Cloud Infrastructure Architecture Case Study

vSphere 5.0 and vShield App 5.0

TECHNICAL MARKETING DOCUMENTATION
v 1.0 December 2011

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	Consulting Architect – Professional Services	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 5 technologies that together expand the capabilities and value customers can realize from a virtualized infrastructure. CIS is designed to help organizations build more intelligent virtual infrastructures, by enabling highly virtualized environments with the automation, self-service and security capabilities customers need to deploy business-critical applications, respond to business needs more quickly, and move to a secure cloud model. CIS is centered around the VMware vSphere platform as it the foundation to pursuing any type of cloud infrastructure. In addition to vSphere, the CIS also includes VMware vShield, VMware vCenter Site Recovery Manager, VMware vCloud Director, and VMware vCenter Operations.

The VMware Cloud Infrastructure Architecture Case Study Series was developed to provide an understanding of the different components of the VMware Cloud Infrastructure Suite. The goal is to explain how the different components of the CIS can be used in specific scenarios. These scenarios are based on real-world customer examples and as such will 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. These considerations and decisions are based on a combination of VMware best practices and specific business requirements and goals to facilitate the requirements of this case study. 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 being 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.

Once 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:

- Financial Institution
- Project initiated by local Insurance business unit
- Vision for the future of IT is to adopt a virtualization first approach and increase agility and availability of service offerings
- This first “foundation infrastructure” to be located at primary site
- 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** – Identifies where an assumption has been made in the absence of factual data

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

Requirements, assumptions and constraints

In this case study the primary requirement for this architecture is lowering 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 sections below.

Case Requirements, assumptions and constraints

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

Table 1. Customer Requirements

ID	Requirement
r101	Increasing agility / flexibility while decreasing cost of doing business
r102	Availability of services defined as 99.9% during core business hours
r103	Security compliancy requires network isolation for specific workloads from other services
r104	Minimal workload deployment time
r105	Separate Management VLAN needs to be used for Management Traffic
r106	Environment should be scalable to allow for future expansion (minimum 1 year, 20% estimated)
r107	Should be able to guarantee resources to groups of workloads as part of internal SLAs
r108	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

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

Table 2. Design Constraints

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 de-duplicated
c105	All volumes, except for database volumes, will be RAID-6
c106	The environment should have 20% 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

Use cases

This design is targeted at the following use cases:

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

Conceptual Architecture Overview Diagram

The VMware Cloud Infrastructure aims to reduce operational overhead and lower Total Cost of Ownership (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 four major pillars:

- Compute
- Networking
- Storage

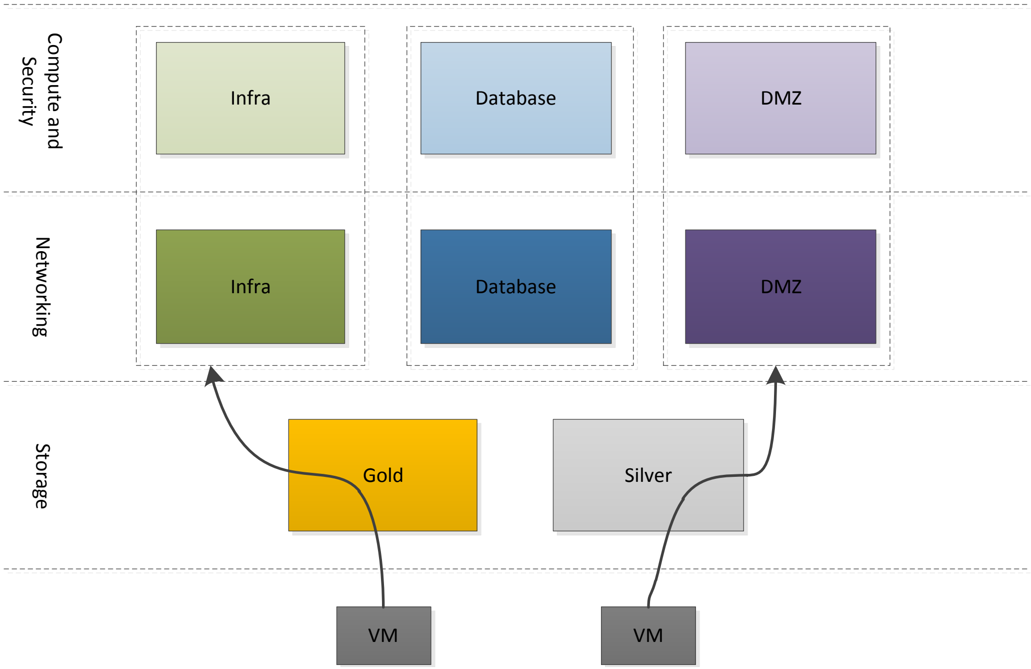
- Security

Each of the pillars will be carved in to 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. In order meet these requirements 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 NIC, 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 will allow for a fast and safe deployment of new workloads.

Figure 1 Conceptual overview



Each of the different types of the pillars are carved up in to different pools for each of the respective workload types. A virtual machine, or vApp (a logical container for 1 or more virtual machines), will be deployed in one of the three different compute resource pools after which a specific Networking & Security pool will be selected as well as a Storage pool (often referred to as tier). Compute, Network & Security pool types are currently defined based on the different type of workloads which this virtual infrastructure will host. In the future, additional blocks may be added based on the requirements of the 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 VMware vSphere 5 environment. By using this approach a consistent experience can be guaranteed for internal customers. This design will allow for both horizontal and vertical scaling when required.

Sizing is based on the evaluation of the virtualization candidates. This section will describe the required amount of hosts based on the analysis and the requirement of this study to have at least 1 year of growth factored in. (Requirement R106)

Workload estimations

To determine the required number of ESXi hosts needed to consolidate the x86 virtualization candidates, the performance and utilization has 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. The following table 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 sufficient resources are available during short resource bursts.

Table 3 ESXi host CPU and memory Requirements

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 % (1000MHz)
Total CPU resources for all VMs 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 VMs at peak (no memory sharing)	*275,000MB*
Assumed memory sharing benefit when virtualized	25% (*)
Total RAM for all VMs at peak (memory sharing)	***206,000 MB***

*** Note:** The estimated savings from memory sharing has been intentionally kept low due to the fact that most Guest operating systems will be 64-bit and as such large memory pages will be used. For more details please read KB 1021095 and 1021896.

