WHITE PAPER

Virtualizing Oracle Database 10g/11g on VMware Infrastructure

Consolidation Solutions with VMware® Infrastructure 3 and EMC Celerra NS40 Multi-Protocol Storage

May 2009

Contents

Executive Overview

This document describes the configuration of three different solutions for virtualizing Oracle database servers on VMware® Infrastructure and EMC Celerra storage. These solutions enable enterprises to take a number of their departmental databases and consolidate them on to a smaller number of servers, thereby reducing administrative requirements, simplifying end-user access and control, and reducing overall system costs associated with complex Oracle environments. Once virtualized, Oracle databases can reap all the benefits of a virtualized platform. These benefits include rapid database server provisioning, simplified high availability and disaster recovery, reduced server hardware requirements and associated costs such as power, cooling, and rack space.

Introduction

This document is intended to provide customers with technical solution and configuration information that can be used as the basis for virtualizing Oracle database servers on the VMware Infrastructure platform. The specific solutions described in this document are the following:

1. Oracle Database on VMware Infrastructure using a pure NFS storage design (on a single VMware® ESX host).
2. Oracle Database on VMware Infrastructure using a pure NFS storage design with a four-node VMware® High Availability (HA)/VMware® Distributed Resource Scheduler (DRS) cluster solution (using multiple ESX servers).
3. Oracle Database on VMware Infrastructure using a blended FCP/NFS storage design (on a single ESX server).

Details are provided for the ESX server and virtual machine configurations as well as Celerra disk storage layouts for the three different configurations. Testing was also done to demonstrate how each of these configurations can provide capabilities for backup, disaster recovery and test/dev. Readers should have a thorough understanding of storage, virtualization and Oracle database concepts to get the most value from this document.

All solution design and testing was done at EMC labs in Raleigh, North Carolina in conjunction with VMware.

Related Documents and Resources

Documents listed here provide additional information relevant to the topics described in this document.

Commercial Publications

- Scalzo, Bert. *Oracle on VMware: Expert Tips for Database Virtualization.* Kittrell, NC: Rampant Techpress, 2008.

VMware Resources

The VMware website for Oracle virtualization (www.vmware.com/oracle) contains a number of different informational resources for customers interested in virtualizing Oracle database applications. These resources include customer case studies, demos, whitepapers and performance studies:

- Whitepaper – *Oracle Database Scalability on VMware ESX*

  http://www.vmware.com/resources/techresources/1055

- Whitepaper – *Simplify Oracle Database Management with VMware Infrastructure 3 and EMC CLARiiON Storage*:

  http://vmware.com/partners/alliances/technology/oracle-database-whitepapers.html

- Whitepaper – *Eliminating Oracle Planned Downtime with VMware® VMotion:*

  http://vmware.com/partners/alliances/technology/oracle-database-whitepapers.html

- Webcast – *Oracle Database Performance on VMware Infrastructure:*

  http://vmware.com/a/webcasts/details/161

- VMware Blog Post – *Performance study*: *100,000 Oracle I/O Operations Per Second, One ESX Host:*

  http://blogs.vmware.com/performance/2008/05/100000-io-opera.html

- VMware Blog Post – *Ten Reasons Why Oracle Databases Run Best on VMware Infrastructure:*

  http://blogs.vmware.com/performance/2007/11/ten-reasons-why.html

- Customer Success Stories:

  http://vmware.com/partners/alliances/technology/oracle-database-customers.html

EMC Publications

The following technical papers are available on the EMC.com and EMC Powerlink websites:

- Reference Architecture: EMC Solutions for Oracle Database 10g/11g for Midsize Enterprises—Virtualized Solutions EMC Celerra NS40 Unified Storage Platform
- White Paper: EMC Solutions for Oracle Database 10g/11g for Midsize Enterprises—EMC Celerra Unified Storage Platform - Applied Technology Guide
- White Paper: EMC Solutions for Oracle Database 10g/11g for Midsize Enterprises —EMC Celerra Unified Storage Platform - Best Practices Planning

Note that access to these documents is based on your login credentials. If you do not have access, contact your EMC representative.

Oracle Support

Oracle provides support on using VMware Infrastructure as described in the Oracle Metalink document, Number 249212.1.

EMC supported configurations for deploying Oracle software (and associated operating systems) on EMC storage hardware and software can be found within EMC's eLab Navigator, available through the EMC PowerLink website.

Business Challenges

Enterprises today face a number of challenges when it comes to managing Oracle database environments. These include:

- Increased end-user demand for on-demand, always-on access to databases and analytics.
- Requirements to support a fragmented environment consisting of multiple departmental servers running different versions of Oracle database and operating systems.
- Rising administrative costs to support this heterogeneous database environment.
- Rising data center costs (power, cooling, floor space, etc.) due to server sprawl.
- Under-utilization of server computing resources.
- Lack of sufficient IT resources to deploy, manage, and maintain complex Oracle database environments at the departmental level.
- Requirements for a simple and affordable consolidation solution of Oracle database servers.

Customers looking to reduce the cost and complexity of their Oracle database server environment are increasingly looking to server virtualization to address the challenges listed above. Oracle databases are mission critical for most organizations which makes designing a consolidated solution that reduces costs while increasing overall availability very challenging.

Virtualization Technology Solution

Pertaining to Oracle virtualization, the solutions described in this document address a number of the challenges listed above:

- For organizations that need to support different departments, each running different versions of database and application software, virtual machines provide an ideal way to maintain isolation of different configurations by deploying each configuration in its own independent virtual machine. These independent virtual machines can then be consolidated on fewer host servers for cost efficiency, while maintaining complete isolation from each other.
- When Oracle database servers are consolidated on VMware Infrastructure, customers can immediately obtain benefits from the high availability features provided by the VMware platform. For enterprise Oracle deployments, VMware high availability features such as VMware VMotion and VMware HA can provide sufficient levels of availability at a fraction of the cost and complexity of traditional cluster solutions such as Oracle RAC.
- Consolidating Oracle database servers using VMware virtual machines can reduce server sprawl and reduce infrastructure costs. The configurations presented in this document demonstrate solutions for consolidation using a single ESX server and how this can scale up to multiple ESX servers for larger environments.
- Running multiple Oracle database virtual machines on the same physical servers can increase the CPU and memory utilization of servers, from what is commonly less than 10%, to upwards of 65% or more, delivering an improved ROI on server hardware capital investments.

- Consolidated virtual infrastructure results in improved server-to-admin ratios. Tasks such as database server provisioning, migrating applications onto newer hardware, and server hardware maintenance can be reduced from days/hours on physical servers to a matter of minutes on virtual systems.

Consolidation with VMware Infrastructure versus Oracle RAC

For enterprise customers evaluating solutions for Oracle database consolidation, the typical recommendation presented by Oracle has been to move many databases into a single, multi-node Oracle RAC implementation. This option works and is well-understood—both the benefits of that effort, as well as the real license and implementation costs associated with this means of database consolidation. However, the process required to move to an Oracle RAC implementation can be very time consuming and complicated, especially in environments that contain large numbers of heterogeneous Oracle database servers (operating system, database versions, applications). As an alternative to the solution using Oracle RAC, consider the approach using VMware virtualization. The VMware approach differs from RAC, and is not an apples-to-apples comparison, but can achieve most of the goals of a RAC solution, with substantial cost and time-to-deployment benefits that are outlined below.

Simplified High Availability

A key issue with consolidation is database availability. With consolidation to a single instance (RAC), clearly uptime is critical. Any offline condition would impact all users. RAC solves this problem with continuous availability and is well-suited for that condition. In the VMware approach, each database remains independent (multiple instances) and any one database failure or even host failure (impacting the databases running on that machine) will impact a limited number of the total user base. With VMware HA, the databases impacted by a server failure are brought back online within minutes, automatically restarted on another ESX server using VMware HA.

Table 1 compares the high availability approach used by VMware Infrastructure to a solution using Oracle RAC.

Table 1. VMware High Availability and Oracle RAC Approaches

Feature/Functionality	RAC	VMware HA Cluster
Failover	Real-time HA w/ continuous database uptime (some loss of connectivity may occur)	Transparent failover for planned downtime using VMotion, but unplanned hardware failure requires reboot (guaranteed loss of connectivity while virtual machine reboots)
Data visibility	Scale-up single database image (e.g. , one monolithic application)	Scale-out many single database instances (e.g., software as a service or database cloud)

The VMware approach provides high availability protection for both planned and un-planned downtime using VMware VMotion and VMware HA.

- **Unplanned downtime due to server hardware failure:** In the event of server hardware failure, VMware HA will restart Oracle virtual machines on a surviving ESX host server. Downtime is encountered as the virtual machines restarts, which is typically measured in minutes in most environments.

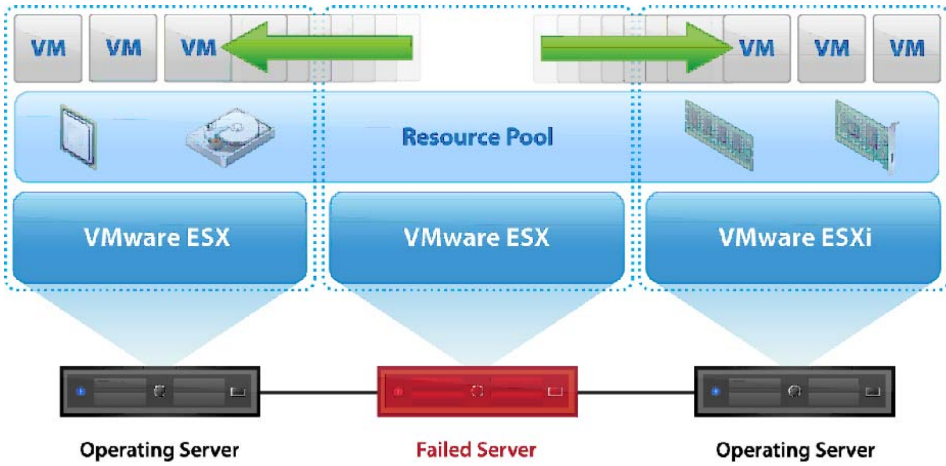

Figure 1. Protection from server hardware failure with VMware HA

- **Planned downtime:** During planned downtime, Oracle database virtual machines can be migrated online at any time to another ESX server, with no loss of service, using VMware vMotion. VMotion can be especially useful when migrating Oracle databases onto newer hardware during server refresh cycles, in hardware troubleshooting scenarios, and managing changes in hardware maintenance windows. All of these can be accomplished with no downtime using VMotion.

