



Cisco UCS and NetApp Solution for Oracle Real Application Clusters (RAC)

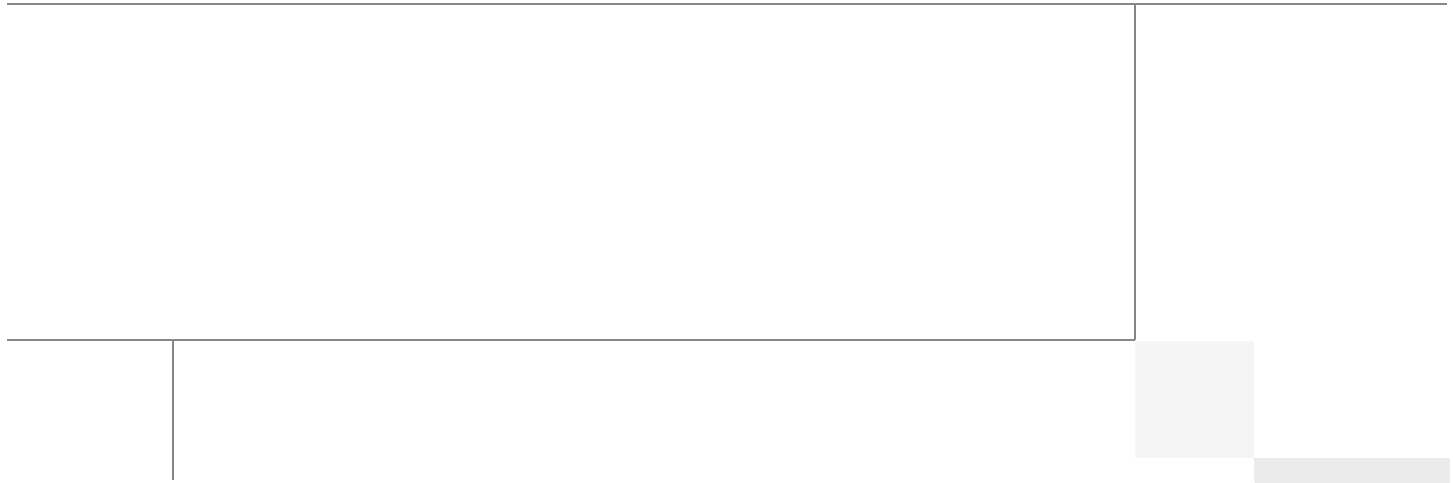
Last Updated: May 18, 2011



Cisco
Validated
Design



Building Architectures to Solve Business Problems



About the Authors



Haseeb Niazi

Haseeb Niazi, Solutions Architect, Systems Architecture and Strategy, Cisco Systems

Haseeb Niazi is a Solutions Architect in Cisco Systems Architecture and Strategy group based in RTP North Carolina. Haseeb has over eleven years of experience dealing in security and data center related technologies. He has helped a large number of enterprise and service provider customers evaluate and deploy various Cisco solutions in their networks. Haseeb holds a masters degree in computer engineering from University of Southern California and has presented to both internal and external audiences at various conferences and customer events.



Tom Simpkins

Tom Simpkins, Senior Database Performance Engineer, NetApp

Tom Simpkins is a Senior Database Performance Engineer concentrating on DW and DSS systems architecture for NetApp's performance engineering group. He has more than 17 years of hands-on Information systems experience mostly concentrated on the design, implementation, and management of DW and DSS systems. Prior to joining NetApp he spent several years architecting and managing data warehouse and decision support solutions for the healthcare industry. Mr. Simpkins holds a Bachelor of Information Systems degree from Marshall University and in addition to his technical background has served in several leadership roles where he managed and mentored DBA's, data modelers, and database developers, as well as DW operations and systems staff.



Niranjan Mohapatra

Niranjan Mohapatra, Oracle RAC RDBMS and NetApp Storage Specialist, NetApp

Niranjan Mohapatra is a specialist on Oracle RAC RDBMS and NetApp Storage and has worked as a Technical Marketing Engineer in the Oracle Alliance Engineering team at NetApp. He has over 11 years of extensive experience on Oracle RDBMS and associated tools working as a TME and a DBA handling production systems in various organizations. He holds a Master of Science (MSc) degree from Maharishi Dayanand University, India and is also an Oracle Certified Professional (OCP -11g DBA) and NetApp accredited storage architect.



Ashwath Narayan

Ashwath Narayan, Technical Marketing Engineer, Oracle and NetApp Storage, NetApp

Ashwath Narayan is working as a TME on Oracle Alliance Engineering team at NetApp. He has five years of extensive experience at NetApp and has worked on various products and solutions in the portfolio.

About Cisco Validated Design (CVD) Program

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

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

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

The Cisco implementation of TCP header compression is an adaptation of a program developed by the University of California, Berkeley (UCB) as part of UCB's public domain version of the UNIX operating system. All rights reserved. Copyright © 1981, Regents of the University of California.

Cisco and the Cisco Logo are trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and other countries. A listing of Cisco's trademarks can be found at <http://www.cisco.com/go/trademarks>. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1005R)

Any Internet Protocol (IP) addresses and phone numbers used in this document are not intended to be actual addresses and phone numbers. Any examples, command display output, network topology diagrams, and other figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses or phone numbers in illustrative content is unintentional and coincidental.

Cisco UCS and NetApp Solution for Oracle Real Application Clusters (RAC)

© 2011 Cisco Systems, Inc. All rights reserved.



Cisco UCS and NetApp Solution for Oracle Real Application Clusters (RAC)

Introduction

This Cisco® Validated Design describes how the Cisco Unified Computing System™ (UCS) can be used in conjunction with NetApp FAS unified storage systems to implement a Decision Support System (DSS) or Online Transaction Processing (OLTP) database utilizing an Oracle Real Application Clusters (RAC) system. The Cisco UCS provides the compute, network, and storage access components of the cluster, deployed as a single cohesive system. The result is an implementation that addresses many of the challenges that database administrators and their IT departments face today, including requirements for a simplified deployment and operation model, high performance for Oracle RAC software, and lower total cost of ownership (TCO). This guide introduces the Cisco UCS and NetApp architecture and provides implementation instructions. It concludes with an analysis of the cluster's performance, reliability characteristics, and data management capabilities.

Data powers essentially every operation in a modern enterprise, from keeping the supply chain operating efficiently to managing relationships with customers. Oracle RAC brings an innovative approach to the challenges of rapidly increasing amounts of data and demand for high performance. Oracle RAC uses a horizontal scaling (or scale-out) model that allows organizations to take advantage of the fact that the price of one-to-four-socket x86-architecture servers continues to drop while their processing power increases unabated. The clustered approach allows each server to contribute its processing power to the overall cluster's capacity, enabling a new approach to managing the cluster's performance and capacity.

Historically, enterprise database management systems have run on costly symmetric multiprocessing servers that use a vertical scaling (or scale-up) model. However, as the cost of one-to-four-socket x86-architecture servers continues to drop while their processing power increases, a new model has emerged. Oracle RAC uses a horizontal scaling, or scale-out, model, in which the active-active cluster uses multiple servers, each contributing its processing power to the cluster, increasing performance, scalability, and availability. The cluster balances the workload across the servers in the cluster and the cluster can provide continuous availability in the event of a failure.

Cisco is the leader in providing network connectivity in enterprise data centers. With the introduction of the Cisco UCS, Cisco is now equipped to provide the entire clustered infrastructure for Oracle RAC deployments. The Cisco UCS provides compute, network, virtualization, and storage access resources that are centrally controlled and managed as a single cohesive system. With the capability to scale up to 14 UCS chassis (112 B200 Blades) in the current release (1.3(1), the Cisco UCS provides an ideal foundation for very large scale Oracle RAC deployments.



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

Copyright © 2011 Cisco Systems, Inc. All rights reserved

All components in an Oracle RAC implementation must work together flawlessly and Cisco and NetApp have worked closely together to create, test, and validate a configuration of Oracle RAC on the Cisco UCS. This configuration provides an implementation of Oracle Database 11g Release 2 with Real Application Clusters technology consistent with industry best practices. To provide storage and data management capabilities this architecture uses NetApp FAS unified storage systems with SAS drives and state-of-the-art Flash Cache to further speed performance and provide space efficient snapshots, DR, and cloning capabilities.

Benefits of the Configuration

Oracle RAC on the Cisco UCS and NetApp offers a number of important benefits, including:

- [Simplified Deployment and Operation](#)
- [Oracle Database 11g Direct NFS Client](#)
- [High-Performance Platform for Oracle RAC](#)
- [Safer Deployments with Validated Configurations](#)
- [Effective System Management and Fault Tolerance](#)

Simplified Deployment and Operation

Because the entire cluster runs on a single cohesive system, database administrators no longer need to painstakingly configure each element in the hardware stack independently. The system's compute, network, and storage-access resources are essentially stateless, provisioned dynamically by Cisco UCS Manager. This role- and policy-based embedded management system handles every aspect of system configuration, from a server's firmware and identity settings to the network connections that connect storage traffic to the destination storage system. This capability dramatically simplifies the process of scaling an Oracle RAC configuration or re-hosting an existing node on an upgrade server. Cisco UCS Manager uses the concept of service profiles and service profile templates to consistently and accurately configure resources. The system automatically configures and deploys servers in minutes, rather than the hours or days required by traditional systems composed of discrete, separately managed components. Indeed, Cisco UCS Manager can simplify server deployment to the point where it can automatically discover, provision, and deploy a new blade server when it is inserted into a chassis.

The system is based on a 10-Gbps unified network fabric that radically simplifies cabling at the rack level by consolidating both IP and Fibre Channel traffic onto the same rack-level 10-Gbps converged network. This wire-once model allows in-rack network cabling to be configured once, with network features and configurations all implemented by changes in software rather than by error-prone changes in physical cabling. This configuration not only supports separate public and private networks as required by Oracle RAC, it also provides redundancy with automatic failover. The notion of public and private networks in Oracle RAC does not necessarily mean secured and unsecured networks as might be commonly understood by network personnel.

Direct NFS (D-NFS) over multiple 10GE connections is utilized to access the NetApp Storage layer where NetApp's unified storage platform provides high performance along with unique data management capabilities for the enterprise. HA clusters provide fault tolerance and automatic failover capabilities. Snapshots and FlexClone allow administrators to take space efficient copies of the database without impacting performance or initially requiring additional space. These can then be used for online point-in-time, read-only copies of the data for backup and recovery as well as writable clones for quickly and efficiently creating development or test databases that only require additional space as data changes.

Oracle Database 11g Direct NFS Client

D-NFS is an Oracle developed, integrated, and optimized client that runs in user space rather than within the operating system kernel. This architecture provides for enhanced scalability and performance over traditional NFS v3 clients. Unlike traditional NFS implementations, Oracle supports asynchronous I/O across all operating system environments with D-NFS. In addition, performance and scalability are dramatically improved with its automatic link aggregation feature. This allows the client to scale across as many as four individual network pathways with the added benefit of improved resiliency when network connectivity is occasionally compromised. It also allows D-NFS to achieve near block level performance. For more information on D-NFS comparison to block protocols, see: <http://media.netapp.com/documents/tr-3700.pdf>.

High-Performance Platform for Oracle RAC

The Cisco UCS B200-M1 Blade Servers used in this validated configuration feature Intel Xeon 5500 series processors that deliver intelligent performance, automated energy efficiency, and flexible virtualization. Intel Turbo Boost Technology automatically boosts processing power through increased frequency and use of hyper-threading to deliver high performance when workloads demand and thermal conditions permit.

The patented Cisco Extended Memory Technology offers twice the memory footprint—up to 384 GB on a B250 Blade Server—of any other server using 8-GB DIMMs or the economical option of a 192-GB memory footprint using less expensive 4-GB DIMMs. Both choices for large memory footprints can help speed database performance by allowing more data to be cached in memory.

The Cisco UCS's 10-Gbps unified fabric delivers standards-based Ethernet and Fibre Channel over Ethernet (FCoE) capabilities that simplify and secure rack-level cabling while speeding network traffic compared to traditional Gigabit Ethernet networks. The balanced resources of the Cisco UCS allow the system to easily process an intensive online transaction processing (OLTP) and/or decision-support system (DSS) workload without resource saturation.

Safer Deployments with Validated Configurations

Cisco and Oracle have been working together to promote interoperability of Oracle's next-generation database and application solutions with the Cisco UCS, helping make the Cisco UCS a simple and safe platform on which to run Oracle software.

Cisco and NetApp have worked together to:

- Complete a Cisco Validated Design for Cisco UCS running Red Hat Enterprise Linux and Oracle Database 11gR2 utilizing NetApp unified storage systems. This system has been designed to ensure interoperability, balanced architectural design, fault tolerance, and scalability.
- Stress test the environment to validate the design objectives and ensure a balanced, reliable, and scalable system that follows the Cisco Data Center 3.0 architecture and is ready to seamlessly integrate into an overall datacenter architecture.

Effective System Management and Fault Tolerance

One of the key success factors in any implementation is the effectiveness of the on-going management and fault tolerance. Cisco's UCS combined with NetApp's unified storage system work together to create a system that is easy to manage, scalable, and provides high levels of fault tolerance.

Document Objectives

This document highlights the Cisco UCS and NetApp storage systems. It discusses how this architecture can address many of the challenges that database administrators and their IT departments face today. This document provides:

- Guidance on how to design a data center architecture that provides network, server, and application-level services that are needed with an enterprise Oracle RAC implementation.
- An overview of the Oracle RAC configuration along with instructions for setting up the Cisco UCS and the NetApp FAS storage system.
- Reports on systems performance measurements for OLTP, DSS, and Interconnect workloads during multi-hour Oracle OAST stress tests.



Note

The reader should refer to the product documentation for Cisco UCS, Oracle RAC, and NetApp FAS for detailed setup and configuration.

- Cisco UCS Configuration Guide:
http://www.cisco.com/en/US/docs/unified_computing/ucs/sw/gui/config/guide/1.3.1/b_UCSM_GUI_Configuration_Guide_1_3_1.html
- NetApp Configuration Guide:
<https://now.netapp.com/NOW/knowledge/docs/hardware/NetApp/syscfg/>
- Oracle RAC Configuration Guide:
http://download.oracle.com/docs/cd/B28359_01/install.111/b28264/racinstl.htm

Solution Architecture

Cisco UCS Overview

The Cisco UCS used for the certified configuration is based on Cisco B-Series Blade Servers; however, the breadth of Cisco's server and network product line suggests that similar product combinations will meet the same requirements. The Cisco UCS uses a form-factor-neutral architecture that will allow Cisco C-Series Rack-Mount Servers to be integrated as part of the system in upcoming releases. The system's core components—high-performance compute resources integrated using a unified fabric—can be integrated manually today using Cisco C-Series servers and Cisco Nexus™ 5000 Series Switches.

The heart of the compute portion of the environment is built from the hierarchy of components illustrated in [Figure 1](#).

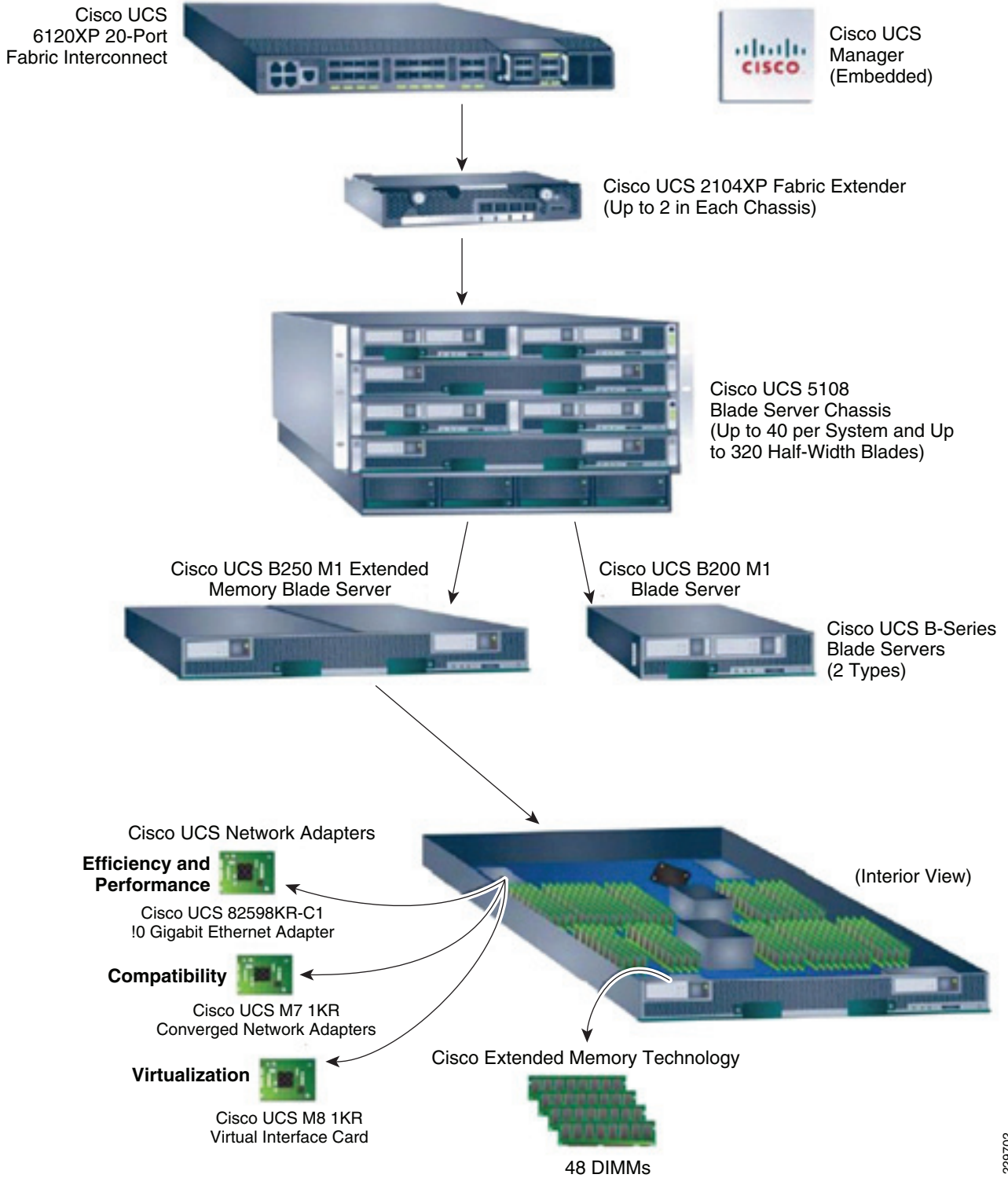
- The Cisco UCS 6120XP 20-Port Fabric Interconnect provides low-latency, lossless, 10-Gbps unified fabric connectivity for the cluster. The interconnect provides connectivity to blade server chassis and the enterprise IP network. Two fabric interconnects are configured in the cluster, providing the capability to utilize separate fabrics for public and private traffic, as well as providing high availability in the event of a failure.
- The Cisco UCS 2104XP Fabric Extender brings the unified fabric into each blade server chassis. The fabric extender is configured and managed by the fabric interconnects, eliminating the complexity of blade server-resident switches. Two fabric extenders are configured in each of the cluster's two blade server chassis. Each one uses two of the four available 10-Gbps uplinks to connect to one of the two fabric interconnects.

- The Cisco UCS 5108 Blade Server Chassis houses the fabric extenders, up to four power supplies, and up to eight blade servers. As part of the system's radical simplification, the blade server chassis is also managed by the fabric interconnects, eliminating another point of management. Two chassis were configured for the Oracle RAC described in this document for chassis redundancy in addition to fabric and server blade redundancy.
- The blade chassis supports up to eight half-width blades or up to four full-width blades. The certified configuration uses four (two in each chassis) Cisco UCS B200 M1 Blade Servers, each equipped with two quad-core Intel Xeon 5500 series processors (the testing process implemented Xeon E5540) at 2.53 GHz. Each blade server was configured with 48 GB of memory. A memory footprint of up to 384 GB can be accommodated through the use of a Cisco UCS B250 M1 Extended Memory Blade Server.
- The blade server form factor supports a range of mezzanine-format Cisco UCS network adapters, including a 10 Gigabit Ethernet network adapter designed for efficiency and performance, the Cisco UCS M81KR Virtual Interface Card designed to deliver the system's full support for virtualization, and a set of Cisco UCS M71KR converged network adapters designed for full compatibility with existing Ethernet and Fibre Channel environments. These adapters present both an Ethernet network interface card (NIC) and a Fibre Channel host bus adapter (HBA) to the host operating system. They make the existence of the unified fabric transparent to the operating system, passing traffic from both the NIC and the HBA onto the unified fabric. Versions are available with either Emulex or QLogic HBA silicon; the certified configuration uses the Cisco UCS M81KR Virtual Interface Card that provides 20-Gbps of connectivity by connecting to each of the chassis fabric extenders.
- The Cisco UCS M8 M81KR Virtual Interface Card is a virtualization-optimized Fibre Channel over Ethernet (FCoE) mezzanine adapter designed for use with Cisco UCS B-Series Blade Servers. The virtual interface card is a dual-port 10 gigabit Ethernet mezzanine card that supports standards-compliant virtual interfaces that can be dynamically configured so that both their interface type (network interface card [NIC] or host bus adapter [HBA] and identity MAC address and worldwide name [WWN]) are established using just-in-time provisioning. The Cisco M81KR VIC is a fully standards-compliant Fiber Channel adapter that delivers cutting edge storage IOPs and throughput performance.

**Note**

The architecture for the test system used and recommends using the M81KR VIC card in production environments for its ability to present multiple vNICs to the system and address the Oracle requirement to have separate interfaces for the public and interconnect networks while also allowing an additional vNIC to isolate the nfs/d-nfs storage traffic.

Figure 1 Cisco Unified Computing System Components



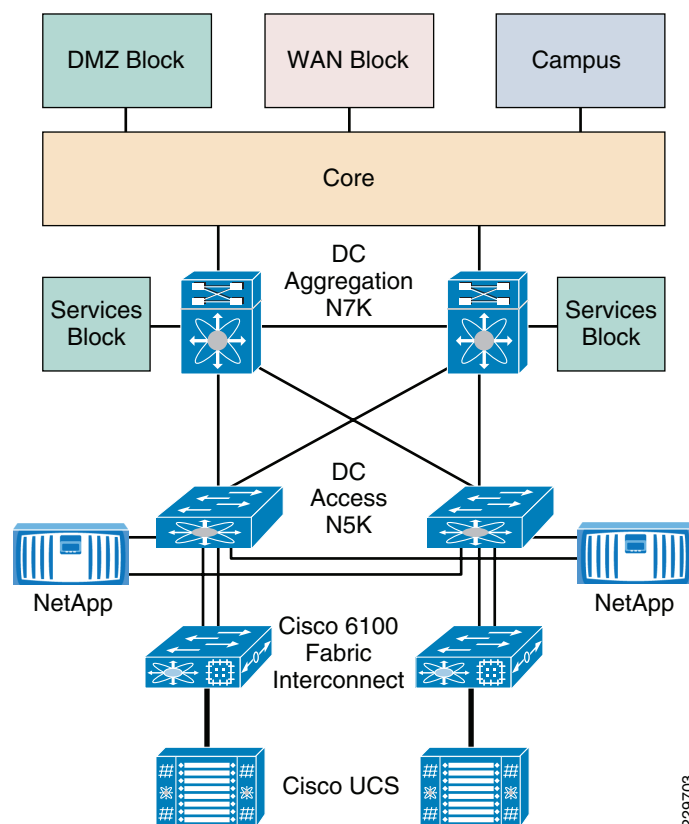
229702

Deployment Topology

At the compute layer, Cisco UCS provides a unified compute environment with integrated management and networking to support compute resources. Each of the UCS blades serves as a node in the Oracle RAC cluster.

At the network layer, we recommend a three-tier architecture enabled with Nexus 5000 as an unified access layer switch and Nexus 7000 as a core/aggregation layer switch as shown in [Figure 2](#).

Figure 2 **Topology Overview**

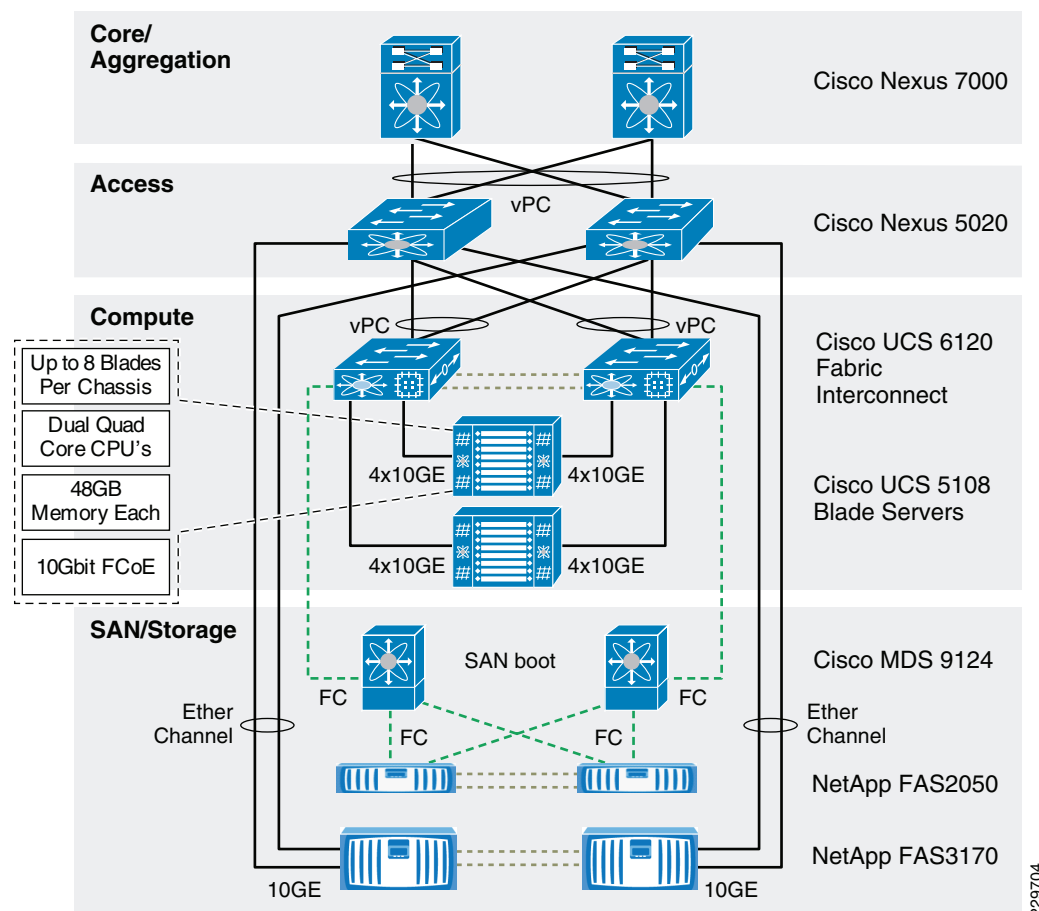


The Oracle RAC testing only included components up to the access layer, i.e., up to the Nexus 5000 switch pair. The two UCS 6120 Fabric Interconnects with dual-fabric topology enable a 10G compute layer. Both the UCS 6120 Fabric Interconnects and NetApp FAS3170 storage controllers are connected to the Nexus 5000 access switch via Port-Channel with dual-10 Gig Ethernet. The NetApp FAS controllers use redundant 10Gb NICs configured in a two-port Virtual Interface (VIF). Each port of the VIF is connected to one of the upstream switches, allowing multiple active paths by utilizing the Nexus vPC feature. This topology combined with vPC provides increased redundancy and bandwidth with a lower required port count.

