Best Practices for Deployment of Oracle® ASM with Dell™ EqualLogic™ PS iSCSI Storage System

Database Solutions Engineering

By

Raghunatha M

Ravi Ramappa

Chen Wendy

Dell Product Group
July 2009

Executive Summary

Storage planning and management are major infrastructure management components in large database deployments. Given that data growth is astronomical in today's environment, it is very important to understand the storage system's ability to meet future storage needs. This paper outlines the different options available to expand the storage capacity in Oracle databases environments that are deployed on PS EqualLogic™ storage and Oracle ASM. It provides examples of the various storage capacity expansion options available for the PS storage that complement Oracle ASM for Windows® and Linux deployments. It also explores the different options available for multipath configuration and provides a comparative analysis of those options for Windows and Linux deployments.

Best practices for deployment of Oracle ASM with Dell EqualLogic™ PS iSCSI storage System

Contents

Introduction

Dell EqualLogic PS Series iSCSI storage arrays are built on a patented peer storage architecture where all arrays in a storage pool are designed to work together to provide the most disk capacity and evenly distribute the load. This architecture enables the arrays to continuous monitor resource and performance, and automatically moves data pages to maximize performance. This combination of features provides simplified storage deployment, provisioning and management. With EqualLogic's unique peer storage architecture, desired performance can be sustained with storage expansion.

Oracle® Automatic Storage Management (ASM) is a feature of the Oracle database that provides vertical integration of the file system, and a volume manager specifically built for the Oracle database files. ASM distributes the I/O load across all available resources to optimize performance, while removing the need for manual I/O tuning. ASM helps DBAs manage a dynamic database environment by allowing them to grow the database size without any database downtime for storage allocation modification. Oracle ASM is easier to manage than conventional file systems, and is designed to provide optimal performance for Oracle database deployments.

Multipathing software is required to provide the storage access redundant paths, and prevents a single point of failure situation. Oracle ASM is a complementary solution to multipathing technologies. ASM does not have any dependencies on host multipathing software, and works transparently in most OS environments.

This white paper outlines the best practices for Oracle database deployment with Oracle ASM and multipathing on Dell EqualLogic storage arrays. Storage capacity management, which includes volume resizing and thin provisioning, is also explained in detail in this paper. The deployment steps and procedures to perform capacity management without any database downtime are explained in detail. Also this paper outlines how to configuring multipathing on Dell EqualLogic PS storage for Windows and Linux. All available multipathing software load balancing policies are evaluated, and best practices are recommended.

Overview of Oracle ASM Technology

In the 10g release, Oracle introduced a storage feature called ASM that provides the database administrator with a simple storage management interface that is consistent across all server and storage platforms. This feature allows a DBA to save time with respect to storage management, as the performance of asynchronous I/O is made easier with file system management and dynamic database environments can be managed efficiently without manual intervention.

ASM provides redundancy by mirroring against disk failures. ASM stripes data evenly across a number of disks in the disk group, and maintains a uniform distribution of extents as the storage configuration changes.

Overview of EqualLogic Storage Technology

iSCSI storage is growing in popularity because of the ability to integrate easily into existing Ethernet infrastructure, and their ease-of-management. All Dell EqualLogic PS Series iSCSI storage arrays are built 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. Individual PS Series arrays that are part of a group are called members, and each member is assigned to a storage pool. A group can contain more than one member with a different capacity and can be configured to use different RAID levels. These characteristics, along with the ability to move volumes into specific pools without any interruption in service, can be used to implement tiered storage without disrupting current operations. PS Series iSCSI storage systems allow storage administrators to expand storage capacity without impacting data availability.

Storage Capacity Management

With the rapid growth of the size of databases, storage capacity expansion has become a frequent maintenance task for administrators. As the demand for uptime continues to increase, it has been increasingly difficult to schedule database downtime in order to perform storage configuration. This section discusses the several different options available for allocating additional Dell EqualLogic storage resources as the database capacity needs increase. It provides detailed procedures of each option, as well as how database availability and performance will be impacted. The examples provided in this section are based on the assumption that the storage volumes to be extended are part of ASM disk groups. The expansion of volumes used for the Oracle RAC quorum disks, such as OCR and CSS, are not covered here. The examples are also based on the assumption that there is only one partition on the volume to be extended.

