EMC Business Continuity for Oracle Database 11*g*/10*g*

Enabled by EMC CLARiiON CX4, EMC Celerra, and VMware vSphere 4.0 High Availability Cluster using FCP and NFS

Reference Architecture

EMC Global Solutions

Copyright © 2009 EMC Corporation. All rights reserved.

Published August, 2009

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

Benchmark results are highly dependent upon workload, specific application requirements, and system design and implementation. Relative system performance will vary as a result of these and other factors. Therefore, this workload should not be used as a substitute for a specific customer application benchmark when critical capacity planning and/or product evaluation decisions are contemplated.

All performance data contained in this report was obtained in a rigorously controlled environment. Results obtained in other operating environments may vary significantly.

EMC Corporation does not warrant or represent that a user can or will achieve similar performance expressed in transactions per minute.

No warranty of system performance or price/performance is expressed or implied in this document. Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

All other trademarks used herein are the property of their respective owners.

Part number: H6392

Table of Contents

Reference architecture overview....4
Solution components....7
Technology solution....10
Key elements....14
Physical architecture....16
Validated environment profile....17
Conclusion....18
Appendix: Hardware and software resources....22

Reference architecture overview

Document purpose

This document provides an overview of the architecture of a virtualized solution for midsize enterprises that uses the EMC® Celerra® unified storage platform (Celerra) and Oracle Database 11*g* and 10*g* on Linux over Fibre Channel Protocol (FCP) and NFS.

Information in this document can be used:

- As the basis for a solution build, white paper, best practices document, or training.
- By other EMC organizations (for example, the technical services or sales organization) as the basis for producing documentation for a technical services or sales kit.

Solution purpose

The purposes of this solution are to:

- Demonstrate the functional, performance, resiliency, and scalability capabilities of an Oracle software stack that:
  - Is booted on virtualized hardware and incorporates a four-node VMware High Availability (HA) cluster
  - Uses FCP and Oracle ASM to access high-demand, low-latency storage elements
  - Uses NFS to access all other storage elements
  - Compare the performance between Virtual Machine File System (VMFS) and Raw Device Mapping (RDM).
- Integrate VMware Distributed Resource Scheduler (DRS) with Oracle Database 11*g* and 10*g*.
- Enable the physical backup and recovery of an Oracle 11*g* or 10*g* production database while offloading all performance impacts of the backup operation off of the production server, using:
  - EMC Replication Manager
  - EMC SnapView™

This demonstrates significant performance and manageability benefits in comparison to normal Oracle Recovery Manager (RMAN) backup and recovery.

- Provide disaster recovery capability using:
  - EMC RecoverPoint with CLARiiON® splitters

This demonstrates significant performance and manageability benefits in comparison to normal Oracle Data Guard disaster recovery.

- Provide the capability to clone a running production database with minimal performance impact and no downtime using EMC SnapView clones.

Continued on next page

Important note on Oracle support

Due to the requirement for RAC qualification, presently there is no support for Oracle 11*g* and 10*g* RAC servers on virtualized devices. For this reason, EMC does not publish such a configuration as a supported and validated solution. However, the use of Oracle Database 11*g* and 10*g* (in single-instance mode) presents far fewer support issues.

The business challenge

The challenge

Midsize enterprises face the same challenges as their larger counterparts when it comes to managing database environments. These challenges include:

- Rising costs
- Control over resource utilization and scaling
- Lack of sufficient IT resources to deploy, manage, and maintain complex environments at the departmental level
- The need to reduce power, cooling, and space requirements

Unlike large enterprises, midsize enterprises are constrained by smaller budgets and cannot afford a custom, one-off solution. This makes the process of creating a database solution for midsize enterprises even more challenging than for large enterprises.

