

Dell Reference Configuration

Deploying Oracle® Database 11g R1 Enterprise Edition Real Application Clusters with Red Hat® Enterprise Linux® 5.1 and Oracle® Enterprise Linux® 5.1 On Dell™ PowerEdge™ Servers, Dell/EMC Storage

Abstract

This white paper provides an architectural overview and configuration guidelines for deploying a two node Oracle Database 11g R1 Real Application Cluster (RAC) on Dell PowerEdge servers with Red Hat Enterprise Linux release 5 update 1 (RHEL5.1) and Oracle Enterprise Linux release 5 update 1 (OEL 5.1) on Dell/EMC storage. Using the knowledge gained through joint development, testing and support with Oracle, this Dell Reference Configuration documents “best practices” that can help speed Oracle solution implementation and help simplify operations, improve performance and availability.

April, 2008

THIS WHITE PAPER IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.

Trademarks used in this text: Intel and Xeon are registered trademarks of Intel Corporation; EMC, Navisphere, and PowerPath are registered trademarks of EMC Corporation; Microsoft, Windows, and Windows Server are registered trademarks of Microsoft Corporation. Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Red Hat is a registered trademark of Red Hat Inc. Linux is a registered trademark of Linus Torvalds. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell Inc. disclaims any proprietary interest in trademarks and trade names other than its own.

April 2008 Rev. A00

Table of Contents

ABSTRACT	1
INTRODUCTION	4
DELL SOLUTIONS FOR ORACLE DATABASE 11G.....	4
OVERVIEW OF THIS WHITE PAPER	4
ARCHITECTURE OVERVIEW - DELL SOLUTIONS FOR ORACLE 11G ON RED HAT ENTERPRISE LINUX 5.1 AND ORACLE ENTERPRISE LINUX 5.1	5
HARDWARE CONFIGURATION	6
STORAGE CONFIGURATION.....	6
<i>Configuring Dell/EMC CX3 Fibre Channel Storage Connections with Dual HBAs and Dual Fibre Channel Switches.....</i>	<i>6</i>
<i>Configuring Disk Groups and LUNs.....</i>	<i>7</i>
SERVER CONFIGURATION	10
<i>Each of the Oracle 11g RAC database cluster nodes should be architected in a highly available manner. The following sections will detail how to setup the Ethernet interfaces, the Fibre Channel host bus adapters (HBAs). These are the two fabrics that the database uses to communicate with each other and to the storage. Ensuring that these interfaces are fault tolerant will increase the availability of the overall system.</i>	<i>10</i>
<i>Configuring Fully Redundant Ethernet Interconnects.....</i>	<i>10</i>
<i>Configuring Dual HBAs for Dell/EMC CX3 storage.....</i>	<i>11</i>
SOFTWARE CONFIGURATION.....	11
OPERATING SYSTEM CONFIGURATION	11
<i>Configuring the Private NIC Teaming.....</i>	<i>11</i>
<i>Configuring the Same Public Network Interface Name on All Nodes.....</i>	<i>12</i>
<i>Configuring SSH.....</i>	<i>12</i>
<i>Configuring Shared Storage for the Oracle Clusterware using the RAW Devices Interface.....</i>	<i>12</i>
<i>Configuring Shared Storage for the Database using the ASM Library Driver.....</i>	<i>13</i>
ORACLE DATABASE 11G R1 CONFIGURATION.....	14
REFERENCE SOLUTION DELIVERABLE LIST – DELL SOLUTION FOR ORACLE 11G R1 ON ORACLE ENTERPRISE LINUX 5.1.....	15
CONCLUSION	16
TABLES AND FIGURES INDEX	17
REFERENCES	17

Introduction

Oracle 11g is the latest evolution of Oracle database technology, which brings many new features and enhancements to its customers including a database replay feature, which allows customers to simulate the production system on a test system by replaying the production workload on the test system, allowing customers to test the potential impact of system configuration changes to the mission critical production system without exposing the impact to the production system. To take advantage of the 11g features, the IT industry is moving towards the adoption of the Oracle 11g technology.

This Reference Configuration white paper is intended to help IT professionals design and configure Oracle 11g RAC database solutions using Dell servers and storage that apply “best practices” derived from laboratory and real-world experiences. This white paper documents Dell’s recommended approach for implementing a tested and validated solution for Oracle 11g RAC database on Dell’s PowerEdge 9th generation servers, Dell/EMC storage running either Red Hat Enterprise Linux release 5 update 1 (RHEL 5.1) or Oracle Enterprise Linux release 5 update 1 (OEL 5.1).

Dell Solutions for Oracle Database 11g

Dell Solutions for Oracle Database 11g are designed to simplify operations, improve utilization and cost-effectively scale as your needs grow over time. In addition to providing price/performance leading server and storage hardware, Dell Solutions for Oracle Database 11g include:

- **Dell Configurations for Oracle 11g** – in-depth testing of Oracle 11g configurations for the most in-demand solutions; documentation and tools that help simplify deployment. This will also include the best practices of implementing solutions using some new 11g core features and enhancements.
- **Integrated Solution Management** – standards-based management of Dell Solutions for Oracle 11g that lower operational costs through integrated hardware and software deployment, monitoring and update
- **Oracle Server Licensing** multiple licensing options that can simplify customer purchase
- **Dell Enterprise Support and Infrastructure Consulting Services for Oracle 11g** – including the planning, deployment and maintenance of Dell Solutions for Oracle 11g

For more information concerning Dell Solutions for Oracle Database 11g, please visit www.dell.com/oracle.