Cisco MDS 9124 provides dual-fabric SAN connectivity at the access layer and both UCS 6120 and NetApp FAS2050 storage controllers are connected to both fabric A and B via Fiber Channel (FC) for SAN Boot. The UCS 6120 has a single FC link to each fabric, each providing redundancy to the other. NetApp FAS2050 is connected to MDS 9124 via dual-controller FC port in a full-mesh topology.

A detailed view of the physical topology is illustrated in [Figure 3](#), which identifies the various levels of the architecture as well as some of the key components and features of the fabric. Although the core/aggregation level is recommended for data center deployments and illustrated in [Figure 3](#), it was not part of the test architecture because the testing and loads were concentrated on the access layer and below and did not involve the larger core/aggregation levels.

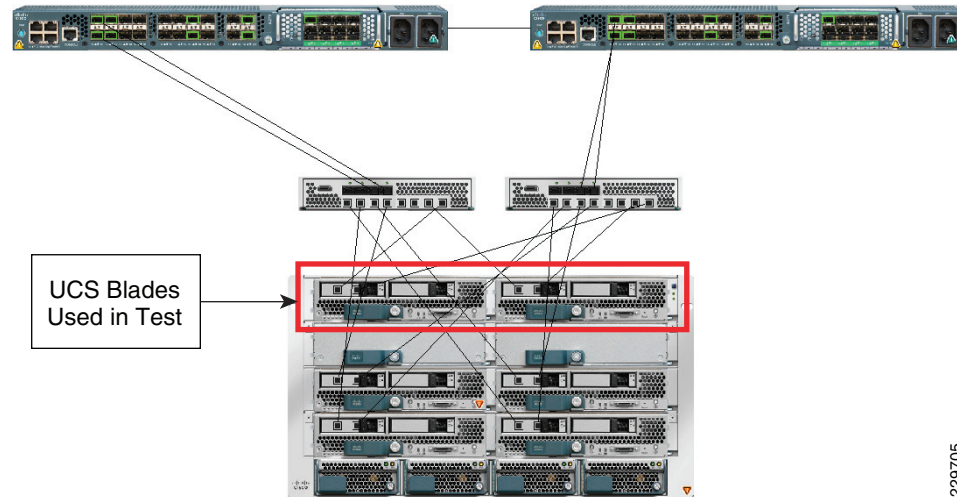
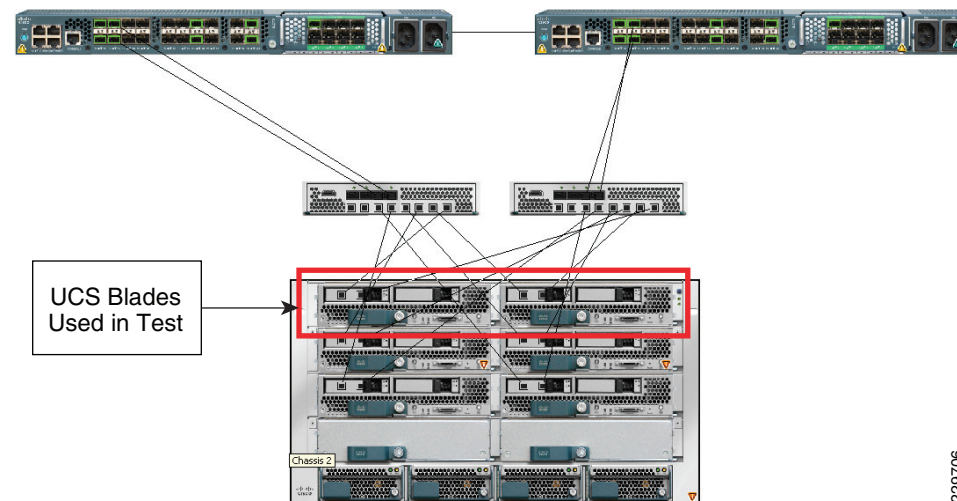
Figure 3 **Detailed Topology**



Infrastructure Deployment

This section describes the steps necessary to build a new Cisco UCS and NetApp environment that is ready to accept an Oracle RAC installation.

After racking the equipment, connect the chassis and 6120 interconnects similar to [Figure 4](#) and [Figure 5](#). Make note of all port assignments, as this will be necessary in later steps. Note that only the first two blades in each chassis were used for the test architecture.

Figure 4 6120 Fabric Interconnect Topology—Chassis 1**Figure 5** 6120 Fabric Interconnect Topology—Chassis 2**Note**

The links between the blade servers and the fabric extenders shown above are for illustration purposes only. These links are internal to the chassis through the backplane.

Network Infrastructure Connectivity

As mentioned previously, the testing did not incorporate the network core and aggregation layers, but it is important to point out that the infrastructure deployment has adopted the best practices recommended in the design guides below, ensuring that the access layer can seamlessly integrate into the Cisco Data Center 3.0 architecture. Any exceptions and specific changes relevant to this deployment are explained in the appropriate sections.

- Cisco Data Center 3.0 infrastructure:
http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DC_3_0/DC-3_0_IPInfra.html
- Cisco SAFE Design Guide:
http://www.cisco.com/en/US/docs/solutions/Enterprise/Security/SAFE_RG/chap4.html

Configuring Infrastructure Management VLAN

The infrastructure devices are managed via separate routed VLANs with appropriate secured access. The VLANs used for VM and application management must be separate from the infrastructure management VLANs. All Cisco devices are configured for secured shell (SSH) access. The management interfaces in NxOS are attached to a dedicated management VRF and thus act as entirely separate out-of-band management instances.

Nexus 5000

N5k_f22

```
vrf context management
ip route 0.0.0.0/0 10.61.160.1
interface mgmt0
vrf member management
ip address 10.61.160.192/24
```

N5k_f22l

```
vrf context management
ip route 0.0.0.0/0 10.61.160.1
interface mgmt0
vrf member management
ip address 10.61.160.194/24
```

MDS 9124

```
mds9124-1
interface mgmt0
switchport speed 100
ip address 10.61.178.237 255.255.255.0
```

```
mds9124-1
interface mgmt0
switchport speed 100
ip address 10.61.178.238 255.255.255.0
```

Similarly, any devices that support dedicated management access can be put in a common subnet. The same subnet is later used for enabling management connectivity for the UCS 5100 blade server management (KVM), and UCS 6100 fiber interconnect.

Port Mode Configuration

- Step 1** All the edge end-point devices connected to Nexus 5000 are configured as an “Edge” port type, which replaces spanning-tree portfast in IOS based devices.

```
interface port-channel52
description uscmA:po52
switchport mode trunk
```

```
spanning-tree port type edge trunk <--
```

Step 2 All inter switch links, typically bridge-to-bridge links, are configured as “Network” port type.

```
interface port-channel500
  switchport mode trunk
  switchport trunk allowed vlan << vlan numbers >>
  spanning-tree port type network <--
```



Note Network ports are connected only to switches or bridges. Bridge Assurance is enabled only on network ports. If you mistakenly configure ports that are connected to hosts or other edge devices, as spanning tree network ports, those ports will automatically move into the blocking state.

Step 3 Any other connectivity (not used in this deployment) is configured as “Normal” port type and considered as generic links in spanning tree.

Network Topology Connectivity

The design guide recommends connecting all the devices in distributed Port Channel-based topology using virtual port-channel (vPC). The vPC-based configuration:

- Allows a single device to use a Port Channel across two upstream devices
- Eliminates Spanning Tree Protocol blocked ports
- Provides a loop-free topology
- Uses all available uplink bandwidth
- Provides fast convergence if either the link or a device fails
- Provides link-level resiliency
- Helps ensure high availability

It is strongly recommend to not disable STP protocol; the topology is loop-free in a normal condition, however accidental loops can be created via misconfiguration, which can lead to a catastrophic collapse of network connectivity if STP is not present to block the alternate path. The configuration steps below cover the essential vPC configuration steps. The following CLI must be enabled on both Nexus 5000s.

Step 1 Enable the vPC feature.

```
N5k_f22(config)# feature vPC
```

Step 2 Enable vPC domain. The domain ID must be unique in the entire network as the domain number is used in LACP system identifier. If LACP system-id is mis-matched, the vPC peers cannot synchronize with each other.

```
! Configure the vPC Domain ID - It should be unique within the network
N5k-f22(config)# vPC domain 500
```

Step 3 Enable primary and secondary role for the vPC domain. Roles are defined under the domain configuration. vPC role defines which of the two vPC peers processes BPDUs. It is recommended to ensure that the vPC primary switch is also the root bridge. **The switch with lower priority will be elected as the vPC primary switch.** If the peer link fails, vPC peer will detect whether the peer switch

is alive through the vPC peer keepalive link. If the vPC primary switch is alive, the vPC secondary switch will suspend its vPC member ports to prevent potential looping while the vPC primary switch keeps all its vPC member ports active.

```
N5k-f22(config-vPC-domain)# role priority ?
<1-65535> Specify priority value
```

- Step 4** You must enable peer keep-alive for detecting a dual-active condition, which can occur if both vPC peer-links are down. Without initial peer keep-alive configuration, the vPC domain will not be active. Peer keep-alive provides an out-of-band heartbeat between vPC peers. It is highly recommended not to carry vPC keep-alive over the peer-link. Peer keep-alive is a routable protocol. A primary design requirement is to have a physically different path than all other vPC traffic. Multiple methods are available to enable keep-alive for the Nexus 5000.



Note If using mgmt 0 interfaces, do **not** connect the supervisor management interfaces back-to-back. In a dual supervisor configuration only one management port is active at a given point in time. Connect both mgmt 0 ports to the OOB network.

The deployment guide uses management interfaces method to enable peer keep-alive.

```
vPC domain 500
  role priority 10
  peer-keepalive destination 10.61.160.194 source 10.61.160.192
```



Note vPC domains remain functional if the vPC peer keep-alive becomes unreachable, however the vPC peer keep-alive must be operational in order to establish a functional vPC connection during the initial configuration. The vPC peer keep-alive should always stay active.

- Step 5** Enable vPC peer-link connectivity between vPC peers. Peer links carry both vPC data and control traffic (STP BPDUs, IGMP updates, etc.) between peer switches. A minimum of two 10GbE ports must be configured to assure high availability, preferably on a separate line card or module. It is not recommended to share vPC and non-vPC VLANs on the same peer-link. The best practice is to allow all the VLANs which are part of vPC domain to be carried over the vPC peer-link. Failing to allow VLANs over the vPC peer-link can disrupt connectivity.

```
interface port-channel500
  description vPC peer-link
  vpc peer-link
  spanning-tree port type network <-- peer-link port role must be of type "network"
```



Note **Always** dual attach devices using vPCs if possible. Singly-attached devices greatly impact the availability of the entire system, create traffic patterns that impact application response time, and complicate system capacity planning.

- Step 6** Add the interface to the PortChannel and then move the PortChannel to the vPC to connect to the downstream device.

[Table 1](#) lists all relevant configurations needed for enabling vPC configuration.

Table 1 vPC configuration on Nexus 5000

Nexus 5000 f22	Nexus 5000 f22l	Comments
feature vpc	feature vpc	
vpc domain 500	vpc domain 500	Unique vPC domain
role priority 10	role priority 20	
peer-keepalive destination 10.61.160.194 source 10.61.160.192	peer-keepalive destination 10.61.160.192 source 10.61.160.194	
interface port-channel500 description vPC peer-link switchport mode trunk vpc peer-link spanning-tree port type network	interface port-channel500 description vPC peer-link switchport mode trunk vpc peer-link spanning-tree port type network	Allow all VLANs on vPC peer-links
interface Ethernet1/36 switchport mode trunk spanning-tree port type network channel-group 500 mode active	interface Ethernet1/36 switchport mode trunk spanning-tree port type network channel-group 500 mode active	Use LACP (mode active)
interface Ethernet1/37 switchport mode trunk spanning-tree port type network channel-group 500 mode active	interface Ethernet1/37 switchport mode trunk spanning-tree port type network channel-group 500 mode active	

End Devices Connectivity with Port-Channel

Once the vPC connectivity between the distribution and access layer is defined, the end devices participating in the vPC domain are configured. In this deployment two critical end devices (UCS 6100 and NetApp FAS-3170s) are configured with port-channel connecting both Nexus 5000s.

Create a port-channel with the same number in both Nexus 5000s and attach a user-defined vPC number (preferably same number as port-channel). [Table 2](#) and [Table 3](#) show the configurations for connecting UCS 6100 fabric interconnects and NetApp 3170 FAS to two Nexus 5000s.



Note

LACP is the default and only port-channel aggregation protocol supported under Nexus series platforms. It is highly recommended to configure “active” mode of operation for port-channels.

Each edge device has specific configuration guidelines for configuring and enabling port-channel. For the UCS 6100 Fiber Interconnect, follow the steps in [Network Connectivity via Port Channels](#). For the NetApp FAS-3170, follow the steps in [Storage Controller Configuration](#).

Table 2 *Cisco 6100 Fabric Interconnect Configuration***UCS 6100 Fabric A—ucsm-A**

5K f22	5K f22I	Comment
interface port-channel52 description ucsmA:po52 switchport mode trunk switchport trunk allowed vlan 101,103-104,160-178,350 vpc 52 speed 10000	interface port-channel52 description ucsmA:po52 switchport mode trunk switchport trunk allowed vlan 101,103-104,160-178,350 vpc 52	
interface Ethernet1/29 description ucsmA:port 19 switchport mode trunk switchport trunk allowed vlan 101,103-104,160-178,350 spanning-tree port type edge trunk channel-group 52 mode active	interface Ethernet1/29 description ucsmA:port 13 switchport mode trunk switchport trunk allowed vlan 101,103-104,160-178,350 spanning-tree port type edge trunk channel-group 52 mode active	Recommended LACP mode "active"
interface Ethernet1/30 description ucsmA:port 20 switchport mode trunk switchport trunk allowed vlan 101,103-104,160-178,350 spanning-tree port type edge trunk channel-group 52 mode active	interface Ethernet1/30 description ucsmA:port 14 switchport mode trunk switchport trunk allowed vlan 101,103-104,160-178,350 spanning-tree port type edge trunk channel-group 52 mode active	

UCS 6100 Fabric B—ucsm-B

5K f22	5K f22I	Comment
interface port-channel53 description ucsmB:po53 switchport mode trunk switchport trunk allowed vlan 101,103-104,160-178,350 vpc 53 speed 10000	interface port-channel53 description ucsmB:po53 switchport mode trunk switchport trunk allowed vlan 101,103-104,160-178,350 vpc 53	
interface Ethernet1/31 description ucsm-B:port 19 switchport mode trunk switchport trunk allowed vlan 101,103-104,160-178,350 spanning-tree port type edge trunk channel-group 53 mode active	interface Ethernet1/31 description ucsm-B:port 13 switchport mode trunk switchport trunk allowed vlan 101,103-104,160-178,350 spanning-tree port type edge trunk channel-group 53 mode active	
interface Ethernet1/32 description ucsm-B:port 20 switchport mode trunk switchport trunk allowed vlan 101,103-104,160-178,350 spanning-tree port type edge trunk channel-group 53 mode active	interface Ethernet1/32 description ucsm-B:port 14 switchport mode trunk switchport trunk allowed vlan 101,103-104,160-178,350 spanning-tree port type edge trunk channel-group 53 mode active	

Table 3 NetApp FAS Configuration

3170 vif1 Configuration		
5K f22	5K f22I	Comment
interface port-channel60 description 3170ucs-1:vif1 switchport access vlan 103 vpc 60	interface port-channel60 description 3170ucs-1:vif1 switchport access vlan 103 vpc 60	
interface Ethernet1/1 description 3170ucs-1:10G_1 switchport access vlan 103 channel-group 60 mode active	interface Ethernet1/1 description 3170ucs-1:10G_2 switchport access vlan 103 channel-group 60 mode active	
3170 vif2 Configuration		
5K f22	5K f22I	Comment
interface port-channel61 description 3170ucs-2:vif1 switchport access vlan 103 vpc 61	interface port-channel61 description 3170ucs-2:vif1 switchport access vlan 103 vpc 61	
interface Ethernet1/2 description 3170ucs-2:10G_1 switchport access vlan 103 channel-group 61 mode active	interface Ethernet1/2 description 3170ucs-2:10G_2 switchport access vlan 103 channel-group 61 mode active	
3170 vif3 Configuration		
5K f22	5K f22I	Comment
interface port-channel62 description 3170ucs-3:vif1 switchport access vlan 103 vpc 62	interface port-channel62 description 3170ucs-3:vif1 switchport access vlan 103 vpc 62	
interface Ethernet1/3 description 3170ucs-3:10G_1 switchport access vlan 103 channel-group 62 mode active	interface Ethernet1/3 description 3170ucs-3:10G_2 switchport access vlan 103 channel-group 62 mode active	
3170 vif4 Configuration		
5K f22	5K f22I	Comment
interface port-channel63 description 3170ucs-4:vif1 switchport access vlan 103 vpc 63	interface port-channel63 description 3170ucs-4:vif1 switchport access vlan 103 vpc 63	
interface Ethernet1/4 description 3170ucs-4:10G_1 switchport access vlan 103 channel-group 63 mode active	interface Ethernet1/4 description 3170ucs-4:10G_2 switchport access vlan 103 channel-group 63 mode active	

A sample output of the state of operational vPC connectivity is shown below:

N5kf22# **show vpc**

Legend:

(*) - local vPC is down, forwarding via vPC peer-link

```
vPC domain id           : 500
Peer status              : peer adjacency formed ok
vPC keep-alive status    : peer is alive
Configuration consistency status: success
vPC role                 : primary
```

vPC Peer-link status

```
-----
id   Port   Status Active vlans
--   -
1    Po500   up      1,6,101,103-104,160-178,350
-----
```

vPC status

```
-----
id   Port   Status Consistency Reason Active vlans
-----
52   Po52    up      success    success    101,103-104,160-178,350
53   Po53    up      success    success    101,103-104,160-178,350
60   Po60    up      success    success    103
61   Po61    up      success    success    103
62   Po62    up      success    success    103
63   Po63    up      success    success    103
-----
```

Storage Controller Configuration

Refer to TR-3633, *NetApp Best Practice Guidelines for Oracle Database 11g*, for additional information on the configuration of the NetApp storage controller and Oracle 11g:
<http://media.netapp.com/documents/tr-3633.pdf>.

-
- Step 1** Referring to the *NetApp Installation and Setup Instructions* (http://now.netapp.com/NOW/knowledge/docs/hardware/hardware_index.shtml), configure the four storage controllers as two HA pairs, ensuring that disk shelves are properly connected and that the controllers are linked with an HA cluster cable. Each controller should have a dual-port 10Gb NIC cabled as indicated in [Figure 2](#).
- Step 2** Referring to the *Active/Active Configuration Guide* (<http://now.netapp.com/NOW/knowledge/docs/ontap/rel734/>), boot the controllers into maintenance mode and configure disk ownership appropriately. Ensure that the controllers are running Data ONTAP version 7.3.4 (see the Upgrade Guide for details on upgrading).
- Step 3** Reboot the controllers and connect to each system via the serial console to complete the setup process. Refer to the Software Setup Guide for details on this procedure. If this is the first time the controller has booted, the following questions are displayed automatically; otherwise, type **setup** at the prompt to begin the process. Unless otherwise noted, all steps should be performed on both storage controllers.

```
Please enter the new hostname? [ ]: 3170-ucs1 (for the second controller, input
"3170-ucs2")
Do you want to enable IPv6? [ ]: n
Do you want to configure virtual network interfaces? [ ]: y
Number of virtual interfaces to configure? [ ]: 1
Name of virtual interface #1? [ ]: vif1
Is vif0 a single [s], multi [m] or lacp [l] virtual interface? [ ]: l (a lowercase 'L')
Is vif0 to use IP based [i], MAC based [m], Round-robin based [r] or Port based [p] load
balancing? [ ]: i
Number of links for vif1? [ ]: 2
Name of link #1 for vif1? [ ]: e2a (the first 10Gb interface)
Name of link #2 for vif1? [ ]: e2b (the second 10Gb interface)
Please enter the IP address for Network Interface vif1 [ ]: (Press Enter)
Should virtual interface vif0 take over a partner virtual interface during failover? [ ]:
```

y

```

Please enter the partner virtual interface name to be taken over by vif1 [ ]: vif1
Please enter the IP address for Network Interface e0a [ ]: 10.61.172.55
(NOTE: If you are NOT using e0a as a separate administration interface, input a
placeholder IP address here, such as 169.254.1.1.)
Please enter the netmask for Network Interface e0a [255.0.0.0]: (Press Enter)
Should interface e0a take over a partner virtual interface during failover? [ ]: n
Please enter media type for e0a {100tx-fd, tp-fd, 100tx, tp, auto (10/100/1000)} [auto]:
(Press Enter)
Please enter flow control for e0a {none, receive, send, full} [full]: (Press Enter)
Do you want e0a to support jumbo frames? [ ]: n
Please enter the IP address for Network Interface e0b [ ]: (Press Enter)
Should interface e0b take over a partner IP address during failover? [ ]: n
Would you like to continue setup through the web interface? [ ]: n
Please enter the name or IP address of the IPv4 default gateway [ ]: (Press Enter)
Please enter the name or IP address of the administration host: (Press Enter)
Please enter the timezone [ ]: (Input the local timezone, e.g., "US/Eastern")
Where is the filer located? [ ]: (Input the controller's location for your reference)
What language will be used for multi-protocol files?: (Press Enter)
Do you want to run DNS resolver? [ ]: y
Please enter the DNS domain name. [ ]: (Input your domain name here)
Please enter the IP address for first nameserver [ ]: (Input your DNS server's IP address)
Do you want another nameserver? [ ]: (Choose 'y' and continue inputting IP addresses of up
to 3 DNS servers, if desired)
Do you want to run NIS client? [ ]: n
Would you like to configure the RLM LAN interface? [ ]: y (The RLM LAN interface is used
for out-of-band management of the controller; input 'y' to enable it)
Would you like enable DHCP on the RLM LAN interface? [ ]: n
Please enter the IP address for the RLM. [ ]: (Input the IP address to use for the RLM
interface. The port must be connected to the appropriate VLAN)
Please enter the netmask for the RLM. [ ]: (Input the netmask for the RLM interface)
Please enter the IP address for the RLM gateway. [ ]: (Input the gateway for the RLM
interface)
Please enter the name or IP address of the mail host. [ ]: (To take advantage of email
notifications, input an SMTP server address.)

```

- Step 4** Once the command prompt is presented (e.g., “3170-ucs1>”), type **reboot** to reboot the controller for the configuration to take effect.
- Step 5** Once the controllers have rebooted, ensure that all purchased licenses have been enabled. Type **license** to list current features, then **license add <code1> <code2>...** to insert any missing licenses. For a list of required licenses, see the NetApp sections in [Appendix C—Bill of Material with Software Versions](#).
- Step 6** To enable HA clustering, type **cf enable** on 3170-ucs1 and 3170-ucs3 only (i.e., one controller from each HA pair). Verify the configuration by typing **cf status** on both controllers:
- ```

3170-ucs1> cf enable
3170-ucs1> cf status
Cluster enabled, 3170-ucs2 is up.
3170-ucs2> cf status
Cluster enabled, 3170-ucs1 is up.

```
- Step 7** Verify that you can access the network from the storage controller with the **ping** command. You should also be able to ping the storage controller from other machines. If this is not the case, verify the cabling, the switch settings (LACP vPC, etc.), and controller’s vif configuration (with the **vif status** command). Once the link is functioning, all subsequent administration can be conducted via this interface.
- Step 8** You are now ready to create aggregates and volumes to hold the Oracle Binaries and data files. The setup of the aggregates and volumes is covered in [Setting up NetApp Storage for Database and Binary](#).

**Note**

When configuring from the command-line, some network configurations within NetApp require editing of the “/etc/rc” file to ensure persistence across reboots. Consult the NetApp NOW site (<http://now.netapp.com>) for more information.

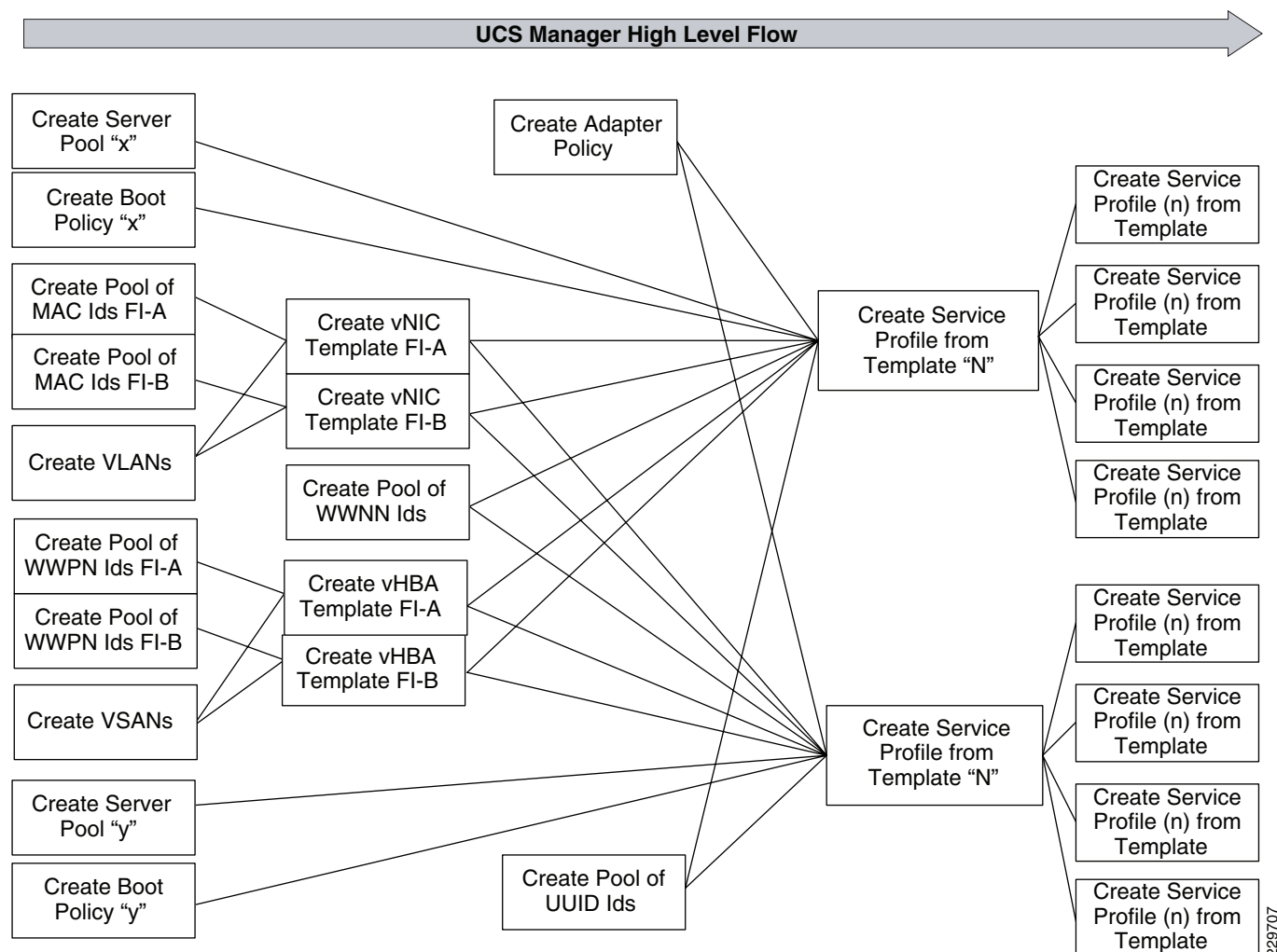
## Unified Computing System

This section describes the configuration steps for the UCS with a brief description of the rationale for each step. The initial setup of the UCS system, including cabling and initial network and chassis configuration, as well as the step-by-step procedures for each operation, are beyond the scope of this document and can be obtained from the *Cisco UCS Manager GUI Configuration Guide*: [http://www.cisco.com/en/US/partner/docs/unified\\_computing/ucs/sw/cli/config/guide/b\\_CLI\\_Config\\_Guide.html](http://www.cisco.com/en/US/partner/docs/unified_computing/ucs/sw/cli/config/guide/b_CLI_Config_Guide.html).

