Cloud Infrastructure Architecture Case Study

VMware vSphere® 5.0 and
VMware vShield App 5.0

TECHNICAL WHITE PAPER
V 1.0 JANUARY 2012

Table of Contents

Design Subject Matter Experts

The following people provided key input into this design:

NAME	TITLE	ROLE
Duncan Epping	Principal Architect	Author
Aidan Dalgleish	Consulting Architect – Center of Excellence	Contributor
Frank Denneman	Technical Marketing Architect – Resource Management	Contributor
Alan Renouf	Technical Marketing Manager – Automation	Contributor
Cormac Hogan	Technical Marketing Manager – Storage	Reviewer
Vyenkatesh Deshpande	Technical Marketing Manager – Networking	Reviewer
Matthew Northam	Security and Compliance Specialist SE	Reviewer

Purpose and Overview

The VMware® Cloud Infrastructure Suite (CIS) consists of five technologies that together expand the capabilities and value that customers can realize from a virtualized infrastructure. CIS is designed to help organizations build more intelligent virtual infrastructures. It does so by enabling highly virtualized environments with the automation, self-service and security capabilities that customers require to deploy business-critical applications, respond to business demands more quickly and move to a secure cloud model. The CIS is based on the VMware vSphere® platform as its foundation in pursuing any type of cloud infrastructure. In addition to vSphere, the CIS also includes VMware vShield App, VMware vCenter Site Recovery Manager™ Server (SRM Server), VMware vCloud® Director™, and VMware® vCenter™ Operations Manager.

The VMware Cloud Infrastructure Architecture Case Study Series was developed to provide an understanding of the various components of the CIS. The goal is to explain how these components can be used in specific scenarios, which are based on real-world customer examples and therefore contain real-world requirements and constraints. This document is the first in a series of case studies, with each case study focusing on a different use case with different requirements and constraints.

This document provides both logical and physical design considerations encompassing components that are pertinent to this scenario. To facilitate the requirements of this case study, these considerations and decisions are based on a combination of VMware best practices and specific business requirements and goals. Cloud infrastructure–related components, including requirements and specifications for virtual machines and hosts, security, networking, storage, and management, are included in this document.

Executive Summary

This architecture was developed to support a virtualization project to consolidate 200 existing physical servers. The required infrastructure that's defined here will be used not only for the first attempt at virtualization but also as a foundation for follow-on projects to completely virtualize the IT estate and to prepare it for the journey to cloud computing.

Virtualization is being adopted to decrease power and cooling costs, reduce the need for expensive datacenter expansion, increase operational efficiency and capitalize on the higher availability and increased flexibility that comes with running virtual workloads. The goal is for IT to be well positioned to respond rapidly to ever-changing business needs.

After this initial foundation architecture has been successfully implemented, it can be horizontally scaled and expanded when cluster limits have been reached, using similar clusters in a building block approach.

Case Background

Company and project background:

- The company is a financial institution.
- The project was initiated by the local insurance business unit.
- The vision for the future of IT is to adopt a "virtualization first" approach and increase the agility and availability of service offerings.
- This first "foundation infrastructure" is to be located at the primary site.
- The initial consolidation project targets 200 x86 servers, including 30 servers currently hosted in a DMZ, out of an estate of 600 x86 servers, which are candidates for the second wave.

Interpreting This Document

The overall structure of this design document is, for the most part, self-explanatory. However, throughout this document there are key points of particular importance that will be highlighted to the user. These points will be identified with one of the following labels:

- **Note** – General point of importance or to add further explanation on a particular section
- **Design decision** – Points of importance to the support of the proposed solution
- **Requirements** – Specific customer requirements
- **Assumption** – Identification of where an assumption has been made in the absence of factual data

This document captures design decisions made in order for the solution to meet customer requirements. In some cases, customer-specific requirements and existing infrastructure constraints might result in a valid but suboptimal design choice.

Requirements, Assumptions and Constraints

In this case study, the primary requirement for this architecture is to lower the cost of doing business. Business agility and flexibility should be increased while operational effort involved with deploying new workloads should be decreased.

Throughout this design document, we will adhere to the standards and best practices as defined by VMware when and where aligned with the requirements and constraints as listed in the following sections.

Case Requirements, Assumptions and Constraints

Requirements are the key demands on the design. Sources include both business and technical representatives.

ID	REQUIREMENT
r101	Business agility and flexibility should be increased; the cost of doing business should be decreased.
r102	Availability of services is defined as 99.9 percent during core business hours.
r103	Security compliance requires network isolation for specific workloads from other services.
r104	Minimal workload deployment time.
r105	A separate management VLAN must be used for management traffic.
r106	The environment should be scalable to enable future expansion (minimum one year, 20 percent estimated).
r107	Resources should be guaranteed to groups of workloads as part of internal SLAs.
r108	The recovery-time objective in the case of a datastore failure should be less than 8 hours.
r109	Servers hosted in the DMZ should be protected within the virtual environment.
r110	N+1 resiliency should be factored in.

Table 1. Customer Requirements

Constraints limit the logical design decisions and physical specifications. They are decisions made independently of this engagement that might or might not align with stated objectives.

ID	CONSTRAINT
c101	Dell and AMD have been preselected as the compute platform of choice.
c102	Eight 1GbE ports will be used per server.
c103	NetApp's NAS offering has been preselected as the storage solution of choice.
c104	All Tier 2 NAS volumes are deduplicated.
c105	All volumes will be RAID-6
c106	The environment should have 20 percent spare capacity for resource bursts.
c107	Physical switches will not be configured for QoS.
c108	Existing Cisco top-of-rack environment should be used for the virtual infrastructure.

Table 2. Design Constraints

Use Cases

This design is targeted at the following use cases:

- Server consolidation (power and cooling savings, green computing, reduced TCO)
- Server infrastructure resource optimization (load balancing, high availability)
- Rapid provisioning (business agility)
- Server standardization

Conceptual Architecture Overview Diagram

The VMware CIS aims to reduce operational overhead and TCO by simplifying management tasks and abstracting complex processes. Throughout this architecture case study, all components will be described in depth, including design considerations for all the components. The focus of this architecture, as indicated by our customer requirements, is resource aggregation and isolation. This will be achieved by using logical containers, hereafter called pools. The environment has the following four major pillars:

- Compute
- Networking
- Storage
- Security

Each of the pillars will be carved into multiple pools to provide different service levels for the various workload types. This will be achieved by leveraging core functionality offered by vSphere 5.0. It was a requirement to provide a secure and shielded environment for the Web farm that is currently hosted in a DMZ. To meet these requirements, VMware vShield App will be implemented to enable the use of security policies on a pool level. Administrators can define and enforce granular policies for all traffic that crosses a virtual network adaptor, increasing visibility over internal virtual datacenter traffic while helping to eliminate detours to physical firewalls.

As a hypervisor-based application-aware firewall solution, vShield App allows defining policies to logical, dynamic application boundaries (security groups) instead of physical boundaries.

This resource and security layering method enables a fast, safe deployment of new workloads.

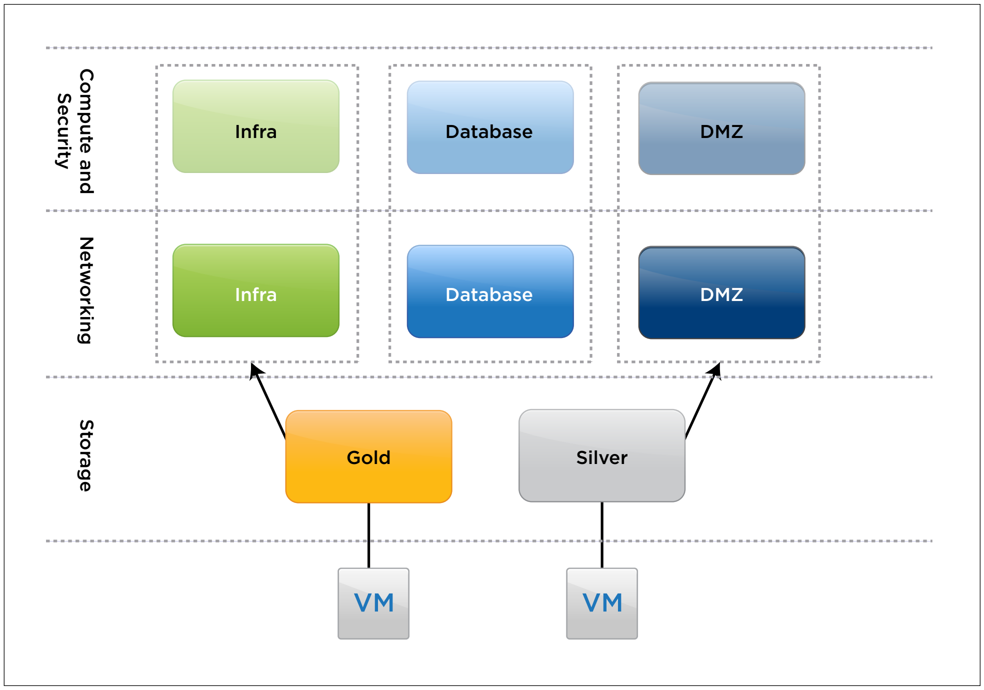

Figure 1. Conceptual Overview

Each different type of pillar is carved up into different pools for each of the respective workload types. A virtual machine, or vSphere vApp (a logical container for one or more virtual machines), will be deployed in one of the three various compute resource pools, after which a specific networking-and-security pool will be selected as well as a storage pool (often referred to as tier). Compute, network and security pool types are currently defined based on the type of workload the virtual infrastructure will host. In the future, additional blocks might be added based on requirements of internal customers and the different types of workloads being deployed.

Sizing and Scaling

VMware recommends using a building block approach for compute resources for this vSphere 5.0 environment. By using this approach, a consistent experience can be guaranteed for internal customers. This design will enable both horizontal and vertical scaling when required.

Sizing is based on evaluation of the virtualization candidates. This section will describe the required number of hosts, based on this analysis and on this study's requirement (r106) to have at least one year of growth factored in.

