Data Warehouse Scalability with Cisco Unified Computing System and Oracle Real Application Cluster with EMC CLARiiON® Storage

What You Will Learn

Designing and scaling an Oracle data warehouse can now be accomplished efficiently, quickly, and in close alignment with evolving business needs through the features of the Cisco Unified Computing System™. These features include the capability to provision servers rapidly with Cisco Unified Computing System service profiles and to meet increasing resource demands with faster processors and a low-latency unified fabric. This document describes a data warehouse scalability study using Oracle Real Application Cluster (RAC), Cisco Unified Computing System, and EMC CLARiiON storage. Data warehouse solutions today face constant challenges related to performance and scalability while they strive to remain cost effective. This study demonstrates the near-linear scalability of the Cisco Unified Computing System as additional server and storage resources are added to an existing cluster. This highly cost-effective and scalable solution is based on:

- The ability to add computing resources incrementally, quickly, and as needed with the Cisco Unified Computing System; using a unified fabric, wire-once model, an additional blade server could be provisioned within minutes using service profiles that are managed by Cisco® UCS Manager
- The modular design of EMC CLARiiON storage, with capacity that scales easily while maintaining I/O service levels for demanding data warehouse workloads

This configuration enables users to scale compute and storage resources horizontally and vertically.

Challenge: Expensive Data Warehouse Scalability and Time-Consuming Server Provisioning

The data warehouse is evolving in many of today's business environments from a passisve reporting database servicing a few users to a mission-critical, real-time repository that must support many users entering random and complex queries that run for hours. Here are the hurdles that are frequently encountered in today's data warehouse deployments:

- Multiple users run random and complex queries in parallel
- Continuous data growth is unavoidable.
- High availability is mandatory for mission-critical applications.
- Growth is difficult to anticipate, with data warehouses frequently growing unexpectedly as a result of their own deployment success.
- Hardware components need to be fully utilized to meet business objectives.
- Achieving a balance between interdependent components is crucial to scaling.

These challenges can be grouped into three categories:

- **Hardware overprovisioning:** To help ensure that data warehouses can accommodate required resource demands now and in the future, organizations commonly deploy more infrastructure than needed, to handle the ever-increasing volumes of data, storage, and transactions. The result is often an oversized hardware configuration and significantly higher costs.
- **Scalability:** The capability to easily, efficiently, quickly, and cost-effectively scale a data warehouse is crucial for businesses that depend on database applications. Typically, scalability is achieved by adding units of computing power and achieving a commensurate improvement in capacity by adding disks. With traditional

© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

symmetric multiprocessors, however, as the infrastructure scales, the linearity of performance and capacity per incremental resource declines. This decline is the result of increased overhead for communication and data transfer between the different data warehouse elements plus increased latency (for example, multiple, simultaneous queries to a database result in a lack of access for some users for periods of time).

- **On-demand resource provisioning:** On-demand provisioning is a critical factor in overcoming the challenge of overprovisioning. In the traditional approach, as each server is added, it must meet hardware compatibility requirements and be individually configured: a cumbersome and time-consuming process. In addition, the compatibility of operating system, firmware, network, Fibre Channel host bus adapters (HBAs), BIOS, and other elements must be maintained.

The Cisco Unified Computing System meets these challenges in a reliable and efficient manner through a simple infrastructure-block approach. This solution is made possible with industry-proven technologies such as Oracle RAC and EMC CLARiiON storage.

Solution: Data Warehouse Using Cisco Unified Computing System and 10-Gbps Unified Fabric with Modular EMC CLARiiON Storage

A data warehouse solution using the Cisco Unified Computing System and 10-Gbps unified fabric with modular EMC CLARiiON storage consists of several main components.

Cisco Unified Computing System

The Cisco Unified Computing System represents a radical simplification of traditional architectures that dramatically reduces the number of devices that must be purchased, cabled, configured, powered, cooled, and secured. As a result, the Cisco Unified Computing System solution delivers end-to-end optimization for x86 environments while retaining compatibility with traditional OS and application stacks. The Cisco Unified Computing System unites network, compute, and virtualization resources into a cohesive system that simplifies setup, improves business metrics, and enables just-in-time resource provisioning. The system integrates a low-latency 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.

Cisco UCS M81KR Virtual Interface Card

A Cisco innovation, the Cisco UCS M81KR Virtual Interface Card (VIC) is a virtualization-optimized Ethernet and Fibre Channel over Ethernet (FCoE) mezzanine adapter designed for use with Cisco UCS B-Series Blade Servers. The VIC 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 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 Fibre Channel adapter that delivers cutting-edge storage I/O operations per second (IOPS) and throughput performance.

Oracle Real Application Cluster

Oracle RAC provides crucial support for transparent deployment of a single database across a cluster of servers providing an active-active configuration that is fault tolerant of hardware failures and planned outages. If a node in the cluster fails, the Oracle database continues running on the remaining nodes. If a need for additional capacity arises, new nodes can be added to the exisiting cluster.

EMC CLARiiON Storage

EMC CLARiiON storage delivers industry-leading innovation in a midrange storage solution. The unique combination of flexible, scalable hardware design and advanced software capabilities enable CLARiiON storage systems to meet the diverse and growing needs of today’s enterprises with unique highly available data integrity features.

© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Goal and Approach for Efficient Implementation

The goal of the data warehouse solution is a modular reference configuration that provides a balance of processing power, throughput, and capability to meet growth demands. This reference configuration is also a foundation for balanced, modular building blocks that can easily scale through the addition of compute nodes, fabric, and spindles as the needs of the data warehouse grow. The components in the base configuration must work efficiently together to achieve the best performance. If any of the components in this basic block are not balanced, wait events are introduced that affect overall performance. The Oracle data warehouse kit has well-established guidelines that allow the solution architect to define balanced processing and I/O subsystem components.