Storage Expansion via Oracle ASM

With the introduction of Oracle ASM, administrators can expand storage capacity without having to take the database offline; additional ASM disks can be added to an existing ASM disk group for increased storage capacity. It is a best practice that all ASM disks belonging to the same disk group are similarly

sized. Therefore, when adding new ASM disks to an existing disk group, it is recommended to configure the new disks with similar sizes as the existing ASM disks of the same disk group.

Adding or dropping ASM disks will automatically trigger ASM to rebalance the I/O that spreads data across all available disks in the disk group for optimal performance and utilization; this rebalancing involves physical movement of file extents. The impact to the online database activity is generally low, because the rebalance is done at one allocation unit at a time. However, Oracle recommends that you schedule the rebalance operation during off-peak hours.

The throughput and the speed of the rebalance operation are influenced by the ASM instance initialization parameter `asm_power_limit` that can be set with values ranging from 0 to 11. The higher the `asm_power_limit`, the faster the rebalancing will complete but with some overhead on performance. Lower values will take longer, but consume fewer processing and I/O resources. An `asm_power_limit` value of 0 turns off the automatic rebalance, and should only be used with caution. The `asm_power_limit` can also be set for a specific rebalance activity using the `alter diskgroup` command. To add ASM disks, following command is used:

```
alter diskgroup <diskgroup> add disk <disk name> rebalance power <1-11>;
```

In the following example, an additional ASM disk is added into the disk group named DATABASEDG, with a rebalance power of 11:

```
SQL> alter diskgroup databasedg add disk 'ORCL:DATA5' rebalance power 11;

Diskgroup altered.
```

The status of the rebalance operation can be queried from the `V$ASM_OPERATION` view as follows:

```
SQL> select * from v$asm_operation;

GROUP_NUMBER OPERA STAT     POWER     ACTUAL      SOFAR    EST_WORK   EST_RATE EST_MINUTES
------------ ----- ---- ---------- ---------- ---------- ---------- ---------- -----------
           1 REBAL RUN          11         11         39     104231        736         141
```

Storage Expansion via Dell EqualLogic Volume Resizing

Dell EqualLogic storage systems allow administrators to increase the size of a volume with an increase in data. To recognize the new volume size, hosts and applications that are currently connected to a volume must support file system expansion. Oracle supports the physical disk expansion. In Metalink document 311619.1 located at the following URS: https://metalink.oracle.com, Oracle documents the

Best practices for deployment of Oracle ASM with Dell EqualLogic™ PS iSCSI storage System

detailed steps required to resize a physical disk at the OS level, and subsequently resize the ASM disk group using the new space. When resizing the physical disks of an ASM disk group, Oracle recommends resizing all of the disks in the disk group to the same size at the same time.

Before extending physical disks, it is important to backup all data on the disk drive. As documented in the Metalink document, the general steps to expand the physical disk and ASM disk group are:

- Resize LUN from storage.
- Pick up new size from OS. A server reboot might be required for OS to recognize the new LUN size or new partition size.
- In rolling fashion, shutdown ASM instance.
- Only when all ASMs have been recycled, resize the ASM disk in one ASM instance.

The Dell engineers tested the expansion of Dell EqualLogic volumes for Oracle databases on Microsoft® Windows Server 2008 with SP2, as well as on Red Hat® Enterprise Linux® 5 Update 3; both of these operating systems support the file system expansion. The procedure is detailed in the following sections.

Resizing Physical Disk and ASM Disk group on Microsoft Windows Server 2008 with SP2

The Oracle database used in this test was a two node Oracle 11.1.0.7 Real Application Cluster (RAC) database running on Microsoft Windows Server 2008 with SP2. Disk expansion was performed on a 10 GB Dell EqualLogic volume on which an ASM diskgroup named DATABASEDG1 was created and populated with data. In this example, the 10 GB Dell EqualLogic volume was formatted as a MBR disk with one extended partition and one logical drive of 10 GB in size.

Increasing the Size of a Dell EqualLogic Volume

The size of a Dell EqualLogic volume can be increased from either the Dell EqualLogic Group Manager Administration GUI tool, or from the Command Line Interface (CLI). In the testing, the 10 GB Dell EqualLogic volume was increased to 12 GB in size, from the Group Manager GUI tool.