Workload Estimations

To determine the required number of VMware ESXi™ hosts needed to consolidate x86 virtualization candidates, performance and utilization have been analyzed using VMware Capacity Planner™. (For this study, we have used data gathered from a real-world study comparable to this case.) The analysis primarily captured the resource utilization for each system, including average and peak CPU and memory utilization. Table 3 summarizes the results of the CPU analysis. It details the overall CPU requirements for the ESXi hosts to support the workloads of the proposed virtualization candidates.

All values have been rounded up to ensure that sufficient resources are available during short resource bursts.

PERFORMANCE METRIC	RECORDED VALUE
Average number of CPUs per physical system	2.1
Average CPU MHz	2,800MHz
Average CPU utilization per physical system	12% (350MHz)
Average peak CPU utilization per physical system	36% (1,000MHz)
Total CPU resources for all virtual machines at peak	**202,000MHz**
Average amount of RAM per physical system	2,048MB
Average memory utilization per physical system	52% (1,065MB)
Average peak memory utilization per physical system	67% (1,475MB)
Total RAM for all virtual machines at peak (no memory sharing)	275,000MB
Assumed memory-sharing benefit when virtualized	25% (*)
Total RAM for all virtual machines at peak (memory sharing)	**206,000 MB**

Table 3. VMware ESXi Host CPU and Memory Requirements

Using the performance data gathered in conjunction with the CPU and RAM requirements analysis, it is possible to derive the high-level CPU and RAM requirements that an ESXi host must deliver. The following tables detail the high-level CPU and memory specifications of the Dell PowerEdge R715 server that are pertinent to this analysis and case study (constraint c101).

ATTRIBUTE	SPECIFICATION
Number of CPUs (sockets) per host	2
Number of cores per CPU (AMD)	8
MHz per CPU core	2,300MHz
Total CPU MHz per CPU	18,400MHz
Total CPU MHz per host	36,800MHz
Proposed maximum host CPU utilization	80%
Available CPU MHz per host	**29,400MHz**

Table 4. ESXi Host CPU Logical Design Specifications

Similarly, the following table details the high-level memory specifications of the Dell PowerEdge R715 server that are pertinent to this analysis. The analysis was performed with both 96GB and 192GB of memory. From a cost perspective, using a configuration with 96GB is recommended because this would provide sufficient capacity for the current estate while still allowing room for growth.

* The estimated savings from memory sharing intentionally has been kept low due to the fact that most guest operating systems (OS) will be 64-bit and, as such, large memory pages will be used. For more details, read VMware knowledge base articles 1021095 and 1021896.

ATTRIBUTE	SPECIFICATION
Total RAM per host	96,000MB
Proposed maximum host RAM utilization	80%
Available RAM per host	**76,800MB**

Table 5. ESXi Host Memory Logical Design Specifications

Using the high-level CPU and memory specifications detailed in Tables 4 and 5, we have derived the minimum number of ESXi hosts required from the perspectives of both CPU and memory. The minimum number of hosts is the higher of the two values. The following table details the number of ESXi hosts necessary from the perspectives of CPU and RAM to satisfy the resource requirements. It should be noted that 20 percent head room for both CPU and RAM has been taken into consideration.

TYPE	TOTAL PEAK RESOURCES REQUIRED	AVAILABLE RESOURCES PER HOST	ESXI HOSTS NEEDED TO SATISFY RESOURCE REQUIREMENTS
CPU	202,000MHz	29,440MHz	7
RAM	206,000MB	76,800MB	3

Table 6. ESXi Host Requirements

Using an anticipated growth rate of 20 percent (r105), the following table shows the required number of hosts for this environment:

NUMBER OF ESXi HOSTS REQUIRED	PERCENTAGE OF GROWTH FACTORED IN	AVAILABILITY REQUIREMENTS	NUMBER OF ESXi HOSTS REQUIRED
7	20%	N+1	10

Table 7. VMware ESXi Hosts Required for Project

Network and Storage

In many cases, the network bandwidth profile of virtual machines is overlooked and a general assumption is made regarding the number of network adaptors required to fulfill the combined bandwidth requirements for a given number of virtual machines. The analysis has shown that the expected average network bandwidth requirement is 4.21Mbps, based on an average of 20 virtual machines per ESXi host. In this case study, network bandwidth is not a limiting factor, because a minimum of eight network adaptors will be used per ESXi host, of which two will be dedicated to virtual machine traffic (constraint c102).

Storage

When designing a storage solution, it is important to understand the I/O profile of the virtual machines that will be placed on the storage. An I/O profile is a description of an application/server I/O pattern. Some applications are heavy on reads and others are heavy on writes; some are heavy on sequential access and others are heavy on random access. In this case, the average I/O requirement for a potential virtual machine has been measured at approximately 42 IOPS.

The sum of the predicted sizes of all of these files for an average virtual machine within a vSphere deployment can be multiplied by the number of virtual machines to be stored per LUN, providing an estimate for the required size of a LUN. The following table provides details of the observed virtualization candidate storage requirements.

AVG C:\ SIZE (GB)	AVG C:\ USED (GB)	AVG 'OTHER':\ SIZE (GB)	AVG 'OTHER':\ USED (GB)
16.59 [34.72]	9.36	93.23	40.74

Table 8. Virtual Machine Storage Capacity Profile

In this case, the average storage requirement for a potential virtual machine has been measured at approximately 50.10GB (9.36GB C:\ and 40.74GB other drive).

Host Design

ESXi is the foundation of every vSphere 5.0 installation. This section will discuss the design and implementation details and considerations to ensure a stable and consistent environment.

Hardware Layout

All of the ESXi hosts have identical hardware specifications and will be built and configured consistently to reduce the amount of operational effort involved with patch management and to provide a building block solution.

Selected Platform

The chosen hardware vendor for this project is Dell. In this instance, the Dell PowerEdge R715 server has been selected as the hardware platform for use with ESXi 5.0. The configuration and assembly process for each system will be standardized, with all components installed identically for all ESXi hosts. Standardizing not only the model but also the physical configuration of the ESXi hosts is critical to providing a manageable and supportable infrastructure by eliminating variability. The specifications and configuration for this hardware platform are detailed in the following table:

ATTRIBUTE	SPECIFICATION
Vendor Model	Dell PowerEdge R715
Number of CPU sockets Number of CPU cores Processor speed	2 8 2.3GHz
Memory	96GB
Number of network adaptor ports Network adaptor vendor(s) Network adaptor model(s) Network adaptor speed	8 Broadcom/Intel 2x Broadcom 5709C dual-port onboard 2x Intel Gigabit ET dual-port Gigabit
Installation destination	Dual SD card
VMware ESXi server version	VMware ESXi 5.0 server; build: latest

Table 9. VMware ESXi Host Platform Specifications

Design/Configuration Considerations

Domain Name Service (DNS) must be configured on all of the ESXi hosts and must be able to resolve short names and Fully Qualified Domain Names (FQDN) using forward and reverse lookup.

Network Time Protocol (NTP) must be configured on each ESXi host and should be configured to share the same time source as the VMware vCenter Server™ to ensure consistency of the overall vSphere solution.

Currently, VMware ESXi 5.0 offers four different solutions (local disk, USB/SD, boot from SAN, stateless) to boot ESXi. During the workshops, customers indicated that they might want to use stateless sometime in the future and therefore wanted to leave that option open, with minimal associated costs. VMware recommends deploying ESXi on SD cards because it will enable a cost-effective migration to stateless when that option is wanted.

For new installations of ESXi, during the auto configuration phase, a 4GB VFAT scratch partition is created if a partition is not present on another disk. In this scenario, SD cards are used as the installation destination, which does not allow for the creation of the scratch partition. Creating a shared volume is recommended. That will hold the scratch partition for all ESXi hosts in a unique, per-server folder. VMware recommends using one of the NFS datastores. Ensure that each server has its own directory and that the scratch partition advanced setting is set as described in VMware knowledge base article 1033696. If wanted, a separate NFS datastore of roughly 20GB can be used to create the scratch partition for each host.

VMware recommends using a minimum of two racks for the 10 hosts and layering the hosts across the racks as depicted in the following diagram. When wiring for redundant power, racks should have two power distribution units (PDUs), each connected to separate legs of a distribution panel or entirely separate panels. The assumption here is that the distribution panels are in turn separately connected to different uninterrupted power supplies. Layering the hosts across the available racks minimizes the impact of a single component failure.

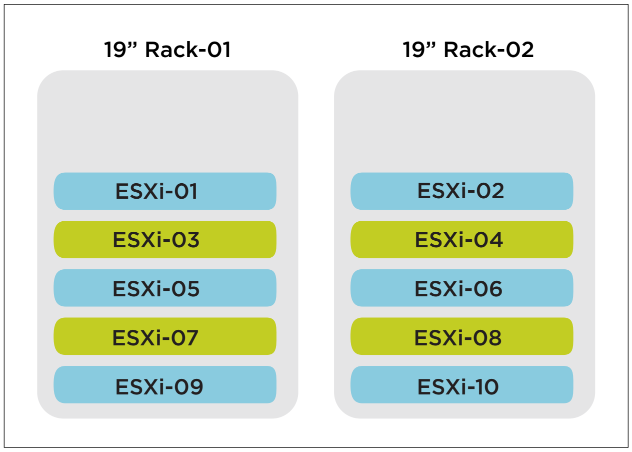

Figure 2. ESXi Host Rack Layout

VMware vCenter Server Design

VMware recommends deploying vCenter Server using a virtual machine as opposed to a standalone physical server. That enables the customer to leverage the benefits available when running in a virtual machine, such as vSphere High Availability (vSphere HA), which will protect the vCenter Server virtual machine in the event of hardware failure.

The specifications and configuration for the vCenter Server virtual machine are detailed in the following table and are based on the recommendations provided in the "vCenter Server Requirements" section of the ESXi and vCenter installation documentation. vCenter Server sizing has grown by 20 percent over the first year. VMware also recommends separating VMware vCenter™ Update Manager from vCenter Server for flexibility during maintenance.