Overview of this White Paper

The balance of this white paper will provide the reader with a detailed view of the Dell Reference Configuration for Oracle Database 11g with Red Hat Enterprise Linux 5 and Oracle Enterprise Linux 5, best practices for configuring the hardware and software components and pointers for obtaining more information.

Architecture Overview - Dell Solutions for Oracle 11g on Red Hat Enterprise Linux 5.1 and Oracle Enterprise Linux 5.1

The Dell Reference Configuration for Oracle 11g on Red Hat Enterprise Linux 5.1 and Oracle Enterprise Linux 5.1 is intended to validate the following solution components:

- Two node cluster comprised of Dell PowerEdge 2950 III quad-core servers.
- Dell/EMC CX3 Fibre-Channel storage system.
- Red Hat Enterprise Linux Release 5 Update 1
- Oracle Enterprise Linux Release 5 Update 1
- Oracle Database 11g R1 Enterprise Edition (11.1.0.6) x86_64.

An architectural overview of the Dell Solution for Oracle 11g on Red Hat Enterprise Linux 5.1 and Oracle Enterprise Linux 5.1 is shown in Figure 1 below. The architectures are made of the following components:

Red Hat Enterprise Linux 5.1 Architecture:

- Dell Optiplex™ desktop systems that will access data stored within the Oracle database
- Client-server network made up of network controllers, cables and switches
- Dell PowerEdge 2950 III servers running RHEL5.1 and Oracle 11g R1 RAC (11.1.0.6)
- Dell/EMC CX3-10, CX3-20, CX3-40, and CX3-80 storage arrays
- Redundant Brocade Fibre Channel switches for a SAN environment

Oracle Enterprise Linux 5.1 Architecture:

- Optiplex desktop that will access data stored within the Oracle database
- Client-server network made up of network controllers, cables and switches
- Dell PowerEdge 2970 III servers running OEL 5.1 and Oracle 11g R1 RAC (11.1.0.6)
- Dell/EMC CX3-10, CX3-20, CX3-40, and CX3-80 storage arrays
- Redundant Brocade Fibre Channel switches for a SAN environment

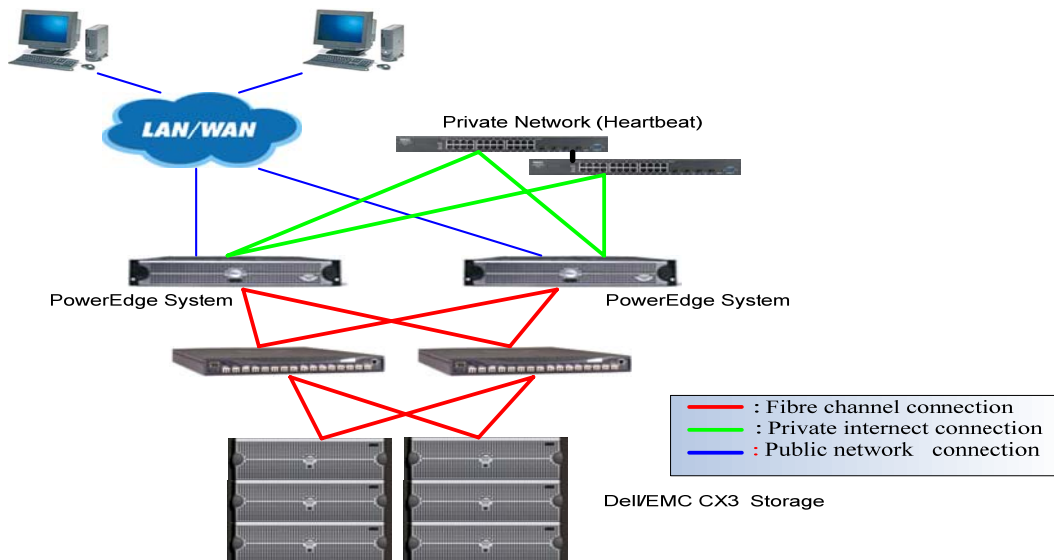


Figure 1 - Architectural Overview of Oracle on RHEL 5.1 OR OEL 5.1 with Dell/EMC Storage

Hardware Configuration

Storage Configuration

Configuring Dell/EMC CX3 Fibre Channel Storage Connections with Dual HBAs and Dual Fibre Channel Switches

Figure 2 illustrates the storage cabling of the two-node PowerEdge cluster hosting Oracle database and the Dell/EMC CX3 storage array where the data resides. Each CX3 storage array has two storage processors (SP), called SPA and SPB, which can access all of the disks in the system. The CX3 storage array provides the physical storage capacity for the Oracle 11g RAC database. Before data can be stored, the CX3 physical disks must be configured into components, known as RAID groups and LUNs. A RAID group is a set of physical disks that are logically grouped together. Each RAID group can be divided into one or more LUNs, which are logical entities that the server uses to store data. The RAID level of a RAID group is determined when binding the first LUN within the RAID group. It is recommended to bind one LUN per RAID group for database workloads to avoid disk spindle contention.¹ For details on LUN configuration, please refer to the “Configuring Disk Groups and LUNs” section below.

In the CX3 array, the LUNs are assigned to and accessed by the Oracle 11g cluster nodes directly through one storage processor. In the event of a storage processor port failure, traffic will be routed to another port on the same SP if the host is connected to more than one SP port and the EMC® PowerPath® multi path software is used. In the event of a storage processor failure, LUNs on the failed processor will trespass to the remaining storage processor. Both events could result in an interrupted service unless multiple I/O paths are configured between the Oracle 11g RAC database hosts and the CX3 array. Therefore, it is crucial to eliminate any single point of failures within the I/O path.

