Abstract

This white paper provides an architectural overview and configuration guidelines for deploying a two-node Oracle 10g R2 Enterprise Edition Real Application Clusters (RAC) database with Microsoft® Windows Server® 2003 R2 Enterprise or Standard Edition with SP2 on Dell™ PowerEdge™ servers with Dell EqualLogic™ PS5000XV iSCSI 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.

Introduction

The Dell EqualLogic PS Series iSCSI storage arrays provide primary and secondary storage capacity to a wide variety of applications with enterprise-class performance and low-cost ownership. The PS Series storage arrays deliver the benefits of consolidated networked storage in a self-managing, iSCSI Storage Area Network (SAN) that is easy to use and affordable. Build on a patented peer storage architecture where all arrays in a storage pool are designed to work together to provide disk capacity and evenly distribute the load, the PS Series SAN offers high performance, reliability, scalability, intelligent automation, simplified deployment, and comprehensive data protection.

The Dell EqualLogic PS Series iSCSI storage arrays are ideal choices to deploy highly reliable and sustainable Oracle 10g RAC databases. This Reference Configuration white paper is intended to help IT professionals design and configure Oracle 10g Real Application Clusters (RAC) database solutions using Dell EqualLogic storage arrays and Dell servers that apply “best practices” derived from laboratory and real-world experiences. This white paper documents the Dell recommended approach for implementing a tested and validated solution for Oracle 10g RAC database on Dell EqualLogic PS5000XV iSCSI storage arrays, and the Dell PowerEdge 9th generation servers running Microsoft Windows Server 2003 Release 2 with Service Pack 2.

Dell Solutions for Oracle Database 10g

Dell Solutions for Oracle Database 10g are designed to simplify operations, improve usability, and cost-effectively scale as your needs grow over time. In addition to providing server and storage hardware, Dell solutions for Oracle 10g include:

- **Dell Configurations for Oracle** – in-depth testing of Oracle 10g configurations for high-demand solutions; documentation and tools that help simplify deployment
- **Integrated Solution Management** – standards-based management of Dell Solutions for Oracle 10g that can lower operational costs through integrated hardware and software deployment, monitoring, and updating
- **Oracle Licensing** – multiple licensing options that can simplify customer purchase
- **Dell Enterprise Support and Infrastructure Services for Oracle** – including offerings for the planning, deployment and maintenance of Dell Solutions for Oracle Database 10g

For more information concerning Dell Solutions for Oracle 10g Database, please visit [dell.com/oracle](http://dell.com/oracle).

Executive Summary

This white paper will provide the reader with a detailed overview of the Dell Reference Configuration for Oracle 10g with Dell EqualLogic iSCSI storage arrays, best practices for configuring the hardware and software components, and pointers for obtaining more information.
Architecture Overview

The Dell Reference Configuration for Oracle 10g on Dell EqualLogic PS5000XV iSCSI storage arrays is intended to validate the following solution components:

- Two-node cluster comprised of Dell PowerEdge 2950 III servers
- Dell EqualLogic PS5000XV iSCSI storage systems
- Microsoft Windows Server 2003 R2 with SP2 Enterprise x64 Edition or Standard x64 Edition
- Oracle Database 10g R2 (10.2.0.3) Enterprise Edition x86_64 or Standard Edition x86_64

An architectural overview of the Dell Solution for Oracle 10g on Dell EqualLogic PS5000XV iSCSI storage arrays is shown in the following Figure 1. The architectures are comprised of the following components:

- 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 Microsoft Windows Server 2003 R2 with SP2 and Oracle 10g R2 RAC (10.2.0.3)
- Redundant Dell Gigabit Ethernet switches for Oracle cluster interconnect network
- Server-storage interconnect using redundant Dell PowerConnect™ 5424 or 5448 Gigabit Ethernet switches
- Three Dell EqualLogic PS5000XV iSCSI storage arrays.
  - Two arrays are configured as RAID 10, which host the Oracle database data volume and the volume required by the Oracle Clusterware (OCR and CSS Voting Disk).
  - One array is configured as RAID 5, which hosts the Oracle database Flash Recovery Area volume.