From the Group Manager, select **Volumes** -> **volume_name** -> **Modify volume settings**- > **Space tab.** The **Modify Volume Settings – Space** dialog box appears. In the Volume size field, enter the new size.

Alternatively, the following CLI command increases the size of a volume:

```
volume select vol_name size [no-snap] new_size
```

Extending a Raw Logic Volume

As documented in Oracle MetaLink note 555273.1, logical volumes in Windows can be extended using the `diskpart` utility.

Best practices for deployment of Oracle ASM with Dell EqualLogic™ PS iSCSI storage System

1. Launch the disk management GUI by entering `diskmgmt.msc` in the **Start** -> **Run** window.
2. Run **Rescan Disks** from the **Action** menu in `diskmgmt` on each node. For this example, we worked with Disk 3 that had 2.01 GB of unallocated space as is displayed in Figure 1.
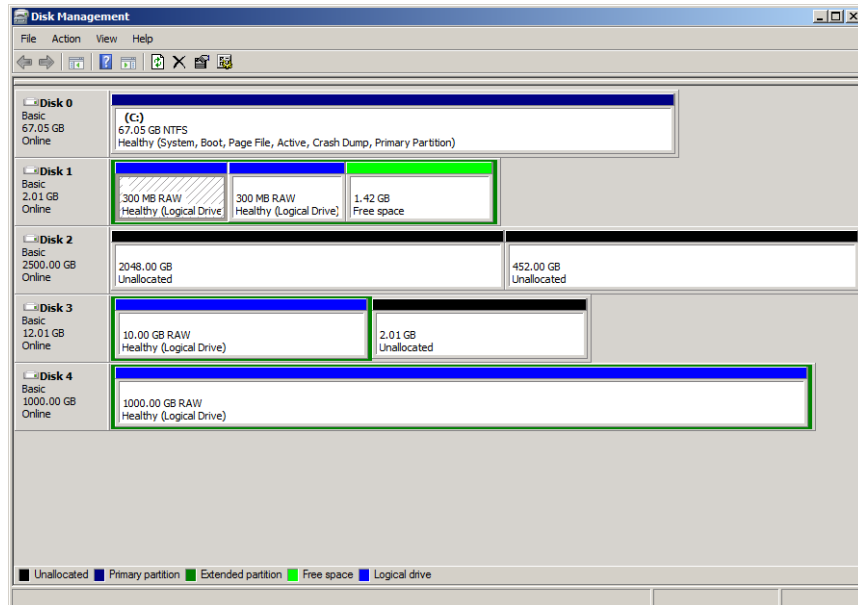
Figure 1: Disk Management for Disk 3 with 2 GB of unallocated space.

3. Launch the `diskpart` utility by entering `diskpart` in the command line.
4. Select the disk to be extended using the `select disk` command in `diskpart`.

```
C:\Users\Administrator>diskpart

Microsoft DiskPart version 6.0.6002
Copyright (C) 1999-2007 Microsoft Corporation.
On computer: W-M610-W1

DISKPART> select disk 3
```

Disk 3 is now the selected disk.

```
DISKPART> list disk

  Disk ###   Status        Size      Free     Dyn  Gpt
  --------   ----------    -------   -------   ---  ---
  Disk 0     Online          67 GB      0 B
  Disk 1     Online        2055 MB   1452 MB
  Disk 2     Online        2500 GB   2048 GB
* Disk 3     Online          12 GB   2054 MB
  Disk 4     Online        1000 GB      0 B
```

5. Display the partitions using the `list part` command in `diskpart`.

```
DISKPART> list part

  Partition ###  Type              Size     Offset
  -------------  ----------------  -------  -------
  Partition 0    Extended          10 GB    1024 KB
  Partition 1    Logical           10 GB    2048 KB
```

6. Select the logical volume by using the `select partition` command in `diskpart`. In this example, the logical volume was partition 1.

```
DISKPART> select partition 1
```

Partition 1 is now the selected partition.

```
DISKPART> list part

  Partition ###  Type              Size     Offset
  -------------  ----------------  -------  -------
  Partition 0    Extended          10 GB    1024 KB
* Partition 1    Logical           10 GB    2048 KB
```

7. Extend the selected partition by using the `extend size=<N>` command in `diskpart`, where N specifies the amount of space in Megabytes (MB) to add to the current volume or partition. If no size is given, all of the contiguous free space that is available on the disk is used.

```
DISKPART> extend
```

`diskpart` successfully extended the volume.