ATTRIBUTE	SPECIFICATION
Vendor Model	VMware virtual machine Virtual hardware version 8
Model	VMware virtual machine
Number of vCPUs	2
Memory	6GB
Number of Local Drives Total Useable Capacity	2 20GB (C:\) and 40GB (E:\)
Operating System	Microsoft Windows 2008 – 64-bit

Table 10. VMware vCenter Server Platform Specifications

Design and Implementation Considerations

When installing vCenter Server 5.0, multiple options are presented. Installing separate components, including syslog and dump functionality, is recommended.

VMware vCenter Update Manager Design

Update Manager will be implemented as a component part of this solution for monitoring and managing the patch levels of the ESXi hosts.

VMware recommends installing Update Manager in a separate virtual machine to enable future expansion to leverage benefits that vSphere 5.0 provides, such as vSphere HA. The specifications and configuration for the Update Manager virtual machine are detailed in the following table and are based on recommendations provided in the Update Manager installation documentation.

ATTRIBUTE	SPECIFICATION
Vendor Model	VMware virtual machine Virtual hardware version 8
Model	VMware virtual machine
Number of vCPUs	2
Memory	2GB

ATTRIBUTE	SPECIFICATION
Number of Local Drives Total Usable Capacity	2 20GB (C:\) and 40GB (D:\)
Operating System	Microsoft Windows 2008 – 64-bit

Table 11. VMware vCenter Update Manager Server Platform Specifications

VMware vCenter Server and vCenter Update Manager Database

vCenter Server and vCenter Update Manager require access to a database. During the installation process, it is possible to install Microsoft SQL Server 2008 Express. However, it is supported only for small deployments not exceeding five ESXi hosts and 50 virtual machines, so there is a requirement to use an alternative supported database. The following table summarizes the configuration requirements for the vCenter and Update Manager databases.

ATTRIBUTE	SPECIFICATION
Vendor and Version	Microsoft SQL 2008 64-bit SP2
Authentication Method	SQL account
vCenter Statistics Level	1
Estimated Database Size – vCenter	10.4GB, 150MB initial + 49–61MB per month
Estimated Database Size – Update Manager	150MB initial + 60–70MB per month
Estimated Disk Utilization – Update Manager	1,050MB initial + 500MB per month

Table 12. VMware vCenter and VMware vCenter Update Manager Database Specifications

To estimate the size requirements of vCenter Server and vCenter Update Manager databases, the VMware vCenter Server 4.1 Database Sizing Calculator for Microsoft SQL Server and VMware vCenter Update Manager 5.0 Sizing Estimator tools have been used.

NOTE: At the time of writing, only the vSphere 4.1—not vSphere 5.0—vCenter database calculator was available.

VMware vSphere Datacenter Design

For this design, there will be a single site, with no secondary sites. Within the vSphere datacenter architecture, the datacenter is the highest-level logical boundary and is typically used to delineate separate physical sites/ locations, or potentially an additional vSphere infrastructure with completely independent purposes.

Cluster

Within a datacenter, ESXi hosts are typically grouped into clusters to provide a platform for different groups of virtual machines requiring different network and storage requirements. Grouping ESXi hosts into clusters also facilitates the use of technologies such as VMware vSphere® vMotion®, vSphere Distributed Resource Scheduler (DRS), vSphere Distributed Power Management (VMware DPM), vSphere High Availability (HA) and vSphere Fault Tolerance (FT). VMware recommends creating a single cluster with all 10 hosts, because multiple clusters would result in a higher overhead from an HA perspective. This means that with N+1 redundancy with two five-node clusters, only eight nodes' (2x 5–1) worth of resources can be used; in a single 10-node cluster, nine nodes' (10–1) worth of resources can be used. This approach also reduces complexity in your environment and avoids the associated operational effort of managing multiple objects.

TYPE	CONFIGURATION VALUE
Number of Hosts	10
DRS	Enabled
HA	Enabled

Table 13. VMware vSphere Cluster Summary

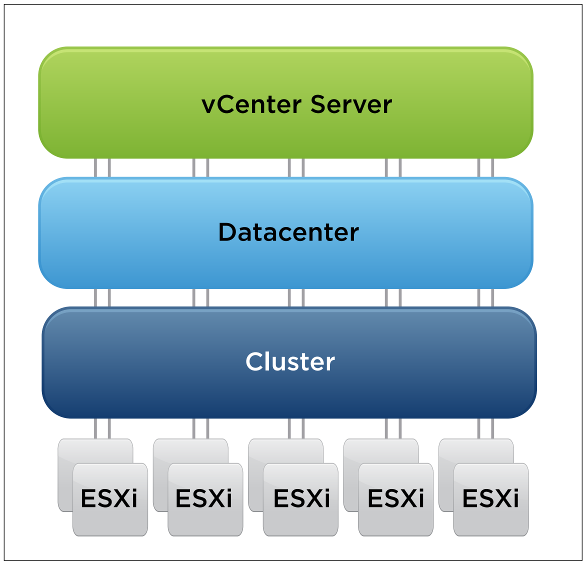

Figure 3. Datacenter and Cluster Overview

VMware vSphere High Availability

HA will be configured on all clusters to provide recovery of virtual machines in the event of an ESXi host failure. If an ESXi host fails, the virtual machines running on that server will go down but will be restarted on another host, typically within a minute. Although there would be a service interruption perceivable to users, the impact is minimized by the automatic restarting of these virtual machines on other ESXi hosts. When configuring a cluster for HA, there are a number of additional properties that require defining.

HA CLUSTER SETTING	CONFIGURATION VALUE
Host Monitoring	Enabled
Admission Control	Prevent virtual machines from being powered on if they violate availability…
Admission Control Policy	Percentage of resources reserved: CPU: 10% Memory: 10%
Default Virtual Machine Restart Priority	Medium
Host Isolation Response	Leave virtual machine powered on
Virtual Machine Monitoring	Enabled
Virtual Machine Monitoring Sensitivity	Medium
Heartbeat Datastores	Select any of the cluster's datastores

Table 14. HA Cluster Configuration Summary

Design Considerations

It is essential that implications of the different types of host isolation responses are understood. A host isolation response is triggered when network heartbeats are not received on a particular host. This might, for instance, be caused by a failed network card or physical switch ports. To avoid unnecessary downtime, VMware recommends using "Leave virtual machine powered on" as the host isolation response. In this scenario, NFS-based storage is used. In case both the management network and the storage network are isolated, one of the remaining hosts in the cluster might restart a virtual machine due to the fact that the file lock on NFS for this virtual machine has expired. When HA detects a split-brain scenario (two identical active virtual machines), it will disable the virtual machine that has lost the lock on the VMDK to resolve this. This is described in greater detail in *vSphere 5.0 High Availability Deployment Best Practices* (http://www.vmware.com/resources/techresources/10232).

Admission control will be enabled to guarantee the restart of virtual machines and that the 10 percent selected for both CPU and memory equals the N+1 requirement (r110) chosen by the customer. Using the percentage-based admission control policy is recommended because some virtual machines will have reservations configured due to the requirements of the applications running within.

Selection of the heartbeat datastores will be controlled by vCenter Server because it makes decisions based on the current infrastructure and because a reelection occurs when required.

Virtual machine monitoring will be enabled to mitigate any guest OS-level failures. It will restart a virtual machine when it has detected that the VMware Tools heartbeat has failed.

FT is not used because the availability requirements of 99.9 percent are fulfilled with HA and because all of the critical virtual machines require multiple virtual CPUs.

vSphere Distributed Resource Scheduler and VMware DPM

To continuously balance workloads evenly across available ESXi hosts, to maximize performance and scalability, DRS will be configured on all clusters. It works with vMotion to provide automated resource optimization, virtual machine placement and migration. When configuring a cluster for DRS, there are a number of additional properties that require defining. They are summarized in Table 15.

DRS CLUSTER SETTING	CONFIGURATION VALUE
DRS	Enabled
Automation Level	Fully Automated
Migration Threshold	Moderate (Default)
VMware DPM	Enabled
Automation Level	Fully Automated
Migration Threshold	Moderate (Default)
Enhanced vMotion Compatibility	Enabled
Swap File Location	Swap File Stored in the Same Directory as the Virtual Machine

Table 15. DRS Cluster Configuration Summary

The VMware DPM feature enables a DRS cluster to reduce its power consumption by powering hosts on and off, based on cluster resource utilization. It monitors the cumulative demand for memory and CPU resources of all virtual machines in the cluster and compares this to the total available resource capacity of all hosts in the cluster. If sufficient excess capacity is found, VMware DPM places one or more hosts in standby mode and powers them off after migrating their virtual machines to other hosts. Conversely, when capacity is deemed to be inadequate, DRS will bring hosts out of standby mode (powers them on) and, using vMotion, will migrate virtual machines to them. When making these calculations, VMware DPM not only factors in current demand but also any user-specified virtual machine resource reservations. VMware recommends enabling VMware DPM. It can use one of three power management protocols to bring a host out of standby mode: Intelligent Platform Management Interface (IPMI), Hewlett-Packard Integrated Lights-Out (iLO), or Wake-On-LAN (WOL). The configuration used for this case study, Dell PowerEdge R715 server, offers remote management capabilities that are fully IPMI 2.0 compliant. They will be configured appropriately to enable VMware DPM to place hosts in standby mode. More details can be found in the *vSphere 5.0 Resource Management Guide* (http://pubs.vmware.com/vsphere-50/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-50-resource-management-guide.pdf).

Enhanced vMotion Compatibility (EVC) simplifies vMotion compatibility issues across CPU generations. It automatically configures server CPUs with Intel Virtualization Technology FlexMigration or AMD-V Extended Migration technologies to be compatible with older servers. After EVC is enabled for a cluster in the vCenter inventory, all hosts in that cluster are configured to present identical CPU features and ensure CPU compatibility for vMotion. EVC can be enabled only if no virtual machine is active within the cluster. VMware recommends enabling EVC during creation of the ESXi cluster.

Resource Pools

During the gathering of initial requirements, it was indicated that all resources should be divided into separate pools (requirement r107) to prevent the various types of workloads from interfering with one another. As such, VMware recommends implementing DRS resource pools for compute resources. Three major types of workloads have been identified; a resource pool will be configured for each.

