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 used to develop policies that increase resource utilization and operational efficiency, and they 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. 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.

For more information, see:
http://www.netapp.com/us/products/management-software/provisioning.html.

NetApp Protection Manager

Using NetApp Protection Manager, 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 designed to suit operational and service-level requirements across the storage infrastructure and within tenant environments. Protection Manager automatically correlates logical datasets and the underlying physical storage resources. This enables administrators to design and apply policies according to business-level or service-level requirements. 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 granting control to tenant administrators. The integration of Provisioning Manager and Protection Manager within a single console allows tenant administrators to seamlessly provision and protect data through unified, policy-driven workflows.

For more information, see: http://www.netapp.com/us/products/management-software/protection.html.

NetApp Operations Manager

NetApp Operations Manager delivers centralized management, monitoring, and reporting tools. These tools enable consolidated, streamlined management of NetApp storage infrastructure. Operations Manager can reduce costs by leveraging comprehensive dashboard views to optimize storage utilization and minimize the IT resources needed to manage shared storage infrastructure. At the same time, administrators can improve the availability and quality of services delivered to their tenant customers. 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 storage infrastructure. Operations Manager gives comprehensive visibility into the storage infrastructure by continuously monitoring storage resources, analyzing utilization and capacity management, and providing insight into the growth trends and resource impact of the tenants.

For more information, see:
http://www.netapp.com/us/products/management-software/operations-manager.html.

Storage Management

NetApp provides comprehensive storage management solutions that enable customers to achieve dramatically improved efficiency, utilization, and availability. NetApp offers a holistic approach focused on simplifying data management that effectively addresses the operational challenges of complex environments.

FilerView® and the Data ONTAP® command line user interfaces are instrumental in the initial build out of the architecture. During subsequent routine service operations, the use of the CLI is discouraged in favor of the policy driven facilities of the NetApp management software portfolio. The following subsections introduce data management structures and concepts to consider when designing a storage management solution for secure separation built on FlexPod.

Resource Groups

Effective data management begins with grouping storage objects according to desired relationships. NetApp Operations Manager allows the flexible creation of resource groups that can be associated with 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 be used to organize each tenant's resources with alerting, reporting, and role-based access control. Tenant organizations and suborganizations can be created to apply access controls or to send reports and alerts to the appropriate administrators.

Resource Pools

Resource pools describe the physical organization of storage systems and aggregates to be sourced for provisioning operations. Resource pools are typically used to group storage according to common attributes such as size, cost, performance, and availability: for example, separating primary from secondary data across varying performance configurations or grouping homogeneous storage types. Within a securely separated data center, resource pools can be used to design and prescribe how storage is sourced from the shared storage infrastructure to accommodate various tenant service requirements.

Datasets

A dataset is a collection of data objects (volumes, qtrees, LUNs) that are managed as a single unit, following the same provisioning and protection requirements. A dataset includes primary data objects as well as all replicas of those objects derived from their data protection configuration. For example, a dataset 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 datasets 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. Protection policies can fully automate the provisioning and configuration of the desired data protection design and processes for primary, secondary, tertiary, and further data replication. Provisioning and protection policies prescribe how storage objects are allocated from underlying resource pools into datasets 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 might apply to multiple vFiler units. The administrator could design a vFiler unit template for each tenant organization or suborganization to define default configuration values to standardize all vFiler units deployed on behalf of a tenant.

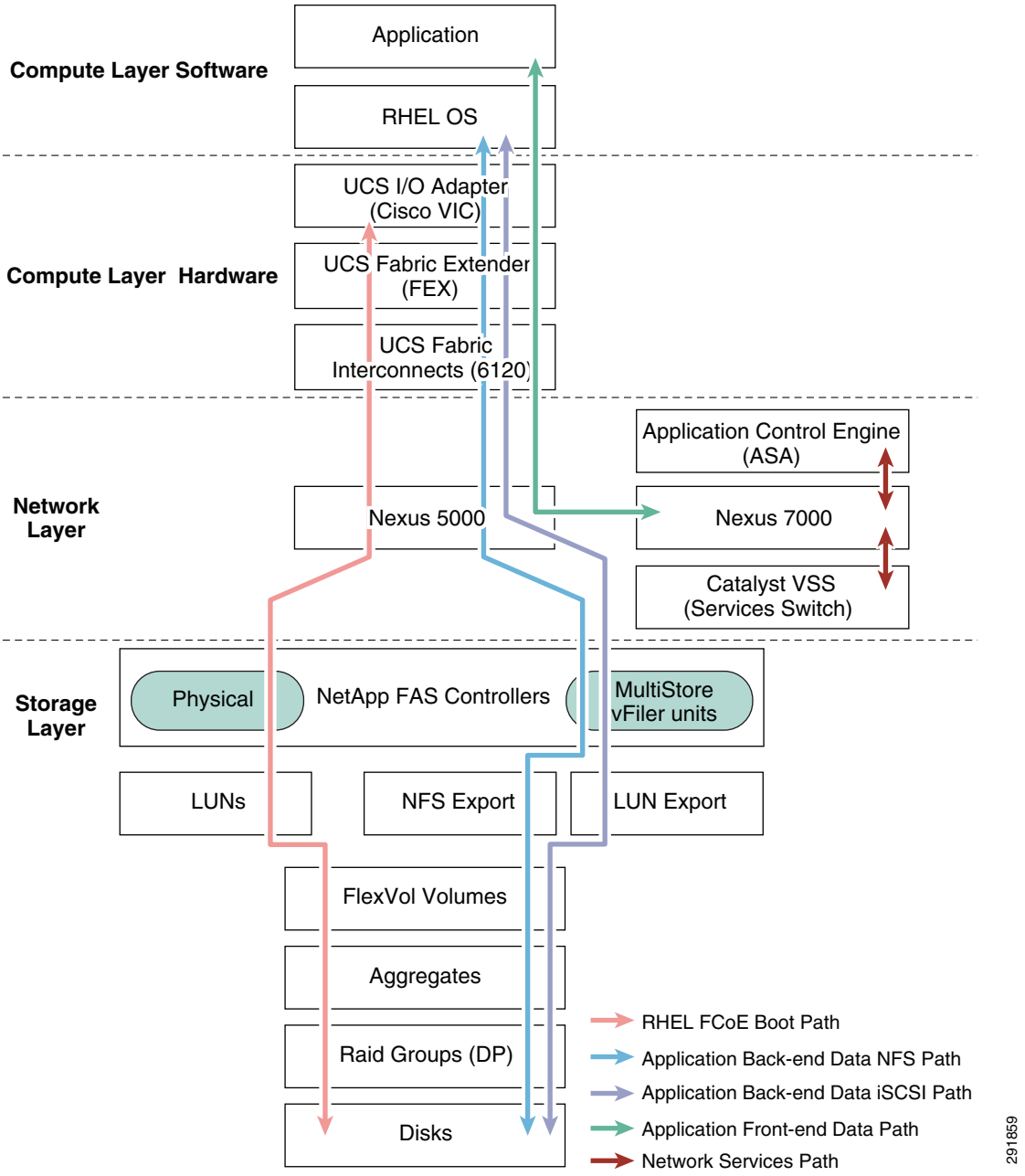
Storage Services

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. When the desired storage service is selected from this catalog, new datasets can automatically be created to comply with desired provisioning and protection policies in a single workflow. Storage services can be designed to deliver tenant data objects that comply with baseline storage offerings defined by the storage administrator. Storage services can also be further tailored to meet the specific needs of a particular tenant, organization, or application.

End-to-End Block Diagram

Figure 3 shows the flow from application to storage in the secure separation test bed. This end-to-end diagram shows the non-virtualized Red Hat Enterprise Linux host FCoE boot, through the network layer, finishing up at the storage foundation.