8. Verify the new partition size using the `list part` command in `diskpart`.

```
DISKPART> list part

  Partition ###  Type              Size     Offset
  -------------  ----------------  -------  -------
  Partition 0    Extended          12 GB    1024 KB
* Partition 1    Logical           12 GB    2048 KB
```

9. In each remaining Oracle RAC database server node, run **Rescan Disks** from disk management. These steps will likely cause Windows to automatically assign drive letters to the newly extended partition. Be sure to remove the drive letters using the **Change Drive Letter and Paths …** from disk management.
10. Restart the remaining cluster nodes in order to recognize the new disk partitions.

Best practices for deployment of Oracle ASM with Dell EqualLogic™ PS iSCSI storage System

Resizing an ASM disk

For Oracle release 10.2.0.3 and earlier, ASM instances need to be restarted in order to see the new extended LUN size. In Oracle release 11g, ASM can detect the new LUN sizes without restarting the ASM instance.

1. Shutdown the database instances and ASM instances except for one.
2. Log into the one ASM instance and enter the following query. The output shows that the ASM disk has 10 GB allocated, and its OS recognized size is 12 GB.

   ```
   SQL> select name, os_mb, total_mb from v$asm_disk;

   NAME                               OS_MB    TOTAL_MB
   ---------------------------- ---------- ----------
   DATABASEDG1_0000                   12297      10242
   ```

3. Enter the following query in ASM instance to resize the ASM disk:

   ```
   alter diskgroup <diskgroup> resize disk <disk name> <size>M rebalance power <1-11>;
   ```

   For example,

   ```
   SQL> alter diskgroup databasedg1 resize disk databasedg1_0000 size 12297 M
   rebalance power 11;

   Diskgroup altered.
   ```

4. Verify that the new ASM disk size is recognized.

   ```
   SQL> select name,os_mb,total_mb from v$asm_disk;

   NAME                               OS_MB    TOTAL_MB
   ---------------------------- ---------- ----------
   DATABASEDG1_0000                   12297      12297
   ```

5. Restart the ASM and database instances on the remaining nodes.

Resizing Physical Disk and ASM Diskgroup on Red Hat Enterprise Linux 5 Update 3

 The Oracle database used in this test was a two node Oracle 11.1.0.7 Real Application Cluster database running on Red Hat Enterprise Linux 5 Update 3. Disk expansion was performed on a 10 GB Dell EqualLogic volume on which an ASM diskgroup named DATABASEDG1 was created and populated with data. In the OS, the 10 GB Dell EqualLogic volume was formatted with one 10 GB primary partition.

Best practices for deployment of Oracle ASM with Dell EqualLogic™ PS iSCSI storage System

Increasing the Size of a Dell EqualLogic Volume

The steps to increase the size of a Dell EqualLogic volume are the same regardless of the host OS. Follow the steps in the "Increasing the Size of a Dell EqualLogic Volume" section of "Resizing Physical Disk and ASM diskgroup on Microsoft Windows Server 2008 with SP2" above.

Extending a Linux Partition

To resize a LUN in Linux, the high level steps involved are to delete the current partition using `fdisk` or `parted`, and then recreate the partition with the new size.

1. Detect the new LUN size using the `sfdisk -R` command on all host devices in the cluster. If multiple devices exist for the same LUN due to multipathing, run this command on all devices. For our example, the 10 GB Dell EqualLogic volume appears as `/dev/sdv` and `dev/sdw`:

   ```
   $ sfdisk -R /dev/sdv

   $ sfdisk -R /dev/sdw

   $ cat /proc/partitions
   major minor  #blocks  name
     65    80   12595200 sdv
     65    81   10490864 sdv1
     65    96   12595200 sdw
     65    97   10490864 sdw1
   ```

2. Recreate the partition table by using the `fdisk` or `parted` utility depending on which utility was used to create the original partition. As documented in Oracle MetaLink note 311619.1, recreating the partition occurs at the partition table level, and that table is stored in the first sectors of the disk. Changing the partition table does not affect the data. The example below used `fdisk` to drop and recreate the only partition on the LUN.