VMware recommends beginning with three lower-level resource pools than the cluster level: one resource pool for infrastructure (infra) servers with a normal priority; one resource pool for database servers with normal priority; and a resource pool for DMZ servers with normal priority. Currently, the split in number of virtual machines per pool is equal over time. VMware recommends setting custom shares values per resource pool, based on the relative priority over other resource pools and the number of virtual machines within the resource pool. VMware recommends recalculating the shares values on a regular basis to avoid a situation where virtual machines in a pool with "low priority" have more resources available during contention than virtual machines in a pool with "high priority." Such a scenario might exist when there is a large discrepancy in the number of virtual machines per pool.

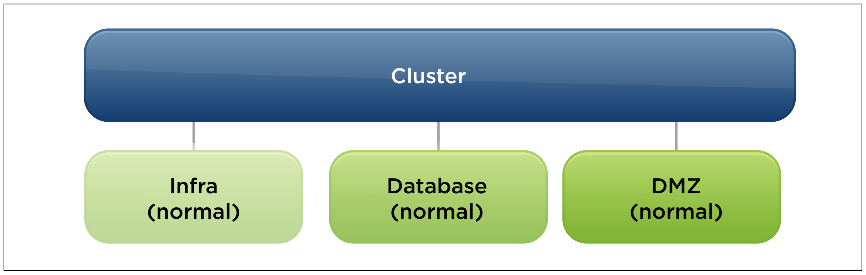

Figure 4. Initial Resource Pool Layout in Consolidated Cluster Design

Network Design

The network layer encompasses all network communications between virtual machines, vSphere management layer and the physical network. Key infrastructure qualities often associated with networking include availability, security and performance.

The network architecture will meet customer requirements with the following best practices:

- Separate networks for vSphere management, virtual machine connectivity, NFS and vMotion traffic
- Distributed switches with at least two active physical adaptor ports
- Redundancy at the physical switch level

vSphere 5.0 offers two different types of virtual switches, the vSphere standard switch (VSS) and the vSphere distributed switch (VDS). VSS must be configured on a per-host basis; VDS spans many ESXi hosts, aggregates networking to a centralized cluster and introduces new capabilities.

For simplicity and ease of management, VMware recommends using VDS in this design, in conjunction with multiple port groups with associated VLAN IDs to isolate ESXi management, vMotion, NFS and virtual machine traffic types. Leveraging network I/O control (NIOC) is recommended to prevent denial-of-service attacks and guarantee fairness during times of contention.

NOTE: The vSphere distributed switch is referred to as a "dvSwitch" in several places in the user interface.

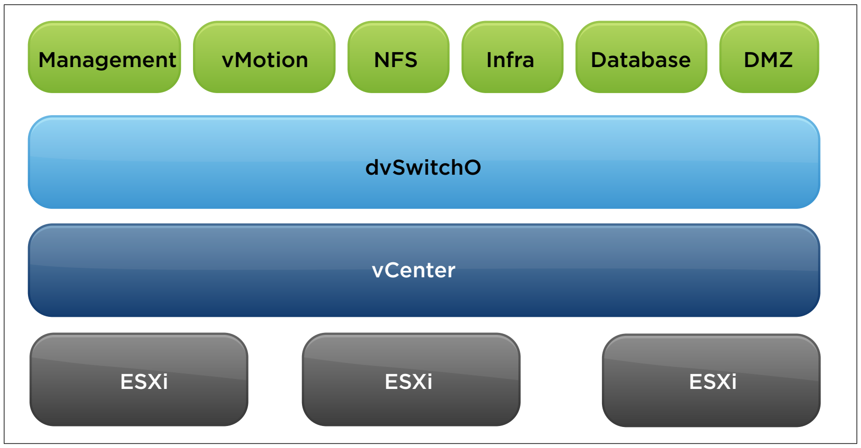

Figure 5. vSphere Distributed Switch

Physical Design

The current physical environment consists of a pair of Cisco 3750E 48-port switches in a stacked configuration per rack. It was indicated (constraint c108) that this was required in order to use the existing physical switching infrastructure. A top-of-rack approach has been taken to limit the use of copper between racks and within the datacenter. The current switch infrastructure has sufficient ports available to enable the implementation of the virtual infrastructure. Each top-of-rack Cisco 3750E pair of switches is connected to the core switch layer, which consists of a pair of Cisco 6500 switches and is managed by the central IT department.

vSphere Distributed Switch Infrastructure

For this case study, a VDS model has been proposed. VMware recommends creating a single VDS that manages all the various traffic streams. The VDS (dvSwitch) will be configured to use eight network adaptor ports (dvUplinks). All physical network switch ports connected to these adaptors should be configured as trunk ports, with Spanning Tree Protocol (STP) configured to PortFast or PortFast trunk depending on the configuration of the physical network switch port. The trunk ports are configured to pass traffic for all VLANs used by the dvSwitch as indicated in Table 16. No traffic-shaping policies will be in place, Load-based teaming "Route based on physical network adaptor load" will be configured for improved network traffic distribution between the physical network adaptors, and NIOC will be enabled.

VIRTUAL SWITCH	NUMBER OF PORTS	PHYSICAL NETWORK ADAPTOR CARDS	DVPORTGROUP (VLAN ID)
dvSwitch0	8	4	Management (10) vMotion (20) NFS (30) Infra (70) Database (75) DMZ (80)

Table 16. VDS Configuration

Table 17 presents the configured failover policies. For the dvSwitch and each of the dvPortgroups, the type of load balancing is specified. Unless otherwise stated, all network adaptors will be configured as active. Each dvPortgroup will overrule the dvSwitch configuration.

VIRTUAL SWITCH	DVPORTGROUP	NETWORK PORTS	LOAD BALANCING
dvSwitch0	Management network (10)	dvUplink0 (active) dvUplink1 (standby)	Route based on virtual port ID
dvSwitch0	vMotion (20)	dvUplink1 (active) dvUplink0 (standby)	Route based on virtual port ID
dvSwitch0	NFS (30) Jumbo frames (9,000 bytes)	vUplink2 dvUplink3	Route based on IP-Hash (*)
dvSwitch0	Infra (70)	dvUplink4, dvUplink5, dvUplink6, dvUplink7	Route based on physical network adaptor load
dvSwitch0	Database (75)	dvUplink4, dvUplink5, dvUplink6, dvUplink7	Route based on physical network adaptor load

* This requires an EtherChannel (Cisco's port link aggregation technology) configuration on the physical switch.

VIRTUAL SWITCH	DVPORTGROUP	NETWORK PORTS	LOAD BALANCING
dvSwitch0	DMZ (80)	dvUplink4, dvUplink5, dvUplink6, dvUplink7	Route based on physical network adaptor load

Table 17. vSphere Distributed Switch Configuration

The following diagram illustrates the dvSwitch configuration:

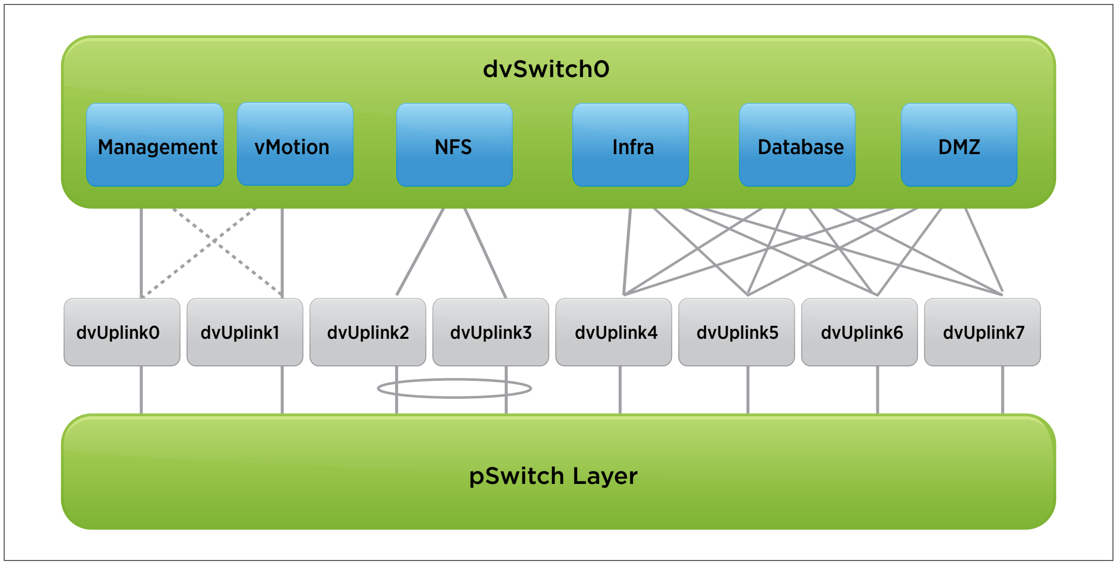

Figure 6. dSwitch Configuration

In this diagram, the physical switch layer has been simplified, consisting of a pair of stacked Cisco switches per rack, as described in the "Physical Design" section. For NFS traffic, as noted earlier, creating an EtherChannel is required to take advantage of enhanced load-balancing mechanisms. In this scenario, a cross-stack EtherChannel is recommended for resiliency. These ports are also required to be configured for jumbo frames (9,000 bytes), to reduce the number of units transmitted and possible processing overhead.

Design Considerations

In addition to the configuration of network connections, there are a number of additional properties regarding security, traffic shaping and network adaptor teaming that can be defined. VMware recommends changing MAC address changes and forged transmits from the default "accept" to "reject." Setting MAC address changes to "reject" at the dvSwitch level protects against MAC address spoofing. If the guest OS changes the MAC address of the adaptor to anything other than what is in the .vmx configuration file, all inbound frames are dropped. Setting forged transmits to "reject" at the dvSwitch level also protects against MAC address spoofing. Outbound frames with a source MAC address that is different from the one set on the adaptor are dropped.

The load-balancing mechanism used will be different per traffic type because each type of traffic has different requirements and constraints. Using an EtherChannel configuration for NFS traffic has been decided upon. This requires IP-Hash to be configured as the load-balancing mechanism used for this dvPortgroup per best practice documented in NetApp's *vSphere 5.0 Storage Best Practices* document tr-3749 (http://media.netapp.com/documents/tr-3749.pdf). More in-depth storage configuration details are provided in the "Storage" section.