All steps are performed using the UCS Manager Java GUI unless otherwise specified. The best practice for implementing UCS is to first engineer elements such as organizations, resource pools, policies, and templates. This flow is shown in [Figure 6](#). There are two types of service profile templates, updating and initial. Initial templates are generally easier to manage in that changes to the original template do not result in changes to the service profiles created from the template. Updating templates results in the downstream service profiles being immediately updated with any change, which may or may not be desirable depending on the use case and the nature of the change to the template. This design used updating templates.

[Figure 6](#) shows a high-level summary of the overall flow of operations. Each action area is explained in detail in this section. The intent of [Figure 6](#) is to show which steps and actions “feed” the subsequent actions, with the flow going from left to right. It can be readily seen that the service profile template is the key construct which, once created, allows rapid creation and provisioning of new service profiles and servers. Not all of the exposed UCS policies and capabilities are shown in [Figure 6](#); it is intended to provide the reader with a guide to the overall flow process.

The final step of creating service profiles from service profile templates is quite trivial as it automatically sources all the attributes you have fed into the templates and associates the right service profiles to the correct blades, powers them on, and boots them according to the boot policies specified.

**Figure 6 UCS Manager High Level Flow**

The following sections outline the steps conducted under each operation tab in UCS Manager. This can be considered a general sequence of steps. Screen shots are shown for some of the steps to help clarify the description, but a screen-by-screen sequence is not in the scope of this document.

## Initial Infrastructure Setup

### UCS Hardware Components Used in the Architecture

A detailed description of all the hardware components is beyond the scope of this document. Detailed documentation for each UCS component can be obtained at:

[http://www.cisco.com/en/US/docs/unified\\_computing/ucs/overview/guide/UCS\\_roadmap.html](http://www.cisco.com/en/US/docs/unified_computing/ucs/overview/guide/UCS_roadmap.html).

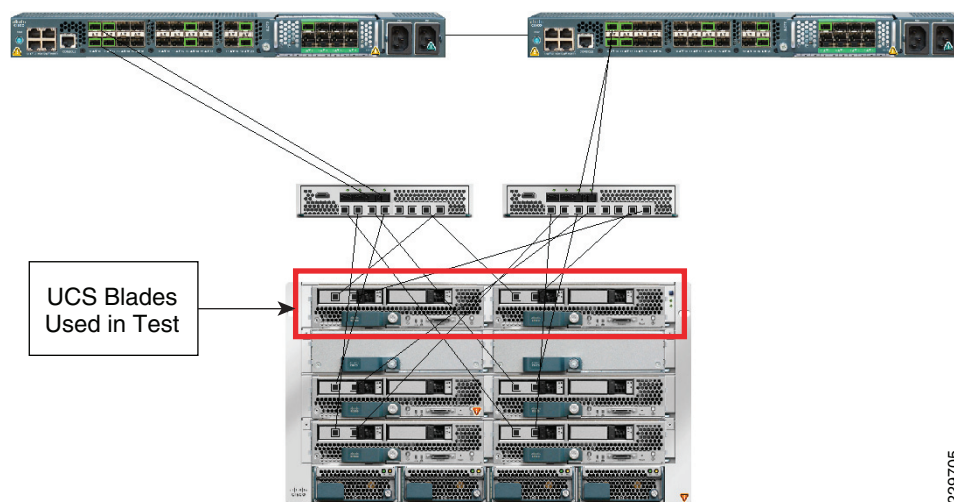
### B200 M1 Blade

The half-width blade was used for the architecture design and validation testing. The blade was populated with 48GB of memory and a single I/O mezzanine card. The 2.53 GHz CPU was used. There were four blades used in two 5108 UCS chassis.

## 6120 Fabric Interconnect

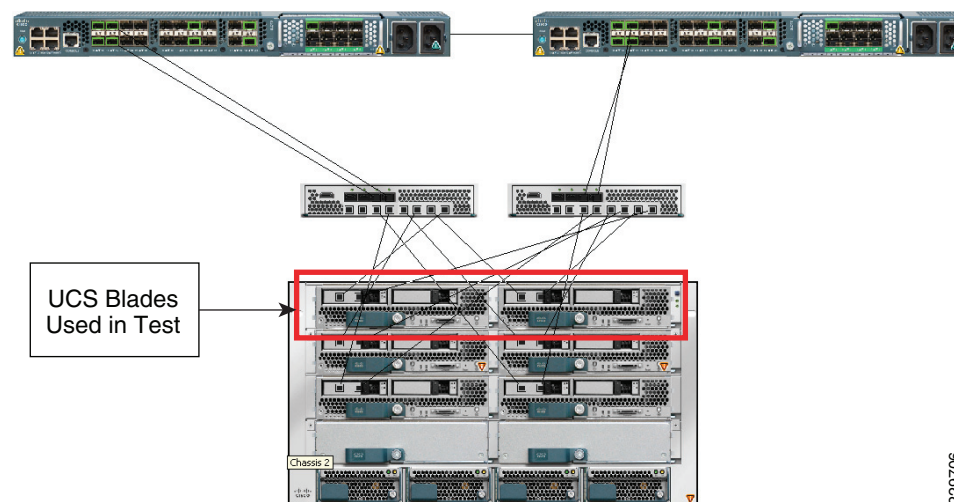
A pair of 6120 fabric interconnects were used for the testing, configured in an HA pair. The FC connections to storage were established using the global expansion modules. Some of the 20 fixed 10GbE ports were used for connecting to the upstream Nexus switches. The “hybrid display” from the UCS Manager GUI is shown here to provide a logical and physical view of the topology.

**Figure 7** 6120 Fabric Interconnect Topology—Chassis 1



229705

**Figure 8** 6120 Fabric Interconnect Topology—Chassis 2



229706

## I/O Mezzanine Card

The I/O card used for this project was the Cisco UCS M81KR Virtual Interface Card, which can present up to 128 vHBAs and 10GbE vNICs to the operating system. We assigned one port or NIC/HBA to each Fabric Interconnect (FI). The card allows seamless inclusion into standard Ethernet and Fibre Channel SAN networks.

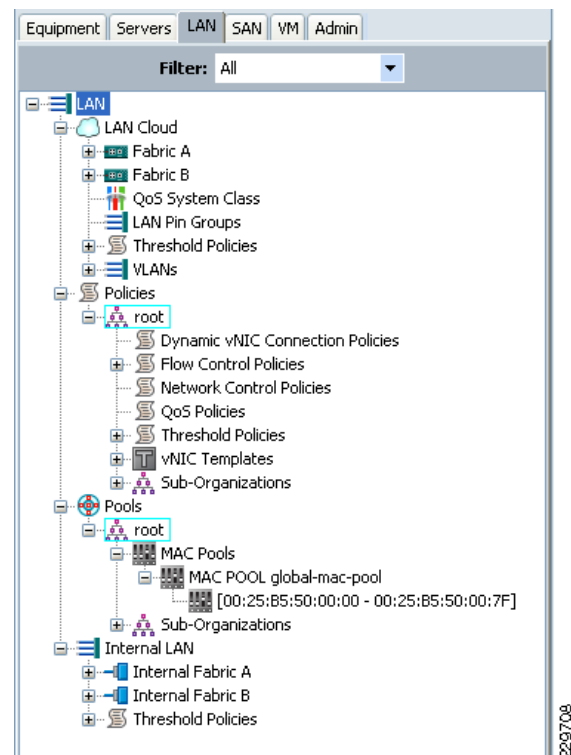
## LAN Configuration (LAN Tab)

The first step in using UCS is to leverage the concept of an organization. Once created, all subsequent steps listed below are done “under” the organization that was created versus at the top of the UCS hierarchy (root). This step is not necessary, but it allows very easy separation of resources and policies, especially if different departments (or tenants) are supported. An organization called “cisco\_ucscvd” is created, under which everything else is constructed and then automatically associated.

### MAC Pool

One global MAC pool was created and used for both Fabric Interconnects, known in UCS terminology and through the remainder of this document as fabric “A” and fabric “B”. The default OUI provided by the UCS manager is used and then the MAC pool is used to seed the different vNIC templates, as shown in [Figure 9](#), which shows an example of the MAC pool created.

**Figure 9** *MAC Pool Created*



#### Note

Two different MAC pools can be created for the two Fabric Interconnects (one each). Unique MAC pools are useful for ease of troubleshooting.

### VLANs

Three different VLANs (VLAN 178, 103, and 104) are created and used throughout the infrastructure. Creating VLANs is simple and only involves specifying a name and associated ID value. Each VLAN is assigned to each fabric such that upon a failure of either 6120 FI, the partner system is able to serve all the VLAN traffic. The VLANs are assigned to both fabrics during the creation of the vNIC templates, which allows different networks to use independent fabric resources and allows for optimal balancing of

loads across the entire system. Creating VLANs is quite trivial (a single GUI button) with UCS Manager revisions equal to or later than 1.0(2d). In the version used in this document (1.3(1i)), you can create a new VLAN under the LAN tab on the VLAN page by clicking **+ New** at the top to access the screen in [Figure 10](#).

**Figure 10**      **Create VLANs**

**Create VLAN(s)**

VLAN Name/Prefix:

☒ Common/Global   ☐ Fabric A   ☐ Fabric B   ☐ Both Fabrics Configured Differently

You are creating global VLANs that map to the same VLAN IDs in all available fabrics.

Enter the range of VLAN IDs.(e.g. "2009-2019", "29,35,40-45", "23", "23,34-45")

VLAN IDs:

Check Overlap   OK   Cancel

229708

## vNIC Templates

Under Policies, three different vNIC Templates are created to separate the public, interconnect, and storage traffic across the two fabrics. [Figure 11](#), [Figure 12](#), and [Figure 13](#) show the configuration information along with the MAC pool each template is assigned to draw from and which VLANs are available. The vNIC templates are used later during the creation of the service profile template to automatically define the connectivity model and attributes for the vNICs, since it has already been engineered at this step.

**Figure 11**      **MAC Pool and VLANS for vNIC Templates—Fabric A**

General vNIC Interfaces Faults Events

**Actions**

- Modify VLANs
- Delete

**Properties**

Name: **tom-vnic-fab-a**

Description: tom-vnic-fab-a

Fabric ID: ☒ Fabric A ☐ Fabric B ☒ Enable Failover

**Target**

☒ Adapter ☒ VM

Template Type: ☐ Initial Template ☒ Updating Template

MTU: 1500

**Policies**

MAC Pool: global-mac-pool

QoS Policy: <not set>

Network Control Policy: <not set>

Pin Group: <not set>

Stats Threshold Policy: default

| Name        | Address | VLAN | Type  | Native VLAN                      |
|-------------|---------|------|-------|----------------------------------|
| Network 178 | derived | 178  | ether | <input checked="" type="radio"/> |

229710

**Figure 12**      **MAC Pool and VLANS for vNIC Templates—Fabric B**

**General** | **vNIC Interfaces** | Faults | Events

**Actions**

- Modify VLANs
- Delete

**Properties**

Name: **tom-vnic-fab-b**

Description: private-vnic-fab-b

Fabric ID: ☐ Fabric A ☒ Fabric B ☒ Enable Failover

**Target**

- ☒ Adapter
- ☒ VM

Template Type: ☐ Initial Template ☒ Updating Template

MTU: 9000

**Policies**

MAC Pool: global-mac-pool

QoS Policy: <not set>

Network Control Policy: <not set>

Pin Group: <not set>

Stats Threshold Policy: default

| Name        | Address | VLAN | Type  | Native VLAN                      |
|-------------|---------|------|-------|----------------------------------|
| Network 103 | derived | 103  | ether | <input checked="" type="radio"/> |

229711

**Figure 13** *MAC Pool and VLANS for vNIC Templates—vNIC C*

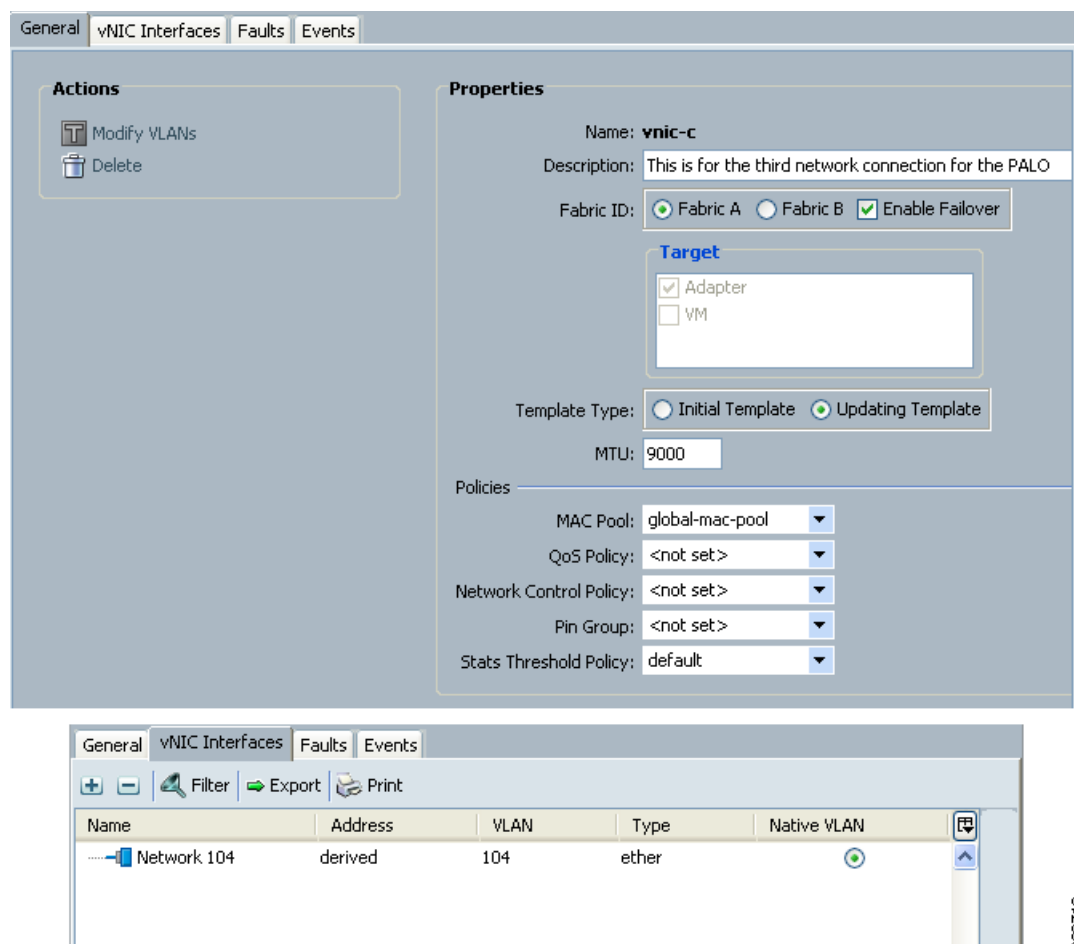
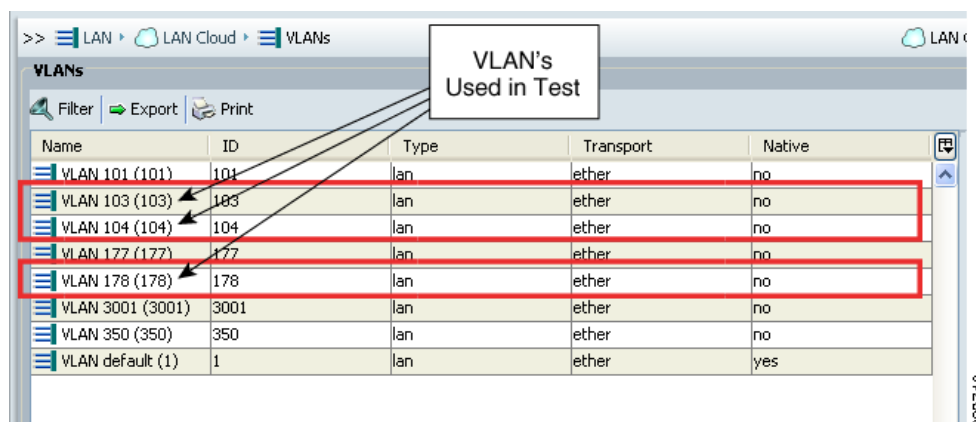


Figure 14 shows all the VLANs used in testing. From a UCS perspective, the VLANs were configured to exist on both fabrics, thus the Fabric ID is “dual” for all of them.

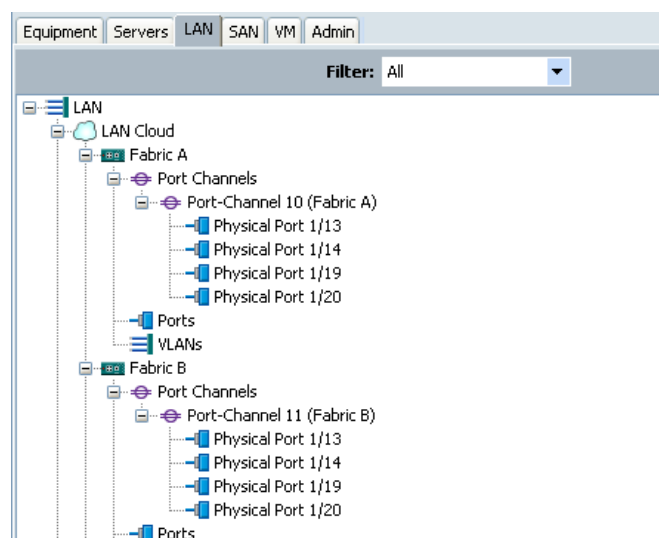
**Figure 14** *VLANs Used in Testing*



## Network Connectivity via Port Channels

A port channel is created and enabled for each fabric using specific uplinks ports on the respective 6120. The port channel is global in scope and not associated to a specific organization. The port channels then connected to upstream Nexus 5000s as shown in [Figure 15](#). The default mode for the UCS port channels is LACP active-active.

**Figure 15** Port Channel Configuration



## Configuring Jumbo Frames

When configuring the system for jumbo frames (MTU sizes greater than the standard 1500) adjustments are made to each of the components (blade OS, UCS, switches, and storage). The test configuration uses frame sizes of 9000 for connectivity to the storage. The M81KR virtual interface cards used will support MTUs up to 14000, while the CNA cards will support values up to 9216. You should make sure all components support the MTU you choose; this system utilized the common 9000 MTU setting that is widely supported. The use of Jumbo frames in an NFS environment can improve performance and allow the system to reach the full potential of the equipment.

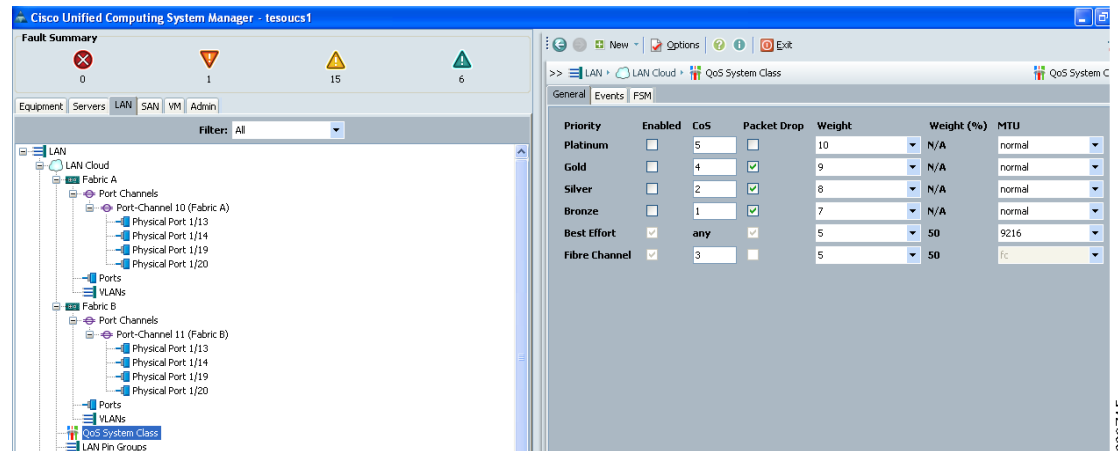
- Step 1** Adjust Nexus5K ports—Adjust the frame sizes on the Nexus 5000 switch to support Jumbo frames. This should be done on both Nexus 5K switches.

```
N5k_f22(config)# policy-map type network-qos jumbo
N5k_f22 (config-pmap-nq)# class type network-qos class-default
N5k_f22 (config-pmap-c-nq)# mtu 9216
N5k_f22 (config-pmap-c-nq)# exit
N5k_f22 (config-pmap-nq)# exit
N5k_f22 (config)# system qos
N5k_f22 (config-sys-qos)# service-policy type network-qos jumbo
```

- Step 2** Modify UCS MTU Settings—Adjust UCS MTU settings for the vNIC in the template (see [Figure 12](#)) and in the QoS System Class folder to 9000 or higher (see [Figure 16](#) for QoS setting).

1. Change on vNIC (change for template). Set MTU to 9000.
2. Change on QoS Service Class (best available). Set MTU to 9000.

**Figure 16** MTU Setting in the QoS Area



**Note**

The QoS definitions are just the standard settings as this architecture does not require QoS mappings.

- Step 3** Modify Storage Interface Settings—Log onto FilerView for each of the controllers and navigate to the Network folder. Click **Manage Interfaces** to get a view of the configured network interfaces. Next click **Modify** beside the vif1 VIF you setup on install. On that tab adjust the MTU setting per [Figure 17](#). Alternatively this can be set at the CLI by using the following command, then adding it to the /etc/rc file to make it persistent. (The GUI adds necessary code to /etc/rc automatically):

```
3170-ucs1> ifconfig vif1 mtusize 9000
```

Figure 17 FilerView MTU Settings

**FilerView®**

Manage  
Configure RAID  
• **Storage** ?  
• **Operations Manager** ?  
• **NFS** ?  
• **HTTP** ?  
• **Network** ?  
Report  
Configure  
Manage Interfaces  
Add Virtual Interface  
Manage Hosts File  
Manage Net Groups  
Configure Host Name Resolution (DNS & NIS)  
• **Security** ?  
• **Secure Admin** ?  
• **NDMP** ?  
• **SNMP** ?  
• **Cluster** ?  
• **Real Time Status** ?  
• **Wizards** ?

### Modify Network Interface ?

Network → Modify Network Interface

Interface: **vif1** Status: Up Type: Virtual Interface

**IPv4 Address:**  
Enter the IPv4 Address address of this interface. 192.168.101.25 ?

**Netmask:**  
Enter the subnet mask for this interface. 255.255.255.0 ?

**Broadcast:**  
Enter the Broadcast address of this interface. If left blank, the default is used. The current broadcast address is on. ?

**Media Type:** ?  
Virtual Interfaces don't have a media type. However, you can modify the media type on the individual interfaces within the virtual interface.

**MTU size:**  
Enter the MTU size (in bytes) for this interface. The default value is 1500. 9000 ?

**Trusted:** ☒ Trust ?  
Check this box if you want this interface to be trusted.

**Partner** ?

Done

- Step 4** Modify Linux Network Interface—Log in using root and adjust the MTU setting for the interface associated with the storage network, which in our case is eth1. To set the MTU in real time you can use: **ifconfig eth1 mtu 9000**. Then to make the change persistent, you should modify the `/etc/sysconfig/network/network-scripts/ifcfg-eth1` to change or add a line that reads `MTU=9000`. Alternatively to the ifconfig step, you can run **service network restart** after changing this file to update the MTU setting.

```
Example ifcfg-eth1 from test configuration:
DEVICE=eth1
BOOTPROTO=static
BROADCAST=192.168.101.255
HWADDR=00:25:B5:50:00:7F
IPADDR=192.168.101.103
NETMASK=255.255.255.0
NETWORK=192.168.101.0
MTU=9000
ONBOOT=yes
```

## SAN Configuration (SAN Tab)

### World Wide Node Name (WWNN) Pool

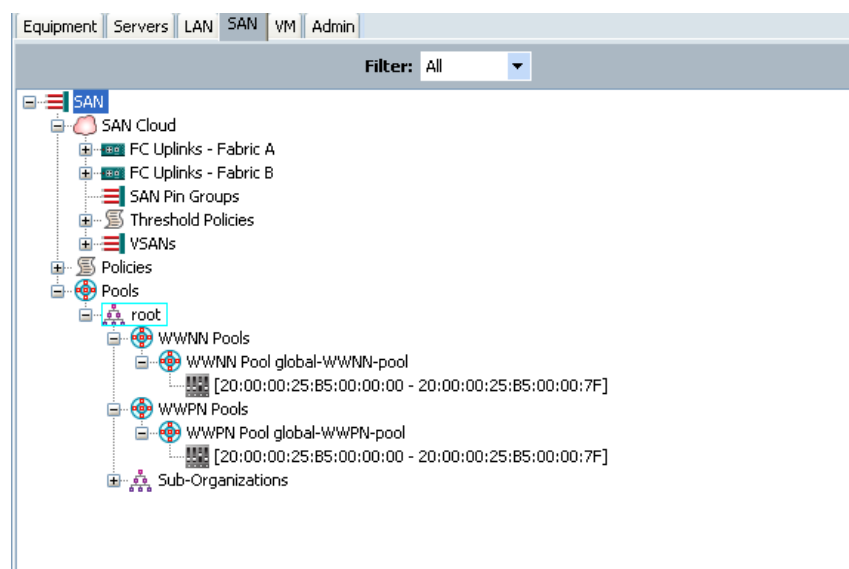
A single WWNN pool is created for all the blades in the system. As this pool is for the parent FC device node (the card itself), it is not necessary to create one for each fabric. This can be seen in [Figure 18](#).