   ```
   $ fdisk -l /dev/sdv

   Disk /dev/sdv: 12.8 GB, 12897484800 bytes
   64 heads, 32 sectors/track, 12300 cylinders
   Units = cylinders of 2048 * 512 = 1048576 bytes

   Device Boot      Start         End      Blocks   Id  System
   /dev/sdv1            1       10245    10490864   83  Linux

   $ fdisk /dev/sdv
   ```

   The number of cylinders for this disk is set to 12300, which could in certain situations cause issues with the following since it is a larger value than 1024:

- software that runs at boot time (e.g., old versions of LILO)
- booting and partitioning software from other OSs (e.g., DOS FDISK, OS/2 FDISK)

```
Command (m for help): d
Selected partition 1

Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First cylinder (1-12300, default 1):
Using default value 1
Last cylinder or +size or +sizeM or +sizeK (1-12300, default 12300):
Using default value 12300

Command (m for help): w
The partition table has been altered!

Calling ioctl() to re-read partition table.
Syncing disks.
```

3. Verify the size of the new partition.

   $ fdisk -l /dev/sdv

   ```
   Disk /dev/sdv: 12.8 GB, 12897484800 bytes
   64 heads, 32 sectors/track, 12300 cylinders
   Units = cylinders of 2048 * 512 = 1048576 bytes

      Device Boot      Start         End      Blocks   Id  System
   /dev/sdv1               1       12300    12595184   83  Linux
   ```

4. Detect the new partition size using the `sfdisk -R` command on all host devices in the cluster.

   $ sfdisk -R /dev/sdw

   $ cat /proc/partitions
   ```
   major minor  #blocks  name
     65    80   12595200 sdv
     65    81   12595184 sdv1
     65    96   12595200 sdw
     65    97   12595184 sdw1
   ```

5. For the device mapper multipathing driver to recognize the new partition size, reboot the servers in the cluster.
6. Verify the correct size of the partition is displayed in the `multipath –ll` output:

```
$ multipath –ll
tplinux-data4 (36090a028706522196ba3741d7fc7e365) dm-12 EQLOGIC,100E-00
[size=12G][features=0][hwhandler=0][rw]
\_ round-robin 0 [prio=0][active]
 \_ 8:0:0:0  sdae 65:224 [active][ready]
 \_ 7:0:0:0  sdaf 65:240 [active][ready]
```

Resizing an ASM disk

The steps to resize an ASM disk are the same regardless of the host OS. Follow the steps in the "Resizing an ASM disk" section of "Resizing Physical Disk and ASM diskgroup on Microsoft Windows Server 2008 with SP2" above.

Storage Expansion via Dell EqualLogic Thin Provisioned Volumes

In today's enterprise environment, excessive amounts of storage capacity are over-allocated to applications to meet volatile growth expectations and avoid future upgrade complexity. This has caused low storage utilization rates, and drives up the capital and related operating expense (power and floor space) associated with keeping unused storage. To help meet this challenge, the Dell EqualLogic PS Series storage systems provide thin provisioning technology that assigns virtual storage volumes to applications and consumes physical capacity on demand. Thin provisioning allows administrators to more efficiently provision storage, while still meeting the application and user storage needs by allocating only the storage space that is required. With thin provisioning, businesses can purchase less storage capacity up front, perform storage capacity upgrades in line with actual business usage, and save operating costs. Thin provisioning helps to increase utilization of storage resources. Since the cost of storage devices continues to decline, thin provisioning helps to reduce the total device purchase cost by making small incremental additions instead of large capital expenditures.

In contrast to a traditional physically-provisioned volume, a thin provisioned volume is initially allocated with only a portion of the volume size. As data is written to the volume, more space is automatically allocated (if available), and the volume reserve increases up to a user-defined limit. Regular event messages are generated as space is used, giving the administrator the opportunity to make adjustments, as needed. Thin provisioning is most effective when the volume growth pattern is predictable, and users do not need immediate, guaranteed access to the full volume size. Administrators must monitor thin provisioned volumes, and be prepared to increase storage space by adding or moving arrays to the pool where the volume resides.