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 R715 (Constraint C101) that are pertinent to this analysis and case study.

Table 4 ESXi host CPU Logical Design Specifications

Attribute	Specification
Number of CPUs (sockets) per host	2
Number of cores per CPU (AMD)	8

Attribute	Specification
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**

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

Table 5 ESXi host Memory Logical Design Specifications

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

Using the high-level CPU and memory specifications detailed above 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 perspective of CPU and RAM to satisfy the resource requirements. It should be noted that 20% head room for both CPU and RAM has been taken in to consideration.

Table 6 ESXi host requirements

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

With an anticipated growth rate of 20% (Requirement R105) the following table shows the required number of host for this environment.

Table 7 ESXi host required for Project

# of ESXi hosts Required	Percentage of Growth factored in	Availability Requirements	# of ESXi hosts Required
7	20%	N+1	10

Network and Storage

In many cases the network bandwidth profile of VMs is overlooked and a general assumption is made regarding the number of NICs required fulfilling the combined bandwidth requirements for a given number of VMs. The analysis has shown that the expected average network bandwidth requirement is **4.21Mbps** based on an average of 20 VMs per ESXi host. In this case study network bandwidth is not a limiting factor, as a minimum of 8 NICs will be used per ESXi host of which 2 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 VMs 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 while others will be heavy on writes, some are heavy on sequential access and others are heavy on random access. In this case the average I/O requirements for a potential VM has been measured at approximately 42 IOPS.

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

Table 8~~9~~ Virtual Machine Storage Capacity Profile

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

In this case the average storage requirement for a potential VM has been measured at approximately 50.10GB (9.36GB C:\ and 40.74GB Other Drive).

Host Design

VMware 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 and in this instance the R715 has been selected as the hardware platform for use with VMware® 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 specification and configuration for this hardware platform is detailed in the following table.

Table 9~~10~~ VMware ESXi host platform specifications

Attribute	Specification
Vendor	Dell
Model	R715
Number of CPU Sockets	2
Number of CPU Cores	8
Processor speed	2.3 GHz
Memory	96 GB
Number of NIC ports	8
NIC Vendor(s)	Broadcom / Intel
NIC Model(s)	2 x Broadcom Dual Port 5709C Onboard 2 x Intel Dual Port Gigabit ET
NIC Speed	Gigabit
Installation destination	Dual SD Card
ESX Server Version	ESXi Server 5.0, Build: Latest

Design / Configuration Considerations

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

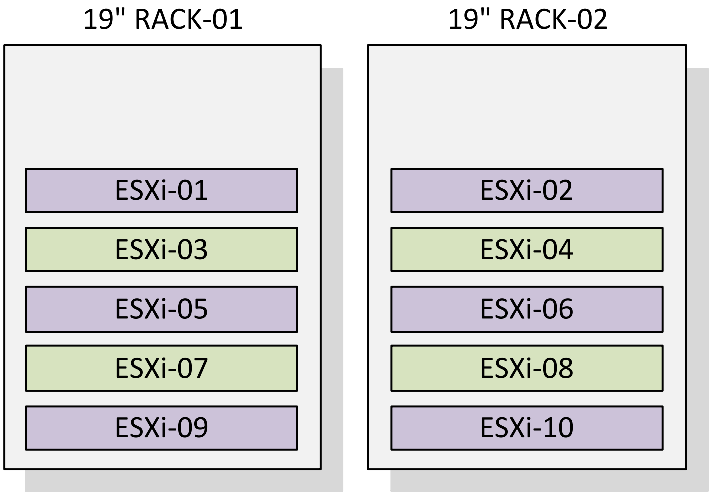
Network Time Protocol (NTP) must be configured on each ESXi host and should be configured to share the same time source as the vCenter Server to ensure consistency of the overall vSphere solution.

Currently VMware ESXi 5.0 offers 4 different solutions (Local disk, USB/SD, Boot From SAN, Stateless) to boot ESXi. During the workshops the customer indicated that they may want to use stateless sometime in the future, and to therefore leave that option open and with minimal cost associated. VMware recommends deploying ESXi on SD cards as this will allow for a cost effective migration to stateless when desired.

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

VMware recommends using two racks at a minimum for the 10 hosts and to layer the hosts across the racks as depicted in the diagram below. When wiring a rack for redundant power, the racks should have two PDU (Power Distribution Units), each connected to at least separate legs of a distribution panel, or entirely separate panels. This is assuming 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



vCenter Server Design

VMware recommends deploying VMware vCenter$^{\text{TM}}$ Server (vCenter Server) using a virtual machine as opposed to a standalone physical server. This enables the customer to leverage the benefits available when running in a virtual machine, such as vSphere HA which will protect the vCenter Server VM in the event of hardware failure.

The specification and configuration for the vCenter Server VM 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. The vCenter Server sizing has taken a 20% growth for the first year in to account. VMware also recommends separating vCenter Update Manager from vCenter Server for flexibility during maintenance.

Table 10~~11~~ VMware vCenter Server platform specifications

Attribute	Specification
Vendor	VMware Virtual Machine
Model	Virtual Hardware 8
Number of vCPUs	2
Memory	6GB
Number of Local Drives	2
Total Useable Capacity	20GB (C:\) and 40GB (E:\)
Operating System	MS Windows 2008 – 64 Bit

Design and implementation considerations

When installing VMware vCenter Server 5.0 multiple different options are presented, it is recommended to install all separate components including the Syslog and Dump functionality.

vCenter Update Manager Design

VMware vCenter 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 vCenter Update Manager in a separate virtual machine to allow for future expansion to leverage the benefits which vSphere 5.0 provides like vSphere HA. The specification and configuration for the vCenter Update Manager VM are detailed in the following table and are based on the recommendations provided in the vCenter Update Manager installation documentation.