Here is the overall approach for establishing a baseline block to enable a scalable and modular solution:

- Baseline a single system block performance running an Oracle RAC instance on EMC CLARiiON CX4-960 storage following EMC best practice guidelines for configuring the CX4-960 for typical Oracle data warehouse deployments.
- Establish a single blade service profile.
- Deploy a Oracle RAC instance on a single blade.
- Validate compute and I/O resource balancing using tools such as Oracle ORION.
- Validate ease of scaling up by deploying multiple blades with matching storage spindles in modular increments.
- Replicate and clone the blade service profile for near-instant compute provisioning.
- Deploy additional Oracle RAC instances.
- Match I/O scaling through fabric expansion and Oracle ORION.
- Test Oracle partitioning and parallel query optimization to scale across multiple Oracle RAC instances.
- Confirm query performance scalability for deployment of additional blades and storage.

Test Workload and Methodology

The data warehouse workload used for this test was based on a Transaction Processing Performance Council Decision Support (TPC-DS) proposed specification and developed by the Oracle data warehouse engineering team to drive real-world point-of-sales data tracking. The toolkit provides queries that simulate real-world data warehouse workloads and range from simple select operations to complex query operations requiring nested joins along with sorting, summing, and aggregation.

Using the toolkit, a data warehouse was created consisting of 11 tables, the largest of which contained about 25 billion rows in the main fact table. The two main fact tables were range-partitioned by time and subpartitioned by hash on join columns. No indexes were created on the tables, so a typical execution plan results in full table or partition scans and partition hash joins. This type of execution plan is a fair representation of ad-hoc, complex queries and queries that have not been explicitly tuned. The toolkit provides the driver for a workload that simulates a scenario in which a user issues all the queries at random as a job sequence.

The first step in the toolkit process is to run a single user against the system block. The step determines the service time needed to complete all the queries by the single user. This step is typically known as the power run. The next step is to establish concurrency, where the number of users is incremented until the service time to complete all the queries exceeds 2.5 times the time of the power run. After a single-block baseline with an appropriate level of concurrency is established, the solution can be scaled by adding blocks and expanding the Oracle RAC cluster.

Solution Design

One of the main features that enables faster server provisioning is the capability to define a service profile template that can be used to provision servers quickly and consistently.

© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Cisco Unified Computing System Service Profiles

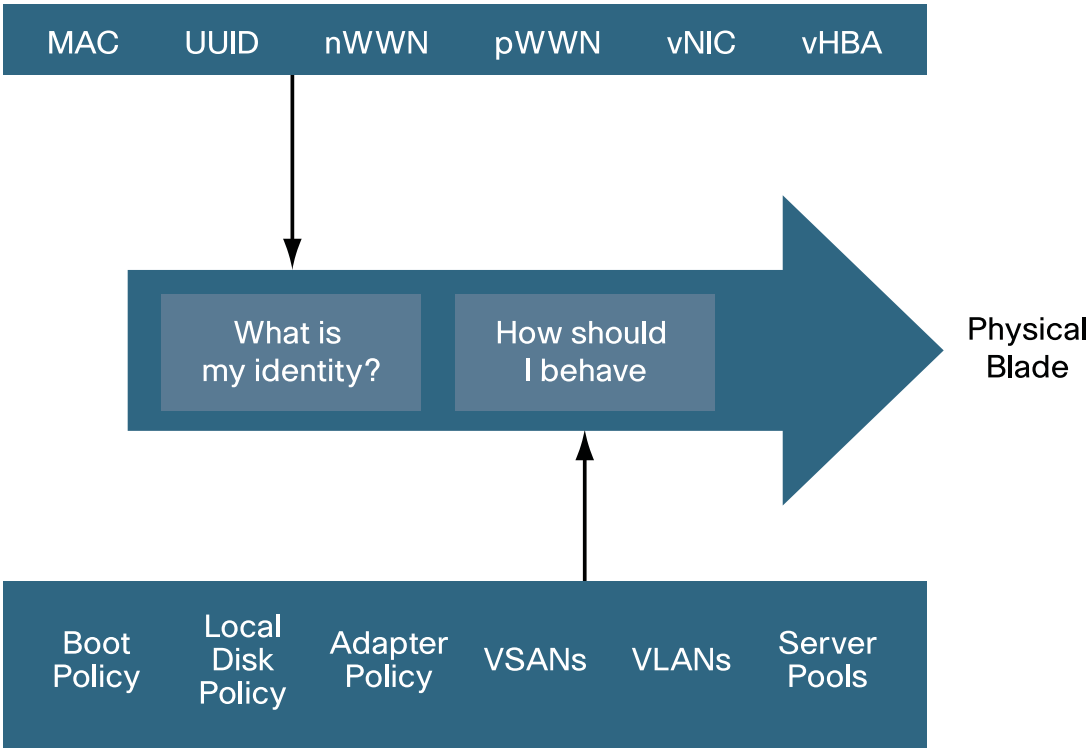
Service profiles are the central concept of the Cisco Unified Computing System. Each service profile serves a specific purpose, helping ensure that the associated server hardware has the configuration required to support the applications it will host. The service profile maintains configuration information about the server hardware, interfaces, fabric connectivity, and server and network identities. This information is stored in a format that administrators can manage through Cisco UCS Manager. All service profiles are centrally managed and stored in a database on the fabric interconnect.

In summary, a service profile:

- Is a self-contained definition of desired server and connectivity configuration and identity
- Applies to compute resources through direct association or through a blade pool
- Can be migrated with no specific local dependencies
- Is easily created or modified using templates for future deployments

Figure 1 provides a conceptual overview of the service profile.