Figure 1 - Architectural Overview of Oracle 10g on Dell EqualLogic PS5000XV iSCSI Storage

*Dell PowerConnect 5424 / 5448 Gigabit Ethernet Switches can support up to three Dell EqualLogic PS5000XV arrays. To support additional PS5000XV arrays, please use Cisco Catalyst 3750 Gigabit Ethernet switches.*
Hardware Configuration

Storage Configuration

Configuring Dell EqualLogic™ PS5000XV iSCSI Storage Connections

The Dell EqualLogic PS5000XV iSCSI storage array offers a highly available hardware design that includes redundancy features to enable a no single point of failures state. Its components are fully redundant and hot swappable with optional dual active / standby control modules, standard dual fan trays, and dual power supplies. As illustrated in Figure 2, the control modules are identified as control module 0 and control module 1. Each control module has three Gigabit Ethernet interfaces, labeled as eth0, eth1, and eth2. Each control module is equipped with 1 GB battery backed write-back cache to ensure cache coherency between the two control modules. In the event of a control module failure, the other control module takes over automatically with no disruption to users.

Figure 2 - Recommended Dual Control Module Network Configuration

Host servers can be attached to the PS5000XV through an IP storage area network (SAN) industry-standard Gigabit Ethernet switch. Figure 2 shows the recommended network configuration for a dual control module PS5000XV array. This configuration includes two Dell PowerConnect 5424 Gigabit Ethernet switches, to provide highest network availability and maximum network bandwidth. It is recommended that two Gigabit Ethernet switches are used because in the event of a switch failure in a single Ethernet switch environment, all hosts will lose access to the storage until the switch is physically replaced and the configuration restored. From each of the control modules, it is recommended that one Gigabit interface connects to one Ethernet switch, and the other two Gigabit interfaces connect to the other Ethernet switch.
Figure 3 - Cabling Three Dell EqualLogic™ PS5000XV Storage Arrays with Two PowerEdge 2950 III Servers

Figure 3 illustrates the cabling of the two-node PowerEdge cluster hosting Oracle database and the three PS5000XV storage arrays where the data resides. The blue colored cables denote the iSCSI storage area network. The red colored cables denote the Oracle RAC private interconnect network. The black colored cables denote the public network. The PS5000XV storage arrays provide the physical storage capacity for the Oracle 10g RAC database. Host servers access the data through a single group IP address. A PS Series group is a storage area network comprised of one or more PS Series arrays. Each array in a group is referred to as a member. All members in a group are managed and accessed as a single storage system using the group IP address. A group or a member has a name assigned. Each member must be specified with a RAID level – either RAID 5, or RAID 10, or RAID 50 – when it is initialized.

As illustrated in Figure 3, the group named oracle-group includes three PS5000XV members: oracle-member01, oracle-member02 and oracle-member03. A PS Series storage group can be segregated into multiple tiers or pools. Tiered storage provides administrators with greater control over how disk resources are allocated. At any one time, a member can be assigned to only one pool. It is easy to assign a member to a pool and also to move a member between pools with no impact to data availability. Pools can be organized according to different criteria, such as disk types or speeds, RAID levels, application types. In Figure 3, pools are organized by member RAID levels – one pool with the name RAID-10 consists of RAID 10 members; one pool with the name RAID-5 consists of RAID 5 members.

Before data can be stored, the PS5000XV physical disks must be configured into usable components, known as volumes. A volume represents a portion of the storage pool, with a specific size, access controls, and other attributes. A volume can be spread across multiple disks and group members and is seen on the network as an iSCSI target. Volumes are assigned to a pool and can be easily moved between pools, with no impact on data availability. In addition, automatic data placement and automatic load balancing occurs within a pool, based on the overall workload of the storage hardware resources within the pool. For details on volume configuration for an Oracle RAC database, please refer to the following section, Configuring Volumes.
At the host level, in order to provide sufficient bandwidth to support three PS5000XV arrays, it is recommended that each node of the two node Oracle 10g RAC have four Gigabit NIC ports with independent paths to both Gigabit Ethernet switches of the iSCSI SAN. With the Dell EqualLogic Multipath I/O Device Specific Module (DSM) software installed on the cluster node, I/O can be balanced across NIC ports as well. As shown in Figure 3, each host server has a separate connection to the redundant Gigabit Ethernet switches and to the dual control modules. In this configuration, there is pathway redundancy at the host, at the two Ethernet switches, and at the dual RAID control modules. On each server, two of the NIC ports for iSCSI traffic are configured on separate NIC cards, leaving two other NIC ports on the two separate NIC cards for Oracle interconnect network. This is to protect against a single PCI bus failure.

At the Ethernet switch interconnection level, the bandwidth of the inter-switch link is critical to support high performing applications. In the Oracle configuration shown in Figure 3, an eight Gigabit inter-switch link is created using the Link Aggregation Group.

