



Deploying Oracle 10gR2 Real Application Clusters on the Cisco Unified Computing System with EMC CLARiiON Storage

Cisco Validated Design

March 4, 2010

Introduction

This design guide describes how the Cisco Unified Computing System™ can be used in conjunction with EMC® CLARiiON® storage systems to implement an Oracle Real Application Clusters (RAC) system that is an Oracle Certified Configuration. The Cisco Unified Computing System 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 needs 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 Unified Computing System and provides instructions for implementing it; it concludes with an analysis of the cluster's performance and reliability characteristics.

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.

Cisco is the undisputed leader in providing network connectivity in enterprise data centers. With the introduction of the Cisco Unified Computing System, Cisco is now equipped to provide the entire clustered infrastructure for Oracle RAC deployments. The Cisco Unified Computing System provides compute, network, virtualization, and storage access resources that are centrally controlled and managed as a single cohesive system. With the capability to



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

© 2010 Cisco Systems, Inc. All rights reserved

scale to up to 320 rack-mount servers and incorporate both blade and rack-mount servers in a single system, the Cisco Unified Computing System provides an ideal foundation for Oracle RAC deployments.

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.

All components in an Oracle RAC implementation must work together flawlessly, and Cisco has worked closely with EMC and Oracle to create, test, and certify a configuration of Oracle RAC on the Cisco Unified Computing System. Cisco's Oracle Certified Configuration provides an implementation of Oracle Database 10g Release 2 and Oracle Database 11g Release 1 with Real Application Clusters technology consistent with industry best practices. For back-end Fibre Channel storage, it uses an EMC CLARiiON storage system with a mix of Fibre Channel drives and state-of-the-art Enterprise Flash Drives (EFDs) to further speed performance.

Benefits of the Configuration

The Oracle Certified Configuration of Oracle RAC on the Cisco Unified Computing System offers a number of important benefits:

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 rehosting 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 Oracle Certified 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.

High-Performance Platform for Oracle RAC

The Cisco UCS B-Series Blade Servers used in this certified 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 hyperthreading to deliver high performance when workloads demand and thermal conditions permit.

The patented Cisco Extended Memory Technology offers twice the memory footprint (384 GB) 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 Unified Computing System'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 Unified Computing System allow the system to easily process an intensive online transaction processing (OLTP) and decision-support system (DSS) workload with no resource saturation.

Safer Deployments with Certified and Validated Configurations

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

In addition to the certified Oracle RAC configuration described in this document, Cisco, Oracle and EMC have:

- Completed an Oracle Validated Configuration for Cisco Unified Computing System running Oracle Enterprise Linux running directly on the hardware or in a virtualized environment running Oracle VM
- Certified single-instance database implementations of Oracle Database 10g and 11g on Oracle Enterprise Linux and Red Hat Enterprise Linux 5.3

Document Objectives

This document introduces the Cisco Unified Computing System and discusses the ways it addresses many of the challenges that database administrators and their IT departments face today. This document provides an overview of the certified Oracle RAC configuration along with instructions for setting up the Cisco Unified Computing System and the EMC CLARiiON storage system, including database table setup and the use of EFDs. The document reports on Cisco's performance measurements for the cluster and a reliability analysis that demonstrates how the system continues operation even when hardware faults occur.

Solution Architecture

Certified Configuration Overview

The Cisco Unified Computing System 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 Unified Computing System uses a form-factor-neutral architecture that will allow Cisco C-Series Rack-Mount Servers to be integrated as part of the system using capabilities planned to follow the product's first customer shipment (FCS). Similarly, 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.

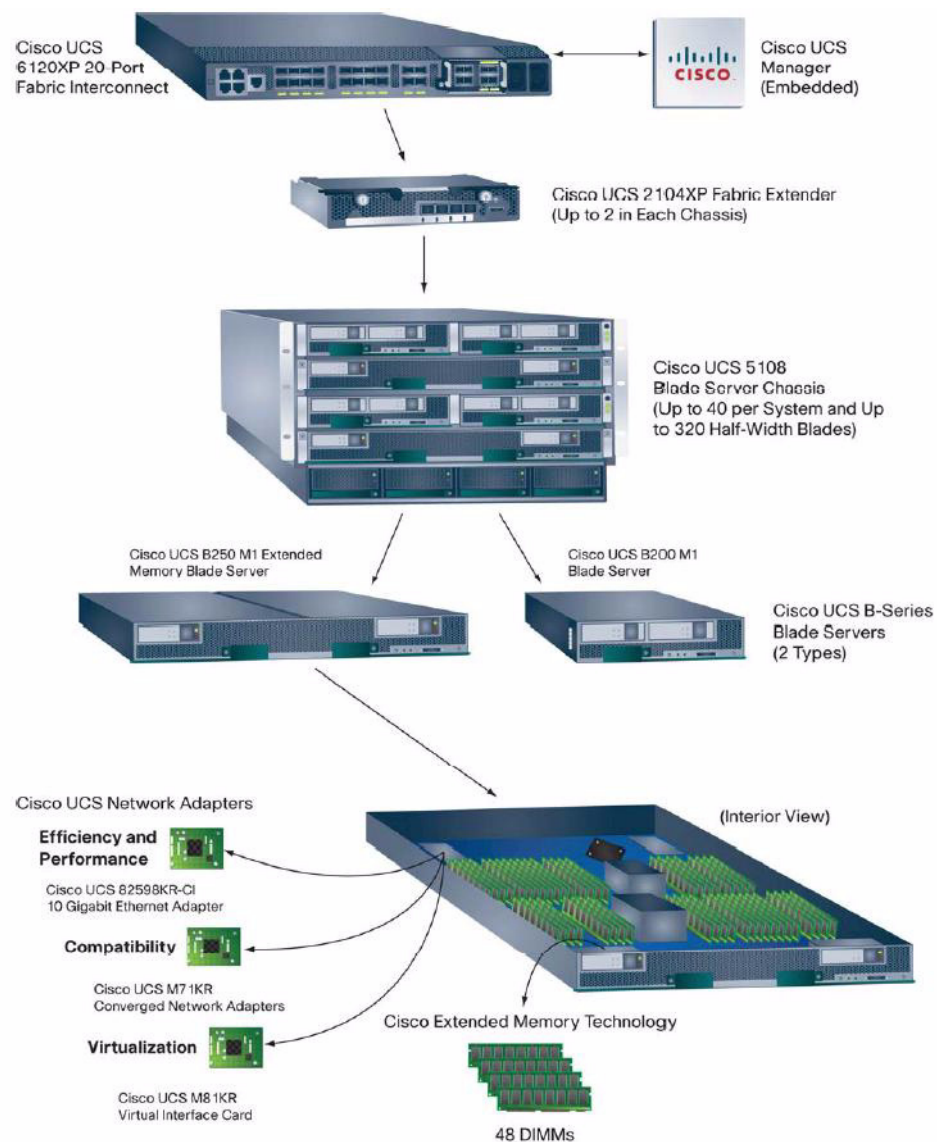
The system used to create the Oracle Certified Configuration is built from the hierarchy of components illustrated in [Figure 1](#):

- 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. Through an 8-port, 4-Gbps Fibre Channel expansion card, the interconnect provides native Fibre Channel access to the EMC CLARiiON storage system. Two fabric interconnects are configured in the cluster, providing physical separation between the public and private networks and also providing the capability to securely host both networks 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.
- The blade chassis supports up to eight half-width blades or up to four full-width blades. The certified configuration uses eight (four 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 5570) at 2.93 GHz. Each blade server was configured with 24 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 a Cisco UCS M71KR-Q QLogic Converged Network Adapter 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. (The testing process implemented the Cisco UCS M71KR-Q QLogic Converged Network Adapter).

Figure 1 Cisco Unified Computing System Components



Architecture Overview

The configuration presented in this document is based on the Oracle Database 10g Release 2 with Real Application Clusters technology certification environment specified for an Oracle RAC and EMC CLARiiON CX4-960 system (Figure 2).

Figure 2 *Oracle Database 10g with Real Application Clusters technology on Cisco Unified Computing System and EMC CLARiiON Storage*

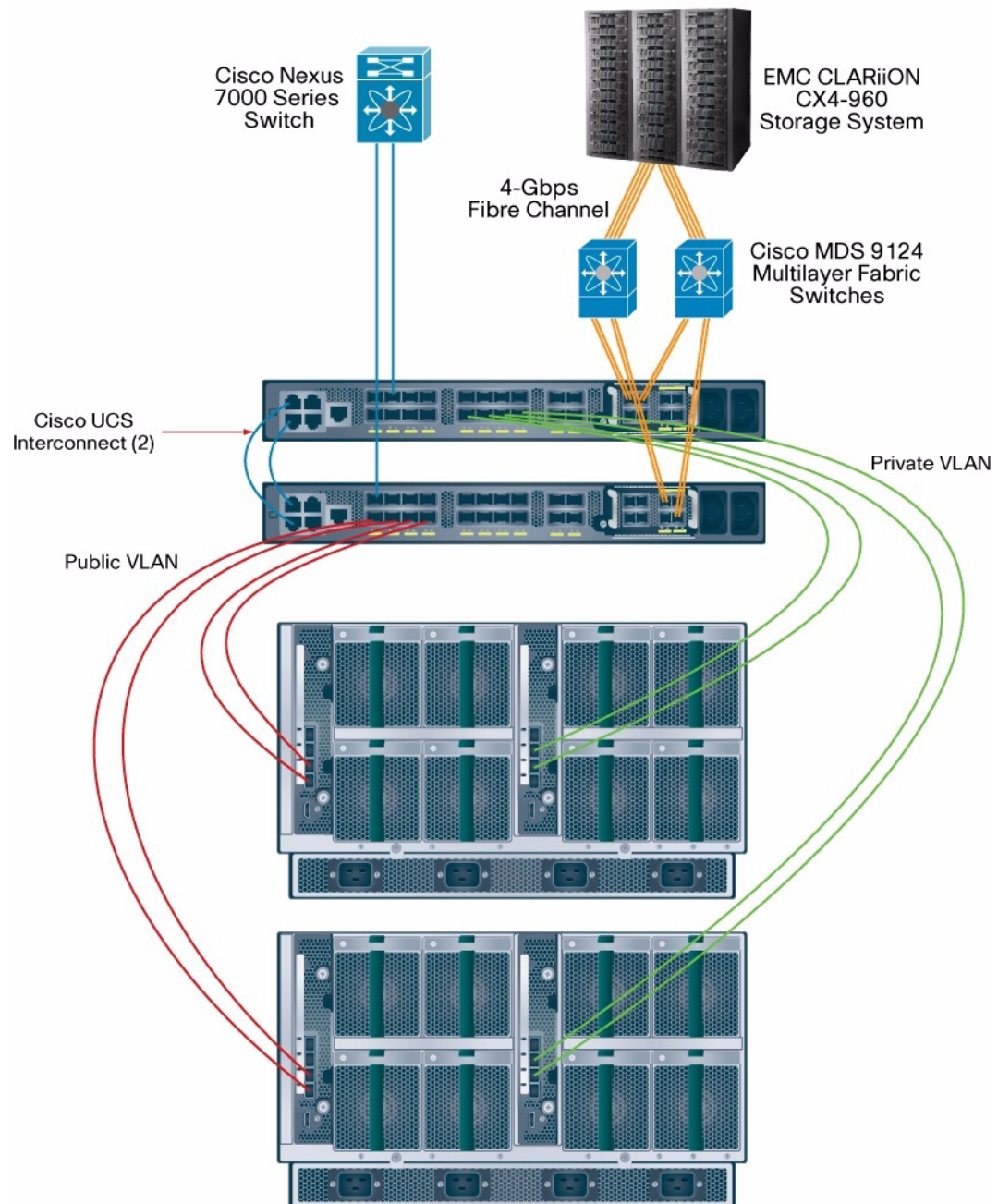


Figure 2 illustrates the 8-node configuration with EMC CLARiiON CX4-960 storage and Cisco Unified Computing System running Oracle Enterprise Linux (OEL) Version 5.3. This is a scalable configuration, that enables users to scale horizontally and internally in terms of processor, memory, and storage.