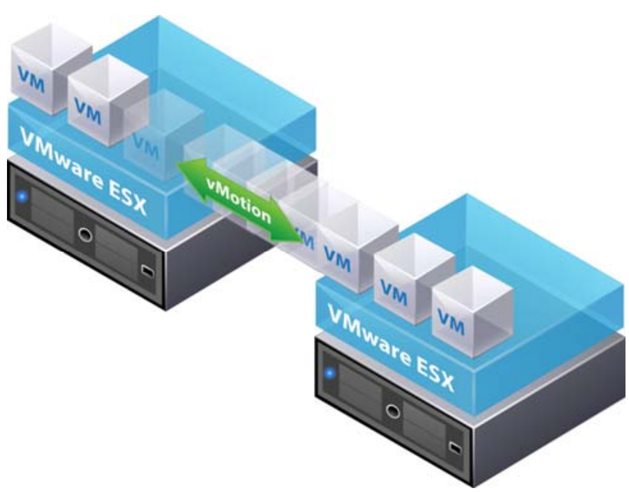

Figure 2. Using VMware VMotion to avoid planned downtime

Most departmental databases can tolerate the minimal downtime associated with VMware HA and a reboot of virtual machines. However, it is important that Oracle database administrators understand the trade-offs between both approaches (Oracle RAC and VMware HA/vMotion) and choose a solution that meets their SLAs. The configurations presented later in this document have all been designed and tested using VMware Infrastructure to provide high availability.

Simplified Project Consolidation

Many organizations currently find themselves supporting a wide range of departmental Oracle database servers, running a wide range of Oracle database versions (everything from 8i to 11g) on a wide range of operating systems (multiple different versions of Windows and Linux as well as Solaris-x86). The approach with Oracle RAC means that all of these database servers (and the applications they support) need to first be upgraded to run on a single version of Oracle database on a single operating system as part of the RAC implementation. The time, cost, complexity, and risk associated with migrating all of these independent servers can be a major barrier to successful, cost-effective consolidation.

Using the VMware approach to Oracle database consolidation allows each database server to remain on its current version of Oracle database and its current operating system. There are no database or operating system migrations to worry about and virtualizing database servers can be as simple as using the free VMware vCenter Converter[1] product to convert a system from a physical machine to a virtual machine running on VMware Infrastructure. Downtime is minimized and each department can continue to maintain its own independent operating system and database instance. Each database can be managed individually in terms of backup/recovery, disaster recovery, patches and upgrades, and test/dev. Figure 3 shows a single VMware ESX server running four Oracle database servers, each running its own version of Oracle database and its own operating system.

[1] For supported operating systems with VMware vCenter Converter, please see http://www.vmware.com/products/converter/

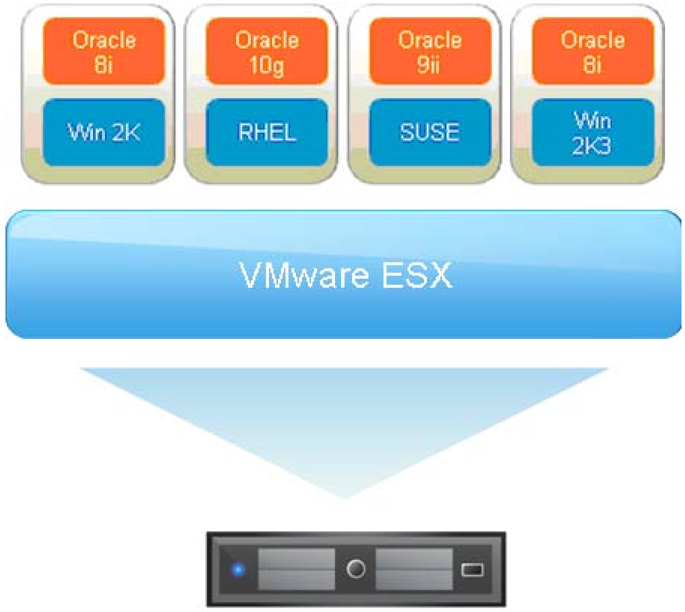

Figure 3. Consolidating heterogeneous Oracle database servers on VMware Infrastructure

Reduced Oracle Licensing Costs

The VMware Infrastructure approach shown in Figure 3, where each Oracle database instance is deployed in its own virtual machine, allows organizations to use the Oracle licenses they already own for each database server. The Oracle RAC approach for consolidation requires upgrading to Oracle RAC Enterprise Edition licenses. This can substantially increase the overall cost of the solution and should be carefully weighed when considering a solution for database consolidation.

Improved Performance

Using the VMware HA cluster approach can actually improve overall Oracle database performance in terms of transactions-per-second (TPS) for a given number of users, while decreasing the overall software license costs per TPS. Figure 4 shows results of performance testing done using an industry standard OLTP workload (Quest Benchmark factory) comparing a 4-node RAC cluster to a 4-node VMware HA cluster[2].

---

[2] 4-node cluster w/ Dell PE2900 servers, each with 2 x 2.66 GHz quad-core Intel Xeon x86-64, 24 GB RAM per node (96 GB total). Eight virtual machines on VMware HA Cluster versus four instances on RAC physical (OLTP workload using Quest BMF).

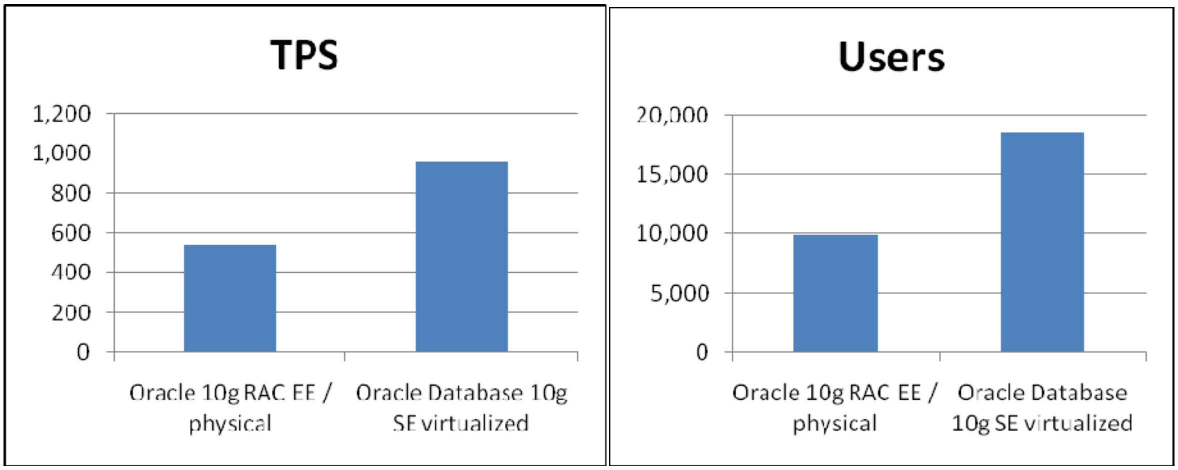

Figure 4. Typical Database Performance (TPS and Users)

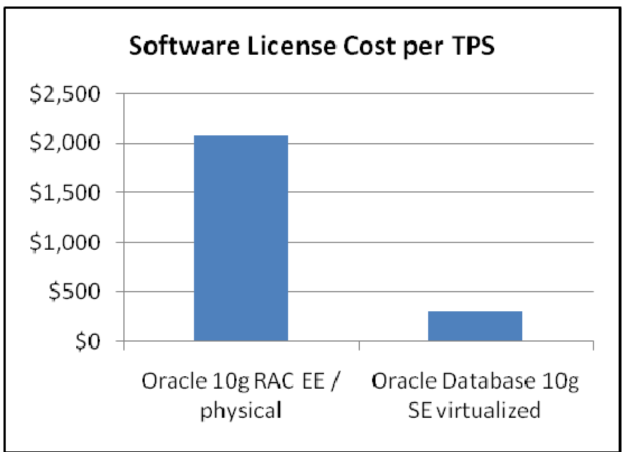

Figure 5. Typical License Costs Per TPS

Running multiple, independent Oracle database virtual machines delivers additional efficiencies over a similar RAC configuration. Overheads associated with RAC in areas such as cache fusion and block pinging are not encountered when using the VMware approach. Additionally, the kernel parameters in /etc/sysctl.conf create hard limits per OS image in a RAC implementation, while the VMware approach creates multiple copies of these limits.

Solution Components

All the solutions described in this document use a platform built using VMware Infrastructure, Oracle Database 10g/11g, and EMC Celerra NS Series Storage.

VMware Infrastructure

In the configurations described in this document, virtualization of all database servers is implemented using the VMware Infrastructure platform.