Table 11~~12~~ VMware vCenter Update Manager Server platform specifications

Attribute	Specification
Vendor	VMware Virtual Machine
Model	Virtual Hardware 8
Number of vCPUs	2
Memory	2GB
Number of Local Drives	2
Total Useable Capacity	20GB (C:\) and 40GB (D:\)
Operating System	MS Windows 2008 – 64 Bit

vCenter Server and vCenter Update Manager Database

vCenter Server and VMware vCenter^{TM} Update Manager (VUM) require access to a database. During the installation process it is possible to install Microsoft (MS) SQL Server 2008 Express, however this is only supported for small deployments not exceeding 5 ESXi hosts and 50 VMs, meaning there is a requirement to use an alternative supported database. The following tables summarizes the configuration requirements for the vCenter and VUM databases.

Table 12~~13~~ vCenter and vCenter Update Manager Databases Specifications

Attribute	Specification
Vendor and version	MS 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 vCenter Update Manager	150MB Initial + 60-70MB per month
Estimated disk utilization vCenter Update Manager	1050MB Initial + 500MB per month

In order to estimate the size requirements of the vCenter Server and VUM databases the VMware vCenter Server 4.1 Database Sizing Calculator for MS SQL Server (at time of writing only vSphere 4.1 vCenter database calculator was available) and vCenter Update Manager 5.0 Sizing Estimator tools have been used.

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.

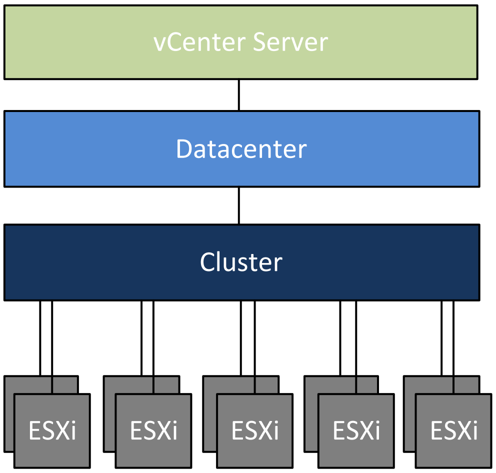
Cluster

Within a Datacenter, ESXi hosts are typically grouped in to clusters to provide a platform for different groups of virtual machines requiring different network and storage requirements. Furthermore grouping ESXi hosts in to Clusters facilitates the use of technologies such as vMotion, vSphere Distributed Resource Scheduler (DRS), vSphere Distributed Power Management (DPM), vSphere High Availability (HA) and vSphere Fault Tolerance (FT). VMware recommends creating a single cluster with all 10 hosts as multiple clusters would result in a higher overhead from an HA perspective. Meaning that with N+1 redundancy with two 5 nodes clusters only 8 nodes (5 – 1) worth of resource can be used while in a single 10-node cluster 9 hosts (10 – 1) worth of resource can be used. This approach also reduces complexity in your environment and avoids the associated operational effort with managing multiple objects.

Table 13~~14~~ vSphere Cluster Summary

Type	Configuration Value
Number of hosts	10
vSphere DRS	Enabled
vSphere HA	Enabled

Figure 3 Datacenter and Cluster overview



vSphere High Availability

vSphere HA will be configured on all Clusters to provide recovery of VMs in the event of an ESXi host failure. If an ESXi host fails, the VMs running on that server will go down, but will be restarted on another host typically within a minute. While 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.

Table 14~~15~~ HA Cluster Configuration Summary

HA Cluster Setting	Configuration Value
Host Monitoring	Enabled
Admission Control	Prevent VMs from being powered on if they violate availability…
Admission Control Policy	Percentage of resources reserved CPU: 10% Memory: 10%
Default VM Restart Priority	Medium
Host Isolation Response	Leave VM powered on
Virtual Machine Monitoring	Enabled
Virtual Machine Monitoring Sensitivity	Medium
Heartbeat Datastores	Select any of the cluster's datastores

Design considerations:

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

Admission control will be enabled to guarantee the restart of virtual machines and the percentage selected of 10% for both CPU and Memory equals the n+1 requirement (r110) as desired by the customer. It is recommended to use the Percentage based admission control policy, as 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 as it makes the decisions based on the current infrastructure and a re-election occurs when required.

VM Monitoring will be enabled to mitigate any Guest OS level failures. VM Monitoring will restart a virtual machine when it has detected the VMware Tools heartbeat has failed.

vSphere Fault Tolerance (FT) is not used as the availability requirements of 99.9% are fulfilled with vSphere HA, and the critical Virtual machines all require multiple virtual CPU.

vSphere DRS and DPM

vSphere DRS will be configured on all Clusters to continuously balance workloads evenly across available ESXi hosts in order to maximize performance and scalability. vSphere DRS works with vMotion to provide automated resource optimization, VM placement and migration. When configuring a Cluster for vSphere DRS there are a number of additional properties that require defining.

Table 15~~16~~ DRS Cluster Configuration Summary

DRS Cluster Setting	Configuration Value
vSphere **DRS**	Enabled
Automation Level	Fully Automated
Migration Threshold	Moderate (Default)
vSphere **DPM**	Enabled
Automation Level	Fully Automated
Migration Threshold	Moderate (Default)
VMware **EVC** (Enhance vMotion Compatibility)	Enabled
Swap file location	Store the swapfile in the same directory as the virtual machine

The vSphere Distributed Power Management (DPM) feature allows a DRS cluster to reduce its power consumption by powering hosts on and off based on cluster resource utilisation. vSphere DPM monitors the cumulative demand of all virtual machines in the cluster for memory and CPU resources and compares this to the total available resource capacity of all hosts in the cluster. If sufficient excess capacity is found, vSphere 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 migrates virtual machines, using vMotion, to them. When making these calculations, VMware DPM considers not only§ current demand, but it also honours any user-specified virtual machine resource reservations. VMware recommends enabling Distributed Power Management. vSphere DPM 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 R715, offers remote management capabilities which are fully IPMI 2.0 compliant. These will be configured appropriately to allow for 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)

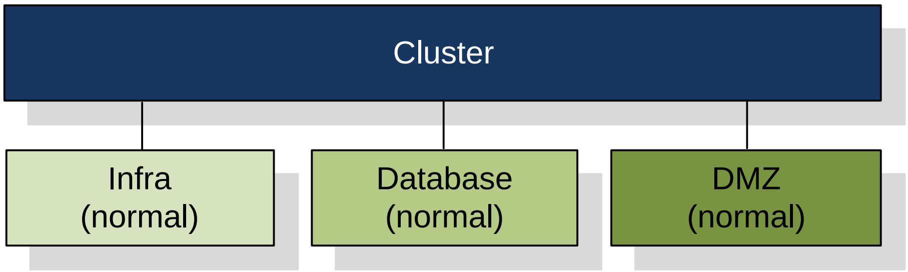
Enhanced vMotion Compatibility (EVC) simplifies vMotion compatibility issues across CPU generations. EVC automatically configures server CPUs with Intel Flex Migration or AMD-V Extended Migration technologies to be compatible with older servers. After EVC is enabled for a cluster in the vCenter inventory,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

Resource Pools

During the initial requirements gather it was indicated that all resources should be divided in separate pools (requirement r107) to avoid the different types of workloads interfering with each other. As such VMware recommends implementing vSphere DRS Resource Pools for compute resources. Three major types of workloads have been identified and for each a resource pool will be configured.