At the interconnect level, it is recommended that each node of the Oracle 11g RAC have two HBAs with independent paths to both storage processors. With the EMC PowerPath software installed on the cluster node, I/O can be balanced across HBAs as well. It is also recommended that two Fibre Channel switches are used because in the event of a switch failure in a single Fibre Channel switch fabric environment, all hosts will lose access to the storage until the switch is physically replaced and the configuration restored.

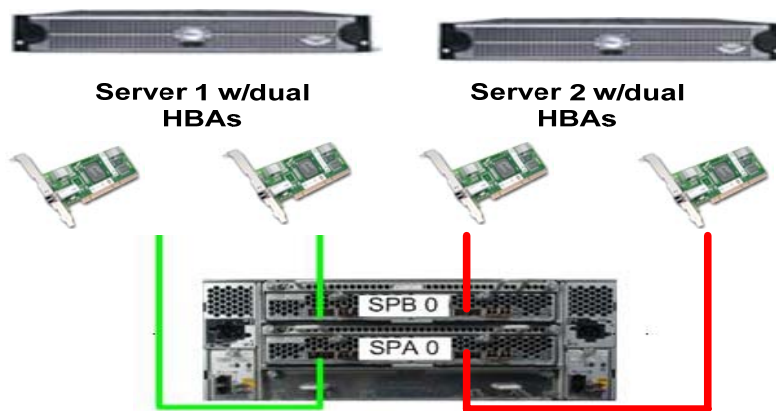


Figure 2 - Cabling a Direct Attached Dell/EMC CX3-80

¹ “Designing and Optimizing Dell/EMC SAN Configurations Part 1”, Arrian Mehis and Scott Stanford, Dell Power Solutions, June 2004. <http://www.dell.com/downloads/global/power/ps2q04-022.pdf>

Figure 3 illustrates the interconnection of a PowerEdge server hosting Oracle 11g RAC and Dell/EMC CX3-80 storage system where the database resides in a SAN environment. This topology introduces a fibre channel switch which provides the means to connect multiple storage subsystems to the host system with limited HBA ports. With the addition of the fibre channel switch additional I/O paths are introduced, which can provide additional redundancy. By using two host bus adapters (HBA) in an Active/Active configuration, commands and data can flow over both HBAs and fibre links between the server and storage system. If an HBA controller, switch, or a CX3-80 storage controller fails, operations continue using the remaining HBA – switch – CX3-80 storage controller combination.

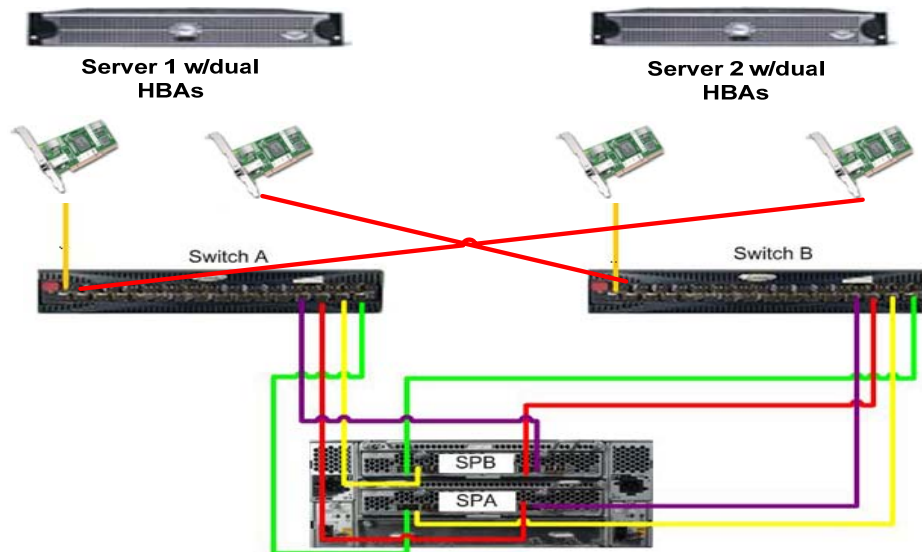


Figure 3 - Cabling a Dell/EMC CX3-80 in a SAN configuration

Note: The different colored connections in Figure 3 are fibre channel connections. The different colors from the Switches to the SPs represent different path options available from switches to the storage.

Configuring Disk Groups and LUNs

Before application data can be stored, the physical storage must be configured into components known as disk groups and LUNs. A LUN is a logical unit of physical disks presented to the host in the CX3 storage. Each disk group created provides the overall capacity needed to create one or more LUNs, which are logical entities that the server uses to store data.

Oracle Automatic Storage Management (ASM) is a feature of Oracle Database 11g which provides a vertical integration of the file system and volume manager specifically built for the Oracle database files. ASM distributes I/O load across all available resource to optimize performance while removing the need for manual I/O tuning such as spreading out the database files to avoid “hotspots.” ASM helps DBAs manage a dynamic database environment by allowing them to grow the database size without having to shutdown the database to adjust the storage allocation.²

The storage for an Oracle 11g RAC database can be divided into three areas of the shared storage. All of these storage areas will be created as block devices which are managed directly by Oracle clusterware or Oracle Automatic Storage Management (ASM) instances, bypassing the host operating system.