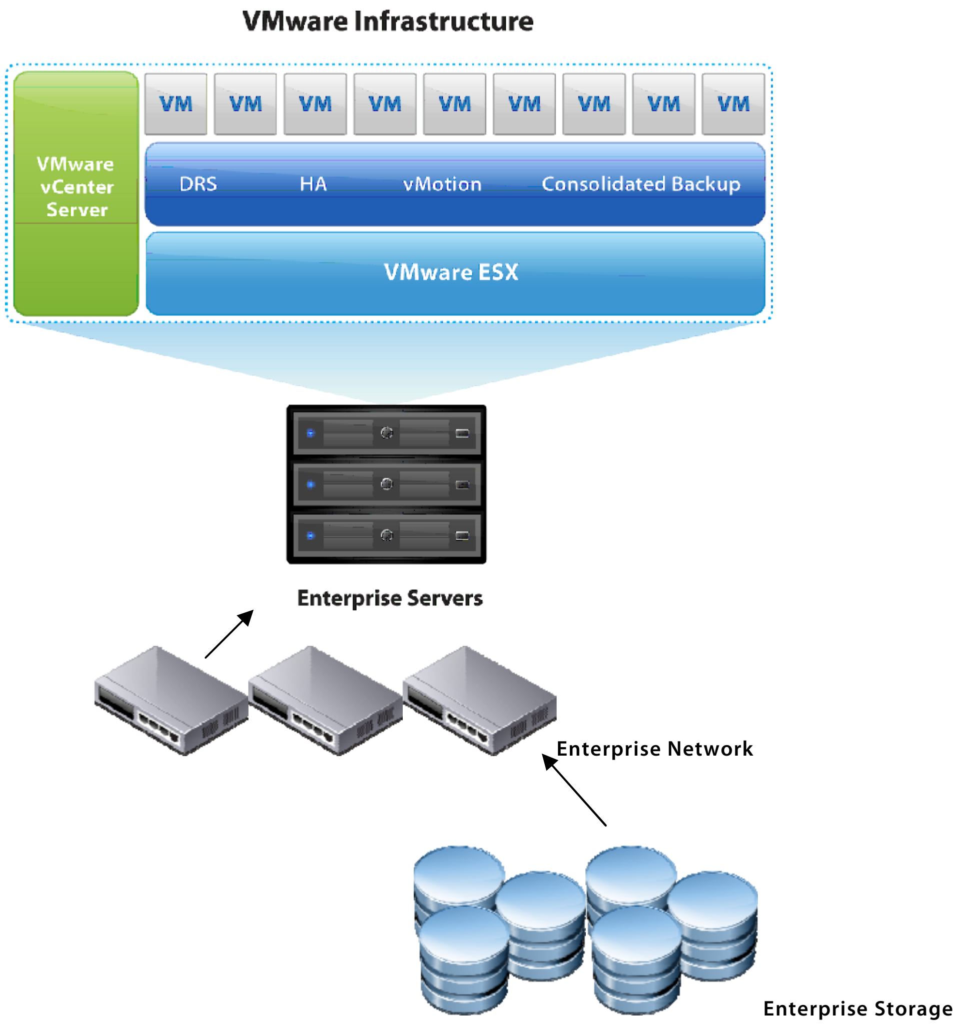

Figure 6. VMware Infrastructure Components

Figure 6 shows the typical components used in a VMware Infrastructure design. VMware Infrastructure is the most widely deployed production virtualization and management platform in the industry. VMware Infrastructure includes the following:

- A high performance hypervisor.

- A distributed file system optimized for virtual machines (VMFS).
- A multi-virtual CPU sub-system (VSMP) that allows vertical scaling of loads as well as the higher level functionality that use the above as a base.

Additional VMware Infrastructure features include:

- VMware® vCenter Management server – provides a management interface to all ESX hosts and virtual machines.
- VMware VMotion and DRS – allow dynamic re-balancing of virtual machine loads across clusters of ESX servers without human intervention.
- VMware High Availability (HA) – provides the capability of automatically re-starting virtual machines and applications after a hardware failure in one ESX host.

These and several other technical features provide a robust virtualization platform for enterprise software like Oracle databases that is in use at many customer sites today.

In the configurations described in this document for virtualizing Oracle database on VMware Infrastructure, the following components were used in the lab for testing:

- **Enterprise Storage:**
  - EMC Celerra NS-series and EMC CLARiiON CX series storage arrays
- **Enterprise Network:**
  - Enterprise class network (1 GB) and Fibre Channel storage (4GB) switches
- **Enterprise Servers:**
  - Enterprise class x86 servers
    - Dual socket quad-core at 2.66 GHz
    - 24GB RAM
- **ESX Server:**
  - All testing was done using ESX Server 3.5.0 (Build 646072)
- **Virtual Machines:**
  - All Oracle database virtual machines were running Oracle Enterprise Linux v5.2.
  - Oracle database 10g/11g Standard Edition.
- **vCenter:**
  - vCenter version 2.5 was deployed in the test lab on a stand-alone system. This system can be deployed as either a stand-alone physical server or in a virtual machine.
- **VMware DRS/VMotion/HA:**
  - VMware HA is used to provide high availability for the Oracle database virtual machines.
  - VMware VMotion was used to migrate live, running database virtual machines across ESX servers.

Oracle Database 10g/11g

Oracle is currently the dominant enterprise-class database software product on the market. Analyst market share studies demonstrate that Oracle has the largest market share percentage of any of the vendors in the same category. As Oracle provides a very reliable, robust, and manageable product, VMware remains committed to creating solutions based upon the Oracle database software stack.

The goal of the solutions presented in this document is to provide comprehensive testing, validation, and documentation of complete environment configurations that

- include the Oracle software stack.
- are enabled by EMC storage hardware, EMC value-added software and VMware Infrastructure.

The Oracle software stack covered by the solutions consists of the following:

- Oracle Enterprise Linux
- Cluster Ready Services (CRS)
- Oracle Database
- Automatic Storage Management (ASM)

EMC Celerra NS Series Storage

Storage hardware used in the solutions described in this document is provided by the EMC® Celerra® NS Series multi-protocol storage array. Celerra includes a Network Attached Storage (NAS) array combined with the functionality and high performance of a Storage Area Network (SAN) array. Celerra provides:

- NAS through the Network File System (NFS) and Common Internet File System (CIFS) protocols
- iSCSI storage through the Celerra's Data Movers
- SAN storage over the Fibre Channel Protocol (FCP) through the back-end EMC CLARiiON® CX3-40f series storage array

This document describes three different approaches for accessing all storage elements in a virtualized Oracle database solution:

1. **Pure NFS** – All of the storage elements are accessed using the NFS protocol.
2. **Pure NFS with VMware High Availability cluster** – All of the storage elements are accessed using the NFS protocol, but incorporating a four-node VMware High Availability (HA) cluster.
3. **Blended FCP/NFS** – The high-demand, low-latency storage elements of Oracle database servers are accessed using Fibre Channel Protocol (FCP) and Oracle ASM. These data elements include data files, online redo log files, control and temp files. All other storage elements are accessed using the NFS protocol. These include flashback recovery area, archive logs, disk-based backups, and CRS files.

Solution Architecture

The next sections describe the configuration of VMware Infrastructure, EMC Celerra NS Series storage and Oracle database 10g/11g components for the three solutions presented in this document:

- **Configuration 1** – Oracle on VMware Infrastructure using a pure NFS storage design (single ESX server).
- **Configuration 2** – Oracle on VMware Infrastructure using a pure NFS storage design with a four-node VMware HA/DRS cluster solution (multiple ESX servers).
- **Configuration 3** – Oracle on VMware Infrastructure using a blended FCP/NFS storage design single ESX server).

Each section also describes specific features that were tested to provide advanced functionality for the Oracle database virtual machines. All configurations were tested for performance and functionality at EMC labs - RTP in Raleigh, North Carolina in conjunction with VMware.

Configuration 1 – Oracle Database on VMware Infrastructure Using a Pure NFS Storage Design

The first configuration is illustrated in Figure 7 below and represents a basic consolidation scenario of four Oracle database servers on a single ESX host and the associated storage layout. Primary storage is hosted on an EMC Celerra array using NFS for all Oracle database components.

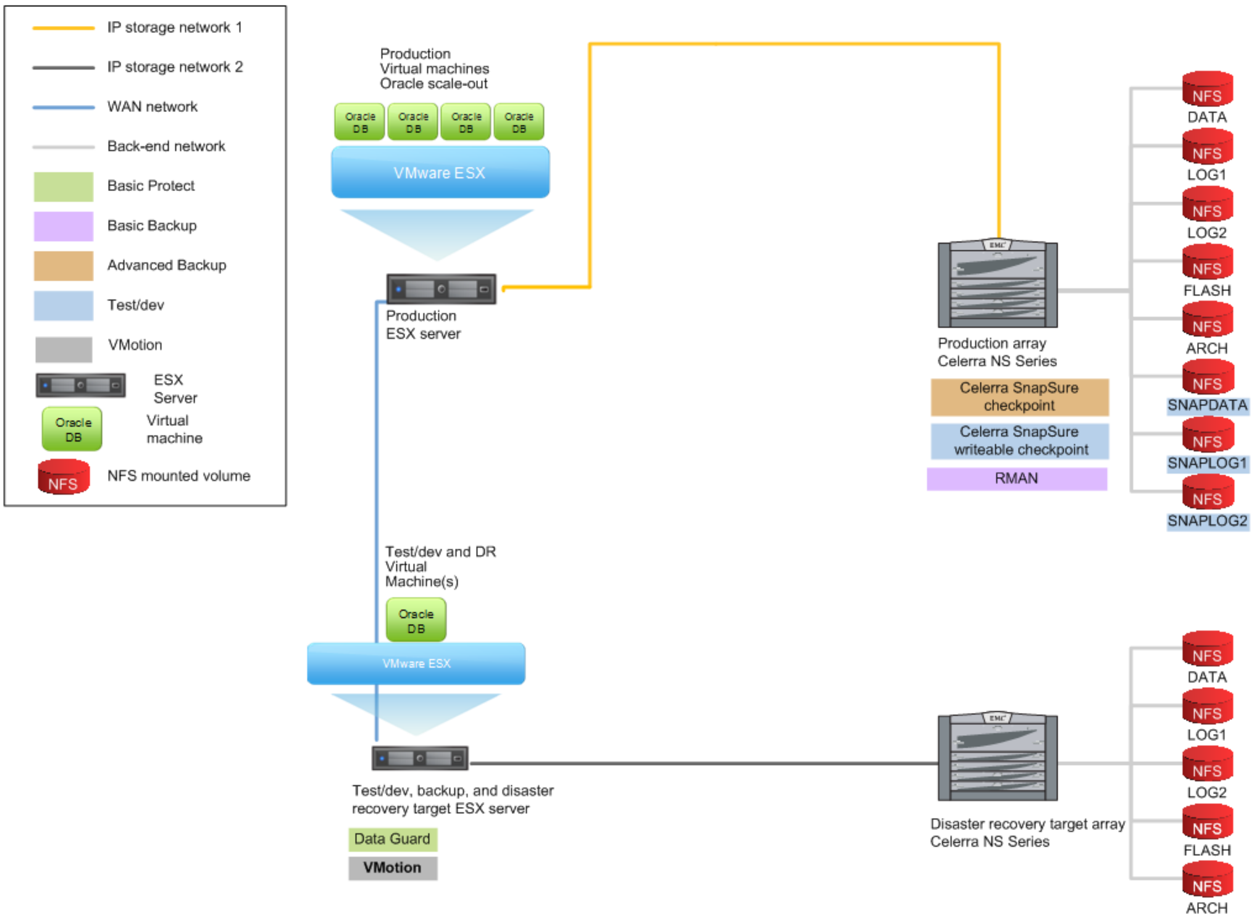

Figure 7. Oracle Database on VMware Infrastructure Using a Pure NFS Storage Design

Table 2 describes the solution features that have been validated for the pure NFS virtualized solution. For a detailed description of the solution features listed, see Appendix A.