Figure 1. Service Profile Overview



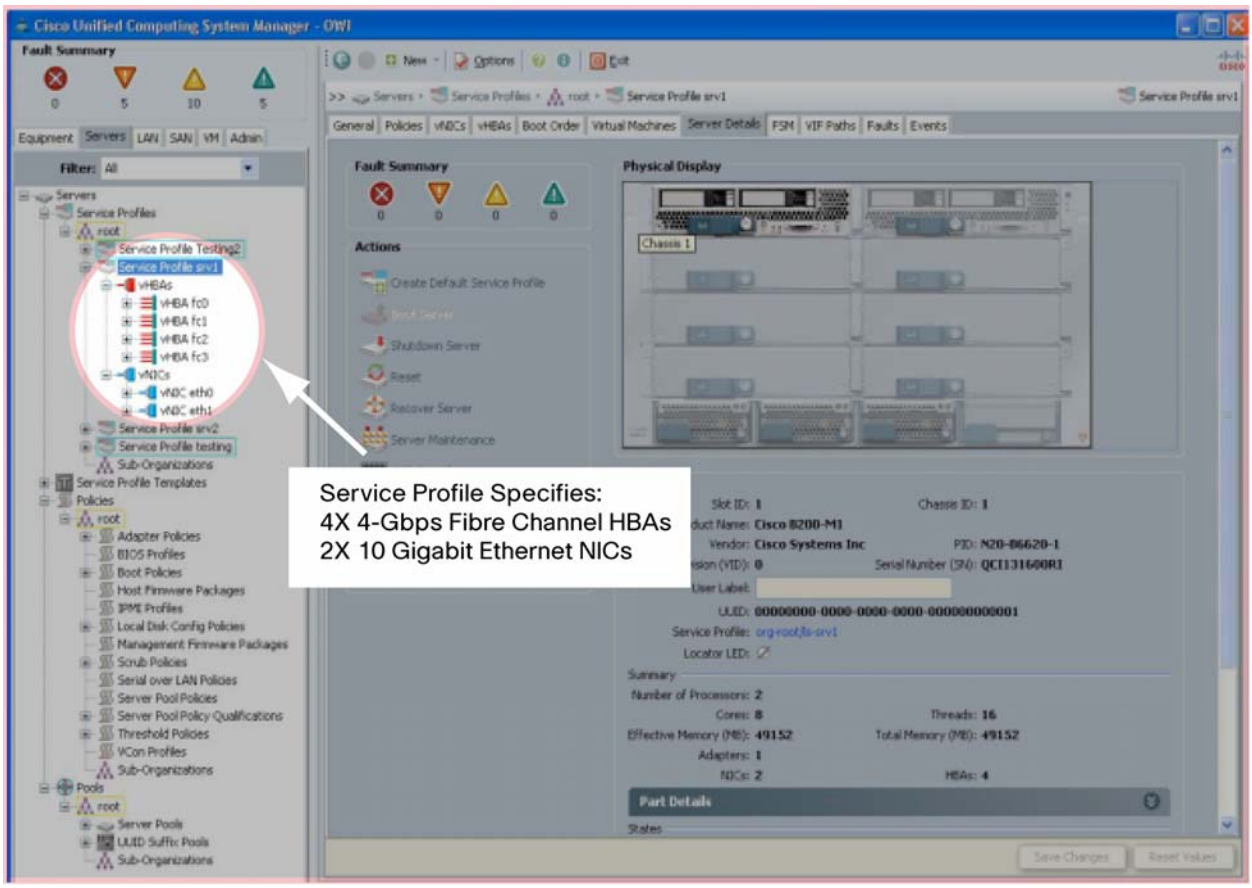
With the service profile, the Cisco Unified Computing System enables a highly dynamic and stateless environment that can be adapted to meet rapidly changing needs for data warehouse deployments. The Cisco Unified Computing System architecture enhances profile portability so that typical tasks such as server identity, LAN- and SAN-based I/O configuration, firmware management, and network connectivity can be applied to any piece of hardware within the system. This feature speeds up just-in-time deployment of new computing resources for data warehouses.

Cisco Unified Computing System Service Profile for a Data Warehouse

For the baseline block compute server, a service profile consisting of 44-Gbps HBAs and two 10-Gbps NICs was deployed as shown in Figure 2.

© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

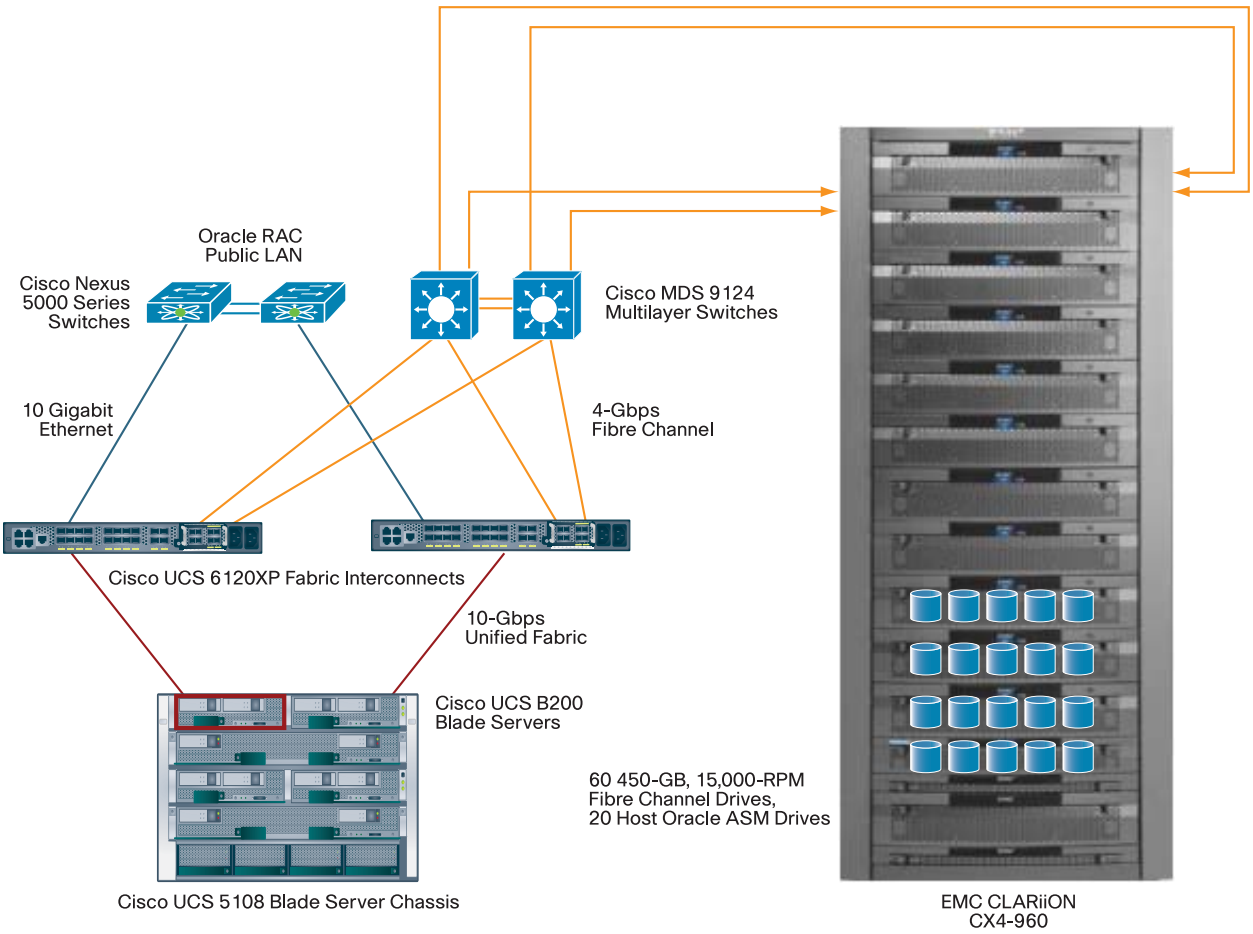
Figure 2. Service Profile for Baseline Server