VMware recommends to start with three resource pools below the cluster level, one resource pool for Infra (workloads supporting management of the infrastructure) 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 VMs per pool is equal, over time VMware recommends setting custom share values per resource pool based on the relative priority over other resource pools and the amount of virtual machines within the resource pool. VMware recommends re-calculating the shares value on a regular basis to avoid the situation where virtual machines in a pool with "low priority" will have more resources available during contention than virtual machines in a pool with "high priority". A scenario like this could exist when there is a large discrepancy in the number of virtual machines per pool.

Figure 4 Initial resource pool layout in consolidate 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.

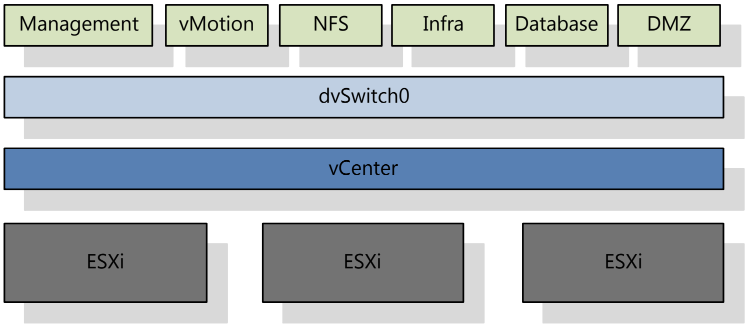
To address customer requirements, the following best practices, the network architecture will meet these requirements:

- Separate networks for vSphere management, VM connectivity, NFS and vMotion traffic
- Distributed Switches with at least 2 active physical adapter ports
- Redundancy at the physical switch level

vSphere 5.0 offers two different type of virtual switches, the vSphere Standard Switch (VSS) and the vSphere Distributed Switch (VDS). The vSphere Standard Switch needs to be configured on a per host basis where the vSphere Distributed Switch spans many ESXi hosts, aggregates networking to a centralised cluster and introduces new capabilities.

In this design VMware recommends using VDS for simplicity and ease of management. The VDS will be used in conjunction with multiple portgroups with associated VLAN IDs to isolate ESXi management traffic, vMotion, NFS traffic and virtual machine traffic. It is recommended to leverage Network I/O Control to avoid 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 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 allow for the implementation of the virtual infrastructure. Each top of rack Cisco 3750E pair is connected to the core switch layer which consists of a pair of Cisco 6500 and is managed by the central IT department.

vSphere Distributed Switch Infrastructure

For this case study a vSphere Distributed Switch model has been proposed. VMware recommends creating a single VDS which manages all the different traffic streams. The VDS (dvswitch) will be configured to use eight NIC ports (dvUplinks). All physical network switch ports connected to these adapters should be configured as trunk ports with spanning tree 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 the table below. No traffic shaping policies will be in place, Load-based teaming “Route based on physical NIC load” will be configured for improved network traffic distribution between the physical NICs and Network I/O Control will be enabled.

Table 16~~17~~ VDS Configuration

Virtual Switch	# Ports	Physical Network Adapters Cards	dvPortgroup (VLAN ID)
dvSwitch0	8	4	Management (10) vMotion (20) NFS (30) Infra (70) Database (75) DMZ (80)

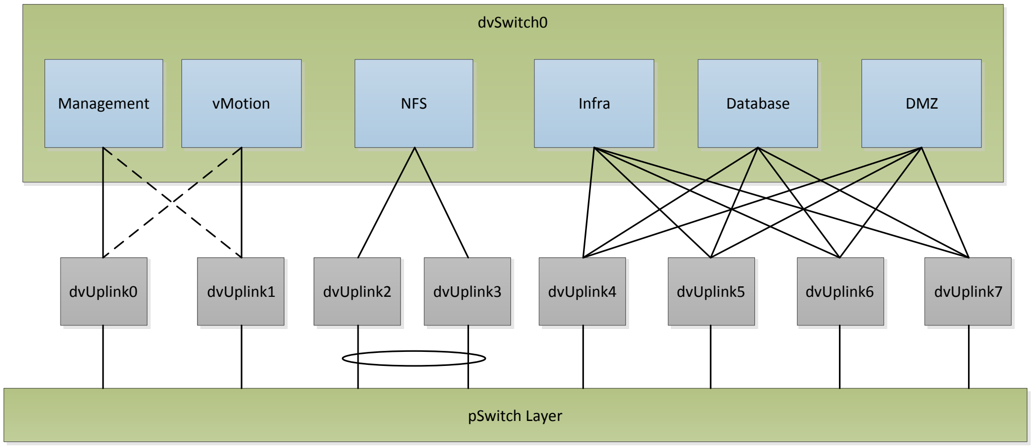
The following table explains the configured failover policies. For the dvSwitch and each of the dvPortgroups the type of load balancing is specified. Unless specifically specified, all NICs will be configured as active. Each dvPortgroup will overrule the dvSwitch configuration.

Table 17~~18~~ vSphere Distributed Switch 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 (9000)	dvUplink2 dvUplink3	Route based on IP-Hash (*)
dvSwitch0	Infra (70)	dvUplink4, dvUplink5, dvUplink6, dvUplink7	Route based on Physical NIC LOAD
dvSwitch0	Database (75)	dvUplink4, dvUplink5, dvUplink6, dvUplink7	Route based on Physical NIC LOAD
dvSwitch0	DMZ (80)	dvUplink4, dvUplink5, dvUplink6, dvUplink7	Route based on Physical NIC LOAD

*** Note:** This requires an etherchannel (Cisco’s port link aggregation technology) configuration on the physical switch.