Table 2. Pure NFS Validated Solution Features

Solution Feature	Details
Scale-Out OLTP	Performance was tested in EMC labs using an industry-standard OLTP database performance benchmark.
Basic Backup and Recovery	Oracle Recovery Manager (RMAN) provides Basic Backup to the primary Celerra storage array.
Advanced Backup and Recovery	EMC Celerra SnapSure™ checkpoints were tested for advanced backup functionality.
Basic Protect	Oracle Recovery Manager (RMAN) to seed DR solution. Oracle Data Guard to standby virtual machine and secondary Celerra array.
Resiliency	Every significant layer of the solution has been tested by introducing faults.
Test/dev	EMC Celerra SnapSure writeable checkpoints to a virtualized single-instance target.
VMotion	VMware VMotion was used to move live Oracle virtual machines from primary ESX host to secondary ESX host while subjected to load testing.

VMware Architecture

This configuration presents a single ESX server for basic database consolidation. The physical ESX server was a standard 2U x86 server with eight processor cores and 24 GB RAM. Performance and functionality testing was done using a total of four Oracle database virtual machines running on the ESX server.

Each virtual machine was allocated 7 GB of memory and two virtual CPUs. While initially this required 28GB of RAM to be allocated to the virtual machines (four virtual machines with 7GB RAM each), the ESX transparent page sharing mechanism was able to eliminate common memory pages on the ESX server and total memory consumption was reduced to 24GB within minutes of powering up all four virtual machines.

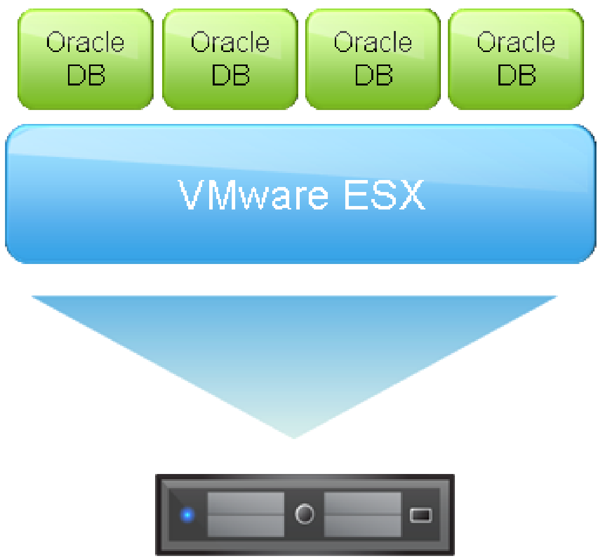

Figure 8. Virtual Machine Configuration on ESX Server

The specific configuration of the ESX server is the following:

- The ESX server is running ESX 3.5 (Build 646072).
- The ESX server is configured with 8 CPUs (2 x quad-core) and 24GB RAM.

The configuration of virtual machines is the following:

- Four database server virtual machines are created on a single ESX server.
- Each virtual machine is allocated 7 GB of memory and 2 vCPUs.
- Oracle Database 10g or 11g for x86 64-bit (SMP Kernel) is run on Oracle Enterprise Linux in all the virtual machines shown in Figure 8.

Storage Architecture

Figure 5 shows how all storage components have been deployed on NFS storage in an EMC Celerra NS-series array:

- Oracle database files and online redo log files reside on their own NFS file system. Online redo log files are mirrored across two different file systems using Oracle software multiplexing. Three NFS file systems are used, one for datafiles and temp files, and two for online redo log files.
- Oracle control files are mirrored across the online redo log file NFS file systems.
- RAID-protected NFS file systems are designed to satisfy the I/O demands of individual database objects. For example, RAID 5 can be used for the datafiles and temp files, but RAID 1 is always used for the online redo logfiles. (See the Pure NFS RAID and Blended RAID configurations described in more detail in Appendixes B and C.)
- The NFS file systems that are used to store the Oracle datafiles, temp files, online redo logfiles, and control files are stored on FC disks.

Target site configuration:

- At the target site, a separate Celerra is connected to the VMware ESX server through the target storage network. The Oracle Database 10g/11g single-instance target server accesses this network through a virtualized connection.

RAID Group Setup

Two different sets of RAID and disk configurations were tested in this configuration. (See Table 3.) For additional information and details on these RAID group configurations, see Appendix B.

Table 3. Pure NFS solution RAID configurations

Figure	Configuration	Description
Figure 14	Pure NFS configuration 1	1 SATA shelf 3 FC shelf RAID 5/RAID 1 AVM using user-defined storage pools
Figure 15	Pure NFS configuration 2	1 SATA shelf 2 FC shelf RAID 5/RAID 1 AVM using user-defined storage pools

Configuration 2 – Oracle Database on VMware Infrastructure Using a Pure NFS Storage Design with a Four-Node VMware HA/DRS Cluster Solution

This configuration is very similar to the previous one in terms of storage architecture, but now introduces the notion of using a set of ESX servers that all belong to a cluster of ESX hosts. The ESX cluster shows how the previous configuration, which used only a single ESX server for consolidation, can be scaled to multiple ESX hosts to accommodate larger database consolidation projects. This configuration was tested using a total of four ESX servers, each running two Oracle database virtual machines.

The ESX cluster also introduces functionality for providing Oracle database high availability using VMware HA. With VMware HA, in the event of a server hardware failure, all virtual machines running on that ESX server will be automatically restarted on a surviving ESX server in the cluster. This is how the VMware approach can provide a high availability alternative to RAC clustering with less complexity.

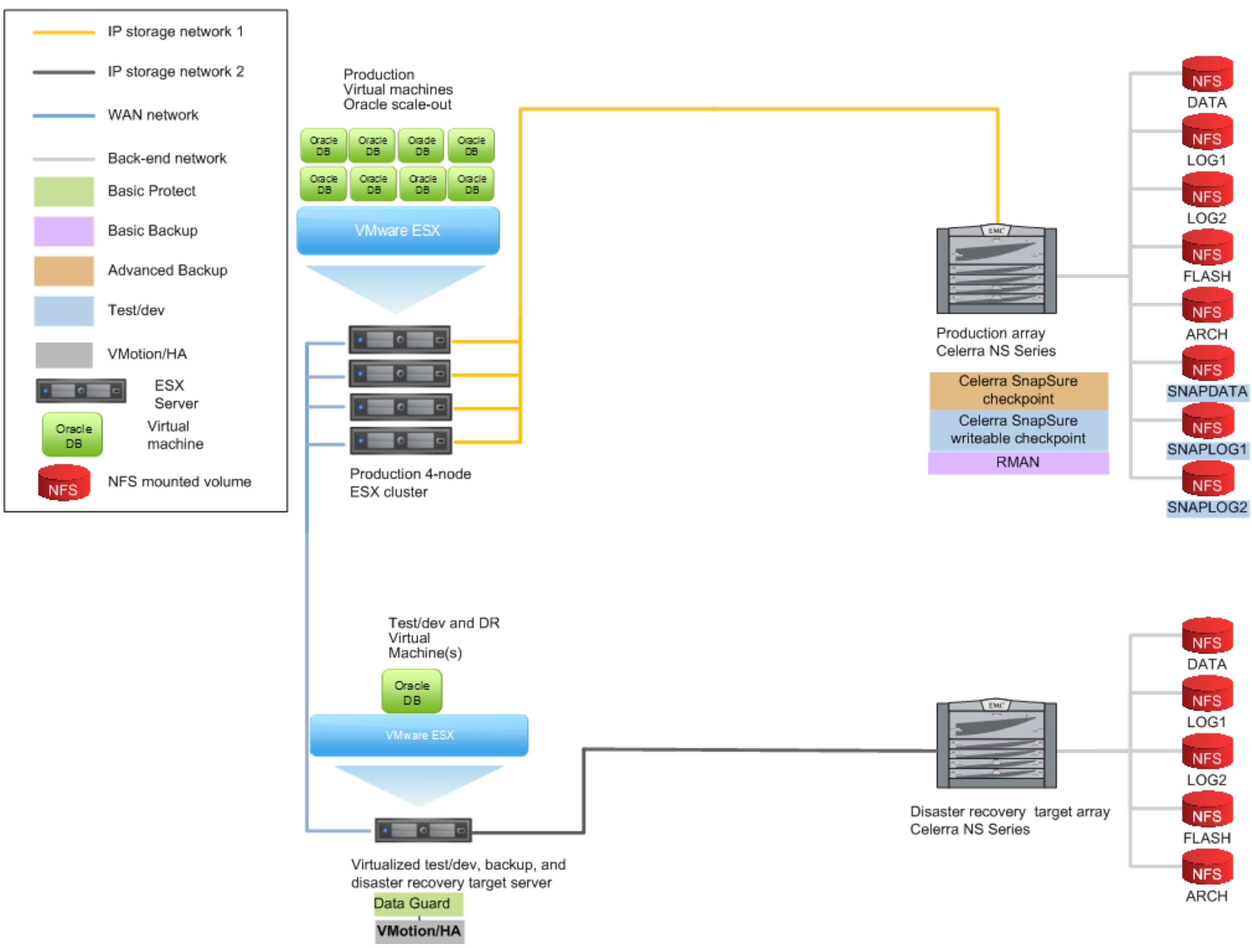

Figure 9. Pure NFS Virtualized Solution with Four-Node VMware HA/DRS Cluster

Table 4 describes the solution features that have been validated for the pure NFS HA cluster solution. For a detailed description of the solution features listed, see Appendix A.

Table 4. Validated Solution Features for Cluster Solution

Solution Features	Details
Scale-Out OLTP	Performance was tested in EMC labs using an industry-standard OLTP database performance benchmark.
Basic Backup and Recovery	Oracle Recovery Manager (RMAN) was used to backup directly to disk on the primary Celerra array.
Advanced Backup and Recovery	EMC Celerra SnapSure™ checkpoints provide instant point-in-time copies for advanced backup and recovery.
Basic Protect	Oracle Recovery Manager (RMAN). Oracle Data Guard.
Resiliency	Every significant layer of the solution has been tested by introducing faults.
Test/dev	EMC Celerra SnapSure writeable checkpoints to a virtualized single-instance target.
VMware VMotion	VMware VMotion was used to move live Oracle virtual machines from primary ESX host to secondary ESX host while subjected to load testing.
VMware HA	VMware HA was tested and validated to provide protection from server hardware failure.

VMware Architecture

This architecture expands on the single ESX server solution presented in the previous section (Configuration 1). An ESX cluster of four nodes is used to provide capacity for additional database consolidation as well as provide high availability.