STP is not supported on virtual switches, so no configuration is required on the dvSwitch. It is important to enable this protocol on the physical switches. STP ensures that there are no loops in the network. VMware recommends enabling PortFast on ESXi host–facing physical switch ports. With this setting, network convergence on these switch ports will take place quickly after the failure because the port will enter the STP forwarding state immediately, bypassing the listening and learning states. VMware also recommends using the Bridge Protocol Data Unit (BPDU) Guard feature to enforce the STP boundary. This configuration protects against any invalid device connection on the ESXi host–facing access switch ports. As mentioned earlier, dvSwitch doesn't support STP, so it doesn't send any BPDU frames to the switch port. However, if any BPDU is seen on these ESXi host–facing switch ports, the BPDU Guard feature puts that particular switch port in error-disabled state. The switch port is completely shut down, which prevents its affecting the STP topology.

In this scenario, management and vMotion traffic do not share dvUplinks with virtual machine traffic. It was decided not to share dvUplinks to prevent the scenario where virtual machine traffic or management/vMotion traffic is impacted by the other. Although NIOC will be implemented, only 1GbE network adaptors are used and NIOC manages resources on a dvUplink level. Considering the burstiness of vMotion traffic and the limitations of 1GbE links, a dedicated dvUplink is recommended.

Network I/O Control

The VDS (dvSwitch0) will be configured with NIOC enabled. After NIOC is enabled, traffic through that VDS is divided into infra, database and DMZ traffic network resource pools. NIOC will prioritize traffic only when there is contention and it is purely for virtual machine traffic, because other traffic streams do not share physical network adaptor ports.

The priority of the traffic from each of these network resource pools is defined by the physical adaptor shares and host limits for each network resource pool. All virtual machine traffic resource pools will be set to "normal." The resource pools are configured to ensure that each of the virtual machines and traffic streams receives the network resources it is entitled to. Each resource pool will have the same name as the dvPortgroup it will be associated with. The shares values specified in Table 18 show the relative importance of specific traffic types. NIOC ensures that each traffic type gets the allocated bandwidth during contention scenarios on the dvUplinks. By configuring equal shares, we are giving equal bandwidth to infra, database and DMZ traffic during contention scenarios. If providing more bandwidth to database traffic during contention is wanted, shares should be configured to "high."

NETWORK RESOURCE POOL	PHYSICAL ADAPTER SHARES	HOST LIMIT
Infra	Normal (50)	Unlimited
Database	Normal (50)	Unlimited
DMZ	Normal (50)	Unlimited

Table 18. Virtual Switch Port Groups and VLANs

Network I/O Settings Explanation

- **Host limits** – These are the upper limits of bandwidth that the network resource pool can use.
- **Physical adaptor shares** – Shares assigned to a network resource pool determine the total available bandwidth guaranteed to the traffic associated with that network resource pool.
  - **High** – This sets the shares for this resource pool to 100.
  - **Normal** – This sets the shares for this resource pool to 50.
  - **Low** – This sets the shares for this resource pool to 25.
  - **Custom** – This is a specific number of shares, from 1 to 100, for this network resource pool.

Storage Design

The most common aspect of datastore sizing discussed today is the limit that should be implemented regarding the number of virtual machines per datastore. In the design for this environment, NetApp is the storage vendor that has been selected by the customer (constraint c103) and NFS is the preferred protocol. With respect to limiting the number of virtual machines per datastore, NetApp has not made any specific recommendations. Datastore sizing is not an easy task and is unique to each individual organization. As such, it is often neglected or forgotten.

In this scenario, a backup solution based on (two) LTO-4 drives is used. The theoretical transfer rate of the LTO-4 drive is 120MB per second, with a theoretical limit of 864GB per hour. This means that based on the defined recovery time objective (RTO) of 8 hours, the maximum size for a given datastore is 13,824GB (theoretical limit for two drives). Analysis has shown that an average of 50GB is required per virtual machine, which would result in a maximum of approximately 276 virtual machines.

In this scenario, the average storage I/O requirement for a potential virtual machine has been measured at approximately 58 IOPS, including the RAID-DP write penalty. Considering the maximum of 40 virtual machines per datastore, this would require each datastore to be capable of providing a minimum of 2,400 IOPS. It is estimated that a 15K RPM disk can generate 175 IOPS, and a 10K RPM disk can drive approximately 125 IOPS. The minimal configuration for each tier from a performance perspective is 14 drives per 15K RPM RAID group and 20 drives per 10K RPM RAID group. The storage configuration will be supplied with 512GB of Flash cache, enabling spare IOPS capacity for read tasks and thereby decreasing the number of disks required to meet the IO profile of this environment. To ensure that SLA requirements are met, our calculation is based on a worst-case scenario.

Physical Design

The chosen storage array to be used is the NetApp FAS 3240. The following tables provide detailed specifications for the arrays intended for use in this vSphere design, based on the data recovery and performance requirements.

ATTRIBUTE	SPECIFICATION
Storage Type	Network-attached storage/NFS
Array Type	NetApp FAS 3240
Firmware	ONTAP 8.1
Flash Cache	512GB
Disk Shelves	4 shelves (2x DS4243, 2x DS2246) 48 disks – 450GB – 15K RPM 48 disks – 600GB – 10K RPM
Number of Switches Number of Ports per Host per Switch	2 (redundant) 4
Frame Size	Jumbo frame (9,000 bytes) – end to end

Table 19. Storage Array Specifications

The suggested aggregate and RAID group configuration presented in Table 20 is based on NetApp's recommendations and meets our requirements. Each aggregate will hold two 24-disk RAID groups, with each a dedicated spare.

TIER SPECIFICATION	AGGREGATE	RAID GROUP	DRIVES
Tier 1	Aggregate 1	RAID Group 1	23 + 1
Tier 1	Aggregate 1	RAID Group 2	23 + 1
Tier 2	Aggregate 2	RAID Group 1	23 + 1
Tier 2	Aggregate 2	RAID Group 2	23 + 1

Table 20. Datastore Specifications

The following diagram depicts the chosen design with two RAID 23+1 RAID groups per aggregate. Each RAID group in Tier 1 can support approximately 75 virtual machines from an IOPS perspective when taking only disk IOPS into account. In Tier 2, this is approximately 52 virtual machines. Because each RAID group can hold more than 100 virtual machines, it has been decided that 512GB of Flash cache will be added to the NetApp FAS 3240 to optimize storage performance.

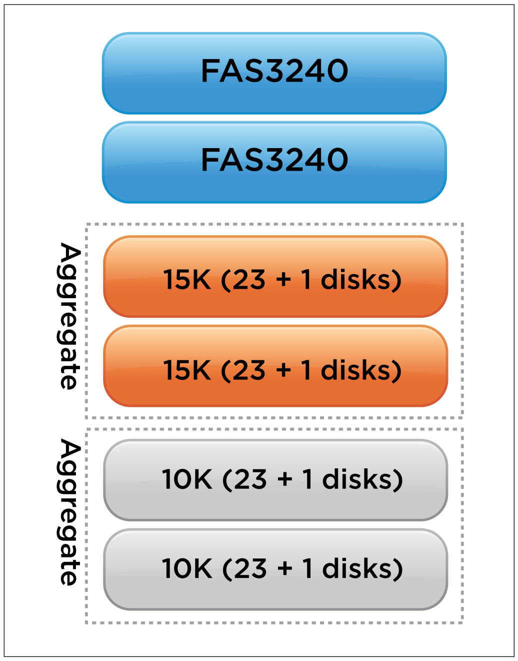

Figure 7. Logical Storage Design

Design and Implementation Considerations

Performance and availability along with reduction of operational effort are key drivers for this design. In this scenario, the decision was made to reduce the number of disks compared to the IOPS requirements per virtual machine by leveraging cache modules. This decision is unique for every organization.

Duplicating all paths within the current storage system environment will ensure that any single points of failure have been removed. Each host will have, at a minimum, two paths to the storage system. Load balancing and configuration of these is done through the NetApp Virtual Storage Console (VSC) and Rapid Cloning Utility (RCU). VSC will also be used to configure the Path Selection Policy (PSP) and the Storage Array Type Plug-in (SATP) according to NetApp best practices.

Deduplication will be enabled on the Fibre Channel 10K volumes, Tier 2. Migrating virtual machines between volumes can impact the effectiveness of the deduplication process, and results can vary.

Per NetApp recommendations, an EtherChannel/IP-Hash configuration is used, leveraging multiple 1GbE network ports and various IP addresses to ensure optimal load balancing.

VMware recommends using jumbo frames for NFS traffic to reduce the number of units transmitted and the possible processing overhead. For the best possible configuration, following storage and network vendor guidelines is recommended.

The average I/O requirements for a potential virtual machine have been measured at approximately 42 IOPS. The read/write ratio for analyzed candidates was typically 62 percent read to 38 percent write. When using RAID-DP (constraint C103), this results in the following IOPS requirements per virtual machine, taking a RAID penalty of 2 into account for RAID-DP because parity must be written to two disks for each write.

% READ	% WRITE	AVERAGE IOPS	RAID PENALTY	I/O PROFILE
62%	38%	42 IOPS	2 IOPS	58 IOPS

Table 21. IOPS Requirements per Virtual Machine

```
IP Profile = (TOTAL IOPS × %READ) + ((TOTAL IOPS × %WRITE) × RAID Penalty)
(42 x 62%) + ((42 x 38%) x 2)
(26.04) + ((15.96) x 2)
26.04 + 31.92 = 57.96
```

The results of this profile have been discussed with the NetApp consultant to determine an optimal strategy from a performance perspective, weighted against the constraints from a recovery-time objective. This will be described in the “Physical Design” paragraph of the “Storage Design” section.