In the figure, the blue lines indicate the public network connecting to Fabric Interconnect A, and the green lines indicate the private interconnects connecting to Fabric Interconnect B. The public and private VLANs spanning the fabric interconnects help ensure the connectivity in case of link failure. Note that the FCoE communication takes place between the Cisco Unified Computing System chassis and fabric interconnects (red and green lines). This is a typical configuration that can be deployed in a customer's environment. The best practices and setup recommendations are described in subsequent sections of this document.

Detailed Topology

As shown in Figure 3, two chassis housing four blades each were used for this eight-node Oracle RAC solution. Tables 1 through 5 list the configuration details for all the server, LAN, and SAN components that were used for testing.

Figure 3 Detailed Topology of the Public Network and Oracle RAC Private Interconnects

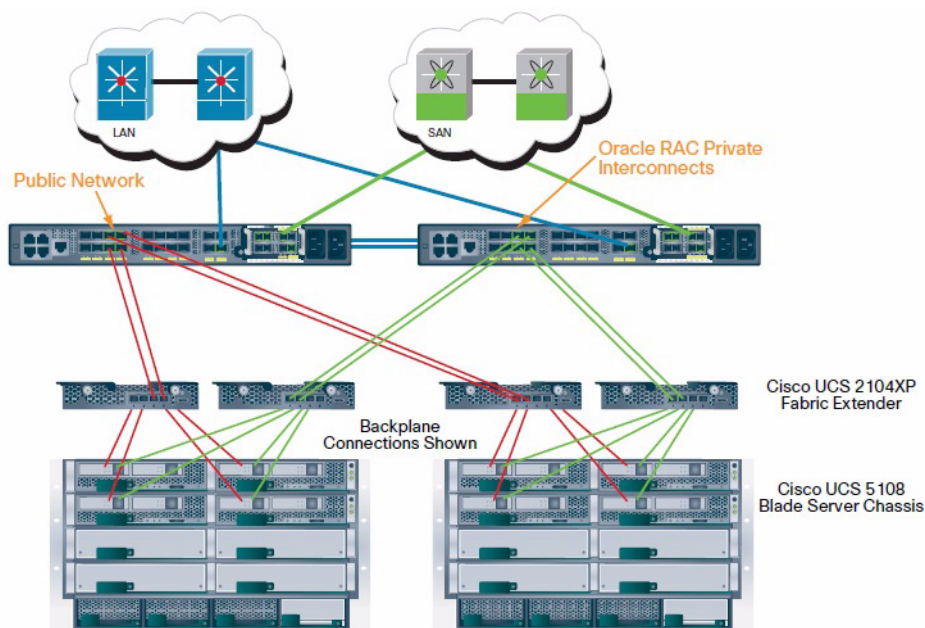


Table 1 **Physical Cisco Unified Computing System Server Configuration**

Quantity	Description
2	Cisco UCS 5108 Blade Server Chassis, with 4 power supply units, 8 fans, and 2 fabric extenders
8	Cisco UCS B200 M1 Blade Server
16	Quad-core Intel Xeon 5500 series 2.93-GHz processor (2 per blade server)
48	4-GB DDR3 DIMM, 1066 MHz (6 per blade, totaling 24 GB per blade server)
8	Cisco UCS M71KR-Q QLogic Converged Network Adapter, PCIe, 2 ports, and 10 Gigabit Ethernet (1 per blade server)
16	73-GB SAS, 15,000 RPM, Small Form-Factor (SFF) hard disk drive (HDD) hot pluggable (2 per blade)
2	Cisco UCS 6120XP 20-Port Fabric Interconnect with 2 power supply units and 2 fans
2	8-port, 4-Gbps Fibre Channel expansion port module
4	4-Gbps Fibre Channel Small Form-Factor Pluggable Plus (SFP+)
8	10GBASE-CU SFP+ cable (5 meters)
8	Fiber cables for connectivity to Fibre Channel and 10 Gigabit Ethernet

Table 2 **LAN Components**

Quantity	Description
1	Cisco Nexus 7000 Series Switch
VLAN Configuration	
1	Public VLAN
100	Private VLAN

Table 3 **SAN Components**

Quantity	Description
2	Cisco MDS 9124 Multilayer Fabric Switch

* This test used only one SAN and LAN switch.

Table 4 **Storage Configuration**

Quantity	Description
1	EMC CLARiiON CX4-960 Storage System
105	450-GB, 15,000 RPM, Fibre Channel spindles
15	73-GB EFDs

Table 5 **Software Components**

Description
Oracle Enterprise Linux 5.3
Oracle Database 10g Release 2 (10.2.0.1) with Real Application Clusters technology with Release 10.2.0.4 patch set

Data Center Server Platform—Cisco Unified Computing System

Cisco Unified Computing System Overview

Today, IT organizations assemble their data center environments from individual components. Their administrators spend significant amounts of time manually accomplishing basic integration tasks rather than focusing on more strategic, proactive initiatives. The industry is in a transition away from the rigid, inflexible platforms that result and moving toward more flexible, integrated, and virtualized environments.

The Cisco Unified Computing System™ is a next-generation data center platform that unites compute, network, storage access, and virtualization into a cohesive system designed to reduce total cost of ownership (TCO) and increase business agility. The system integrates a low-latency, lossless 10 Gigabit Ethernet unified network fabric with enterprise-class, x86-architecture servers. The system is an integrated, scalable, multichassis platform in which all resources participate in a unified management domain.

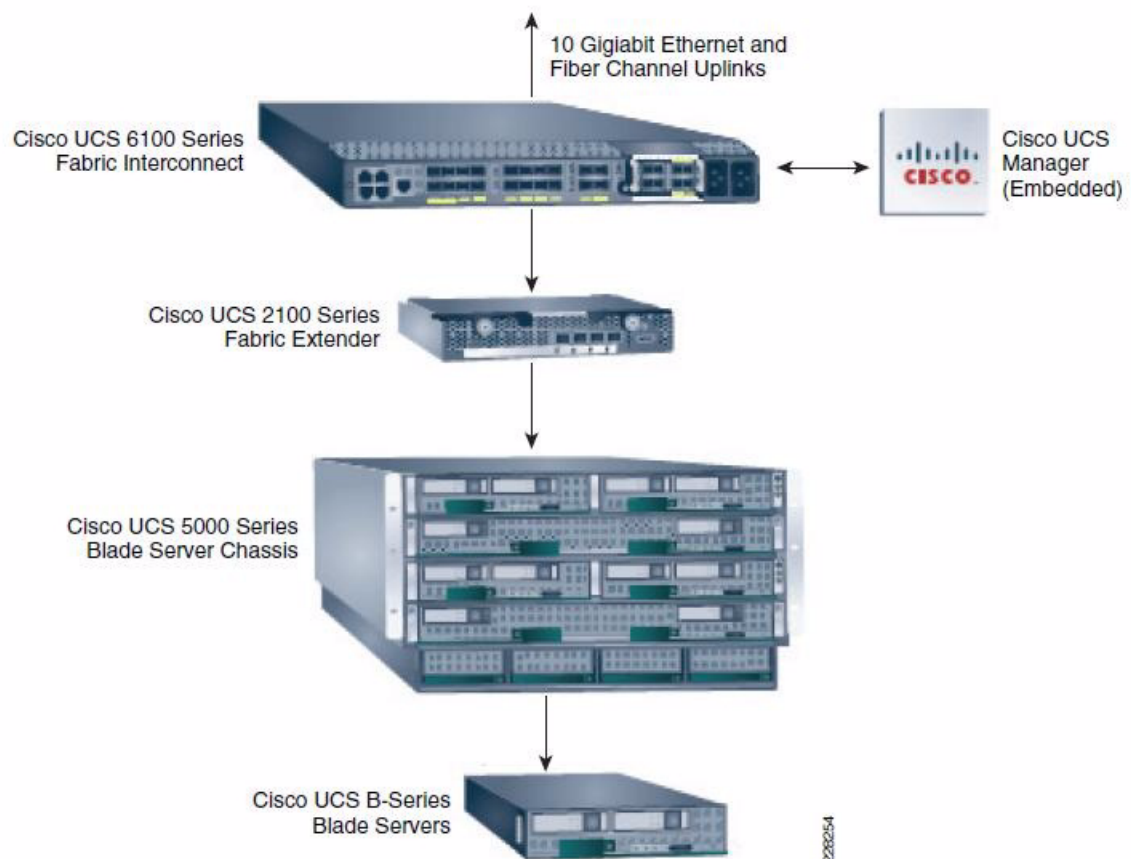
Managed as a single system whether it has one server or 320 servers with thousands of virtual machines, the Cisco Unified Computing System decouples scale from complexity. The Cisco Unified Computing System accelerates the delivery of new services simply, reliably, and securely through end-to-end provisioning and migration support for both virtualized and nonvirtualized systems. It provides the following benefits:

- **Embedded system management**—Management is uniquely integrated into all the components of the system, enabling the entire solution to be managed as a single entity through Cisco UCS Manager. Cisco UCS Manager provides an intuitive GUI, a command-line interface (CLI), and a robust API to manage all system configuration and operations. Cisco UCS Manager enables IT managers of storage, networking, and servers to collaborate easily on defining service profiles for applications.
- **Just-in-time provisioning with service profiles**—Cisco UCS Manager implements role- and policy-based management using service profiles and templates. Infrastructure policies—such as power and cooling, security, identity, hardware health, and Ethernet and storage networking—needed to deploy applications are encapsulated in the service profile. This construct improves IT productivity and business agility. Now infrastructure can be provisioned in minutes instead of days, shifting IT's focus from maintenance to strategic initiatives.
- **Unified fabric**—Cisco's unified fabric technology reduces cost by eliminating the need for multiple sets of adapters, cables, and switches for LANs, SANs, and high-performance computing networks. The system's fabric extenders pass all network traffic to parent fabric interconnects, where it can be processed and managed centrally, improving performance and reducing points of management. The unified fabric is a low-latency lossless 10-Gbps Ethernet foundation that enables a "wireonce" deployment model in which changing I/O configurations no longer means installing adapters and recabling racks and switches.
- **State-of-the-art performance**—Intel® Xeon® 5500 series processors automatically and intelligently adjust server performance according to application needs, increasing performance when needed and achieving substantial energy savings when not. Performance and power settings can also be manually configured.
- **Energy efficiency**—The system is designed for energy efficiency. Power supplies are 92 percent efficient and the Intel Xeon 5500 series processors use automated low-power states to better match power consumption with workloads. The simplified design of the Cisco UCS B-Series Blade Servers improves airflow efficiency and can reduce the number

of components that need to be powered and cooled by more than 50 percent compared to traditional blade server environments; similar component reduction can be achieved with the Cisco UCS C-Series Rack-Mount Servers.

