Technical Report

Oracle 11*g*R2 DSS/DW Performance on Cisco UCS and NetApp Storage

J. Thomas Simpkins and Chris Lemmons, NetApp

June 2011 | TR-3926

ABSTRACT

This technical report describes the performance of a series of tests using Oracle® 11*g*™R2 Real Application Cluster (RAC) running on Cisco® Unified Computing System™ (UCS) B200 M1 blade servers and NetApp® 3170 clusters with Oracle Direct Network File System (D-NFS) over 10-Gigabit Ethernet (GbE). The tests use a common workload generator designed to emulate decision support systems/data warehouse (DSS/DW) workloads.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION AND EXECUTIVE SUMMARY

Data powers essentially every operation in a modern enterprise. This is especially evident in the big-data space of DSS/DW environments. As the size and demands of data warehouse systems continue to grow, so do the challenges of meeting system availability and performance criteria. Adding to the difficulty, these challenges must be met in environments with shortened maintenance windows under pressure to continually lower total cost of ownership (TCO).

With the introduction of the Cisco UCS, Cisco provides state-of-the-art technology for supporting clustered infrastructures for Oracle RAC database deployments. The Cisco UCS provides compute, network, virtualization, and storage access resources that are centrally controlled and managed as a single cohesive system. With the capability to scale up to 14 UCS chassis (112 B200 blades), the Cisco UCS provides an ideal foundation for very large-scale Oracle RAC deployments to support DSS/DW environments.

NetApp unified storage helps you meet the growing challenges of decision support systems. Our innovative storage solutions deliver the high levels of reliability, availability, and serviceability that today's decision support systems demand. With NetApp, you can perform frequent space-efficient backups of your data warehouses and recover data in minutes. Efficient, low-cost replication makes disaster recovery economically viable and our unique cloning technologies eliminate the need for additional physical copies for data marts, development and test, or other functions. Because of NetApp's superior storage efficiency, you can raise your overall storage use significantly while delivering the flexibility, scalability, and performance you need for continued growth.

The configuration defined in this technical report is based on a Cisco Validated Design (CVD), Cisco UCS and NetApp Solution for Oracle Real Application Clusters. It combines Cisco's revolutionary UCS architecture and networking with NetApp's unified storage systems to create a high-performance, reliable, and well-balanced system for Oracle 11*g*R2 RAC. NetApp's unified storage system and Cisco's UCS platform work together to create a system that is easy to manage, is scalable, and provides high levels of fault tolerance.

All components in a DSS implementation must work together seamlessly. Cisco and NetApp have worked together closely to create, test, and validate a balanced platform for Oracle RAC DSS/DW environments. This configuration provides an implementation of Oracle Database 11*g*R2 with RAC technology consistent with industry best practices. To provide storage and data management capabilities, this architecture uses NetApp fabric-attached storage (FAS) unified storage systems with SAS drives to further speed performance while providing space-efficient Snapshot® copies, disaster recovery, and cloning capabilities. The result is an implementation that addresses many of the challenges that database administrators and their IT departments face, including the need for a simplified deployment and operation model, high performance for Oracle RAC software, and lower TCO.

1.1 EXECUTIVE SUMMARY

This technical report provides high-level guidelines for configuring a balanced environment and provides links to more detailed installation and configuration guides for the various layers, including UCS, network, storage, and software. Also, we measured the performance of the system using a common DSS/DW workload generation suite. The workload generation suite comprises three distinct query sets chosen specifically to mimic the various types of DSS workloads that might occur in a production environment. These workload generation suites are:

- **Throughput workload.** We chose the first workload to represent users running multiple iterations of reports or queries based on the same table sets with different parameters to filter the data to different geographic regions or different time periods. We used this workload to drive the throughput to a high level while allowing some queries that could use at least a portion of cached data.
- **Analytic workload.** We chose the second workload to highlight the balanced nature of the configuration by using processes and queries that not only scanned the database with large

sequential reads but also performed more compute-intensive functions on the row sets that might aid in data analysis. In most DSS/DW installations, functions are used to analyze and summarize the vast amount of data returned from the storage system.

- **Scan workload.** We chose the third and final workload to represent a simple full scan of the centralized fact table, which represented returning a subset of data for further analysis. Typical of this type of query/process, the host-side CPU use is expected to be extremely low. Because it is a sequential scan of a large fact table, this workload does not use cache but instead relies on the NetApp storage system to service the requests directly from disks.

Oracle D-NFS was used as the primary mode for database storage access. D-NFS is an Oracle-developed, -integrated, and -optimized NFS client that runs in user space rather than within the operating system kernel. D-NFS is a baseline feature that is available in the Oracle Database 11*g* platform. This architecture can, in the right circumstance, provide enhanced scalability and performance compared to traditional NFSv3 clients. D-NFS generally requires lower amounts of host-side CPU resources and can scale across as many as four individual network pathways automatically to each mount point, providing the added benefit of improved resiliency when network connectivity is occasionally compromised.

The test results demonstrate the potential level of performance achievable from this configuration in a production environment using D-NFS over 10GbE. All performance tests were executed using the three workloads previously described and restored to a base point in time using aggregate-level Snapshot copies between each run. The tests were not designed to stress the system to its maximum but rather to demonstrate the performance for each given type of simulated DSS workload. Overall, the performance testing showed the configuration is capable of delivering enterprise-class performance for sustained periods with no observed errors or other issues.

Using information gathered from the Automatic Workload Repository (AWR) reports during the Throughput workload test, this configuration sustained an average I/O throughput of approximately 3.1GB/s from the four NetApp FAS3170 storage controllers to the four Oracle RAC nodes, with sustained load periods observed on the storage side of approximately 3.6GB/s delivered to the same RAC nodes.

During testing with the Analytic workload, the four Oracle RAC node hosts averaged approximately 43% CPU usage while the CPU usage on the NetApp FAS3170 storage controllers maintained an average of approximately 79% during an extended test period in which the full query load was running. This demonstrates how the system can effectively service workloads that have demands on both the host and storage resources.

During the Throughput and Scan workloads, the host-side CPU usage hovered around 12%, demonstrating that these workloads required little host-side processing and were almost entirely I/O-centric.

In our testing, D-NFS delivered performance comparable to that available using block protocols such as Fibre Channel Protocol (FCP). This suggests D-NFS is a viable choice for DSS/DW configurations, which have typically been deployed on dedicated storage area network (SAN) environments using block protocols such as FCP. D-NFS can provide similar high levels of performance with the added benefit of being easier to manage and possibly reducing TCO by allowing, among other things, the use of a common type of infrastructure to support both network and storage traffic.

The remainder of this report discusses the architecture of the Cisco-NetApp solution and provides the details of the performance testing. It provides a high-level overview of the configuration and specific settings for kernel parameters, `init.ora` settings, and NFS mount options.

2 ARCHITECTURE

The architecture for this test is based on the Cisco 3.0 data center architecture and follows best practices put forth by Cisco, NetApp, and Oracle. Any deviations from best practices are explained in the appropriate sections. This architecture section provides an overview of the configuration's overall architecture. Links are provided in the various sections to more detailed configuration guides. The CVD document, Cisco UCS and NetApp Solution for Oracle Real Application Clusters (RAC), also provides more detailed configuration information for this specific environment.