Profile-Driven Storage

Managing datastores and matching the SLA requirements of virtual machines with the appropriate datastore can be a challenging and cumbersome task. One of the focus areas of this architecture case study is to reduce the amount of operational effort associated with management of the virtual infrastructure and provisioning of virtual machines. vSphere 5.0 introduces Profile-Driven Storage, which enables rapid and intelligent placement of virtual machines based on SLA, availability, performance or other requirements and provided storage capabilities. It has been determined that two storage tiers will be created, each with different characteristics, as presented in Table 22.

TIER	PERFORMANCE	CHARACTERISTICS	RAID
Gold	15K RPM	n/a	RAID-DP
Silver	10K RPM	Deduplication	RAID-DP

Table 22. Virtual Machine Storage Profiles (Profile-Driven Storage) Specifications

VMware recommends using Profile-Driven Storage to create two virtual machine storage profiles, representing each of the offered tiers. These virtual machine storage profiles can be used during provisioning, cloning and VMware vSphere® Storage vMotion to ensure that only those datastores that are compliant with the virtual machine storage profile are presented.

vSphere Storage DRS

vSphere Storage DRS (SDRS) is a new feature introduced in vSphere 5.0. It provides smart virtual machine placement and load-balancing mechanisms based on I/O and space capacity. It helps decrease the operational effort associated with the provisioning of virtual machines and monitoring of the storage environment. VMware recommends implementing Storage DRS.

Two tiers of storage, each with different performance characteristics as shown in Table 22, will be provided to internal customers. Each of these tiers should be grouped in datastore clusters to prevent a degradation of service when a Storage DRS migration recommendation is applied (for example, when being moved from "Gold" to "Silver," the maximum achievable IOPS likely is lower).

VMware advises turning on automatic load balancing for the entire "Gold-001" datastore cluster after having become comfortable with the manual recommendations made by SDRS. VMware also advises keeping the "Silver-001" datastore cluster configured as manual, due to the fact that the datastores that form this cluster are deduplicated natively by the array. Because deduplication is a scheduled process, it is possible that more storage will temporarily be consumed after the Storage DRS–recommended migration than before. Applying recommendations for the "Silver-001" datastore cluster is advised during off-peak hours, with a scheduled deduplication process to run afterward.

DATASTORE CLUSTER	I/O METRIC	AUTOMATION LEVEL	DATASTORES
Gold-001	Enabled	Fully automated/ manual	2
Silver-001	Enabled	Manual	2

Table 23. Storage DRS Specifications

Design and Implementation Considerations

By default, the Storage DRS latency threshold is set to 15ms. Depending on the workload, types of disks and SLA requirements, modification of this value might be required. When I/O load balancing is enabled, Storage DRS automatically enables Storage I/O Control (SIOC).

The Storage DRS out-of-space avoidance threshold is configured to 80 percent by default, so Storage DRS will be invoked if more than 80 percent of a datastore is consumed. Storage DRS will then determine whether recommendations must be made—and, if so, what they should be—based on growth patterns, risks and benefits. VMware recommends using the default value for out-of-space avoidance.

NOTE: During the initial conversation, there was discussion that array-based replication might be implemented in the future. VMware emphasizes that if array-based replication is implemented, the migration recommendations made by Storage DRS should be applied during off-peak hours and that the workload being migrated will be temporarily unprotected.

Storage I/O Control

vSphere 5.0 extends SIOC to provide cluster-wide I/O shares and limits for NFS datastores, so no single virtual machine will be able to create a bottleneck in any environment regardless of the type of shared storage used. SIOC automatically throttles a virtual machine that is consuming a disparate amount of I/O bandwidth when the configured latency threshold has been exceeded. This enables other virtual machines using the same datastore to receive their fair share of I/O. When Storage DRS I/O metric is enabled, SIOC is enabled by default with a latency of 30ms as the threshold.

To prevent a single virtual machine from creating a bottleneck for all virtual machines in the environment, VMware recommends leaving SIOC enabled.

Design Considerations

When the Storage DRS latency is changed to a different value, it might make sense in some scenarios to also increase or decrease the latency value specified for SIOC. At the time of writing, this is a manual process and is not automatically done by Storage DRS. If the storage environment is expanded with Flash-based drives, reevaluation of the SIOC latency threshold is recommended. Whereas SIOC mitigates short-term I/O bottlenecks, Storage DRS mitigates medium- to long-term I/O bottlenecks. Therefore, configuring Storage DRS to at least half the specified SIOC latency threshold is recommended.

vSphere Storage APIs

At the time of writing, NetApp has not made its vSphere Storage APIs for Array Integration (VAAI) plug-in or the vSphere Storage APIs for Storage Awareness (VASA) provider available for NAS. VMware recommends tracking the status of the VAAI plug-in and the VASA storage provider to evaluate these when they are made available.

The VAAI plug-in for NAS will enable the creation of a thick-provisioned disk and for the off-loading of the cloning process to the NAS device. This will speed up the provisioning process. The VASA storage provider will enable storage characteristics such as RAID type, replication, deduplication and more to be surfaced to vCenter. VASA will make the provision process and the creation of virtual machine storage profiles and datastore clusters easier by providing the necessary details to make decisions based on performance and availability.

Security Design

VMware offers the most robust and secure virtualization platform available through vSphere 5.0 and vShield App. In this section, we will discuss the various design considerations and recommendations regarding vCenter, ESXi and virtual machines. The security considerations are divided into implementation and operational tables where applicable.

During the gathering of requirements, it was indicated that physical separation was not desired due to the inflexibility and inefficiency of this solution. It was decided to leverage vShield App to enable security on a vNIC boundary instead of only through logical separation with VLANs or on the outside of the virtual infrastructure.

Leveraging vShield App will decrease operational effort associated with securing the environment, due to the ability of applying policies to the various virtual objects. vShield App will also enable the isolation of traffic through the configuration of different security zones as specified in requirement r103.

vShield App includes the following components:

- VMware® vShield Manager™ – The centralized network management component of vShield. It is installed as a virtual appliance on any ESXi host in the vCenter environment. It centrally manages the firewall rules and distributes them to vShield App appliances across the environment.
- vShield App – Installs as a hypervisor module and firewall service virtual appliance on each ESXi host. It provides firewalls between virtual machines by placing a firewall filter on every virtual network adaptor.

- A vSphere plug-in option that can register vShield Manager as a VMware vSphere® Client™ plug-in, which enables configuration of most vShield options from the vSphere Client. vShield Manager includes a Web client that enables the security team to manage security aspects without the need to log in to vCenter.

VMware vShield App

To overcome the current challenges of providing a secure virtual infrastructure, a new vShield App–based solution will be implemented. vShield App enables the creation of secure zones with associated security policies based on various vCenter objects. To provide a secure, scalable and easy-to-manage environment, VMware recommends creating security groups based on resource pools for this environment.

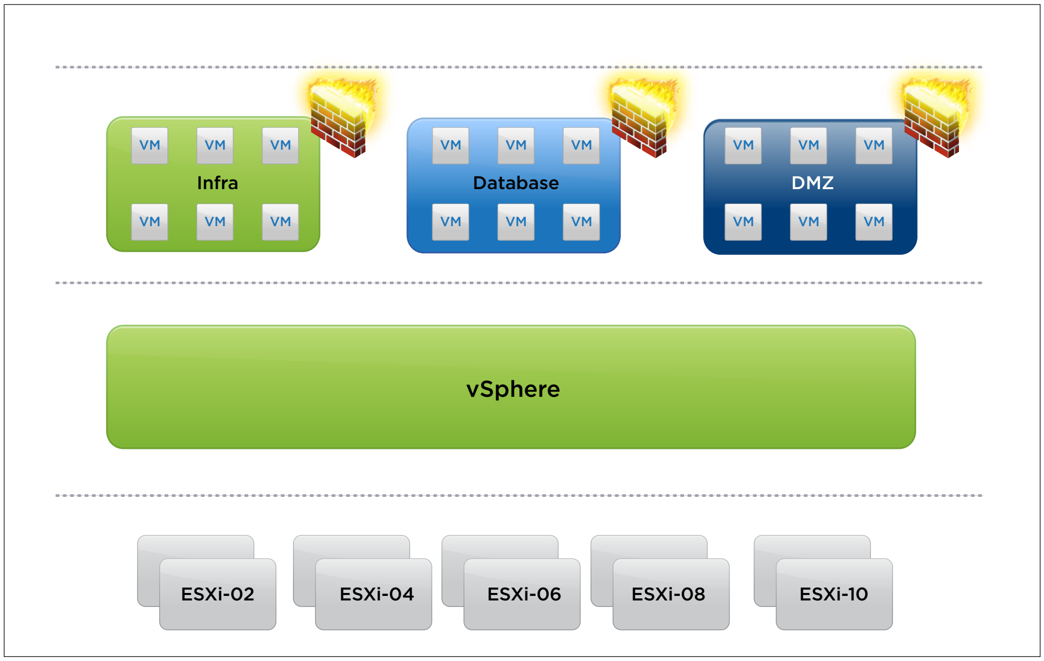

Figure 8. Security Groups Based on Resource Pools

This approach, with security zones based on resource pools, enables us to block certain types of traffic between the pools without the need to specify IP addresses. Each virtual machine within the resource pools will inherit the specified rules/policies from its parent. All traffic, unless explicitly specified, is blocked. If a virtual machine is migrated between resource pools, its security policy will change based on the policy applied to its destination resource pool.

vShield App is a hypervisor-based, vNIC-level technology that enables the inspection of all traffic leaving and entering a virtual machine and the virtual environment in general. Flow monitoring decodes this traffic into identifiable protocols and applications. This feature provides the ability to observe network activity between virtual machines, to define and refine firewall policies. VMware recommends leveraging this feature to create the appropriate policies based on actual network activity. vShield App consists of a vShield App virtual machine and a vShield module per ESXi host.

VMware recommends creating three security zones based on the current network and compute topology. Applying security policies on a resource-pool level is recommended. This will enable applying policy changes to a complete zone in a single click as well as applying policies for new virtual machines by simply dragging them into the correct resource pool.

Table 24 describes the minimal zones and port configuration based on our current understanding of the infrastructure. Currently, only Internet – DMZ – Database is secured. Further investigation must result in a more restrictive policy and configuration of the security zones.

