



**EMC Unified Storage for Oracle Database 11g/10g  
Physically Booted Solution**

Enabled by EMC Celerra and Linux  
using NFS and DNFS

Reference Architecture

**EMC Global Solutions**



Copyright © 2009 EMC Corporation. All rights reserved.

Published May, 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: H6294

EMC Unified Storage for Oracle Database 11g/10g – Physically Booted Solution  
Enabled by EMC Celerra and Linux using NFS and DNFS Reference Architecture

**Table of Contents**

---

Table of Contents ..... 3  
Reference architecture overview ..... 4  
Solution components ..... 6  
Technology solution ..... 8  
Key elements ..... 11  
Physical architecture ..... 14  
Validated environment profile ..... 15  
Conclusion ..... 16  
Appendix: Hardware and software resources ..... 18

---

## Reference architecture overview

---

### Document purpose

This document provides an overview of the architecture of a physically booted solution for midsize enterprises that uses the EMC® Celerra® Unified Storage Platform (Celerra) and Oracle RAC 11g and 10g on Linux.

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 purposes

The purposes of this solution are to:

- Demonstrate the functional, performance, resiliency, and scalability capabilities of an Oracle software stack that is physically booted on normal hardware, and that uses either the Oracle Direct NFS (DNFS) protocol (Oracle 11g) or the kernel NFS (KNFS) protocol (Oracle 10g) to access the storage elements for the Oracle database.
- Demonstrate the use of EMC SnapSure™ to enable the physical backup and recovery of an Oracle 11g/10g production database while offloading all performance impacts of the backup operation off of the production server. This demonstrates significant performance and manageability benefits in comparison to normal Oracle Recovery Manager (RMAN) backup and recovery.
- Demonstrate the significant performance, manageability, and efficiency benefits of Oracle DNFS over kernel NFS (KNFS).
- Demonstrate the use of a specialized hardware deduplication array, the Quantum DXi5500, to increase the density of storage of Oracle database backups.
- Provide the capability to clone a running production database with minimal performance impact and no downtime using SnapSure writeable checkpoints. The solution covers bringing up the clone database as RAC on physically booted servers and also as single instance on a virtualized target.
- Support disaster recovery using RMAN and Oracle Data Guard.

The Celerra is used for storage and consolidation. Load-balancing and high availability are managed by the Oracle DNFS client.

---

*Continued on next page*

## Reference architecture overview, Continued

---

### The business challenge

#### *The challenge*

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

---

### The solution

This solution demonstrates how organizations can:

- Maximize the use of the database-server CPU, memory, and I/O channels by offloading the performance impacts of backup, restore, and recovery operations off of the production server.
  - Use DNFS to:
    - Simplify network setup and management by taking advantage of DNFS automated management of tasks, such as setting up network subnets, LACP bonding, and tuning of Linux NFS parameters.
    - Increase the capacity and throughput of their existing infrastructure. Transactions per second and user load are both higher with DNFS than with KNFS, enabling more output from the same infrastructure.
  - Use EMC SnapSure to free up the database server's CPU, memory and I/O channels from the effects of operations relating to backup, restore and recovery. SnapSure writeable checkpoints also help in creating test/development systems without any impact on the production environment.
  - Reduce the costs of storage, power, and space by using the Quantum DXi5500 deduplication array to provide very dense storage for Oracle RMAN backup files.
  - Ensure business continuity by using Celerra and Oracle Data Guard to provide disaster recovery capability.
-

## 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 both 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. For example, 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. For example, EMC SnapView™ for backup and recovery, and EMC MirrorView™ for disaster recovery.

---

*Continued on next page*

## Solution components, Continued

### Core solution components

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

Component	Description
Scale-up OLTP	Using an industry-standard OLTP benchmark against a single database instance, comprehensive performance testing is performed to validate the maximum achievable performance using the solution stack of hardware and software.
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: Oracle RAC node, Oracle RAC node interconnect port, storage processors, Data Mover.

### 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 is used to provide Advanced Backup functionality: <ul style="list-style-type: none"><li>• EMC SnapSure</li><li>• Data deduplication using the Quantum DXi 5500</li></ul>
Basic Protect	This is disaster recovery using Oracle Data Guard, Oracle's built-in remote replication tool.
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.