## World Wide Port Name (WWPN) Pools

A single WWPN pool was created for all the blades in the system. The WWPNs are assigned to the ports when the vHBA templates are created. Use the best practice of starting the OUI prefix with the 20 convention, which is common for host initiators in Fiber Channel SANs.

The naming convention and specific values of the WWPN are coordinated with the NetApp storage controller such that the host WWPNs would be defined in their initiator groups, resulting in the correct LUN masking assignments. This is important not only for booting from SAN, but for general data access as well.

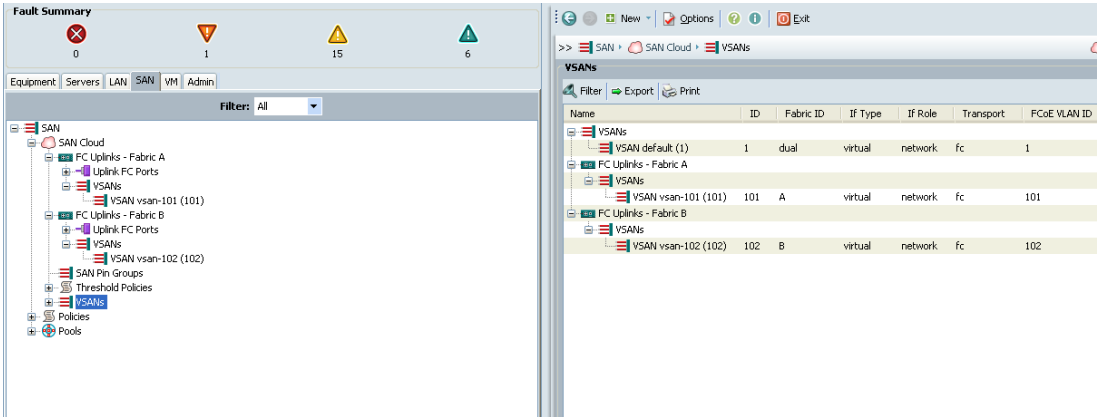
**Figure 18** *World Wide Port Name Pool Configuration*



## VSANs

Two VSANs are created in addition to the default to allow FC traffic engineering to occur, assigning different VSANs to different upstream MDS FC switches. This also allows manual balancing of boot traffic across the two different NetApp storage controllers. FC storage is only used in this architecture for boot traffic and the root partition. This architecture utilized a separate set of controllers for the FC boot, however that was due to the nature of the shared test environment in which the tests were conducted. This is not a requirement and the same NetApp controllers used for data storage could also be utilized for the SAN boot by creating LUNs. This is made possible by the multi-protocol support of the NetApp storage controllers.

Figure 19 VSAN Configuration

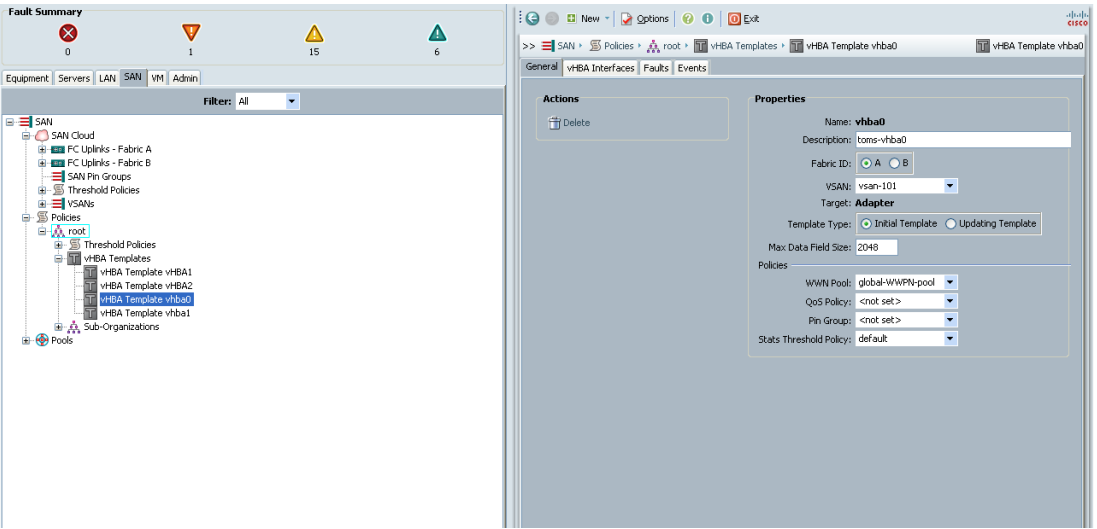


229718

vHBA Templates

Two vHBA templates are created, one for each fabric. The template is directed to draw upon the corresponding WWPN pool described earlier.

Figure 20 vHBA Template Configuration

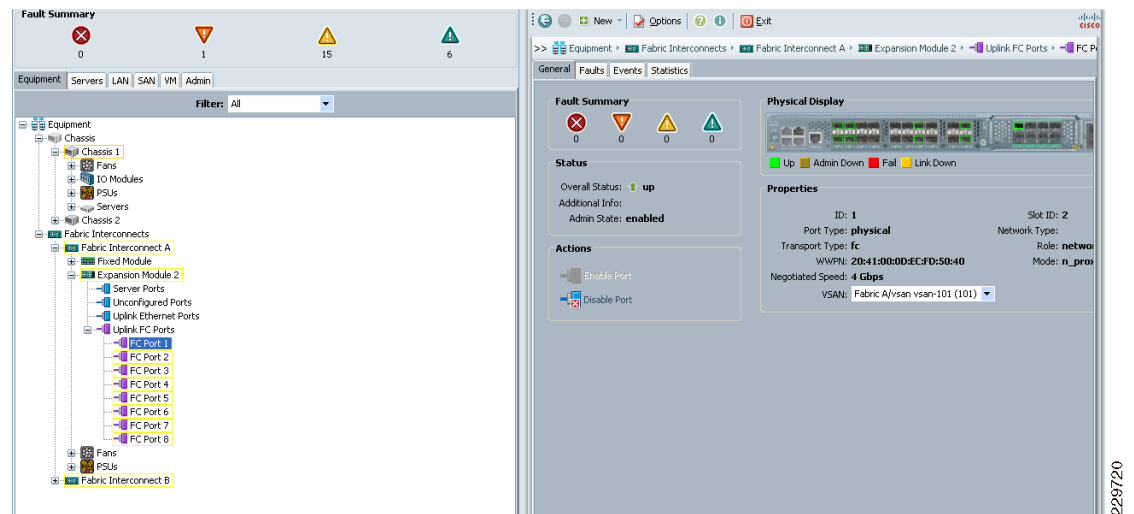


229719

FC Uplink to VSAN Assignment (Equipment Tab)

In the UCS-M equipment tab, each FC uplink used in the configuration is assigned to the correct VSAN.

**Figure 21 FC Uplink to VSAN Assignment Configuration**

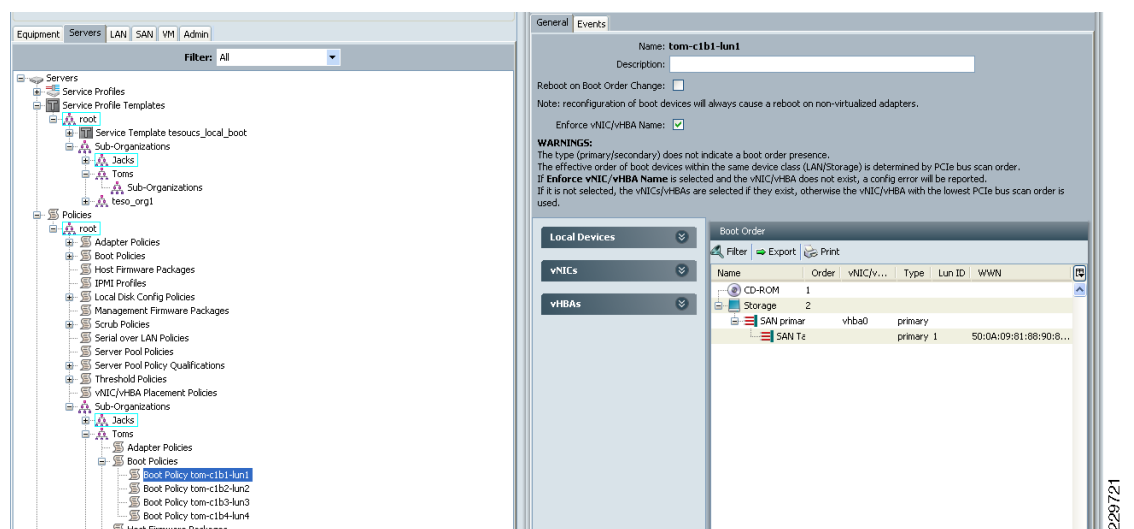


## Server Configuration (Servers Tab)

### Boot Policies

Four different boot policies were created to facilitate the balancing of booting the blades across both controllers. The boot policies specify an HBA to use in a primary/secondary concept, which target the WWN to connect to and then which LUN ID to use. With larger numbers of blades to balance you may want to create Server Pools first, then create a boot policy for each server pool.

**Figure 22 Boot Policy #1 Configuration**



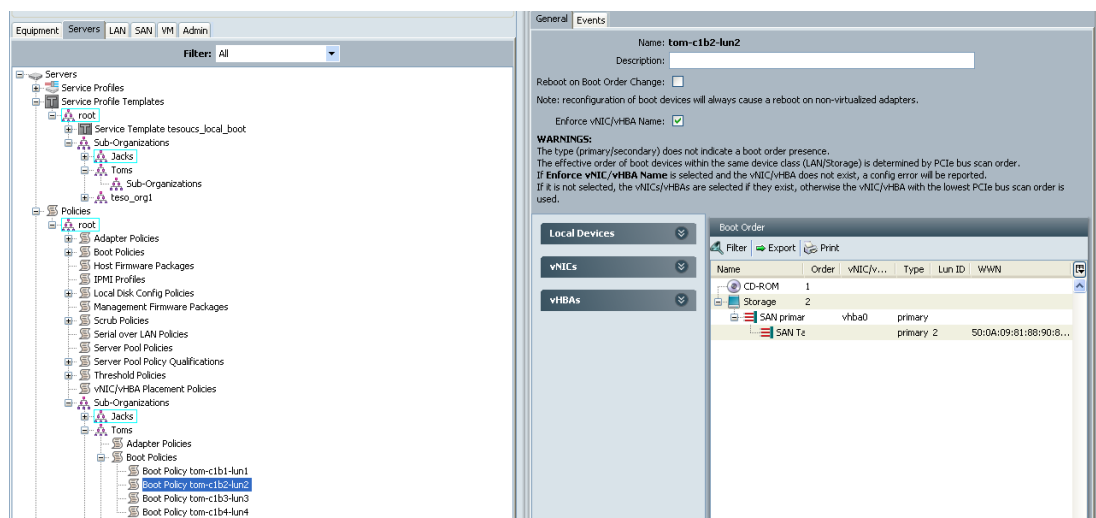
**Figure 23**      **Boot Policy #2 Configuration**

Figure 22 and Figure 23 show the boot path. For the purposes of the test environment we only utilized a primary boot path. However, in a production environment we recommend you define a secondary boot path as well. This will instruct the BIOS on a given server to use the designated WWPN target first for the boot location and, if not available, to fall back to the secondary. Having different policies allows for balancing of the boot traffic across fabric interconnects.

Later when a service profile template and associated individual service profile are created, they use this boot policy, as shown in Figure 24.

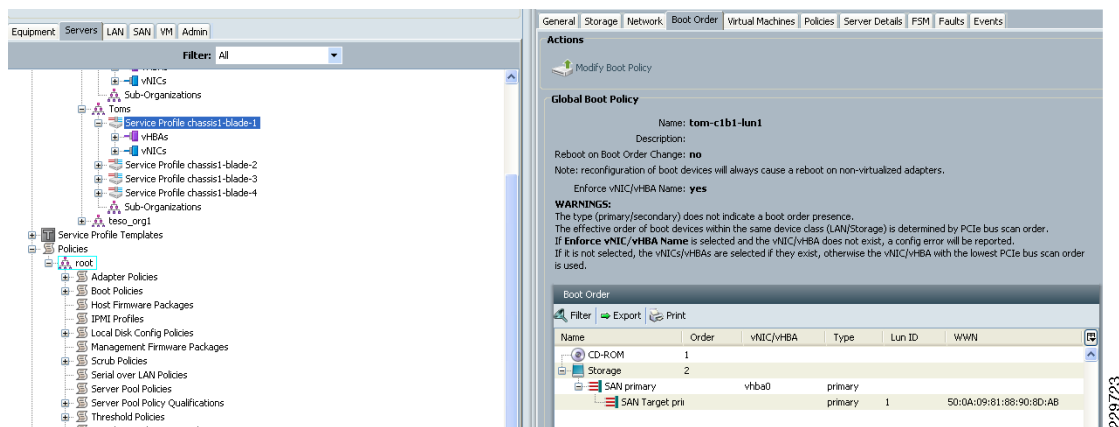
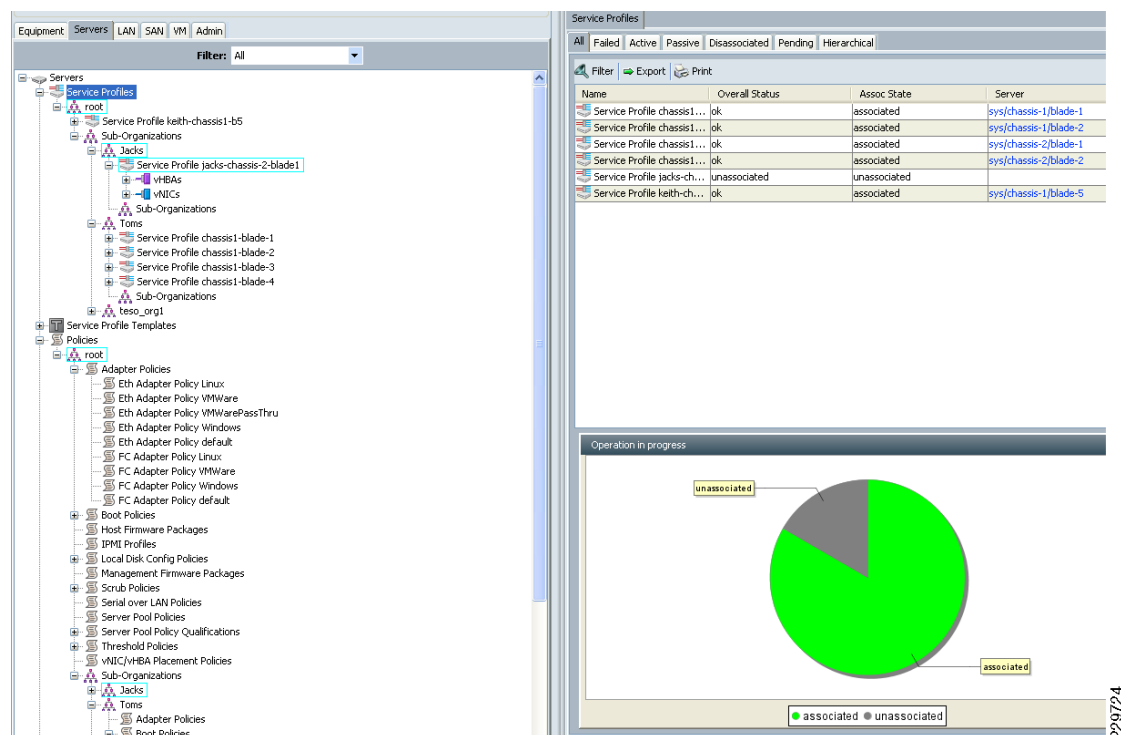
**Figure 24**      **Service Profile Template Configuration for Boot Policy**

Figure 25 shows a summary of the system and the associations between service profiles and blades. The blue “sys/chassis-1/blade-x” next to a service profile shows that the profile is active on that indicated physical blade.

**Figure 25**      **Service Profile Summary**

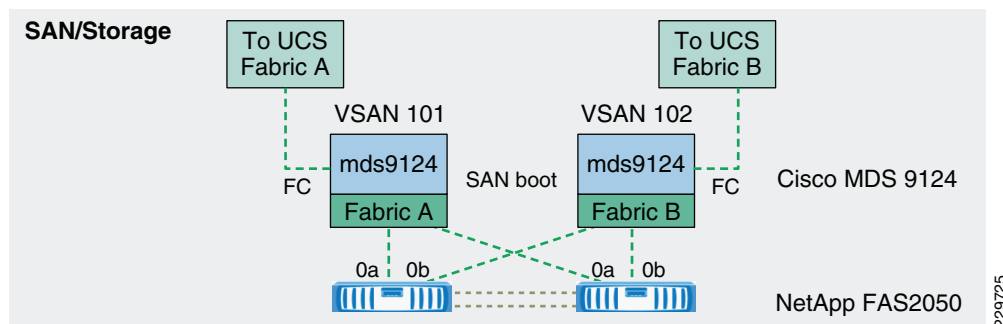
## SAN Boot Setup

This section details the configuration steps involved in setting up the boot fabric for the SAN-booted RHEL hosts in the environment. SAN design best practices dictate redundancy in the fabric in both link and path. Therefore, the configuration steps are performed on two MDS switches to ensure redundancy and high availability. Fabric A of the UCS is linked to the primary MDS switch, while fabric B is connected to the secondary MDS switch. VSANs are used to create a logical segmentation of the physical infrastructure: the primary fabric is configured with access to VSAN 101, the secondary fabric with VSAN 102. The NetApp controllers have connectivity to both fabrics in an active/passive configuration. For load balancing purposes, you can have the hosts boot off of alternating NetApp controllers. The cabling and VSAN layout are illustrated in [Figure 26](#).



### Note

While SAN boot provides statelessness, configuring SAN boot is optional. Customers can select to install the RHEL on the local drives on B200 Servers.

**Figure 26 SAN Boot Setup**

## Configuration Procedure

With respect to SAN fabrics, there may be many variables that are unique to each customer environment. Therefore, initial setup of the SAN switch is not covered in this document. Rather, aspects specific to the SAN boot are detailed. If an action is to be performed on redundant equipment, then it is indicated as such. Unless otherwise noted, configuration examples are only given for fabric A. For clarity, each step includes a prefix to denote the equipment being referenced.

### Ensure SAN Connectivity

Verify connectivity within the SAN on the UCSM, MDS switch, and NetApp controller outlined in the following steps:

#### Step 1 (UCSM) Verify connectivity to the MDS fabric.

Earlier, Fibre Channel uplink ports were configured for the primary and secondary VSANs in UCSM. Verify that uplink ports are “up” under the SAN tab within UCSM.

#### Step 2 (NetApp) Enable connectivity to the MDS fabric.

Fibre channel has been licensed in the initial setup. Use the **fcp start** command to enable connectivity on both NetApp controllers:

```
NetApp1> fcp start
NetApp1> fcp status
 FCP service is running
```

#### Step 3 (MDS) Ensure connectivity to the UCS fabric interconnect.

Enable NPIV on both fabric switches.

```
mds9124-fabA# config t
mds9124-fabA(config)# npiv enable
```

#### Step 4 (MDS) Create VSANs to be used in the environment and associate target and server ports.

In this example, ports fc1/1,5,6 on each switch correspond to the both UCS and both NetApp connections. VSAN 101 is used for the primary fabric and VSAN 102 for the secondary.

```
mds9124-fabA# config t
mds9124-fabA(config)# vsan database
mds9124-fabA(config-vsan-db)# vsan 101 name fc_boot_primary
mds9124-fabA(config-vsan-db)# vsan 101 interface fc1/1,fc1/5,fc1/6
mds9124-fabB# config t
mds9124-fabB(config)# vsan database
```

```
mds9124-fabB(config-vsan-db)# vsan 102 name fc_boot_secondary
mds9124-fabB(config-vsan-db)# vsan 102 interface fc1/1,fc1/5,fc1/6
```

- Step 5** (MDS) Configure ports to be used in the environment and assign descriptions on both fabric switches. In this example, NetApp FCP port 0a is used for the primary fabric and 0b is used for the secondary fabric.

```
mds9124-fabA(config)# int fc1/1
mds9124-fabA(config-if)# switchport description ucs-fabA-fc2/1
mds9124-fabA(config-if)# int fc1/5
mds9124-fabA(config-if)# switchport description 2050san-1-0a
mds9124-fabA(config-if)# int fc1/6
mds9124-fabA(config-if)# switchport description 2050san-2-0a
```

- Step 6** (MDS) Verify connectivity on both fabric switches.

```
MDS9124-2# show fcns database
```

```
VSAN 101:
```

```

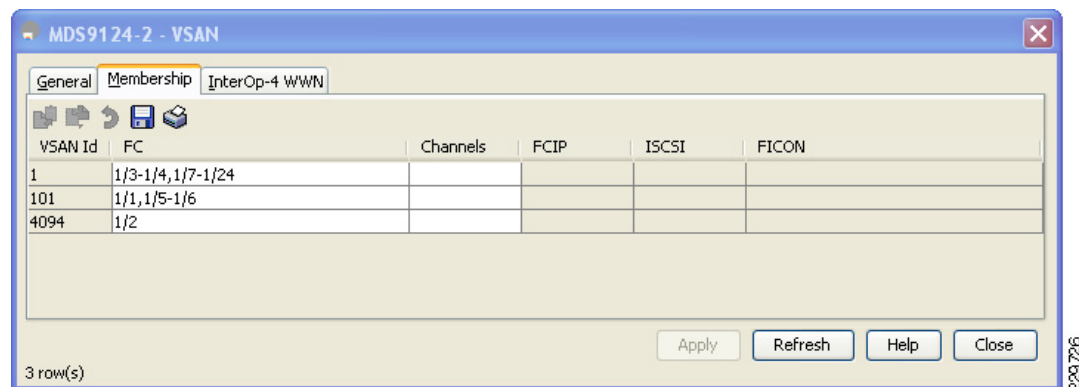
FCID TYPE PWWN (VENDOR) FC4-TYPE:FEATURE

0xc90000 N 20:41:00:0d:ec:fd:50:40 (Cisco) npv
0xc90001 N 20:00:00:25:b5:00:00:7f scsi-fcp fc-gs
0xc90002 N 20:00:00:25:b5:00:00:2f scsi-fcp fc-gs
0xc90003 N 20:00:00:25:b5:00:00:1f scsi-fcp fc-gs
0xc90004 N 20:00:00:25:b5:00:00:4e scsi-fcp fc-gs
0xc90007 N 20:00:00:25:b5:00:00:0d scsi-fcp fc-gs
0xc90700 N 50:0a:09:81:88:90:8d:ab scsi-fcp
0xc90800 N 50:0a:09:81:98:90:8d:ab scsi-fcp
Total number of entries = 8
MDS9124-2#
```

## Configure Boot Path on MDS Switch

Configure the primary boot path for the RHEL hosts. Utilizing the GUI interface on the MDS-9124 switches, you can assign the associated ports as seen in [Figure 27](#). Port 1/1 is for connection to the 6120 interconnect and 1/5-1/6 are for connection to the storage controllers. You would follow a similar process for the second MDS switch.

**Figure 27** MDS-9124 Interface



## Configure Boot Target Information

Create target LUNs and map them on the NetApp storage arrays:

- Step 1** (NetApp) Create initiators for both vHBA ports of all hosts in the environment. Load balancing is employed for target configuration: half the hosts use the first controller and half use the second.

Command syntax: **igroup create** { **-f** | **-i** } **-t** <ostype> [ **-a** <portset> ] <initiator\_group> [ <node> ... ]

```
2050san-1> igroup create -f -t linux chassis1-b1 WWPN_of_initiator_c1b1
2050san-1> igroup create -f -t linux chassis1-b2 WWPN_of_initiator_c1b2
2050san-1> igroup create -f -t linux chassis2-b1 WWPN_of_initiator_c1b3
2050san-1> igroup create -f -t linux chassis2-b2 WWPN_of_initiator_c1b4
```

- Step 2** (NetApp) Create boot storage for all hosts that will attach to the given target.

Command syntax: **lun create** **-s** <size> **-t** <ostype> [ **-o** noreserve ] [ **-e** space\_alloc ] <lun\_path>

```
2050san-1> lun create -s 40g -t linux -o noreserve /vol/sanboot/chassis1-b1
2050san-1> lun create -s 40g -t linux -o noreserve /vol/sanboot/chassis1-b2
2050san-1> lun create -s 40g -t linux -o noreserve /vol/sanboot/chassis2-b1
2050san-1> lun create -s 40g -t linux -o noreserve /vol/sanboot/chassis2-b2
```

- Step 3** (NetApp) Map initiators for a given host to the allocated boot storage with a lun id of 0.

Command syntax: **lun map** [ **-f** ] <lun\_path> <initiator\_group> [ <lun\_id> ]

```
2050san-1> lun map /vol/sanboot/chassis1-b1 chassis1-b1 0
2050san-1> lun map /vol/sanboot/chassis1-b2 chassis1-b2 0
2050san-1> lun map /vol/sanboot/chassis2-b1 chassis1-b1 0
2050san-1> lun map /vol/sanboot/chassis2-b2 chassis1-b2 0
```

## Setting up NetApp Storage for Database and Binary

This section of the document provides a general overview of the storage configuration for the database layout, database binary, clusterware binary, OCR file, and voting disk. It also discusses the virtual interface (VIFs) configuration used to achieve high availability and load balancing access from the database host to the NetApp storage. We have used two cluster pair of FAS-3170 to test the entire setup. For more information about NetApp FAS storage, refer to <http://www.netapp.com>.

The NetApp storage model in Table 4 was used for this testing.