Configuring Volumes

Oracle® Automatic Storage Management (ASM) is a feature of Oracle Database 10g, which provides vertical integration of the file system and volume manager specifically built for the Oracle database files. ASM distributes I/O load across all available resources 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.1

The storage for an Oracle 10g RAC database can be divided into three areas of the shared storage.

- The first area of the shared storage is for the Oracle Cluster Registry (OCR), and the Clusterware Cluster Synchronization Services (CSS) Voting Disk. 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. Unlike traditional database files, these files cannot be placed on the disks managed by ASM because they must be accessible before the ASM instance starts. These files can be placed on RAW partitions that are shared by all the RAC nodes.2

- The second area of the shared storage is for the actual ORACLE database that is stored in the physical files including the data files, online redo log files, control files, SPFILE for the database instances, and temp files for the temporary table spaces. The volumes(s) on this area are used to create the ASM diskgroup and managed by ASM instances. Although the minimal configuration is one volume per ASM diskgroup, multiple volumes can be created for one ASM diskgroup and more than one ASM diskgroup can be created for a database.

- The third area of the shared storage is for the Oracle Flash Recovery Area, which is an optional storage location for all recovery-related files, as recommended by Oracle. If configured, 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 database data area and the flash recovery area onto their separate volumes 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.

---

2 “Oracle ClusterWare and Oracle Real Application Clusters Installation Guide”, 10g Release 2 (10.2) for Microsoft Windows, B14207-06 http://download-west.oracle.com/docs/cd/B19306_01/install.102/b14207.pdf

Dell Reference Configuration for Oracle 10g R2 on Dell EqualLogic PS5000XV iSCSI Storage Arrays
Table 1 shows a sample volume configuration with volumes for each of the three previously described storage areas.

<table>
<thead>
<tr>
<th>Volume</th>
<th>Minimum Size</th>
<th>RAID</th>
<th>Number of Partitions</th>
<th>Used For</th>
<th>OS Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Area Volume</td>
<td>1024 MB</td>
<td>10</td>
<td>One extended partition with two logical drives – 50 MB for Voting Disk and 120 MB for OCR</td>
<td>Voting disk and Oracle Cluster Registry (OCR)</td>
<td>One extended partition with two logical drives: 1 x Voting Disk, 1 x OCR</td>
</tr>
<tr>
<td>Second Area Volume(s)</td>
<td>Larger than the size of your database</td>
<td>10</td>
<td>One</td>
<td>Data</td>
<td>ASM disk group DATABASEDG</td>
</tr>
<tr>
<td>Third Area Volume(s)</td>
<td>Minimum twice the size of your second area volume(s)</td>
<td>5</td>
<td>One</td>
<td>Flash Recovery Area</td>
<td>ASM disk group FLASHBACKDG</td>
</tr>
</tbody>
</table>

Table 1 - Volumes for the Oracle RAC Configuration

RAID 10 is considered the optimal choice for Oracle 10g RAC LUN implementation because it offers fault tolerance, excellent read performance, and outstanding write performance. The PS5000XV array member 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 5 provides a cost-effective alternative, especially for predominantly read-only workloads such as a data warehouse database. The Flash Recovery Area is recommended as two times the size of the data. If the space becomes an issue, RAID 5 can be used for the Flash Recovery Area. However, RAID 5 is not suitable for data within the database with heavy write workloads, such as 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.

In a PS Series storage group, access control records are used to control which hosts can access a volume. Each PS Series volume has a single list of access control records. In each record, you can specify an IP address, iSCSI initiator name or Challenge Handshake Authentication Protocol (CHAP) user name (or any combination of the three). A server must match all the requirements in one record in order to access the volume. The most secure way to control access to your volumes is to use a combination of an IP address and a CHAP user name. For example, if a record includes both an IP address and a CHAP user name, a server must present the IP address and supply the CHAP user name and its associated password in order to match the record.

Each storage volume will be presented to all the Oracle 10g RAC hosts and configured at the OS level. For details on the shared storage configuration at the OS level, please refer to the “Configuring Host Access to the iSCSI Volumes,” “Configuring Shared Storage for the Oracle

---

Clusterware using the RAW Partitions,” and “Configuring Shared Storage for the Database Using the Automatic Storage Management (ASM)” sections.