In a Dell EqualLogic PS Series group, there are four user-defined parameters that control space utilization in a thin provisioned volume:

- Volume size – size that the group reports to iSCSI initiators. The system connected to the volume will associate this amount of space with the volume. The volume size is determined by the administrator based on the current and future maximum sizes needed by applications. This volume size can be increased in the same manner as traditional physically-provisioned volumes.
- Minimum volume reserve – minimum amount of pool space that the group will allocate to the volume; the default is 10% of the volume size. When the free volume reserve is less than 6 GB, the group will attempt to allocate additional pool space to the volume reserve in increments of 10 GB, depending on the amount of free pool space. If there is not enough free pool space to increase the volume reserve, the volume reserve will not increase. If in-use space consumes all of the volume reserve, the volume will be taken offline. If free pool space becomes available allowing the volume reserve to increase automatically, the volume will be automatically brought online.
- In-use space warning value – when in-use space reaches this percentage of the volume size, a warning event message is generated. The default is the group-wide setting of 60% of volume size. After the in-use warning value is reached, additional warning messages are generated according to the following policy:
    - For volumes larger than 200 GB, additional warnings are generated when the in-use volume reserve increases by every additional 10 GB.
    - For volumes smaller than 200 GB, additional warnings are generated when the in-use volume reserve increases by every additional 5%.
- Maximum in-use space – maximum volume size percentage that can be in use. The default is the group-wide setting of 100% of volume size. If the in-use space reaches the maximum allowed, the volume will be taken offline. If pool space is available and the volume size or the maximum in-use space setting is increased, the volume will be automatically brought online.

Dell EqualLogic Thin Provisioning and ASM Diskgroup

When a thin provisioned Dell EqualLogic volume is presented to an Oracle database 11.1.0.7 on Microsoft Windows 2008 and Red Hat Enterprise Linux 5, both the OS and Oracle associates its defined volume size as the size of the volume, even in situations where the actual physical capacity of the storage system is less than the defined volume size. As the Oracle storage needs grow, additional Dell EqualLogic arrays can be added into the pool where the thin provisioning volume resides. The thin provisioned volume helps to reduce or eliminate the Oracle database downtime spent dealing with storage capacity expansion.

Multipath and Oracle Automatic Storage Management

Multipathing software is implemented at the OS device driver layer by creating an abstraction layer that allows I/O operations to be shared across all available I/O paths. If any one component fails, an application still can access their data thereby eliminating a single point of failure. Multipathing solutions can also improve system performance by distributing I/O across all available paths. It provides a higher level of data availability through automatic failover and failback with Active-Active (all path actively doing I/O) or Active-Passive (one path active the other standby) configurations.

Oracle ASM does not have multipathing functionality, but is a complementary solution to Multipathing technologies. ASM does not have any dependencies on host multipathing software and works transparently in most OS environments. Oracle ASM encounters an error when it discovers the same disk using multiple paths, and a single disk can have multiple device names in multipath configuration. Therefore, Oracle ASM should be configured with the discovered pseudo device names. When I/O is requested to the pseudo device, the multipath driver intercepts it, provides load balancing and failover capabilities to the other existing paths.

Dell EqualLogic Multipath Solution

Microsoft has MPIO multipathing software that supports multiple I/O paths from a server to SAN storage. For increased redundancy and performance, Microsoft MPIO allows servers to have multiple connected paths to the same volume in the SAN. Using Microsoft MPIO, users can configure each server network port to the SAN one at a time. This process can become complex as the number network port increases.

Dell EqualLogic DSM on Windows

To fully utilize Microsoft's MPIO capabilities, Dell EqualLogic has DSM (Device Specific Module) that connects all the available I/O paths from server to storage based on designated IP subnets.

The Dell EqualLogic DSM has two components:

1. A kernel mode component that works in conjunction with the Microsoft MPIO driver to route I/O to the desired path.
2. A user mode service that manages the connections. The connection manager makes the appropriate connections automatically to the target volume when a single connection is made.

Above two components help administrators to easily install and configure multipath for iSCSI networks. The MPIO DSM is part of the Host Integration Tool Kit, can be downloaded without extra cost at www.support.dell.com/EqualLogic .