In the initial phase, a single Oracle RAC instance block was deployed in a Cisco UCS B200 M1 Blade Server in a Cisco UCS 5108 Blade Server Chassis (Figure 3).

© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Figure 3. Single Server Running Oracle RAC Block



Note: Red lines represent unified fabric, blue lines represent 10 Gigabit Ethernet, and orange lines represent 4-Gbps Fibre Channel.

In the power-run test, one user stream took 83 seconds to complete. Based on established rules for concurrency, one blade server can support 16 concurrent users while maintaining a response time of less than 2.5 times that of a power run. For this run, CPU utilization was noted to be consistently at approximately 90 percent. For Fibre Channel I/O, the peak data rate per blade was observed to be approximately 1490 MBps, with 1400 MBps typically sustained using four Fibre Channel HBAs. This test demonstrates the unified fabric's ability to handle multiple parallel traffic streams without any affect on performance.

Data Warehouse Scale Out Using Block Addition

After a baseline block is established, the next step is to test the solution' scalability through the addition of building blocks. The steps involved in block expansion include:

- **Add compute nodes and extend the fabric:** With the service profile, the Cisco Unified Computing System speeds up just-in-time deployment of new computing resources. The Cisco Unified Computing System architecture is built on a unified fabric that enables a wire-once model. In this model, adding compute nodes does not require any change in I/O configuration and no longer requires installiation of adapters or recabling of racks and switches. With this approach, compute infrastructure can be provisioned in minutes instead of days.

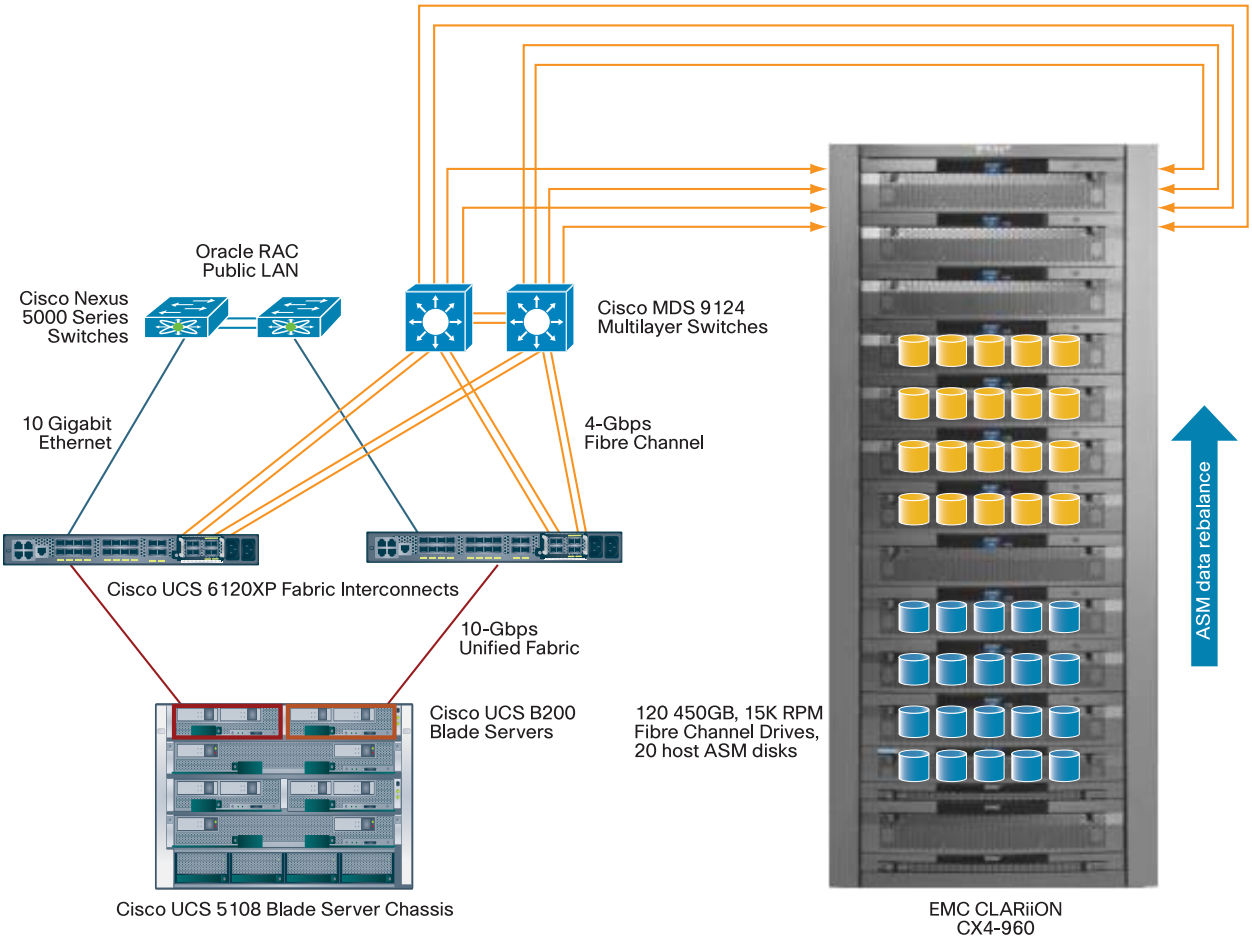
© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

- **Add new storage logical unit numbers (LUNs) and rebalance using Oracle Automatic Storage Management (ASM):** The EMC CLARiiON's flexible, modular design makes LUN provisioning extremely fast and simple. After the LUNs from a new disk array are provisioned, they can easily be added to an existing Oracle ASM disk group, and the data can be rebalanced across the newly added LUNs. Oracle ASM makes LUN expansion and rebalancing simple and reliable.