Figure 3 End-to-End Block Diagram



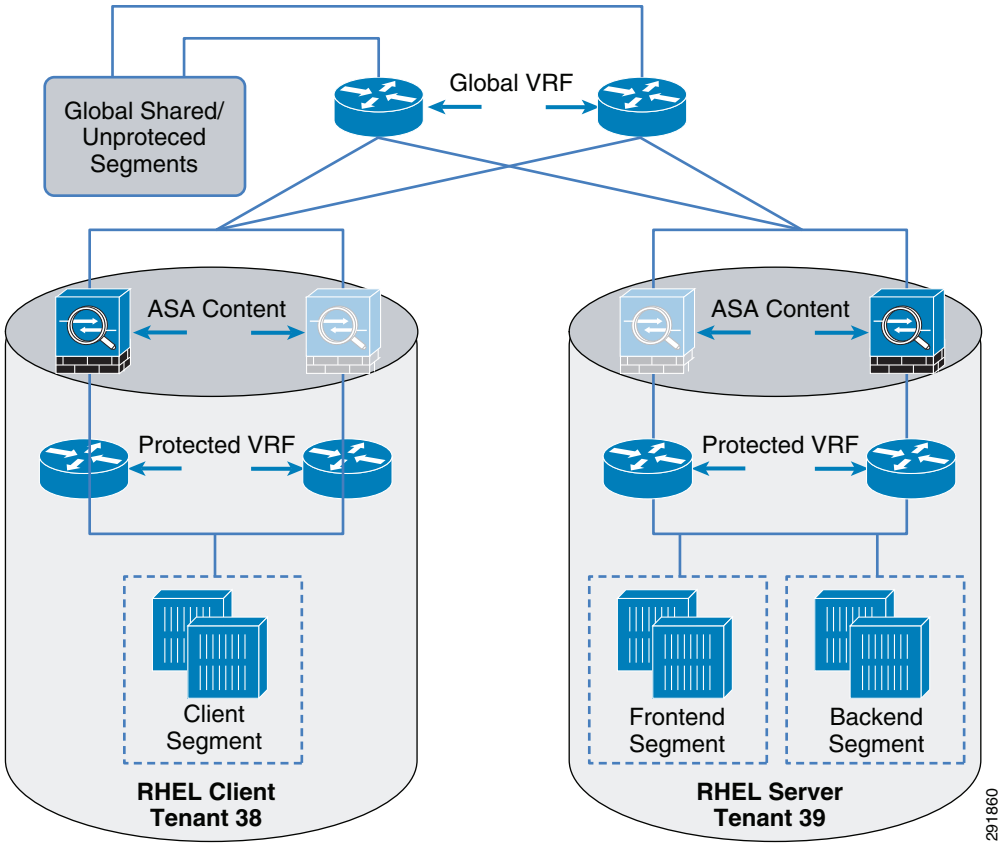
System Design

This section provides design best practices for implementing secure separation in a non-virtualized data center built on FlexPod.

Tenant Model

The secure separation test bed utilizes a flexible multi-tenant architecture. The tenant model describes the logical configuration of compute and network resources on a standardized infrastructure. The tenant containers exist as resources connected to a protected VRF behind a firewall context.

Figure 4 Tenant Model



The tenant model leverages the following features and services:

- Aggregate global VRF for Layer 3 services between core, tenant, and globally shared services
- An unprotected segment at the global or organizational level may support services which do not require or support firewall-based security services.
- Dedicated virtual firewall context to enforce security policy on tenant ingress and egress traffic flows
- Dedicated protected VRF for Layer 3 tenant-specific services (typically default gateway for server farm)
- Layer 2 segmentation via VLANs (grouped into segments)
- Protected segments employ a firewall virtual context.

Administrator Roles

Administrator roles are implemented with role-based access control, which is configurable within the management applications in each infrastructure layer of a multi-tenant environment built on FlexPod. Consistency of RBAC should be maintained through Lightweight Directory Access Protocol (LDAP) or Active Directory (AD) to consolidate common tenant resources. These common tenant resources can be managed by:

- Tenant Administrator—A role defined to manage resources specific to a particular tenant.
- Primary Administrator—A role created to have management visibility across the entire infrastructure. This role is responsible for creating tenants, managing tenant boundaries, and monitoring the service assurance/prioritization between the tenants.

Secure Separation

Secure separation ensures one tenant does not have access to another tenant's resources, such as operating system, network bandwidth, and storage. Each tenant must be securely separated from every other tenant and from other threats 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 may be 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 often effective because these protocols are usually allowed to flow through the enterprise network and enter the intranet data center. Threat vectors which affect the multi-tenant data center include:

- 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 and URL), cross-site scripting attacks, SQL attacks, malformed packets, viruses, worms, and man-in-the-middle attacks. 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 1 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 1 Data Center Threats and Mitigation Resources

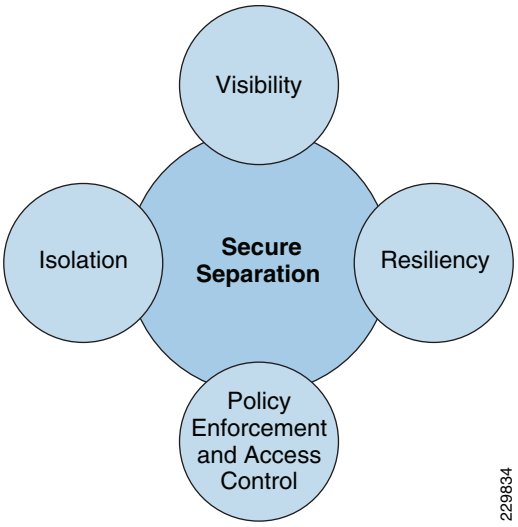
Network Services	Botnets	DOS	Spyware/Malware	Data Leakage	Visibility	Network Abuse	Control
Policy Enforcement					Yes	Yes	Yes
Application Control Engine (ACE)		Yes		Yes	Yes		Yes
IPS integration	Yes		Yes		Yes	Yes	Yes
Switching Security		Yes		Yes		Yes	
Secure Device Access					Yes	Yes	Yes
Telemetry	Yes	Yes			Yes	Yes	
Firewall Services	Yes	Yes			Yes	Yes	Yes

Introduction to Security 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 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 flexible in the way they scale to accommodate new services and new servers (both virtualized and non-virtualized). 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 to achieve the desired level of resiliency. Figure 5 shows the security architecture framework implemented in this design.

Figure 5 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).

Traffic Flow within the Data Center

The characterization of traffic patterns is critical as one begins to implement multiple tenants and multi-tier applications within a data center. Understanding these flows is instrumental to the development of a comprehensive set of security and application policies. One can categorize traffic flows into two distinct categories: east-west and north-south.