As previously discussed, the Flash Recovery Area is an optional disk-based recovery area that can be used to store files for backup and recovery operations. If users choose not to configure the disk-based Flash Recovery Area, PS Series arrays provide alternative methods to perform disk-based backups of the Oracle database. Snapshots, clones, and replications are the storage-based protection features included with the purchase of the PS Series storage arrays. Snapshots enable point in time copies of volume data, which can be used for backups. Snapshots are the most space efficient method of volume protection and utilize reserve space to hold the delta between the original base volume data and the snapshot data. Clones are a complete point-in-time copy of the base volume. In addition, the PS Series replication feature allows volume data to be replicated to a secondary or remote PS Series storage group.

Configuring Challenge Handshake Authentication Protocol (CHAP)

As mentioned in the “Configuring Volumes” preceding section, the Challenge Handshake Authentication Protocol (CHAP) can be used to restrict host access to PS Series storage volumes. Challenge Handshake Authentication Protocol (CHAP) is an iSCSI authentication method that authenticates access between the storage volumes (targets) and the iSCSI initiators on the host servers. CHAP is an optional feature and is not required to use iSCSI. However, with CHAP authentication, volume access can be restricted to hosts that supply the iSCSI initiator with the correct user name and password. This information must match an access control record for the volume, in addition to an entry in a CHAP database, in order to gain access to the volume.5

iSCSI SAN Gigabit Ethernet Switch Configuration

Configuring iSCSI SAN Network

For the best performance, the iSCSI SAN network should be isolated away from other networks. This can be achieved by dedicating Gigabit Ethernet switches for iSCSI traffic or using VLANs to separate networks within a switch. As illustrated in Figure 1, the iSCSI network is physically isolated from the client network using two dedicated Dell PowerConnect 5424 Gigabit Ethernet switches.

For optimal performance, following these SAN network guidelines for PS Series iSCSI storage network:

- Use the Rapid Spanning Tree Protocol (RSTP) and enable the PortFast setting on the switch ports between switches. The Spanning Tree Protocol (STP) is a link management protocol that prevents loops in an Ethernet network by ensuring that only one active path exists between switches. Upon linkup, a switch performs a 30 to 50 second STP calculation to transition ports into forwarding or blocking state. Because STP can increase the time it takes to recover from a PS Series array control module failover or a network switch failure, it is recommended to enable switch port settings that allow the immediate transition of the port into STP forwarding state upon linkup. This functionality can reduce network interruptions that occur when devices restart. For example, the Dell PowerConnect 5424 Gigabit Ethernet switch includes a feature called PortFast, which immediately transitions a port into STP forwarding state upon linkup. It is also preferable to use Rapid Spanning Tree Protocol (RSTP) instead of STP. RSTP allows a switch port to bypass the Spanning-Tree listening and learning states and quickly enter the STP forwarding state.6

---


Dell Reference Configuration for Oracle 10g R2 on Dell EqualLogic PS5000XV iSCSI Storage Arrays 11
• Enable flow control on switch ports and NIC ports. When the data transmissions from network senders exceed the throughput capacity of network receivers, the receivers may drop packets, forcing senders to retransmit the data after a delay. Although this will not result in any data loss, latency will increase due to the data packets re-transmissions, resulting in I/O performance degradation. The flow control feature is designed to allow the network receivers to slow down network senders to avoid data re-transmissions. This amount of delay is much less than the overhead of re-transmitting packets. It is recommended to enable flow control on all switch ports and NIC ports that handle the iSCSI traffic.6

• Disable unicast storm control on switch ports. Many switches have traffic storm control features that prevent ports from being disrupted by broadcast, multicast, or unicast traffic storms on physical interfaces. These features typically work by discarding network packages when the traffic on an interface reaches a percentage of the overall load. Because iSCSI traffic is unicast traffic and can typically use the entire link, it is recommended that the unicast storm control feature is disabled on switches that handle iSCSI traffic. However, the use of broadcast and multicast storm control is encouraged on switches.7

• Enable Jumbo Frames on switches and NICs. Jumbo Frames extend the standard Ethernet frame size to 9000 bytes to allow more data to be transferred with each Ethernet transaction. PS Series storage arrays support standard Ethernet frames and Jumbo Frames up to 9000 bytes (9018 including the TCP header). It is recommended to enable Jumbo Frames on switches that handle iSCSI traffic. To obtain performance benefits and ensure consistent behavior, Jumbo Frames on the NICs must also be enabled.7


As mentioned in the “Configuring Dell EqualLogic PS5000XV iSCSI Storage Connections” section, a high bandwidth inter-switch link is critical to provide optimal performance of the iSCSI SAN. In a PS Series iSCSI SAN environment with two Gigabit Ethernet switches, it is recommended that multiple links between the switches are aggregated together to form a Link Aggregation Group (LAG). The LAG is treated as if it were a single link at a higher bandwidth. The link aggregation between switches is also called trunking. As illustrated in Figure 3, eight ports from each switch are connected and they form an 8-Gigabit LAG between the two switches. The Dell PowerConnect 5424 / 5448 Gigabit Ethernet switches can support up to eight links per Link aggregation Group.