Two-Node Oracle RAC Data Warehouse Block

Figure 4 shows the architecture for two blocks.

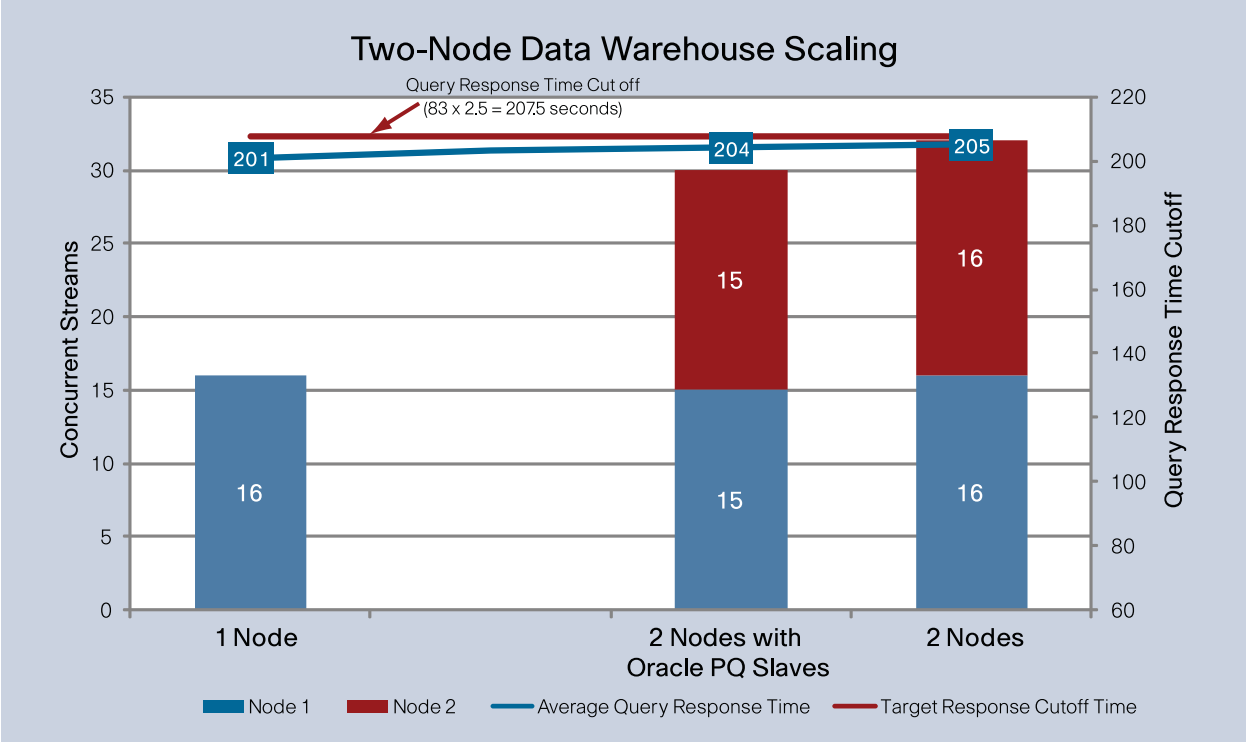
Figure 4. Scaling from One to Two Oracle RAC Blocks with Two Blade Servers



For this two-node Oracle RAC solution, a new block consisting of an additional Cisco UCS blade server and the same number of storage spindles was added to the baseline configuration. The provisioning of the second blade from Cisco UCS Manager using a service profile was extremely simple, and the server was online in minutes. The data was rebalanced across the newly added disks (Figure 5) using Oracle ASM. Observed results from testing the two-node block follow.

© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Figure 5. Performance Results Based on Users and Oracle RAC Instances



For the two-node Oracle RAC configuration, two kinds of tests were run:

- Manual distribution of queries (load) across the nodes
- Use of Oracle Parallel Query feature to distribute the load across the cluster

For the manual approach, 16 user query load streams were started up on each of the two RAC nodes of the two-node cluster. Linear scalability was observed, with twice the number of users serviced while achieving nearly the same service response time as with a single blade.

Instead of a manual approach, Oracle Parallel Query may be used to automatically balance the load among clustered servers. 30 user query load streams were started all from node 1 of the two-node cluster, with Oracle Parallel Query automatically load distributing and balancing the query subtasks across both RAC nodes supporting the database service. The parallel query feature requires slightly higher overhead, but the benefits of this approach are well documented. Use of Oracle Parallel Query results in higher Oracle RAC interconnect traffic as data blocks are transferred across instances. However, the lossless and low-latency 10 Gigabit Ethernet network and superior design of the Cisco Unified Computing System reduced the impact on performance, and this configuration could still handle 30 users.

Figures 6 and 7 show system resource utilization charts for the two-node cluster.

© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Figure 6. Two-Node Oracle RAC Server CPU Utilization

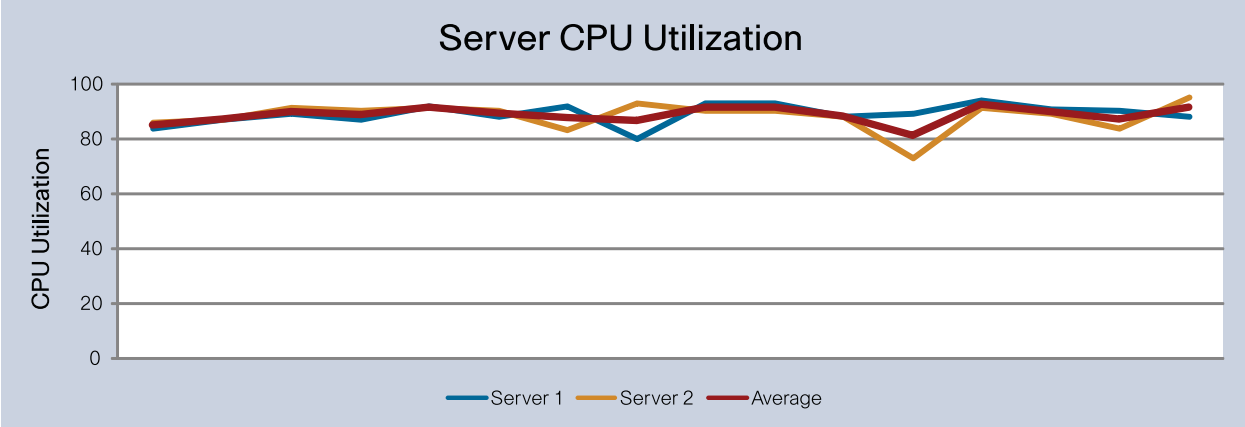
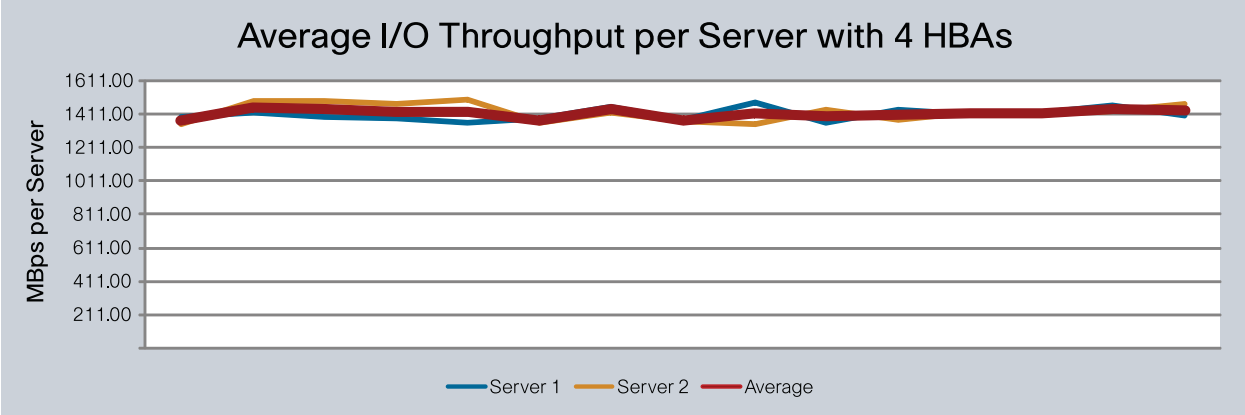


Figure 7. Two-Node Oracle RAC Server Average I/O Throughput



The main observations for the two-node Oracle RAC solution include:

- Average CPU utilization was approximately 90 percent (user plus system) with very low CPU idle time.
- Each server sent approximately 1400 MBps of I/O simultaneously using shared database LUNs.
- Fibre Channel I/O traffic was evenly distributed across four HBAs using EMC PowerPath path management solution.
- Testing involved 16 concurrent complex user reporting streams on each node: 32 in total.

The data warehouse solution was then scaled to a four-node block.

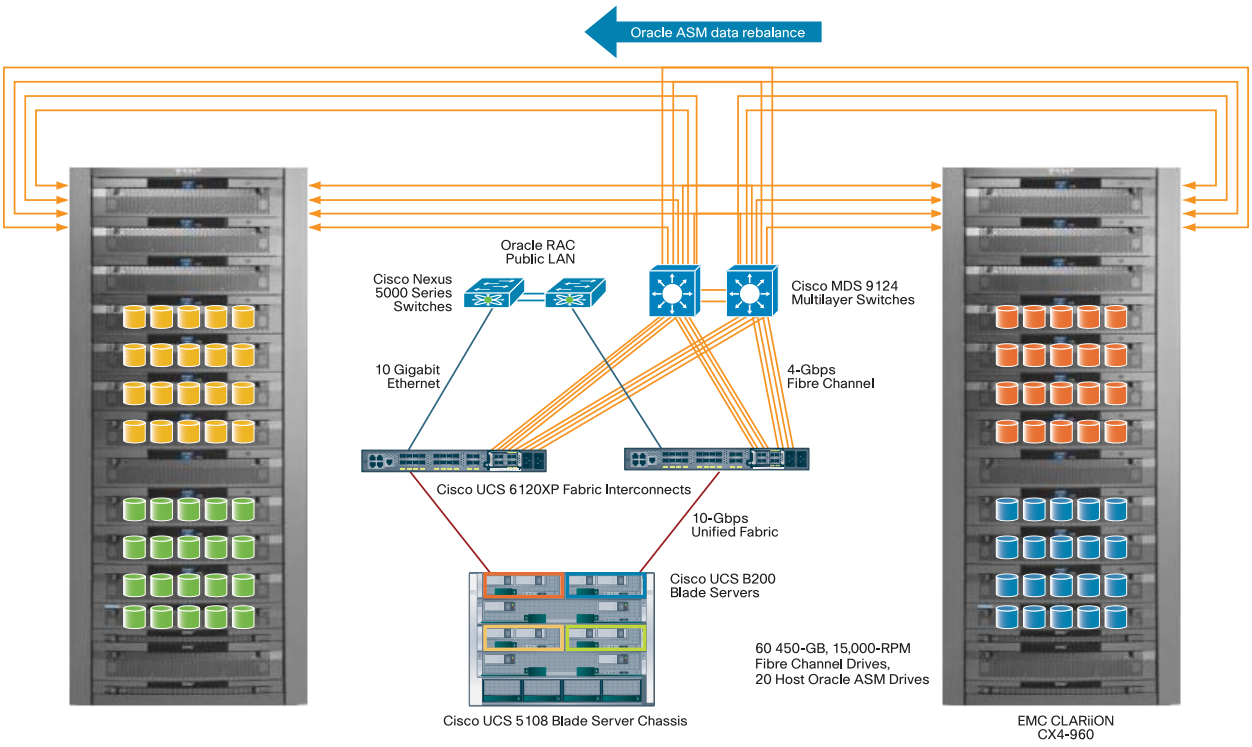
Four-Node Oracle RAC Data Warehouse Block

For a four-node Oracle RAC, a second EMC CLARiiON storage array was introduced to accommodate the data warehouse I/O servicing demand from 240 drive spindles. Oracle ASM does a good job of evenly redistributing the data across both storage systems. However, Oracle ASM cannot predict how the data as distributed will actually be accessed by the subsequent queries. For example, consider a yearly sales forecast table partition of 100 GB that is distributed evenly across 240 drives and across two storage systems so that the first storage system has data for the odd months (January, March, etc.) and the second storage system has data for the even months (February, April, etc.). In this case, there is no guarantee that a query will always retrieve data from both storage systems. These kinds of data patterns are extremely common and nearly impossible to avoid.