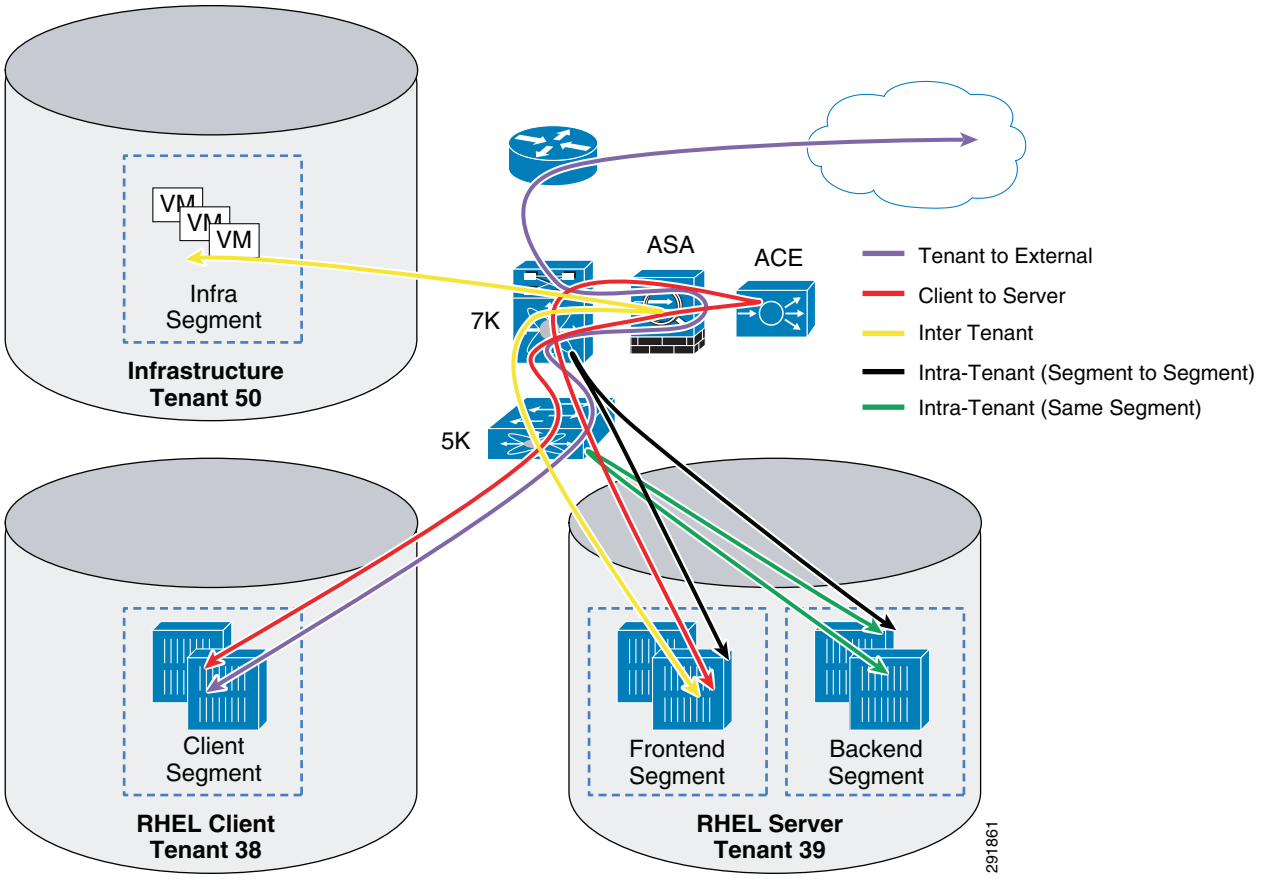
North-South Traffic

Flows are either ingress or egress in relation to the infrastructure and are commonly client-to-server in nature. This traffic traverses the data center and is readily exposed to a number of services in its path, including firewalling, load balancing, intrusion detection, and network analysis devices. This ensures each tenants policy is uniformly applied between every tenant. The security services applied to ingress-egress traffic flows in the data center depends upon application specific requirements and the overall security policies of the enterprise.

East-West Traffic

This traffic flow refers to the communication between servers within the data center. 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 optimize east-west traffic patterns within a multi-tenant data center using non-virtualized servers, it is recommended to leverage security services in the access layer switch and host-based security capabilities within the host operating system.

Figure 6 Traffic Flows



Typical deployments include additional traffic flows, based on business requirements. Figure 6 shows the additional traffic flows within this infrastructure which were considered in the lab validation design:

- Tenant-to-External
- Client-to-Server
- Inter-Tenant
- Intra-Tenant Single Segment
- Intra-Tenant Segment-to-Segment

Tenant-to-External

Tenant-to-external traffic flows are necessary for Web access, software updates, and OS patching. This is a north-south flow that is internally initiated. Highly restrictive policies on these flows are important as these connections to external sources are the primary culprit for most network attacks.

Client-to-Server

Client-to-server traffic flows are typically client-to-server conversations and therefore one end of the conversation is usually outside the data center. These north-south flows traverse the data center and may be exposed to additional network services, such as intrusion sensors and load balancers. The policy applied to these flows is typically dependent upon the nature of the application and security policy of the enterprise.

Inter-Tenant

In the multi-tenant environment, east-west traffic flows occur between servers in each tenant container and as such are subject to the specific security and application policies of each tenant. This allows each tenant entity to apply service policies addressing their specific security and application needs.

Intra-Tenant Single Segment

Traffic internal to a tenant segment is contained within the access layer in the case of non-virtualized servers.

Typically, this type of traffic occurs between server roles within an enterprise n-tier application. In this case—where the servers reside in the same subnet—the traffic is switched at the access layer and not through the routing layer at the core. Security policies for such traffic are implemented in the host (host-based security features include the iptables, ip6tables, ebtables, and arptables network filters) and the access layer switch. Server to server traffic does not traverse the appliances in the services layer.

Intra-Tenant Segment-to-Segment

Intra-tenant segment-to-segment traffic requires routing services between the disparate segments within a single tenant container. Traffic flows traverse the virtual edge access layer as well as the access layer to consume Layer 3 services at the tenant VRF. The application and security policies of one tenant are enforced by the unified services offered by the container and defined by the tenant. In this case server to server traffic traverses the appliances in the services layer.

Security Components

Securely separating tenants in a data center built on FlexPod requires components in addition to the base FlexPod (firewall, load balancer, intrusion prevention system, and network analysis module). Red Hat Enterprise Linux provides features which enhance the system security capabilities. Securely separated tenants in a data center built on FlexPod may be implemented with other operating systems or hypervisors.

Cisco Firewall Services

Firewall services provide stateful firewall security capabilities within the architecture. The virtual Cisco firewall security contexts 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 firewall context are consistently applied to protect the related networks. The firewall can also provide excellent visibility into the network. One can use the Cisco Adaptive Security Device Manager’s real time log viewer and monitoring dashboards and packet trace capabilities to gain visibility into the traffic flows and detect and mitigate malicious or unauthorized traffic.

- Cisco Application Control Engine
- Cisco Intrusion Prevention System
- Cisco Network Analysis Module Products
- Cisco Security Manager
- NetApp MultiStore and IP Spaces

Linux Netfilter and IPTables

The Linux kernel features a powerful networking subsystem called Netfilter. The Netfilter subsystem provides stateful or stateless packet filtering as well as NAT and IP masquerading services. Netfilter also has the ability to mangle IP header information for advanced routing and connection state management. Netfilter is controlled using the iptables tool.

The iptables tool uses the Netfilter subsystem to enhance network connection, inspection, and processing. The iptables tool provides command line interface to configure advanced logging, pre- and post-routing actions, network address translation, and port forwarding. It can be used to enable access to the trusted network services and disable everything else.

Security-Enhanced Linux

Security-Enhanced Linux (SELinux) is an implementation of a mandatory access control mechanism in the Linux kernel, checking for allowed operations after standard discretionary access controls are checked. SELinux can enforce rules on files and processes in a Linux system, and on their actions, based on defined policy.

Using SELinux reduces the system's vulnerability to privilege escalation attacks. Because SELinux policy rules define how processes access files and other processes, if a process is compromised, the attacker only has access to the normal functions of that process and to files to which the process has been configured to have access. For example, if the Apache HTTP Server is compromised, an attacker can not use that process to read files in user home directories, unless a specific SELinux policy rule was added or configured to allow such access.

Red Hat Enterprise Linux enables SELinux by default and provides policies protecting network services.

Mapping Security Principals to Features

Secure separation is one of the key design concepts within a multi-tenant environment. Together, the network, compute, storage, and management components within this architecture provide features and capabilities which form the security framework and ensure secure separation within tenants. Table 2 provides the 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 2 Mapping of Security Principals to Features

Principal	Network	Compute	Storage
Isolation	ASA/FWSM firewall services Virtual Firewall ACE virtual contexts	SELinux IPtables	NetApp Data ONTAP, MultiStore, IP spaces and VLAN interfaces
Visibility	Netflow, ERSPAN Syslog ASA/FWSM/IPS/ACE Event Notifications ASA/FWSM/IPS/ACE Telemetry Data Export	Audit logs Security logs	NetApp Data ONTAP audit logs

Table 2 Mapping of Security Principals to Features

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

Network Security Design Considerations

A defense-in-depth design approach is essential to implement a resilient end-to-end architecture. This approach implements security features at every layer of the network and compute infrastructure. The security considerations differ considerably for different traffic flows as outlined below.

Client-to-Server Security Considerations

The client-server traffic flow passes through the Cisco ASA firewall appliance, the inline IPS, and the Cisco ACE load balancer.

The firewall is used:

- To restrict traffic to servers from specific branches within the enterprise.

  In cases where the servers are only to be accessed by specific remote users, firewalls can be used to protect the servers from unauthorized users from within or outside the enterprise.

- To restrict incoming traffic flow to specific ports

  One can use firewalls to lockdown traffic to specific ports to reduce possibility of malicious traffic entering the data center from unauthorized users

- To segment different tenants into separate security zones

  One can use the firewall to restrict client traffic to separate firewall contexts, where each context can be mapped to a particular tenant.

- For e-commerce applications, by placing the servers in a separate network segment and using the firewall to separate e-commerce servers from the rest of the enterprise network.

IPS is used to mitigate attacks from clients as follows:

- IPS multi-context capability can be leveraged to implement tenant-based policies depending on the tenants requirements. Different tenants can be mapped to different contexts and different policies can be applied to those contexts.
- Signatures within each context can be tuned to correspond to the tenant's requirements.
- E-commerce applications need a highly secure environment. One can force all e-commerce client traffic to pass through the IPS and ACE appliances. IPS provides attack detection and mitigation capabilities and ACE provides load-balancing and compute resiliency services.