Cisco Unified Computing System Component Details

Figure 4 *Cisco Unified Computing System*



The main system components include:

- UCS 5108 blade server chassis that fits on a standard rack and is 6RU high. Each UCS chassis can hold either eight half slot or four full slot blade servers, two redundant fabric extenders, eight cooling fans, and four power supply units. The cooling fans and power supply are hot swappable and redundant. The chassis requires only two power supplies for normal operation; the additional power supplies are for redundancy. The highly-efficient (in excess of 90%) power supplies, in conjunction with the simple chassis design that incorporates front to back cooling, makes the UCS system very reliable and energy efficient.
- UCS B-Series blade servers are based on an entirely new class of computing system that incorporates blade servers based on Intel Xeon 5500 Series processors. This design uses the UCS B200-M1 half-slot two socket blade servers, which has 12 DIMM slots (up to 96GB) per blade server and one dual-port Converged Network Adapter (CNA). Each UCS B200-M1 also has the following system features:
 - Two Intel Xeon 5500 series processors (quad cores)

- Two optional SAS/SATA hard drives
- Hot pluggable blades and hard disk drive support
- Blade service processor
- Stateless blade design
- 10 Gb/s CNA and 10 Gb/s Ethernet adapter options

I/O Adapters

The blade server has various Converged Network Adapters (CNA) options. The following two CNA options were used in this Cisco Validated Design:

- Efficient, High-Performance Ethernet with the Cisco UCS 82598KR-CI 10 Gigabit Ethernet Adapter

The Cisco UCS 82598KR-CI 10 Gigabit Ethernet Adapter is designed to deliver efficient, high-performance Ethernet connectivity. This adapter uses Intel silicon to present two 10 Gigabit Ethernet NICs to the peripheral component interconnect (PCI) device tree, with each NIC connected to one of the two fabric extender slots on the chassis. Like all the mezzanine cards available for the Cisco Unified Computing System, this card supports the Cisco DCE features needed to manage multiple independent network traffic streams over the same link. The adapter's MAC addresses are just-in-time configured by Cisco UCS Manager and the adapter is designed for:

- Network-intensive workloads, such as Web servers, in which all content is accessed over Network File System (NFS) or iSCSI protocols
- Environments in which efficiency and performance are important considerations
- QLogic converged network adapter (M71KR-Q) and Emulex converged network adapter (M71KR-E)

For organizations needing compatibility with existing data center practices that rely on Emulex or QLogic Fibre Channel HBAs, the Cisco UCS M71KR-E Emulex and UCS M71KR-Q QLogic Converged Network Adapters provide compatibility with interfaces from Emulex and QLogic, respectively. These CNAs use Intel silicon to present two 10 Gigabit Ethernet NICs and either two Emulex or two QLogic HBAs to the PCI device tree. The operating system sees two NICs and two HBAs and the existence of the unified fabric is completely transparent. A Cisco application-specific integrated circuit (ASIC) multiplexes one Ethernet and one Fibre Channel traffic stream onto each of the two midplane connections to the fabric extender slots. These CNAs are most appropriate for:

- Organizations that want to continue to use Emulex or QLogic drivers in the Cisco Unified Computing System
- Organizations that want to streamline the qualification process for new Fibre Channel hardware; use of standard HBA silicon allows use of HBA vendor-provided drivers
- Both traditional physical and virtualized environments

Cisco UCS 2100 Series Fabric Extenders

The Cisco UCS 2104XP Fabric Extender brings the I/O fabric into the blade server chassis and supports up to four 10-Gbps connections between blade servers and the parent fabric interconnect, simplifying diagnostics, cabling, and management. The fabric extender multiplexes and forwards all traffic using a cut-through architecture over one to four 10-Gbps unified fabric connections. All traffic is passed to the parent fabric interconnect, where network profiles are managed efficiently and effectively by the fabric interconnects. Each of up to two

fabric extenders per blade server chassis has eight 10GBASE-KR connections to the blade chassis midplane, with one connection to each fabric extender from each of the chassis' eight half slots. This configuration gives each half-width blade server access to each of two 10-Gbps unified fabric connections for high throughput and redundancy.

The benefits of the fabric extender design include:

- **Scalability**—With up to four 10-Gbps uplinks per fabric extender, network connectivity can be scaled to meet increased workload demands simply by configuring more uplinks to carry the additional traffic.
- **High availability**—Chassis configured with two fabric extenders can provide a highly available network environment.
- **Reliability**—The fabric extender manages traffic flow from network adapters through the fabric extender and onto the unified fabric. The fabric extender helps create a lossless fabric from the adapter to the fabric interconnect by dynamically throttling the flow of traffic from network adapters into the network.
- **Manageability**—The fabric extender model extends the access layer without increasing complexity or points of management, freeing administrative staff to focus more on strategic than tactical issues. Because the fabric extender also manages blade chassis components and monitors environmental conditions, fewer points of management are needed and cost is reduced.
- **Virtualization optimization**—The fabric extender supports Cisco VN-Link architecture. Its integration with VN-Link features in other Cisco UCS components, such as the fabric interconnect and network adapters, enables virtualization-related benefits including virtual machine-based policy enforcement, mobility of network properties, better visibility, and easier problem diagnosis in virtualized environments.
- **Investment protection**—The modular nature of the fabric extender allows future development of equivalent modules with different bandwidth or connectivity characteristics, protecting investments in blade server chassis.
- **Cost savings**—The fabric extender technology allows the cost of the unified network to be accrued incrementally, helping reduce costs in times of limited budgets. The alternative is to implement and fund a large, fixed-configuration fabric infrastructure long before the capacity is required.

UCS 6100 XP Series Fabric Interconnect

The UCS 6100 XP fabric interconnect is based on the Nexus 5000 product line. However, unlike the Nexus 5000 products, it provides additional functionality of managing the UCS chassis with the embedded UCS manager. A single 6140 XP switch can support up to 40 chassis or 320 servers with half slot blades.

Some of the salient features provided by the switch are:

- 10 Gigabit Ethernet, FCoE capable, SFP+ ports
- 20 and 40 fixed port versions with expansion slots for additional Fiber Channel and 10 Gigabit Ethernet connectivity
- Up to 1.04 Tb/s of throughput
- Hot pluggable fan and power supplies with front-to-back cooling system
- Hardware based support for Cisco VN-Link technology
- Can be configured in a cluster for redundancy and failover capabilities

In this solution, two UCS 6120 Fabric Interconnects were configured in a cluster pair for redundancy and were configured in switch mode to be comparable with traditional Layer 2 switches in data centers today.

Configuring Cisco Unified Computing System for the Eight-Node Oracle RAC

This section describes how to configure the solution components and shows the specific configurations that were implemented during the validation of this solution. The information is presented in the order in which the solution should be implemented based on dependencies among the solution components.

While this section provides details on CLI commands and GUI steps, it is not meant to be a configuration guide for any of the products in this solution as existing product documentation details how to configure the various features of each of these products. Therefore, this section details commands and GUI screenshots for configuration steps that will save the reader time, but where the instructions would be too lengthy for this document, URLs are provided for existing documentation.

Configuring the Cisco UCS 6120XP Fabric Interconnect

The Cisco UCS 6120XP Fabric Interconnect is configured in a cluster pair for redundancy. It provides resiliency and access to the system configuration data in the rare case of hardware failure.

For fabric interconnects, the configuration database is replicated from the primary switch to the standby switch. All operations are transaction-based, keeping the data on both switches synchronized.

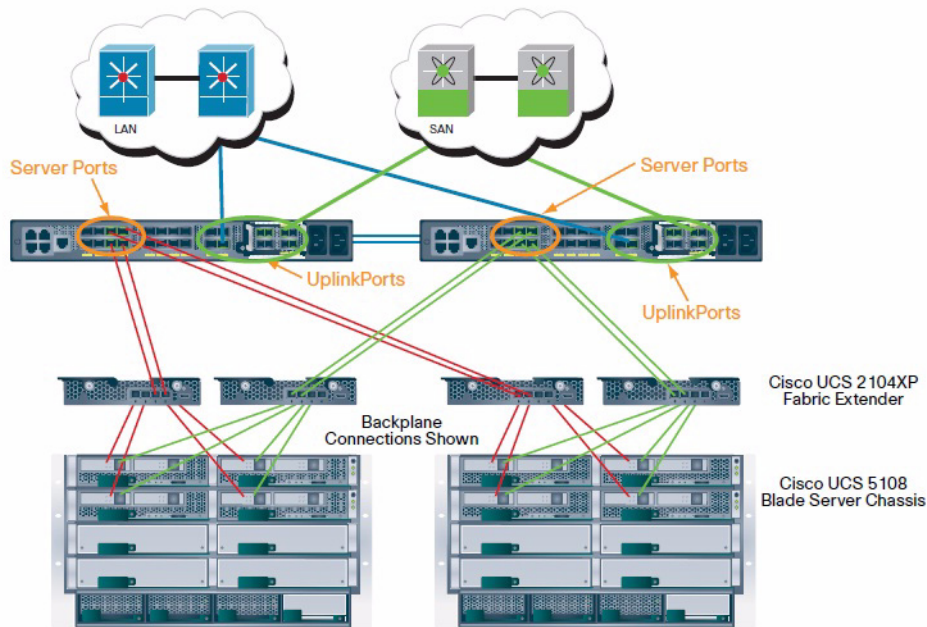
**Note**

Detailed information about the fabric interconnect configuration is beyond the scope of this document. For more information, refer to the Cisco Unified Computing System documentation at:
http://www.cisco.com/en/US/docs/unified_computing/ucs/sw/gui/config/guide/b_GUI_Config_Guide.html.

Configuring the Server Ports

The first step is to establish connectivity between the blades and fabric interconnects. As shown in [Figure 5](#), four public (two per chassis) links go to Fabric Interconnect A (ports 5 through 8). Similarly, four private links go to Fabric Interconnect B. Configure these ports as server ports, as shown in [Figure 5](#).

Figure 5 *Physical Connectivity and Port Configuration*



Configuring Uplinks to the SAN and LAN

At this time, configure the uplink Fibre Channel ports (Expansion Module 2). SAN connectivity is discussed later in this document.

Configuring the SAN and LAN on Cisco UCS Manager

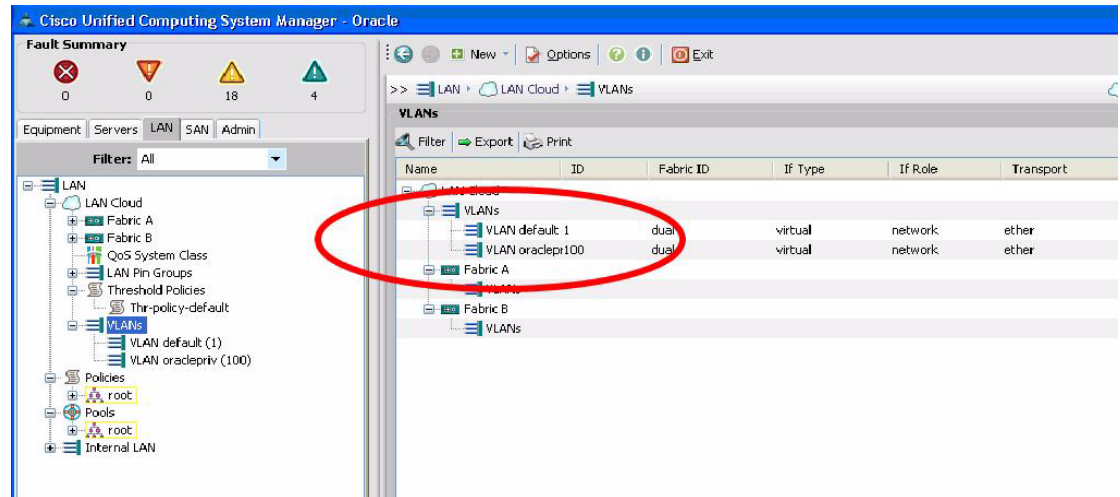
Before configuring the service profile, do the following:

- Configure the SAN: On the SAN tab, set the VSANs to be used in the SAN (if any). You should also set up pools for world wide names (WWNs) and world wide port names (WWPNs) for assignment to the blade server virtual HBAs (vHBAs). For detailed information, go to http://www.cisco.com/en/US/docs/unified_computing/ucs/sw/gui/config/guide/GUI_Config_Guide_chapter20.html.
- Configure the LAN: On the LAN tab, set the VLAN assignments to the virtual NICs (vNICS). You can also set up MAC address pools for assignment to vNICS. For this setup, the default VLAN (VLAN ID 1) was used for public interfaces, and a private VLAN (VLAN ID 100) was created for Oracle RAC private interfaces



Note

It is very important that you create a VLAN that is global across both fabric interconnects. This way, VLAN identity is maintained across the fabric interconnects in case of failover.

Figure 6 Two VLANs

After these preparatory steps have been completed, you can generate a service profile template for the required hardware configuration. You can then create the service profiles for all eight nodes from the template.

Setting up Service Profiles

Service profiles are the central concept of the Cisco Unified Computing System. Each service profile serves a specific purpose: to help ensure that the associated server hardware has the configuration required to support the applications it will host.

The service profile maintains configuration information about:

- Server hardware
- Interfaces
- Fabric connectivity
- Server and network identity

This information is stored in a format that can be managed through Cisco UCS Manager. All service profiles are centrally managed and stored in a database on the fabric interconnect.