Continued on next page

The solution

This solution demonstrates how organizations can:

- Deploy a solution using a combination of the NFS and FCP protocols on the Celerra.
  - FCP is used for high-I/O and low-latency database objects (notably the datafiles, tempfiles, online redo logfiles, and controlfiles).
  - NFS is used for all other database objects (consisting basically of the flashback recovery area, disk-based backups, and archive logs).
  - Manageability advantages are obtained by using a combination of FCP and NFS. Specifically, archived logs and backups can be accessed through a normal file system interface rather than ASM. This simplifies the software installation and configuration on the database servers.
- Avoid investing in additional FC infrastructure by implementing a blended solution that uses both FCP and NFS to access storage elements.
- Get the most out of hardware by using VMware DRS to automatically reallocate virtual machines, according to need, across servers.
- Maximize the use of the database-server CPU, memory, and I/O channels by offloading performance impacts off of the production server during:
  - Backup and recovery by using EMC Replication Manager
  - Disaster recovery operations by using EMC RecoverPoint
- Implement a disaster recovery solution with RecoverPoint that reduces costs and complexity by using IP as the network protocol.
- Reduce the complexity of backup operations and eliminate the need to implement scripted solutions by using EMC Replication Manager.
- Use SnapView to free up the database server's CPU, memory, and I/O channels from the effects of operations relating to backup, restore, and recovery. SnapView clones also help in creating test/development systems without any impact on the production environment.

Solution components

Introduction

This section describes the components of the solution, and explains some of the key terminology and concepts that are used in this document.

Key terms defined

Solution

A solution is a complete stack of hardware and software upon which a customer would choose to run their entire business or business function.

Core solution component

A core solution component addresses the entire solution stack, but does so in a way relating to a discrete area of testing. For example, performance testing is a core solution component.

Functionality solution component

A functionality solution component addresses a subset of the solution stack that consists of a discrete set of hardware or software, and focuses on a single IT function. For example, backup and recovery, and disaster recovery are functionality solution components.
A functionality solution component can be either “basic” or “advanced.”

Basic solution component

A basic solution component uses only the features and functionality provided by the Oracle stack. Examples are RMAN for backup and recovery, and Data Guard for disaster recovery.

Advanced solution component

An advanced solution component uses the features and functionality of EMC hardware or software. Examples are EMC SnapView for backup and recovery, and EMC MirrorView for disaster recovery.

Continued on next page

Core solution components

The following table describes the core solution components that are included in this solution:

Component	Description
Scale-out OLTP	This is the baseline performance test presently used for virtualized solutions. It consists of an industry-standard TPC-C workload running on a cluster of VMware ESX servers. Multiple database images are used for the workload, emulating a software-as-a-service (SaaS) customer. This is a scale-out, because all users in the database are not able to access all rows in the combined databases but can only see the data in the database to which it is connected. Thus, the database is federated.
Resiliency	The purpose of resiliency testing is to validate the fault-tolerance and high-availability features of the hardware and software stack. Faults are inserted into the configuration at various layers in the solutions stack. Some of the layers where fault tolerance is tested include: • VMware HA cluster interconnect port • Storage processor

Continued on next page

Functionality solution components

The following table describes the functionality solution components that are included in this solution:

Component	Description
Basic Backup	This is backup and recovery using Oracle RMAN, the built-in backup and recovery tool provided by Oracle.
Advanced Backup	This is backup and recovery using EMC value-added software or hardware. In this solution the following are used to provide Advanced Backup functionality: • EMC Replication Manager • EMC SnapView
Basic Protect	This is disaster recovery using Oracle Data Guard, Oracle's built-in remote replication tool.
Advanced Protect	This is disaster recovery using EMC value-added software and hardware. In this solution the following is used to provide Advanced Protect functionality: • EMC RecoverPoint with CLARiiON splitters
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. EMC Replication Manager is used to automate the Test/dev process.

Technology solution

Overview

Two sites connected by a WAN are used in the solution, one site is used for production and the other site is used as a disaster recovery target. A Celerra is present at each site. Oracle Database 11*g* or 10*g* for x86-64 is run on Red Hat Enterprise Linux or on Oracle Enterprise Linux.