Inter-Tenant Security Considerations

As shown in Figure 6, all inter-tenant traffic passes through the ASA firewall. Each tenant can be mapped to a separate security zone. As outlined above, security zones for each tenant can be implemented using the firewall appliance at the aggregation layer. The following summarizes some of the considerations for designing inter-tenant security policies:

- A shared infrastructure can be implemented as a separate tenant. In this case, the ASA firewall can implement rules to allow access from all tenants to the shared infrastructure, but at the same time locking down traffic flows to only allow access to the specific shared resources and nothing else.
- In some implementations, a tenant can be mainly devoted to certain applications, such as Microsoft Exchange or SharePoint. In this case, one can fine tune security policies to implement application-based access control to those resources. ACE load balancer can be used to load balance traffic across multiple servers that host these applications. The IPS functionality can be used to mitigate any application/host-based attacks to those servers.
- As an extension of above use cases, one may implement different tenant containers within the same organization. Also some organizations may need to further subdivide their tenant structures into separate logical and functional components. The ASA firewall can be used to implemented security zones and access-control policies to this segmented network infrastructure.

Intra-Tenant Single Segment

In many n-tier applications, different components of an application may reside on different physical servers. Many application user guides recommend the separation of front-end Seb portals and back-end database servers. In many instances these components may reside in the same VLAN. One may use Access Control Lists (ACL) to control access to these application components. The intra-VLAN traffic is limited to within the Cisco Nexus 5000 switch and does not flow through the Cisco Nexus 7000.

- The ACL feature set within the Cisco Nexus 5000 can be leveraged to implement access control policies based on IP address or MAC address
- Port-Control feature at the UCS can be used to disallow spoofed MAC addresses
- Host-based security features can also be implemented:
  - iptables
  - ip6tables
  - ebtables
  - arptables

Intra-Tenant Segment-to-Segment

Some organizations may require the ability to further subdivide their tenant structures into separate logical and functional components. For example, corporate users can be subdivided into office users and lab users. These users may reside in the same tenant and require different policies. Placing office users and lab users in different VLAN segments forces inter-segment traffic to flow through the Cisco Nexus 7000, the firewall, and the IPS. One can use the firewall and IPS to enforce unique security policies for each of the sub-tenants.

E-Commerce Use Case

In e-commerce application environments, it is customary to place servers within the data center infrastructure, while the front end customer-facing portal servers are normally placed in the internet edge DMZ zone. The customer-facing portal servers communicate with internal servers in the data center. Securely separating these data center servers is a critical element of the overall security strategy. This secure separation can be achieved as follows:

- Place all the e-commerce data center servers in a separate VRF container.
- Define firewall rules to allow access to the e-commerce data center servers only from authorized front-end portal servers that normally reside in the internet DMZ.
- Use ACE load balancer to achieve high availability and resiliency of the e-commerce servers.
- Implement IPS in the inline mode to promptly mitigate attacks, worms, and other harmful traffic. The Global Correlation feature of IPS can provide up-to-date threat information that can be used to mitigate the latest viruses and attacks.

Storage Layer Resiliency

The NetApp FAS controller in a securely separated environment focuses on restricting access to systems and data. Achieving this important goal requires following proper security best practices, including defining a set of policies and procedures, and significant planning. This document does not cover detailed security design, but does illustrate the pertinent security guidelines to harden the NetApp FAS controller. Following these guidelines, in addition to the use of MultiStore, provides secure separation and confinement for the storage layer.

Tenant users should not have direct access to the NetApp FAS controller. Direct login access should only be provided to the primary administrator and tenant administrators. Even then, the primary administrator is the only role with complete access to the NetApp FAS controller. Tenant administrators only need access to their own vFiler units. Primary and tenant administrators should be granted access based on the group policy as opposed to using a shared login. This enables tracking the specific actions of specific users. Do not create a single “administrator” account to which multiple people have access.

By default, storage administrators can login from any network and any host. This must be secured as soon as the administrator accounts are created. Plan a limited number of management networks and administrative hosts. Allow administrative access only from those networks and hosts, for example, one per vFiler unit and one for the physical NetApp FAS controller. Finally, the administrative access should use LDAP over SSL for centralized authentication and authorization rather than NIS.

After primary and tenant administrators are established, VLAN use needs to be planned. VLANs are one of the primary means of segregating traffic in the network layer and the underlying storage needs to follow suit. Proper planning is required before setting up the vFiler units to ensure 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 with the same level of separation.

The next step is to disable 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. SSH2 should be used instead of SSH1. 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.

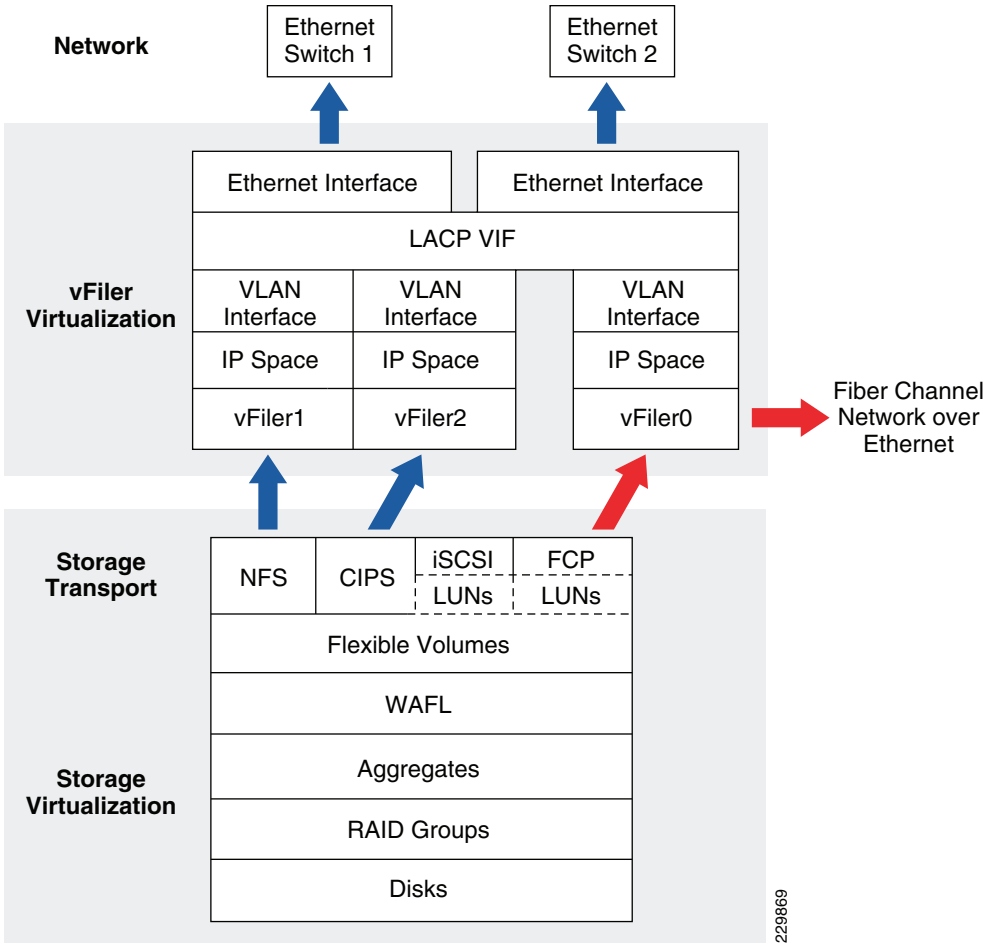
Access to the iSCSI LUNs requires entering the exact initiator name when creating the igroup. From there, the network interfaces that accept iSCSI traffic should be limited as well. An additional layer of security can be added to the iSCSI LUN by enforcing a CHAP login/password challenge. For FCP LUNs, NetApp highly recommends 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 (although we are using Data ONTAP 8.0.1, the vast majority of concepts and commands are identical).

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 7 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.

In the following example, we see that the NetApp controller has three vFiler instances running:

```
dc22-netapp1> vfiler status
vfiler0                          running
client_vfiler_1                  running
server_vfiler_1                  running
```

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 and iSCSI) are served by unprivileged vFiler units. In this case, this refers to the NFS exports to the front-end and back-end servers, as well as the iSCSI LUN used by the back-end servers.