**Table 4** *FAS-3170 Cluster A and FAS-3170 Cluster B*

| Cluster Name | Controller   | Disk Array / Number of Disks                                 | Total Storage Size | Flash Cache Size |
|--------------|--------------|--------------------------------------------------------------|--------------------|------------------|
| Cluster A    | FAS-3170_A_1 | Array_A1, Array_A2 / 48 Disks (410GB 15K RPM SAS drive each) | 19 TB              | 512 GB           |
| Cluster A    | FAS-3170_A_2 | Array_A3, Array_A4 / 48 Disks (410GB 15K RPM SAS drive each) | 19 TB              | 512 GB           |
| Cluster B    | FAS-3170_B_1 | Array_B1, Array_B2 / 48 Disks (410GB 15K RPM SAS drive each) | 19 TB              | 512 GB           |
| Cluster B    | FAS-3170_B_2 | Array_B3, Array_B4 / 48 Disks (410GB 15K RPM SAS drive each) | 19 TB              | 512 GB           |

**Table 5** Maximum Values for Selected System Elements

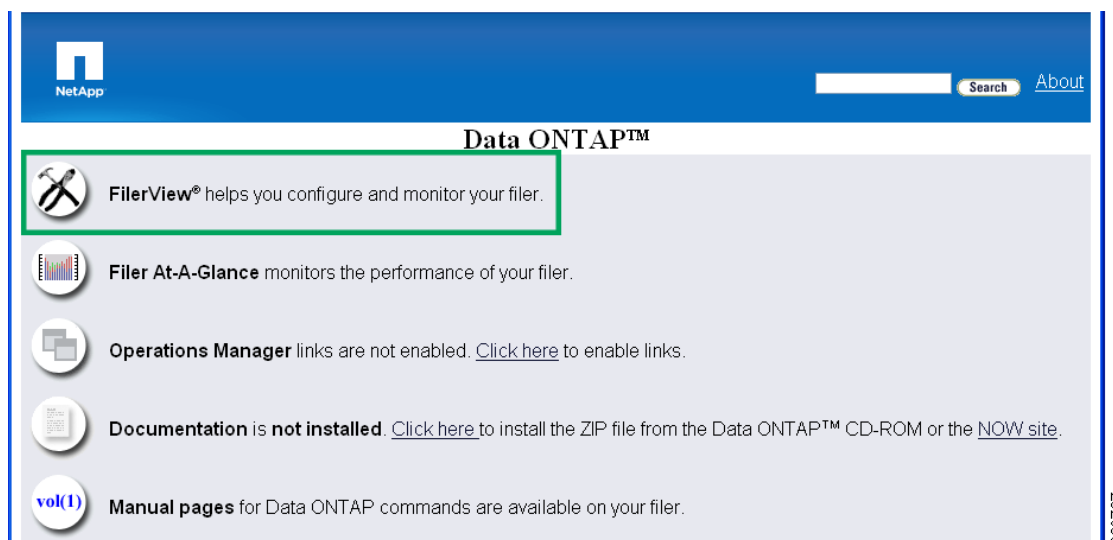
| Element Value    | Per Storage Controller                              |
|------------------|-----------------------------------------------------|
| Snapshot® copies | No more than 10 times the number of FlexVol volumes |
| CPU utilization  | No greater than 50%                                 |
| Disk utilization | No greater than 50%                                 |

## Network Configuration of Storage for Database Access

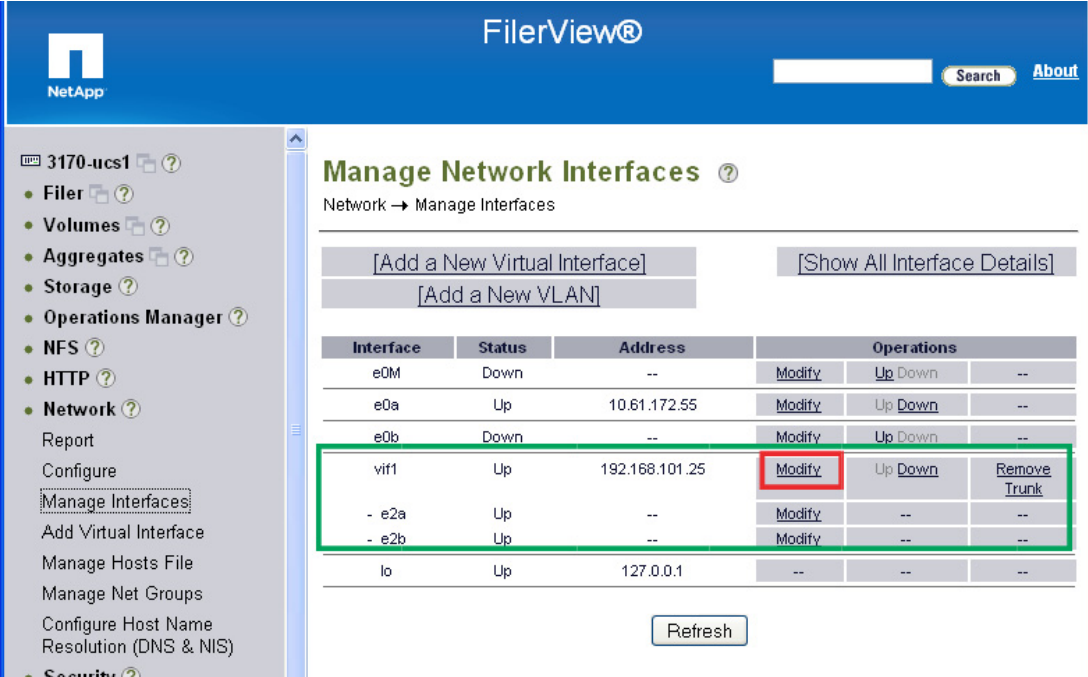
In order to setup, manage, or view the configuration, you can utilize the FilerView tool available as part of OnTap. You can access FilerView with:

`http://IP Address of Controller/na_admin`

Then click **FilerView** as shown in [Figure 28](#).

**Figure 28** FilerView

To achieve network high availability and load balancing, virtual interface (VIF) should be configured during initial setup. The VIFs use multiple physical ports of the NetApp storage for the database access from the host. If modifications need to be made or to just view the configuration, you can select **Network** and **Manage Network Interfaces** from the menu, as shown in [Figure 29](#).

**Figure 29** *Manage Network Interfaces*


**Manage Network Interfaces** ?

Network → Manage Interfaces

[Add a New Virtual Interface] [Show All Interface Details]

[Add a New VLAN]

| Interface | Status | Address        | Operations |         |              |
|-----------|--------|----------------|------------|---------|--------------|
| e0M       | Down   | --             | Modify     | Up Down | --           |
| e0a       | Up     | 10.61.172.55   | Modify     | Up Down | --           |
| e0b       | Down   | --             | Modify     | Up Down | --           |
| vif1      | Up     | 192.168.101.25 | Modify     | Up Down | Remove Trunk |
| - e2a     | Up     | --             | Modify     | --      | --           |
| - e2b     | Up     | --             | Modify     | --      | --           |
| lo        | Up     | 127.0.0.1      | --         | --      | --           |

[Refresh]

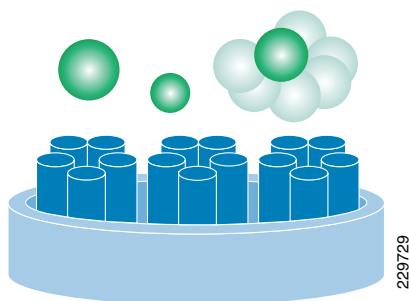
By clicking the **Modify** button of the virtual interface, you can change IP address, subnet mask, etc.; click **Apply** to complete the virtual interface configuration.

**Note**

The VIF configuration is required in each NetApp storage controller connected to the UCS and utilizing vPC.

## Create Aggregates and Volumes on the NetApp Storage

An aggregate consists of a pool of many disk drives from which space is allocated to volumes. You must specify a number of disks to be added to the new aggregate and you can optionally specify a list of specific disks that are to be added to the aggregate.

**Figure 30** *Aggregate Is of a Pool of Many Disks From Which Space Is Allocated to Volumes*

When you create an aggregate, you can control the size of a RAID group. Generally, larger RAID groups maximize your data storage space by providing a greater ratio of data disks to parity disks. For information on RAID group size guidelines, see Considerations for sizing RAID groups

(<http://now.netapp.com/NOW/knowledge/docs/ontap/rel713/html/ontap/mgmtsag/4raid2.htm#1236408>). When creating an aggregate, you can specify the aggregate's RAID type (RAID-DP, RAID4) for RAID groups on the aggregate and the maximum number of disks that can be included in a RAID group. Data ONTAP supports two levels of RAID protection, which you can assign on a per-aggregate basis: RAID4 and RAID-DP. For more information on RAID4 and RAID-DP, see Levels of RAID protection (<http://now.netapp.com/NOW/knowledge/docs/ontap/rel713/html/ontap/mgmtsag/4raid2.htm#1275615>).

Syntax for creating an aggregate:

```
aggr create <aggr-name>
 [-r <raid-group-size>]
 [-t {raid4 | raid_dp}] [-v] <disk-list>
Storage1> aggr create aggr1 -r 21 -t raid_dp -v d1 d2 d3 d4
```

Create an aggregate to hold all binary and database files. To do so, first type **aggr status -s** to determine the number of spare disks currently available. Best practice calls for no fewer than two spare disks per controller plus one additional spare for every 56 disks on the controller. For example, if 48 disks (two shelves) are attached, then two disks should be set aside as spares plus one additional. The root volume always occupies three disks, leaving 42 disks for the aggregate. If 96 disks (four shelves) are attached, then an additional spare would be needed, giving a total of four spares, leaving three disks for the root volume and 89 disks for the aggregates. Given this, type **aggr create <aggregate-name> -r <raid-grp-size> -t <raid-type> <number-of-disks>** to create the aggregate on each controller. The following example assumes a two-shelf configuration:

```
3170-ucs1> aggr create AGGR_ORA_1_A_1 -r 21 -t raid_dp 42
3170-ucs1> aggr status
 Aggr State Status Options
 aggr0 online raid_dp, aggr root
 AGGR_ORA_1_A_1 online raid_dp, aggr raidsize=21
```

Alternatively you can create the aggregate using the FilerView Wizard. The aggregates used for this test environment can be seen in [Figure 31](#).

**Figure 31** Manage Aggregates

**FilerView®**

NetApp

Search About

3170-ucs1

- Filer
- Volumes
- Aggregates
  - Add
  - Manage
  - Configure RAID
- Storage
- Operations Manager
- NFS
- HTTP
- Network
- Security
- Secure Admin

### Manage Aggregates

Aggregates → Manage

Filter by: All Aggregates View

| Name                                    | Status         | Root | Avail   | Used | Total  | Disks | Files | Max Files | Checksums |
|-----------------------------------------|----------------|------|---------|------|--------|-------|-------|-----------|-----------|
| <input type="checkbox"/> AGGR_ORA_1_A_1 | online,raid_dp |      | 8.87 TB | 31%  | 13 TB  | 42    | 118   | 31.1 k    | block     |
| <input type="checkbox"/> aggr0          | online,raid_dp | ✓    | 15.9 GB | 95%  | 349 GB | 3     | 106   | 31.1 k    | block     |

Select All - Unselect All Online Restrict Offline Destroy

Aggregates: 1-2 of 2

Refresh

2297/31

Table 6 shows the aggregates created on each of the FAS3170 storage controllers for installing the Oracle DSS/OLTP database including the aggregate name, RAID group type and size, the usable capacity, and the purpose for the specific aggregate.

**Table 6**      **Aggregate Layout**

| Controller   | Aggregate Name | Option /RG Size | # of Disks/ Usable Size | Purpose                                                                |
|--------------|----------------|-----------------|-------------------------|------------------------------------------------------------------------|
| FAS-3170_A_1 | aggr0          | RAID-DP, RG-16  | 3 no's/350GB            | DOT and root volume                                                    |
| FAS-3170_A_1 | AGGR_ORA_1_A_1 | RAID-DP, RG-21  | 42 no's/13TB            | Data files, Redo logs, control files, Database and Cluster Ware Binary |
| FAS-3170_A_2 | aggr0          | RAID-DP, RG-16  | 3 no's/350GB            | DOT and root volume                                                    |
| FAS-3170_A_2 | AGGR_ORA_2_A_1 | RAID-DP, RG-21  | 42 no's/13TB            | Data files, Redo logs, control files, FRA                              |
| FAS-3170_B_1 | aggr0          | RAID-DP, RG-16  | 3 no's/350GB            | DOT and root volume                                                    |
| FAS-3170_B_1 | AGGR_ORA_1_B_1 | RAID-DP, RG-21  | 42 no's/13TB            | Data files, Redo logs, control files                                   |
| FAS-3170_B_2 | aggr0          | RAID-DP, RG-16  | 3 no's/350GB            | DOT and root volume                                                    |
| FAS-3170_B_2 | AGGR_ORA_2_B_1 | RAID-DP, RG-21  | 42 no's/13TB            | Data files, Redo logs, control files, Archive Log                      |

### Flexible Volume (FlexVol)

A FlexVol volume is a volume that is loosely coupled to its containing aggregate. A FlexVol volume can share its containing aggregate with other FlexVol volumes. Thus, a single aggregate can be the shared source of all the storage used by all the FlexVol volumes contained by that aggregate. Since a FlexVol volume is managed separately from the aggregate, you can create small FlexVol volumes (20 MB or larger) and you can increase or decrease the size of FlexVol volumes in increments as small as 4 KB. By using FlexVol volumes, you can separately manage file systems stored in an aggregate.

Syntax for creating a volume:

```
vol create f_vol_name aggr_name size{k|m|g|t}
```

For example:

```
Storage1> vol create datavol aggr1 size 20g
```

On each controller, use the **vol create** command to create volumes within aggregate (For example: 3170-ucs1 uses aggregate AGGR\_ORA\_1\_A\_1) to contain the nfs shares.

For example (Volumes for 3170-ucs1):

```
3170-ucs1> vol create CRS_HOME -s volume AGGR_ORA_1_A_1 100g
3170-ucs1> vol create ORA_HOME -s volume AGGR_ORA_1_A_1 100g
3170-ucs1> vol create OCR_CSS -s volume AGGR_ORA_1_A_1 50g
3170-ucs1> vol create VOL_DATA_1_A_1 -s volume AGGR_ORA_1_A_1 9400g
3170-ucs1> vol create VOL_LOG_1_A_1 -s volume AGGR_ORA_1_A_1 200g
```

These volumes have volume space reserve (**-s volume**) and have a maximum capacity defined. Repeat this process on 3170-ucs2...4 using the storage layout in [Figure 32](#).

Alternatively you can again use the FilerView GUI to create the volumes.

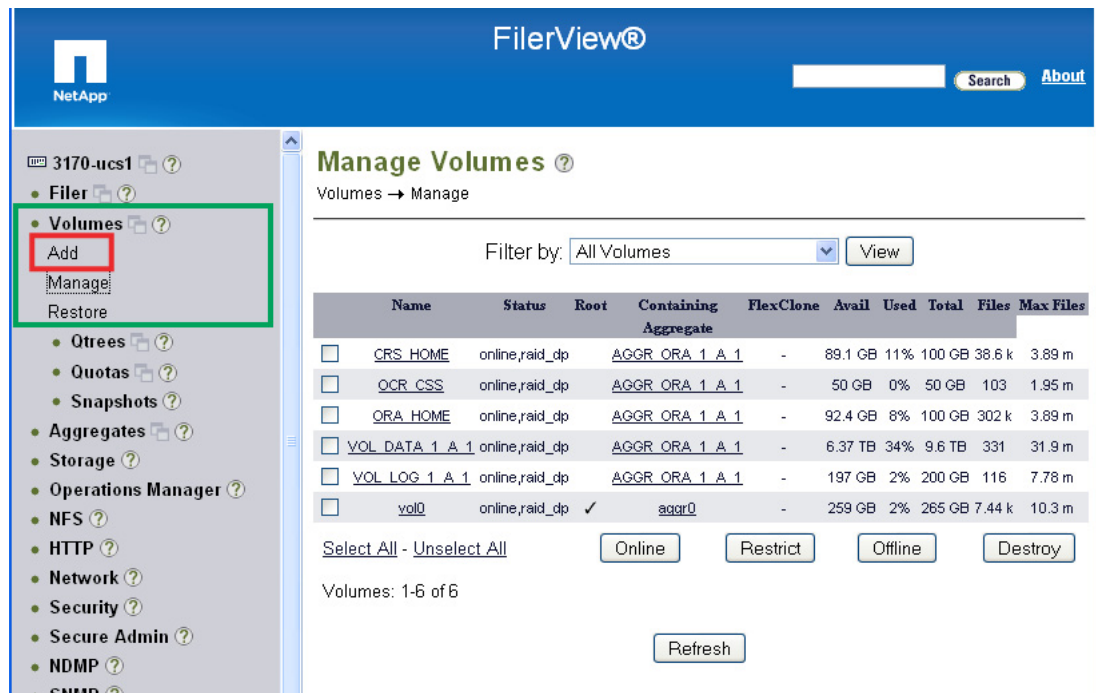
**Figure 32**      **Manage Volumes**

Table 7 shows the volumes created for the test environment. It includes the controller on which the volume is created, volume name and containing aggregate, volume size, and the purpose associated with the specific volume.

**Table 7**      **Volume Layout**

| Controller   | Volume name    | Aggregate Name | Size  | Purpose                 |
|--------------|----------------|----------------|-------|-------------------------|
| FAS-3170_A_1 | ORA_HOME       | AGGR_ORA_1_A_1 | 100GB | Database Binary         |
| FAS-3170_A_1 | CRS_HOME       | AGGR_ORA_1_A_1 | 100GB | Cluster ware Binary     |
| FAS-3170_A_1 | OCR_CSS        | AGGR_ORA_1_A_1 | 50GB  | OCR & Voting Disks      |
| FAS-3170_A_1 | VOL_DATA_1_A_1 | AGGR_ORA_1_A_1 | 12TB  | Datafiles, control file |
| FAS-3170_A_1 | VOL_LOG_1_A_1  | AGGR_ORA_1_A_1 | 200GB | Redo log Files          |
| FAS-3170_A_2 | VOL_DATA_2_A_1 | AGGR_ORA_2_A_1 | 10TB  | Datafiles, Control File |
| FAS-3170_A_2 | VOL_LOG_2_A_1  | AGGR_ORA_2_A_1 | 200GB | Redo log Files          |
| FAS-3170_A_2 | VOL_FRA_2_A_1  | AGGR_ORA_2_A_1 | 2TB   | FRA                     |
| FAS-3170_B_1 | VOL_DATA_1_B_1 | AGGR_ORA_1_B_1 | 12TB  | Datafiles, Control file |
| FAS-3170_B_1 | VOL_LOG_1_B_1  | AGGR_ORA_1_B_1 | 200GB | Redo log Files          |
| FAS-3170_B_2 | VOL_DATA_2_B_1 | AGGR_ORA_2_B_1 | 10TB  | Datafiles, Control File |
| FAS-3170_B_2 | VOL_LOG_2_B_1  | AGGR_ORA_2_B_1 | 200GB | Redo log Files          |
| FAS-3170_B_2 | VOL_ARC_2_B_1  | AGGR_ORA_2_B_1 | 2TB   | Archive Log             |

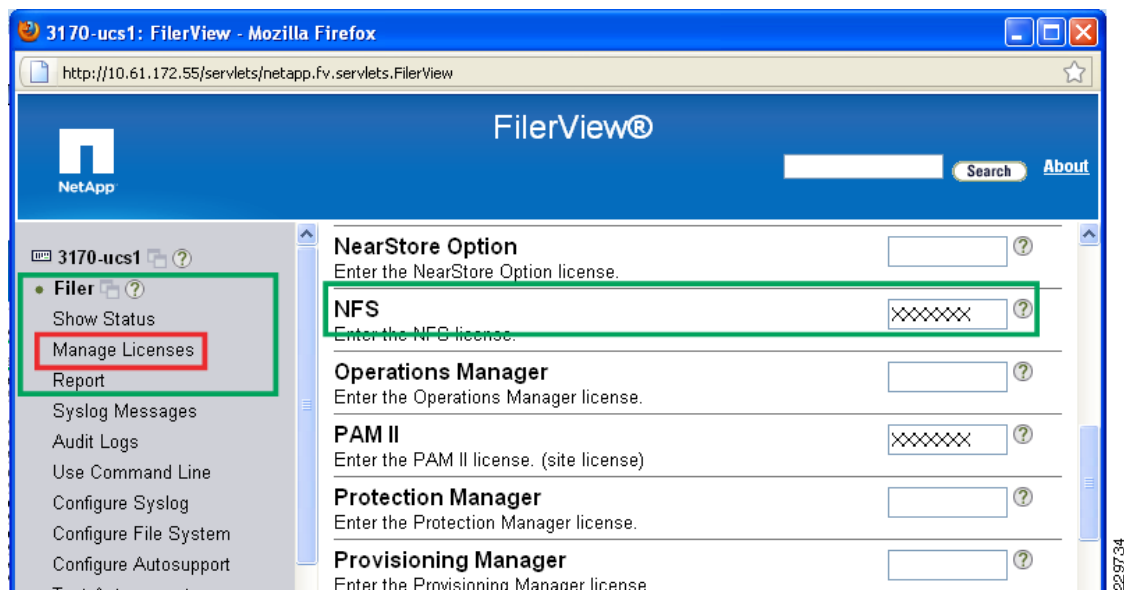
At this point, the storage controllers are ready to provide NFS shares for an Oracle installation.

## Exports the Volumes to Mount Them on Database Hosts

Before you exports the volumes created, make sure the NFS license has been provided in each NetApp filer view as follows:

Click **Filer** in the top-left corner and click the **Manage Licenses** option, as shown in [Figure 33](#), to view whether or not the storage has the NFS license installed. If the NFS license is not installed, enter a valid NFS license code in the box and click **Apply**.

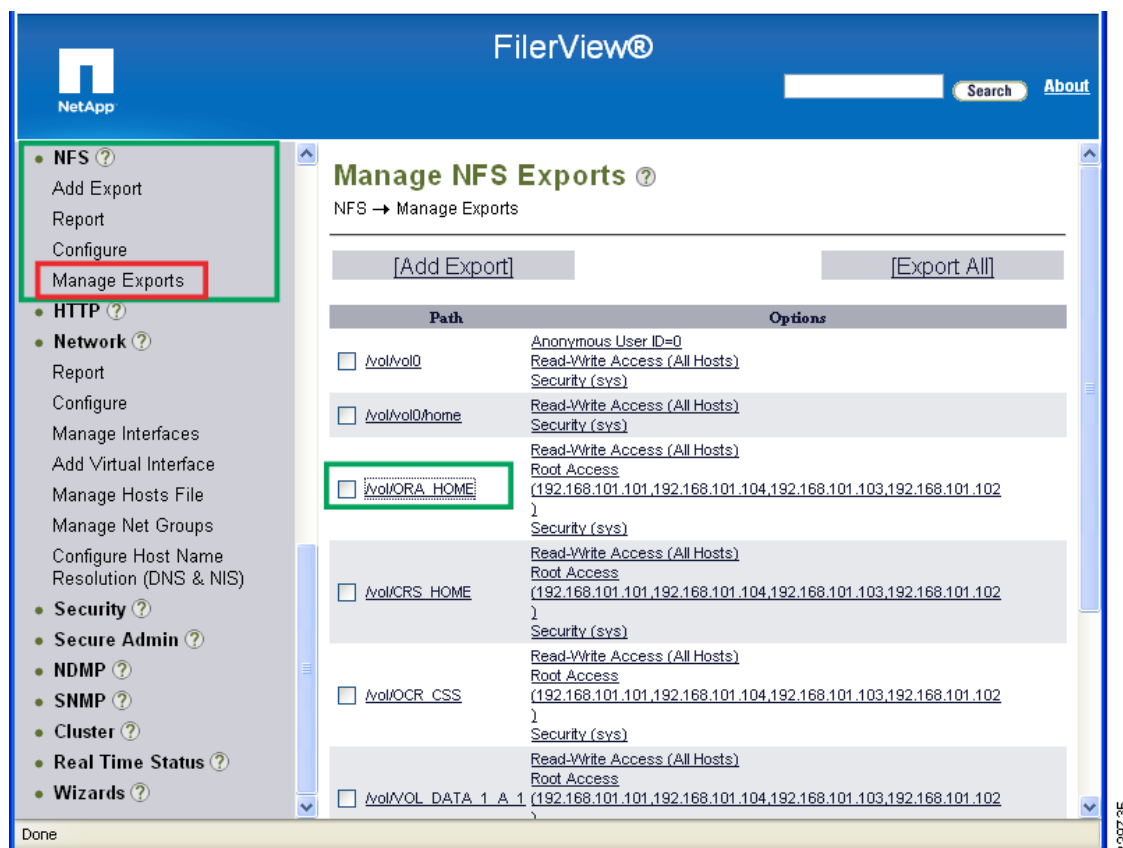
**Figure 33** *FilerView-->Manage Licenses*



After verifying NFS license, exports the volumes as follows:

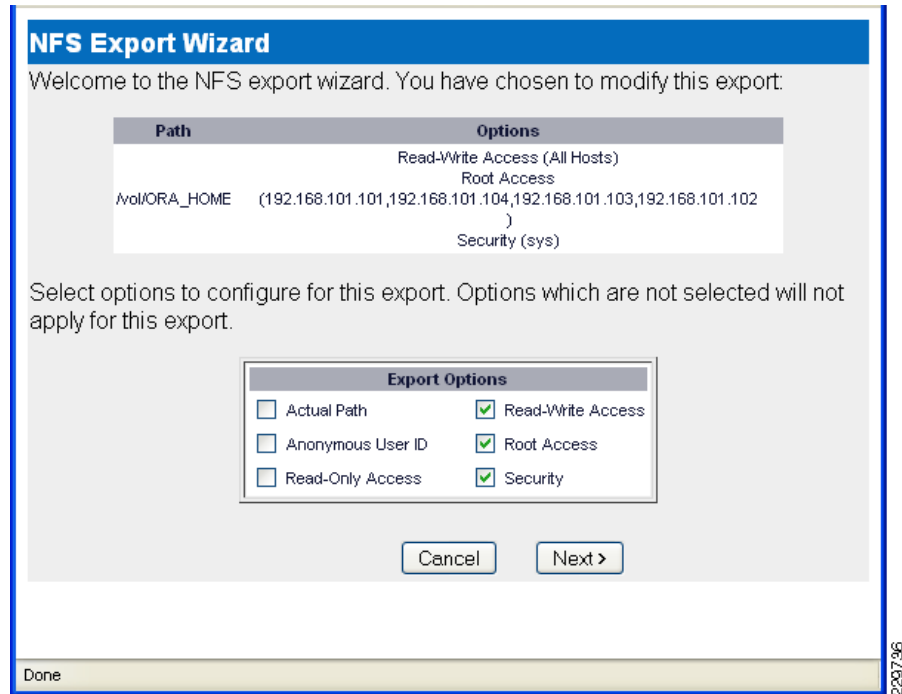
Go to the “FilerView” and expand the NFS menu by clicking on it at the left side of the window. Next click **Manage exports** to see all the volumes created and the default export parameters, as shown in [Figure 34](#).

**Figure 34**      **Manage NFS Exports**



Click each volume shown on the “Manage NFS Export” window to provide appropriate export parameters required for Oracle database. Select the checkbox as shown in the [Figure 35](#). Click the **Next** button to move forward.

Make sure the root user on each host will be able to write to each of the NFS exports. You can do this by adding the appropriate IP(s) to the Root Access list when configuring the exports on each storage controller (see [Figure 35](#)).

**Figure 35** *NFS Export Wizard*

## Applying Patches, Environment, and OS Settings

After completing the configuration of the Cisco UCS, the SAN, and storage, you can install the OS.