The following diagram illustrates the dvSwitch configuration.



In this diagram the physical switch layer has been simplified. The physical switch layer consists of a pair of stacked Cisco switches per rack as described in the physical design section. For the NFS traffic as noted earlier it is required to create an etherchannel to take advantage of enhanced load balancing mechanisms. In this scenario for resiliency purposes a cross-stack etherchannel is recommended. These ports also have the requirement to be configured for jumbo frames (9000) to reduce the number of units transmitted and the 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 NIC 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 Operating System (OS) changes the MAC address of the adapter 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 protects against MAC address spoofing. Outbound frames with a source MAC address that is different from the one set on the adapter are dropped.

The load balancing mechanism used will be different per traffic type as each type of traffic has different requirements and constraints. It has been decided to use an etherchannel configuration for NFS traffic, 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 described in the storage section.

Spanning Tree protocol is not supported on virtual switch and thus no configuration is required on the dvSwitch. It is important to enable this protocol on the physical switches. STP makes sure 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 happen fast after the failure because the port will enter the spanning tree forwarding state immediately, bypassing the listening and learning states. VMware also recommends using "BPDU guard" to enforce STP boundary. This configuration protects from any invalid device connection on the ESXi host facing access switch ports. As mentioned earlier, dvSwitch doesn't support Spanning Tree protocol and thus doesn't send any Bridge Protocol Data Unit (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 and prevents affecting the Spanning Tree Topology.

In this scenario Management and vMotion traffic do not share dvUplinks with Virtual Machine traffic. It was decided not to share dvUplinks to avoid the scenario where Virtual Machine traffic or Management/vMotion traffic is impacted by either. Although Network I/O Control will be implemented, only 1GbE NICs are used and Network I/O Control 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 vSphere Distributed Switch (dvSwitch0) will be configured with Network I/O Control (NetIOC) enabled. After NetIOC is enabled, traffic through that distributed switch is divided into the following network resource pools: Infra, Database and DMZ traffic. It should be noted that NetIOC will only prioritize traffic when there is contention and that it is purely for virtual machine traffic as other traffic streams do not share physical NIC ports.

The priority of the traffic from each of these network resource pools is defined by the physical adapter 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 each of the virtual machine's 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 the following table show the relative importance of specific traffic type, and NetIOC ensures that during contention scenarios on the dvUplinks each traffic type gets the allocated bandwidth. By configuring equal shares we are giving equal bandwidth to Infra, Database, and DMZ traffic during contention scenarios. If it is desired to provide more bandwidth to Database traffic during contention then shares should be configured to high.

Table 18~~19~~. Virtual Switch Port Groups and VLANs

Network Resource Pool	Physical Adapter Shares	Host Limit
Infra	Normal (50)	Unlimited
Database	Normal (50)	Unlimited
DMZ	Normal (50)	Unlimited

Network I/O Settings Explanation

- **Host Limits**. Host limits are the upper limit of bandwidth that the network resource pool can use.
- **Physical Adapter Shares**. Shares assigned to a network resource pool determine the total available bandwidth guaranteed to the traffic associated with that network resource pool.
  - **High**. Sets the shares for this resource pool to 100.
  - **Normal**. Sets the shares for this resource pool to 50.
  - **Low**. Sets the shares for this resource pool to 25.
  - **Custom**. 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 what limit should be implemented regarding the number of VMs per Datastore. In the design for this environment, NetApp is the storage vendor that has been selected by the customer (constraint c103) and NFS the preferred protocol. With respect to limiting the number of VMs 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 LTO-4 drives (two) is used. The theoretical transfer rate of the LTO-4 drive is 120MB/s and a theoretical limit of 864GB per hour. This means that based on the defined RTO of 8 hours the maximum size for a given datastore is 13824GB (two drives * theoretical limit). Analysis has shown 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 VM 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 2400 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. It should be noted that the storage configuration will be supplied with 512GB of Flash Cache allowing for spare IOps capacity for read tasks and as such decreasing the amount of disks required to meet the IO profile of this environment. In order to ensure 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.

Table 19~~20~~ Storage Array Specifications

Attribute	Specification
Storage type	Network Attached Storage / NFS
Array type	FAS 3240
Firmware	ONTAP 8.1
Flash Cache	512GB
Disk shelves	4 shelves (2 x DS4243, 2 x DS2246) 48 disks - 450GB - 15k RPM 48 disks – 600GB – 10k RPM
Number of switches	2 (redundant)
Number of ports per host per switch	4
Frame Size	Jumbo Frame (9000) – end to end

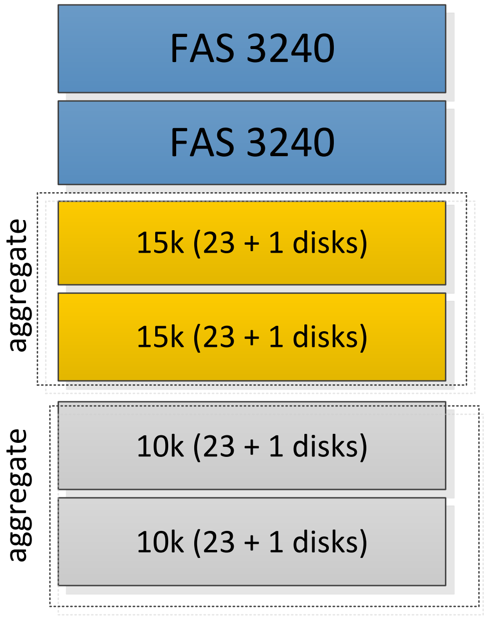
The suggested Aggregate and RAID group configuration as described in the table below is based on NetApp’s recommendations and meets our requirements. Each Aggregate will hold two 24 Disk RAID Groups with each a dedicated spare.