The Dell PowerConnect 5424 Gigabit Ethernet switch has 24 ports that can support up to three Oracle nodes with three PS5000XV arrays; The Dell PowerConnect 5448 Gigabit Ethernet switch has 48 ports which can support up to 15 Oracle nodes with three PS5000XV arrays. In order to support additional Oracle nodes or additional PS5000XV arrays, alternative switches with larger port counts, such as the Cisco Catalyst 3750 Gigabit Ethernet switch, can be used. For more information on the Cisco Catalyst 3750 switch, please visit http://www.cisco.com/en/US/products/hw/switches/ps5023/.

Server Configuration

Each of the Oracle 10g RAC database cluster nodes should be architected for optimal availability. The following sections will detail how to set up the Ethernet interfaces for Oracle private

interconnect network and the Ethernet interfaces for iSCSI SAN. These are the two channels 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**

Besides the Ethernet interfaces connecting to iSCSI storage area network, each Oracle 10g RAC
database server needs at least three additional network interface cards (NIC) ports: one NIC port
for the external interface and two NIC ports for the private interconnect network. The servers in
an Oracle 10g 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 a non-redundant deployment, if the private interconnect or a network switch fails, the server
communication to the master node is lost, and the master node will initiate recovery of the failed
database instance on a different server. This recovery is initiated to ensure that the critical data
contained in the database will remain consistent and not become corrupted. 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, Dell recommends that users implement a fully redundant interconnect network
configuration, with redundant private NICs on each server and redundant private network
switches.8

Figure 3 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, as shown in the red colored cables that represent the Oracle
cluster private interconnect network. 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
independently 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.8 For
details on configuring NIC teaming, please refer to the “Configuring the Private NIC Teaming” in
the following section.

**Configuring Multiple Ethernet Interfaces for iSCSI Storage Area Networks**

For high availability and optimal performance, it is recommended that multiple NIC ports are used
on each host of the Oracle RAC to communicate with the PS Series iSCSI storage arrays. A
minimum of two NIC ports are required to provide redundant links to the storage arrays. This is
because in the event of a NIC port failure in a single NIC port environment, the host will lose
access to the storage until the failed NIC is physically replaced. As illustrated in Figure 3, in order
to provide sufficient host-side bandwidth to support three PS5000XV arrays with a two-node
Oracle 10g RAC database. Dell recommends using four NIC ports on each of the PowerEdge
servers hosting the Oracle 10g RAC database.

---

8 Dyke and Shaw, op. cit.
Software Configuration

Operating System Configuration

Configuring Host Access to iSCSI Volumes

As discussed in “Configuring Volumes” a number of volumes are created in the PS5000XV storage array to use as the Oracle storage space. In order to access these volumes, the Microsoft iSCSI initiator software needs to be installed and configured. The Microsoft iSCSI initiator enables the connection of a Windows® host to an external iSCSI storage array. During the installation of the Microsoft iSCSI initiator, ensure that the Initiator Service, the Software Initiator, and the Microsoft MPIO Multipathing Support for iSCSI installation options are all selected.

The MPIO multipath support feature of the Microsoft iSCSI initiator enables multiple I/O paths to iSCSI SAN storage. MPIO allows hosts to have multiple connected paths to the same SAN volume for increased redundancy and performance. However, with many hosts and multiple NICs per host, the configuration of iSCSI initiator to access volumes can be lengthy and tedious. To simplify this process, Dell has developed a Device Specific Module (DSM) to wrap around Microsoft’s MPIO capabilities. The Dell EqualLogic MPIO DSM takes the confusion out of connecting all available I/O paths from the server to the storage and automatically connects these paths based on the designated subnets. The Dell EqualLogic DSM consists of two components, a kernel mode component that works in conjunction with the Microsoft MPIO driver to route I/O to the desired path, and a user mode service that manages connections. The Dell EqualLogic MPIO DSM provides performance load balancing and high availability by helping remove single points of failure between the host and storage. The Dell EqualLogic MPIO DSM is easy to install as part of the Dell EqualLogic Host Integration Tool Kit. The Host Integration Tool Kit is included with the purchase of a PS Series array at no extra cost.