² “Oracle Database 10g – Automatic Storage Management Overview”, Oracle TechNet.
<http://www.oracle.com/technology/products/manageability/database/pdf/asmov.pdf>

- The first area of the shared storage is for the Oracle Cluster Registry (OCR), the Clusterware Cluster Synchronization Services (CSS) Voting Disk, and the Server Parameter File (SPFILE) for the Oracle Automatic Storage Management (ASM) instances. The OCR stores the details of the cluster configuration including the names and current status of the database, associated instances, services, and node applications such as the listener process. The CSS Voting Disk is used to determine which nodes are currently available within the cluster. The SPFILE for the ORACLE ASM instances is a binary file which stores the ASM parameter settings. Unlike traditional database files, these files mentioned above cannot be placed on the disks managed by ASM because they need to be accessible before the ASM instance starts. These files can be placed on block devices or RAW devices that are shared by all the RAC nodes.³ If the shared storage used for the OCR and Votingdisk does not provide external redundancy, it is a best practice to create two copies of the OCR and three copies of voting disk, configured in such a way that each copy does not share any hardware devices, avoiding the creation of single points of failure.⁴
- The second area of the shared storage is for the actual ORACLE database that is stored in the physical files including the datafiles, online redo log files, control files, SPFILE for the database instances (not ASM), and temp files for the temporary tablespaces. The LUN(s) on this area are used to create the ASM diskgroup and managed by ASM instances. Although the minimal configuration is one LUN per ASM diskgroup, multiple LUNs can be created for one ASM diskgroup and more than one ASM diskgroups can be created for a database.
- The third area of the shared storage is for the Oracle Flash Recovery Area which is a storage location for all recovery-related files, as recommended by Oracle. The disk based database backup files are all stored in the Flash Recovery Area. The Flash Recovery Area is also the default location for all archived redo log files. It is a best practice to place the databases data area and the flash recovery area onto their separate LUNs that do not share any common physical disks this separation can enable better I/O performance by ensuring that these files do not share the same physical disks.⁴

Table 1 shows a sample LUN configuration with LUNs for each of the three storage areas described above in a best practice and alternative configuration.⁴

Figure 4 illustrates a sample disk group configuration on a Dell/EMC CX3-80 with two Disk Array Enclosure (DAE). There are separate partitions for the OCR, QUORUM, and SPFILE, data for user defined databases, and flash recovery area on distinct physical disks. Spindles 0 through 4 in Housing 0 of CX3-80 contain the operating system for the storage. These spindles are also used during power outage to store the system cache data. It is not recommended to use the operating system spindles for as data or flash recovery area drives. As the need for the storage increases additional DAE can be added to the storage subsystem. With the use of Oracle Automated Storage Management (ASM), expansion of the DATA and the Flash Recovery Area can be simple and quick.

³ “Oracle Clusterware 11g”, Oracle Technical Whitepaper
<http://www.oracle.com/technologies/grid/docs/clusterware-11g-whitepaper.pdf>

⁴ Oracle Clusterware Installation Guide for Linux, Oracle 11g document, B28263-03
http://download.oracle.com/docs/cd/B28359_01/install.111/b28263.pdf

LUN	Minimum Size	RAID	Number of Partitions	Used For	OS Mapping
First Area LUN (Best Practice)	1024 MB	10, or 1	three of 300 MB each if RAID 10 or 1 used [6]	Voting disk, Oracle Cluster Registry (OCR) SPFILE for ASM instances	Three raw devices: 1 x Votingdisk, 1 x OCR, 1 x SPFILE if RAID 10 or 1 used
First Area LUN (Alternative)	2048 MB		Six of 300 MB each if no RAID 10 or 1 used		Six raw devices:3 x mirrored Voting disk on different disks 2 x mirrored OCR on different disks 1 x SPFILE if no RAID 10 or 1 used
Second Area LUN(s)	Larger than the size of your database	10, or 5 for read-only	One	Data	ASM disk group DATABASEDG
Third Area LUN(s)	Minimum twice the size of your second Area LUN(s)	10, or 5	One	Flash Recovery Area	ASM disk group FLASHBACKDG