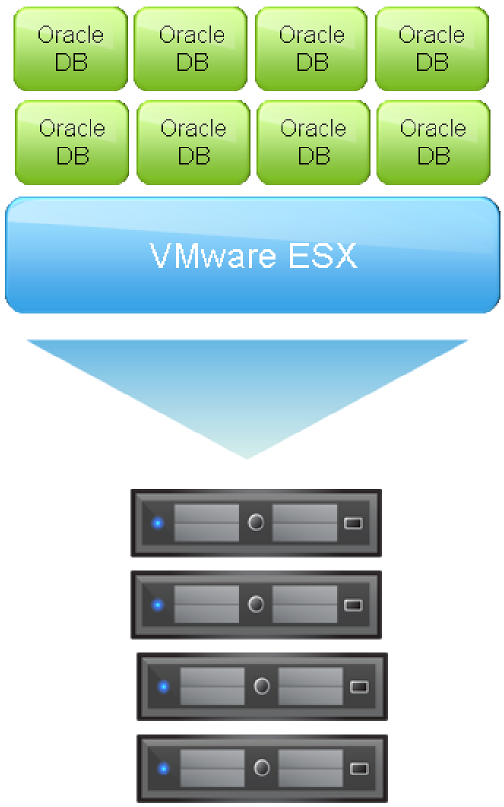

Figure 10. ESX 4-Node Cluster Configuration for Database Consolidation

The specific configuration of the ESX servers is as follows:

- The ESX servers are running ESX 3.5 (Build 646072).
- The ESX servers are configured with 8 CPUs (2 x quad-core) and 24GB RAM.

The configuration of virtual machines is the following:

- Two database server virtual machines are created per ESX server.
- Each virtual machine is allocated 7 GB of memory and 2 vCPUs.
- Oracle Database 10g or 11g for x86 64-bit (SMP Kernel) is run on Oracle Enterprise Linux in all the virtual machines shown in Figure 6.

In this configuration, a total of eight Oracle database virtual machines are consolidated onto four ESX servers. Figure 11 below demonstrates how VMware HA will automatically restart virtual machines on a surviving ESX host in the cluster in the event of server hardware failover. This

functionality was tested and validated in the EMC labs while under load and was shown to restart database virtual machines in approximately 2 minutes. The VMware HA solution for high availability provides protection from server hardware failure for each virtual machine without requiring cluster solutions such as Oracle RAC. For a comparison of VMware HA and RAC functionality, please see Table 1.

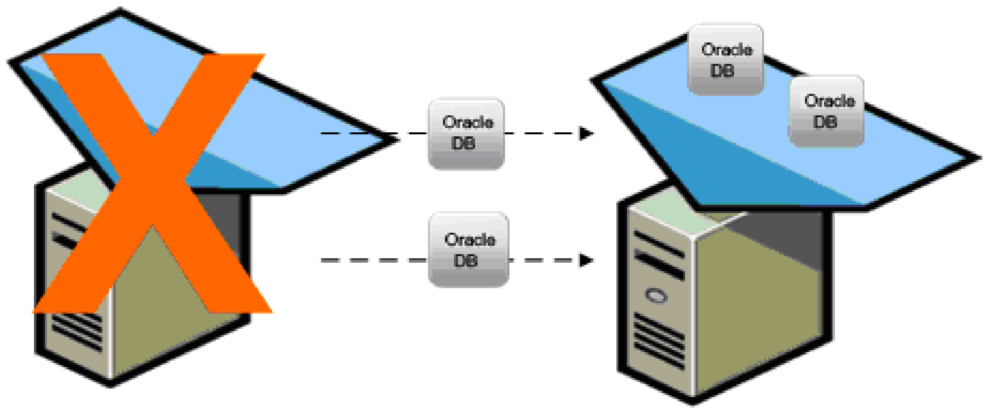

Figure 11. Oracle Database Consolidation on ESX Host Servers

Storage Architecture

The storage architecture for this solution is identical to Configuration 1 where all database components are deployed on NFS storage presented to the ESX server. Additional information and details on this storage design can be found in Appendix B.

Configuration 3 – Oracle Database on VMware Infrastructure Using Blended FCP/NFS Storage Design

For this solution, the server architecture is similar to that of configuration 1, using a single ESX server to illustrate basic consolidation on standard x86 hardware. The storage is configured using a combination of both FCP and NFS. This hybrid design is intended for database environments with high disk I/O requirements and places high I/O components of the Oracle database on a Fibre Channel SAN based on an EMC CLARiiON array, while leaving remaining components on Celerra with NFS.

The data and redo logs reside on a single EMC CLARiiON CX3-40f series array using a 4GB/sec Fibre Channel storage area network, and the archive and flash area reside on a Celerra using NFS. A separate set of LUNs are created on the EMC CLARiiON CX3-40f series array and presented accordingly using RDM to the respective virtual machines. The archive and flash file systems are directly mounted using NFS on the virtual machines.

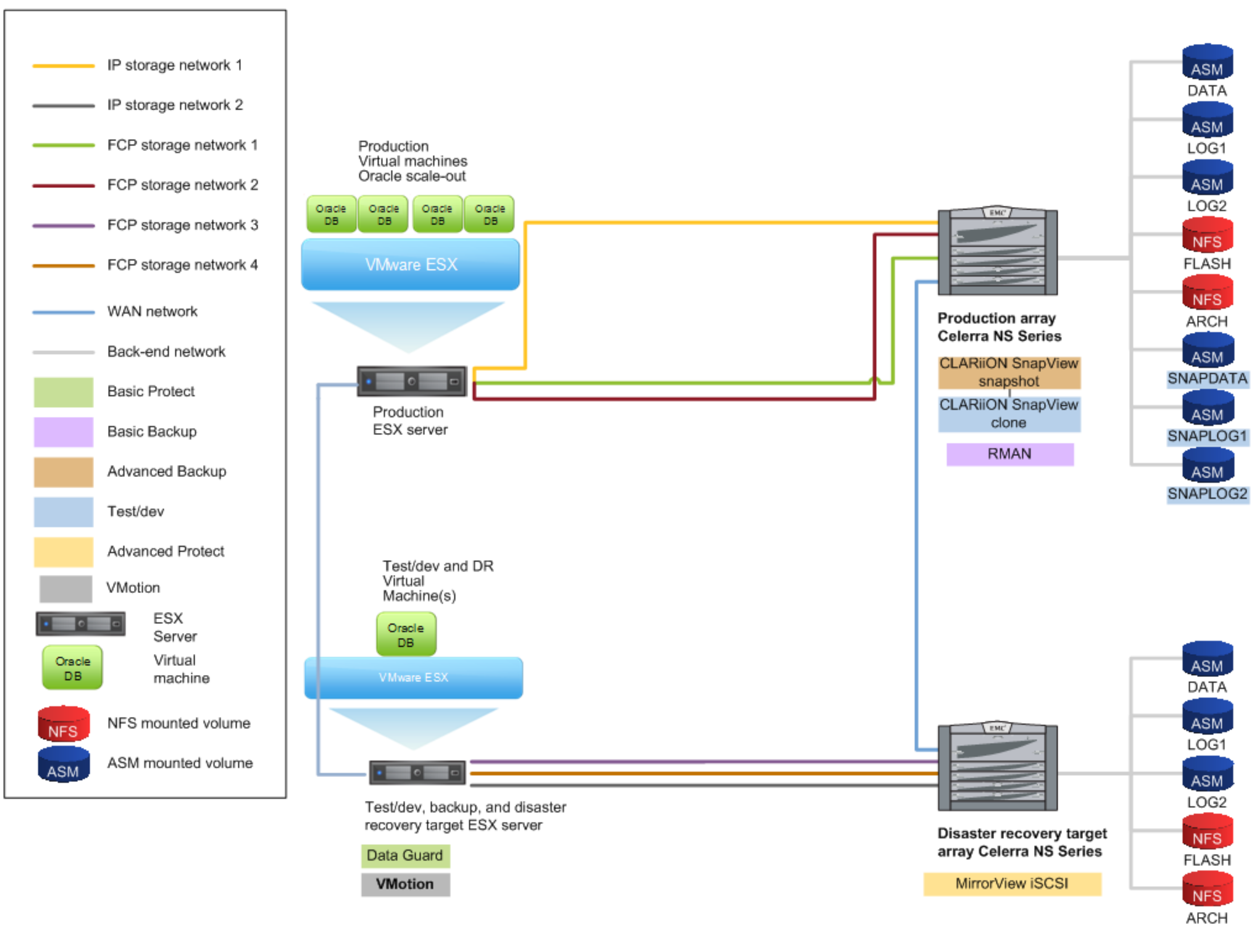

Figure 12. Blended FCP/NFS Virtualized Solution Reference Architecture

Table 5 describes the solution features that have been validated for the blended FCP\NFS solution. For a detailed description of these components, please see Appendix A.

Table 5. Validated Solution Features

Solution Features	Details
Scale-Out OLTP	Utilizes an industry-standard OLTP database performance benchmark.
Basic Backup and Recovery	Oracle Recovery Manager (RMAN).
Advanced Backup and Recovery	EMC CLARiiON SnapView® snapshots.
Basic Protect	Oracle Recovery Manager (RMAN). Oracle Data Guard.
Advanced Protect	MirrorView/A through iSCSI.
Test/dev	EMC CLARiiON SnapView clones.
VMware VMotion	VMware VMotion was used to move live Oracle virtual machines from primary ESX host to secondary ESX host while subjected to load testing.

Figure 8 shows the architecture for the blended FCP/NFS virtualized solution.

Note: For simplicity, Fibre Channel switches are not shown in the diagram. In the labs, two Fibre Channel switches were used in this configuration for connecting the CLARiiON array to the Fibre Channel ESX server.

VMware Architecture

This configuration is similar to configuration 1 and presents a single ESX server for basic database consolidation. The physical ESX server was a standard 2U x86 server with eight processor cores and 24 GB RAM. Performance and functionality testing was done using a total of four Oracle database virtual machines running on the ESX server.

Each virtual machine was allocated 7 GB of memory and two virtual CPUs. While initially this required 28GB of RAM to be allocated to the virtual machines (four virtual machines with 7GB RAM each), the ESX transparent page sharing mechanism was able to eliminate common memory pages on the ESX server and total memory consumption was reduced to 24GB within minutes of powering up all four virtual machines.