2.1 TOPOLOGY OVERVIEW

Figure 1 provides a graphic description of the overall configuration, including all compute, network, and storage resources. This figure identifies the various levels of the architecture as well as some of the key components and features of the fabric. Although the core/aggregation level is recommended for data center deployments and is illustrated in the diagram, it was not part of the test architecture. This is because the testing and loads were concentrated on the access layer and below and did not involve the larger core/aggregation levels.

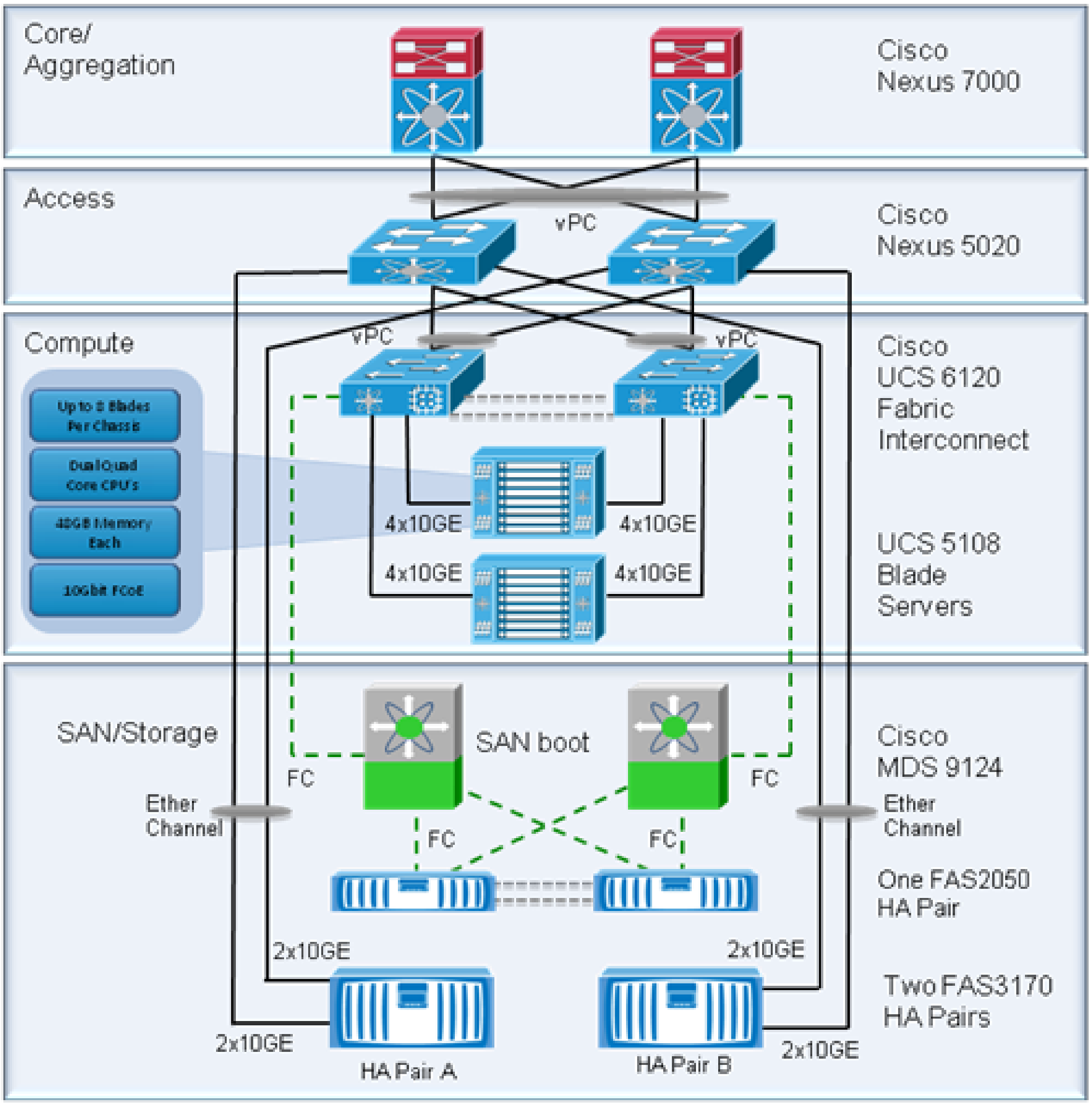
The Oracle RAC configuration we used for these tests included only components up to the access layer; that is, up to the Cisco Nexus® 5000 switch pair. The two UCS 6120 fabric interconnects (FIs) with dual-fabric topology enable a 10GbE compute layer. Both the UCS 6120 FIs and NetApp FAS3170 storage controllers are connected to the Nexus 5000 access switch through port channel with dual-10GbE. The NetApp FAS3170 controllers use redundant 10GbE connections configured in a two-port virtual interface (VIF). Each port of the VIF is connected to one of the upstream Nexus 5020 switches, allowing multiple active paths by using the Nexus virtual port channel (vPC) feature. This topology, combined with vPC, provides increased redundancy and bandwidth with a lower required port count.

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

Note: The NetApp FAS2050 was only used for SAN boot functionality and is not a required part of this architecture because the NetApp FAS3170s could be used for SAN boot functionality as well.

At the compute layer, Cisco UCS provides a unified environment with integrated management and networking to support compute resources. Our configuration included a total of four B200 M1 blade servers. Each of the four UCS blades served as a node in the Oracle RAC cluster.

Figure 1) Full topology (graphic supplied by Cisco).



Note: This architecture uses two FAS3170 high-availability (HA) pairs (four controller heads). The interconnect for the FAS3170 is internal to the chassis, where the FAS2050 has an external cable for the interconnect. Each controller had two 10GbE connections configured in a VIF with one wire going to each of the FAS5020 switches.

2.2 UCS COMPONENTS

This section provides the details of the various Cisco UCS components used in the test architecture. Descriptions of the components as well as diagrams of how they were connected are provided.

A detailed description of all the hardware components is beyond the scope of this document. Cisco UCS B-Series Servers Documentation Roadmap provides detailed documentation for each UCS component.

B200 M1 BLADE

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

6120 FABRIC INTERCONNECT

We used a pair of 6120 FIs for the testing, configured in an HA pair. The FC connections to storage were established using the global expansion modules. Some of the 20 fixed 10GbE ports were used for connecting to the upstream Nexus switches. The hybrid display from the UCS Manager GUI is shown in Figure 2 and Figure 3 to provide a logical and physical view of the topology. Note that the UCS fabric extenders are internal to the UCS chassis and are shown separately for illustrative purposes only.

Figure 2) 6120 FI topology, UCS chassis 1.