Table 1 - LUNs for the Cluster Storage Groups / RAID Groups

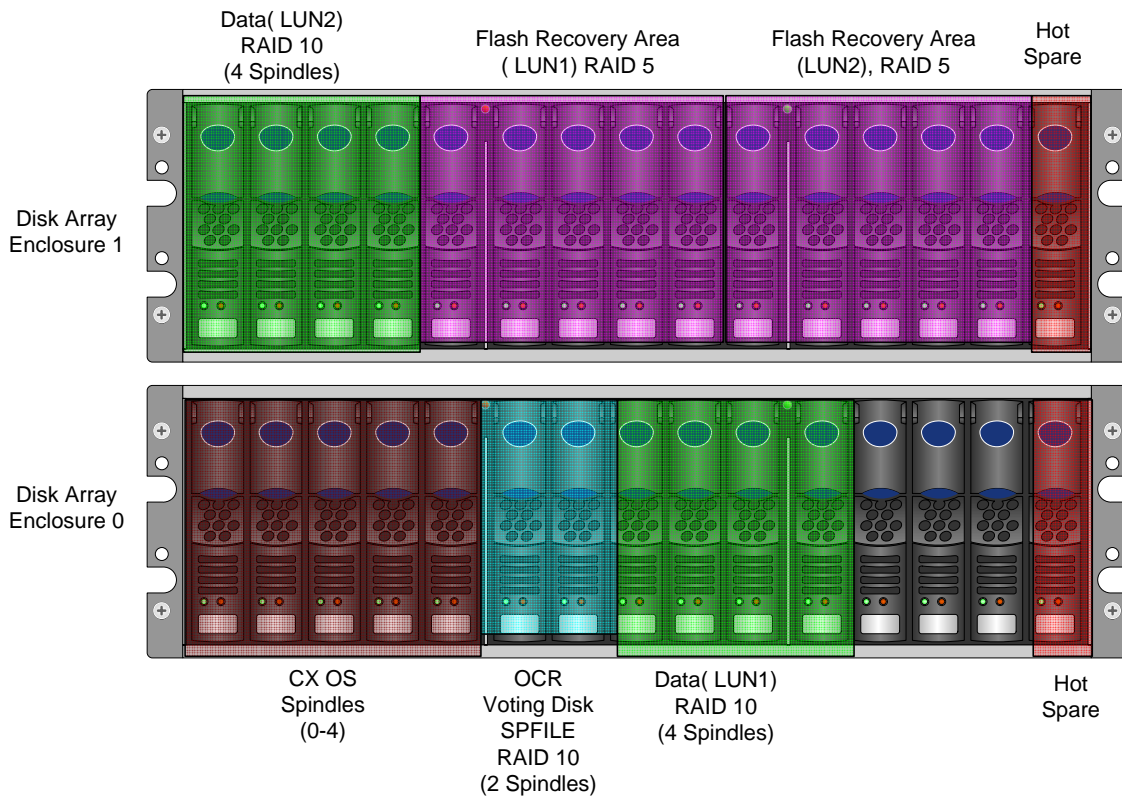


Figure 4 - Separation of Disk Groups and LUNs within a Dell/EMC CX 3-80 Storage Array

RAID 10 is considered the optimal choice for Oracle 11g RAC LUN implementation because it offers fault tolerance, greater read performance, and greater write performance.⁵ The disk group / RAID group on which the data is allocated should be configured with RAID 10.

Because additional drives are required to implement RAID 10, it may not be the preferred choice for all applications. In these cases, RAID 1 can be used for the OCR, Voting Disk, and SPFILE, which provides protection from drive hardware failure. However, RAID 0 should never be considered as an option as this configuration does not provide any fault tolerance. For the disk group / RAID group of the LUN for the data storage area, RAID 5 provides a cost effective alternative especially for predominantly read-only workloads such as a data warehouse database. Since the Flash Recovery Area is recommended as 2 times of size of Data, if the space becomes an issue, RAID5 can be used for Flash Recovery Area. However, RAID 5 is not suitable for data for the database with heavy write workloads, such as in an OLTP database, as RAID 5 can have significantly lower write performance due to the additional read and write operations that come with the parity blocks on top of the load generated by the database.

From the example above:

- 2 spindles are the minimal number of disks to form a RAID1 to ensure the physical redundancy.
- Initially we allocate two RAID 10 LUNs of 4 disks each for data, and two RAID 5 LUNs of 5 disks each for the Flash Recovery.
- Then we use 4 disks for data LUN in RAID 10 and 5 disks for Flash Recovery LUN in RAID 5 as a unit or a building block to add more storage as the database grows in future.

Each LUN created in storage will be presented to all the Oracle 11g RAC hosts and configured at the OS level. For details on the shared storage configuration at the OS level, please refer to the “Configuring Shared Storage for the Oracle Clusterware using the RAW Devices Interface” section and the “Configuring Shared Storage for the Database using the ASM Library Driver” section below.

Server Configuration

Each of the Oracle 11g RAC database cluster nodes should be architected in a highly available manner. The following sections will detail how to setup the Ethernet interfaces, the Fibre Channel host bus adapters (HBAs). These are the two fabrics that the database uses to communicate with each other and to the storage. Ensuring that these interfaces are fault tolerant will help increase the availability of the overall system.

Configuring Fully Redundant Ethernet Interconnects

Each Oracle 11g RAC database server needs at least three network interface cards (NICs): one NIC for the external interface and two NICs for the private interconnect network. The servers in an Oracle 11g RAC are bound together using cluster management software called Oracle Clusterware which enables the servers to work together as a single entity. Servers in the cluster communicate and monitor cluster status using a dedicated private network also known as the cluster interconnect or private interconnect. One of the servers in the RAC cluster is always designated as the master node.

In the event of a private interconnect failure in a single interconnect NIC environment, the server communication to the master node is lost, and the master node will initiate recovery of the failed database instance on the server.

In the event of a network switch failure in a single private network switch environment, a similar scenario will occur, resulting in a failure of every single node in the cluster except for the designated master node.

⁵ “Pro Oracle Database 11g RAC on Linux”, Julian Dyke and Steve Shaw, Apress, 2006.

The master node will then proceed to recover all of the failed instances in the cluster before providing a service from a single node which will result in a significant reduction in the level of service and available capacity.

Therefore, it is recommended to implement a fully redundant interconnect network configuration, with redundant private NICs on each server and redundant private network switches.⁶

Figure 5 illustrates the CAT 5E/6 Ethernet cabling of a fully redundant interconnect network configuration of a two-node PowerEdge RAC cluster, with two private NICs on each server, and two private network switches. For this type of redundancy to operate successfully, it requires the implementation of the Link Aggregation Group, where one or more links are provided between the switches themselves. These two private interconnect network connections work independent from the public network connection.

To implement a fully redundant interconnect configuration requires the implementation of NIC teaming software at the operating system level. This software operates at the network driver level to provide two physical network interfaces to operate underneath a single IP address.⁷ For details on configuring NIC teaming, please refer to the “Configuring the Private NIC teaming” section below.