The Oracle software stack is booted on virtualized hardware and incorporates a four-node VMware HA cluster. For the purpose of high availability, VMware VMotion is used to migrate a VM running the production database from one ESX server to another without any downtime. Each ESX server is connected to both the production storage and to the cluster-interconnect TCP/IP networks. VMware DRS is used to manage the allocation of resources automatically to different VMs based on the need.

FCP storage networks consisting of dedicated, redundant FCP switches are present at both sites.

The solution also includes virtualized servers for use as Test/dev, Advanced Backup, Basic Protect, and Advanced Protect targets. Virtualization of the test/dev and disaster recovery (DR) target servers is supported using VMware ESX Server.

Blended FCP/NFS

This is a blended FCP/NFS solution. Depending on the nature of the database object, either FCP or NFS is used to access it. FCP is used to access high-demand, low-latency objects. NFS is used to access low-demand, high latency objects.

“Storage layout” describes in detail how each Oracle file type and database object is stored and accessed for this solution.

Continued on next page

Production site

The following components are present at the production site and are connected to the production FCP storage network and to the WAN:

- Virtualized Oracle Database 11*g* or 10*g* scale-out servers running on VMs hosted on a four-node VMware HA cluster. The virtualized servers are connected to the production FCP storage network using virtualized connections.
- A Celerra (actually the CLARiiON back-end array) is connected to the four-node VMware HA cluster through the production storage network. EMC SnapView is used to provide an advanced backup solution, and SnapView clones are used to create a test/dev target database.
- A RecoverPoint cluster connected to the FCP storage network, and to the WAN

The virtualized Oracle Database 11*g* or 10*g* servers are also connected to the client and VMware HA cluster interconnect networks.

Disaster recovery target site

The disaster recovery target site consists of:

- A virtualized single-instance Oracle Database 11*g* or 10*g* server that is used as:
  - The disaster recovery target for Basic Protect
  - The target for test/dev

  The virtualized single-instance server is connected to the client, WAN, and target storage networks through virtualized connections on the virtualization server.
- A Celerra connected to the VMware ESX server through the disaster recovery storage network. The Oracle Database 11*g*/10*g* single-instance target server accesses this network through a virtualized switch on the ESX server.
- A RecoverPoint cluster connected to the FCP storage network, and to the WAN.

Connected to both sites

- A VMware ESX server is connected to both the production and target FCP storage networks.
- A virtualized single-instance Oracle 11*g* or 10*g* server is used as:
  - The disaster recovery target for Basic Protect and Advanced Protect (DR site)
  - The target for Test/dev (production site)

  The virtualized single-instance Oracle 11*g* or 10*g* target server accesses both the production and target FCP storage networks and is connected to the client WAN through virtualized connections on the virtualization server.

The LUNs on the Celerra are formatted using the VMFS file system; VMDK

files are provisioned on the target VMs within this file system.

Storage layout

The following table describes how each Oracle file type and database object is stored and accessed for this solution:

What	Object type	Protocol	Stored on	File-system type
Oracle datafiles	High demand Low latency	FCP	FC disk (LUNs)	RAID-protected ASM diskgroup
Oracle tempfiles				
Oracle online redo logfiles				
Oracle controlfiles				
VMDK pointer files				
Archived logfiles	Low demand High latency	NFS	SATA II	RAID-protected NFS
Flashback recovery area				
Disk-based backups				

- High-performance database objects are accessed over an FCP network using redundant network switches.
- FCP storage networks consist of dedicated, redundant FCP switches. Four separate sets of LUNs (for data and redo logs) are created on the back-end storage and each corresponding set is presented to the VM using RDM and VMFS. The FCP LUNs were presented to the VM using RDM on one of the configurations and using VMFS on the other. The performance on these configurations was validated and a comparative analysis performed.
- Advanced Backup and Test/Dev solutions were validated by integrating Replication Manager on the configuration where the LUNs were presented using RDM.
- The disaster recovery solution by integrating RecoverPoint was performed on the configuration, where the LUNs were presented using VMFS.