Table 20~~21~~ Datastore Specifications

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

The following diagram depicts the chosen design with two RAID 23+1 Raid Groups per Aggregate. Each RAID Group in Tier 1 could cater for approximately 75 virtual machines from an IOps perspective when only taking disk IOps in to account. In Tier 2 this is approximately 52 VMs. As each RAID Group can hold over 100 VMs and it has been decided that 512GB of Flash Cache will be added to the FAS 3240 to optimize storage performance.

Figure 6 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 lower the amount of disks compared to the IOps requirements per VM by leveraging cache modules. This decision is unique for every organization.

Duplicating all paths within the current storage system environment will ensure any single points of failure have been removed. Each host will have at a minimum two paths to the storage system and load balancing and configuration of these is done through NetApp’s Virtual Storage Console and the 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’s best practices.

Deduplication will be enabled on the FC10k volumes, Tier 2. It should be noted that migrating virtual machines between volumes can impact the effectiveness of the deduplication process and results may vary.

Per NetApp’s 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. It is recommended to follow the guidelines from both the storage and network vendor around the optimal configuration.

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

Table 21 IOps requirements per VM

% Read	% Write	Avg IOps	RAID Penalty	IO Profile
62%	38%	42 IOps	2 IOps	58 IOps

```
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 and weighted against the constraints from a recovery time objective. This will be described in the Physical Design paragraph of the Storage section.

Profile-Driven Storage

Managing datastores and matching the SLA requirements of virtual machines with the appropriate datastore can be a challenging and a cumbersome task. One of the focus areas of this Architecture Case Study is reducing 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 will allow for 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 with each different characteristics as displayed in the following table.

Table 22 VM Storage Profiles (Profile-Driven Storage) specifications

Tier	Performance	Characteristics	RAID
Gold	15k RPM	n/a	RAID-DP
Silver	10k RPM	deduplication	RAID-DP

VMware recommends using Profile-Driven Storage to create two VM Storage Profiles which represent each of the offered tiers. These VM Storage Profiles can be used during provisioning, cloning and Storage vMotion to ensure only those datastores that are compliant with the VM Storage Profile are presented.

Storage DRS

Storage DRS (SDRS) is a new feature introduced in vSphere 5.0 providing smart virtual machine placement and load balancing mechanisms based on I/O and space capacity. Storage DRS will help 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 will be provided to the internal customers, each with different performance characteristics as shown in Table 22~~Table 22~~. Each of these tiers should be grouped in Datastore Clusters to avoid 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 is likely lower.)

VMware recommends turning on automatic load balancing for all the Gold-001 Datastore Cluster after having become comfortable with the manual recommendations made by SDRS. VMware recommends keeping the Datastore Cluster “Silver-001” configured as manual due to the fact that the datastores which form this Datastore Cluster are deduplicated natively by the array. After applying a Storage DRS recommendation the possible impact will be that temporarily more storage is consumed after the migration than before, as the deduplication process is a scheduled process. It is recommended to apply recommendations for the Datastore Cluster “Silver-001” during off-peak hours and to schedule the deduplication process to run afterwards.

Table 23 Storage DRS Specifications

Datastore Cluster	I/O Metric	Automation level	Datastores
Gold-001	enabled	Fully Automated / Manual	2
Silver-001	enabled	Manual	2

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 it could be required to modify this value. It should be noted that when I/O Load Balancing is enabled Storage DRS automatically enables Storage I/O Control.

The Storage DRS out-of-space avoidance threshold is configured to 80% by default. Meaning that if more than 80% of a datastore is consumed Storage DRS will be invoked. Storage DRS will then decide based on growth patterns, risk and benefits if recommendations need to be made and what these should be. VMware recommends using the default value for out-of-space avoidance.

During the initial conversation it was discussed that array based replication might be implemented in the future. VMware wants to stress 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 Storage I/O Control (SIOC) to provide cluster-wide I/O shares and limits for NFS datastores. This means that 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 is to allow other virtual machines using the same datastore to receive their fair share of I/O. When Storage DRS I/O Metric is enabled Storage I/O Control is enabled by default with a latency of 30ms as the threshold.

In order to avoid a single virtual machine creating a bottleneck for all virtual machines in the environment VMware recommends leaving Storage I/O Control enabled.

Design Considerations

When the Storage DRS latency is changed to a different value it could make sense in some scenarios to also increase or decrease the latency value specified for Storage I/O Control. It should be noted that this is a manual process and is not automatically done by Storage DRS at the point of writing. If the storage environment is expanded with Flash based drives it is recommended to re-evaluate the SIOC latency threshold. It should be noted that SIOC mitigates short term I/O bottlenecks where Storage DRS mitigates medium to long term I/O bottlenecks and as such it is recommended to configure Storage DRS to at least half of the specified SIOC latency threshold.

vSphere Storage APIs

At the time of writing NetApp has not made their vSphere Storage APIs for Array Integration (VAAI) plugin or the vSphere Storage APIs for Storage Awareness (VASA) provider available for NAS. VMware recommends tracking the availability of the VAAI plugin and the VASA Storage Provider to evaluate these when made available.

The VAAI plugin for NAS will allow for the creation of a thick-provisioned disks and for the offloading of the cloning process to the NAS device. This will speed up the provisioning process The VASA Storage Provider will allow storage characteristics like RAID type, replication, deduplication and more to be surfaced to vCenter. VASA will make the provision process and the creation of VM 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 with regards to vCenter, ESXi and virtual machines. The security considerations are divided into implementation and operational tables where applicable.

During the requirements gathering 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 allow for security on a vNIC boundary instead of only through logical separation with VLANs or on the outside of the virtual infrastructure.

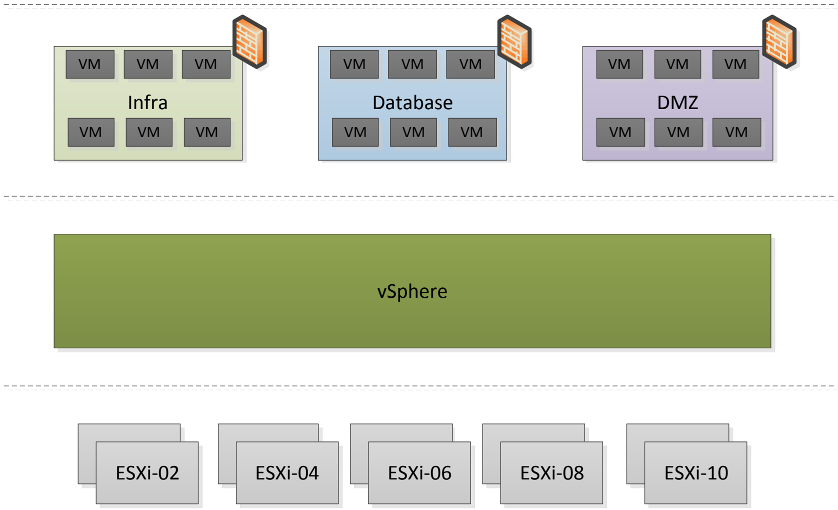
Leveraging vShield App will decrease operational effort associated with securing the environment due to the ability of applying policies to the different virtual objects. vShield App will also allow the isolation of traffic through the configuration of different security zones as specified in requirement R103.

VMware 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. vShield App provides firewalls between virtual machines by placing a firewall filter on every virtual network adapter.