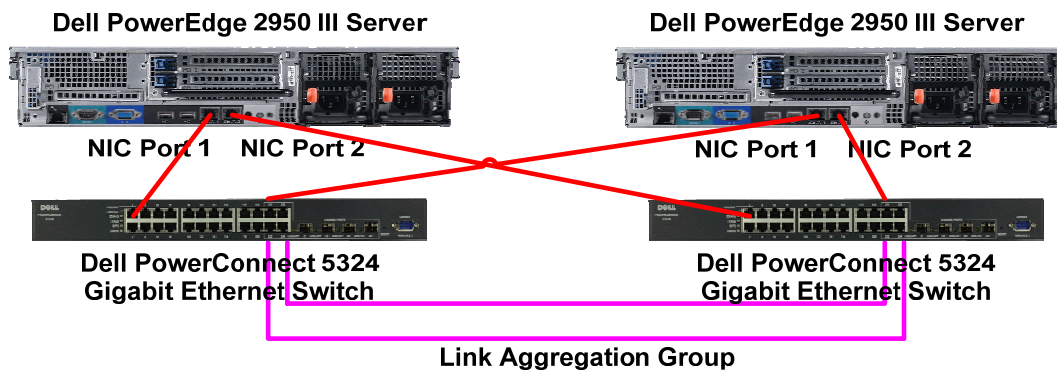


Figure 5 - Ethernet Cabling a Fully Redundant Private Interconnect Network

Configuring Dual HBAs for Dell/EMC CX3 storage

As illustrated in Figure 2 and Figure 3, it is recommended that two HBAs be installed on each of the PowerEdge server hosting the Oracle 11g RAC database because in the event of a HBA failure in a single HBA fabric environment, the host will lose access to the storage until the failed HBA is physically replaced. Using dual HBAs provides redundant links to the CX3 storage array. If the dual port HBAs is required to achieve the IO throughput, use two dual port HBAs connecting to two switches respectively to provide redundant links.

Software Configuration

Operating System Configuration

Configuring the Private NIC Teaming

As mentioned in the Section “Configuring Fully Redundant Ethernet Interconnects” above, it is recommended to install two physical private NICs (the onboard NICs can serve this purpose) on each of the

⁶ Dyke and Shaw, op. cit.

⁷ Dyke and Shaw, op. cit.

Oracle 11g RAC cluster servers to help guard against private network communication failures. In addition to installing the two NICs, it is required to use NIC teaming software to bond the two private network interfaces together to operate under a single IP address. Both Intel NIC teaming software and Broadcom® NIC teaming software are supported. The NIC teaming software provides failover functionality. If a failure occurs, affecting one of the NIC interfaces – examples include switch port failure, cable disconnection, or failures of the NIC itself – network traffic is routed to the remaining operable NIC interface. Failover occurs transparent to the Oracle 11g RAC database with no network communication interruption or changes to the private IP address.

Configuring the Same Public Network Interface Name on All Nodes

It is important to ensure that all nodes within an Oracle 11g RAC cluster have the same network interface name for the public interface. For example, if “eth0” is configured as the public interface on the first node, then “eth0” should also be selected as the public interface on all of the other nodes. This is required for the correct operation of the Virtual IP (VIP) addresses configured during the Oracle Clusterware software installation.⁸ For the purpose of installation, the public IP configured for the RAC node has to be a routable IP. They cannot be 192.xxx.xxx.xxx, 172.xxx.xxx.xxx, or 10.xxxx.xxx.xxx. However, this configuration can be changed post Oracle RAC installation.

Configuring SSH

During the installation of Oracle 11g RAC software, the Oracle Universal Installer (OUI) is initiated on one of the nodes of the RAC cluster. OUI operates by copying files to and running commands on the other servers in the cluster (is this before or after RAC is installed on the other nodes). In order to allow OUI to perform properly, the secure shell (SSH) must first be configured, so no prompts or warnings are received when connecting between SSH hosts as the oracle user. To prevent unauthorized users from accessing the systems, it is recommended that RSH be disabled after the Oracle software installation.

Configuring Shared Storage for the Oracle Clusterware using the RAW Devices Interface

Before installing Oracle 11g RAC Clusterware software, it is necessary for shared storage to be available on all cluster nodes to create the Oracle Cluster Registry (OCR) and the Clusterware Cluster Synchronization Services (CSS) Voting Disks. The OCR file and the CSS Voting disk file can be placed on a shared raw device file. As discussed in the Section “Configuring Disk Groups and LUNs” above, two LUNs are created for the OCR and three LUNs are created for Voting Disk, along with a SPFILE for the ASM instances. These LUNs should be configured as raw disk devices.

Support for raw devices has been deprecated in the Linux kernel 2.6 for Red Hat Enterprise Linux 5 and Oracle Enterprise Linux 5. Earlier versions Linux like RHEL 4 or OEL 4 allowed access RAW devices by binding the block device or partition to a character-mode device node such as /dev/raw/raw1. With the release of OEL5 and RHEL5, this technique is no longer supported. Raw device binding cannot be created using /etc/sysconfig/rawdevices file. The process for OEL5 and RHEL5 is to add custom rules to the standard rule file provided by the UDEV that comes from OS install. The UDEV facility is now the standard way to provide persistent file naming and to map RAW devices to block devices. For example, a mapping from a block device to /dev/raw/rawN can be created using a custom UDEV rule in /etc/udev/rules.d in OEL5 and RHEL5.

NOTE:

For detailed procedure of setting up custom UDEV rule to map RAW devices in RHEL5 and OEL5, please see Oracle Metalink Notes # 443996.1 and #465001.1 at <http://metalink.oracle.com>.