Once the Microsoft iSCSI initiator and the Dell EqualLogic MPIO DSM are installed, administrators can start configuring connections to a PS Series storage array. For detailed information on the configuration procedure, please refer to the Dell EqualLogic white paper “Configuring and Deploying the Dell EqualLogic Multipath I/O Device Specific Module (DSM) in a PS Series SAN” available at http://www.equallogic.com/psp/PDF/TR1036-MPIO_EQLX-DSM.pdf.

Configuring the Oracle Private Network NIC Teaming

As mentioned in the Section “Configuring Fully Redundant Ethernet Interconnects,” Dell recommends that users install two physical private NICs on each of the Oracle 10g RAC cluster servers to prevent private network communication failures. In addition to installing the two NICs, it is a requirement to use NIC teaming software to bond the two private network interfaces together to operate under a single IP address. 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 10g RAC database with no network communication interruption or changes to the private IP address.

Configuring Environment Variables

As a requirement for Oracle installation on Windows, both the TEMP and TMP environment variables must be set to the same directory path on all nodes in the cluster.\textsuperscript{10}

Configuring the Network Requirements on All Nodes

In order to install Oracle Clusterware, it is important to ensure all nodes within an Oracle 10g RAC cluster have the same network interface name for the public interface as well as the private interface.\textsuperscript{10} The public interface on each node of the 10g RAC cluster must be listed first in the bind order.\textsuperscript{10} Oracle supports the TCP/IP protocol for the public and private networks and requires that Windows Media Sensing is disabled by setting the value of the \textit{DisableDHCPMediaSense} parameter to 1.\textsuperscript{10}

An IP address with an associated public network name must be configured and registered in the DNS or entered in the local system host’s file on each node of the cluster. One virtual IP (VIP) address with an associated VIP network name must be registered in the DNS or entered in the local system host file on each node of the cluster. The VIP address must be on the same subnet as the public IP address. One private IP address with an associated private network name must be configured and entered in the local system host file on each node of the cluster.\textsuperscript{10}

It is important to verify the network configuration via ping tests prior to the Oracle installation. You should ensure the public IP address and the private IP address are resolvable by pinging from each node to any other nodes of the cluster. Oracle provides the Cluster Verification Utility (CVU) to verify the configuration of cluster at any stage during the installation process and also when the database is in production. The following command checks the general hardware and operating system configuration. You may use it to verify network connectivity between all nodes for which the cluster is configured.

\texttt{cluvfy stage -post hwos -n node_list [-verbose]}

Configuring Host Equivalence

During Oracle RAC software installation, software gets distributed by copying from local node to the remote nodes. All nodes within an Oracle 10g RAC cluster must have user equivalency for the Administrative privileges account that installs the database. This means that the administrative privileges user account and password must be the same on all nodes if you use the local Administrator account. If you have a domain user name, the domain user must be explicitly declared as a member of the local Administrators on all nodes in the cluster.\textsuperscript{10} You can use the following command to verify the host equivalence is set up correctly.

On node 1 issue the following command:

\texttt{net use * \<node2 name>\c$}

On node 2 issue the following command:

\texttt{net use * \<node1 name>\c$}

You may also use the following CVU command to check host equivalence configuration.

\texttt{10 “Oracle Clusterware and Oracle Real Application Clusters Installation Guide”, op. cit.}
cluvfy stage -post hwos -n node_list [-verbose]

Configuring Shared Storage for Oracle Clusterware using Raw Partitions

Before installing Oracle 10g RAC Clusterware software, it is necessary, at a minimum, for shared storage to be available on all cluster nodes for use by the Oracle Cluster Registry (OCR) and the Clusterware Cluster Synchronization Services (CSS) Voting Disk. As a prerequisite to Oracle installation, the write caching on all disks that will be used to share data between nodes in the Oracle cluster, must be disabled. As another prerequisite, disk automounting must be enabled on Windows 2003 servers hosting the Oracle RAC database, in order to correctly install Oracle Clusterware and the Oracle database. Oracle recommends enabling automatic mounting before creating any logical partitions for use by the Oracle Clusterware and database. Each node must be restarted after enabling disk automounting.