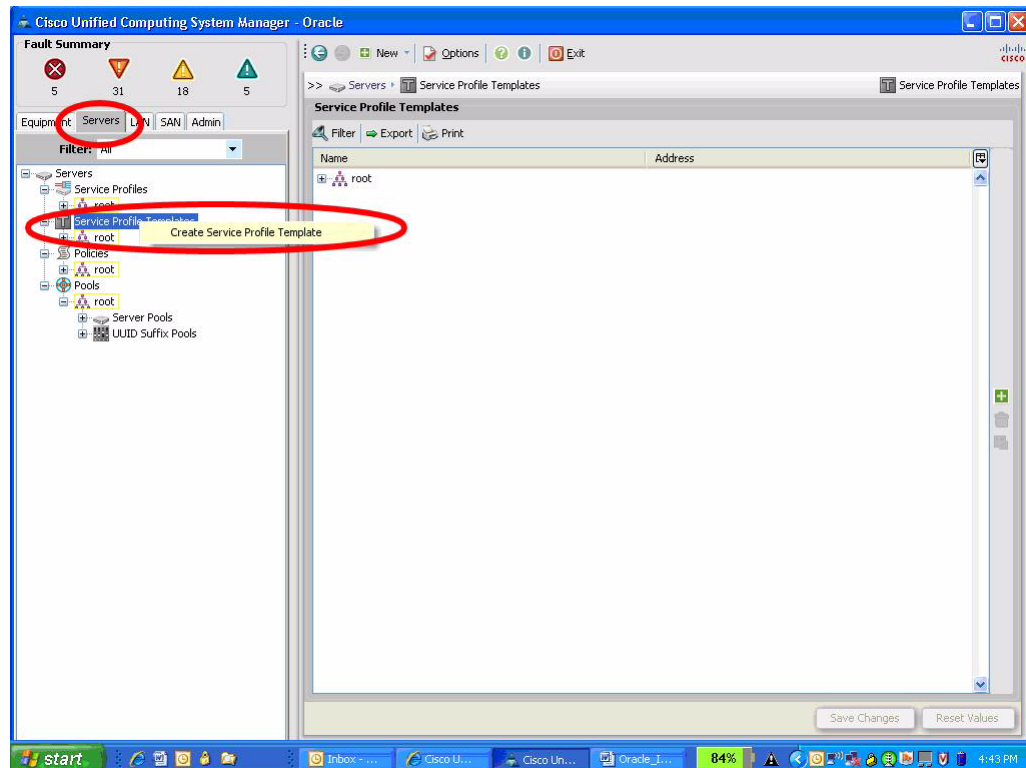
The service profile consists of the following information:

- Identity and personality information for the server
 - Universally unique ID (UUID)
 - World wide node name (WWNN)
 - Boot order
- LAN and SAN configuration (through the vNIC and vHBA configuration)
 - NIC and HBA identity (MAC addresses and WWN and WWPN information)
 - Ethernet NIC profile (flags, maximum transmission unit [MTU], etc.)
 - VLAN and VSAN connectivity information
- Various policies (disk scrub policy, quality of service [QoS], etc.). For Oracle certification testing, no policies were used.

Creating the Service Profile Template

To create the service profile template, follow these steps:

- Step 1** From the Service Profile Templates screen:
- Click the **Servers** tab.
 - Right-click **Service Profile Template**.



The Identify Service Profile Template screen displays.

Create Service Profile Template

Unified Computing System Manager

Create Service Profile Template

1. ☒ **Identify Service Profile Template**
2. ☐ Storage
3. ☐ Networking
4. ☐ Server Boot Order
5. ☐ Server Assignment
6. ☐ Operational Policies

Identify Service Profile Template

You must enter a name for the service profile template and specify the template type. You can also specify how a UUID will be assigned to this template and enter a description.

Name:

The template will be created in the following organization. Its name must be unique within this organization.

Where: **org-root**

The template will be created in the following organization. Its name must be unique within this organization.

Type: ☒ Initial Template ☐ Updating Template

Specify how the UUID will be assigned to the server associated with the service generated by this template.

UUID

UUID Assignment:

Select UUID assignment option.
If no selection is made, the UUID will be assigned from the default pool.
WARNING: The selected pool does not contain any available entities.
You can select it, but it is recommended that you add entities to it.

Optionally enter a description for the profile. The description can contain information about when and where the service profile should be used.

< Prev **Next >** Finish Cancel

Step 2 In the **Name** field, enter the template name (such as Oracle_RAC).

Step 3 For the template type, select **Initial Template**.

Initial templates create new service profiles with the same attributes, but the child service profiles are not updated when a change is made to the original template. If you select Updating Template, child profiles will immediately be updated when a change to the template is made, potentially making all the dependent child profiles to cause servers to reboot, so you should use updating templates with care.

Step 4 Click **Next**.

The Storage screen displays.

Create Service Profile Template

1. ☒ Identify Service Profile Template
 2. ☒ **Storage**
 3. ☒ Networking
 4. ☐ Server Boot Order
 5. ☐ Server Assignment
 6. ☐ Operational Policies

Unified Computing System Manager

Storage

Optionally specify disk policies and SAN configuration information.

Select a local disk configuration policy.

Local Storage: If nothing is selected, the default Local Storage configuration policy will be assigned to this service profile.

Create Local Disk Configuration Policy

Scrub Policy: Create Scrub Policy

How would you like to configure SAN storage? ☐ Simple ☒ **Expert** ☐ No vHBAs

A server is identified on a SAN by its World Wide Node Name (WWNN). Specify how the system should assign a WWNN to the server associated with this profile.

World Wide Node Name

WWNN Assignment:

Select WWNN assignment option.
 If nothing is selected, the WWNN will be assigned from the default pool.
WARNING: The selected pool does not contain any available entities.
 You can select it, but it is recommended that you add entities to it.

Name	WWPN	Order

Move Up Move Down Delete **Add** Modify

- Step 5** To create vHBAs for SAN storage:
- In the How would you like to configure SAN storage? options, select **Expert**.
 - Click **Add** to add an HBA.

The Create vHBA screen displays.

Create vHBA

Name: **vHBA1**

World Wide Port Name: default(16/32)

Use SAN Connectivity Template: ☐

WWPN Assignment: default(16/32)

The WWPN will be assigned from the selected pool.
The available/total WWPNs are displayed after the pool name.

+ Create vHBA Template

Fabric ID: ☒ A ☐ B

Select VSAN: **VSAN default**

Pin Group: <not set>

Persistent Binding: ☒ disabled ☐ enabled

Operational Parameters

Adapter Performance Profile

Adapter Policy: <not set>

OK Cancel

Step 6 In the Name field, enter **vHBA1**.

Step 7 In the Select VSAN drop-down list, choose **VSAN default**.

For simplicity, this configuration uses the default VSAN for both HBAs. You may need to make a different selection depending on what is appropriate for your configuration.

Step 8 If you have created SAN pin groups for pinning Fibre Channel traffic to a specific Fibre Channel port, specify appropriate pin groups, using the Pin Group drop-down list.

Pinning in a Cisco Unified Computing System is relevant only to uplink ports, where you can pin Ethernet or FCoE traffic from a given server to a specific uplink Ethernet (NIC) port or uplink (HBA) Fibre Channel port. When you pin the NIC and HBA of both physical and virtual servers to uplink ports, you get finer control over the unified fabric. This control helps ensure better utilization of uplink port bandwidth. However, manual pinning requires an understanding of network and HBA traffic bandwidth across the uplink ports. The configuration described here does not use pin groups.



Note

The screenshot shows the configuration for vHBA1 assigned to Fabric Interconnect A.

Step 9 Click **OK**.

Step 10 From the Storage screen, create the second vHBA for SAN storage:

- a. Click **Add** to add an HBA.

Step 11 From the Create vHBA screen, create the second vHBA:

- a. In the Name field, enter **vHBA2**.
- b. In the Select VSAN drop-down list, choose **VSAN default**.

For simplicity, this configuration uses the default VSAN for both HBAs. You may need to make a different selection depending on what is appropriate for your configuration.

- c. If you have created SAN pin groups for pinning Fibre Channel traffic to a specific Fibre Channel port, specify appropriate pin groups, using the Pin Group drop-down list.
- d. Click **OK**.

This screenshot shows the configuration for vHBA2 assigned to Fabric Interconnect B.

Step 12 From the Storage screen, click **Finish**.

Two vHBAs have now been created, which completes the SAN configuration.

Creating vNICs and Associating with VLANs

To create the vNICs and associate them with the appropriate VLANs, follow these steps:

- Step 1** From the Networking screen:
- In the How would you like to configure LAN connectivity? options, select **Expert**.
 - Click **Add**.

The screenshot shows the 'Unified Computing System Manager' interface. On the left, a sidebar lists steps for 'Create Service Profile Template', with 'Networking' selected. The main panel is titled 'Networking' and contains the question 'How would you like to configure LAN connectivity?' with three radio buttons: 'Simple', 'Expert' (selected), and 'No vNICs'. Below this, a table with columns 'Name', 'MAC Address', 'Order', 'Fabric ID', and 'Native VLAN' is shown. At the bottom of the table, the 'Add' button is circled in red. Navigation buttons at the bottom include '< Prev', 'Next >', 'Finish', and 'Cancel'.

The Create vNICs screen displays.

Step 2 In the Name field, enter **vNIC1**.

Step 3 b) For the Fabric ID options, select **Fabric A** and **Enable Failover**.

Step 4 c) For the VLAN Trunking options, select **Yes**.

VLAN trunking allows multiple VLANs to use a single uplink port on the system.

Step 5 In the VLANs area, select the associated check boxes for **default** and **oraclepriv**.

Step 6 Click **OK**.

vNIC1 is now assigned to use Fabric Interconnect A for the public network.

Step 7 From the Networking screen, create the second vNIC.

a. Click **Add** to add vNIC2.

b. From the Create vNICs screen, in the Name field, enter **vNIC2**.

c. b) For the Fabric ID options, select **Fabric B** and **Enable Failover**.

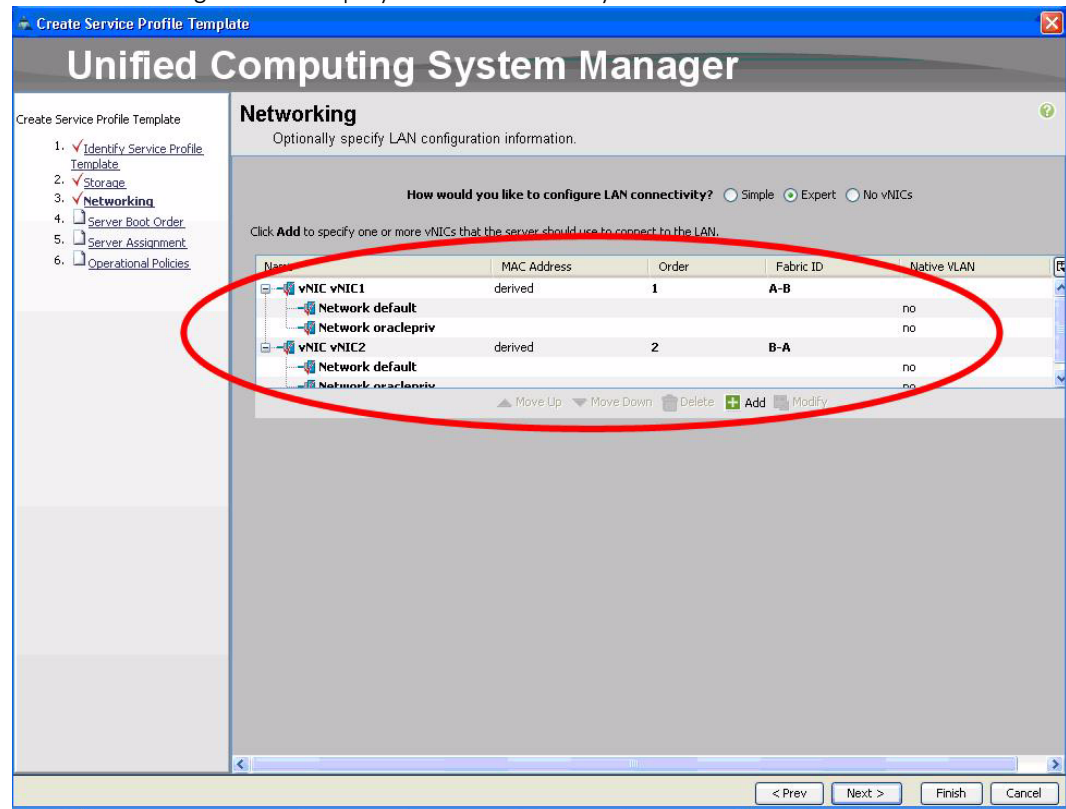
d. c) For the VLAN Trunking options, select **Yes**.

e. d) In the VLANs area, select the associated check boxes for **default** and **oraclepriv**.