Oracle 11g RAC requires special ownership and permissions for the OCR and Voting Disk devices. On Red Hat Enterprise Linux 5.1 and Oracle Enterprise Linux 5.1, the udev is the default method through

⁸ Dyke and Shaw, op. cit.

which the kernel controls the creation of the special files that represent objects such as block devices. This can lead to problems because udev sets the permissions of the raw devices at every boot. A recommended solution is to alter the udev configuration so that the permissions on the raw devices are set appropriately.

NOTE:

For detailed procedures of setting up udev for OCR and Voting Disk ownership and permission, please see Oracle Metalink Note #414897.1 at <http://metalink.oracle.com>.

Configuring Shared Storage for the Database using the ASM Library Driver

As discussed in the Section “Configuring LUNs” above, two separate LUNs are created for the data storage area, and the Flash Recovery Area, respectively. It is recommended that these two LUNs be configured as ASM disks to benefit from the capabilities of ASM. For Oracle 11g R1 databases, ASM requires the installation of a number of additional RPM packages including the following rpms that work for both Enterprise Linux 5.1 and Oracle Enterprise Linux 5.1 both of which are running on Kernel 2.6.18-53.el5

- oracleasm-2.6.18-53.el5-2.0.4-1.el5.x86_64.rpm : the kernel module for the ASM library that is specific to kernel 2.6.18-53.el5
- oracleasm-support-2.0.4-1.el5.x86_64.rpm : the utilities needed to administer ASMLib
- oracleasm-lib-2.0.3-1.el5.x86_64.rpm: the ASM libraries.

In order to configure a disk to participate in an ASM diskgroup using the ASM library driver, it is required to create the disk as an Automatic Storage Management disk using the Oracleasm createdisk command. However, due to the issue that EMC PowerPath version 5.0.1 doesn't properly support the i/o calls that oracleasm makes, oracleasm createdisk command fails with the error: marking disk "/dev/emcpowera11" as an ASM disk: asmtool: Device "/dev/emcpowera11" is not a partition [FAILED]. The solution is to asmtool command directly. For the detailed information, please see Oracle MetaLink Note 469163.1 at <http://metalink.oracle.com>.

Automatic Storage Management allows the DBA to define a pool of storage called a disk group; the Oracle kernel manages the file naming and placement of the database files on that pool of storage. The DBA can change the storage allocation, adding or removing disks with SQL commands such as “create diskgroup”, “alter diskgroup” and “drop diskgroup”. The disk groups can also be managed by Oracle Enterprise Manager (OEM) and the Oracle Database Configuration Assistant (DBCA). As shown in figure 6, each Oracle 11g RAC node will contain an ASM instance that has access to the backend storage. The ASM instance, similar to database instance, communicates to other instances in the RAC environment and also features failover technology.

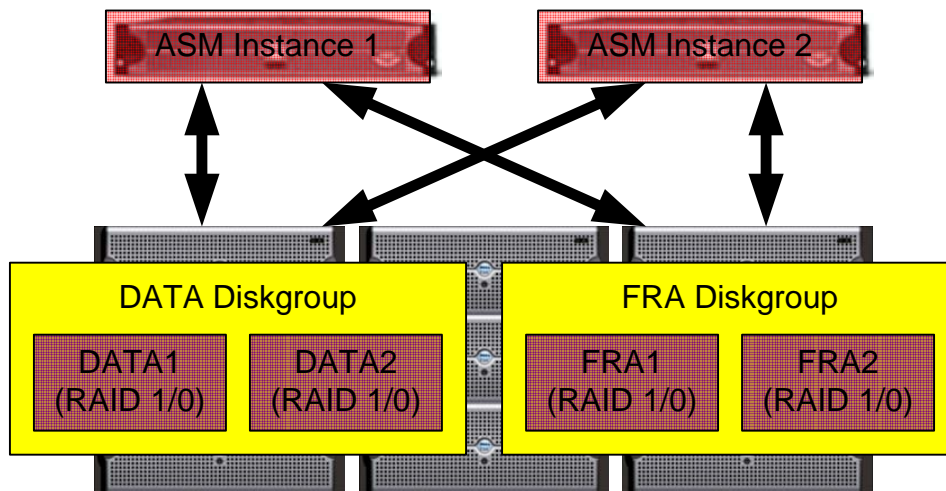


Figure 6 – ASM Instance, ASM Disks and ASM Diskgroup Layout

Oracle Database 11g R1 Configuration

The preferred method to install Oracle Cluster Ready Service (CRS) and Oracle Database is to use the Oracle Universal Installer (OUI). OUI provides a simple wizard like installation mechanism to install Oracle CRS and DB binaries on the RHEL 5.1 and OEL 5.1. During the CRS and Oracle installation the OUI will ask for general information such as paths for inventory directory, multi-node information, etc. The RAC deployment feature of OUI is further enhanced with the ability to push the required binaries to multiple nodes of a RAC from one master server.

The general installation guidelines are as follows:

1. Install Oracle 11g R1 (11.1.0.6) Clusterware.
2. Install Oracle 11g R1 (11.1.0.6) and configure ASM in an ASM Home Directory.
3. Install Oracle 11g R1 (11.1.0.6) Oracle Database software and create a cluster database.

Reference Solution Deliverable List – Dell Solution for Oracle 11g R1 on Oracle Enterprise Linux 5.1

This section contains the Solution Deliverables List (SDL) for the Dell Solution for Oracle 11g on RHEL 5.1 and OEL 5.1. It contains detailed listing of server and storage hardware configurations, firmware, driver, OS and database versions.