Figure 8 shows the architecture for four blocks.

© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Figure 8. Scaling to Four-Node Oracle RAC Data Warehouse Blocks with Four Blade Servers



Overall Results of Four-Node Oracle RAC Data Warehouse Testing

For the four-node Oracle RAC configuration, the same tests used for the two-node cluster were repeated, including:

- Manual distribution of load across the nodes
- Use of the Oracle parallel query feature to distribute the load across the cluster

For four-nodes, 60 users could be run using manual load distribution and 56 users with the Oracle parallel query feature while staying within query response time cutoff limits (Figures 9, 10, and 11).

© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Figure 9. Performance Results Based on Users and Oracle RAC Instances

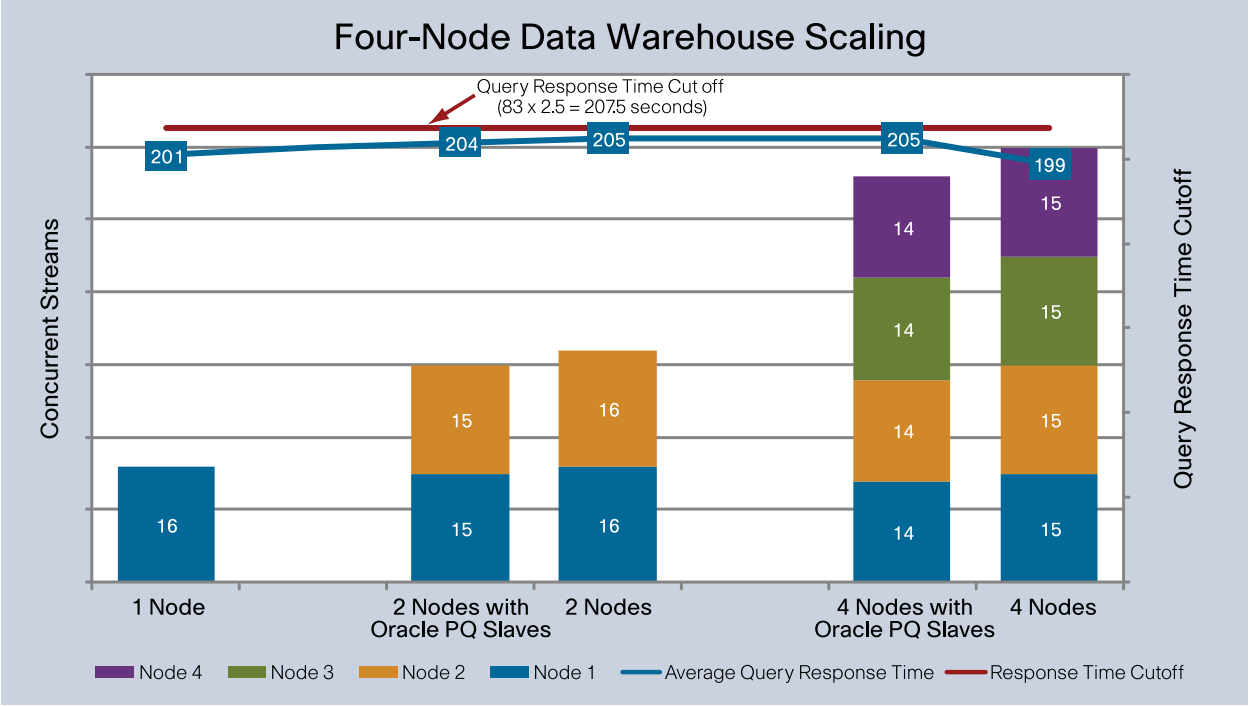
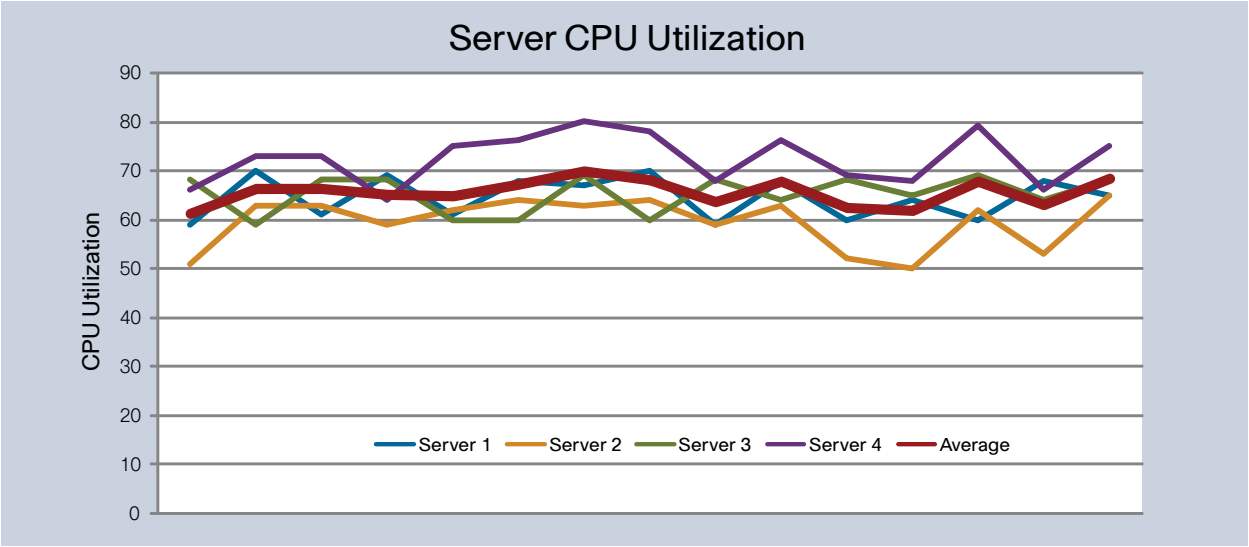
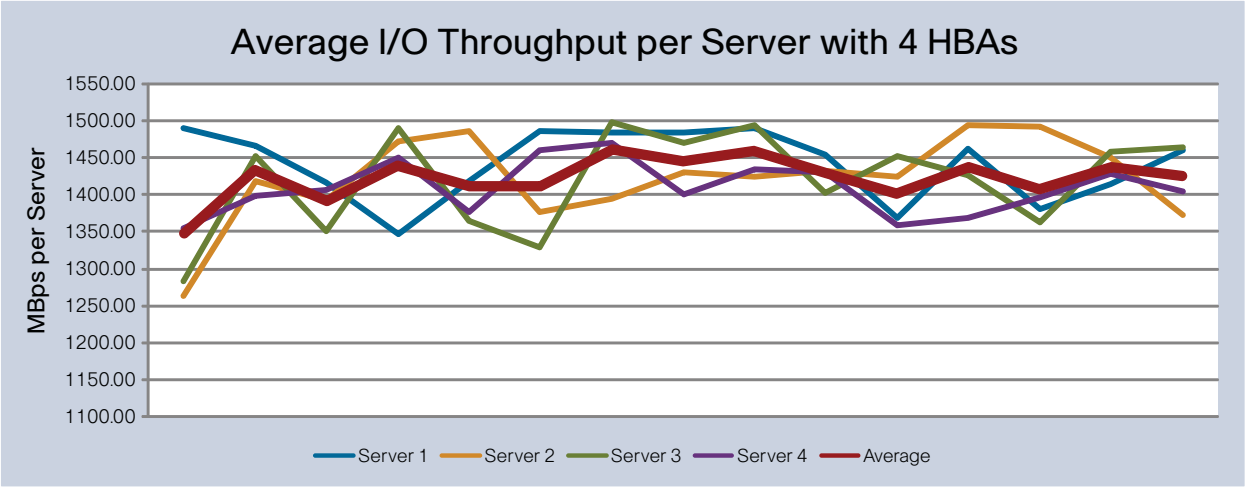