The OCR file and the CSS Voting disk file can be placed on a shared raw partition. As discussed in the Section “Configuring Volumes” above, one volume is created for the OCR file and the Voting Disk file. This volume is required to be a basic disk as dynamic disks are not supported with Oracle Clusterware. This volume should be configured as one extended raw partition with two raw logical drives each for the OCR file and the Voting Disk file. Users must ensure that the raw partition does not have a drive letter assigned and is not formatted during the partition creation.

Configuring Shared Storage for the Database Using the Automatic Storage Management (ASM)

As discussed in the preceding section, two separate volumes are created for the data storage area and the Flash Recovery Area, respectively. It is recommended that these two volumes be configured as ASM disks to benefit from the capabilities of ASM. During disk initialization, these two volumes are required to be configured as basic disks. Then one extended raw partition with a logical drive should be created on each of the two volumes. The raw partitions must not have drive letters assigned and must not be formatted during the partition creation. After the raw partition configurations, the two volumes must be manually stamped as ASM disks either by using the asmtool (GUI version) or using asmtool (command line version). The asmtool and asmtool commands associate meaningful, persistent names with disks to facilitate using those disks with Automatic Storage Management.

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). Each Oracle 10g 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.

You may use the following CVU command to check the shared storage accessibility.

cluvfy stage -post hwos -n node_list [-verbose]

Oracle 10g R2 Configuration

Installing Oracle Clusterware and Database Software

The preferred method to install Oracle Clusterware and Oracle Database is to use the Oracle Universal Installer (OUI). OUI provides a simple wizard-like installation mechanism to install Oracle Clusterware and database binaries on the Windows Server 2003. During the Clusterware and Oracle installation, the OUI will ask for general information such as paths for inventory directory and multi-node information. 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 10g R1 (10.2.0.1) Clusterware software.
2. Install Oracle 10g R1 (10.2.0.1) database software.
3. Upgrade Oracle Clusterware software to 10g R2 (10.2.0.3) patch set.
4. Upgrade Oracle database software to 10g R2 (10.2.0.3) patch set.
5. Create listener and cluster database.

Installing Oracle Process Manager (OPMD)

Upon Oracle database host server reboots, Oracle RAC systems on Windows platforms may experience delays in starting cluster related services. This is due to the Oracle Clusterware processes started by the operating system before the required storage or network devices are fully initialized, and which causes the Clusterware layer to fail to start up upon server reboots. To work around this issue, Oracle provides a utility OPMD.exe, which allows the creation of a new service called the Oracle Process Manager. The Oracle Process Manager will control the startup of the clustering services to prevent them from starting up before the rest of the operating system services have started. For more information, please refer to Oracle MetaLink notes #358156.1 at http://metalink.oracle.com.
## Configuration Deliverables List - Dell Solution for Oracle 10g on Dell EqualLogic PS5000XV iSCSI Storage

This section contains the Solution Deliverables List (SDL) for the Dell Solution for Oracle 10g on Dell EqualLogic PS5000XV iSCSI storage arrays. It contains a detailed listing of server and storage hardware configurations, firmware, driver, operating system, and database versions.

### Recommended Hardware/Software Requirements

<table>
<thead>
<tr>
<th>Validated Component(s)</th>
<th>Minimum Oracle RAC Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PowerEdge Nodes</strong></td>
<td></td>
</tr>
<tr>
<td>PowerEdge 2950 III</td>
<td>2</td>
</tr>
<tr>
<td><strong>Memory</strong></td>
<td></td>
</tr>
<tr>
<td>All valid PowerEdge 2950 III memory configurations</td>
<td>1GB (per node)</td>
</tr>
<tr>
<td><strong>PS Series Storage Array</strong></td>
<td>PS5000XV</td>
</tr>
<tr>
<td><strong>Ethernet Ports</strong></td>
<td></td>
</tr>
<tr>
<td>Intel or Broadcom Gigabit NICs</td>
<td>5</td>
</tr>
<tr>
<td><strong>Ethernet Switches for Oracle Private Interconnect Network</strong></td>
<td></td>
</tr>
<tr>
<td>Any Dell PowerConnect Gigabit-only Switches</td>
<td>2</td>
</tr>
<tr>
<td><strong>Ethernet Switches for iSCSI Storage Area Network</strong></td>
<td></td>
</tr>
<tr>
<td>Dell PowerConnect 5424 or 5448 Gigabit Switches Only with iSCSI optimization feature turned off</td>
<td>2</td>
</tr>
<tr>
<td><strong>Raid Controllers (Used for internal storage only)</strong></td>
<td></td>
</tr>
<tr>
<td>PERC 6/i</td>
<td>1 (Per Node)</td>
</tr>
<tr>
<td><strong>Internal Drive</strong></td>
<td></td>
</tr>
<tr>
<td>All valid PowerEdge 2950 III internal storage configurations</td>
<td>73 GB/node</td>
</tr>
<tr>
<td><strong>Oracle Software &amp; Licenses</strong></td>
<td></td>
</tr>
<tr>
<td>Oracle 10g R2 10.2.0.1 Enterprise or Standard Edition (Base) + Oracle Patchset 10.2.0.3</td>
<td>RAC</td>
</tr>
<tr>
<td><strong>Operating System</strong></td>
<td></td>
</tr>
<tr>
<td>Microsoft Windows Server 2003 R2 with SP2 Enterprise x64 Edition or Standard x64 Edition</td>
<td></td>
</tr>
<tr>
<td>PowerEdge Servers</td>
<td>Model</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------</td>
</tr>
<tr>
<td></td>
<td>PE2950 III</td>
</tr>
</tbody>
</table>