## Technology solution

---

**Overview** All database objects are stored on an NFS mount. In the case of Oracle RAC 11g, datafiles, tempfiles, control files, online redo logfiles and archive log files are accessed using the DNFS protocol.

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. Oracle RAC 11g or 10g for x86-64 is run on Red Hat Enterprise Linux or on Oracle Enterprise Linux.

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

A RAC-to-RAC cloning test/dev solution component is included that consists of four physically booted, target database servers.

---

**Production site** The production site consists of:

- A physically booted four-node Oracle RAC 11g or 10g cluster
  - A Celerra connected to the Oracle RAC 11g/10g servers through the production storage network. EMC SnapSure is used to provide an advanced backup solution and SnapSure writeable checkpoints are used to create a test/dev target database.
  - A physically booted, RAC-to-RAC cloning test/dev solution component. This consists of a four-node physically booted Oracle 11g or 10g RAC cluster that is used as a test/dev target. The Oracle RAC 11g or 10g servers are connected to the client, RAC interconnect, WAN, and production storage networks.
- 

**Disaster recovery target site**

The disaster recover site consists of:

- A virtualized single-instance Oracle Database 11g or 10g server that is used as the disaster recovery target for Basic Protect and Advanced Protect. The virtualized single-instance server is connected to the client WAN and target storage networks through virtualized connections on the virtualization server.
  - A Celerra that is connected to a VMware ESX server through both the production and the disaster recovery storage networks. The Oracle Database 11g/10g single-instance target server accesses these networks through a virtualized switch on the ESX server.
- 

*Continued on next page*

## Technology solution, Continued

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

What	Where	File-system type
Oracle datafiles	FC disks	RAID-protected NFS file system
Oracle tempfiles		
Oracle online redo logfiles		
Oracle controlfiles		
Voting disk		
OCR files		
Archived logfiles	SATA II	
Flashback recovery area		
Backup target		

For implementations using Oracle Database 11g, all files are accessed using DNFS.

For implementations using Oracle Database 10g, all files are accessed using KNFS.

- RAID-protected NFS file systems are designed to satisfy the I/O demands of particular database objects. For example, RAID 5 is sometimes used for the datafiles and tempfiles, but RAID 1 is always used for the online redo logfiles. Two separate RAID configurations are supported. For more information refer to *EMC Solutions for Oracle Database 10g/11g for Midsize Enterprises Physically Booted Solutions EMC Celerra NS40 Unified Storage Platform Reference Architecture*.
- Oracle datafiles and online redo logfiles reside on their own NFS file system. Online redo logfiles are mirrored across two different file systems using Oracle software multiplexing. Three NFS file systems are used — one file system for datafiles and tempfiles, and two file systems for online redo logfiles.
- Oracle control files are mirrored across the online redo logfile NFS file systems.
- A specialized deduplication array, the Quantum DXi5500, provides a backup target for the Advanced Backup solution component.

*Continued on next page*

## Technology solution, Continued

---

### **Network architecture**

- TCP/IP provides network connectivity
  - DNFS provides file system semantics for Oracle RAC 11g. KNFS provides file system semantics for NFS file systems on Oracle RAC 10g.
  - Client virtual machines run on a VMware ESX server. They are connected to a client network.
  - Client, RAC interconnect, and redundant TCP/IP storage networks consist of dedicated network switches and virtual local area networks (VLANs).
  - The RAC 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 the Celerra are described in the following table:

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

---

### EMC SnapSure

SnapSure creates a logical point-in-time image (checkpoint) of a production file system (PFS) that reflects the state of the PFS at the point in time when the checkpoint is created. SnapSure can maintain up to 96 PFS checkpoints while allowing PFS applications continued access to the real-time data.

#### ***How SnapSure works***

The principle of SnapSure is “copy on first write” (COFW). When a block within the PFS is modified, a copy containing the block's original content is saved to a separate volume called the SavVol. Subsequent changes made to the same block in the PFS are not copied into the SavVol. The original blocks from the PFS (in the SavVol) and the unchanged PFS blocks (remaining in the PFS) are read by SnapSure according to a bitmap and blockmap data-tracking structure. These blocks combine to provide a complete point-in-time file system image called a checkpoint.