After switching vFiler contexts (essentially logging into a different vFiler instance), we see that "server_vfiler_1" has its own NFS exports as well as an iSCSI LUN:

```
dc22-netapp1*> vfiler context server_vfiler_1
Fri Aug 26 07:51:20 EDT [server_vfiler_1@dc22-netapp1: cmds.vfiler.console.switch:notice]:
Console context was switched to a vFiler(tm) unit server_vfiler_1.
/vol/fp_server01
-sec=sys,rw=192.168.88.150:192.168.88.151,root=192.168.88.150:192.168.88.151
/vol/fp_server02
-sec=sys,rw=192.168.90.150:192.168.90.151,root=192.168.90.150:192.168.90.151
server_vfiler_1@dc22-netapp1*> lun show
    /vol/fp_lunvol/fp_lun      120.0g (128861601792)  (r/w, online, mapped)
```

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.

Primary 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 architecture include various physical and virtual objects, including storage containers and storage controllers. These objects play an important role in maintaining overall operations. However, from a security perspective, they are treated exactly the same as tenant resources and are isolated from other tenants as such. 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 primary 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 primary administrator owns all storage allocations, tenants can only use the storage directly allocated to their vFiler unit. If additional storage is needed, the primary administrator may resize the allocated flexible volume(s) for that tenant or assign an additional flexible volume. Tenants cannot use more than their allocated storage. Only the primary administrator, who can responsibly manage storage resources among the tenants, has the ability to allocate storage capacity.

Tenant Perspective

Each tenant possesses their 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 their assigned storage resources or export storage by using LUNs and leverage iSCSI with CHAP between the application systems and the storage. The method by which application 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 primary administrator) and storage deployment (managed by the tenant administrator).

Availability

Secure Separation built on FlexPod is designed as a highly available environment that provides active and standby redundancy in all components. Eliminating planned downtime and preventing unplanned downtime are key aspects in the design of the multi-tenant shared services infrastructure. This section discusses availability design considerations and best practices related to compute, network, and storage.

Highly Available Topology

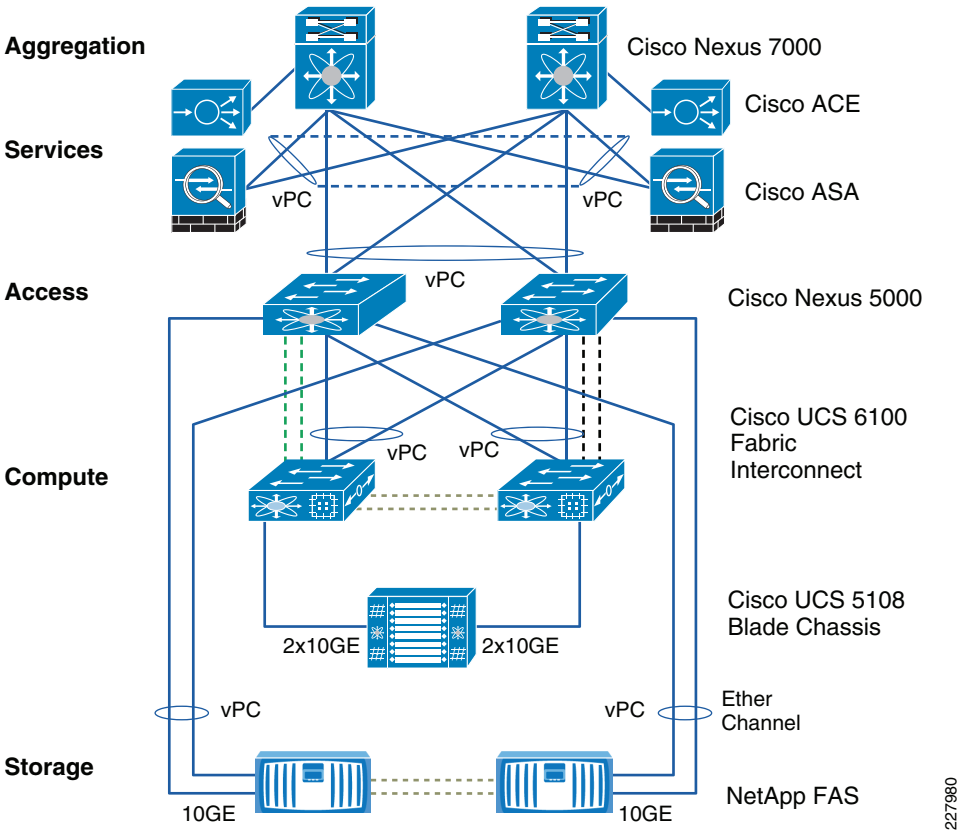
Red Hat Enterprise Linux creates resiliency for the backend of the e-commerce workflow with the Red Hat Enterprise Linux High Availability Add-On. Cisco UCS provides a unified compute environment with integrated management and networking to support compute resources. At the network layer, the architecture is enabled with Cisco Nexus 5000 as a unified access layer switch and Cisco Nexus 7000 as a virtualized aggregation layer switch. The two UCS 6120 Fabric Interconnects provides a robust compute layer platform. Via Virtual Port-Channel (vPC), a topology with redundant chassis, card, and links with Cisco Nexus 5000 and Cisco Nexus 7000 provides a loop less topology.

Both the UCS 6120 Fabric Interconnects and NetApp FAS storage controllers are connected to the Cisco Nexus 5000 access switch via EtherChannel with dual-10 Gig Ethernet. The NetApp FAS controllers use redundant 10Gb NICs configured in a two-port VIF. Each port of the VIF is connected to one of the upstream switches, allowing multiple active paths by utilizing the Cisco 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 port-channel 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 Cisco Nexus 5000 via dual-controller FCoE adapters in a full mesh topology.

The Cisco Nexus 7000 provides redundant paths to the Cisco Nexus 5000 access layer and ASA via vPC. vPC provides a logically loopless topology with convergence times based on Etherchannel and not spanning tree.

Figure 8 'High Availability System Design



Compute Availability Design Considerations

Red Hat Enterprise Linux High Availability Add-On

The High Availability Add-On to Red Hat Enterprise Linux is a cluster system that provides reliability, scalability, and availability to critical production services. A cluster created with the High Availability Add-On can include two or more nodes. It supports highly available services by eliminating single points of failure and by failing over services from one cluster node to another in case of a node becomes inoperative.

The High Availability Add-On is an integrated set of software components that can be deployed in a variety of configurations. It has the following major components:

- Cluster infrastructure—Provides fundamental functions for nodes to work together as a cluster: distribution of the cluster configuration changes, supporting heartbeat and quorum disk voting, membership management, lock management, and fencing.
- High availability service management—Provides failover of services and associated resources like IP addresses and file systems from one cluster node to another.
- Cluster administration tools—Configuration and management tools for setting up, configuring, and managing the cluster.

On FlexPod the cluster nodes are implemented as UCS service profiles that can be deployed on any physical blades. A typical cluster for one securely separated tenant requires configuration of at least four VLANs:

- Application tier network for client access. The cluster framework provides failover of IP addresses assigned to the highly available services between cluster nodes making failover transparent to the clients.
- Cluster interconnect is used for heartbeats and lock management between cluster nodes. This network must be configured to support multicast traffic.
- iSCSI or NFS data access network(s) providing connectivity to the shared storage devices. Optionally one small iSCSI attached LUN can be used as a quorum disk.
- Cisco UCS Manager access network for fencing.

Fencing is one of the most important features of the High Availability Add-On. Fencing is ultimately responsible for guaranteeing the integrity of the cluster and preventing data corruption on the shared storage. Red Hat Enterprise Linux High Availability Add-On includes a fencing agent that can work through Cisco UCS Manager. It can identify and fence cluster nodes by the name of the corresponding UCS service profiles and does not depend on the physical blade address. If the Cisco UCS Manager access VLAN is shared between tenants, it is recommended to protect each cluster node with the iptables packet filter rules allowing communication only with UCS Manager host.

Red Hat Enterprise Linux provides network high availability with the channel bonding driver. The high availability of the iSCSI storage access can be implemented with the device-mapper-multipath driver. It is recommended to use these core operating system features on the cluster nodes.