SOURCE	DESTINATION	PORT	ACTION
Datacenter			Allow all
External	Infra RP		Allow all
Infra RP	Database RP		Allow all
DMZ RP	Infra RP		Disallow all
Infra RP	DMZ RP		Disallow all
External	Infra RP	TCP/UDP 902 TCP 903 TCP 80 TCP/UDP 88 TCP 443 TCP 8000 TCP 8080 (vCenter)	Allow
Infra	External	TCP/UDP 902 TCP 903 TCP 80 TCP/UDP 88 TCP 443 TCP 8000 TCP 8080 (vCenter)	Allow
DMZ RP	Database RP	TCP 5432 (PostgreSQL)	Allow
External	DMZ	TCP 80	Allow
External	DMZ	TCP 443	Allow

Table 24. Security Zones

Datacenter-level rules have higher priority than cluster-level rules. Because rules are inherited from parents, it should be ensured that no conflicts exist within the policy hierarchy.

It was decided to use an “allow all” approach and leverage flow monitoring to observe network traffic between the various machines and define firewall policies. This applies to both the infra and database resource pools but not to the DMZ resource pool. The “allow all” approach will be used for three weeks to enable extensive testing and planning, after which the policy will be changed to a “deny all” approach to maximize security.

NOTE: More details about ports used by vCenter Server and ESXi can be found in VMware knowledge base article 1012382. Depending on the features and functionality used, opening additional ports might be required.

VMware vShield Deployment Considerations

When deploying vShield Manager, a 3GB memory reservation is configured on the vShield Manager virtual machine. This memory reservation can impact HA admission control. In this design, we have decided to use the percentage-based admission control policy to avoid issues caused by a large memory reservation with the default admission control policy. When using the “host failures tolerated” admission control policy, the reservation might skew the slot size used by this algorithm.

Using NTP timekeeping is recommended for both the vShield Manager and vShield App appliances and syslog for vShield App, as described in the *VMware vShield Administration Guide*.

vShield is capable of enforcing security on different layers. Within vCenter, multiple levels can be leveraged to define and apply security policies, such as datacenter, resource pools, vApps and virtual machines. In this scenario, resource pools are used.

NOTE: Because vCenter Server is deployed as a virtual machine on the protected ESXi hosts, it is possible to lock it out. To avoid all risks, vCenter can be deployed on a physical server or on a nonprotected ESXi host.

For this design, various best-practices documents have been leveraged, VMware recommends reading the following for additional background information:

- *VMware vShield App Protecting Virtual SAP Environments*
  http://www.vmware.com/resources/techresources/10213
- *VMware vShield App Design Guide*
  http://www.vmware.com/resources/techresources/10185
- *Technology Foundations of VMware vShield*
  http://www.vmware.com/resources/techresources/10187

VMware vShield Manager Availability

VMware recommends enabling HA virtual machine monitoring on the HA cluster to increase availability of the vShield components. VMware also recommends regularly backing up the vShield Manager data, which can include system configuration, events and audit log tables. Backups should be saved to a remote location that must be accessible by the vShield Manager and shipped offsite.

vSphere Security

VMware has broken down the *vSphere Security Hardening Guide*. Only sections pertinent to this design are included.

VMware recommends applying all of the following guidelines. These recommendations refer to the *VMware vSphere 4.1 Security Hardening Guide* (http://www.vmware.com/resources/techresources/10198). At the time of writing, the *VMware vSphere 5.0 Security Hardening Guide* was not available.

Periodically validating compliance through the use of the VMware Compliance Checker for vSphere (http://www.vmware.com/products/datacenter-virtualization/vsphere-compliance-checker/overview.html) or custom VMware vSphere® PowerCLI scripts is recommended.

vCenter Server Security

vCenter Server is the cornerstone of the VMware cloud infrastructure. Securing it adequately is key to providing a highly available management solution. All recommendations pertinent to this design have been listed in the tables on the following page.

CODE	NAME
VSH03	Provide Windows system protection on the vCenter Server host.
VSH05	Install vCenter Server using a service account instead of a built-in Windows account.
VSC01	Do not use default self-signed certificates.
VSC03	Restrict access to SSL certificates.
VSC05	Restrict network access to the vCenter Server system.
VSC06	Block access to ports not being used by vCenter.
VSC07	Disable managed object browser.
VSC08	Disable Web access.
VSC09	Disable datastore browser.
VSD01	Use least privileges for the vCenter Server database user.

Table 25. Security Hardening vCenter – Implementation

CODE	NAME
VSH01	Maintain supported OS, database and hardware for vCenter.
VSH02	Keep the vCenter Server system properly patched.
VSH04	Avoid user login to the vCenter Server system.
VSH06	Restrict usage of vSphere administrator privilege.
VSC02	Monitor access to SSL certificates.
VSC04	Always verify SSL certificates.
VCL01	Restrict the use of Linux-based clients.
VCL02	Verify the integrity of the vSphere Client.

Table 26. Security Hardening vCenter – Operational

Encryption and Security Certificates

VMware recommends changing the default encryption and security certificates as recommended in Table 27. Security Hardening ESXi – Implementation," code HCM01. Host certificate checking is enabled by default. SSL certificates are used to encrypt network traffic. ESXi automatically uses certificates that are created as part of the installation process and are stored on the host. These certificates are unique and make it possible to begin using the server. However, they are not verifiable and are not signed by a trusted, well-known certificate authority (CA).

This design recommends installing new certificates that are signed by a valid internal certificate authority. This will further secure the virtual infrastructure to receive the full benefit of certificate checking. For more details, refer to the *vSphere 5.0 Security Guide*.

ESXi Security

The ESXi architecture is designed to ensure security of the ESXi system as a whole. ESXi is not a general-purpose OS. It has very specific tasks and therefore very specific software modules, so there is no arbitrary code running and it is not susceptible to common threats. It has a very small footprint and therefore provides a much smaller attack surface and requires fewer patches.

The ESXi architecture consists of the following three major components:

- The virtualization layer (VMkernel) – This layer is designed to run virtual machines. The VMkernel controls the hardware of the host and is responsible for scheduling its hardware resources for the virtual machines.
- The virtual machines – By design, all virtual machines are isolated from each other, so multiple virtual machines can run securely while sharing hardware resources. Because the VMkernel mediates access to the hardware resources, and all access to the hardware is through the VMkernel, virtual machines cannot circumvent this isolation.
- The virtual networking layer – ESXi relies on the virtual networking layer to support communication between virtual machines and their users. The ESXi system is designed so that it is possible to connect some virtual machines to an internal network, some to an external network and some to both—all on the same host and all with access only to the appropriate machines.

The following are core components of the ESXi security strategy:

- Enable lockdown mode

Lockdown mode restricts which users are authorized to use the following host services: VIM, CIM, local Tech Support Mode (TSM), remote Tech Support Mode (SSH) and the Direct Console User Interface (DCUI) service. By default, lockdown mode is disabled. When lockdown mode is enabled, no users other than vpxuser have authentication permissions, nor can they perform operations against the host directly. Lockdown mode forces all operations to be performed through vCenter Server.

- Security banner

The default security banner will be added via the following procedure:

1. Log in to the host from the vSphere Client.

2. From the **Configuration** tab, select **Advanced Settings**.

3. From the **Advanced Settings** window, select **Annotations**.

4. Enter a security message.

- Syslog server will be used to maintain log files for data retention compliance.

CODE	NAME
HCM01	Do not use default self-signed certificates for ESXi communication.
HCM02	Disable managed object browser.
HCM03	Ensure that ESXi is configured to encrypt all sessions.
HLG01	Configure remote syslog.
HLG02	Configure persistent logging.
HLG03	Configure NTP time synchronization.
HMT01	Control access by CIM-based hardware-monitoring tools.
HMT02	Ensure proper SNMP configuration.
HCN02	Enable lockdown mode to restrict root access.
HCN04	Disable Tech Support Mode.
HCP01	Use a directory service for authentication.
NAR01	Ensure that vSphere management traffic is on a restricted network.
NAR02	Ensure that vMotion traffic is isolated.
NAR04	Maintain strict control of access to management network.
NCN03	Ensure that the “MAC address change” policy is set to “reject.”
NCN04	Ensure that the “forged transmits” policy is set to “reject.”
NCN05	Ensure that the “promiscuous mode” policy is set to “reject.”

Table 27. Security Hardening ESXi – Implementation

CODE	NAME
HST03	Mask and zone SAN resources appropriately.
HIN01	Verify integrity of software before installation.
HMT03	Establish and maintain configuration file integrity.
HCN01	Ensure that only authorized users have access to the DCUI.
HCN03	Avoid adding the root user to local groups.
NCN06	Ensure that port groups are not configured to the value of the native VLAN.
NCN07	Ensure that port groups are not configured to VLAN 4095 except for virtual guest tagging.
NCN08	Ensure that port groups are not configured to VLAN values reserved by upstream physical switches.
NCN10	Ensure that port groups are configured with a clear network label.
NCN11	Ensure that all dvSwitches have a clear network label.
NCN12	Fully document all VLANs used on dvSwitches.
NCN13	Ensure that only authorized administrators have access to virtual networking components.
NPN01	Ensure that physical switch ports are configured with STP disabled.

CODE	NAME
NPN02	Ensure that the *non-negotiate* option is configured for trunk links between external physical switches and virtual switches in VST mode.
NPN03	Ensure that VLAN trunk links are connected only to physical switch ports that function as trunk links.

Table 28. Security Hardening ESXi – Operational

Directory Service Authentication

For direct access via the vSphere Client for advanced configuration and troubleshooting of an ESXi host, ESXi supports directory service authentication to a Microsoft Active Directory (AD) domain. Rather than creating shared privileged accounts for local host authentication, AD users and groups can be assigned to ESXi local groups and authenticated against the domain. Although ESXi cannot use AD to define user accounts, it can use AD to authenticate users. In other words, individual user accounts can be defined on the ESXi host, and passwords and account status can be managed with AD.

Lockdown mode should be applied to the ESXi hosts. It must first be disabled to log in locally. Because lockdown mode is enabled, AD authentication will not work until lockdown mode is disabled. Disabling it to enable troubleshooting locally on an ESXi host must be included in the operational procedures.