---

*Continued on next page*

## Key elements, Continued

---

- Oracle software stack** The Oracle software stack covered by this solution consists of:
- Oracle Database 11g/10g using Real Application Cluster (RAC)
  - Oracle DNFS (11g only, KNFS is used on 10g)
  - Oracle Enterprise Linux
  - Cluster Ready Services (CRS)
- 

**Oracle DNFS** DNFS is a new feature introduced in Oracle RAC 11g; it integrates the NFS client directly inside the database kernel instead of the operating system kernel. DNFS provides significant performance, manageability, and efficiency benefits over KNFS.

***Better performance***

Transactions per second and user load are both higher with DNFS than with KNFS, and this enables organizations to gain more output from the same infrastructure. CPU cost on both the database server and the file server are lower. In addition, port scaling with DNFS is much better, enabling higher bandwidth and thus higher scaling.

***High availability***

Load balancing and high availability (HA) are managed internally within the DNFS client itself.

***Concurrent I/O***

The DNFS client performs concurrent I/O by bypassing the operating system. The benefits of this are:

- Better performance due to the reduction of memory consumption and CPU utilization
- Consistent NFS performance is observed across all operating systems

***Optimized for DB workloads***

DNFS is optimized for DB workloads and supports asynchronous I/O, which is suitable for most databases; it delivers optimized performance by automatically load balancing across the available paths. Load balancing in DNFS is almost invariably superior to the conventional Linux kernel NFS (KNFS).

---

*Continued on next page*

## Key elements, Continued

---

### **Quantum DXi5500**

The Quantum DXi5500 disk array uses data deduplication to optimize the storage space required to store backups.

The array detects when a duplicate block is being stored, and simply stores a pointer in the place of the duplicate block. As a result, significant space savings can be achieved. While this requires a modest increase in backup time, no additional impact is imposed on the performance of the backup.

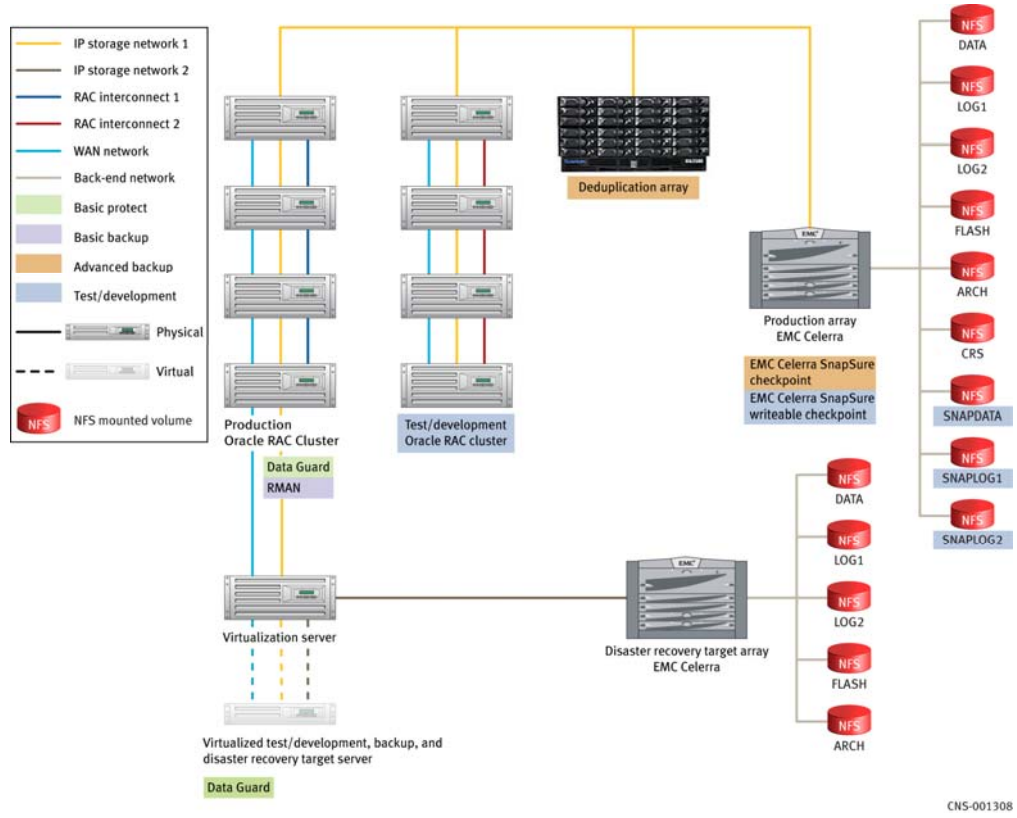
The DXi5500 backup solution addresses some of the main challenges faced by the customer such as:

- Time required to complete the backup
  - Time required to restore the backup
  - Space required to store the Oracle backup sets
-

# 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<sup>®</sup>.

---

## 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 RAC 11g and 10g 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 SnapSure to carry out a physical backup of an Oracle 11g or 10g production database while offloading all performance impacts of the backup operation off of the production server.
  - Oracle DNFS to achieve better performance due to the reduction of memory consumption and CPU utilization.
- 

### Improved performance

The Direct NFS (DNFS) client performs concurrent I/O by bypassing the operating system. The benefits of this are:

- Consistent NFS performance is observed across all operating systems.
  - DNFS is optimized for DB workloads and supports asynchronous I/O which is suitable for most databases; it delivers optimized performance by automatically load balancing across the available paths. Load balancing in DNFS is frequently superior to the conventional Linux kernel NFS (KNFS).
- 

### Ease of use

The use of DNFS simplifies network setup and management by eliminating administration tasks such as:

- Setting up network subnets
- LACP bonding
- Tuning of Linux NFS parameters

Load balancing and high availability (HA) are managed internally within the DNFS client itself

---

*Continued on next page*

## Conclusion, Continued

---

### **Business continuity**

#### ***Advanced backup and recovery***

Advanced backup and recovery with EMC SnapSure dramatically improves the mean time to recovery (MTTR) by reducing the time required for the restore operation. Further, as the backup operation has minimal impact on the database server performance, the backup can be run more often. This means that the recovery operation is also optimized since fewer archived logs must be applied.

#### ***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.

---

### **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 11g or 10g 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 Platform (includes an EMC CLARiiON CX4 back-end storage array)	2	<ul style="list-style-type: none"> <li>• 2 Data Movers</li> <li>• 4 GbE network connections per Data Mover</li> <li>• 2 or 3 FC shelves</li> <li>• 1 SATA shelf</li> <li>• 30 or 45 73 GB FC disks (depending on configuration)</li> <li>• 15 500 GB SATA disks</li> <li>• 1 Control Station</li> <li>• 2 storage processors</li> <li>• DART version 5.6.37</li> </ul>
Gigabit Ethernet switch	5	24 ports per switch
Database servers (Oracle RAC 11g/10g servers)	4	<ul style="list-style-type: none"> <li>• 2 2.66 GHz Intel Pentium 4 quad-core processors</li> <li>• 24 GB of RAM</li> <li>• 2 146 GB 15k internal SCSI disks</li> <li>• 2 onboard GbE Ethernet NICs</li> <li>• 2 additional Intel PRO/1000 PT quad-port GbE Ethernet NICs</li> <li>• 2 SANblade QLE2462-E-SP 4 Gb/s dual-port FC HBAs (4 ports in total)</li> </ul>
Deduplication array	1	Capacity: 12.26 TB
Virtualization server (VMware ESX server)	1	<ul style="list-style-type: none"> <li>• 4 2.86 GHz AMD Opteron quad-core processors</li> <li>• 32 GB of RAM</li> <li>• 2 146 GB 15k internal SCSI disks</li> <li>• 2 onboard GbE Ethernet NICs</li> <li>• 3 additional Intel PRO/1000 PT quad-port GbE Ethernet NICs</li> <li>• 2 SANblade QLE2462-E-SP 4 Gb/s dual-port FC HBAs (4 ports in total)</li> </ul>

*Continued on next page*

## Appendix: Hardware and software resources, Continued

---

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

<b>Software</b>	<b>Version</b>
Oracle Enterprise Linux	5.2
VMware ESX Server	3.5
Microsoft Windows Server 2003 Standard Edition	2003
Oracle RAC Enterprise Edition	11g or 10g
Oracle Database Standard Edition	11g or 10g
Quest Benchmark Factory for Databases	5.0.1
EMC Celerra Manager Advanced Edition	5.6.37-6
EMC Navisphere <sup>®</sup> Agent	6.26.0.2.24
EMC PowerPath <sup>®</sup>	5.0.0 (build 157)
EMC FLARE <sup>®</sup>	6.26, patch level 14
EMC DART	5.6.37
EMC Navisphere Management	6.26

---