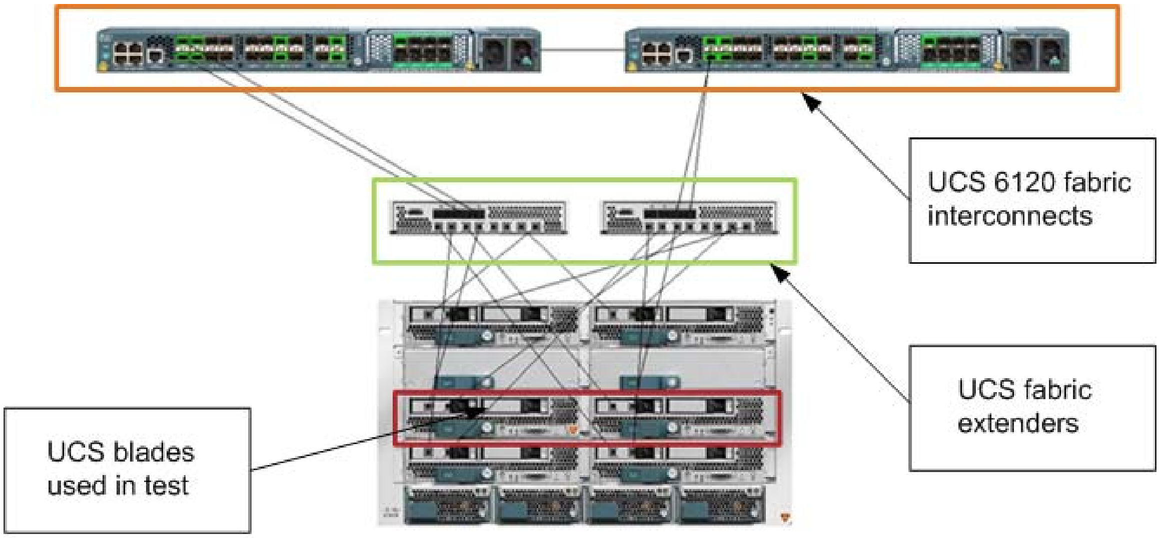
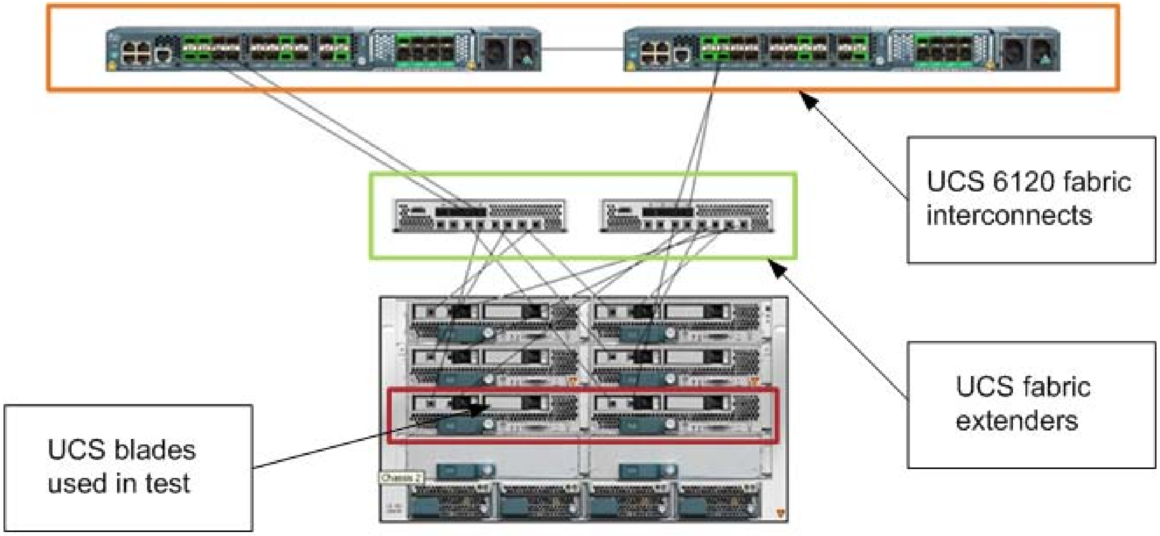


Figure 3) 6120 FI topology, UCS chassis 2.



I/O MEZZANINE CARD

We used the Cisco UCS M81KR Virtual Interface Card (VIC) as the I/O card for this project. It can present up to 128 virtual host bus adapters (vHBAs) and 10GbE virtual network interface cards (vNICs) to the operating system. We assigned one port or NIC/HBA to each FI. The card allows seamless inclusion into standard Ethernet and FC SAN networks.

2.3 NETWORK COMPONENTS

This section provides the details of the main networking components used in the architecture. These include descriptions of the switches for both the Ethernet and FCP. These components connect the blade servers to the network for connectivity as well as the storage for data access and SAN boot.

NEXUS 5020 SWITCH

The Cisco Nexus 5000 Series switches comprise a family of line-rate, low-latency, lossless 10GbE and FC over Ethernet (FCoE) switches for data center applications. This switch also delivers more than 1Tbps of switching capacity with 40 fixed wire-speed 10GbE ports that support Data Center Bridging (DCB) and FCoE.

MDS 9124 SWITCH

The Cisco MDS 9124 24-Port Multilayer Fabric switch features 24 ports that are capable of FC speeds of 4, 2, and 1 Gbps. It also offers outstanding value by providing a flexible, highly available, and secure solution that is both affordable and easy to use, all in a compact one-rack-unit (1-RU) form factor. The Cisco MDS 9124 supports the same industry-leading MDS 9000 SAN OS software that is supported on the entire Cisco MDS 9000 family of products.

2.4 NETAPP COMPONENTS

This section provides the details of the core NetApp storage components used in the architecture. It provides descriptions of the storage controllers and disk shelves for the main data access storage.

FAS3170

NetApp FAS systems simplify data management, enabling enterprise customers to reduce costs and complexities, minimize risks, and control change. NetApp FAS systems are the most versatile storage systems in the industry for storage consolidation. The FAS3170 addresses the core requirements of the midrange enterprise storage market, delivering a superb blend of price, performance, and scalability for DSS/DW systems. The compact, modular design provides native support for FCoE, FC, iSCSI, and network-attached storage (NAS) storage with scalability to over 800 disk drives. The FAS3170 storage controller supports FC, SAS, and SATA disk drives. FAS3170 systems support as many as 40 FC ports or 36 Ethernet ports, including support for 4GB FC, 8GB FC, and 10GbE.

DS4243 DISK SHELVES

The DS4243 is NetApp's next-generation SAS-architecture disk shelf, providing customers with increased storage density, resiliency, and power efficiency. The 4U-high 24-drive enclosure supports either 15,000 RPM SAS or 7,200 RPM SATA drives and uses native SAS architecture.

3 SYSTEM CONFIGURATION

This section provides a high-level overview of the overall system configuration. It touches on the network, UCS server, and storage setup. This is only a high-level view and provides high-level guidelines for the system configuration. Links to more detailed information are in the various subsections and can also be found in section 7, References.

3.1 NETWORK CONFIGURATION

The testing did not incorporate Network Core and Aggregation layers; however, the infrastructure deployment has adopted the best practices recommended in the following design guides. Adopting the best practices enabled the access layer to be seamlessly integrated into the Cisco DC 3.0 architecture. Any exceptions and specific changes relevant to this deployment are explained in the appropriate sections.