Minimum Hardware/Software Requirements (For details, see below)		
	Validated Component(s)	Minimum Oracle RAC Configuration
PowerEdge Nodes	PowerEdge 2950 III	2
Memory	All valid PowerEdge 2950 III memory configurations	1GB (per node)
Dell/EMC FC Storage Array	CX3-10, CX3-20, CX3-40, CX3-80	1
Fibre Channel Switch	Brocade SW4100	N/A (For Direct Attached)
HBAs	QLE 2460	2 ports (Per Node)
Ethernet Ports	Intel or Broadcom Gigabit NICs	3 (Per Node)
Ethernet Switches (For Private Interconnect)	Gigabit-only Switches	2
Raid Controllers (Used for internal storage only)	PERC 6/i	1 (Per Node)
Internal Drive	All valid PowerEdge 2950 III internal storage configurations	73 Gig/node
Oracle Software & Licenses	Oracle 11g R1 11.1.0.6 Enterprise Edition	RAC
Operating System	Red Hat Enterprise Linux 5 Update 1 Or Oracle Enterprise Linux 5 Update 2	

Table 2 - Solution Minimal Hardware/Software Requirements

Validated Servers				
PowerEdge Servers	Model	BIOS[*]	ESM/BMC Firmware[*]	Notes
		PE2950 III	2.0.1	1.77
Internal Disks RAID				
PERC 6/i	Firmware version = 6.0.1-0080 Driver version = 00.00.03.10			
Network Interconnect				
Intel NIC Drivers (1000MT)	Driver version = (e1000) 7.3.20-k2-NAPI			
Broadcom NIC Drivers (5708)	Driver version = (bnx2) 1.5.11			
Broadcom NIC Drivers	Driver version = (tg3) 3.80			
NIC Bonding	Ethernet Channel Bonding Driver, Version = 3.1.2			
Fibre Channel Host Bus Adapter (HBA)				
Qlogic HBA QLE2460	BIOS = 1.08 ; Firmware = 4.00.150; Driver = (QLE2460) version 4.00.150			
Fibre Channel Switches				
Brocade Fibre Channel Switch (SW4100)	Firmware = v5.2.1 or higher			
Fibre Channel Storage				
Dell/EMC	Storage Arrays Supported (with Software)			
	CX3-10, CX3-20, CX3-40, CX3-80 (Release 24 or later)			
Database Software				
Oracle	11g R1 11.1.0.6 Enterprise Edition			
ASMLib	oracleasm- 2.6.18-53, oracleasm-lib- 2.0.4-1			
Operating systems	RHEL 5 QU1 (kernel-2.6.18-53) OEL 5 QU1(kernel-2.6.18-53)			
EMC PowerPath	5.1.0.0-194 (available at www.emc.com)			

Table 3 – Solution Detailed Firmware, Driver and Software Versions

NOTES:

*: Minimum BIOS and ESM/BMC versions. For the latest BIOS updates go to <http://support.dell.com>

Conclusion

Dell Solutions for Oracle Database 11g are designed to simplify operations, improve utilization and cost-effectively scale as your needs grow over time. This reference configuration white paper provides a blueprint for setting up a Oracle 11g RAC database on Dell PowerEdge servers and Dell/EMC storage arrays. Although we used a two nodes RAC as an example, the deployment method applies to multiple nodes RAC configuration.

The best practices described here are intended to help achieve optimal performance of Oracle 11g on Red Hat Enterprise Linux 5.1 and Oracle Enterprise Linux 5.1. To learn more about deploying Oracle 11g RAC on PowerEdge servers and Dell storage, please visit www.dell.com/oracle or contact your Dell representative for up to date information on Dell servers, storage and services for Oracle 11g solutions.

Tables and Figures Index

Table 1 - LUNs for the Cluster Storage Groups / RAID Groups.....	9
Table 2 - Solution Minimal Hardware/Software Requirements	15
Table 3 – Solution Detailed Firmware, Driver and Software Versions	16
Figure 1 - Architectural Overview of Oracle on RHEL5.1 and OEL 5.1 with Dell/EMC Storage	5
Figure 2 - Cabling a Direct Attached Dell EMC Dell/EMC CX3-80	6
Figure 3 - Cabling a Dell/EMC CX3-80 in a SAN configuration	7
Figure 4 - Separation of Disk Groups and LUNs within a Dell/EMC CX 3-80 Storage Array.....	9
Figure 5 - Ethernet Cabling a Fully Redundant Private Interconnect Network.....	11
Figure 6 – ASM Instance, ASM Disks and ASM Diskgroup Layout.....	13

References

1. “Designing and Optimizing Dell/EMC SAN Configurations Part 1”, Arrian Mehis and Scott Stanford, Dell Power Solutions, June 2004.
<http://www.dell.com/downloads/global/power/ps2q04-022.pdf>
2. “Oracle Database 10g – Automatic Storage Management Overview”, Oracle TechNet.
<http://www.oracle.com/technology/products/manageability/database/pdf/asmov.pdf>
3. “Pro Oracle Database 11g RAC on Linux”, Julian Dyke and Steve Shaw, Apress, 2006.
4. “Oracle Clusterware 11g”, Oracle Technical Whitepaper
<http://www.oracle.com/technologies/grid/docs/clusterware-11g-whitepaper.pdf>
5. Oracle Clusterware Installation Guide for Linux, Oracle 11g document, B28263-03
http://download.oracle.com/docs/cd/B28359_01/install.111/b28263.pdf