To test the Cisco UCS data center solution, 64-bit RedHat Enterprise Linux (OEL) 5.4 was used as the OS.



### Note

If you plan to use RHEL 5.3, be aware of the following bug in the Linux Kernel: <https://access.redhat.com/kb/docs/DOC-16041>. As a workaround, you can turn off Hyper Threading for the B200 Servers. This bug only affects Servers with Dual Xeon Processors.

## Installing the OS and Setting up the Environment

To install the OS and enable the environment settings:

- 
- Step 1** Install 64-bit RHEL 5.4 on all four nodes.
  - Step 2** Install the required RPM package for Oracle 11g R2 Grid and database on all the four nodes.  
Use the following Oracle document for pre-installation tasks, such as setting up the kernel parameters, RPM packages, user creation, etc.

*Oracle Grid Infrastructure Pre-Installation Tasks*

([http://download.oracle.com/docs/cd/E11882\\_01/install.112/e10812/prelinux.htm#BABHJHCJ](http://download.oracle.com/docs/cd/E11882_01/install.112/e10812/prelinux.htm#BABHJHCJ))

See [Appendix A—System Configuration Settings](#) for a list of kernel settings if you decide to set them up manually. [Appendix A—System Configuration Settings](#) also includes settings useful for improved NFS performance.

**Step 3** Create required oracle users and groups.

```
groupadd -g 1000 oinstall
groupadd -g 1200 dba
useradd -u 1100 -g oinstall -G dba oracle
passwd oracle
```

**Step 4** Create local directories on each node and give the ownership of these directories to oracle user.

```
mkdir -p /u1/app/11.2.0/grid
mkdir -p /u1/app/oracle
mkdir -p /u1/oradata
mkdir -p /u1/oralog
mkdir -p /u1/app/11.2.0/ocr
chown -R oracle:oinstall /u1/app /u1/app/oracle /u1/oradata /u1/oralog /u1/app/ocr
chmod -R 775 u1/app /u1/app/oracle /u1/oradata /u1/oralog /u1/app/ocr
```

For this configuration, each controller is mapped under a corresponding /u# base directory. For example, /u2/oradata will be mounted from 3170-ucs2, /u3/oradata would be mounted from 3170-ucs3, etc.

**Step 5** Edit /etc/fstab file in each node and add the entry for all volumes and its corresponding local directory created above with the proper mount options.

```
[root@clb1 ~]# vi /etc/fstab
fas1a1:/vol/ORA_HOME/orahome /u1/app/oracle nfs
rw,bg,hard,rsz=65536,wsz=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp 0 0
fas1a1:/vol/CRS_HOME/crshome /u1/app/11.2.0/grid nfs
rw,bg,hard,rsz=65536,wsz=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp 0 0
fas1a1:/vol/OCR_CSS/ocr_css /u1/app/11.2.0/ocr nfs
rw,bg,hard,rsz=65536,wsz=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp 0 0
Etc.....
```

To find proper mount options for different file systems of Oracle 11g R2, see:

<https://now.netapp.com/Knowledgebase/solutionarea.asp?id=kb7518>.



**Note**

An rsz and wsz of 65536 is supported by NFS v3 and used in this configuration to improve performance.

**Step 6** Mount all the local directories created to store database and binary, once you complete the editing of /etc/fstab in each node using root user.

```
node1# mount /u1/app/11.2.0/grid
node1# mount /u1/app/oracle/product/11.2.0/db_1
...
node2# mount /u1/app/11.2.0/grid
...
nodeN#
```

Give the ownership of all local directories to oracle user created to store database and binary after you mount them in all nodes using root user.

```
chown -R oracle:oinstall /u1/app /u1/app/oracle /u1/oradata /u1/oralog /u1/app/ocr
```

**Step 7** Configure the private and public NICs with the appropriate IP addresses.

**Step 8** Identify the virtual IP addresses and SCAN IPs and have them setup in DNS per Oracle's recommendation. Alternatively, you can update the /etc/hosts file with all the details (private, public, and virtual IP) if you do not have DNS services available.

- Step 9** Configure the ssh option (with no password) for the Oracle user.  
For more information about ssh configuration, refer to the Oracle installation documentation.
- 

You are now ready to install Oracle grid infrastructure and the database.

## Installing Oracle Grid Infrastructure and the Database

It is not within the scope of this document to include the specifics of an Oracle RAC installation; you should refer to the Oracle installation documentation for specific installation instructions for your environment.

To install Oracle, follow these steps:

- 
- Step 1** Download the Oracle Database 11g Release 2 Grid Infrastructure (11.2.0.1.0) and Oracle Database 11g Release 2 (11.2.0.1.0) for Linux x86-64.
- Step 2** Install Oracle Database 11g Release 2 Grid Infrastructures (11.2.0.1.0). See *Grid Infrastructure Installation Guide for Linux* for detailed instructions ([http://www.oracle.com/pls/db112/to\\_toc?pathname=install.112/e10812/toc.htm](http://www.oracle.com/pls/db112/to_toc?pathname=install.112/e10812/toc.htm)).
- Step 3** Install Oracle Database 11g Release 2 Database “Software Only”; do not create the database. See *Real Application Clusters Installation Guide for Linux and UNIX* for detailed installation instructions ([http://www.oracle.com/pls/db112/to\\_toc?pathname=install.112/e10813/toc.htm](http://www.oracle.com/pls/db112/to_toc?pathname=install.112/e10813/toc.htm)).
- Step 4** Configure Direct NFS client.

For improved NFS performance, Oracle recommends using the Direct NFS Client shipped with Oracle 11g. The direct NFS client looks for NFS details in the following locations:

1. \$ORACLE\_HOME/dbs/oranfstab
2. /etc/oranfstab
3. /etc/mtab

Since the NFS mount point details were defined in the “/etc/fstab”, and therefore the “/etc/mtab” file also, there is no need to configure any extra connection details. When setting up your NFS mounts, reference the Oracle documentation for guidance on what types of data can/cannot be accessed via Direct NFS.

For the client to work we need to switch the libodm11.so library for the libnfsodm11.so library, as shown below.

```
srvctl stop database -d oastdb

cd $ORACLE_HOME/lib
mv libodm11.so libodm11.so_stub
ln -s libnfsodm11.so libodm11.so

srvctl start database -d oastdb
```

With the configuration complete, you can see the direct NFS client usage via the following views:

- v\$dnfs\_servers
- v\$dnfs\_files
- v\$dnfs\_channels

- v\$dnfs\_stats

For example:

```
SQL> SELECT svrname, dirname FROM v$dnfs_servers;
```

| SVRNAME  | DIRNAME                     |
|----------|-----------------------------|
| fas1b1   | /vol/VOL_DATA_2_A_1/oradata |
| fas2a1   | /vol/VOL_DATA_1_B_1/oradata |
| fas1a1   | /vol/VOL_DATA_1_A_1/oradata |
| fas2b1   | /vol/VOL_DATA_2_B_1/oradata |
| Etc..... |                             |

SQL>



**Note** The Direct NFS Client supports direct I/O and asynchronous I/O by default.

**Step 5** Apply the latest Oracle and Grid patch sets.

For our test environment, we applied the following patches:

- Upgrade Opatch to 11.2.0.1 before applying patches.
- Apply patch 8898852 as a prerequisite for the cumulative patch.
- Apply patch 9654983 cumulative patch release July 13, 2010.

**Step 6** Now you are ready to create the database and perform the workload setup.

## Configuring the Oracle OAST Workload

The performance testing utilized the OCE v4.0 OAST software from Oracle. The test suite was installed following the *Oracle® Certification Environment OAST Installation and User's Guide for Oracle Database* document. The latest version of OAST kit, install documentation, and README can be downloaded from: <http://www.oracle.com/technology/software/ocf/ocfurls/ocf4kiturls.html>.

The Oracle Automated Stress/System Testing suite provides the following five workloads that can be used to validate the performance capabilities of a storage platform with Oracle database products:

- IO—IO bound workload, can be used as a load generator for stressing the system I/O throughput using a mixture random reads and writes.
- CPU—CPU bound workload used to saturate the CPU on the servers running Oracle.
- IO and CPU ERP—Medium I/O and high CPU workload.
- Interconnect—Workload to stress Interconnect traffic.
- DSS—DSS workload characterized by large sequential reads. Can be used to identify storage/system throughput.

According to Oracle, the following workloads are essential for certifying a specific server/storage combination running Oracle:

- IO and CPU ERP load (24 hour run)
- Interconnect/Cache Run Stress (3 hour run)

In addition to the two tests above required for certification, we also employed the DSS workload for 12 hours to measure performance of the configuration using workloads that are common in enterprise-class data warehouse environments.

For all our tests we used a single Oracle database. The database was created by OAST using a default 8k block size and 3000 GB of seed data for a total of 45000 warehouses. An example of the installation selections is:

```
OAST: Release Release 6.2.0_11gr2_061410
Copyright (c) 2009, 2010, Oracle and/or its affiliates. All rights reserved.

OAST HOME: /home/oracle/oast/home
\nClient/Server Environment?
1) Yes
2) No
#? 2
You chose: No
\nGetting Server(s) information...\n
Your OS: Linux \n
ORACLE_HOME=/u1/app/oracle/product/11.2.0/db_1
ORACLE_BASE [/u1/app/oracle]: \c
Please enter the OAST database name [oastdb]: \c
oastdb
\nWould you like to enable database archive mode?
1) enable
2) disable
#? 1
You chose: enable
\nSelect one of Oracle options?
1) Cluster
2) Non-Clustered
#? 1
You chose: Cluster
c1b1 c1b2 c1b3 c1b4
The owner of your server(s) ORACLE HOME: oracle
of nodes: 4
login : oracle
checking connection to system c1b1 ...
checking connection to system c1b2 ...
checking connection to system c1b3 ...
checking connection to system c1b4 ...
creating oast_cluster directory on system c1b1 ...
creating oast_cluster directory on system c1b2 ...
creating oast_cluster directory on system c1b3 ...
creating oast_cluster directory on system c1b4 ...
c1b1 c1b2 c1b3 c1b4
of CPUs =16
Free System Memory (Megabytes) for Oracle is 32768 Megabytes
NOTE: 1G of seed data = 15 Warehouses = 1.3G in tablespace size (free and used)
Target OLTP SEED data size (Gigabytes)[10 Gigabytes] : \c
3000
db_block_size=8192
Calculating the size of database...
Are your database files stored on a filesystem or ASM?
1) filesystem
2) ASM (Automatic Storage Management)
#? 1
You chose: filesystem
\nPlease enter (Cluster) filesystem for datafiles.
Filesystem mount point: \c
/u2/oradata
\nPlease enter (Cluster) filesystem for log group member 1 .
Filesystem mount point: \c
/u1/oralog
\nPlease enter (Cluster) filesystem for log group member 2 .
Filesystem mount point: \c
/u2/oralog
filesystem_loc: /u2/oradata
```

```

loggroupm1_loc: /u1/oralog
loggroupm2_loc: /u2/oralog
Disk Storage requirement for Oracle Database :
of warehouses: 45000
Total Size : 14824636 Megabytes
of members per log group: 2
Total size for log group member 1 : 4000 Megabytes
Total size for log group member 2 : 4000 Megabytes
Press <return> to confirm and <e> to exit. \c

```

The install process creates a kit that must be compiled by running the following:

```

cd $OAST_HOME
./nrnoastoltpXXX.sh -kitcompile

```

Once this process was complete and the logs verified the database was created by executing the following command on the primary node. This created the database using the Oracle Database Configuration Assistant and loads data into the database for use during the performance testing.

```

$ cd $OAST_HOME
$ nrnoastoltpXXX.sh -d y -dbca y

```



#### Note

Prior to running the above database creation process, some modifications were made to the OAST templates to generate scripts to distribute the data and log files across multiple nfs mounts and thus across each of the NetApp FAS3170 storage controllers. This allows for the workload to be evenly distributed across all storage nodes. The `.../oast/install/lib/oastinstall.lib` file was also modified to ignore the Filesystem space check during the above installation so that the data files could be distributed evenly across the storage controllers.

The database created by the OAST software contained the following tablespaces and associated tables. The data files for each tablespace were spread across the four ORADATA nfs mounts evenly.

| TABLESPACE_NAME | TOT_SIZE_MB | USED_SPACE | PCT_USED | FREE_SIZE_MB | PCT_FREE |
|-----------------|-------------|------------|----------|--------------|----------|
| CUST_0          | 1620000     | 882000     | 54.4     | 738000       | 45.6     |
| DIST_0          | 3728        | 3544       | 95.1     | 184          | 4.9      |
| HIST_0          | 200000      | 107600     | 53.8     | 92400        | 46.2     |
| ICUST1_0        | 96000       | 41560      | 43.3     | 54440        | 56.7     |
| ICUST2_0        | 128000      | 91700      | 71.6     | 36300        | 28.4     |
| IDIST_0         | 800         | 40         | 5        | 760          | 95       |
| IITEM_0         | 600         | 25         | 4.2      | 575          | 95.8     |
| IORDR1_0        | 80000       | 41620      | 52       | 38380        | 48       |
| IORDR2_0        | 120000      | 76100      | 63.4     | 43900        | 36.6     |
| ISTOK_0         | 200000      | 125200     | 62.6     | 74800        | 37.4     |
| ITEM_0          | 2400        | 30         | 1.3      | 2370         | 98.8     |
| IWARE_0         | 240         | 16         | 6.7      | 224          | 93.3     |
| NORD_0          | 24000       | 6460       | 26.9     | 17540        | 73.1     |
| ORDL_0          | 1600000     | 971000     | 60.7     | 629000       | 39.3     |
| ORDR_0          | 160000      | 73900      | 46.2     | 86100        | 53.8     |
| SP_0            | 200         | 20         | 10       | 180          | 90       |
| STOK_0          | 2100000     | 1574000    | 75       | 526000       | 25       |
| SYS_AUX         | 17610       | 1616       | 9.2      | 15994        | 90.8     |
| SYSTEM          | 1600        | 342        | 21.4     | 1258         | 78.6     |
| UNDOTBS1        | 32768       | 272        | .8       | 32496        | 99.2     |
| UNDOTBS2        | 3410        | 528        | 15.5     | 2882         | 84.5     |
| UNDOTBS3        | 3255        | 322        | 9.9      | 2933         | 90.1     |
| UNDOTBS4        | 995         | 352        | 35.4     | 643          | 64.6     |
| USERS           | 5           | 1          | 20       | 4            | 80       |
| WARE_0          | 388         | 368        | 94.8     | 20           | 5.2      |

```
sum 6395999 3998616 2397383
```

The core tables for the test environment with number of rows are:

| TABLE_NAME | NUM_ROWS    | BYTES         |
|------------|-------------|---------------|
| CUST       | 1349974230  | 893386752000  |
| DIST       | 450032      | 3707764736    |
| HIST       | 1354243174  | 112407347200  |
| ITEM       | 113538      | 10485760      |
| NORD       | 411361491   | 6815744000    |
| ORDL       | 13651505881 | 995622912000  |
| ORDR       | 1344299226  | 82103500800   |
| STOK       | 4492010764  | 1606418432000 |
| WARE       | 45136       | 377487360     |

## Workload Performance Testing

To evaluate basic performance of the total configuration, we utilized a subset of the default test suites that were part of the OAST test suite as follows:

- IO and CPU ERP load (24 hour run)
- Interconnect/Cache Run Stress (3 hour run)
- DSS (12 hour run)

All performance tests were executed using durations recommended by the OAST test kit. Each of the workloads used is designed to stress the system in a different way. Overall the performance testing showed the configuration is capable of delivering enterprise-class performance for sustained periods with no observed errors or other issues.

The following general observations were noted during the performance testing:

- CPU utilization—During the IO\_CPU workload the hosts consistently maintained approximately 90% utilization, while the CPU utilization on the NetApp FAS3170 storage controllers maintained an average of approximately 95%. This demonstrates the balanced nature of the architecture, showing both hosts and storage were running at similar levels of CPU utilization during the performance tests.
- Lack of saturation points within the subsystems (CPU, disk, I/O, or networking) demonstrates the balance of the architecture.
- DSS Throughput—Per the Oracle AWR reports, this configuration sustained an average DNFS-based I/O throughput of approximately 3GB/s from the four NetApp FAS3170 storage controllers to the four RAC nodes over a 12 hour period (see [Figure 37](#)). This throughput does not include Oracle interconnect traffic on the network.
- OLTP Workload—Per the Oracle AWR reports, this configuration sustained an average of approximately 360k IOPs measured from the four NetApp FAS3170 storage controllers to the four RAC nodes over a 24 hour IO\_CPU workload run. Note that the size of the working set dictated that the majority of the IOPs were serviced from the 2TB of Flash Cache on the NetApp storage controllers.

The observed workload performance can be attributed to several factors, including but not limited to:

- The simplified, performance-oriented architectural design of the Cisco UCS based on a 10-Gbps unified fabric.
- The pairing of the Cisco UCS with NetApp FAS storage with Flash Cache.

- The balance between UCS hosts, NetApp Storage, and Cisco network to allow for high levels of performance and the elimination of performance limiting bottlenecks.
- Use of the Oracle 11gR2 and the D-NFS client in place of the kernel-based NFS client.

## Test Results Using the IO CPU Workload

The IO\_ERP/IO\_CPU OAST workload was ran for a period of 24 hours per the recommendations in the Oracle OCE documentation. This is an OLTP type workload and is characterized by small random reads/writes (8k in this case). A typical OLTP Oracle application has some level of write activity because of Data Manipulation Language (DML) operations, such as updates, inserts, and deletes. This workload utilized 8k blocks and 625 users per node for a total of 2500 concurrent simulated user processes.

Table 8 and Table 9 provide the throughput and host CPU utilization taken from the Oracle OAST AWR reports for each of the four RAC nodes. This data validates the throughput data shown later from the NetApp storage. Additionally, this data shows that, on average, the four RAC nodes consumed approximately 90% of the available CPU resources.



### Note

The test results presented here should not be used for comparison purposes. It is worth noting that limited tuning was performed and a common database was used for both OLTP and DSS workloads. Additionally, because of the size of the associated working set, the OLTP workload was serviced largely from the FlashCache in the NetApp FAS3170 storage controllers. This results in higher levels of IOPS than would generally be observed had the data been served from the disks.

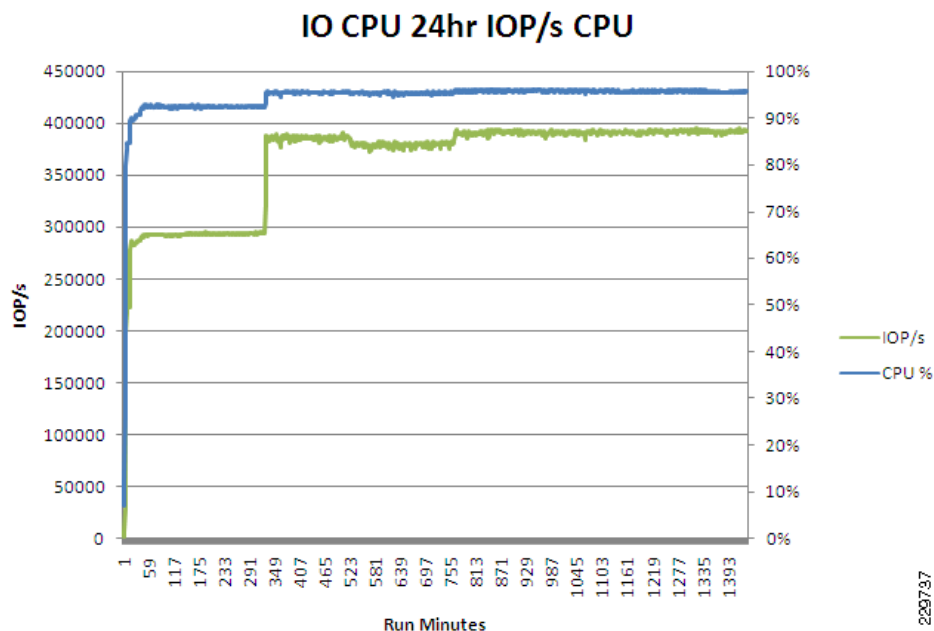
**Table 8** *IO\_CPU Read/Write Statistics from AWR Reports*

| Instance | DataReqs per sec. | MB Reads sec. | DataReqs per sec. | MB Writes sec. |
|----------|-------------------|---------------|-------------------|----------------|
| oastdb1  | 92363.66          | 722.215       | 1293.08           | 10.8031        |
| oastdb2  | 87606.55          | 685.07        | 1365.21           | 11.4904        |
| oastdb3  | 88086.77          | 688.8         | 1337.13           | 11.2258        |
| oastdb4  | 90501.6           | 707.651       | 1345.48           | 11.2879        |
| Totals   | 358558.58         | 2803.736      | 5340.9            | 44.8072        |

**Table 9** *IO\_CPU Average Host CPU Utilization*

| Host    | %user | %nice | %sys  | %iowait | %irq   | %soft | %steal | %idle   |
|---------|-------|-------|-------|---------|--------|-------|--------|---------|
| c1b1    | 60.22 | 0     | 21.83 | 0       | 0.26   | 6.74  | 0      | 10.95   |
| c1b2    | 60.61 | 0     | 21.91 | 0       | 0.37   | 6.34  | 0      | 10.76   |
| c1b3    | 60.73 | 0     | 21.87 | 0       | 0.26   | 6.46  | 0      | 10.67   |
| c1b4    | 60.64 | 0     | 21.83 | 0       | 0.26   | 6.6   | 0      | 10.67   |
| Average | 60.55 | 0     | 21.86 | 0       | 0.2875 | 6.535 | 0      | 10.7625 |

Figure 36 shows the throughput in IOPs and average CPU utilization observed by the four FAS3170 storage controllers during this test. The NetApp storage delivered an aggregate average throughput of approximately 365K IOPS over the 24 hour test duration with a sustained average CPU utilization of approximately 95%. Each IOP is 8k in size making for an average of 2.9GB/s of read/write throughput.

**Figure 36** Combined IOP/s and Average Storage CPU Utilization

## Test Results Using the DSS Workload

For simplicity, the same database was used for the DSS tests and the IO\_CPU tests. The OAST DSS workload was ran over a 12 hour period utilizing 50 users per RAC node (200 total) to drive the load. The number of users in a DSS environment is typically less than you would find in an OLTP implementation. Unlike the OLTP workload, DSS workloads are generally made up of mostly large sequential reads with minimal amounts of writes and random reads. The size of the workload was chosen to represent a configuration used by power users and/or analysts that are executing long running queries or performing analysis against the database.



### Note

FlashCache was not used for the DSS workload tests in order to increase the amount of data being serviced from disk and thus create a more realistic workload scenario.

[Table 10](#) and [Table 11](#) provide the host CPU utilization and read throughput statistics taken from the Oracle OAST AWR reports for each of the four RAC nodes. For this test, the OAST DSS workload does not significantly stress any of the four RAC nodes from a CPU utilization perspective utilizing an average of approximately 13% of the CPU resources in each of the RAC nodes.

In terms of throughput, Oracle reported an average of approximately 3GB/s of data read from the NetApp FAS3170 storage controllers over the 12 hour duration of the test.

As can be seen by the host CPU utilization ([Table 10](#)), the OAST DSS workload does not stress the host side CPU and as such the UCS is able to handle the load with ease, leaving an abundance of capacity available for processing. This available capacity demonstrates the UCS's ability to service environments where processes may be manipulating the vast amounts of data returned and thus putting a heavier load on the host side CPU.

**Table 10 DSS Workload Host Level CPU Utilization**

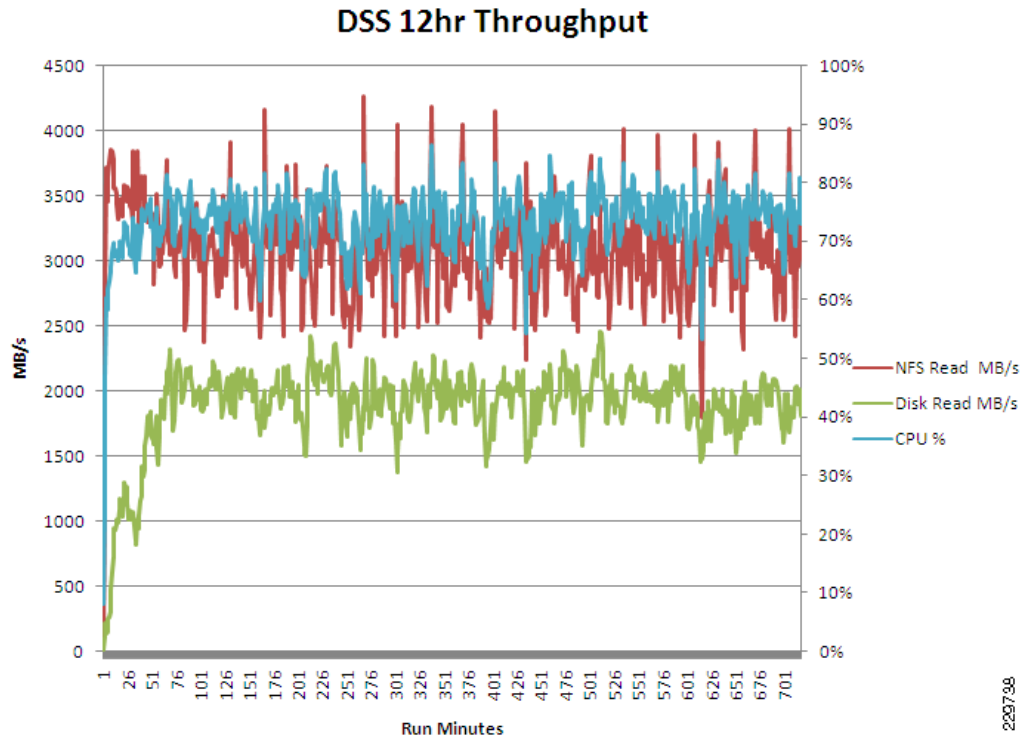
| Host    | %user | %nice | %sys  | %iowait | %irq   | %soft | %steal | %idle   |
|---------|-------|-------|-------|---------|--------|-------|--------|---------|
| c1b1    | 8.34  | 0     | 2.91  | 0.01    | 0.13   | 1.46  | 0      | 87.14   |
| c1b2    | 8.15  | 0     | 3.28  | 0       | 0.13   | 1.63  | 0      | 86.8    |
| c1b3    | 7.38  | 0     | 3.2   | 0.01    | 0.14   | 1.62  | 0      | 87.65   |
| c1b4    | 8.01  | 0     | 3.27  | 0.01    | 0.13   | 1.63  | 0      | 86.94   |
| Average | 7.97  | 0     | 3.165 | 0.0075  | 0.1325 | 1.585 | 0      | 87.1325 |

**Table 11 AWR Read Statistics for RAC Instances**

| Instance | statistic           | Total bytes         | Bytes Per Second |
|----------|---------------------|---------------------|------------------|
| oastdb1  | physical read bytes | 48,414,141,341,696  | 750,783,794.68   |
| oastdb2  | physical read bytes | 47,784,039,522,304  | 741,023,227.21   |
| oastdb3  | physical read bytes | 48,065,921,507,328  | 745,388,030.14   |
| oastdb4  | physical read bytes | 47,808,913,203,200  | 741,403,949.27   |
| Total    |                     | 192,073,015,574,528 | 2,978,599,001.30 |

Figure 37 shows the overall average CPU utilization as well as the overall NFS read throughput in MB/s delivered to the RAC nodes during the test from the perspective of the NetApp FAS3170 storage controllers. This throughput value matches the amount of throughput reported by the Oracle AWR reports. Additionally, Figure 37 shows the amount of data read from the disks of the NetApp storage controllers during this test in MB/s. The DSS workload tended to be bursty in nature. However, this test shows that the NetApp FAS3170 storage controllers delivered high performance while maintaining an average of 73% CPU utilization. This suggests there is untapped capacity available at the storage level as well as the host side to support higher levels of throughput should the need arise.