ASM and ASMLib

- Oracle ASM is used as the file system and volume manager. ASMLib is used to virtualize the LUNs on the database server.
- Oracle datafiles, tempfiles, and online redo logfiles are stored on separate LUNs that are mounted on the database server using ASM over FCP.
- Three ASM diskgroups are used — one diskgroup for datafiles and tempfiles, and two diskgroups for online redo logfiles.
- The online redo logfiles are mirrored across the two ASM diskgroups using

Oracle software multiplexing. The controlfiles are mirrored across the online redo log ASM diskgroups.

- Each ASM diskgroup and its underlying LUNs are designed to satisfy the I/O demands of individual database objects. For example, RAID 5 is used for the datafiles but RAID 1 is used for the logfiles. All of these diskgroups are stored on FC disks.

IP network architecture

- TCP/IP and NFS provide network connectivity and file system semantics for NFS file systems on Oracle Database 11*g* or 10*g*.
- Client virtual machines run on the VMware ESX server. They are connected to a client network.
- Client, VMware HA cluster interconnect and redundant TCP/IP storage networks consist of dedicated network switches and virtual local area networks (VLANs).
- The VMware HA cluster interconnect and storage networks consist of trunked IP connections to balance and distribute network I/O. Jumbo frames are enabled on these networks.

Key elements

Introduction

This section provides an overview of the technologies that are used in this solution.

EMC Celerra unified storage platform

The Celerra is a remarkably versatile device. Celerra includes a world-class network-attached storage (NAS) array combined with the functionality and performance of the leading midrange storage area network (SAN) array. Celerra provides both NAS and SAN functionality and performance without compromise.

The key features provided by Celerra are described in the following table:

Feature	Provided by
SAN storage	Fibre Channel Protocol (FCP) through the back-end EMC CLARiiON CX4 storage array
NAS	Network File System (NFS) and Common Internet File System (CIFS) protocols
iSCSI storage	Celerra's Data Movers

EMC SnapView

EMC SnapView lets you create local point-in-time snapshots and complete data clones for testing, backup, and recovery operations. With SnapView, you can create multiple copies of production data on your CLARiiON quickly and easily.
SnapView enables you to:

- Carry out nondisruptive backup operations with space-saving SnapView snapshots
- Restore data quickly and easily with SnapView clones
- Accelerate application development by quickly providing the most current data for testing

EMC Replication Manager

EMC Replication Manager manages EMC point-in-time replication technologies through a centralized management console. Replication Manager coordinates the entire data replication process – from discovery and configuration to the management of multiple application consistent disk-based replicas. Replication Manager can auto-discover your replication environment and enable streamlined management by scheduling, recording, and cataloging replica information including auto-expiration.

Replication Manager integrates with the Oracle database server and provides an easy interface to create and manage Oracle replicas.