- Cisco Data Center Design—IP Network Infrastructure
- Cisco SAFE Reference Guide

3.2 UCS CONFIGURATION

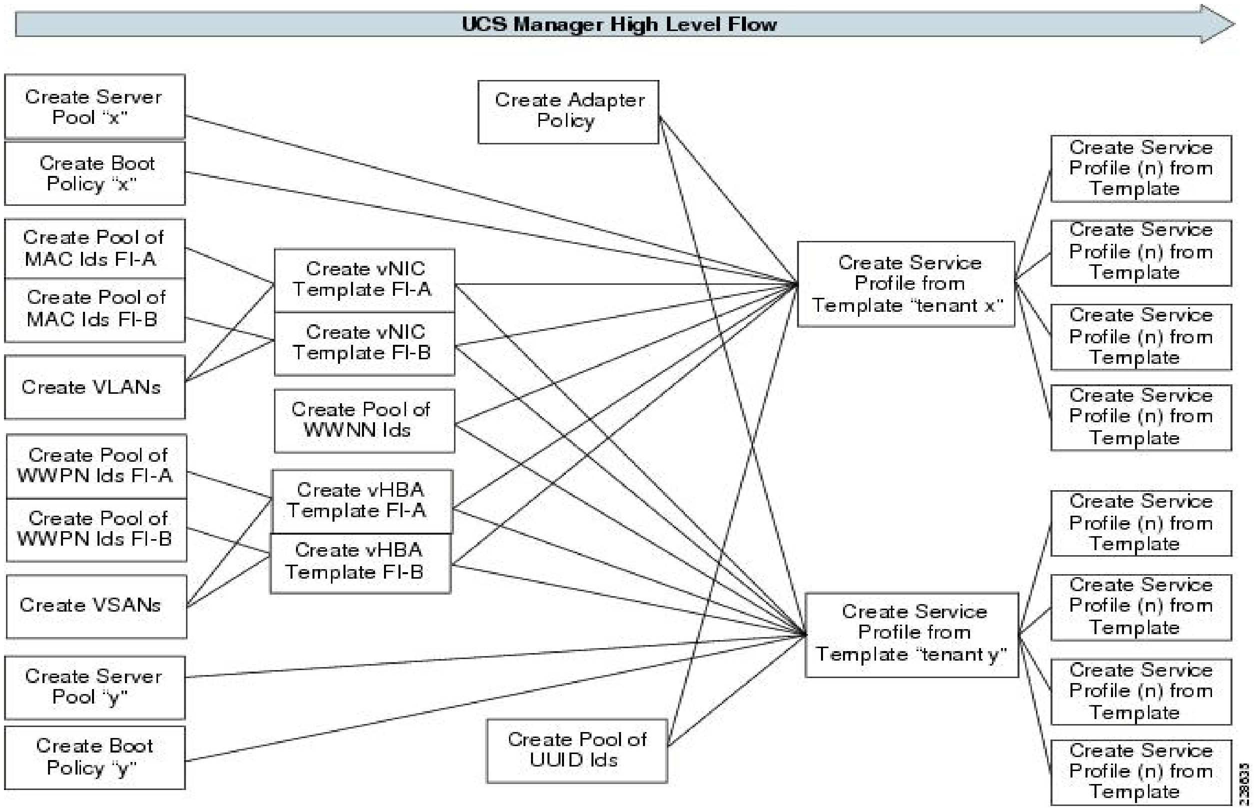
This section describes the configuration steps for the UCS with a brief description of the rationale for each step. The step-by-step procedures for each operation are beyond the scope of this document. The initial setup of the UCS system, including cabling and initial network and chassis configuration, is also beyond the scope of this document. These procedures can be accessed from Cisco UCS Manager GUI Configuration Guide.

All steps are performed using the UCS Manager Java GUI unless otherwise specified. The best practice for implementing UCS is first to engineer elements such as organizations, resource pools, policies, and templates. This flow is shown graphically in Figure 4. There are two types of service profile templates: updating and initial. Initial templates are generally easier to manage because changes to the original template do not result in changes to the service profiles created from the template. Updating templates result in the downstream service profiles being immediately updated with any change, which might or might not be desirable depending on the use case and the nature of the change to the template. This design used updating templates.

Figure 4 shows a high-level summary of the overall flow of operations. The intent of Figure 4 is to show which steps and actions feed the subsequent actions, with the flow going from left to right. The service profile template is the key construct that, once created, allows rapid creation and provisioning of new service profiles and servers. Not all of the exposed UCS policies and capabilities are shown in Figure 4; the figure is intended to provide the reader with a guide to the overall flow process. This figure provides a general sequence of steps. A screen-by-screen sequence is not in the scope of this document.

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

Figure 4) UCS manager high-level flow (graphic supplied by Cisco).



3.3 STORAGE CONFIGURATION

This section of the document provides a general overview of the storage configuration used to support the database layout. It also discusses the VIFs configuration used to achieve HA and load-balancing access from the database host to the NetApp storage. We used two FAS3170 HA pairs in the configuration described in this report. For more information about NetApp FAS storage, refer to netapp.com.

The following NetApp storage model was used for this testing.

Table 1) FAS3170 HA pair A and FAS3170 HA pair B.

HA Pair	Controller	Disks	Total Storage Size	Flash Cache Size
HA pair A	FAS-3170_A_1	48 disks (450GB 15K RPM SAS drives)	19TB	512GB
HA pair A	FAS-3170_A_2	48 disks (450GB 15K RPM SAS drives)	19TB	512GB
HA pair B	FAS-3170_B_1	48 disks (450GB 15K RPM SAS drives)	19TB	512GB
HA pair B	FAS-3170_B_2	48 disks (450GB 15K RPM SAS drives)	19TB	512GB

Table 2 provides the details of the aggregates created on each of the FAS3170 storage controllers, including the aggregate name, RAID group type and size, the usable capacity, and the purpose for the specific aggregate.

Table 2) Aggregate layout.

Controller	Aggregate Name	Option /RG Size	# of Disks/ Usable Size	Purpose
FAS-3170_A_1	aggr0	RAID-DP, RG-3	3 disks/350GB	DOT and root volume
FAS-3170_A_1	AGGR_ORA_1_A_1	RAID-DP, RG-21	42 disks/13TB	Data files, redo logs, control files, database binary
FAS-3170_A_2	aggr0	RAID-DP, RG-3	3 disks/350GB	DOT and root volume
FAS-3170_A_2	AGGR_ORA_2_A_1	RAID-DP, RG-21	42 disks/13TB	Data files, redo logs, control files, Flash Recovery Area (FRA)
FAS-3170_B_1	aggr0	RAID-DP, RG-3	3 disks/350GB	DOT and root volume
FAS-3170_B_1	AGGR_ORA_1_B_1	RAID-DP, RG-21	42 disks/13TB	Data files, redo logs, control files
FAS-3170_B_2	aggr0	RAID-DP, RG-3	3 disks/350GB	DOT and root volume
FAS-3170_B_2	AGGR_ORA_2_B_1	RAID-DP, RG-21	42 disks/13TB	Data files, redo logs, control files, archive log

Table 3 provides the details of the volumes created for the test environment. This information includes the controller on which the volume is created, the volume name and containing aggregate, the volume size, and the purpose associated with the specific volume. For performance purposes, the data and logs were spread across all four controller heads. Spreading the data across the heads allows maximum performance from the entire system. This layout might need to be adjusted to better address backup, restore, and mirroring requirements in specific environments. TR-3633: NetApp Best Practice Guidelines for Oracle Database 11*g* provides thorough coverage of the best practices for Oracle Database 11*g*.