UCS Adapter HA Implementation

The test bed used Cisco UCS M81KR Virtual Interface Card Adapter or VIC in all of the UCS B-Series blades used. The VIC is a Converged Network Adapter (CNA) that virtualizes adapter instances which are presented to the host as physical Ethernet and Fiber Channel interfaces. Each VIC virtual Network Interface Card (vNIC) is configured to talk through a specific fabric side (A/B), or in the case of Ethernet vNICs, an option that can failover between the fabrics.

These vNIC interfaces are seen as 10G capable, but can be egress rate limited, as well as QoS prioritized.

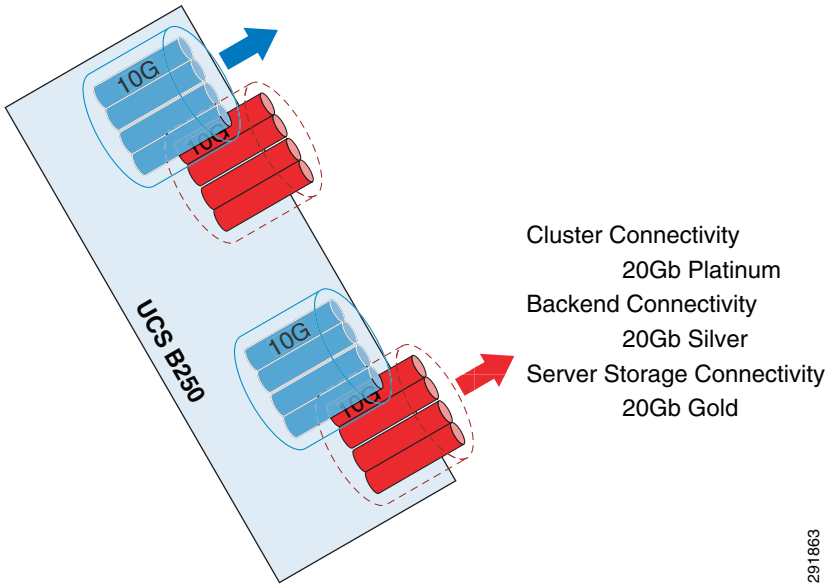
For the secure separation deployment, there were three configurations used on the blades to provide examples of the options available in differing use cases. Each of these configurations had QoS prioritizations assigned to the vNICs of the hosts based on the tenant and application needs. These options were:

- OS interface bonding without QoS egress
- Fabric Failover with QoS egress
- Fabric Failover without QoS egress

Backend Server Example Adapters

The OS interface bonding without QoS egress configuration was used on the Backend Server blades hosting the database. These blades were UCS B250 M2s with two VIC adapters installed.

Figure 9 Backend Server Example Adapter

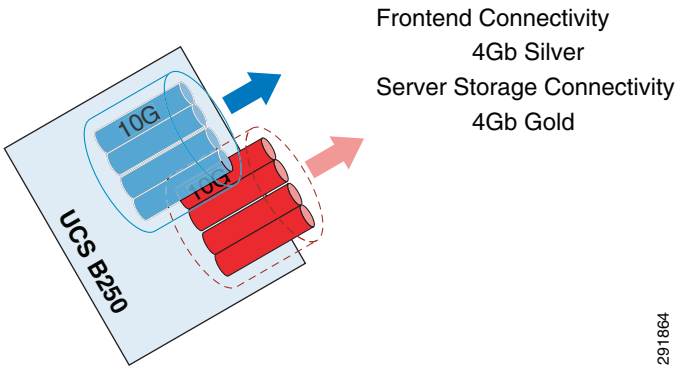


With multiple adapters in place, the vNICs were deployed in pairs bonded within the OS. These bonds have a higher bandwidth potential within the OS, but also an added layer of resiliency in being able to function through an adapter level fault event.

Frontend Server Example Adapters

The Fabric Failover with QoS egress configuration was used on the Frontend Server blades hosting the web application. The Frontend Servers ran from UCS B200s, each with a single VIC adapter.

Figure 10 Frontend Server Example Adapter

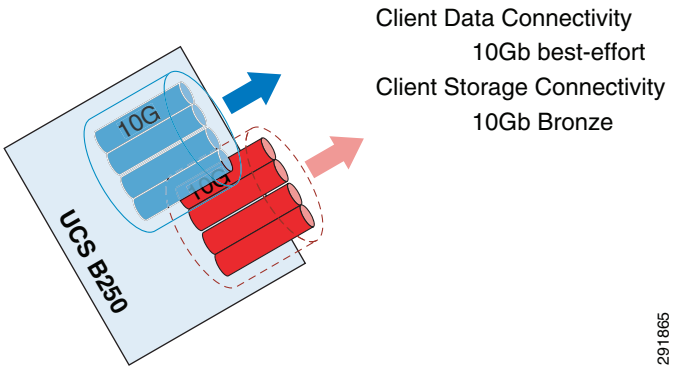


This implementation used fabric failover for the vNICs provisioned and QoS egress as an example of committing outbound traffic to the blade to not exceed levels that would potentially conflict with the other vNICs assigned to the host. (Fiber Channel data is not pictured in Figure 10, but makes up 2Gb of the cumulative bandwidth of VIC CAN.)

Client Example Adapters

Fabric Failover without QoS egress was used on our simulated Client hosts in the secure separation test bed.

Figure 11 Client Example Adapter



This third configuration type utilizes concepts similar to the Frontend Servers, but has a much lower QoS priority for its traffic and no QoS egress considerations between the vNICs.

Unified Computing System Service Profile

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 in UCS Manager, 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.

Network Availability Design Considerations

Fiber Interconnect and Fabric Extender Redundancy

The UCS architecture includes Fabric Interconnects (6100 XP) to attach each chassis to the network infrastructure. The Fabric Interconnects also house the UCS manager. The Fabric Extenders (2100 XP) connect the 6100 to the blade servers. With so much functionality in these components, it is desirable to build high availability into the design involving these components. Redundant pairs of the both the 6100 XP and 2100 XP provide the following for this design:

- Fabric Availability—The UCS provides two completely independent fabric paths A and B.
- 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 Link Aggregation Control Protocol (LACP) “active-active” mode to Cisco Nexus 5000.

Cisco Nexus 5000 Distribution Redundancy

In this design the Cisco Nexus 5000 acts as the access layer device connecting edge layer devices to the infrastructure as well as connecting any storage devices to the architecture. The feature capabilities of the Cisco Nexus 5000 facilitate the creation of a highly available design at this layer. It allows for a loop-less topology via vPC technology. The two-tier vPC design is enabled such that all paths from end-to-end are available for forwarding. The Cisco Nexus 5000 can be connected to the aggregation layer as well as any edge devices using vPC. This design recommendation is employed in this architecture.

To allow for the use of vPC going to edge devices, a pair of Cisco Nexus 5000 switches are inserted in the access layer. This provides device redundancy as well as path redundancy for connecting to both the aggregation layer and edge layer.

Cisco Nexus 7000 Aggregation Redundancy

The Cisco Nexus 7000 is employed as the aggregation layer device. 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 Cisco Nexus 7000. Device redundancy is recommended to ensure high availability. Virtual port channels can be used to connect the aggregation layer to the access or services layer to achieve path redundancy. vPC allows for redundant paths between entities while simultaneously removing the sub-optimal blocking architecture associated with traditional spanning tree designs.