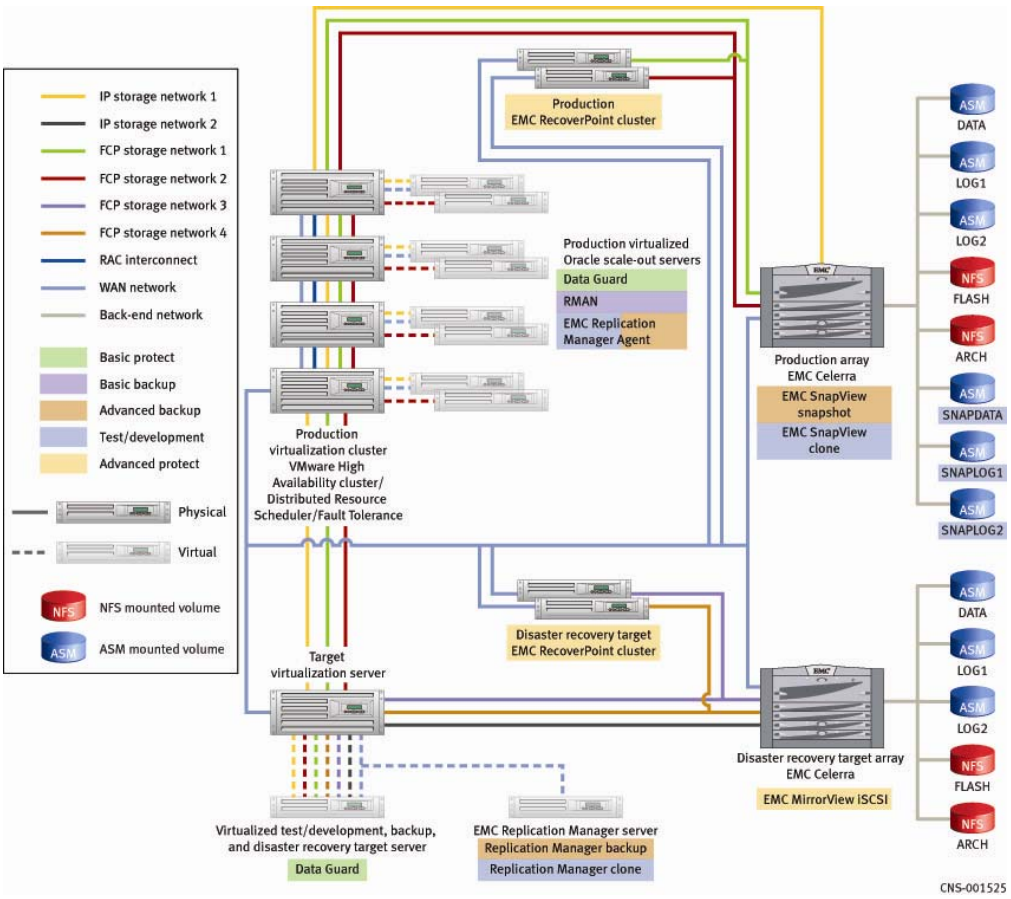
Continued on next page

EMC RecoverPoint	EMC RecoverPoint provides integrated continuous remote replication (CRR) and continuous data protection (CDP) for on-demand protection and recovery at any point in time. RecoverPoint's advanced capabilities include policy-based management, application integration, and WAN acceleration. RecoverPoint can be implemented to provide a single, unified solution to protect and replicate data across heterogeneous storage. This enables the customer to simplify management and reduce costs, recover data at a local or remote site at any point in time, and ensure continuous replication to a remote site without impacting performance.
EMC CLARiiON splitter	The CLARiiON (CX) splitter is a driver that can be directly installed on a CLARiiON. When an application writes to storage, the CX splitter splits the data and sends one copy of the data over FCP to the source-site RecoverPoint appliance (RPA), and a second copy to the storage array. These requests are acknowledged by the source RPAs and are then replicated to the remote RPAs over the IP WAN network. The main advantages of CLARiiON array-based splitters are that they: • Reduce the cost associated with the additional host-based splitter agents or specialized fabric • Provide concurrent local and remote (CLR) protection. CLR replication can be performed on the same LUNs.
Oracle software stack	The Oracle software stack covered by this solution consists of: • Oracle Database 11*g* or 10*g* • Oracle Enterprise Linux • Automatic Storage Management (ASM)
VMware ESX Server	VMware ESX is the flagship virtualization server product from VMware and is the market leader in server virtualization and cloud computing. VMware vSphere provides a high-performance, robust, fault-tolerant and high-availability virtualization solution. This also helps in leveraging virtualization features to enable cloud computing, which will help in better manageability of IT infrastructure and lowest total cost per application workload.
VMware Distributed Resource Scheduler	VMware Distributed Resource Scheduler (DRS) is a cluster feature of VMware vCenter Server; it provides dynamic load balancing and resource sharing for multiple virtual machines across ESX servers. DRS uses VMotion as the underlying transport feature to move VMs from one ESX server to another.