Table 3) Volume layout.

Controller	Volume name	Aggregate Name	Size	Purpose
FAS-3170_A_1	ORA_HOME	AGGR_ORA_1_A_1	100GB	Database binary
FAS-3170_A_1	OCR_CSS	AGGR_ORA_1_A_1	50GB	OCR and voting disks
FAS-3170_A_1	VOL_DATA_1_A_1	AGGR_ORA_1_A_1	12TB	Data files, control file
FAS-3170_A_1	VOL_LOG_1_A_1	AGGR_ORA_1_A_1	200GB	Redo log files
FAS-3170_A_2	VOL_DATA_2_A_1	AGGR_ORA_2_A_1	10TB	Data files, control file
FAS-3170_A_2	VOL_LOG_2_A_1	AGGR_ORA_2_A_1	200GB	Redo log files

Controller	Volume name	Aggregate Name	Size	Purpose
FAS-3170_A_2	VOL_FRA_2_A_1	AGGR_ORA_2_A_1	2TB	FRA
FAS-3170_B_1	VOL_DATA_1_B_1	AGGR_ORA_1_B_1	12TB	Data files, control file
FAS-3170_B_1	VOL_LOG_1_B_1	AGGR_ORA_1_B_1	200GB	Redo log files
FAS-3170_B_2	VOL_DATA_2_B_1	AGGR_ORA_2_B_1	10TB	Data files, control file
FAS-3170_B_2	VOL_LOG_2_B_1	AGGR_ORA_2_B_1	200GB	Redo log files
FAS-3170_B_2	VOL_ARC_2_B_1	AGGR_ORA_2_B_1	2TB	Archive log

4 PERFORMANCE TESTS AND RESULTS

4.1 PERFORMANCE TEST WORKLOADS

The workload for this set of performance tests and comparisons uses the database and query processes from an industry-standard DW workload test suite. The individual workloads described in the following list are subsets of the larger DW test suite and were chosen to represent specific real-world scenarios. For our performance testing, we chose a 1TB data warehouse.

- **Throughput workload.** We chose the first workload to represent users running multiple iterations of reports or queries based on the same table sets with different parameters to filter the data to different geographic regions or different time periods. This workload was used to drive higher levels of throughput while allowing some queries to be satisfied using data in the storage controller cache. Subsequently, approximately 75% of this workload is serviced directly from disk with the remaining 25% serviced from the storage controller cache in the baseline configuration.
- **Analytic workload.** We chose the second workload to highlight the balanced nature of the configuration by using processes and queries that not only scanned the database with large sequential reads but also performed more compute-intensive functions on the row sets that might aid in data analysis. In most DSS/DW installations, functions are used to analyze and summarize the vast amount of data returned from the storage system. As the server executes these functions on the data, this workload consumes more CPU resources in the database server compared to the other workloads.
- **Scan workload.** We chose the third and final workload to represent a simple full scan of the centralized fact table to represent returning a subset of data for further analysis. As is typical with this type of query/process, the host-side CPU use is expected to be extremely low. Because it is a sequential scan of a large fact table, this workload does not use significant cache but instead relies on the NetApp storage system to service the requests directly from disks.

The configuration we tested is designed to represent a sizable percentage of customer installations. Although there are many variances, this configuration should represent many common configuration choices made by NetApp customers. This configuration consists of Oracle Database 11*g*R2 on Red Hat Enterprise Linux® (RHEL) 5.4. It uses D-NFS over 10GbE and incorporates jumbo frames.

All performance tests were executed using the three workloads defined in the prior list. Aggregate-level Snapshot copies were used between tests to restore the database to the same condition for each test. The tests were not designed to stress the system to its maximum but were designed to demonstrate the performance of each given type of simulated DSS workload. Overall, the performance testing showed the configuration is capable of delivering enterprise-class performance for sustained periods with no observed errors or other performance issues.

The following general observations were noted during the performance testing for each of the specific workloads under test:

- **DSS throughput.** The information gathered from the AWR reports shows that this configuration sustained an average I/O throughput of approximately 3.1GB/s from the four NetApp FAS3170 storage controllers to the four RAC nodes over the test duration, with sustained load periods observed on the storage side of approximately 3.6GB/s read from the storage. This data demonstrates the system's ability to handle sustained bursts in demand successfully.
- **Analytic workload.** During the Analytic workload test, the four RAC node hosts maintained approximately 34% CPU usage for the duration of the test while the CPU usage on the NetApp FAS3170 storage controllers maintained an average of approximately 77%. The CPU usage for the hosts was 43% and the storage was 79% during an extended test period in which the full query load was running. This demonstrates the system's ability to address more CPU-intensive workloads that are common in DSS environments.
- During the Throughput and Scan workloads, the host-side CPU usage hovered around 10–12%, demonstrating that these workloads required little host-side processing and were almost entirely I/O-centric.
- **Scan workload.** During this test, the RAC nodes consumed approximately 2.5GB/s with sustained periods of over 3.5GB/s observed from the storage side. As expected, CPU usage was low (approximately 10%) on the database server side with the bulk of the load being placed on the storage system.
- Lack of saturation points within the subsystems (CPU, disk, I/O, or networking) demonstrates the balance of the architecture while maintaining available resources to address additional loads on the system.

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

- The simplified, performance-oriented architectural design of the Cisco UCS based on a 10Gbps unified fabric
- The pairing of the Cisco UCS with NetApp FAS storage
- The balance between UCS hosts, NetApp storage, and Cisco network to allow high levels of performance and the elimination of performance-limiting bottlenecks
- Use of the Oracle Database 11*g*R2 RAC and the D-NFS protocol

4.2 THROUGHPUT WORKLOAD

We chose the DSS Throughput workload to mimic a set of analysis queries or reports run against the DSS for various result sets relating to date ranges, geographies, and so on. These queries were run in parallel to increase the load and used some of the same tables and data. This allowed some of the information to be sourced from storage cache while the vast majority (approximately 75%) was sourced from the storage controller disks.

These particular tests were mostly I/O driven and did not tax the host-side CPU. These queries had minimal joins and functions and required only minimal host-side CPU resources. During the tests, the host-side CPU usage was approximately 12%. These tests demonstrated the system's ability to handle a heavy load over a sustained period of time without errors or performance degradation.