**Figure 37** *DSS Workload Combined Throughput and CPU Utilization*



## Interconnect Workload

The OAST Interconnect workload was run for a period of three hours without errors or incident. The run time was set per the recommendation in the OAST documentation. This workload is designed to stress the interconnect between RAC nodes and the Cisco UCS fabric performed without flaws or any excessive latency.

## Backup, Recovery, and Cloning of Oracle Database 11gR2

The purpose of a backup and recovery strategy is to protect the business against data loss and to reconstruct the database after data loss. The major contributors for a database recovery scenario are User errors, Application errors, and Media failure. For Oracle database administrators, backup, recovery, and cloning is a complex task that may require detailed negotiations with network, system, and storage administrators, all of whom are involved in architecting and implementing backup, recovery, and cloning solutions for Oracle database systems. An Oracle database environment faces significant data related challenges in terms of backup, recovery, and cloning.

Some of the top challenges to being successful in these tasks are:

- Backing up data frequently without disrupting the production environment
- Enhancing the options of recoverability
- Managing the backup/recovery credentials
- Managing consistent data across primary and secondary environments/application consistent backups

- Identifying the data that needs to be backed up
- Easier and effective way of recovering the database in less time
- Provisioning sufficient current copies (clones) of the production database for dev/test and training purposes
- Managing the infrastructure resources for database clones is particularly critical in large database environments

## Solutions for Oracle Backup, Recovery, and Cloning

There are various enterprise backup and recovery options available for Oracle database environments. Oracle Corporation provides several native options, with the most widely used being:

- [Oracle Recovery Manager \(RMAN\)](#)
- [Oracle Data Guard](#)
- [Oracle Flashback](#)
- [Oracle 11gR2 CloneDB](#)
- [Oracle Rapid Clone](#) (for E-Business Suite full stack provisioning and replication)

In addition, NetApp, as a leading data management platform provider, offers enhanced backup, recovery, and cloning solutions for Oracle environments to the customers:

- [NetApp SnapManager for Oracle \(SMO\)](#)
- [NetApp SnapCreator for Oracle](#)
- [NetApp SnapMirror](#)

### Oracle Recovery Manager (RMAN)

Oracle Recovery Manager (RMAN) is a command line and Oracle Enterprise Manager (OEM) based tool. It is the Oracle preferred method for efficiently backing up and recovering an Oracle database. RMAN is a built-in component of the Oracle database server and is available since Oracle version 8. RMAN is designed to work intimately with the database server, providing block-level corruption detection during backup and recovery. It integrates with Oracle Secure Backup and third-party media management products for tape backups. It provides a common interface for backup tasks across different host operating systems and offers features not available through user-managed methods, such as backup-files, retention policy, and detailed history of all backups. RMAN also has a feature to duplicate (clone) a database.

Since many companies have extended retention requirements, RMAN supports a proxy copy extension. Many third-party companies have developed media management libraries (MML) that facilitate the use of tape management systems and native snapshot backups. NetApp has developed an MML that can be used to backup, recover, and duplicate databases using RMAN as an interface to the native data management capabilities of Data ONTAP.

The following are the backup types offered by RMAN:

- Full or Incremental Backup
- Open or Closed Backup
- Consistent or Inconsistent Backup
- Backup via proxy copy

The restore and recovery options available with RMAN are:

- Whole database restore and recovery
- Restore and recovery of individual tablespaces or datafiles to default or new location
- Restoring of lost database files such as Control file and SPFILE to default or new location
- Performing media recovery of a database
- Point-in-time recovery of a database or tablespace
- Restoring archived redo logs from backup to default or new location
- Data block recovery

#### **Database Duplication (Cloning) with RMAN**

RMAN has a built-in feature to duplicate the production database. It creates a physical standby database which can be provisioned for training purposes. It also leverages to take backups from this physical standby database, thus relieving the load on the production database and enabling efficient use of system resources on the standby site. As part of the duplication, RMAN manages:

- Restores the target datafiles into the duplicate database and performs incomplete recovery using all available archived log and incremental backups.
- Opens the duplicate database with the RESETLOGS option after incomplete recovery to create the online redo logs.
- Generates a new, unique database identifier for the duplicate database.

### **Oracle Data Guard**

Oracle Data Guard ensures high availability, data protection, and disaster recovery for enterprise data. Data Guard provides a comprehensive set of services that create, maintain, manage, and monitor one or more standby databases to enable production Oracle databases to survive disasters and data corruptions. Data Guard maintains these standby databases as transaction consistent copies of the production database. Then, if the production database becomes unavailable because of a planned or an unplanned outage, Data Guard can switch any standby database to the production role, thus minimizing the downtime associated with the outage. Data Guard can be used with traditional backup, restoration, and cluster techniques to provide a high level of data protection and data availability.

The advantages of Data Guard are:

- Disaster recovery, data protection, and high availability  
Data Guard provides an efficient and comprehensive disaster recovery, data protection, and high availability solution. Easy-to-manage switchover and failover capabilities allow role reversals between primary and standby databases, minimizing the downtime of the primary database for planned and unplanned outages.
- Complete data protection  
With its standby databases, Data Guard guarantees no data loss, even in the face of unforeseen disasters. A standby database provides a safeguard against data corruption and user errors. Storage level physical corruptions on the primary database do not propagate to the standby database. Similarly, logical corruptions or user errors that cause the primary database to be permanently damaged can be resolved. Finally, the redo data is validated when it is applied to the standby database.
- Efficient use of system resources  
The standby database tables that are updated with redo logs received from the primary database can be used for other tasks such as backup operations, reporting, summations, and queries, thereby reducing the primary database workload necessary to perform these tasks, saving valuable CPU and

I/O cycles. With a logical standby database, users can perform normal data manipulation operations on tables in schemas that are not updated from the primary database. A logical standby database can remain open while the tables are updated from the primary database and the tables are simultaneously available for read-only access. Finally, additional indexes and materialized views can be created on the maintained tables for better query performance and to suit specific business requirements.

- Flexibility in data protection to balance availability against performance requirements

Oracle Data Guard offers maximum protection, maximum availability, and maximum performance modes to help enterprises balance data availability against system performance requirements.

- Centralized and simple management

The Data Guard broker provides the Data Guard Manager graphical user interface and the Data Guard command line interface to automate management and operational tasks across multiple databases in a Data Guard configuration. The broker also monitors all of the systems within a single Data Guard configuration.

- Automatic gap detection and resolution

If connectivity is lost between the primary and one or more standby databases (for example, due to network problems), redo data being generated on the primary database cannot be sent to those standby databases. Once connectivity is re-established, the missing log sequence (or the gap) is automatically detected by Data Guard and the necessary redo logs are automatically transmitted to the standby databases. The standby databases are resynchronized with the primary database, with no manual intervention by the DBA.

## Oracle Flashback

Oracle Flashback Technology is a group of Oracle database features that let you view past states of database objects or to return database objects to a previous state without using point-in-time media recovery. Oracle Flashback Technology reduces recovery time from hours to minutes. Oracle Flashback Technology provides a set of new features to view and rewind data back and forth in time. The Flashback features offer the capability to query historical data, perform change analysis, and perform self-service repair to recover from logical corruptions while the database is online. Oracle Flashback features uses the Automatic Undo Management (AUM) system to obtain metadata and historical data for transactions. They rely on undo data, which are records of the effects of individual transactions.

The features of Oracle Flashback are:

- Perform queries that return past data
- Perform queries that return metadata that shows a detailed history of changes to the database
- Recover tables or rows to a previous point in time
- Automatically track and archive transactional data changes
- Roll back a transaction and its dependent transactions while the database remains online

**Caveat**—Oracle Flashback is based on a copy-on-write (CoW) methodology which means additional I/O operations are required for every DML statement that is executed. The net result is that the operational efficiency of the database instance is reduced. Therefore, it is imperative to characterize the cost (performance impact) versus benefit (flexible recovery options) of Oracle Flashback so that the infrastructure can be “sized” to accommodate the additional load that accompanies the use of this feature. Moreover, the need for frequent logical recoveries may be an indication of operational challenges associated with the management of the system. For example, if tables are frequently dropped in production, by mistake, then safeguards can be developed or enhanced to ensure this situation is mitigated by the infrastructure or by enhanced procedures.

## Oracle 11gR2 CloneDB

Oracle 11gR2 CloneDB leverages both RMAN and general purpose copy-on-write (CoW) semantics to create a replica of an Oracle database from an RMAN copy. Refer to the caveat in the list of features of Oracle Flashback for a summary of the “cost” of CoW when running an application against a live database.

## Oracle Rapid Clone

Oracle Rapid Clone was developed to facilitate the cloning of all or part of the Oracle E-Business Suite software stack. The two use cases for this are “clone” and configure a “middle tier” and “clone” and configure both the “middle” and “database” tiers for any number of reasons, including full stack provisioning for dev/test and application hosting. Refer to NetApp (Technical Report) TR-3840 for details on how to easily integrate NetApp FlexClone into this paradigm.

## NetApp SnapManager for Oracle (SMO)

Backup, recovery, and cloning are all complicated tasks that are synonymous with Oracle database management. NetApp SnapManager for Oracle simplifies and automates these complex operations by leveraging NetApp Snapshot, SnapRestore, and FlexClone technologies to provide fast, space-efficient, disk-based backups; rapid, granular restore and recovery; and quick, space-efficient cloning of Oracle databases.

NetApp SnapManager for Oracle automates and simplifies the complex, manual, and time-consuming processes associated with the backup, recovery, and cloning of Oracle databases. SnapManager leverages NetApp technologies like Snapshot, SnapRestore, and FlexClone while integrating with the latest Oracle database releases. SnapManager also integrates seamlessly with native Oracle technology (such as Oracle Real Application Clusters [RAC], Oracle Recovery Manager [RMAN], Automatic Storage Management [ASM], and Direct NFS) and across FC, iSCSI, and NFS protocols to allow IT organizations to scale their storage infrastructure, meet increasingly stringent SLA commitments, and improve the productivity of database and storage administrators across the enterprise.

### What does SMO Offer?

SMO eliminates the single point of failure for recovering Oracle databases by integrating with NetApp Protection Manager and NetApp Operations Manager provides the following:

Backup options:

- NetApp Snapshot copies for space efficient backup on primary storage
- NetApp SnapVault for longer term backups on secondary storage
- NetApp SnapMirror for high availability and/or disaster recovery

For flexibility, SMO provides the option of full or partial and/or application-/crash-consistent backups:

- A full backup puts all tablespaces into “hot backup” mode at once, creates Snapshot copies of the associated volumes, returns the tablespaces to normal operation, and then performs a log switch, archives the logs, and makes a Snapshot copy of the archived logs. Full backups use the least Snapshot copies to get the job done, but may keep tablespaces in “backup mode” for a longer period.
- A partial backup does essentially the same thing, but iterates through the tablespaces one at a time. This ensures that each tablespace is in “backup mode” for the shortest possible time, but uses more Snapshot copies since a volume may be used by more than one tablespace.
- By default, SMO does application consistent backups. It should be noted that recovery options are very limited with crash consistent backups, i.e., “Instance Recovery” only. Refer to MyOracleSupport document ID 604683.1 for Oracle’s supportability guidelines on this topic.

Restore options:

- The entire backup/full data volume restore
- User-specified tablespaces or data files
- Only control files
- Control files along with data files or tablespaces

Recovery options:

- The last transaction that occurred in the database
- A specific date and time
- A specific Oracle SCN
- The time of the backup
- Logical Recovery (Table, Schema, Data, etc.):
  - Create a database clone from a backup containing the latest version of the affected object
  - Use the Oracle supplied tool of your choice to restore the object
  - Optionally, drop the clone

## Database Clone Using SMO

With the use of NetApp FlexClone technology, SMO can create an instantaneous copy of the production database without using any additional physical space on the storage system. FlexClone makes the whole process much more time and space efficient because instead of storing a new physical copy, FlexClone simply stores the changed blocks. These clone copy of the database can be utilized for dev/test, reporting, root cause analysis, and training purposes.

**Table 12 Comparing SnapManager for Oracle and RMAN—Feature and Function Only**

|                                               | <b>SnapManager for Oracle</b>                               | <b>RMAN</b> |
|-----------------------------------------------|-------------------------------------------------------------|-------------|
| <b>Incremental backups</b>                    | No                                                          | Yes         |
| <b>Full database backup/restore</b>           | Yes                                                         | Yes         |
| <b>Tablespace backup/restore</b>              | Yes                                                         | Yes         |
| <b>Data file backup/restore</b>               | Yes (partial file SnapRestore)                              | Yes         |
| <b>Control file backup/restore</b>            | Yes                                                         | Yes         |
| <b>Backup verification</b>                    | Yes                                                         | Yes         |
| <b>Inconsistent/crash consistent backups</b>  | No (refer to MyOracleSupport document ID 604683.1)          | Yes         |
| <b>Online/offline backups</b>                 | Yes                                                         | Yes         |
| <b>Logical recovery (table, schema, etc.)</b> | Yes (clone from backup, then “transfer” object from backup) | No          |
| <b>Data block recovery</b>                    | No                                                          | Yes         |
| <b>Database clone/duplicate</b>               | Yes                                                         | Yes         |
| <b>Point-in-time or SCN recovery</b>          | Yes                                                         | Yes         |
| <b>Unstructured data, outside the DB</b>      | Yes (scripted)                                              | No          |
| <b>Register backups with RMAN</b>             | Yes                                                         | Yes         |

**Table 12 Comparing SnapManager for Oracle and RMAN—Feature and Function Only**

|                                               | SnapManager for Oracle | RMAN |
|-----------------------------------------------|------------------------|------|
| <b>ASM Integration (host file-to-storage)</b> | Yes                    | Yes  |
| <b>Virtual and Physical environments</b>      | Yes                    | Yes  |

### NetApp SnapCreator for Oracle

SnapCreator provides a central framework that integrates NetApp Snapshot technology with enterprise applications. Normally this requires a customized script that would then interface with the Oracle database application and the NetApp storage system. These customized scripts are written over and over every day. SnapCreator saves time and provides our customers with the best, most reliable solution possible. SnapCreator provides Oracle integration through modules or plug-ins. SnapCreator offers a framework which can be integrated with Oracle database application consistent backup scripts or use the built-in SnapCreator for Oracle module.

### NetApp SnapMirror

NetApp SnapMirror software combines disaster recovery and data distribution in a streamlined solution that supports today's global enterprises. SnapMirror is a very cost-effective solution with efficient storage and network bandwidth utilization and provides additional value by enabling you to put the DR site to active business use.

SnapMirror allows a data set to be replicated between NetApp storage systems over an IP or fibre channel network for backup or disaster recovery purposes. The destination file system can be made available for read-only access or can be made writable using either NetApp FlexClone technology or by “breaking” the mirror relationship. When using FlexClone technology to make the destination copies writable, the replication is uninterrupted. In case the replication relationship is broken, it can be reestablished by synchronizing the changes made to the destination back to the source file system.

With the use of FlexClone on SnapMirror destination volumes, the read-only replicas can be made writable without consuming space and without interrupting replication operations from the primary to DR site. The clone creation typically takes only a few seconds. The writable clones can then be used to develop applications or for testing (functional or performance) before production deployment.

**Table 13 NetApp SnapMirror versus Oracle Data Guard**

|                 | NetApp SnapMirror                                                                                                                                                                   | Oracle Data Guard                                                                     |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| <b>Unit</b>     | Changed blocks between storage controllers—no impact on hosts                                                                                                                       | “redo transport” between DB hosts—“steals” cycles from transaction processing         |
| <b>Benefits</b> | <ul style="list-style-type: none"> <li>• Application independent</li> <li>• Both structured and unstructured data</li> <li>• Scales linearly with additional controllers</li> </ul> | Ideal for workloads where large amounts of data are modified by a few DML operations. |
| <b>General</b>  | 1:M and Bi-Directional                                                                                                                                                              | 1:M and Bi-Directional                                                                |



#### Note

In the above section, we discussed the most widely used backup, recovery, and cloning solutions offered by application vendors and NetApp. However, the testing was executed only on NetApp SMO and Oracle RMAN.

# Test Environment

There were few changes made to the test environment used in the Phase 1 testing. The environment was reconfigured to simulate production and dev/test environments. The configuration changes made to the Phase I test environment are indicated below.

## Production

- A Cisco UCS blade server with Oracle Database 11gR2 single instance database hosted on it.
- NetApp FAS 3170 HA pair attached to the database host.
- Database size—8TB (the same OAST [Oracle Automated Stress Test] workload database used in Phase I testing).
- SnapManager for Oracle (SMO)—version installed 3.1.
- SnapDrive for UNIX (SDU)—version installed 4.2

## Dev/Test

- A Cisco UCS blade server with Oracle Database 11gR2 single instance hosted on it.
- NetApp FAS 3170 HA pair attached to the dev/test host.
- SnapManager for Oracle (SMO)—version installed 3.1.
- SnapDrive for UNIX (SDU)—version installed 4.2

## SnapManager for Oracle (SMO) Repository

- A Cisco UCS blade server with Oracle Database 11gR2 database (sample DB) hosted on it.  
This is a repository database to store backup information and backup credentials used by SnapManager for Oracle (SMO).



### Note

Apart from the above configuration changes, every other component used in Phase I of the testing was unchanged. SnapDrive for UNIX (SDU) is a pre-requisite for SMO as SMO communicates with the storage system via SDU.

## Test Procedure

1. SnapMirror the production (source) site database volumes to dev/test (destination) site over IP.  
The time taken to do a SnapMirror base line transfer of database size 8TB from source to destination is 10:20:00.  
Break the SnapMirror relationship and the production site database copy is available at the dev/test site.
2. Backup the production database using SnapManager for Oracle.

The SMO backup is a snapshot-based, point-in-time copy of the 8TB size database and stored at the production site storage controllers. This backup does not occupy additional space on the storage controllers.

The time taken by SnapManager for Oracle to backup the database size 8TB is 00:01:47.

3. Corrupt a couple of data files at the production site and restore/recover the same through SMO backup.

For example: `dd if=/dev/zero of='data file path' bs=8k count=1`

The time taken by SnapManager for Oracle to restore/recover the corrupt data files is 00:04:27.

4. Backup the production database by Oracle Recovery Manager (RMAN).

Full Database backup plus archive logs.

The time taken by RMAN to backup the database size 8TB is 06:38:00.

5. Corrupt a couple of data files at the production site and restore/recover the same through RMAN backup.

For example: `dd if=/dev/zero of='data file path' bs=8k count=1`

The time taken by RMAN to restore/recover the corrupt data files is 00:09:12.

6. Backup the production database by Oracle Recovery Manager (RMAN) using binary compression.

Full database backup plus archive log using binary compression.

The time taken by RMAN to backup the database size 8TB is 26:48:40.

7. Corrupt a couple of data files at the production site and restore/recover the same through RMAN compressed backup.

For example: `dd if=/dev/zero of='data file path' bs=8k count=1`

The time taken by RMAN to restore/recover the corrupt data files from binary compressed backup is 00:16:28.

8. Backup the database at dev/test site by snap mirrored database volumes using SnapManager for Oracle.

The backup is taken on the dev/test site using SMO and stored on the dev/test storage controllers.

9. Create the clone of production database from backup at dev/test site using SMO.

Created two clones at the dev/test site from the backup at dev/test site and stored on storage controllers attached to dev/test site.

The time taken by SnapManager for Oracle to create a clone of the 8TB size database is 00:06:06.

The clone database does not occupy additional space on the residing storage controller and only the delta occupies physical disk space.

## Results and Conclusions

SnapManager for Oracle provides a rich feature set that allows IT organizations to take advantage of fast, space-efficient, disk-based backups; rapid, granular restore and recovery; and quick, space-efficient cloning.

Table 14 shows the observations made while testing SMO backup, recovery, and cloning and Oracle Recovery Manager backup and recovery.

**Table 14** Time Taken by SMO Backup, Recovery, and Cloning—RMAN Backup and Recovery

|                                                          | Backup Time (Full Database Backup) | Restore and Recovery Time (Two Corrupt Data Files) |
|----------------------------------------------------------|------------------------------------|----------------------------------------------------|
| <b>SnapManager for Oracle( SMO)</b>                      | 00:01:47                           | 00:04:27                                           |
| <b>Oracle Recovery Manager (RMAN)</b>                    | 06:38:00                           | 00:09:12                                           |
| <b>Oracle Recovery Manager (RMAN)—Binary Compression</b> | 26:48:40                           | 00:16:28                                           |

**Note**

The Oracle Recovery Manager (RMAN) time taken to backup and recover is subject to the test environment used. The RMAN backup and recovery was performed with default parameters and none of the parameters were tuned in any manner.

## Software Licenses for NetApp Storage Controllers

- NFS
- SnapManager for Oracle
- SnapVault
- Snap Drive for UNIX
- SnapMirror
- SnapRestore
- FlexClone
- SnapMirror Synchronous

**Table 15** References

| Document                               | Topic                                                  |
|----------------------------------------|--------------------------------------------------------|
| NetApp TR-3633                         | Best Practices for Oracle                              |
| NetApp TR-3840                         | Cloning Oracle E-Business Ste                          |
| NetApp TR-3861                         | DevTest with VMware vSphere                            |
| MyOracleSupport 604683.1               | Crash Consistent Snapshots                             |
| MyOracleSupport 249212.1               | Oracle on VMware                                       |
| Cisco search “NetApp, Oracle”          | Full stack deployment                                  |
| NetApp search “SnapManager for Oracle” | Simplified and efficient backup, recovery, and cloning |
| NetApp search “ENOSPC”                 | ENOSPC errors when running Oracle on a NetApp          |

# Sensitive Data Discovery and Masking with Dataguise

This section describes the test procedure for Dataguise DgDiscover and DgMasker within the Cisco UCS and NetApp Solution for Oracle Real Application Clusters (RAC). It provides a high-level overview of the Dataguise products, customer use cases, and value proposition. It then describes the prerequisites, setup, and test plan used for deploying the products in the Cisco UCS and NetApp environment. A summary of the results of the testing is presented at the end.

## Benefits of the Solution

The Cisco UCS and NetApp Solution for Oracle Real Application Clusters (RAC) provides a number of benefits for data intensive computing environments, including:

- Simplified deployment and operation
- Oracle Database 11g Direct NFS Client
- High-performance platform for Oracle RAC
- Safer deployments with validated configurations
- Effective system management and fault tolerance
- Fast and efficient cloning of production data to create multiple copies for non-production use

The addition of DgDiscover and DgMasker to the solution enhances these benefits by providing the ability to identify and protect sensitive data from breach or unauthorized disclosure when leveraged for test, development, QA, and business analytics uses.

## Solution Overview

Deploying and supporting business applications involves continuous patch and upgrade cycles. As a result, for every application in production, enterprises may require an additional six to eight copies of the application and data for development, testing, and support. While organizations typically maintain strict access controls on production systems, data security in non-production instances is generally less rigorous, making this data potentially vulnerable to breaches due to negligence or loss from insider threats.

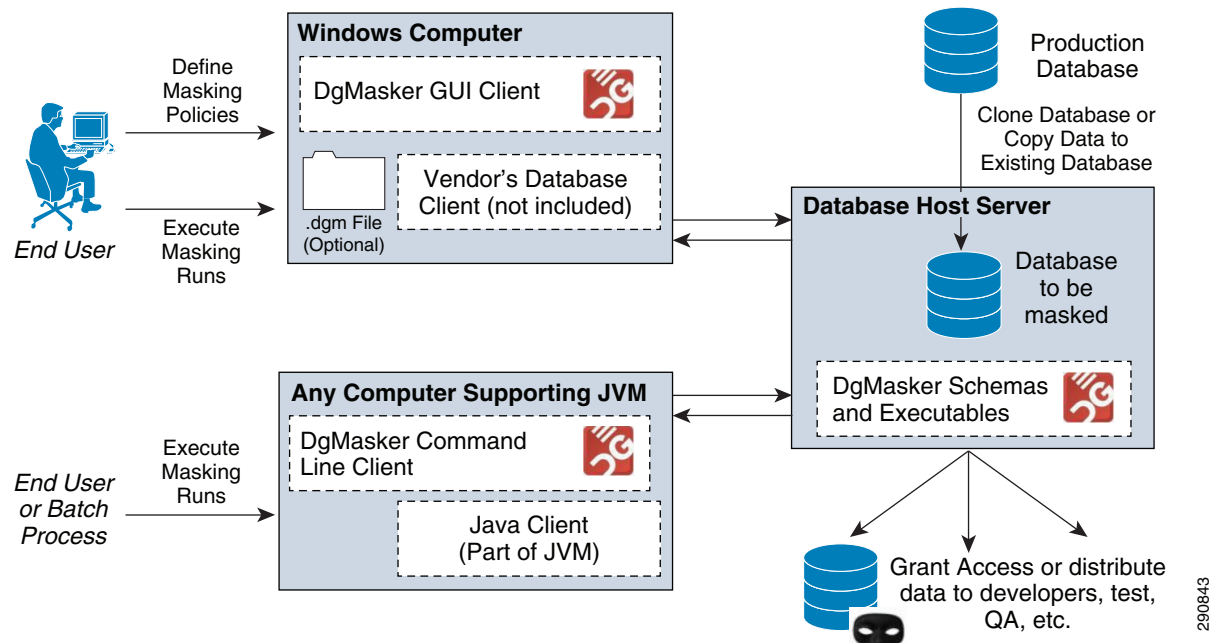
Dataguise DgDiscover and DgMasker work together to help administrators protect sensitive data in non-production use. DgDiscover finds databases deployed in enterprise networks and scans the data in them to identify sensitive information such as credit card numbers, social security numbers, names, addresses, telephone numbers, and other personally identifiable or proprietary information. DgMasker enables the rapid definition of policies for securely masking the sensitive data in a way that is completely transparent to downstream applications and uses.

Data masking is the process of de-identifying (masking) specific data elements within data stores. When used properly, sensitive data within application data sets is replaced with realistic, but not real, data. Data masking is typically done while provisioning non-production environments so that copies created to support test and development processes are not exposing sensitive information. The goal is to ensure that sensitive information is not available outside authorized environments, thus avoiding risks of breach or disclosure.