Physical architecture

Architecture diagram

The following illustration depicts the overall physical architecture of the solution.



Validated environment profile

Environment profile and test results	For information on the validated environment profile and performance results, refer to the Validation Test Reports for this solution. This information can be accessed on EMC Powerlink®.

Conclusion

Summary

This section provides a summary of the solution and of the business challenges that it addresses.

Reduced total cost of ownership

In any reasonable configuration, the database server's CPU is the most precious component of the entire architecture. Therefore, the over-arching principle of EMC's Oracle Database 11*g* and 10*g* solutions for midsize enterprises is to free up the database server's CPU (as well as memory and I/O channels) from utility operations such as backup and recovery, disaster recovery staging, test/dev, and cloning. The highest and best use of the database server’s CPUs is to parse and execute the SQL statements that are required by the application user.

CPU usage

This solution reduces the load on the database server CPU by using:

- EMC Replication Manager and EMC SnapView to carry out a physical backup of an Oracle 11*g* or 10*g* production database, while offloading all performance impacts of the backup operation off of the production server. EMC RecoverPoint for disaster recovery operations.

Blended FCP/NFS architecture

The use of a blended FCP/NFS architecture provides very significant ease-of-use advantages. Provided that the customer has access to a Celerra unified storage platform, the amount of software that the customer is required to install and configure on the database server is significantly reduced.

The NFS protocol provides access to storage elements such as archive logs and disk-based backups using normal file-system semantics. This is a significantly simpler form of management than that provided through FCP and ASM.

The use of FCP allows for high-performance, low-latency I/O for databases that require high-performance storage.

Continued on next page

Virtualization

The use of virtualization for the production database server provides manageability and ease-of-use advantages. In a scale-out context, virtualization can provide superior performance and scalability compared to physically booted configurations – even when using hardware identical to that used in the physically booted configuration.

Utility servers, such as a Test/dev target and Basic Protect target, are more easily and conveniently managed as virtual machines than as physically booted Oracle database servers. The advantages of consolidation, flexible migration and so on, which are the mainstays of virtualization, apply to these servers as well.

Reduced costs

One of the main challenges faced by the customer is to reduce cost by utilizing infrastructure effectively. Virtualization enables reduction in the number of servers and related IT hardware in the data center. The other main feature is the ability to move a running VM production database from one physical sever to another physical server without any downtime.

Distributed resource scheduling

VMware DRS provides the ability to distribute the workload across multiple ESX servers by migrating virtual machines according to resource consumption and demand. The VMware DRS helps in dynamically managing the resources pooled together from multiple ESX servers. VMware helps in aggregating and centrally managing the resources of multiple hosts in the form of resource pools. Different sets of rules and priority can be defined for individual Oracle VMs based on the criticality. When an Oracle database server VM operating within a resource pool experiences increased utilization of resources, DRS will try to allocate the resources from the centralized pool based on the rules and priorities.

High availability

VMotion

The use of VMware VMotion to migrate an Oracle database instance from one piece of hardware to another was achieved in our tests with virtually no performance impact and no downtime on the running Oracle database instance. This provides a very high level of manageability and downtime reduction for tasks such as software and hardware upgrades.

VMware HA cluster

The use of a VMware HA cluster provides further high-availability advantages to the virtualized solution. EMC has validated the use of a VMware HA cluster with Oracle Database 11*g* and 10*g* in a single-instance scale-out environment. In addition, a VMware HA cluster can be used to automate failover in the event of a hardware or software failure.