Virtual Machine Security

Virtual machine security is provided through several different components of the VMware cloud infrastructure. In addition to using vShield App, applying the following guidelines is recommended. They should be applied after P2V migrations, where and when applicable, on each of the created templates.

CODE	NAME
VMX02	Prevent other users from spying on administrator remote consoles.
VMX10	Ensure that unauthorized devices are not connected.
VMX11	Prevent unauthorized removal, connection and modification of devices.
VMX12	Disable virtual machine–to–virtual machine communication through VMCI.
VMX20	Limit virtual machine log file size and number.
VMX21	Limit informational messages from the virtual machine to the VMX file.
VMX24	Disable certain unexposed features.
VMP03	Use templates to deploy virtual machines whenever possible.
VMP05	Minimize use of the virtual machine console.

Table 29. Security Hardening Virtual Machine – Operational

Operations

Operational efficiency is an important factor in this architecture case study. We will leverage several vSphere 5.0 Enterprise Plus components to decrease operational cost and increase operational efficiency. Similar to the building-block approach taken for this architecture, it is recommended that all operational procedures be standardized to guarantee a consistent experience for customers and decrease operational costs.

Process Consistency

During the initial workshops, it was indicated that automation was done mainly through Microsoft Windows PowerShell. VMware recommends using vSphere PowerCLI to automate common virtualization tasks that will help reduce operational expenditure and enable the organization to focus on more strategic projects. Alternatively, VMware vCenter Orchestrator™ can be used to create predefined workflows for standard procedures. Automating the management of a vSphere environment can help save time and increase virtual machine–to–administrator ratio. Automation with vSphere PowerCLI enables faster deployment and more efficient maintenance of the virtual infrastructure. Its structured model removes the risk of human error involved when performing repetitive tasks, ensuring operational consistency. vSphere PowerCLI is recommended to automate the following tasks:

- Virtual machine deployment and configuration
- Host configuration
- Reporting on common issues (snapshot usage and so on)
- Capacity reporting
- vCenter configuration
- Patching of hosts
- Upgrade and auditing of VMware Tools for virtual machines
- Auditing all areas of the virtual infrastructure

Host Compliance

Host Profiles provide the ability to capture a “blueprint” of an ESXi host configuration and ensure compliance with this profile. They also offer the ability to remediate hosts that are not compliant, ensuring a consistent configuration within the vSphere environment. VMware recommends taking advantage of Host Profiles to ensure compliance with a standardized ESXi host build. Host Profiles can be applied at datacenter level and at cluster level. VMware recommends applying Host Profiles at the cluster level. Although only a single cluster is used, this will enable both a scale-up and a scale-out approach from a cluster perspective.

The assumption has been made that ESXi hosts within a cluster will be identically configured. Exporting the Host Profile after the creation to a location outside of the virtual infrastructure is recommended.

NOTE: When a host profile is exported, administrator passwords are not exported. This is a security measure. You will be prompted to reenter the values for the password after the profile is imported and the password is applied to a host.

Virtual Machine Storage Compliance

Virtual machine storage profiles enable you to check compliance of all virtual machines and the associated virtual disks in a single pane of glass. As a result, managing storage tiers, provisioning, migrating, cloning virtual machines and correct virtual machine placement in vSphere deployments have become more efficient and user friendly. This eliminates the need to maintain complex and tedious spreadsheets and to manually validate compliance during every migration or creation of a virtual machine or virtual disk. VMware recommends leveraging virtual machine storage profiles (Table 23. Virtual Machine Storage Profiles (Profile-Driven Storage) Specifications) to ensure that virtual machines reside on the correct storage tier.

Virtual Machine Provisioning

Because this design focuses on providing a running platform for virtual machines that will be produced using a P2V process, it is not easy to define a standard virtual machine configuration model. To reduce the complexity and standardize the environment, three different virtual machine configurations will be offered, and each can be provisioned on two different tiers of storage. Each virtual machine will be placed in a resource group, for both compute and networking, to ensure that each group will receive its fair share. During the P2V process, each virtual machine will be sized according to these standards. During the course of four months, each virtual machine will be monitored and, where applicable, reduced in size to further maximize consolidation ratio and cost savings.

The following table provides the virtual machine sizes that VMware recommends for use during the P2V process and the provisioning of new virtual machines.

TIER	CONFIGURATION	TYPICAL USE CASE
Bronze	1 vCPU, 2GB memory	DNS, AD, print servers
Silver	2 vCPU, 4GB memory	Application servers
Gold	2 vCPU, 8GB memory	Database servers

Table 30. Virtual Machine Tiering Model

Analysis of the current estate has shown that 90 percent of all deployed servers will fit this model. In scenarios where more memory or vCPUs are required, the request will be evaluated on a case-by-case basis.

Patch/Version Management

Update Manager will be implemented as a component of this solution for monitoring and managing the patch levels of the ESXi hosts only, with virtual machines being updated in line with the existing processes for physical servers. (Guest OS updates are not supported with Update Manager 5.0.)

Update Manager will be installed on separate virtual machines and will be configured to patch/upgrade the ESXi hosts. VMware Tools will be installed within the virtual machines and will upgrade them to a higher virtual hardware version when required.

The following table summarizes the Update Manager configuration settings that will be applied in this case:

ATTRIBUTE	SPECIFICATION
Patch Download Sources	**Download vSphere ESXi and ESX patches** Deselect **Download ESX 3.x patches** and **Download virtual appliance upgrades**
Shared Repository	n/a
Proxy Settings	TBD
Patch Download Schedule	Every day at 03:00 a.m. CET
Email Notification	TBD
Update Manager Baselines to Leverage	Critical and noncritical ESX/ESXi host patches VMware Tools upgrade to match host
Virtual Machine Settings	n/a

ATTRIBUTE	SPECIFICATION
ESX Host/Cluster Settings	Host maintenance mode failure: **Retry** Retry interval: **5 Minutes** Number of retries: 3 Temporarily disable: **VMware DPM**
vApp settings	Select **Enable smart reboot after remediation**

Table 31. vCenter Update Manager Server Configuration

vCenter Server and vSphere Client Updates

VMware recommends routinely checking for and evaluating new vCenter Server and vSphere Client updates. These will be installed in a timely fashion following release and proper testing. vSphere Client updates should be manually installed whenever vCenter Server is updated. Using conflicting versions of the vSphere Client and vCenter Server can cause unexpected results. The vSphere Client will automatically check for and download an update if it exists when connecting to an updated vCenter Server.

Monitoring

To implement an effective monitoring solution for a vSphere infrastructure, there is a requirement to monitor the health of the ESXi hosts, vCenter Server, storage infrastructure and network infrastructure. VMware recommends keeping monitoring relatively simple by leveraging the monitoring and alarms available by default in vCenter. VMware recommends monitoring the following alarms at a minimum because they will directly impact the stability and availability of the virtual infrastructure. In this default configuration, there would be a requirement for an administrator to log in to vCenter to view alerts, so it is recommended that an email alert be configured for these specific events:

- Cannot connect to storage
- Cannot find HA master
- Datastore cluster is out of space
- Datastore usage on disk
- Health status monitoring
- Host CPU usage
- Host error
- Host hardware
- Host memory status
- Host memory usage
- Network uplink redundancy degraded
- Network uplink redundancy lost
- Virtual machine error
- HA insufficient failover resources
- HA cannot find master
- HA failover in progress
- HA host HA status
- HA virtual machine monitoring error
- HA virtual machine monitoring action
- HA failover failed

VMware Syslog Collector

Currently no syslog server is in place. VMware recommends installing the VMware Syslog Collector because it integrates with vCenter Server and is relatively easy to install and use. Each ESXi host can subsequently be configured to use the configured syslog server as output for its logs. Increasing the size of the log files is recommended before rotation to 5MB (default 2MB). Increasing the number of rotations to a minimum of 16 (default 8) is also recommended.

VMware ESXi Dump Collector

VMware ESXi™ Dump Collector is most useful for datacenters where ESXi hosts are configured using the Auto Deploy process. It can also be installed for ESXi hosts that have local storage as an additional location where VMkernel memory dumps can be redirected when critical failures occur. For this environment, one of the NFS datastores will be used as the scratch location. As such, it is not required to implement ESXi Dump Collector.

Storage and Network Monitoring

All storage and networking infrastructure will be monitored in line with existing monitoring processes. However, two key alarms (“Cannot connect to network” and “Cannot connect to storage”) within the default vCenter alarms for monitoring ESXi hosts extend to monitoring storage and networking. VMware recommends monitoring these alarms because they provide early indication of issues with ESXi host storage and networking.

Conclusion

In this VMware Cloud Infrastructure Case Study, based on real-world requirements and constraints, we have shown an example of how to implement VMware vSphere 5.0 Enterprise Plus in combination with VMware vShield App. Using a building-block approach, this environment can easily be horizontally and vertically scaled. The building-block approach ensures consistency and reduces operational costs. A similar approach for operational tasks, using standardized and scripted operational procedures, guarantees a decrease in administrative overhead. Leveraging vSphere 5.0 features such as network I/O control, Profile-Driven Storage and Storage DRS simplifies management and monitoring and also ensures that service-level agreements can be met.

Security is crucial in every environment. But in many environments, hardening of the respective layers is often neglected. VMware vCenter and VMware ESXi are hardened using various best practices, and VMware vShield App helps protect the respective workloads from internal and external threats.

About the Author

Duncan Epping is principal architect in the Technical Marketing group at VMware and is focused on cloud infrastructure management architecture. Previously, he worked at Oracle and multiple consultancy companies, where he had more than 10 years of experience designing and developing infrastructures and deployment best practices. He was among the first VMware certified design experts (VCDX 007). He is the co-author of several books, including best sellers *vSphere 5.0 Clustering Technical Deepdive* and *VMware vSphere 4.1 HA and DRS Technical Deepdive*. He is the owner and main author of the leading virtualization blog, yellow-bricks.com.

- Follow Duncan Epping’s blogs at http://www.yellow-bricks.com and http://blogs.vmware.com/vSphere.
- Follow Duncan Epping on Twitter: @DuncanYB.

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com