- A vSphere plug-in option that can register vShield Manager as a VMware vSphere Client plug-in, after which, you can configure most vShield options from the vSphere Client. vShield Manager includes a web client which allows the security team to manage the security aspects without the need to log into vCenter.

vShield App

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



This security zones based on resource pools approach enables us to block certain traffic between the pools without the need to specify specific IP addresses. All virtual machines within the resource pools will inherit the specified rules / policies from its parent. All traffic, unless explicitly specified, is blocked. It should be noted that if virtual machines are migrated between resource pools their security policy will change based on the policy applied to the destination resource pool.

vShield App is a hypervisor-based, vNIC level technology which allows 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. It is recommended to apply security policies on a Resource Pool level. This will allow for applying policy changes to a complete zone in a single click and will also enable applying policies for new virtual machines by simply dragging them in to the correct resource pool. Table 24~~Table 24~~ describes the minimal zones and port configuration based on our current understanding of the infrastructure. It should be pointed out that currently only Internet – DMZ – Database is secured. Further investigation will need to result in a more restrictive policy and configuration of the security zones.

Table 24 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

Source	Destination	Port	Action
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

It should be noted that datacenter-level rules have higher priority than cluster-level rules. As rules are inherited from its parents it should be ensured that no conflicts exists 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 to define firewall policies. This applies to both the Infra and Database resource pool but not to the DMZ resource pool. The “Allow All” approach will be used for three weeks to allow for 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 KB 1012382. Depending on features and functionality used, it might be required to open additional ports.

vShield Deployment considerations

When deploying vShield Manager it should be noted that a 3GB memory reservation is configured on the vShield Manager virtual machine. This memory reservation can impact your vSphere HA Admission Control algorithm. In this design we have decided to use the Percentage Based admission control policy and as such will not be impacted by the reservation. When using the “Host failures tolerated” admission control policy the reservation could possibly skew the slot size used by this algorithm.

It is recommended to NTP time keeping for both the vShield Manager and vShield App appliances and Syslog for vShield App as described in the 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 Datacenter, Resource Pools, vApps and VMs. In this scenario Resource Pools are used.

Note: As vCenter Server is deployed as a virtual machine on the protected ESXi hosts it should be noted that it is possible to lock out your vCenter Server. To avoid all risks, vCenter can be deployed on a physical server or on a non-protected ESXi host.

For this design various best practices documents have been leveraged, VMware recommends reading this 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

vShield Manager Availability

VMware recommends enabling vSphere HA VM Monitoring on the vSphere 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 off site.

vSphere Security

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

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

It is recommended to periodically validate compliance through the use of the vSphere Compliance Checker or custom PowerCLI scripts. The vSphere Compliance Checker can be found here: http://www.vmware.com/products/datacenter-virtualization/vsphere-compliance-checker/overview.html.

vCenter Security

vCenter Server is the cornerstone of the VMware Cloud Infrastructure. Securing vCenter Server adequately is key to providing a highly available management solution. All recommendations pertinent to this design have been listed in the tables below.

Table 25 Security Hardening vCenter - Implementation

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 vCenter Server system
VSC06	Block access to ports not being used by vCenter

Code	Name
VSC07	Disable Managed Object Browser
VSC08	Disable Web Access
VSC09	Disable Datastore Browser
VSD01	Use least privileges for the vCenter Server Database user

Table 26 Security Hardening vCenter - Operational

Code	Name
VSH01	Maintaining supported operating system, database, and hardware for vCenter
VSH02	Keep vCenter Server system properly patched
VSH04	Avoid user login to 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 vSphere Client

Encryption and Security Certificates

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

Formatted: F

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 we would like to refer to the vSphere 5 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 Operating System, and has only very specific tasks, and thus very specific software modules. This means there is no arbitrary code running, and is not susceptible to common threats. ESXi has a very small footprint and thus provides a much smaller attack surface and requires fewer patches. The ESXi architecture consists of 3 major components:

- The virtualisation layer (VMkernel) – this layer is designed to run Virtual Machines. The VMkernel controls the hardware of the host and is responsible for the scheduling it's hardware resources for the virtual machines.
- The Virtual Machines – by design, all Virtual Machines are isolated from each other meaning that multiple Virtual Machines can run securely while sharing hardware resources. As the VMkernel mediates the 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 communications 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 only with access 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 you enable lockdown mode, 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 compliancy.

Table 27 Security Hardening ESXi - Implementation

Code	Name
HCM01	Do not use default self-signed certificates for ESXi Communication
HCM02	Disable Managed Object Browser
HCM03	Ensure 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 vSphere management traffic is on a restricted network.
NAR02	Ensure vMotion Traffic is isolated
NAR04	Strict control of access to Management network
NCN03	Ensure the “MAC Address Change” Policy is set to Reject

Code	Name
NCN04	Ensure the “Forged Transmits” Policy is set to Reject.
NCN05	Ensure the “Promiscuous Mode” Policy is set to Reject.

Table 28 Security Hardening ESXi - Operational

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 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 physical switch ports are configured with spanning tree disabled
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

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 domain. Rather than creating shared privileged accounts for local host authentication. Active Directory users and groups can be assigned to ESXi local Groups and authenticated against the domain. Whilst ESXi cannot use Active Directory (AD) to define user accounts, it can use AD to authenticate users. In other words, you can define individual user accounts on the ESXi host and use AD to manage passwords and account status.

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

VM Security

VM Security it provided through several different components of the VMware Cloud Infrastructure. In addition to vShield App it is recommend to apply the recommendations listed below.

It should be noted that these should be applied after P2V Migrations where and when applicable and should be applied on each of the created templates.

Table 29 Security Hardening Virtual Machine - Operational