Fault Tolerance

The use of VMware Fault Tolerance provides true zero downtime for Oracle VMware-based database servers. This means that a database server running on VMware can be protected from unplanned downtime. This is a significant improvement over the previous version of VMware, provided on the HA cluster. The use of FT is presently limited to one vCPU, making FT only applicable to smaller Oracle database servers.

Business continuity

Advanced backup and recovery

EMC Replication Manager is a comprehensive graphical application that provides Oracle storage replication using EMC storage technology. This eliminates the requirement for the customer to write scripts or to manually perform replication tasks. These tasks can now be fully automated and managed by Replication Manager.

Advanced Protect using EMC RecoverPoint and CX splitters

The advantages of Advanced Protect using RecoverPoint through the CX splitter driver are as follows:

- **Cost savings:** RecoverPoint provides data compression of between two and 15 times before shipping data over the wire (typical compression is approximately five times).
- **Superior data integrity:** RecoverPoint logs every change, and maintains those changes as long as the journal volume is capable of tracking them (this depends on change rate and the size of the journal).
- **No disruption to replication of production data:** RecoverPoint allows the customer to rapidly provision a copy of the database to test disaster recovery or other tasks, without any disruption to the ongoing replication of the production data.
- **Simultaneous local and remote replication:** RecoverPoint is unique in that it allows for simultaneous replication targets from the same source.
- **Minimal performance impact:** The performance testing carried out by EMC demonstrates that RecoverPoint has minimal, if any, impact on performance.

Test/dev

The ability to deploy a writeable copy of the production database is required by many customers. The process of provisioning this copy must create minimal, if any, performance impact on the production database server. Also, absolutely no downtime can be tolerated. The Test/dev solution documented here provides this.

Continued on next page

Conclusion, Continued

Robust performance and scaling

The resiliency testing carried out by EMC ensures that the database configuration is reliable. High availability is used at every major layer of the solution, including the database server, NAS file server, and back-end SAN array. By testing the fault tolerance of all of these layers, the ability of the application to withstand hardware failures with no downtime is assured.

The performance testing carried out by EMC utilizes an industry-standard OLTP benchmark but does so without exotic tunings that are not compliant with best practices. In addition, real-world configurations that the customer is likely to deploy are used. This enables the customer to be reasonably assured that the configuration that they choose to run their application will do so predictably and reliably.

Next steps

EMC can help accelerate assessment, design, implementation, and management while lowering the implementation risks and costs of an end-to-end solution for an Oracle Database 11*g* or 10*g* environment.

To learn more about this and other solutions contact an EMC representative or visit http://www.emc.com/solutions/application-environment/oracle/solutions-for-oracle-database.htm

Appendix: Hardware and software resources

Hardware The hardware used to validate the solution is listed below.

Equipment	Quantity	Configuration
EMC Celerra unified storage platforms (includes an EMC CLARiiON CX4 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 • DART version 5.6.44-4
Gigabit Ethernet switches	5	24 ports per switch
FCP switches	2	• 16 ports • 4 Gb throughput
VMware ESX HA cluster servers	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)
Virtualization server (VMware ESX server)	1	• 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)

Continued on next page

Software The software used to validate the solution is listed below.

Software	**Version**
Oracle Enterprise Linux	4.7
VMware ESX Server/VMware vSphere	4.0
Microsoft Windows Server 2003 Standard Edition	2003
Oracle Database Standard Edition	11*g (11g Ver 11.1.0.7)* or 10*g*
Quest Benchmark Factory for Databases	5.8.1
EMC Celerra Manager Advanced Edition	5.6
EMC Navisphere® Agent	6.26.0.2.24
EMC PowerPath®	5.0.0 (build 157)
EMC FLARE®	04.28.000.5.504
EMC DART	5.6.44-4
EMC Navisphere Management	6.28
EMC MirrorView	6.7
EMC RecoverPoint	3.0 SP1
EMC CLARiiON splitter driver	04.28.006.6.003
EMC Replication Manager	5.2.0.0