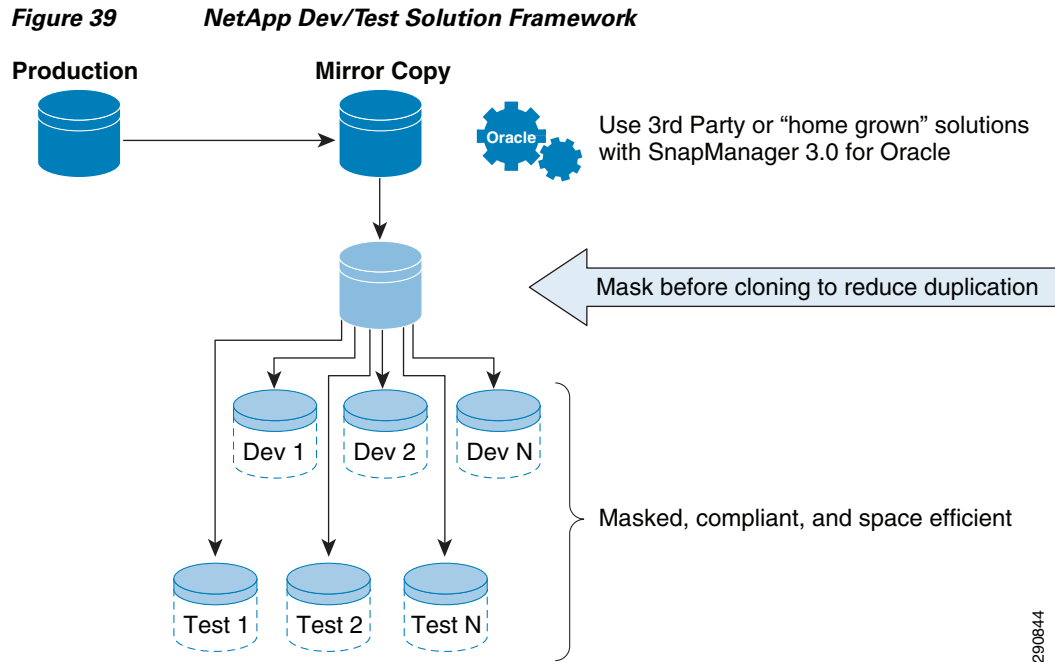
DgDiscover and DgMasker help administrators de-identify sensitive data for use in non-production environments. DgMasker's masking technology lets users create policies specific to their needs and makes the masking process repeatable. DgMasker allows data to be altered in a number of ways such as shuffling, character masking, Intellimasking™, randomization, format preserving masking, and

custom-defined algorithms. DgMasker's CUPS (Consistent, Unique, Persistent, and Synchronous) features further enhance masking options to generate masked data sets that preserve application and referential integrity and various other data constraints during the masking operation.

**Figure 38** *How Does DgMasker Work?*



NetApp and Dataguise products work together to provide a solution for rapidly sharing production data for non-production use while transparently protecting sensitive information from disclosure. NetApp SnapMirror® enables quick duplication of production data without disrupting running production applications. Dataguise sensitive data protection solutions, DgDiscover and DgMasker, reliably analyze the production data to locate and mask sensitive data. Writable clones of the masked data set can then be distributed quickly and efficiently with NetApp FlexClone®. The result is a solution for supporting key business processes such as test, development, and business analytics with fresh production data safely, efficiently, and with no impact on production systems.



DgMasker masks sensitive application data sets with a highly scalable, masking-in-place technology that optimally leverages the computing power and features of the database platform. Unlike ETL approaches, Dataguise's unique masking-in-place technology utilizes storage space most efficiently by de-identifying data within the rows and columns of existing database schemas. As a result, masking is performed in a remarkably short period of time to meet business SLA requirements.

With NetApp and Dataguise solutions organizations enjoy increased productivity in supporting their key business processes while simultaneously managing and reducing their sensitive data exposure risks.

## Product Installation Pre-requisites

DgDiscover and DgMasker support Oracle databases running on Linux, Unix, or Windows platforms.

DgDiscover and DgMasker require a Windows client for schema discovery and masking policy definition. Required configuration of the client includes:

- Microsoft Windows 2003/2008 or Windows XP/Vista/7
- Java 1.6+
- .net framework 3.5
- Oracle client (Administrator type) to connect to an Oracle database

## Setup

The test environment had the following configuration:

Dataguise DgDiscover and DgMasker are installed on a Microsoft Windows machine connected to the Master Clone database created using SnapManager. The objective was to create a masked copy of the Master database which could be efficiently cloned and distributed to support numerous test, development, and analytics uses.

- Install DgDiscover and DgMasker on Microsoft Windows machine.
- Create a “DG” user with the appropriate privileges using a script provided by Dataguise.

DgDiscover was used to perform a discovery operation against the target database to identify sensitive data and to guide in the creation of an appropriate masking policy. Steps included:

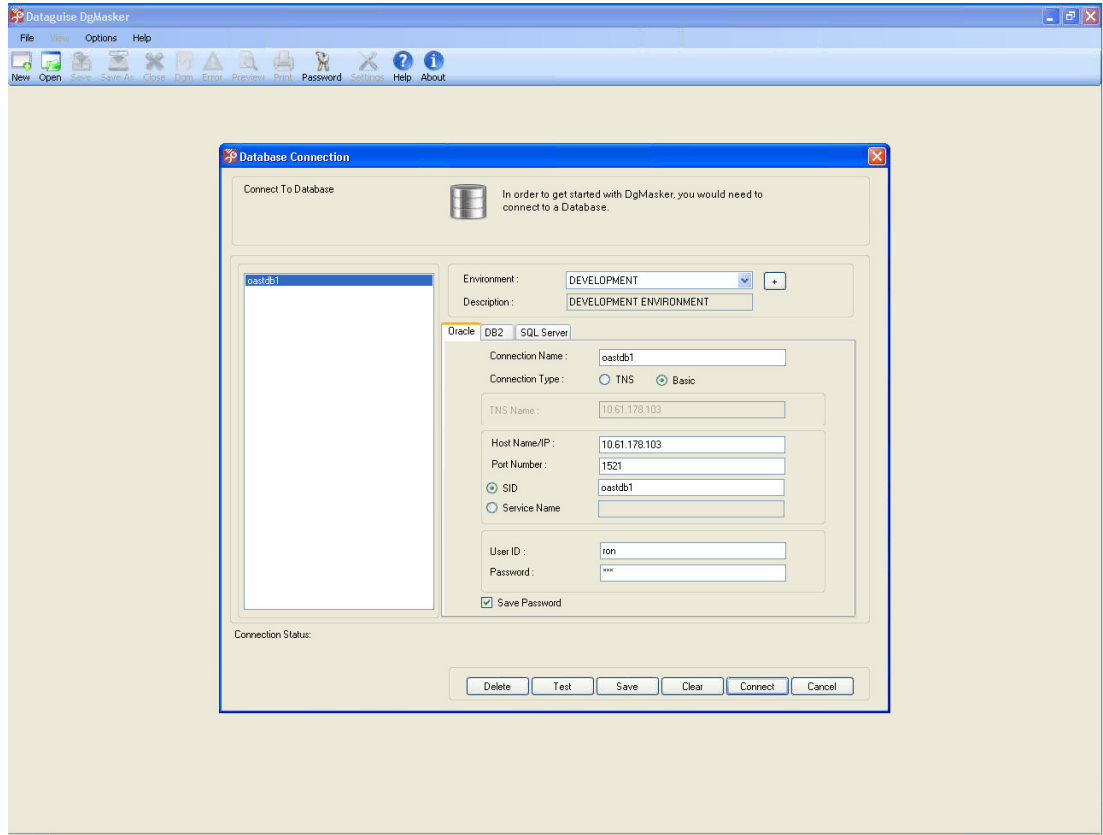
- Figure 40** *DgDiscover Results Against the Test Dataset*



DgMasker was used to connect to the database and define a masking policy, which allowed testers to:

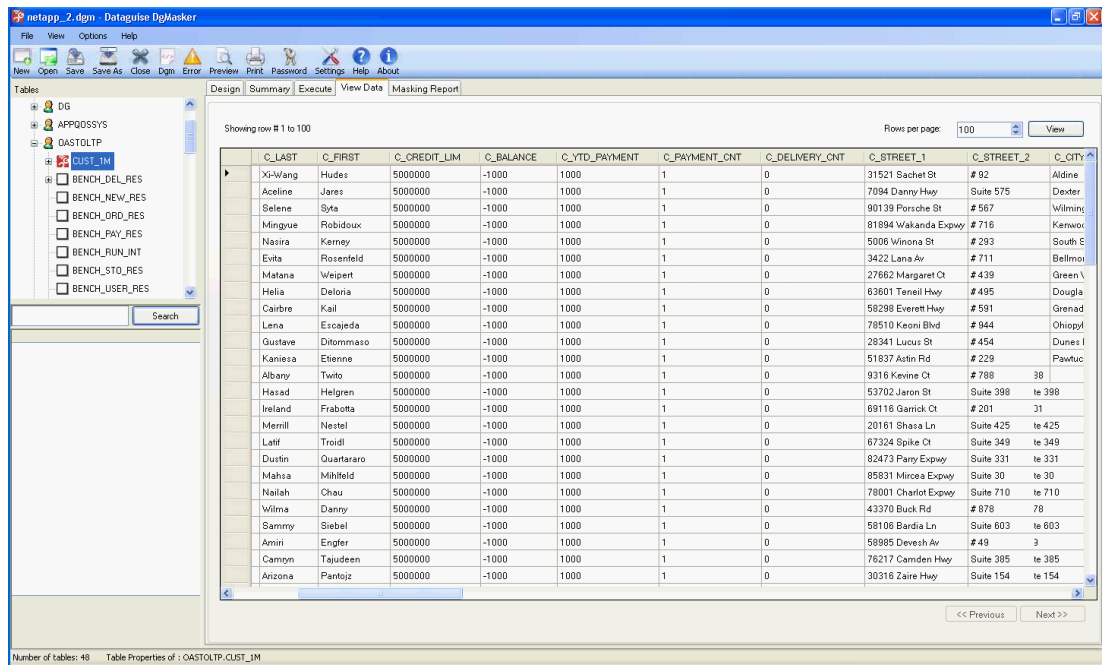
- 71

**Figure 41** Connecting to the Target Database Through DgMasker

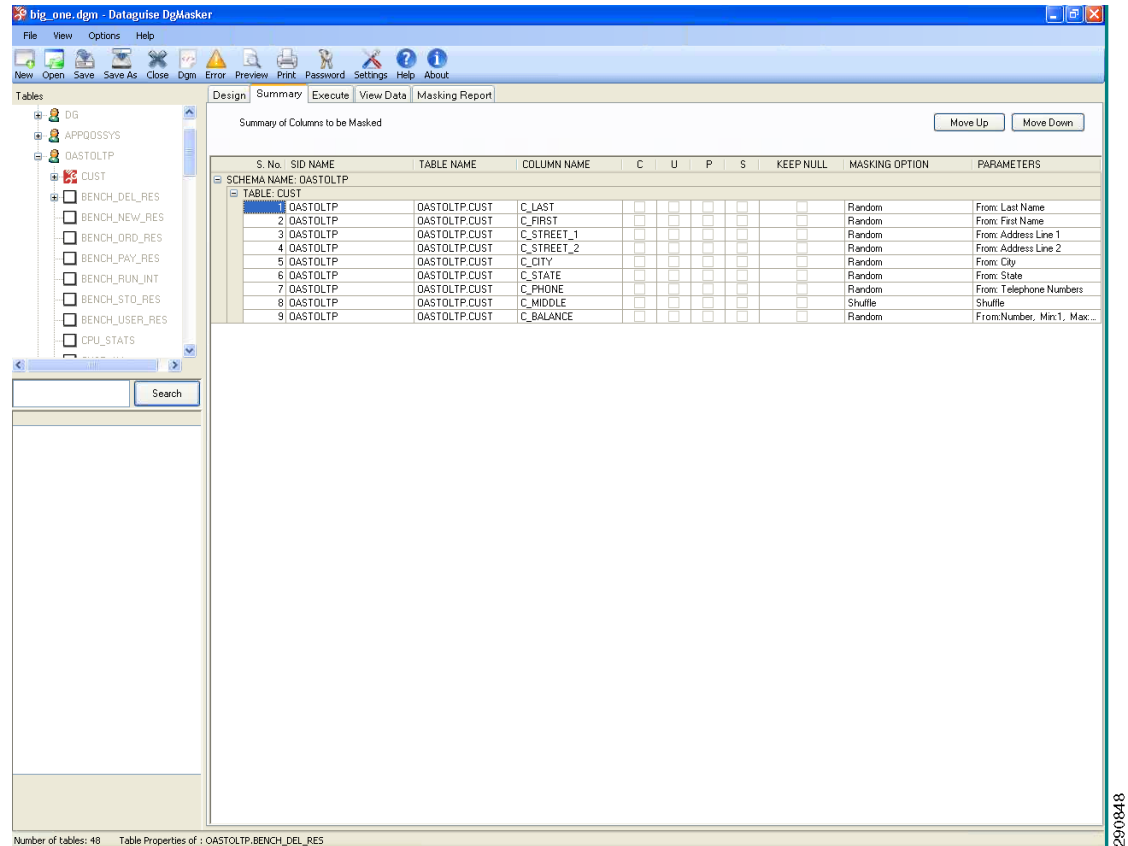


290846

**Figure 42** Viewing Table Data Through DgMasker



290847

**Figure 43** Summary of DgMasker Masking Policy

- Step 3** The objective of this step was to demonstrate generation and execution of a masking policy which could be integrated to run seamlessly with NetApp's cloning process:
1. Masking policy created in Step 2 was used to generate and load a native, PL/SQL language masking script into the target database.
  2. The masking operation was run interactively from the UCS console using the DgMasker Java client.
  3. The results were verified by inspecting the masked table using the DgMasker client.

## Results

Masking of the selected columns concluded without errors. DgMasker successfully masked nine columns of a 1.3 billion row database table in less than nine hours at the rate of over 150 million rows per hour. At the conclusion of the masking operation, NetApp FlexClone could be used to quickly generate writable, space-efficient clones of the masked data to support test, development, and analytics uses.

The testing demonstrated how Dataguise's sensitive data protection solutions complement NetApp's state-of-the-art, rapid cloning of Oracle data for distribution on the Cisco UCS platform. Dataguise DgDiscover and DgMasker products reliably identified sensitive data in the dataset for masking and generated a script suitable for integration as part of an automated production data cloning process. Furthermore, Dataguise DgMasker demonstrated the ability to mask large production datasets in a

remarkably short period of time to suit business process SLAs. By transparently protecting sensitive data in the data set with masking, enterprises using the solution have the ability to efficiently support their critical business applications while responsibly controlling their sensitive data exposure risks.

## Conclusion

Designed using a new and innovative approach to improve data center infrastructure, the Cisco UCS unites compute, network, storage access, and virtualization resources into a scalable, modular architecture that is managed as a single system.

For the Cisco UCS, Cisco has partnered with Oracle because Oracle databases and applications provide mission-critical software foundations for the majority of large enterprises worldwide. In addition, the architecture and large memory capabilities of the Cisco UCS connected to the industry proven and scalable NetApp storage system enable customers to scale and manage Oracle database environments in ways not previously possible.

Both database administrators and system administrators will benefit from the Cisco UCS combination of superior architecture, outstanding performance, and unified fabric. They can achieve demonstrated results by following the documented best practices for database installation, configuration, and management outlined in this document.

The workload performance testing included a realistic mix of OLTP and DSS which generated a sustained load on the four-node Oracle RAC configuration for periods of 24 and 12 hours. In addition, per the Oracle OAST recommendations, a three hour interconnect stress test was conducted. These workloads are designed to exceed the demands of typical database deployments and are meant to stress the system.

Despite the strenuous workload, the following high-performance metrics were achieved:

- The quad-core Intel Xeon 5500 series processors ramped up to nearly 90% of capacity during the IO\_CPU stress test without issues. Observed performance was consistent and at an expected level.
- The I/O demands generated by the load were effectively supported by the capabilities of the balanced NetApp storage array configuration. The array featured SAS 15K drives and NetApp's FlashCache (for OLTP tests). The system was designed to be a balanced configuration with the goal of eliminating obvious points of performance limiting bottlenecks.

In summary, the Cisco UCS is a game-changing computing model that uses integrated management and combines a wire-once unified fabric with an industry-standard computing platform.

The platform:

- Optimizes database environments
- Reduces total overall cost of the data center
- Provides dynamic resource provisioning for increased business agility

The benefits of the Cisco UCS include:

- Reducing total cost of ownership at the platform, site, and organizational levels
- Increasing IT staff productivity and business agility through just-in-time provisioning and mobility support for both virtualized and non-virtualized environments
- Enabling scalability through a design for up to 320 discrete servers and thousands of virtual machines in a single highly available management domain
- Using industry standards supported by a partner ecosystem of innovative, trusted industry leaders

## For More Information

Visit <http://www.cisco.com/en/US/netsol/ns944/index.html#>.

# Appendix A—System Configuration Settings

## Cisco UCS Kernel Settings (/etc/sysctl.conf)

This appendix provides the parameters for the Cisco UCS with 48 GB of RAM.

```
Kernel sysctl configuration file for Red Hat Linux
#
For binary values, 0 is disabled, 1 is enabled. See sysctl(8) and
sysctl.conf(5) for more details.

Controls IP packet forwarding
net.ipv4.ip_forward = 0

Controls source route verification
net.ipv4.conf.default.rp_filter = 1

Do not accept source routing
net.ipv4.conf.default.accept_source_route = 0

Controls the System Request debugging functionality of the kernel
kernel.sysrq = 0

Controls whether core dumps will append the PID to the core filename
Useful for debugging multi-threaded applications
kernel.core_uses_pid = 1

Controls the use of TCP syncookies
net.ipv4.tcp_syncookies = 1

Controls the maximum size of a message, in bytes
kernel.msgmnb = 65536

Controls the default maximum size of a message queue
kernel.msgmax = 65536

Controls the maximum shared segment size, in bytes
kernel.shmmax = 68719476736

Controls the maximum number of shared memory segments, in pages
kernel.shmall = 4294967296

Local Mods: Changes to support Oracle 11gR2 RAC
fs.aio-max-nr = 1048576
fs.file-max = 6815744
semaphores: semmsl, semmns, semopm, semmni
kernel.sem = 250 32000 100 128
net.ipv4.ip_local_port_range = 9000 65500
sunrpc.tcp_slot_table_entries = 128
net.core.rmem_default = 1342177
net.core.rmem_max = 16777216
net.core.wmem_default = 1342177
net.core.wmem_max = 16777216
net.ipv4.tcp_rmem = 4096 1342177 16777216
```

```
net.ipv4.tcp_wmem = 4096 1342177 16777216
net.ipv4.tcp_window_scaling = 1
net.ipv4.tcp_syncookies = 0
net.ipv4.tcp_sack = 0
net.ipv4.tcp_dsack = 0
net.core.netdev_max_backlog = 300000
```

## Appendix B—References

Cisco Unified Computing System: <http://www.cisco.com/en/US/partner/netsol/ns944/index.html>

Cisco Nexus 7000: <http://www.cisco.com/en/US/products/ps9402/index.html>

Cisco Nexus 5000: <http://www.cisco.com/en/US/products/ps9670/index.html>

Cisco MDS: <http://www.cisco.com/en/US/products/hw/ps4159/index.html>

Cisco DCNM:

[http://www.cisco.com/en/US/docs/switches/datacenter/sw/4\\_1/dcnm/fundamentals/configuration/guide/fund\\_overview.html](http://www.cisco.com/en/US/docs/switches/datacenter/sw/4_1/dcnm/fundamentals/configuration/guide/fund_overview.html)

NetApp ONTAP: <http://www.netapp.com/us/products/platform-os/data-ontap/>

NetApp Snapshot: <http://www.netapp.com/us/products/platform-os/snapshot.html>

NetApp FAS Platforms: <http://www.netapp.com/us/products>

NetApp Ethernet Storage: [http://www.netapp.com/us/company/leadership/ethernet-storage/NetApp Ethernet Storage](http://www.netapp.com/us/company/leadership/ethernet-storage/NetAppEthernetStorage)

NetApp 11g Best Practices: <http://media.netapp.com/documents/tr-3633.pdf>

NetApp Scalable DW: <http://media.netapp.com/documents/tr-3760.pdf>

## Appendix C—Bill of Material with Software Versions

This appendix includes a listing of the primary equipment and software required to build this solution.



### Note

This deployment guide follows the instructions and set up procedures specific to the software versions listed in this appendix. For the generic deployment scenario, it is recommended to consider the most recent software release available for each product. In general, the latest published software releases reduce known caveats. However, the published procedures and configuration guidelines may not always be directly applicable to the latest software releases.

**Table 16** *Bill of Material with Validated Software Versions*

| Part Number                         | Description            | SW Version | Quantity |
|-------------------------------------|------------------------|------------|----------|
| <b>UCS Solution: UCS-B Baseline</b> |                        | 1.3(1i)    | 1        |
| N20-S6100                           | Fabric Interconnect    |            | 2        |
| N20-C6508                           | Blade Server Chassis   |            | 2        |
| N20-I6584                           | 2104XP Fabric Extender |            | 4        |

**Table 16** *Bill of Material with Validated Software Versions*

| Part Number                    | Description                                                                              | SW Version  | Quantity |
|--------------------------------|------------------------------------------------------------------------------------------|-------------|----------|
| N20-B6620-1                    | UCS B200-M1 Blade Servers; dual 2.53 GHz CPU, 48 GB RAM (DDR3 1333 MHz)                  |             | 4        |
| N20-AC0002                     | UCS M81KR Virtual Interface Card/PCIe/2-port 10Gb                                        |             | 4        |
| <b>Nexus 5020</b>              |                                                                                          | 4.1(3)N2(1) | 2        |
| N5K-C5020P-B-S                 | Nexus 5020 Solutions Kit, 8G FC Unified Fabric (req SFP+)                                |             | 2        |
| N5KUK9-413N2.1                 | Nexus 5000 Base OS Software Rel 4.1(3)N2(1)                                              | 4.1(3)N2(1) | 2        |
| CAB-C13-C14-JMPR               | Recessed receptacle AC power cord 27                                                     |             | 4        |
| DS-SFP-FC8G-SW                 | 8 Gbps Fibre Channel SW SFP+, LC                                                         |             | 8        |
| N5020-ACC-KIT                  | N5020 Accessory Kit, Option                                                              |             | 2        |
| N5020-SSK9-LAB                 | Storage Protocol Services license for N5020 LAB Bundle                                   |             | 2        |
| N5K-M1060-LAB                  | N5K 8G FC Expansion Module for LAB Bundle                                                |             | 2        |
| N5K-PAC-750W                   | Nexus 5020 PSU module, 100-240VAC 750W                                                   |             | 4        |
| SFP-10G-SR                     | 10GBASE-SR SFP Module                                                                    |             | 8        |
| SFP-H10GB-CU5M                 | 10GBASE-CU SFP+ Cable 5 Meter                                                            |             | 32       |
| CON-SNT-N52SKL                 | SMARTNET 8X5XNBD Storage Protocol Ser                                                    |             | 2        |
| CON-SNT-N502BS                 | SMARTNET 8X5XNBD Nexus 5020 Solutions Kit, 8G FC Unified                                 |             | 2        |
| CON-SNT-N506L                  | SMARTNET 8X5XNBD N5K 8G FC Expansion Module for LAB                                      |             | 2        |
| <b>MDS 9124</b>                |                                                                                          | 3.2(1a)     | 2        |
| DS-C9124AP-K9                  | Cisco MDS 9124 4G Fibre Channel 24 port Switch                                           |             | 2        |
| DS-C24-300AC=                  | MDS 9124 Power Supply                                                                    |             | 4        |
| DS-C34-FAN=                    | FAN Assembly for MDS 9134                                                                |             | 4        |
| DS-SFP-FC4G-SW=                | 4 Gbps Fibre Channel-SW SFP, LC, spare                                                   |             | 48       |
| CON-SNT-24EV                   | SMARTNET MDS9124 8x5xNBD                                                                 |             | 2        |
| <b>NetApp Storage Hardware</b> |                                                                                          |             | 1        |
| FAS3170A-CHASSIS-R5-C          | FAS3170,ACT-ACT,Chassis,AC PS,-C,R5                                                      |             | 2        |
| FAS3170A-IB-BASE-R5            | FAS3170A,IB,ACT-ACT,OS,R5                                                                |             | 4        |
| X1938A-R5                      | PAM II 512GB Perf Accel Module FLASH PCIe, SupportEdge INC (Optional for OLTP workloads) |             | 4        |
| X1941A-R6-C                    | Cluster Cable 4X, Copper, 5M                                                             |             | 2        |
| DS4243-1511-24S-R5-C           | Disk Shelf, 24x450GB, 15K, 3Gb SAS,IOM3,-C,R5, SupportEdge INC                           |             | 8        |
| X6521-R6-C                     | Loopback, Optical, LC                                                                    |             | 4        |
| X6530-R6-C                     | Cable, Patch, FC SFP to SFP, 0.5M                                                        |             | 12       |
| X6539-R6-C                     | SFP, Optical, 4.25Gb                                                                     |             | 8        |
| X6553-R6-C                     | Optical Cable, 50u, 2GHz/KM, MM, LC/LC, 2M                                               |             | 12       |
| X1107A-R6                      | 2pt, 10GbE NIC, BareCage SFP+ Style, PCIe                                                |             | 4        |

**Table 16**      *Bill of Material with Validated Software Versions*

| Part Number                    | Description                                                | SW Version       | Quantity |
|--------------------------------|------------------------------------------------------------|------------------|----------|
| X-SFP-H10GB-CU5M-R6            | Cisco N50XX 10GBase Copper SFP+ cable, 5m                  |                  | 4        |
| X6536-R6                       | Optical Cable, 50u, 2000MHz/Km/MM, LC/LC, 5M               |                  | 8        |
| X6539-R6                       | Optical SFP, 4.25Gb                                        |                  | 8        |
| CS-O-4HR                       | SupportEdge Premium, 7x24, 4hr Onsite - 36 months          |                  | 1        |
| <b>NetApp Storage Software</b> |                                                            | Data ONTAP 7.3.4 |          |
| SW-T7C-NFS-C                   | NFS Software                                               |                  | 4        |
| SW-T7C-FLEXCLN-C               | Flexclone Software                                         |                  | 4        |
| SW-T7C-PAMII-C                 | PAMII Software (required only if Flash Cache is purchased) |                  | 4        |
| SW-T7C-SRESTORE-C              | SnapRestore Software                                       |                  | 4        |
| SW-T7C-DFM-OPSMGR              | Operations Manager                                         | 3.8              | 4        |
| SW-SSP-T7C-OPSMGR              | SW Subs, Operations Manager - 25 months                    |                  | 4        |