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 VM-to-VM communication through VMCI.
VMX20	Limit VM log file size and number.
VMX21	Limit informational messages from the VM to the VMX file.
VMX24	Disable certain unexposed features
VMP03	Use templates to deploy VMs whenever possible.
VMP05	Minimize use of the VM console.

Operations

Operational efficiency is an important factor in this architecture case study. We will leverage several vSphere 5.0 Enterprise Plus components to lower operational cost and to 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 mainly done through Microsoft Powershell. VMware recommends using VMware PowerCLI to automate common virtualization tasks that will help you reduce your operational expenditure and allow for the organization to focus on more strategic projects. Alternatively vSphere Orchestrator can be used to create pre-defined workflows for standard procedures. Automating the management of your vSphere environment can help you save time and increase your Virtual Machine to Administrator ratio. Automation with VMware PowerCLI allows your team to deploy and maintain your virtual infrastructure faster and in a structured and automated model removing the risk of human error involved when performing repetitive tasks ensuring operational consistency. It is recommended to use PowerCLI to automate the following tasks at a minimum:

- Virtual Machine deployment and configuration
- Host Configuration
- Reporting on common issues (Snapshot usage etc)
- Capacity Reporting
- vCenter Configuration
- Patching of hosts
- Upgrade and auditing of VMware Tools for virtual machines
- Auditing all areas of the virtual infrastructure

Host Compliancy

Host profiles provide the ability to capture a “blue print” of an ESXi host configuration, and ensure compliance with this profile. It also provides the ability to remediate hosts that are not compliant ensuring a consistent host configuration within the vSphere environment. VMware recommends taking advantage of Host Profiles to ensure compliancy with a standardized ESXi host build. Host Profiles can be applied on several levels, at datacenter level and cluster level. VMware recommends applying Host Profiles at the cluster level, although only a single cluster is used this approach will allow for both a scale up and scale out approach from a cluster perspective.

Assumption: The assumption has been made that ESXi hosts within a cluster will be identically configured.

It is recommended to export the Host Profile after the creation to a location outside of the virtual infrastructure.

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

VM Storage Compliancy

VM 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 has become more efficient and user-friendly. This removes the need for maintaining complex and tedious spreadsheets and validating compliance manually during every migration or creation of a virtual machine or virtual disk. VMware recommends leveraging VM Storage Profiles (Table 22 VM Storage Profiles (Profile-Driven Storage) specifications~~Table 22 VM Storage Profiles (Profile-Driven Storage) specifications~~) to ensure virtual machines reside on the correct storage tier.

Virtual Machine Provisioning

Since this design is focused on providing a running platform for virtual machines that will be produced using a P2V process it is not easy to define a standard VM configuration model. In order to reduce the complexity and to standardize the environment three different virtual machine configurations will be offered which 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 each group will receive what it is entitled to. During the P2V process each virtual machine will be sized according to these standards. During the course of 4 months each virtual machine will be monitored and where applicable be reduced in size to further maximize consolidation ratio and cost savings

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

Table 30 VM Tiering Model

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

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

Patch / Version Management

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

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

The following table summarize the VUM configuration settings that will be applied in this case.

Table 31 vCenter Update Manager Server Configuration

Attribute	Specification
Patch download sources	**Download vSphere ESXi and ESX patches** Unselect **Download ESX 3.x patches** and **Download virtual appliance upgrades**
Shared repository	n/a
Proxy settings	TBD

Attribute	Specification
Patch download schedule	Every day at 03:00AM CET
Email notification	TBD
Update Manager baselines to leverage	Critical and non-critical ESX/ESXi host patches VMware Tools upgrade to match host
Virtual machine settings	N/A
ESX Host / Cluster settings	Host maintenance mode failure: **Retry** Retry interval: **5 Minutes** Number of retries: **3** Temporarily disable: Distributed Power Management
vApp settings	Select **Enable smart reboot after remediation**

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. VMware 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. vSphere Client will automatically check for and download an update if it exists when connecting to an updated vSphere 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 as they will directly impact the stability and availability of the virtual infrastructure. It should be noted that in this default configuration there would be a requirement for an administrator to log in to vCenter to view alerts as such it is recommended that for these specific events an email alert is configured.

- Cannot connect to Storage
- Cannot find vSphere 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
- vSphere HA Insufficient failover resources
- vSphere HA Cannot find master
- vSphere HA Failover in progress
- vSphere HA Host HA status
- vSphere HA VM monitoring error
- vSphere HA VM monitoring action

- vSphere HA Failover failed

VMware Syslog Collector

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

VMware ESXi Dump Collector

The Dump Collector is most useful for datacenters where ESXi hosts are configured using the Auto Deploy process. You can also install the Dump Collector for ESXi hosts that do 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 and as such it is not required to implement the ESXi Dump Collector.

Storage and Network Monitoring

All storage and networking infrastructure will be monitored in line with existing monitoring process, 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 as 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 vSphere 5.0 Enterprise+ in combination with vShield App. Using a building block approach this environment can easily be horizontally and vertically scaled. The building block approach ensures consistency and lowers operational costs. A similar approach for operational tasks, using standardized and scripted operational procedures, guarantees a decrease in administrational overhead. Leveraging vSphere 5.0 features like Network I/O Control, Profile-Driven Storage and Storage DRS simplifies management and monitoring but also ensures Service Level Agreements can be met.

In every environment security is crucial, but in many environments hardening of the respective layers is often neglected. VMware vCenter and ESXi are hardened using the various best practices and vShield App is used to shield 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 architecure. 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. Duncan was among the first VMware Certified Design Experts (VCDX 007). He is the co-author of multiple books including best sellers “vSphere 5.0 Clustering Technical Deepdive” and “VMware vSphere 4.1 HA and DRS Technical Deepdive”. Duncan is the owner and main author of leading virtualization and VMware blog Yellow-bricks.com.

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

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