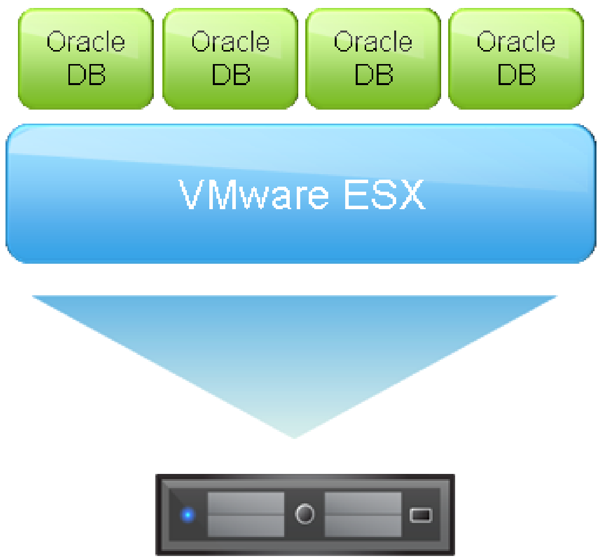

Figure 13. Virtual Machine Configuration on ESX Server

The specific configuration of the ESX server is the following:

- The ESX server is running ESX 3.5 (Build 646072).
- The ESX server is configured with 8 CPUs (2 x quad-core) and 24GB RAM.

The configuration of virtual machines is the following:

- Four database server virtual machines are created on a single ESX server.
- Each virtual machine is allocated 7 GB of memory and 2 vCPUs.
- Oracle Database 10g or 11g for x86 64-bit (SMP Kernel) is run on Oracle Enterprise Linux in all the virtual machines shown in Figure 6.

Storage Architecture

For the blended solution the data and redo logs reside on a single EMC CLARiiON CX3-40f series array, and the archive and flash area reside on a Celerra. A separate set of LUNs are created on the EMC CLARiiON CX3-40f series array and presented accordingly using RDM to the respective virtual machines. The archive and flash file systems are directly mounted using NFS on the virtual machines.

High-performance files (data files, temp files, control files, and online redo logfiles) reside on the EMC CLARiiON CX3-40f series array and are accessed through FCP. Oracle ASM is used to manage these files. All other database files (archive logfiles, backup target, and the flashback recovery area) reside on the Celerra storage array, and are accessed through NFS.

The following are the characteristics of the FCP network used in testing the blended FCP/NFS solution:

- Fully redundant FCP switches.
- Two target ports on each of the CLARiiON CX3-40f storage processors are utilized. Each target port is connected to a separate FCP switch.
- The HBA ports on the ESX server are zoned to all the LUNs on the back-end storage, but only the required LUNs are presented to each virtual machine using RDM.
- Zoning is employed on the FCP switches to ensure that all Oracle hosts have redundant paths to access all LUNs on the CLARiiON CX3-40f array.

Table 6 outlines which storage elements are accessed using each protocol – FCP or NFS.

Table 6. Protocol Access of Different Storage Elements

Accessed using FCP	Accessed using NFS
Datafiles	Flashback recovery area
Online redo log files	Archive logs
Control files	Disk-based backups
Temp files	CRS files

Two sets of RAID and disk configurations have been tested. (See Table 7.)

Table 7. Oracle RAC 10g/11g FCP solution RAID configurations

Figure	Configuration	Description
Figure 16	Blended configuration 1	1 SATA shelf 3 FC shelf RAID 1-0/RAID 1
Figure 17	Blended configuration 2	1 SATA shelf 3 FC shelf RAID 5/RAID 1

For more information and details on RAID group layouts for these configurations, see Appendix C.

Conclusion

Enterprises must get the maximum possible value from their IT infrastructure to stay in business in today's competitive landscape. VMware Infrastructure helps them to do that—particularly, by hosting multiple Oracle databases together on the VMware Infrastructure 3 platform. This provides the following benefits:

- Reduced costs through server consolidation that uses fewer servers
- Higher availability through automatic restart of critical servers
- Better load balancing
- No downtime for maintenance through the use of VMotion to evacuate servers when fixes are required
- Easier planning and implementation of disaster recovery

This paper provided details on three different designs of the server and storage layout for a virtualized set of Oracle databases. Each configuration was also tested with solutions for Basic and Advanced backup, as well as disaster recovery and test/dev functionality using various EMC, VMware and Oracle data protection and high availability tools. This testing should give customers additional confidence to virtualize their own Oracle database environments and also enable DBAs to see exactly what kind of technical choices they can make to optimize the use of their IT infrastructure. For most, if not all, of the enterprise Oracle database implementations that currently exist, there should be no obstacles to running those systems on VMware Infrastructure and achieving the benefits that come with it.

Appendix A. Oracle Solution Features and Capabilities

Table 8. Oracle Solution Features and Capabilities

Solution Features	Description
Scale-Out OLTP	Real-world performance and capacity testing. Utilizes an industry-standard OLTP database performance benchmark, while providing only real-world tuning on a reasonably priced and configured platform. Scalability is provided by adding additional database instances that are not clustered and that access their own physical database. This assumes that the database application can be broken down into many small, independent databases, and that no single user needs to see the data of any other user outside of the database associated with that user. A typical example would be Software as a Service (SaaS).
Basic Backup and Recovery	Uses only the functionality provided by the database server and the operating system software to perform backup and recovery. Uses the database server's CPU, memory, and I/O channels for all backup, restore, and recovery operations.
Advanced Backup and Recovery (snapshot)	Uses additional software components at the storage layer to free up the database server's CPU, memory, and I/O channels from the effects of operations relating to backup, restore, and recovery. Provides high-performance backup and restore operations, improved space efficiency, or other benefits in comparison to Basic Backup and Recovery.
Basic Protect	Uses tools provided by the operating system and database server software (in the same sense as basic backup) to provide disaster recovery. Uses the database server's CPU, memory, and I/O channels for all operations relating to the disaster recovery configuration.
Advanced Protect	Uses additional hardware or software components at the storage layer to enable disaster recovery, thereby freeing up the database server's CPU, memory, and I/O channels from the effects of these operations. Enables the creation of a writeable copy of the production database on the disaster recovery target, allowing this database to be used for operations such as backup, test/dev, and data warehouse staging.
Resiliency	Every significant layer of the solution is tested by introducing faults in an effort to cause the solution to fail. In the process, the entire solution is shown to be resilient to faults at every layer, including database clustering, networking, and storage.
Test/dev	A running production OLTP database is cloned with minimal, if any, performance impact on the production server, as well as no downtime. The resulting dataset is provisioned on another server for use for testing and development.
VMware VMotion	VMware VMotion was used to move live Oracle virtual machines from primary ESX host to secondary ESX host while subjected to load testing.
VMware HA	VMware HA was tested and validated to provide protection from server hardware failure.

Appendix B. Storage Layouts for Pure NFS Configuration

The following storage configurations represent the RAID group layouts for the pure NFS storage configuration. Two different RAID configurations were tested for this configuration:

Table 9. Pure NFS solution RAID configurations

Figure	Configuration	Description
Figure 14	Pure NFS configuration 1	1 SATA shelf 3 FC shelf RAID 5/RAID 1 AVM using user-defined storage pools
Figure 15	Pure NFS configuration 2	1 SATA shelf 2 FC shelf RAID 5/RAID 1 AVM using user-defined storage pools

These configurations have the following characteristics:

- Both configurations use a single SATA shelf using two 6+1 RAID groups (RAID 5). These RAID Groups contain the flashback recovery area and archive dump areas.
- RAID 1 groups are used for online redo log files and control files. The first configuration contains four of these RAID groups, while the second contains two.
- The remaining RAID groups are 4+1 (RAID 5) and are used to contain the Oracle data files. The two configurations differ in the number of these RAID groups and thus in total usable capacity and performance. Configuration 1 contains an additional DAE and thus provides spindles for additional capacity and performance over configuration 2.