Step 8 Click **OK**.

vNIC2 is now assigned to use Fabric Interconnect B for the Oracle RAC private network.

The Networking screen displays the vNICs that you have created.



The setup created here did not use SAN boot or any other policies. You can configure these in the screens that follow the Networking screen. You may be required to configure these policies if you choose to boot from the SAN or if you associate any specific policies with your configuration.

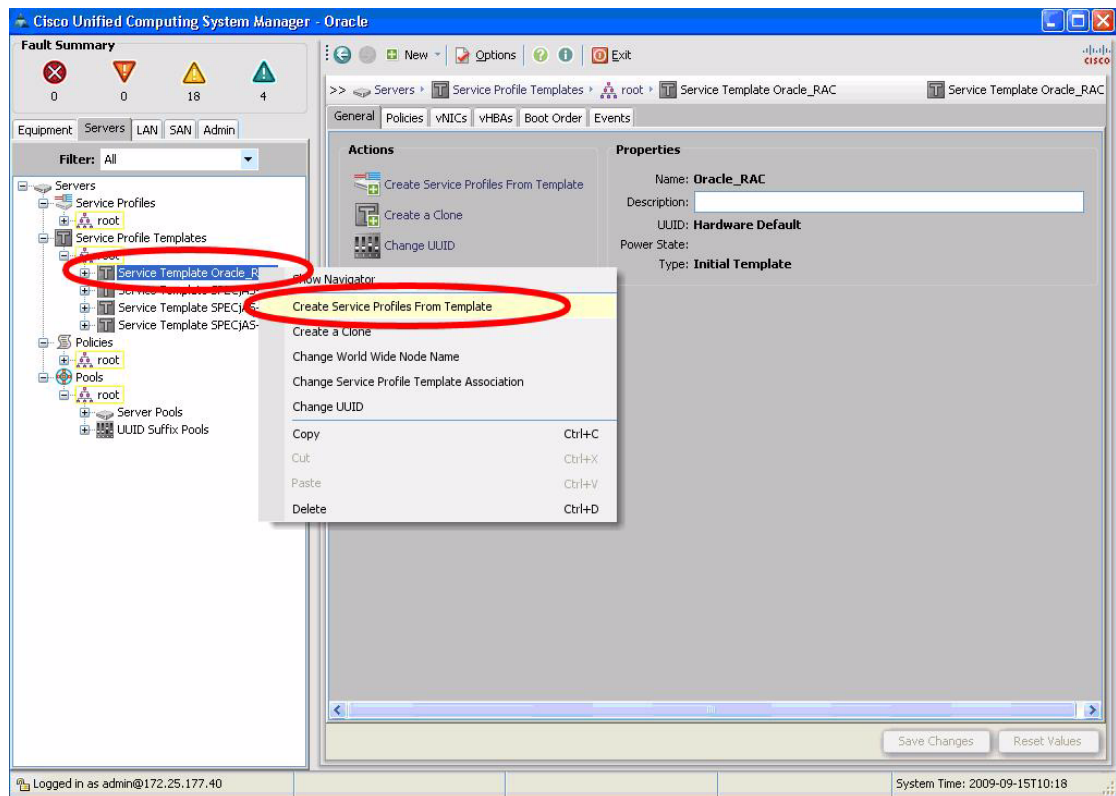
Step 9 Click **Finish** to complete the service profile template.

Creating and Associating Service Profiles with Blade Servers

To create eight service profiles and associate them with individual Blade Servers, follow these steps:

Step 1 From the Cisco Unified Computing System Manager screen:

- a. Right-click **Service Template Oracle_RAC**.
- b. Select **Create Service Profiles From Template**.

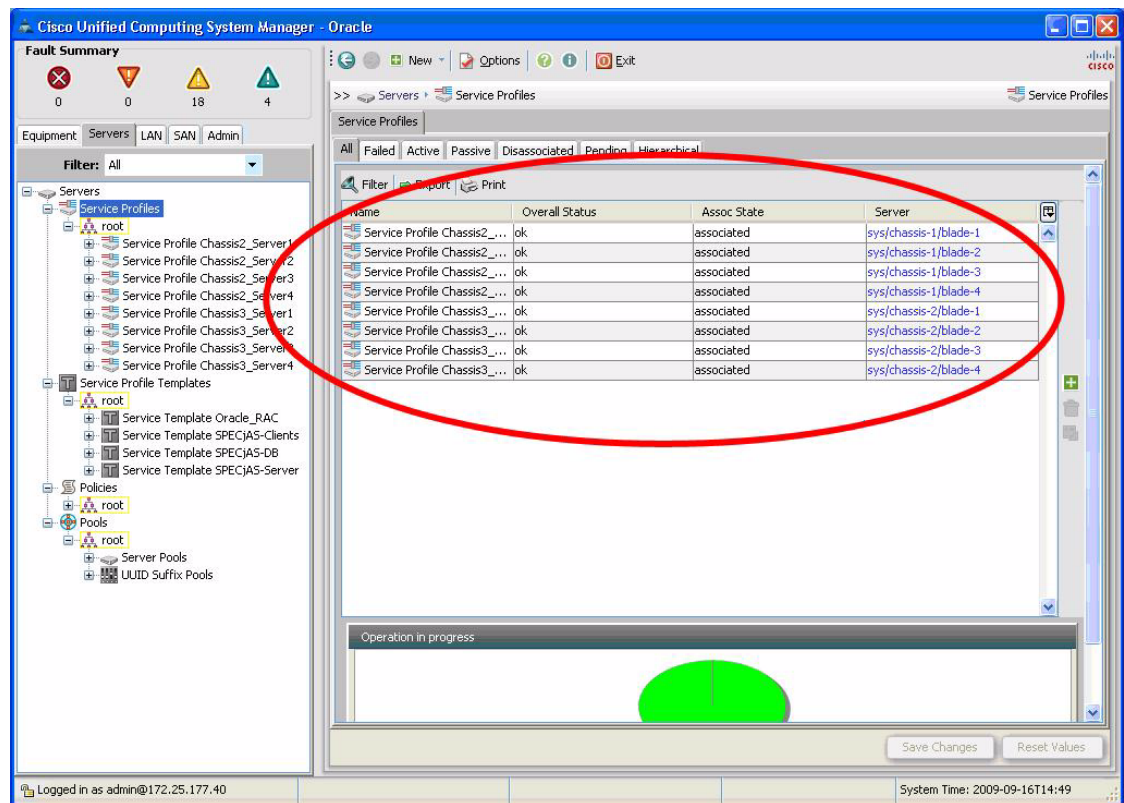


The Create Service Profiles From Template dialog displays.

- Step 2** In the Naming Prefix field, enter **RAC**.
- Step 3** In the Number field, enter **8**.
- Step 4** Click **OK**.

This step creates service profiles for all eight blade servers (see screenshot below). When the service profiles are created, they will pick unique MAC address, WWN, and WWPN values from the resource pools created earlier.

Now you can associate the profiles with the appropriate blade servers in the chassis.



Configuring the SAN Switch and Zoning

The fabric interconnects are connected to a SAN switch that also provides connectivity to storage.

To configure the SAN switch, follow these steps:

Step 1 Ensure the following configuration details are implemented:

- The NPIV feature must be enabled on the Cisco MDS 9124 Multilayer Fabric Switch.
NPIV allows a Fibre Channel host connection or N-Port, to be assigned multiple N-Port IDs or Fibre Channel IDs (FCIDs) over a single link. All FCIDs assigned can now be managed on a Fibre Channel fabric as unique entities on the same physical host. Different applications can be used in conjunction with NPIV. In a virtual machine environment where many host operating systems or applications are running on a physical host, each virtual machine can now be managed independently from zoning, aliasing, and security perspectives. For detailed information, go to:
http://www.cisco.com/en/US/prod/collateral/ps4159/ps6409/ps5989/ps9898/white_paper_c11-459263.html.
- The 4-GB SPF+ modules must be connected to the Cisco UCS 6100 Series Fabric Interconnect with the port
- Mode and speed set to AUTO.
- If you have created different VSANs, be sure to associate each Fibre Channel uplink with the correct VSAN.

Step 2 Refer to established SAN and zoning best practices for your setup.

Step 3 Complete the zoning.

Table 6 lists the zones and their associated members that are used in the testing and discussed in this document.

Table 6 Zones for Oracle RAC Setup

Zone Name	Host (HBA)	Storage Visible from Zone
RAC1_Zone1	RAC1_HBA1	SPA0, SPA4, SPB2, and SPB6
RAC2_Zone1	RAC2_HBA1	
RAC3_Zone1	RAC3_HBA1	
RAC4_Zone1	RAC4_HBA1	
RAC5_Zone1	RAC5_HBA1	
RAC6_Zone1	RAC6_HBA1	
RAC7_Zone1	RAC7_HBA1	
RAC8_Zone1	RAC8_HBA1	
RAC1_Zone2	RAC1_HBA2	SPA2, SPA6, SPB0, and SPB4
RAC2_Zone2	RAC2_HBA2	
RAC3_Zone2	RAC3_HBA2	
RAC4_Zone2	RAC4_HBA2	
RAC5_Zone2	RAC5_HBA2	
RAC6_Zone2	RAC6_HBA2	
RAC7_Zone2	RAC7_HBA2	
RAC8_Zone2	RAC8_HBA2	

After you complete the zoning, you are ready to configure storage.

Setting up EMC CLARiiON Storage

This document provides a general overview of the storage configuration for the database layout. However, it does not supply details about host connectivity and logical unit number (LUN)—that is, RAID—configuration. For more information about EMC CLARiiON storage, refer to <http://powerlink.emc.com>.

Configuring Storage

To configure storage for the Cisco Unified Computing System data center solution, follow these steps:

Step 1 Ensure host connectivity.

If each host has the EMC Navisphere Agent® package installed, the agent automatically registers the HBA initiators.

Step 2 If the package is not installed, make sure that all initiators are registered properly to complete the host registration.

Step 3 Create the RAID groups.

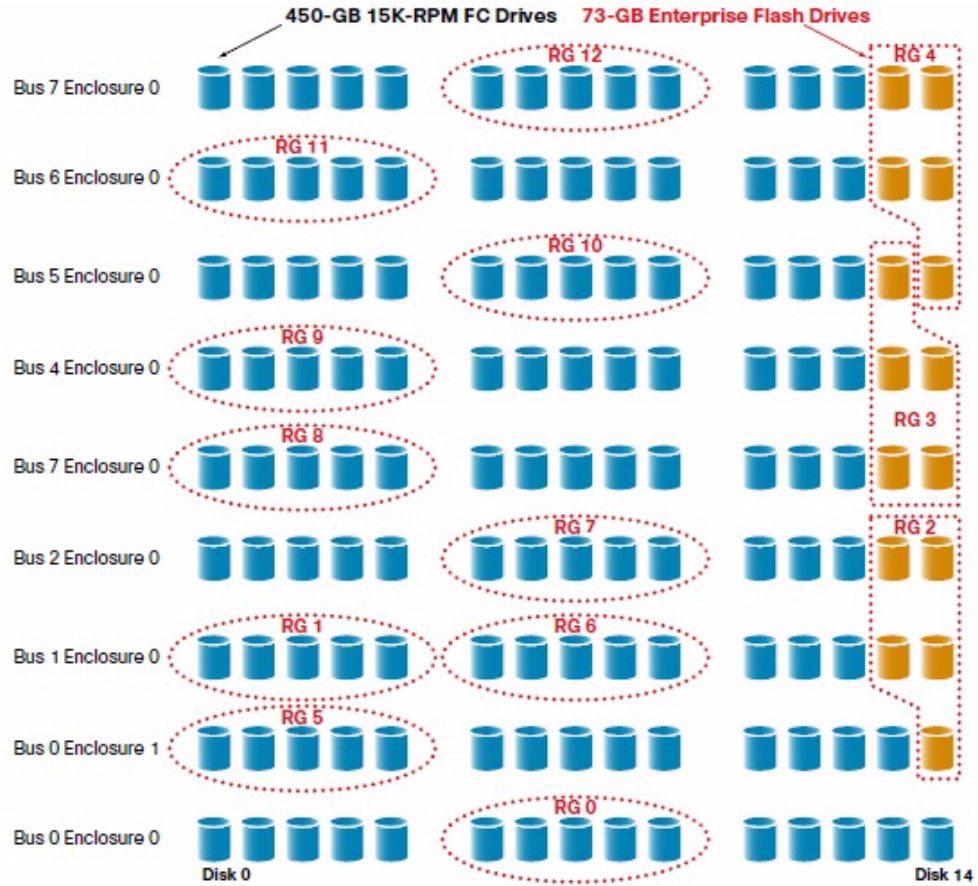
Testing for the Cisco Unified Computing System solution used:

- EMC CLARiiON CX4-960 with 105 Fibre Channel spindles

- 15 EFDs

Figure 7 illustrates the RAID groups created for database testing.