### Internal Disks RAID
- PERC 6/i: Firmware version = 6.0.1-0080; Driver version = 2.14.0.64
- SAS/SATA Backplane 0:0 Backplane: Firmware version = 1.0.5;

### Network Interconnect
- Intel PRO/1000 Network Drivers: Driver version = 9.9.12.0
- Broadcom NetXtreme II Gigabit Ethernet Driver: Driver version = 4.1.3.0
- Ethernet Teaming Driver: Driver version = 8.4.2.0

### iSCSI SAN Switches
- Dell PowerConnect 5424 / 5448 Gigabit Ethernet Switches: Firmware = v1.0.0.35

### iSCSI Storage
- Dell EqualLogic iSCSI Storage: PS5000XV; Firmware = V4.0.1

### Software
- Oracle: 10g R2 10.2.0.3 Enterprise Edition x86_64 or Standard Edition x86_64
- Operating systems: Microsoft Windows Server 2003 R2 with SP2 Enterprise x64 Edition or Standard x64 Edition
- iSCSI Initiator: Microsoft iSCSI Initiator 2.0.7
- Dell EqualLogic Host Integration Tool Kit: 3.1.1

#### Table 3 – Dell 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](http://support.dell.com)
Conclusion

Dell Solutions for Oracle Database 10g are designed to simplify operations, improve usage and cost-effectively scale as your needs grow over time. This reference configuration white paper provides a blueprint for setting up an Oracle 10g RAC database on Dell EqualLogic iSCSI storage arrays and Dell PowerEdge servers. Although we used a two-node RAC as an example, the deployment method applies to multiple nodes on the RAC configuration.

The best practices described here are intended to help achieve optimal performance of Oracle 10g on Microsoft Windows Server 2003. To learn more about deploying Oracle 10g RAC on Dell storage systems and PowerEdge servers, please visit www.dell.com/oracle or contact your Dell representative for the most current information on Dell servers and storage and services for Oracle 10g solutions.
Tables and Figures

Table 1 - Volumes for the Oracle RAC Configuration ................................................................. 10
Table 2 – Dell Solution Minimal Hardware/Software Requirements ......................................... 18
Table 3 – Dell Solution Detailed Firmware, Driver and Software Versions ................................ 19

Figure 1 - Architectural Overview of Oracle 10g on Dell EqualLogic PS5000XV iSCSI Storage ... 6
Figure 2 - Recommended Dual Control Module Network Configuration ................................... 7
Figure 3 - Cabling Three EqualLogic PS5000XV Storage Arrays with Two PowerEdge 2950 III
Servers ........................................................................................................................................... 8

References


2. “Oracle Clusterware and Oracle Real Application Clusters Installation Guide”, 10g Release 2
   (10.2) for Microsoft Windows, B14207-06.
   http://download-west.oracle.com/docs/cd/B19306_01/install.102/b14207.pdf


   Dell EqualLogic white paper.

5. “PS Series Groups: Storage Array Network Performance Guidelines”, a Dell EqualLogic white
   paper.

6. “PS Series Best Practice: Configuring and Deploying the Dell EqualLogic Multipath I/O
   Device Specific Module (DSM) in a PS Series SAN”, a Dell EqualLogic white paper.