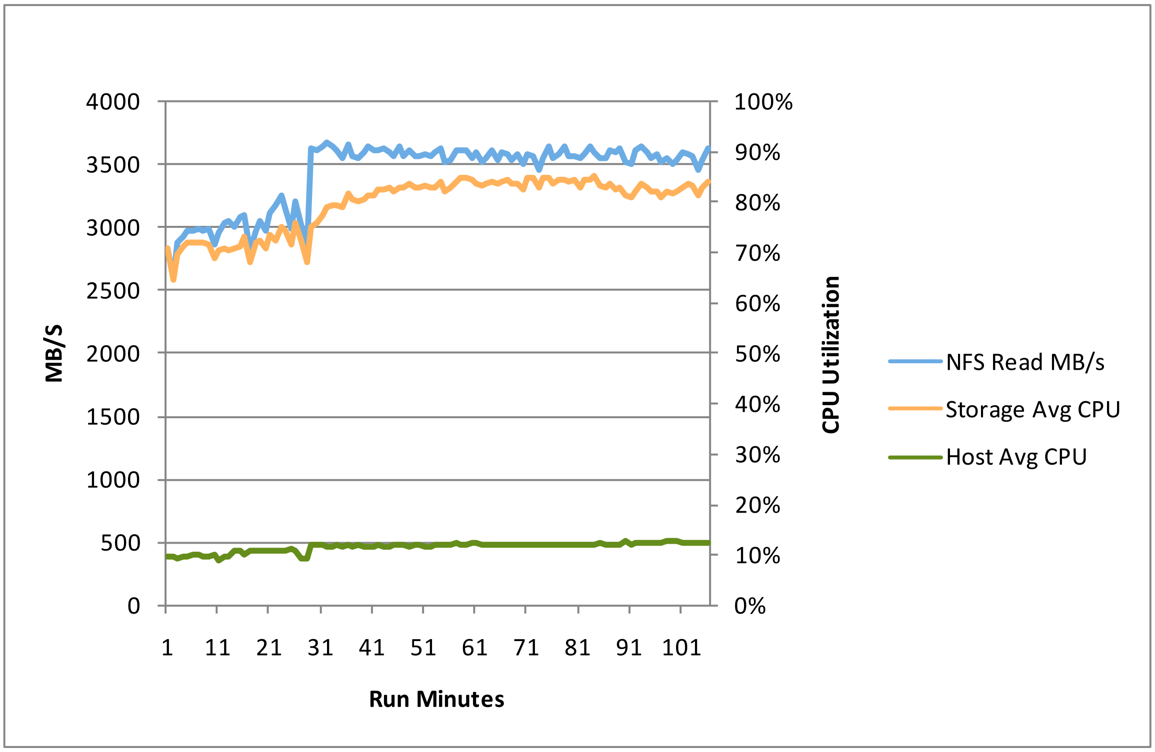
Table 4 shows the Physical Read MB/s value derived from the Oracle AWR report for each of the four RAC nodes (trdss[1-4] are the RAC instances) as well as the total throughput calculated for the RAC cluster. The results show how the load was spread evenly across the four Oracle RAC nodes.

Table 4) AWR physical read MB/s (Throughput workload).

trdss1	trdss2	trdss3	trdss4	Total
755MB/s	790MB/s	783MB/s	792MB/s	3120MB/s

Figure 5 shows the average NFS read throughput and CPU usage from the FAS3170 storage controllers as well as the average UCS host-side CPU usage during the same period. Approximately 75% of the total throughput was serviced by the storage controller disks. This test also demonstrates a sustained period in which the FAS3170 storage controllers are serving up over 3.6GB/s of throughput with no errors.

Figure 5) FAS3170 read throughput with storage and host CPU use during Throughput workload.



4.3 ANALYTIC WORKLOAD

The second workload we tested was chosen to mimic a more balanced load requiring more host-side processing of the data read from the storage controllers. This set of queries not only scans through the centralized fact table but it also performs multiple table joins and several functions on the data that require some host-side processing. This set of queries is typical of many DSS/DW environments in which a centralized fact table might be joined to several dimension tables in a star schema or just joins several tables to perform an initial pass into a deeper analytic exploration.

Again, this test was not designed to push the throughput of the configuration to the limits, but, rather, to demonstrate the performance with a load that not only taxed the storage but also required a significant amount of server-side processing. During the test, both host and storage systems maintained CPU resources available for handling additional workload if necessary.

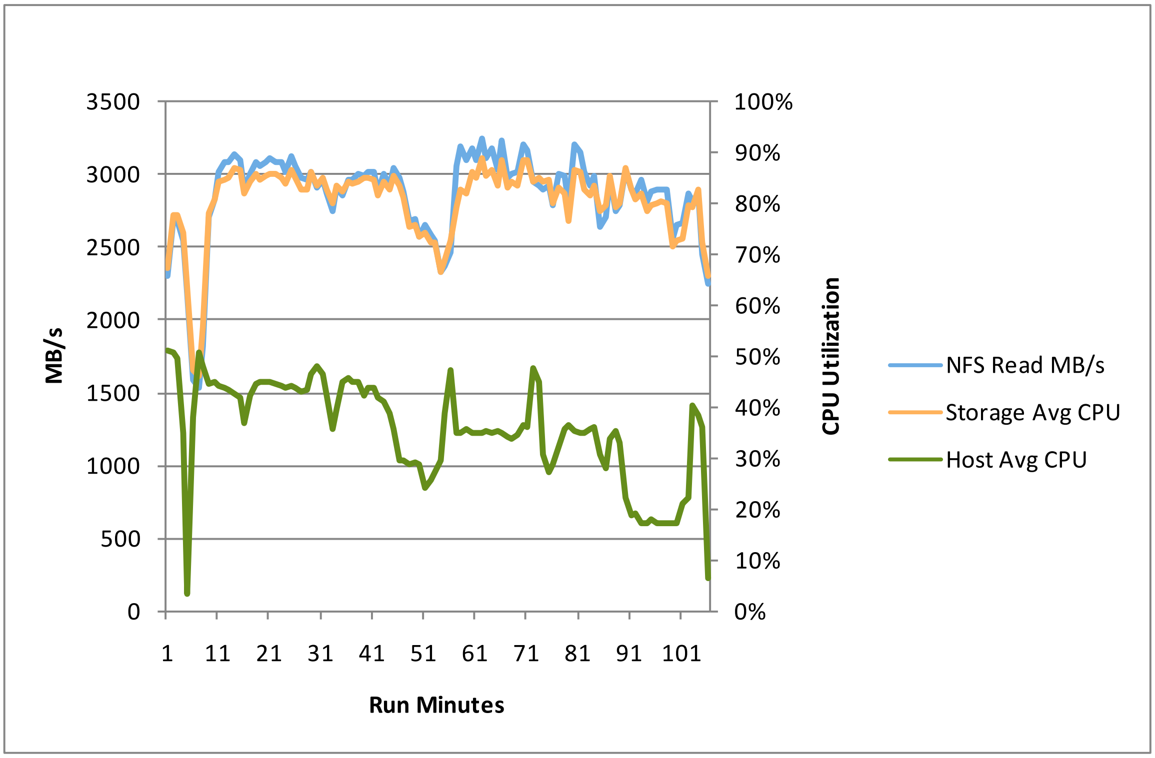
Table 5 shows the Physical Read MB/s value derived from the Oracle AWR report for each of the four RAC nodes (trdss[1-4] are the RAC instances) as well as the total throughput calculated for the RAC. These results again show how the load was spread evenly across the four RAC nodes. During the test, throughput consumed by the four RAC nodes was approximately 2.4GB/s with sustained periods observed on the storage side of approximately 3GB/s. The server-side CPU usage averaged 35% across the four RAC nodes with usage hitting 43% for a sustained period, representing a significant portion of the test run. The storage controller CPU usage averaged 74% for the duration of the test with a significant high load period averaging approximately 79%. In all cases, there was no indication of saturation of the network component or database errors.