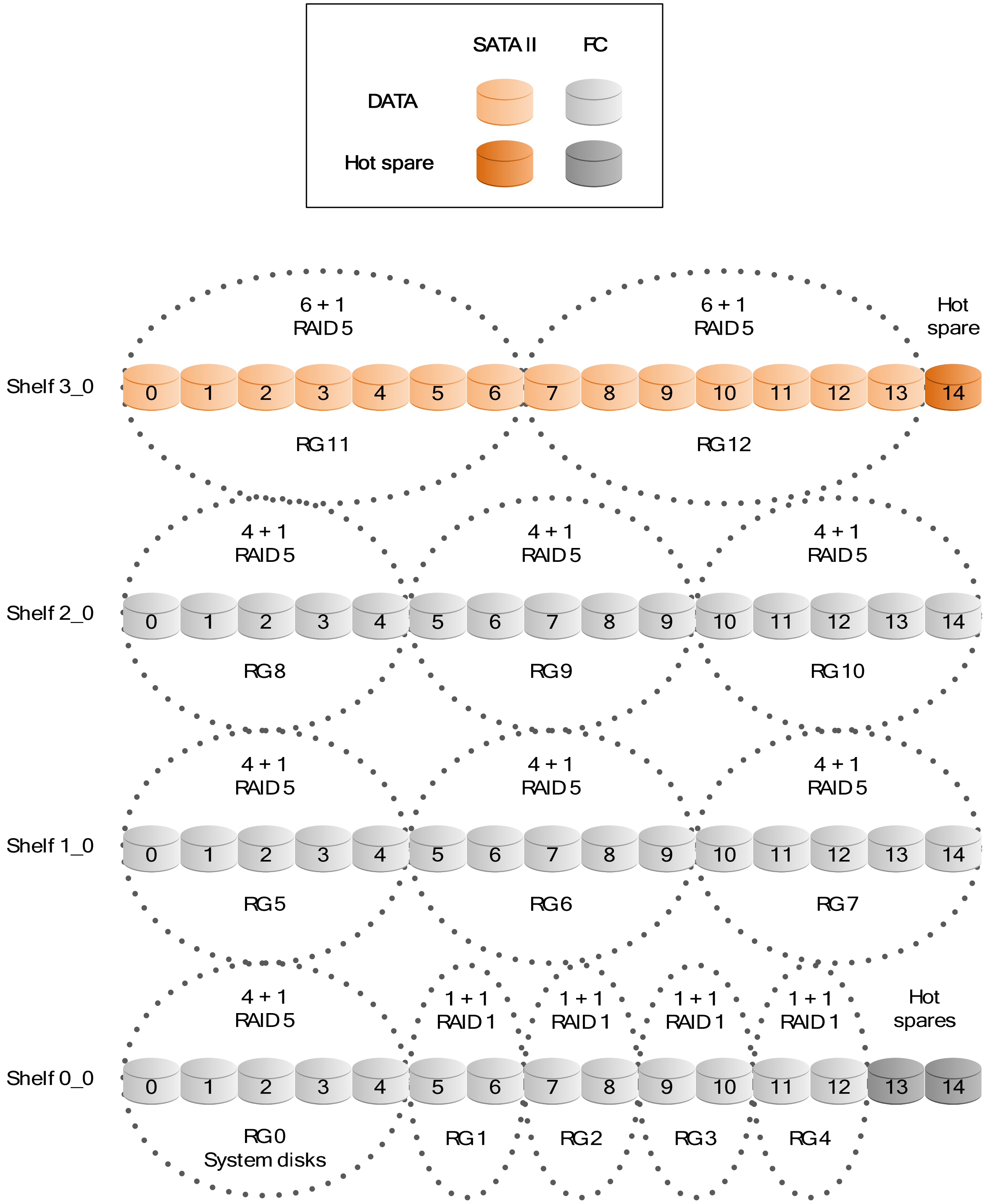

Figure 14. Pure NFS configuration 1: 3 FC shelf RAID configuration

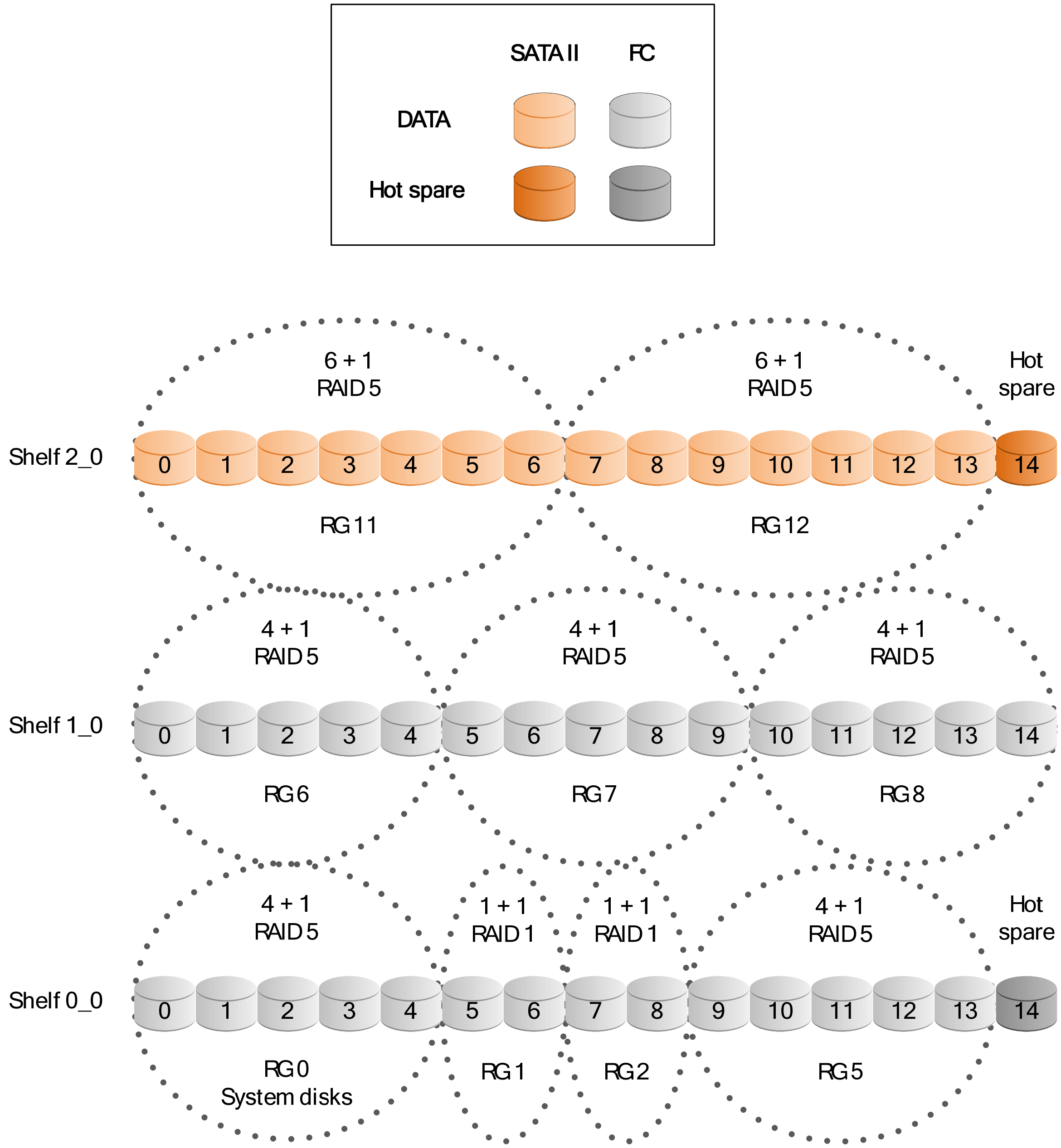

Figure 15. Pure NFS configuration 2: 2 FC shelf RAID configuration

Table 10 describes various features and capabilities provided by the RAID configurations.

Table 10. RAID Configuration Features and Capabilities

Solution Feature	Description
Disk volume setup	Once the RAID groups are created, the Celerra automatically creates disk volumes that are accessible to the Data Movers.
Automatic Volume Management	Automatic Volume Management (AVM) is a way for the user to configure Celerra volumes from RAID groups. User-defined pools allow the user to control exactly which RAID groups are used for a given volume. With AVM having user-defined pools, the pool can be automatically expanded if additional RAID groups are added later. Thus, this configuration option provides the greatest flexibility and control for configuring Celerra volumes. AVM with user-defined pools was used to configure all volumes in the solution. Figure 17 (Blended configuration 2: 3 FC shelf RAID 5/RAID 1) contains a description of all the user-defined pools used.
File systems, exports, and mount points	The file systems shown in Figure 14/15 (Pure NFS configuration 2: 3 FC shelf RAID 5/RAID 1) were created on the AVM user defined pools, exported on the Celerra NS, and mounted on the database servers. For pure NFS virtualized solutions, the file systems for hosting the databases are located on a single Celerra. The file systems are created as shown in Figure 17 (Blended configuration 2: 3 FC shelf RAID 5/RAID 1). Separate directories are created under each file system, which map to each separate database that is managed by the virtual machines. Only the relevant file systems are mounted on the virtual machines so that the virtual machines will not have access to other databases.

Table 11 describes the layout of file system volumes in RAID configurations.

Table 11. File System Layout

File system/export	AVM user-defined pool	Volumes
/vm	Default pool	/VMwareBoot
/datafs	datapool (user-defined storage pool created using datastripe volume)	datastripe (metavol consisting of all available FC 4+1 RAID 5 groups)
/log1fs	log1pool (user-defined storage pool created using log1stripe volume)	log1stripe (metavol using half of the RAID 1 groups)

File system/export	AVM user-defined pool	Volumes
/log2fs	log2pool (user-defined storage pool created using log2stripe volume)	log2stripe (metavol using half of the RAID 1 groups)
/archfs	archpool (user-defined storage pool created using archstripe volume)	archstripe (metavol using the SATA 6+1 RAID 5 group)
/flashfs	flashpool (user-defined storage pool created using flashstripe volume)	flashstripe (metavol using the SATA 6+1 RAID 5 group)
/snapdatafs	datapool (user-defined storage pool created using datastripe volume)	snapdatafs (SnapSure writeable checkpoint of the datafs volume)
/snaplog1fs	log1pool (user-defined storage pool created using log1stripe volume)	snaplog1fs (SnapSure writeable checkpoint of the log1fs volume)
/snaplog2fs	log2pool (user-defined storage pool created using log2stripe volume)	snaplog2fs (SnapSure writeable checkpoint of the log2fs volume)

Appendix C. Storage Layouts for Blended NFS/FCP Configuration

The following storage configurations represent the RAID group layouts for the blended FCP/NFS storage configuration. Two different RAID configurations were tested for this configuration:

Table 12. Oracle RAC 10g/11g FCP solution RAID configurations

Figure	Configuration	Description
Figure 16	Blended configuration 1	1 SATA shelf 3 FC shelf RAID 1-0/RAID 1
Figure 17	Blended configuration 2	1 SATA shelf 3 FC shelf RAID 5/RAID 1

These configurations have the following characteristics:

- Both configurations use a single SATA shelf using two 6+1 RAID groups (RAID 5). These RAID Groups contain the flashback recovery area and archive dump areas.
- RAID 1+1 groups are used for online redo log files and control files. The first configuration contains four of these RAID groups, while the second contains two.
- The remaining RAID groups are 4+1 (RAID 5) and are used to contain the Oracle data files. The two configurations differ in the RAID level protection. Configuration 1 uses a RAID 1/0 configuration for the Oracle data files and configuration 2 uses RAID 5. RAID 1/0 provides additional performance over RAID 5; however it provides less usable capacity.