Figure 7 RAID Groups used in database testing



Step 4 Create the LUNs.



Note

It is extremely important that you choose an appropriate storage processor as the default owner so that the service processors are evenly balanced. The Cisco Unified Computing System data center solution creates one LUN per RAID group for Fibre Channel drives and four LUNs per RAID group for EFDs.

Table 7 provides the LUN configuration data.

Table 7 LUN Configuration Data

RAID Group and Type	LUN	Size	Purpose	Owner (Storage Processor)
RAID Group 0 (RAID-5 4+1)	LUNs 0 and 1	256 MB	Voting disks	SP-A
	LUN 2	256 MB	OCR disk	SP-B
RAID Group 1 (RAID-5 4+1)	LUN 3	256 MB	Voting disk	SP-B
	LUN 4	256 MB	OCR disk	SP-A
RAID Group 2 EFD drives (RAID-5 4+1)	LUNs 5 and 6	66 GB	Data disks for Oracle ASM	SP-A
	LUNs 7 and 8	66 GB	Data disks for Oracle ASM	SP-B
RAID Group 2 EFD drives (RAID-5 4+1)	LUNs 9 and 10	66 GB	Data disks for Oracle ASM	SP-A
	LUNs 11 and 12	66 GB	Data disks for Oracle ASM	SP-B
RAID Group 4 EFD drives (RAID-5 4+1)	LUNs 13 and 14	66 GB	Data disks for Oracle ASM	SP-A
	LUNs 15 and 16	66 GB	Data disks for Oracle ASM	SP-B
RAID Group 5 (RAID-5 4+1)	LUN 24	512 GB	Redo logs	SP-A
RAID Group 6 (RAID-5 4+1)	LUN 17	512 GB	Redo logs	SP-B
RAID Group 7 (RAID-5 4+1)	LUN 18	512 GB	Redo logs	SP-A
RAID Group 8 (RAID-5 4+1)	LUN 19	512 GB	Redo logs	SP-B
RAID Group 9 (RAID-5 4+1)	LUN 20	768 GB	Temp	SP-A
RAID Group 10 (RAID-5 4+1)	LUN 21	768 GB	Temp	SP-B
RAID Group 11 (RAID-5 4+1)	LUN 22	768 GB	Temp	SP-A
RAID Group 12 (RAID-5 4+1)	LUN 23	768 GB	Temp	SP-B

Step 5 Follow these additional recommendations for configuring storage and LUNs:

- a. Turn off the read and write caches for EFD-based LUNs. In most situations, it is better to turn off both the read and write caches on all the LUNs that reside on EFDs, for the following reasons:
 - The EFDs are extremely fast: When the read cache is enabled for the LUNs residing on them, the read cachelookup for each read request adds more overhead compared to Fibre Channel drives. This scenario occurs in an application profile that is not expected to get many read cache hits at any rate. It is generally much faster to directly read the block from the EFD.
 - In typical situations, the storage array is also shared by several other applications and the database. This situation occurs particularly when storage deploys mixed drives, which may also consist of slower SATA drives. The write cache may become fully saturated, placing the EFDs in a force-flush situation, which adds latency. Therefore, it is better in these situations to write the block directly to the EFDs than to the write cache of the storage system.
- b. Distribute database files for EFDs. Refer to [Table 8](#) for recommendations about distributing database files based on the type of workload.

Table 8 *Distribution of Data Files Based on Type of Workload*

EFD-Friendly Database Workloads	Expensive Workloads (Not as Cost Effective on EFD)
User Tablespace-Based Data Files Random Reads <ul style="list-style-type: none"> • B-tree leaf access • Row ID lookup in table • Access to out-of-line LOB • Access to overflowed row • Index scan of unclustered table • Compression: Increases I/O intensity (I/O operations per second [IOPS] per GB) Serial Reads Random Writes <ul style="list-style-type: none"> • Row update by PK • Index maintenance • Reduction of checkpoint interval Temporary Tablespace Files: Sort Areas and Intermediate Tables <ul style="list-style-type: none"> • Sequentially read and written, but I/O is performed in 1-MB units; not enough to amortize seeks • Lower latency: Get in and get out 	Redo Log Files <ul style="list-style-type: none"> • Sequential I/O • Read and write and commit latency already handled by cache in storage controller Undo Tablespace <ul style="list-style-type: none"> • Sequential writes and random reads by Oracle Flashback. • Generally, reads are for recently written data that is likely to be in the buffer cache Large Table Scans (If Single Stream)

The configuration described here employs most of EMC's best practices and recommendations for LUN distribution in the database. It also adopts the layout for a mixed storage environment consisting of Fibre Channel disks and EFDs.

For more information about Oracle database best practices for flash-drive-based EMC CLARiiON storage, refer to the document "Leveraging EMC CLARiiON CX4 with Enterprise Flash Drives for Oracle Database Deployments" go to:

<http://www.emc.com/collateral/hardware/white-papers/h5967-leveraging-clariion-cx4-oracle-deploy-wp.pdf>.

Applying Patches, Environment, and OS Settings

After completing the configuration of the Cisco Unified Computing System, the SAN, and storage, you can install the OS.

To test the Cisco Unified Computing System data center solution, 64-bit Oracle Enterprise Linux (OEL) 5.3, Update 3, was used as the OS.

Installing the OS and Setting Up the Environment

To install the OS and enable the environment settings, follow these steps:

-
- Step 1** Install 64-bit OEL 5.3, Update 3, on all eight nodes.
- Step 2** Update the Intel ixgbe driver by applying the latest errata kernel.
- Because of a bug in OEL and Red Hat Enterprise Linux (RHEL) 5.3, systems with 16 or more logical processors that use network devices requiring the ixgbe driver have intermittent network connectivity or can experience a kernel panic. To help ensure network stability, follow the recommendations in the article at <http://kbase.redhat.com/faq/docs/DOC-16041>.
- Step 3** Install the Oracle Validated RPM package.

Use of this RPM package can simplify preparation of Linux for Oracle Clusterware and RAC installation. The RPM downloads (or updates) all necessary RPM packages on the system, resolves dependencies, and creates Oracle users and groups. It also sets all appropriate OS and kernel specifications, depending on the system configuration.

The appendix lists kernel settings if you decide to set them manually.

Step 4 Install the Oracle Automatic Storage Management (ASM) RPM package.

The Oracle ASM Library (ASMLib) enables ASM-based I/O to Linux disks without the limitations of the standard UNIX I/O API. For information about downloading and installing Oracle ASMLib based on your kernel version, go to

<http://www.oracle.com/technology/tech/linux/asmlib/install.html>.

Step 5 Configure Oracle ASM; then create and label the disks that the ASM is to manage.

The following disks were created for the test environment:

- 12 data disks managed by Oracle ASM
- 4 temporary disks
- 4 log disks (archive, logs, etc.)

Step 6 Install the EMC Navisphere RPM package.

It is highly recommended that you install this RPM package because the package helps ensure automatic host registration with EMC CLARiiON storage.

Step 7 Configure the multipathing software.

The test environment uses the Linux Device Mapper utility for multipathing and device-naming persistence.

For the test environment, a total of 24 LUNs were created and divided into the following components:

- 2 Oracle Cluster Registry (OCR) disks
- 3 voting disks
- 12 data disks managed by Oracle ASM
- 4 temporary disks
- 4 log disks (for archives, logs, etc.)

For more information about multipathing software and device setup, refer to Oracle MetaLink document 564580.1 (Oracle service contract required).

Step 8 Create disk partitions for Oracle Clusterware disks (OCR and voting):

```
root@rac1 downloads]# fdisk /dev/dm-1
n                new partition
p                primary partition
1                partition 1
(CR)             start from beginning of device or LUN
(CR)             use all the available sectors
w                commit changes
```

Step 9 Create a partition for ASM-managed data disks at the offset of 1 MB (or 2048 sectors).

This step is useful because Oracle ASM performs I/O operations at 1-MB boundaries. Setting this offset aligns host I/O operations with the back-end storage I/O operations.

Use the following setup for ASM-managed data disks:

```
[root@rac1 downloads]# fdisk /dev/dm-8
n                new partition
p                primary partition
1                partition 1
(CR)             start from beginning of device or LUN
(CR)             use all the available sectorsx go into EXPERT mode
b                adjust partition header data begin offset
1                for partition 1
2048             to sector 2048 from beginning of LUN, or 1MB
w                commit changes
```

- Step 10** Configure the private and public NICs with the appropriate IP addresses.
- Step 11** Identify the virtual IP addresses for each node and update the /etc/hosts file with all the details (private, public, and virtual IP).
- Step 12** 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 Clusterware and the database.

Installing Oracle Clusterware and the Database

For more information about the Oracle RAC installation, refer to the Oracle installation documentation.

To install Oracle, follow these steps:

-
- Step 1** Download the Oracle Database 10g Release 2 (10.2.0.1.0) software.
- Step 2** Install Oracle Database 10g Release 2 Clusterware.
- Step 3** Install Oracle Database 10g Release 2 Database “Software Only”; do not create the database.
- Step 4** Download the Oracle Database 10g Release 2 (10.2.0.4) Patch Set 3 bundle and install it.

During the installation, you may encounter several known issues, described in the following Oracle MetaLink notes:

- **Note 414163.1:** 10gR2 RAC install issues on Oracle EL5 or RHEL5 or SLES10 (VIPCA failures)
- **Note 443617.1:** (32-bit) libXp-1.0.0-8.1.el5.i386.rpm is required to avoid OUI error during install

Now you are ready to create the database and the workload setup.

Configuring the Oracle SwingBench Workload

Two databases were created for the configuration discussed here: for Oracle SwingBench OLTP and DSS workloads.

To set up the Oracle SwingBench workloads, follow these steps:

Step 1 Create two databases using Oracle Database Creation Assistant (DBCA):

- OLTP (Order Entry) workload
- DSS (Sales History) workload

Step 2 Populate the databases.

Both databases were populated with the data shown in the next sections: "OLTP (Order Entry) Database" and "DSS (Sales History) Database."

OLTP (Order Entry) Database

The OLTP database was populated with the following data:

```
[oracle@rac1 ~]$ sqlplus soe/soe
```

```
SQL*Plus: Release 10.2.0.4.0 - Production on Sat Sep 19 17:34:46 2009
Copyright (c) 1982, 2007, Oracle. All Rights Reserved.
```

```
Connected to:
```

```
Oracle Database 10g Enterprise Edition Release 10.2.0.4.0 - 64bit
Production
```

```
With the Partitioning, Real Application Clusters, OLAP, Data Mining and
Real Application Testing options
```

```
SQL> select table_name, num_rows from user_tables;
```

```
TABLE_NAME NUM_ROWS
```

```
-----
```

```
WAREHOUSES 264
```

```
PRODUCT_INFORMATION 288
```

```
PRODUCT_DESCRIPTIONS 288
```

```
LOGON 50033
```

```
INVENTORIES 77815
```

```
ORDERS 25459388
```

```
ORDER_ITEMS 89147570
```

```
CUSTOMERS 25305070
```

```
8 rows selected.
```

DSS (Sales History) Database

The DSS database was populated with the following data:

```
[oracle@rac1 ~]$ sqlplus sh/sh
```

```
SQL*Plus: Release 10.2.0.4.0 - Production on Sat Sep 19 17:43:42 2009
Copyright (c) 1982, 2007, Oracle. All Rights Reserved.
```

```
Connected to:
```

```
Oracle Database 10g Enterprise Edition Release 10.2.0.4.0 - 64bit
Production
```

```
With the Partitioning, Real Application Clusters, OLAP, Data Mining
```



```

and Real Application Testing options
SQL> select table_name, num_rows from user_tables;
TABLE_NAME NUM_ROWS
-----
COUNTRIES 23
COSTS 132528
SALES 2034392904
SUPPLEMENTARY_DEMOGRAPHICS 204426381
CUSTOMERS 204400080
TIMES 1826
CHANNELS 5
PRODUCTS 72
PROMOTIONS 503

9 rows selected.