Table 5) AWR physical read MB/s (Analytic workload).

trdss1	trdss2	trdss3	trdss4	Total
652MB/s	605MB/s	603MB/s	551MB/s	2411MB/s

Figure 6 shows the average NFS read throughput and CPU usage from the FAS3170 storage controllers as well as the average UCS host-side CPU usage during the same period.

Figure 6) FAS3170 read throughput with storage and host CPU use during Analytic workload.



4.4 SCAN WORKLOAD

The final workload was designed to represent a simple scan and aggregate on a single large dataset. This query set did not include joins and only included a single aggregate function. This query might represent a case in which users were attempting to extract a single answer from a centralized fact table or other aggregate to address a business question. Large sequential scans are common in DSS/DW environments, with this workload representing one of many possible scenarios.

During this test, the RAC nodes consumed approximately 2.5GB/s with sustained periods of over 3.5GB/s observed from the storage side. As expected, CPU usage was low (approximately 10%) on the database server side with the bulk of the load being placed on the storage system. During the test runs, the CPU usage on the storage controllers was approximately 81%; demonstrating available capacity to handle additional work.

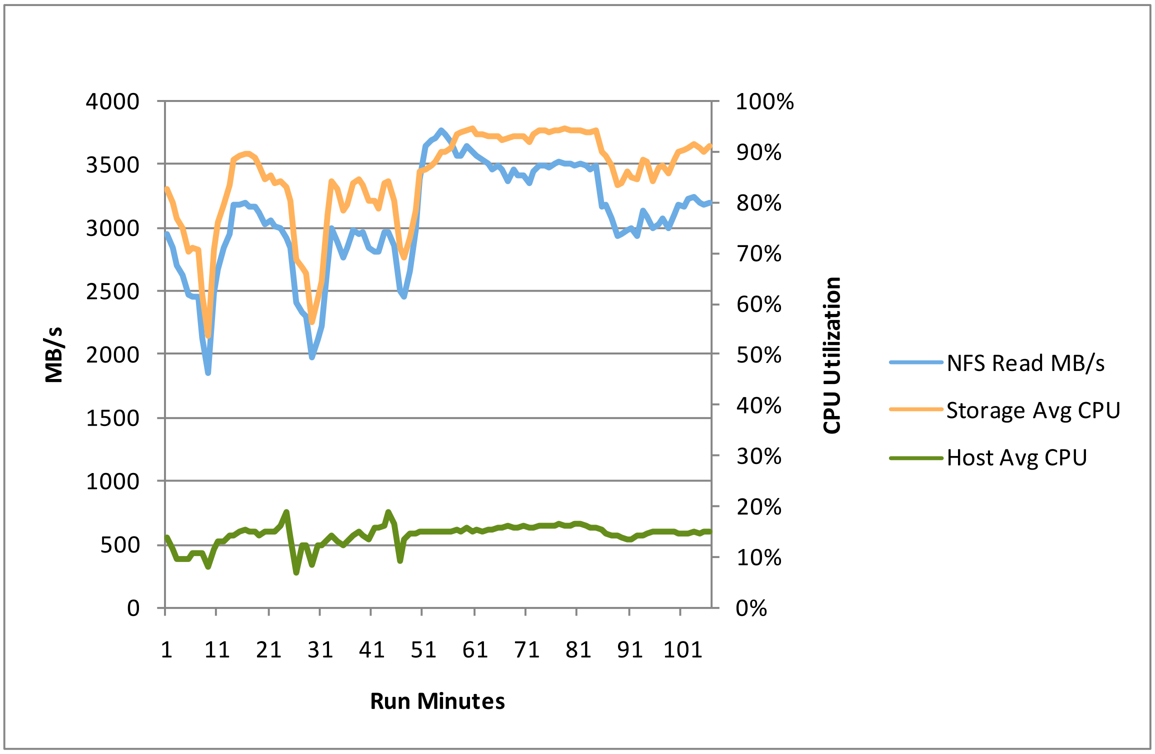
Table 6 shows the Physical Read Bytes per Second value from the Oracle AWR report for each of the four RAC nodes as well as the total throughput calculated for the RAC. The results show the load was once again balanced evenly across the RAC nodes for the Scan workload.

Table 6) AWR physical read MB/s (Scan workload).

trdss1	trdss2	trdss3	trdss4	Total
592MB/s	653MB/s	608MB/s	669MB/s	2522MB/s

The graph in Figure 7 shows the average NFS read throughput as observed on the FAS3170 storage controllers. For the majority of the test, nearly the entire total throughput is serviced by the storage controller disks. This test also demonstrates a sustained period in which the FAS3170 storage controllers are serving over 3.5GB/s of throughput with no errors or other issues observed. The Scan workload did not produce the same level of system load or cache usage as the Throughput workload.

Figure 7) FAS3170 read throughput with storage and host CPU use during Scan workload.



5 CONCLUSION

Cisco UCS was designed using a new and innovative approach to improve data center infrastructure. When combined with NetApp's unified storage, the combined architecture unites compute, network, virtualization, and storage access resources into a scalable, modular architecture that provides enterprise-class performance in an easy-to-manage package.

In addition, the architecture and large memory capabilities of the Cisco UCS connected to the industry-proven and scalable NetApp storage system enable customers to scale and manage Oracle Database environments in ways not previously possible.

Both database administrators and system administrators benefit from the NetApp and Cisco UCS combination of superior architecture, outstanding performance, and unified fabric. They can achieve demonstrated results by following the documented best practices for database installation, configuration, and management outlined in the Cisco UCS and NetApp Solution for Oracle Real Application Clusters (RAC) CVD for the configuration on which this technical report is based.

The workload performance testing included a realistic mix of DSS workloads, which generated sustained loads on the four-node Oracle RAC configuration. These workloads were designed to demonstrate three different types of loads that are common in a DSS/DW environment. These include a Throughput workload that mimics parameter-driven reports, an Analytic workload that demonstrates a load that puts stress on both the storage and server systems, and a simple Scan workload that represents scanning through a large fact table to gain a simple set of facts or aggregates.

The following high-performance metrics were achieved during the challenging workload tests:

- The quad-core Intel® Xeon® 5500 series processors in the UCS B200 M1 blades handled the load without issue. Observed performance was consistent and at an expected level because none of the workloads was designed specifically to stress the CPUs, but, rather, to show how easily they can handle large volumes of data and the demands of a DSS/DW system.
- The I/O demands generated by the load were effectively supported by the capabilities of the balanced NetApp storage array configuration. The system was designed with the goal of eliminating obvious points of performance-limiting bottlenecks.
- While delivering enterprise-level throughput numbers without issue, all subcomponents of the architecture showed sufficient headroom available to address sustained bursts or additional load.