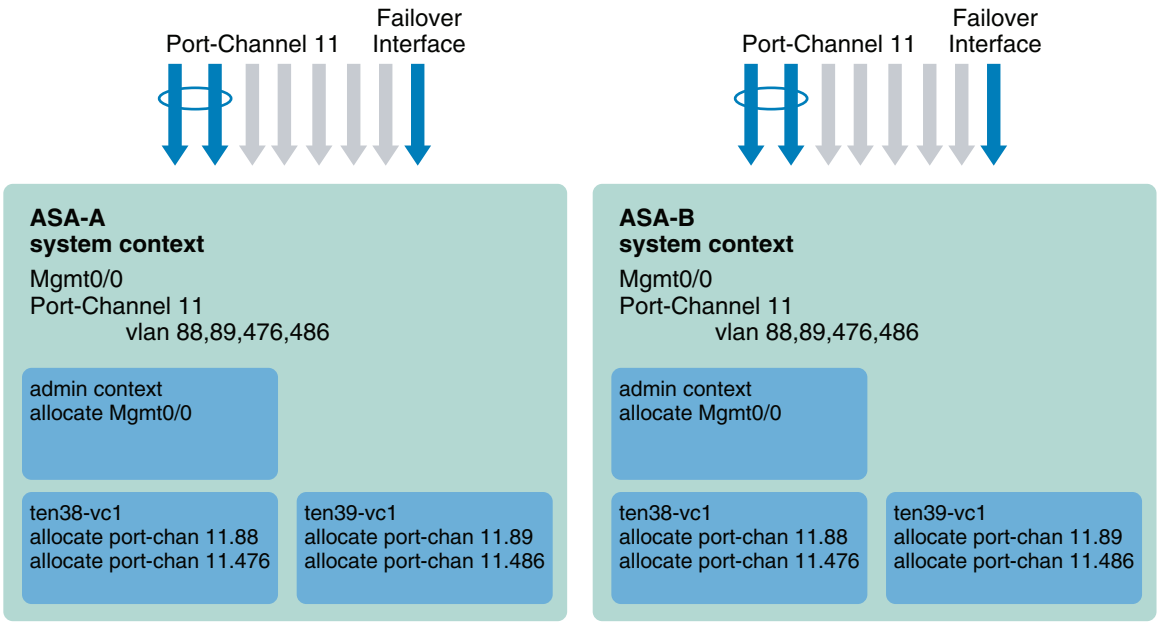
At the Layer 3 level, redundant pairs of VRF instances provide Layer 3 services for their associated tenant VLAN segments. First hop redundancy can also be provided using HSRP. HSRP can be used to provide gateway redundancy for the edge devices in the data center. Each switch can become an HSRP peer in however many groups are required for the design. Ideally, each tenant would have an HSRP group.

Advanced Security Appliance Redundancy

This design incorporates services such as firewall and server load balancing via appliances. To ensure a highly available design at this layer, redundant devices are essential. For the ASA, an Active/Active or Active/Standby failover pair can be configured. With Active/Active, it is ensured that each device in the pair will be utilized by dividing the load according to virtual contexts. Virtual contexts allow for each tenant to have their own virtual firewall. With Active/Standby each tenant can still have their own virtual firewall or context, but all contexts are handled by the same ASA. To ensure that every resource is utilized, the Active/Active configuration is used in this design.

The ASA is also capable of recognizing port channels. This allows for the opportunity to create path redundancy when connecting to the Cisco Nexus 7000s in the aggregation layer. A vPC can be configured going to each ASA to ensure HA at the path level.

Figure 12 Port Channel Design



In this configuration, the VLANs for the security contexts are carried on a common vPC. The Client tenant (ten38-vc1) context for the outside traffic is carried on the global VRF within VLAN 88. The inside traffic for the Client tenant is carried through VLAN 476, which is a member of the Client tenant VRF. The Server tenant (ten39-vc1) is similarly configured with allocations of outside traffic through VLAN 89 and inside traffic shown on VLAN 486. Within failover groups (not pictured) the active states of each tenant security contexts are split between the two ASAs.

Application Control Engine Redundancy

The ACE appliance can be inserted into the design with the same options available. As with the ASA, Active/Active or Active/Standby failover pair can be used. Once again the Active/Active design is chosen to ensure full utilization of each of ACE appliances in the pair. It is also capable of recognizing port channels so a vPC from the aggregation Cisco Nexus 7000s is also possible.

Of course the main purpose of choosing the ACE for this design is to allow for server redundancy and load balancing. Two or more servers performing the same role in an application environment share the load of client requests with the ACE facilitating the administration of the arrangement. The ACE can designate a virtual IP address or VIP for the servers to share and load balance the requests coming to the VIP among the servers designated by the administrator. This ensures that no one server gets overloaded with client requests and that services remain available if one of the servers becomes incapacitated.

Data Availability with RAID Groups and Aggregates

RAID groups are the fundamental building blocks when constructing resilient storage arrays containing any type of application datasets or host 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 host boot LUNs, guest virtual machines, and application datasets 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 datasets or systems 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 host's 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.

Here is an example of an aggregate:

```
dc22-netapp1*> aggr status
           Aggr State           Status            Options
        fp_aggr online          raid_dp, aggr
                                32-bit
```

Here we see the volumes within that aggregate:

```
dc22-netapp1*> aggr status fp_aggr
           Aggr State           Status            Options
        fp_aggr online          raid_dp, aggr
                                32-bit

       Volumes: fp_client, fp_server01, client_data_1, fp_bootvol,
                server_data_1, fp_lunvol, fp_server02,
                fp_backend_iscsi

                Plex /fp_aggr/plex0: online, normal, active
                    RAID group /fp_aggr/plex0/rg0: normal
```

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, including 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, see: http://media.netapp.com/documents/ds-3100.pdf.

In the following example, we see that the cluster status is good, followed by an example of an interface that has been configured with a "partner". That is to say, if the NetApp controller were to fail, the surviving node would take over that interface and service any storage requests.

```
dc22-netapp1*> cf status
Cluster enabled, dc22-netapp2 is up.
Interconnect status: up.
dc22-netapp1*> ifconfig CNA-VIF0-489
CNA-VIF0-489: flags=0x4b48863<UP,BROADCAST,RUNNING,MULTICAST,TCPCKSUM,NOWINS> mtu 9000
    inet 192.168.89.100 netmask 0xffffff00 broadcast 192.168.89.255
    partner CNA-VIF0-489 (not in use)
    ether 02:a0:98:14:47:6a (Enabled interface groups)
```

Storage Network Connectivity (Virtual Interfaces) Using Link Aggregation Control Protocol

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:

- Single-mode
- 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 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.

This example shows a dynamic multimode VIF configured from the 10GB Ethernet portion of a converged network adapter (CNA). The VIF is composed of devices e1a and e1b:

```
CNA-VIF0: 2 links, transmit 'IP Load balancing', Ifgrp Type 'lacp' fail 'default'
     Ifgrp StatusUp Addr_set
    up:
    e1b: state up, since 19Aug2011 17:19:16 (6+15:24:30)
        {snipped for brevity}
    e1a: state up, since 19Aug2011 17:08:41 (6+15:35:05)
        {snipped for brevity}
```

Service Assurance

Network Service Assurance Design Considerations

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. 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 Cisco 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 the Source of Origin

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

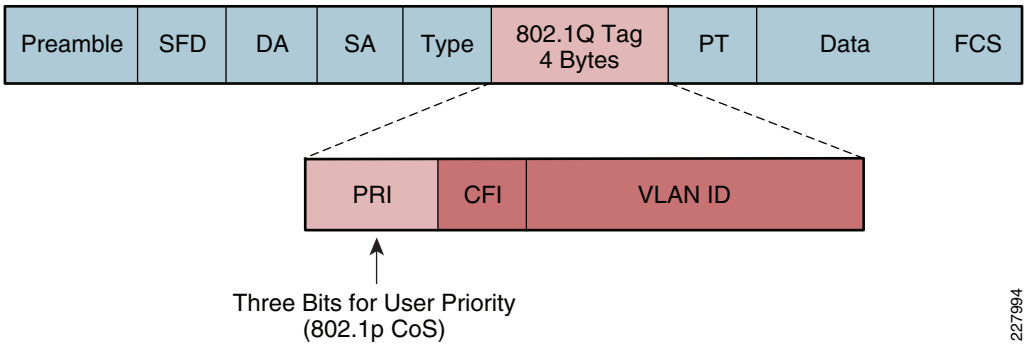
Classification Capability of Layer 2 Network

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 Data Center Bridging (DCB). 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 802.1Q/p CoS Bits. 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 13 802.1Q/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 Figure 13.

Table 3 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 primary network administrator 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 two major types of traffic flow under the infrastructure category:

  - Control Plane Traffic—This can include traffic required for non-virtualized server administration such as cluster connectivity traffic. 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 server management access, storage and network device management, and per-tenant traffic (application and server 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.

- 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 server-to-server 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 4. 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 4 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 4 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

Table 4 Services Levels for Back-End User Data Traffic

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 server 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 require immediate response from the application in either direction (examples 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 applications 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 5.

Table 5 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 6.

Table 6 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).