```

For more information about creating and populating databases for OLTP (Order Entry) and DSS (Sales History) workloads, refer to the Oracle SwingBench documentation at <http://dominicgiles.com/swingbench.html>.

Testing Workload Performance

To evaluate workload performance, the cluster was stressed for 24 hours with a sustained load. During the 24-hour run of both the OLTP (Order Entry) and the DSS (Sales History) workloads, no crashes or degradation of performance was observed.

The following workload performance metrics were detected and recorded:

- Very consistent CPU utilization: around 40 percent on all eight nodes
- No saturation levels of any subsystems (CPU, disk, I/O, or networking)
- Sustained FCoE-based I/O ranging between 1.8 and 2.0 GB per second, which could be further divided into 1.4 GB per second of Fibre Channel I/O and approximately 450 MB per second of interconnect communication
- No occurrence of I/O bottlenecks or wait times
- Excellent I/O service times for storage

The consistent workload performance can be attributed to:

- The simplified, excellent architectural design of the Cisco Unified Computing System based on a 10-Gbps unified fabric
- The pairing of the Cisco Unified Computing System with EMC CLARiiON storage with high-performance EFDs



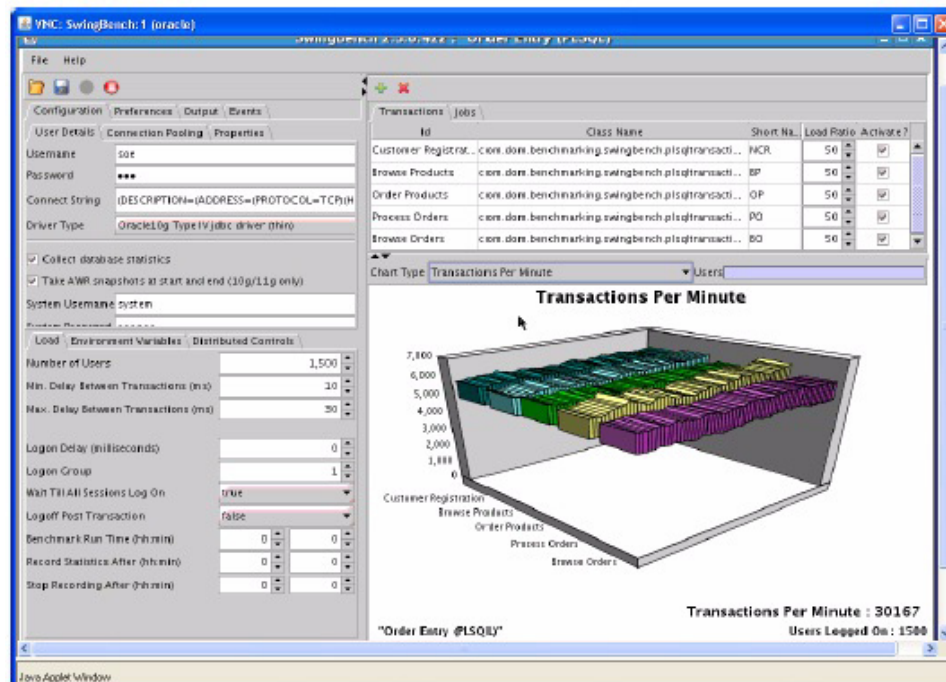
Note

This is a testing, not a performance benchmarking, exercise. The numbers presented here should not be used for comparison purposes. The intent here is to look at the Cisco Unified Computing System supporting a sustained load over a long time period. Note that no tuning was performed, and the lack of resource saturation indicates that significant headroom is available to support greater performance than that shown here.

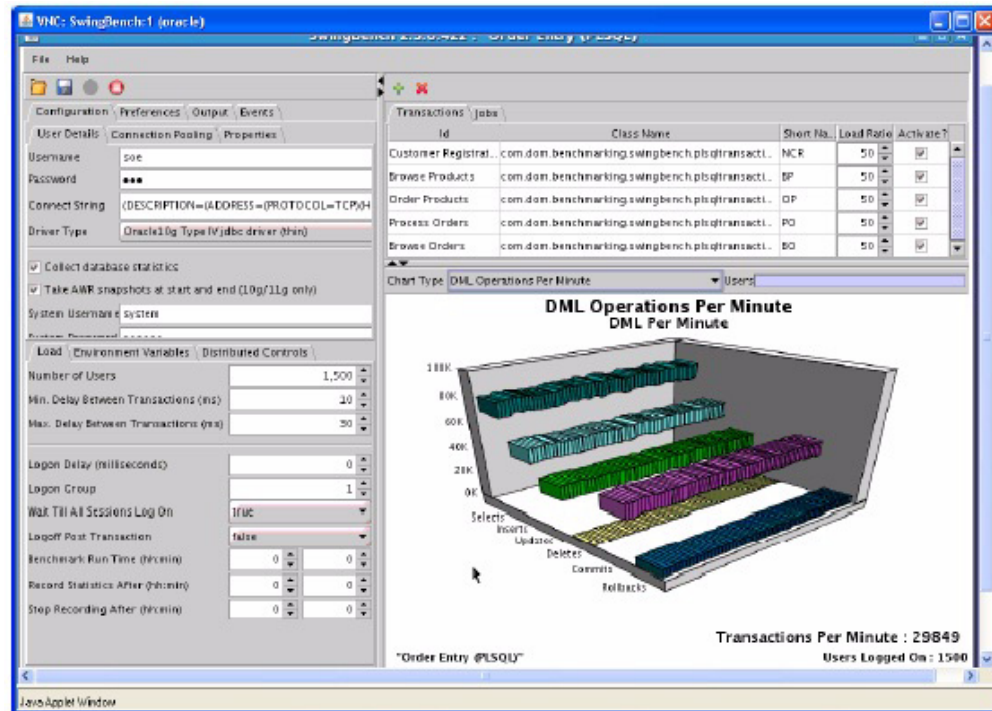
OLTP Workload

Figure 8 shows the Order Entry workload running 1500 users in the eighth hour of a 24-hour run.

Figure 8 *Order Entry Workload*



A typical OLTP Oracle application has some write activity because of Data Manipulation Language (DML) operations such as updates and inserts. Figure 9 shows the DML operations per minute for the OLTP workload.

Figure 9 DML Operations Breakdown for OLTP Workload

DSS Workload

Unlike the OLTP workload, the DSS workload is set to run from the command line. DSS workloads are generally very sequential and read intensive. For DSS workloads, it is common practice to set the parallel queries and the degree of parallelism on heavily read tables. This practice was followed in the test environment and achieved excellent performance, as indicated in the Tablespace and File IO Stats information from the Oracle Automated Workload Repository (AWR) report (90-minute duration) shown in [Table 9](#) and [Table 10](#).

Table 9 Oracle AWR Report Tablespace IO Stats Information

Tablespace	Reads	Av Reads/s	Av Rd(ms)	Av Blks/Rd	Writes	Av Writes/s	Buffer Waits	Av Buf Wt(ms)
SH	847,901	134	0.03	122.44	0	0	28	14.29
SYS_AUX	695	0	36.23	2.19	2,340	0	33	954.85

As [Table 9](#) shows, 134 read operations occur per second. Each read fetches about 122 data blocks, and each data block is 8 KB in size. Consequently, each read operation fetches about 1 MB (122.44 x 8 KB). In other words, this particular instance performs about 134 MB of read operations on SH tablespace. Similar behavior was observed across all eight nodes. The result is about 130 MB x 8 instances at 1 GB per second of read operations for the DSS workload.

Table 10 Oracle AWR Report File IO Stats Information

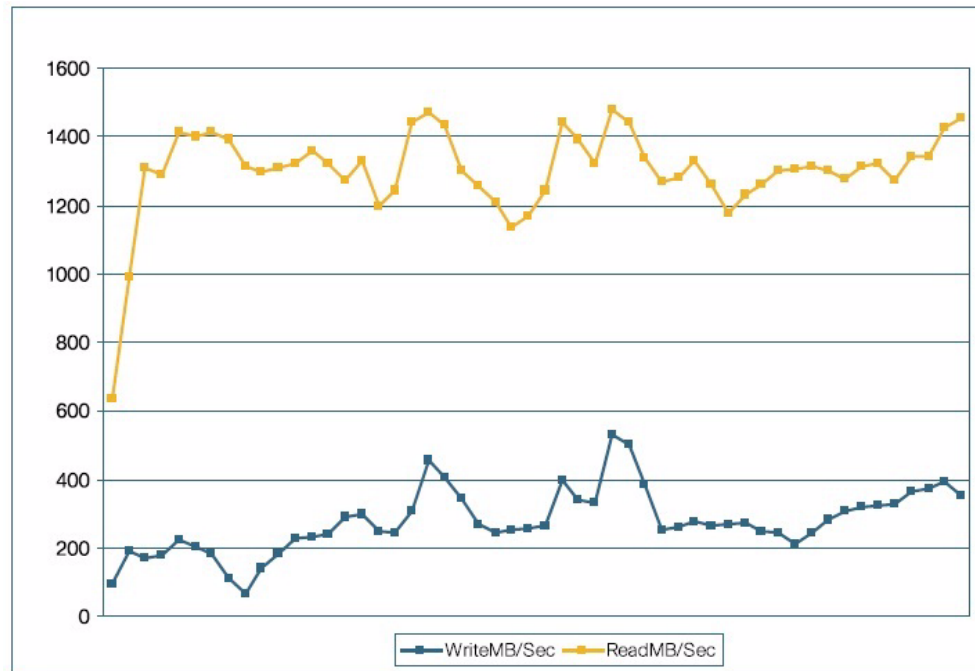
Tablespace	Filename	Reads	Av Reads/s	Av Rd(ms)	Av Blks/Rd	Writes	Av Writes/s	Buffer Waits	Av Buf Wt(ms)
SH	+USER_DG/sh1	122,795	19	0.00	122.32	0	0	0	0.00
SH	+USER_DG/sh2	122,239	19	0.00	122.66	0	0	0	0.00
SH	+USER_DG/sh3	124,802	20	0.08	122.24	0	0	16	25.00
SH	+USER_DG/sh4	120,709	19	0.01	122.48	0	0	2	0.00
SH	+USER_DG/sh5	117,037	18	0.00	122.24	0	0	6	0.00
SH	+USER_DG/sh6	116,416	18	0.10	122.84	0	0	0	0.00
SH	+USER_DG/sh7	123,903	20	0.00	122.34	0	0	4	0.00

The File IO Stats information indicates that all ASM-managed files have evenly spread read operations (18 to 20 operations per second). However, the benefit of the EFD drives is clearly reflected in the Av Rd(ms) column.

Generally speaking, rotating Fibre Channel drives perform well in a single stream of queries. However, addition of multiple concurrent streams (or parallel queries) causes additional seek and rotational latencies, thereby reducing the overall per-disk bandwidth. In contrast, the absence of any moving parts in EFDs enables sustained bandwidth regardless of the number of concurrent queries running on the drive.