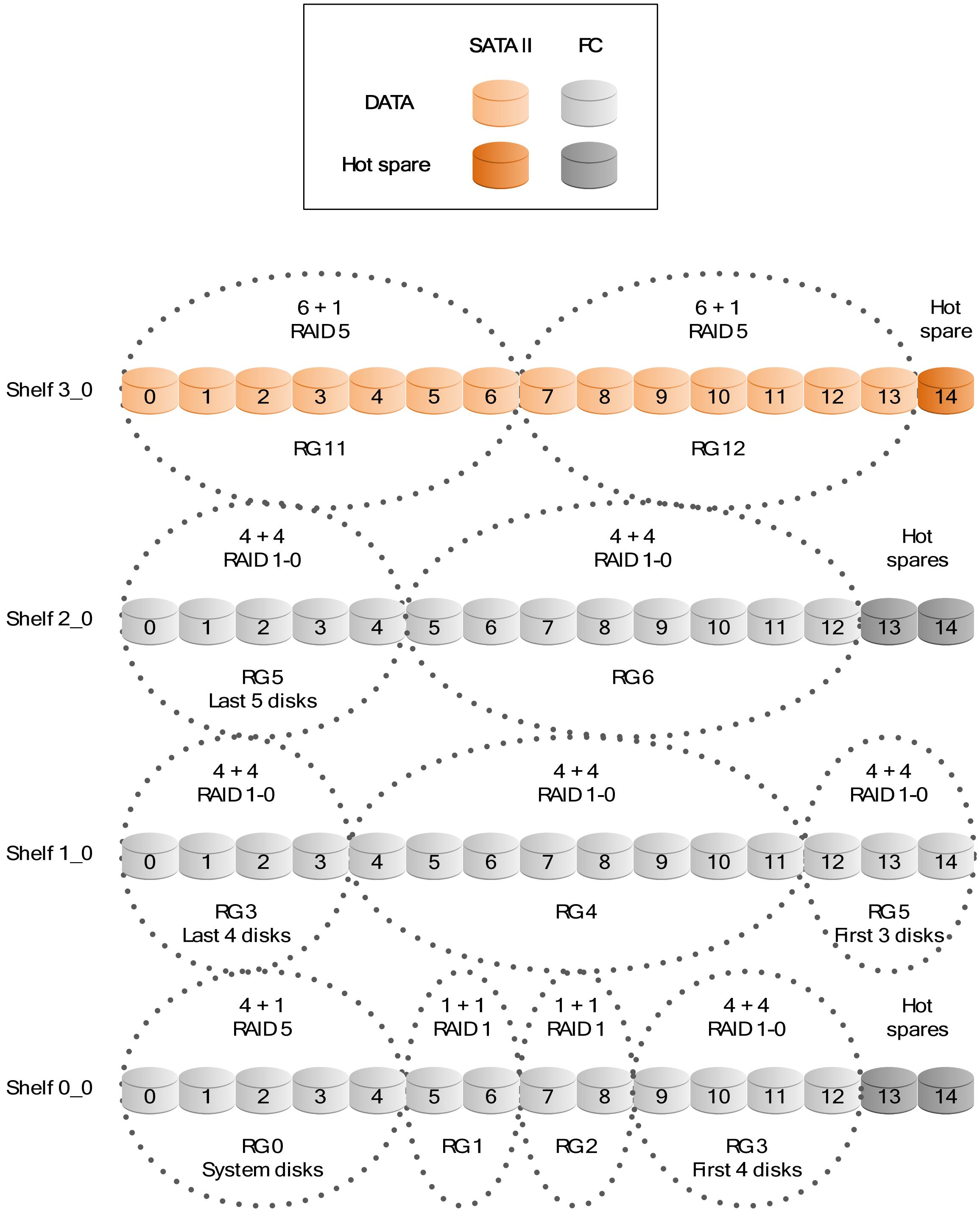

Figure 16. Blended configuration 1: 3 FC shelf RAID 1-0/RAID 1

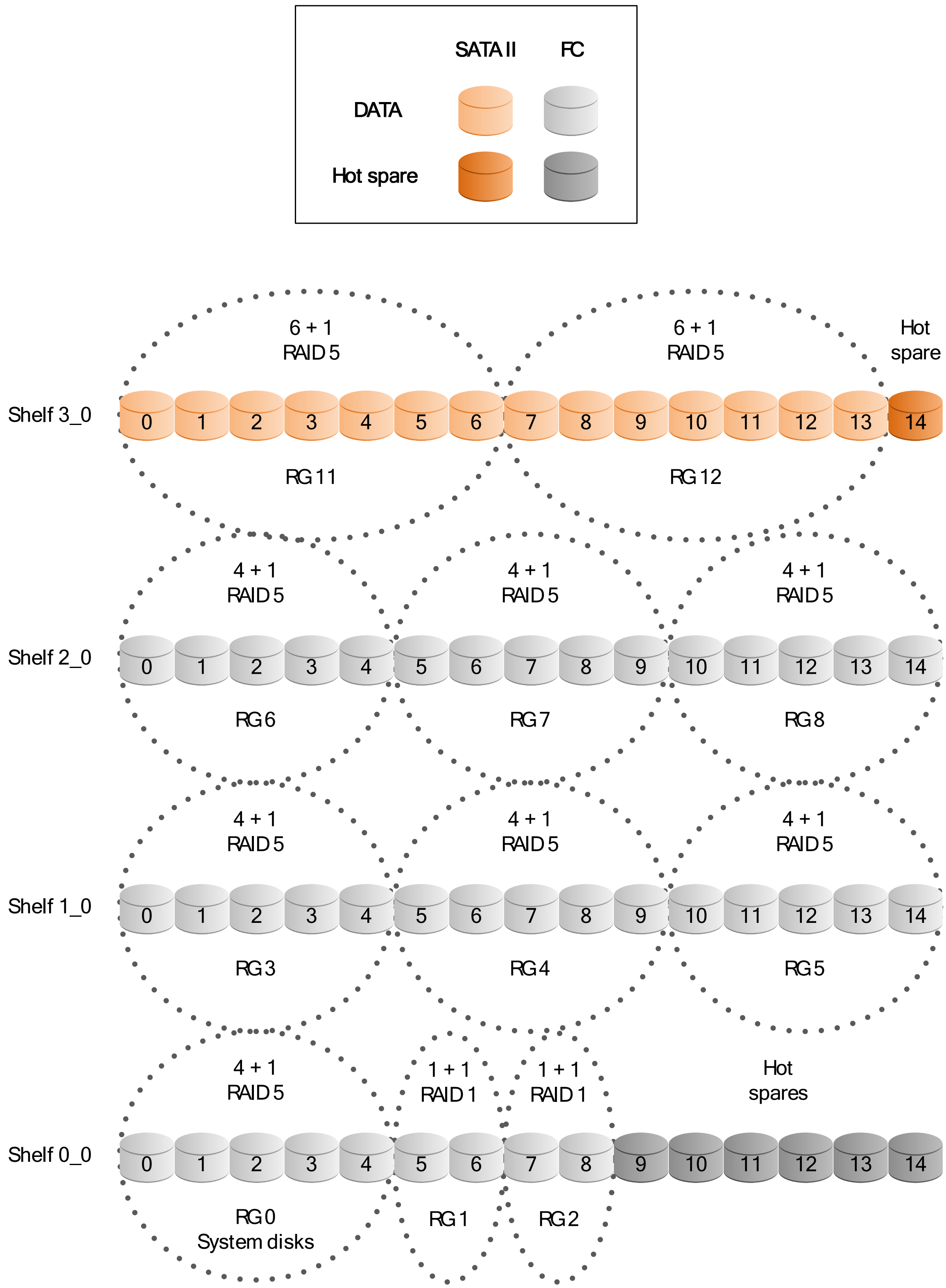

Figure 17. Blended configuration 2: 3 FC shelf RAID 5/RAID 1

Table 13 describes various features and capabilities of the blended RAID configurations.

Table 13. Blended configuration 2: 3 FC shelf RAID 5/RAID 1

Solution Feature	Description
LUN setup	Once the RAID groups are created, EMC Navisphere® software is used to create the LUNs that are used to store the database. These LUNs are added to a storage group accessible to the RAC hosts.
ASM and NFS	ASM is used to store the database objects requiring high performance. In addition, files not requiring high performance are stored on NFS. These NFS file systems are in turn stored on low-cost SATA II drives. This brings down cost in terms of storage, while improving manageability.

Table 14 provides a detailed description of all the database objects in the RAID file system and where they are stored.

Table 14. File system layout

File system/ mount point	File system type	Data stored on	Contents
/vm	NFS	/VMwareBoot volume on Celerra	Virtual machine boot images
/RDM_PointerMap	VMFS	LUN 20	VMware Virtual Disk (VMDK) pointer files
+DATA	ASM	LUNs 3 through 8	Oracle datafiles
+LOG1 and +LOG2	ASM	LUNs 1 and 2	Online redo logs and control file (mirrored copies)
/u03	NFS	LUN 9	Flashback recovery area (all backups stored here)
/u04	NFS	LUN 10	Archived log dump destination
+CLONEDATA	ASM	LUNs 53 through 58	Cloned Oracle datafiles
+CLONELOG1 and +CLONELOG2	ASM	LUNs 51 and 52	Cloned online redo logs and control file (mirrored copies)

For the blended virtualized solution, four different sets of DATA and LOG LUNs are created on the same storage. Each set is presented to a single virtual machine using RDM. Similarly, four different directories are created on the archive and flash file systems so that they can be individually mounted on the corresponding virtual machines.

Appendix D. Hardware and Software Resources

Hardware Resources

Hardware used by the various solutions is described in Table 15.

Table 15. Solution Hardware Resources

Equipment	Quantity	Configuration
EMC Celerra NS40 Series multi-protocol storage array (includes an EMC CLARiiON CX3-40f back-end storage array)	2	▪ 2 Data Movers ▪ 4 GbE network connections per Data Mover ▪ 2 or 3 FC shelves ▪ 1 SATA shelf ▪ 30 or 45 73 GB FC disks (depending on configuration) ▪ 15 500 GB SATA disks ▪ 1 Control Station ▪ 2 storage processors
Gigabit Ethernet switch	5 (client, RAC interconnect, and storage networks)	▪ 24 ports per switch
FCP Switch	2	▪ 16 ports ▪ 4 Gb throughput
VMware ESX HA cluster server	4	▪ 2 2.66 GHz Intel Pentium 4 quad-core processors ▪ 24 GB of RAM ▪ 2 146 GB 15k internal SCSI disks ▪ 2 onboard GbE Ethernet NICs ▪ 2 additional Intel PRO/1000 PT quad-port GbE Ethernet NICs ▪ 2 SANblade QLE2462-E-SP 4 Gb/s dual-port FC HBAs (4 ports in total)
De-duplication array	2	Capacity: 12.26 TB
Virtualization server	2	▪ 4 2.86 GHz AMD Opteron quad-core processors ▪ 32 GB of RAM ▪ 2 146 GB 15k internal SCSI disks ▪ 2 onboard GbE Ethernet NICs ▪ 3 additional Intel PRO/1000 PT quad-port GbE Ethernet NICs ▪ 2 SANblade QLE2462-E-SP 4 Gb/s dual-port FC HBAs (4 ports in total)

Software Resources

Software used by the various solutions (and number of required licenses) is described in Table 16.

Table 16. Solution Software Resource Requirements

Software Title	Number of licenses
Red Hat Enterprise Linux version 4.5 or Oracle Enterprise Linux version 5.2	▪ 1 per virtualized production or target database server
VMware ESX Server 3.5	▪ 1 per VMware server
Microsoft Windows Server 2003 Standard Edition	▪ 1 per virtual client
Oracle Database 10g or 11g Standard Edition	▪ 1 per virtualized production or target database server
Quest Benchmark Factory for Databases 5.0.1	▪ 1 per virtual client
EMC Celerra Manager Advanced Edition version 5.6.37-6	▪ 1 per Celerra NS40
EMC Navisphere Agent version 6.26.0.2.24	▪ 1 per database server
EMC FLARE® version 6.26 patch level 14	▪ 1 per CLARiiON storage processor
EMC Navisphere Management version 6.26	▪ 1 per CLARiiON storage processor

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