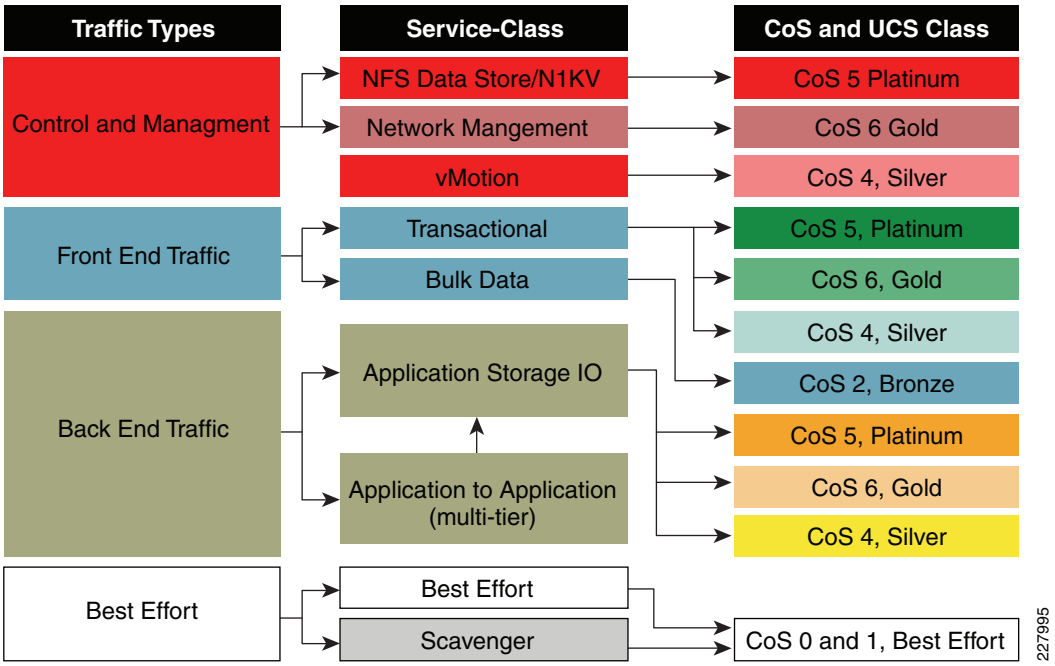
  In this design guide the best effort and scavenger classes are merged into one class called "best-effort" in UCS 6100 and "class-default" in Cisco Nexus 5000.

Table 7 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 14 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 14 Service Class Model



Classify the Packet Near 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 Traffic Originating from Hosts
- Classification for Traffic Originating External to the Data Center
- Classification for Traffic Originating from Networked Attached Devices

Classification for Traffic Originating from Hosts

Because this design is operating in a non-virtualized environment on UCS, QoS marking must take place on the adapter of the server. When using the Cisco M81KR VIC, all traffic leaving the vNIC can be classified as a singular CoS value. For a server housing multiple levels of traffic, a separate vNIC template can be used to specify the particular CoS value required for that level of traffic.

Classification for Traffic Originating External to the Data Center

The Cisco 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 Cisco 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 Cisco 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 Traffic Originating from Networked Attached Devices

In this design the Cisco 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 Cisco 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 Cisco Nexus 5000 QoS model consists of a separate class and policy map for each type of QoS function. QoS is an umbrella framework of many functionalities and the QoS functionality in Cisco 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 are implemented with the Modular QoS CLI (MQC) framework.

For more information, see: http://www.cisco.com/en/US/technologies/tk543/tk545/technologies_white_paper09186a0080123415.html.

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. 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 has the capability to provide bandwidth control at the vNIC level. With the Cisco VIC, each vNIC can be policed to a desired bandwidth.

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 8 is developed based on the minimum queuing class available based on end-to-end network capability.

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:

- Application to UCS Blades—The density of the application driving server network activity can oversubscribe 10G interface.
- 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-server to server (either server residing on a separate blade or to other UCS systems in a domain)
  - Server to storage (NFS datastore and Application IO per tenant)—Front end user data traffic-server 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 8. 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; for example, if Platinum class is assigned a 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 8. Notice that the weight multiplier can range from 1 to 10. The multiplier automatically adjusts the total bandwidth percentage to 100%. Table 8 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 one (1) is referred as best-effort, however that does not mean the traffic in the respective class is treated as best-effort.

Table 8 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 8 Unified Computing System—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 Cisco Nexus 5000 with two 10Gb/second links forming a single EtherChannel. 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-controllers 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 server (read IO) is managed at the Cisco Nexus 5000 with bandwidth control.
- Front-end user traffic—In this design the application flows from server 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 server) bandwidth control can be managed per class at either the Cisco Nexus 7000 or Cisco Nexus 5000. In this design the Cisco Nexus 5000 is used as the bandwidth control point.

The Cisco Nexus 5000 QoS components are described in QoS-Based Classification. The queuing and bandwidth capability reflecting the above requirements are shown in Table 9. In the Cisco 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 a WRR (weighted round robin) schedule. For more information on Cisco 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 9 shows the mapping of CoS to queue and bandwidth allocation. Table 9 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 9 Cisco 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 Cisco 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 Cisco 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. 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. The Cisco VIC can be used to rate limit certain vNICs as needed.

Storage I/O Service Assurance Design Considerations

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 Data ONTAP operating system. 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 datasets 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 operations 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 affect 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, see: http://www.netapp.com/us/products/platform-os/flexshare.html and 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 systems 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 thin-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.

Here is an example of a thin-provisioned (space reservation is disabled) LUN:

```
dc22-netapp1*> lun show -v /vol/fp_bootvol/back_01
    /vol/fp_bootvol/back_01     40.0g (42953867264)   (r/w, online, mapped)
        Comment: "FlexPod back end 2 boot lun"
        Serial#: dfX/N4eR8/m9
        Share: none
        Space Reservation: disabled
        Multiprotocol Type: linux
        Maps: iBack_01=0
        Occupied Size:    7.5g (8075890688)
        Creation Time: Fri Aug  5 09:33:12 EDT 2011
        Alignment: aligned
        Cluster Shared Volume Information: 0x0
```

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

As multi-tenant environments grow, so do the challenges of managing them. The components in the secure separation system design have element managers with rich feature sets and open APIs. The element managers are listed below as cross references to the product descriptions earlier in this document.

- Compute
  - Cisco UCS Manager
  - Red Hat Network Satellite
- Network
  - Cisco Data Center Network Manager
  - Cisco Application Network Manager
  - Cisco Security Manager
- Storage
  - NetApp Provisioning Manager

  - NetApp Protection Manager
  - NetApp Operations Manager

The management APIs allow for customization of configuration, monitoring, and process automation through third-party or independently-created interfaces. This enables customers to custom build a management environment which meets their business needs.

References

- FlexPod Deployment Guide CVD
  http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/Virtualization/flexpod_deploy.html
- Red Hat Enterprise Linux built on FlexPod Deployment Guide CVD
  http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/Virtualization/flexpod_rhel.html
- Enhanced Secure Multi-tenant Design Guide CVD
  http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/Virtualization/securecldg_V2.html
- Cisco Virtual Multi-tenant Data Center Design Guide CVD
  http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/VMDC/2.1/design_guide/vmdc21DesignGuide.html
- Red Hat Online Documentation: http://docs.redhat.com
- Red Hat Customer Portal: https://access.redhat.com
- Cisco Unified Computing System: http://www.cisco.com/en/US/netsol/ns944/index.html
- NetApp Support (formerly NetApp on the Web [NOW®]) site: http://.now.netapp.com

Glossary

Term	Meaning
ACE	Cisco Application Control Engine
ACL	Access Control List
ASA	Cisco Adaptive Security Appliance
CDP	Cisco Discovery Protocol
CNA	Converged Network Adapter
CoS	Class of Service
CVD	Cisco Validated Design
DCNM	Data Center Network Manager
DNS	Domain Name Services
FAS	Fabric Attached Storage
FC	Fibre Channel
FCoE	Fibre Channel over Ethernet

Term	Meaning
Gbps	Giga (10^9) bits per second
IPS	Cisco Intrusion Prevention System
LAN	Local Area Network
LUN	Logical Unit Number
NAM	Network Analysis Module
NFS	Network File System
QoS	Quality of Service
RBAC	Role Based Access Control
RHN	Red Hat Network
SAN	Storage Area Network
Tbps	Tera (10^{12}) bits per second
TCO	Total Cost of Ownership
UCS	Unified Compute System
VDC	Virtual Device Contexts
VIC	Virtual Interface Card (Cisco UCS M81KR)
VLAN	Virtual Local Area Network
VM	Virtual Machine
vNIC	virtual Network Interface Card
VPN	Virtual Private Network
VRF	Virtual Routing and Forwarding