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

This platform enables IT professionals to:

- Optimize database environments
- Reduce total overall cost of the data center
- Provide dynamic resource provisioning for increased business agility

The benefits of the Cisco UCS include:

- Reducing TCO 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 nonvirtualized 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

6 APPENDIXES

6.1 KERNEL SETTINGS

```
# 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 maxmimum size of a mesage queue
kernel.msgmax = 65536

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

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

# Changes to support Oracle Database 11gR2 RAC and tuning
fs.aio-max-nr = 1048576
fs.file-max = 6815744
# semaphores: semmsl, semmns, semopm, semmni
kernel.sem = 250 32000 100 128
net.ipv4.ip_local_port_range = 9000 65500
#net.core.rmem_default=262144
#net.core.rmem_max=4194304
#net.core.wmem_default=262144
#net.core.wmem_max=1048586
#sunrpc.tcp_slot_table_entries=128
sunrpc.tcp_slot_table_entries = 128
net.core.rmem_default = 1342177
net.core.rmem_max = 16777216
net.core.wmem_default = 1342177
net.core.wmem_max = 16777216
net.ipv4.tcp_rmem = 4096 1342177 16777216
net.ipv4.tcp_wmem = 4096 1342177 16777216
net.ipv4.tcp_window_scaling = 1
net.ipv4.tcp_syncookies = 0
net.ipv4.tcp_sack = 0
net.ipv4.tcp_dsack = 0
net.core.netdev_max_backlog = 300000
```

6.2 INIT.ORA ADJUSTED SETTINGS

```
NAME                                 TYPE        VALUE
------------------------------------ ----------- ----------------------
parallel_degree_policy               string      LIMITED
parallel_max_servers                 integer     160
parallel_servers_target              integer     64
parallel_threads_per_cpu             integer     2
pga_aggregate_target                 big integer 0
resource_manager_plan                string      FORCE:INTERNAL_PLAN
sga_max_size                         big integer 19328M
sga_target                           big integer 0
sort_area_size                       integer     10485760
```

6.3 NFS MOUNT SETTINGS

```
fas1a1:/vol/ORA_HOME/orahome        /u1/app/oracle   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp     0 0
fas1a1:/vol/OCR_CSS/ocr_css        /u1/app/11.2.0/ocr   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp  0 0
fas1a1:/vol/VOL_DATA_1_A_1/oradata        /u1/oradata   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp  0 0
fas1a1:/vol/VOL_LOG_1_A_1/oralog        /u1/oralog   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp     0 0
fas1b1:/vol/VOL_DATA_2_A_1/oradata        /u2/oradata   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp  0 0
fas1b1:/vol/VOL_LOG_2_A_1/oralog        /u2/oralog   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp     0 0
fas1b1:/vol/VOL_FRA_2_A_1/orafra        /u2/orafra   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp     0 0
fas1b1:/vol/VOL_ARC_2_A_1/oraarch        /u2/oraarch   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp   0 0
fas2a1:/vol/VOL_DATA_1_B_1/oradata        /u3/oradata   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp  0 0
fas2a1:/vol/VOL_LOG_1_B_1/oralog        /u3/oralog   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp     0 0
fas2b1:/vol/VOL_DATA_2_B_1/oradata        /u4/oradata   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp  0 0
fas2b1:/vol/VOL_LOG_2_B_1/oralog        /u4/oralog   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp     0 0
fas2b1:/vol/VOL_ARC_2_B_1/oraarch        /u4/oraarch   nfs
rw,bg,hard,rsize=65536,wsize=65536,vers=3,actimeo=0,nointr,timeo=600,suid,tcp   0 0
```

7 REFERENCES

7.1 CISCO REFERENCES

- Cisco DCNM Fundamentals Configuration Guide, Release 4.1
  www.cisco.com/en/US/docs/switches/datacenter/sw/4_1/dcnm/fundamentals/configuration/guide/fund_overview.html
- Cisco MDS 9000 Software
  www.cisco.com/en/US/products/hw/ps4159/index.html
- Cisco Nexus 5000 Series Switches
  www.cisco.com/en/US/products/ps9670/index.html

- Cisco Nexus 7000 Series Switches
  www.cisco.com/en/US/products/ps9402/index.html
- Cisco SAFE Reference Guide
  www.cisco.com/en/US/docs/solutions/Enterprise/Security/SAFE_RG/chap4.html
- Cisco UCS and NetApp Solution for Oracle Real Application Clusters (RAC)
  www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/App_Networking/UCS_Oracle_NetApp.html
- Cisco UCS B-Series Servers Documentation Roadmap
  www.cisco.com/en/US/docs/unified_computing/ucs/overview/guide/UCS_roadmap.html
- Cisco UCS Manager GUI Configuration Guide
  www.cisco.com/en/US/docs/unified_computing/ucs/sw/gui/config/guide/1.4/b_UCSM_GUI_Configuration_Guide_1_4.html
- Cisco Unified Computing System
  www.cisco.com/en/US/netsol/ns944/index.html
- Data Center Design—IP Network Infrastructure
  www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DC_3_0/DC-3_0_IPInfra.html

7.2 NETAPP REFERENCES

- Data ONTAP® 7G
  www.netapp.com/us/products/platform-os/data-ontap/
- netapp.com
  www.netapp.com/us/
- Products
  www.netapp.com/us/products
- Snapshot
  www.netapp.com/us/products/platform-os/snapshot.html
- TR-3633: NetApp Best Practice Guidelines for Oracle Database 11*g*
  https://fieldportal.netapp.com/search/?qsb=1&query=tr-3633&within=&thumb=on&dscp=on
- TR-3760: Building a Scalable Data Warehouse Solution
  http://media.netapp.com/documents/tr-3760.pdf

NetApp provides no representations or warranties regarding the accuracy, reliability, or serviceability of any information or recommendations provided in this publication, or with respect to any results that may be obtained by the use of the information or observance of any recommendations provided herein. The information in this document is distributed AS IS, and the use of this information or the implementation of any recommendations or techniques herein is a customer's responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. This document and the information contained herein may be used solely in connection with the NetApp products discussed in this document.

Go further, faster®

www.netapp.com

© 2011 NetApp, Inc. All rights reserved. No portions of this document may be reproduced without prior written consent of NetApp, Inc. Specifications are subject to change without notice. NetApp, the NetApp logo, Go further, faster, Data ONTAP, RAID-DP, and Snapshot are trademarks or registered trademarks of NetApp, Inc. in the United States and/or other countries. Cisco and Cisco Nexus are registered trademarks and Unified Computing System is a trademark of Cisco Systems. Intel and Xeon are registered trademarks of Intel Corporation. Linux is a registered trademark of Linus Torvalds. Oracle is a registered trademark and Oracle11*g* is a trademark of Oracle Corporation. All other brands or products are trademarks or registered trademarks of their respective holders and should be treated as such. TR-3926-0511