Figure 10. Four-Node Oracle RAC Server CPU Utilization



© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

Figure 11. Four-Node Oracle RAC Server Average I/O Throughput



Figures 10 and 11 show the server resources for the four-node cluster. During the four-node Oracle RAC data warehouse testing, the following was observed:

- Average CPU utilization across all four node nodes was approximately 65 percent (user plus system).
- Each server read approximately 1425 MBps simultaneously from shared database LUNs that spanned two EMC CLARiiON 960 storage systems.
- For each server, the Fibre Channel I/O traffic was evenly distributed across four HBAs using EMC PowerPath.

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 EMC and Oracle because Oracle databases and CLARiiON storage systems 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.

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, and provides dynamic resource provisioning for increased business agility.

The benefits of the Cisco Unified Computing System include:

- **Reduced total cost of ownership** (TCO), enabling up to 20 percent reduction in capital expenditures (CapEx) and up to 30 percent reduction in operating expenses (OpEx)
- **Improved IT productivity** and business agility by enabling IT to provision applications in minutes instead of days and shifting the focus from IT maintenance to IT innovation
- **Increased scalability** without added complexity, since the data warehouse is managed as a single system, whether the system has one server or 320 servers with thousands of virtual machines
- **Improved energy efficiency** with significantly reduced power and cooling costs
- **Interoperability and investment protection** from an infrastructure based on industry standards

© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.

For More Information

- Cisco Unified Computing System: http://www.cisco.com/en/US/netsol/ns944/index.html
- Deploying Oracle Real Application Clusters on the Cisco Unified Computing System with EMC CLARiiON Storage: http://www.cisco.com/en/US/prod/collateral/ps10265/ps10280/white_paper_c11-562881_ps9402_Products_White_Paper.html
- Cisco Unified Computing System: Benchmarks, Industry Recognition, Software Interoperability: http://www.cisco.com/en/US/prod/ps10265/at_work_promo.html
- EMC CLARiiON Storage Systems: http://www.emc.com/products/family/clariion-family.htm

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

CCDE, CCENT, CCSI, Cisco Eos, Cisco Explorer, Cisco HealthPresence, Cisco IronPort, the Cisco logo, Cisco Nurse Connect, Cisco Pulse, Cisco SensorBase, Cisco StackPower, Cisco StadiumVision, Cisco TelePresence, Cisco TrustSec, Cisco Unified Computing System, Cisco WebEx, DCE, Flip Channels, Flip for Good, Flip Mino, Flipshare (Design), Flip Ultra, Flip Video, Flip Video (Design), Instant Broadband, and Welcome to the Human Network are trademarks; Changing the Way We Work, Live, Play, and Learn, Cisco Capital, Cisco Capital (Design), Cisco:Financed (Stylized), Cisco Store, Flip Gift Card, and One Million Acts of Green are service marks; and Access Registrar, Aironet, AllTouch, 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 Lumin, Cisco Nexus, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Collaboration Without Limitation, Continuum, EtherFast, EtherSwitch, Event Center, Explorer, Follow Me Browsing, GainMaker, iLYNX, IOS, iPhone, IronPort, the IronPort logo, Laser Link, LightStream, Linksys, MeetingPlace, MeetingPlace Chime Sound, MGX, Networkers, Networking Academy, PCNow, PIX, PowerKEY, PowerPanels, PowerTV, PowerTV (Design), PowerVu, Prisma, ProConnect, ROSA, SenderBase, SMARTnet, Spectrum Expert, StackWise, WebEx, and the WebEx logo are registered trademarks of Cisco 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. (1002R)

Printed in USA C11-588286-00 03/10

© 2010 Cisco and/or its affiliates. All rights reserved. This document is Cisco Public Information.