FCoE Traffic and the Read-to-Write Ratio

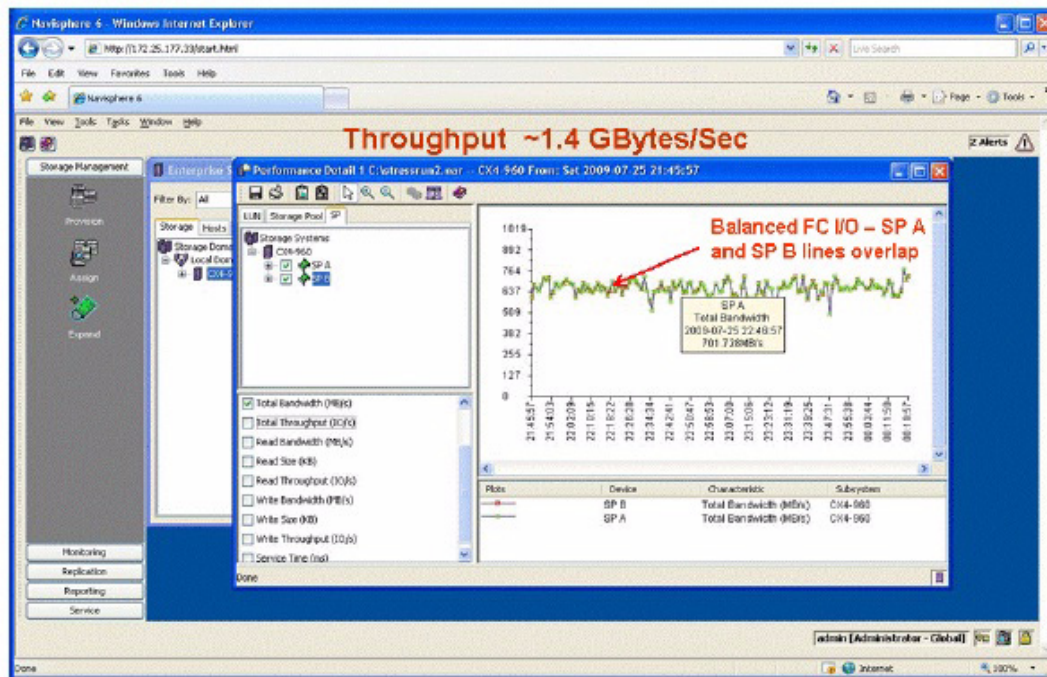
Figure 10 provides a sample from a 24-hour stress run using the workload. It shows the combined FCoE read and write traffic observed at the fabric interconnects. This I/O is the combination of Oracle RAC interconnect traffic (approximately 450 MB) and Fibre Channel I/O (1.4 GB).

Figure 10 *FCoE Traffic Observed at Fabric Interconnects*

Fibre Channel Throughput

A sample from a 24-hour stress run ([Figure 11](#)) shows the Fibre Channel I/O serviced by the storage.

Figure 11 *Fibre Channel Traffic Serviced by Storage*



Availability after Hardware Failures

Previous sections described Cisco Unified Computing System installation, configuration, and performance. This section examines the Cisco Unified Computing System's nearly instant failover capabilities to show how they can improve overall availability after unexpected, but common, hardware failures attributed to ports and cables.

Figure 12 shows some of the failure scenarios (indicated by numbers) that were tested under the stress conditions described in the preceding section, "Testing Workload Performance."

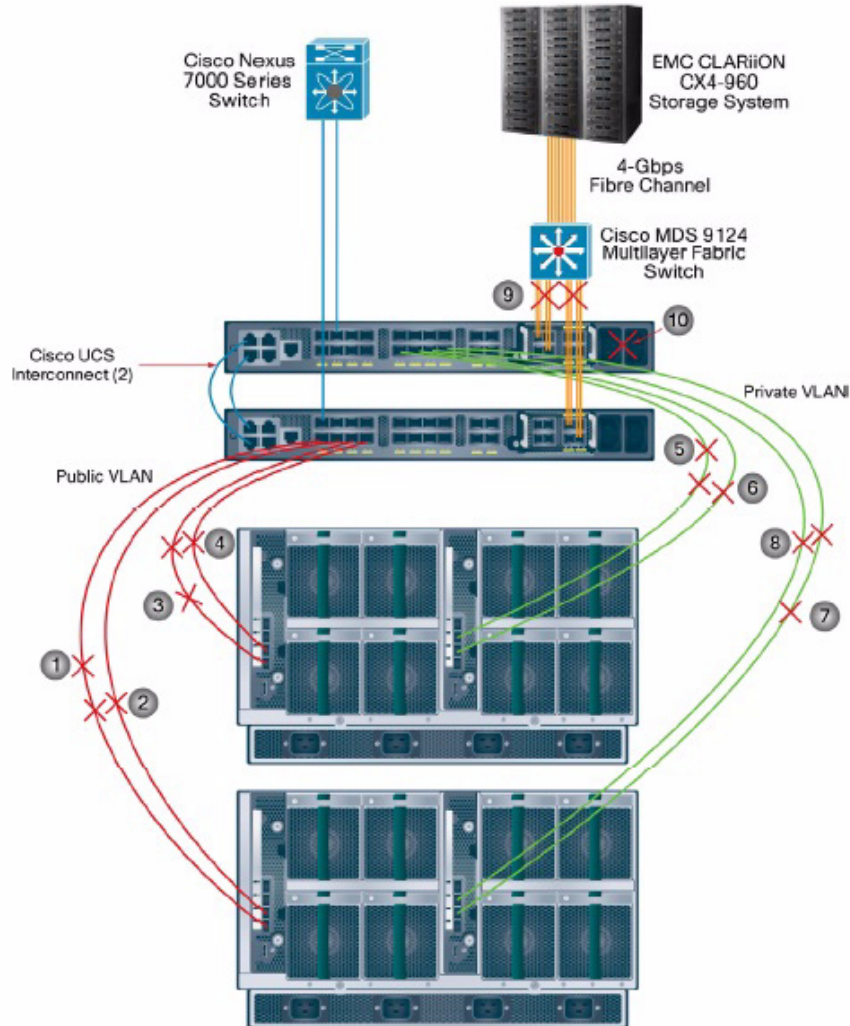
Figure 12 **Sample Failure Scenarios**

Table 11 summarizes the failure scenarios (each indicated by a number in Figure 10) and describes how the Cisco Unified Computing System architecture sustains unexpected failures related to ports, links, and the fabric interconnect (a rare occurrence).

Table 11 **Failure Scenarios and Cisco Unified Computing System Response**

Failure Scenario (Figure 10)	Description	Response
1 or 3	Single-link failure (public interface)	Represents port failure; nodes continue to work
2 or 4	All-links failure (public interface)	Nodes continue to work through other I/O module
5 or 7	Single-link failure (private interconnect)	Represents port failure; nodes continue to work
6 or 8	All-links failure (private interconnect)	Nodes continue to work through other I/O module
9	Storage path failure	Single I/O path loss have no effect; if all I/O paths fail, nodes reboot
2 and 6 or 4 and 8	Failure of both I/O modules on a single chassis	All nodes in failed chassis reboot; nodes in other chassis continue to work
10	Fabric interconnect failure	All nodes continue to work through other fabric interconnect

Conclusion

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

For the Cisco Unified Computing System, 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 Unified Computing System connected to the industry proven and scalable CLARiiON 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 Unified Computing System 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 workloads, which generated a sustained load on the eight-node Oracle RAC configuration for a period of 72 hours. This type of load far exceeds the demands of typical database deployments.

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

- The quad-core Intel Xeon 5500 series processors barely reached 50 percent of their capacity, leaving substantial headroom for additional load.
- The average 10 Gigabit Ethernet port utilization at the fabric interconnect was about 40 percent.
- The I/O demands generated by the load were supported very efficiently by the capabilities of the minimally configured EMC CLARiiON storage array. The array featured a mix of Fibre Channel and EFDs.

In summary, the Cisco Unified Computing System is a new 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 Unified Computing System 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

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

Appendix

Cisco Unified Computing System Kernel Settings (/etc/sysctl.conf)

This appendix provides the parameters for the Cisco Unified Computing System with 24 GB of RAM.



Note

It is highly recommended that you use the Oracle Validated RPM to derive kernel settings that are most suitable for your system.

```
# 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 = 4398046511104

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

# User added
kernel.shmmni = 4096
#kernel.sem = 250 32000 100 142
kernel.sem = 250 32000 100 256
fs.file-max = 6553600
```

```

net.ipv4.ip_local_port_range = 1024 65000
net.core.wmem_default = 262144
net.core.wmem_max = 262144
kernel.panic_on_oops = 1
kernel.panic = 30
vm.nr_hugepages=7200
#vm.nr_hugepages=0

kernel.msgmni = 2878
net.core.rmem_default = 262144
net.core.rmem_max=2097152
net.core.wmem_default = 262144
net.core.wmem_max = 262144
fs.aio-max-nr = 3145728

```

Multipath Settings (/etc/multipath.conf)

The setup described in this document used the Linux Device Mapper.

The entries from the multipath.conf file are as follows:

```

# This is a basic configuration file with some examples, for device mapper
# multipath.
# For a complete list of the default configuration values, see
# /usr/share/doc/device-mapper-multipath-0.4.7/multipath.conf.defaults
# For a list of configuration options with descriptions, see
# /usr/share/doc/device-mapper-multipath-0.4.7/multipath.conf.annotated

# Blacklist all devices by default. Remove this to enable multipathing
# on the default devices.
#blacklist {
#    devnode "*"
#}

## By default, devices with vendor = "IBM" and product = "S/390.*" are
## blacklisted. To enable multipathing on these devices, uncomment the
## following lines.
#blacklist_exceptions {
#    device {
#        vendor "IBM"
#        product "S/390.*"
#    }
#}

## Use user friendly names, instead of using WWIDs as names.
defaults {
    user_friendly_names yes
}

multipaths {
    multipath {
        wwid 360060160aef72200b84551be8672de11
        alias ocr1
    }

    multipath {
        wwid 360060160aef72200229b6dc98672de11
        alias ocr2
    }

    multipath {
        wwid 360060160aef72200c804f1218c72de11
        alias voting1
    }
}

```

```

multipath {
    wwid 360060160aef72200c904f1218c72de11
    alias voting2
}

multipath {
    wwid 360060160aef72200248d682f8c72de11
    alias voting3
}

multipath {
    wwid 360060160aef7220048576cb0cf70de11
    alias usr_disk1
}

multipath {
    wwid 360060160aef7220049576cb0cf70de11
    alias usr_disk2
}

multipath {
    wwid 360060160aef72200449fb70cd070de11
    alias usr_disk3
}

multipath {
    wwid 360060160aef72200459fb70cd070de11
    alias usr_disk4
}

multipath {
    wwid 360060160aef72200de51f950d070de11
    alias usr_disk5
}

multipath {
    wwid 360060160aef72200df51f950d070de11
    alias usr_disk6
}

multipath {
    wwid 360060160aef722004a576cb0cf70de11
    alias usr_disk7
}

multipath {
    wwid 360060160aef722004b576cb0cf70de11
    alias usr_disk8
}

multipath {
    wwid 360060160aef7220046c52c1ad070de11
    alias usr_disk9
}

multipath {
    wwid 360060160aef7220047c52c1ad070de11
    alias usr_disk10
}

multipath {
    wwid 360060160aef72200f857cc5dd070de11
    alias usr_disk11
}

multipath {
    wwid 360060160aef72200f957cc5dd070de11
    alias usr_disk12
}

multipath {
    wwid 360060160aef722008e27404bd770de11
    alias redo_disk1
}

multipath {
    wwid 360060160aef7220010fe640fd770de11
    alias redo_disk2
}

```

```

    }

    multipath {
        wwid 360060160aef722002cde712ed770de11
        alias redo_disk3
    }

    multipath {
        wwid 360060160aef7220008351e59d770de11
        alias redo_disk4
    }

    multipath {
        wwid 360060160aef72200ca4cd6c6d770de11
        alias temp_disk1
    }

    multipath {
        wwid 360060160aef7220090066da1d770de11
        alias temp_disk2
    }

    multipath {
        wwid 360060160aef72200066886bad770de11
        alias temp_disk3
    }

    multipath {
        wwid 360060160aef72200660708b2d770de11
        alias temp_disk4
    }
}
devices {
    device {
        vendor "DGC "
        product "*"
#        path_grouping_policy group_by_prio
        path_grouping_policy multibus
        getuid_callout "/sbin/scsi_id -g -u -s /block/%n"
        prio_callout "/sbin/mpath_prio_emc /dev/%n"
        path_checker emc_clariion
        path_selector "round-robin 0"
        features "1 queue_if_no_path"
        no_path_retry 300
        hardware_handler "1 emc"
        failback immediate
    }
}

```

Cisco Validated Design

The Cisco Validated Design Program consists of systems and solutions designed, tested, and documented to facilitate faster, more reliable, and more predictable customer deployments. For more information visit www.cisco.com/go/validateddesigns.

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

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

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

All other trademarks mentioned in this document or Website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0807R)