Best practices for deployment of Oracle ASM with Dell EqualLogic™ PS iSCSI storage System

MPIO DSM Load-Balance Policy

Microsoft MPIO DSM allows the initiator (server) to login multiple sessions to the same target (storage), and then aggregate that into a single device. Multiple target sessions can be established using different NICs to the target ports. If one of the sessions fails, then another session continues to process I/O without interrupting the application.

Dell EqualLogic MPIO DSM supports following balancing policies.

- **Fail Over Only**: Data is sent in one path, while other paths are standby. This connection is used for routing data until it fails or times out. If the active connection fails, then one of the available paths is chosen until the former is available. This load balancing policy is the default configuration when MPIO DSM is disabled.
- **Round Robin**: All available paths are used to perform I/O in a rotating sequence (round robin sequence). There is no disruption in sending I/O even if any of the paths fails. Using this policy, all paths are used effectively.
- **Least Queue Depth**: I/O is sent to the path that has least queue length.

The performance analyses for the above load balancing policies are presented in the following sections.

Test Configuration

To evaluate the Dell EqualLogic PS iSCSI storage host-side load balancing policies, a 2-node Oracle 11g R1 (11.1.0.7) RAC cluster was deployed. PS EqualLogic storage was configured with RAID 10 layout. Table 2 below has the hardware and software component details used in the test configuration.

Hardware	Components	
	Server	2xR710
	Storage	4xPS6000XV (1 Storage Pool)
	Network Switch	Power Connect 5448
Software		
	OS	RHEL 5U3
	Oracle RDBMS (with RAC)	EE 11g R1 (11.1.0.7)
	Microsoft Initiator	2.0.8
Database		
	Database Volume Size	200 GB
Benchmark Tools		
	ORION	Version 11.1.0.7

Table 2- Hardware and Software Component List

Best practices for deployment of Oracle ASM with Dell EqualLogic™ PS iSCSI storage System

Test Cases

To evaluate the Dell EqualLogic PS iSCSI storage MPIO DSM load balancing policies, an iSCSI initiator was configured with one load balancing policy at a time. Benchmark Factory TPC-H workload was used to simulate the OLAP workload. The OLAP workload processes a lot of data with large I/O size (1MB), and generates heavy iSCSI traffic on all available paths. For these types of workloads, throughput (measured in megabytes per second) is the critical measure for performance.

Analysis

Different performance metrics (such as latency, throughput) were captured for the OLAP workload from Benchmark Factory TPC-H runs. These metrics were analyzed to determine the optimal configuration. Figure 2 and 3 contain the detailed findings.
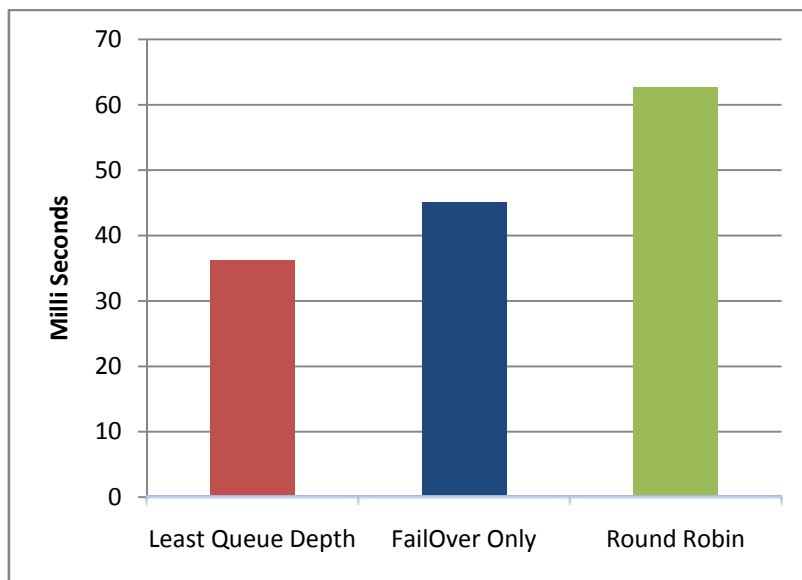

Figure 2: Load Balancing Policies Latency

Figure 2 displays the latency for different load balancing policies. From the above figure we can observe that the "Least Queue Depth" policy provides the least amount of OLAP workload latency.
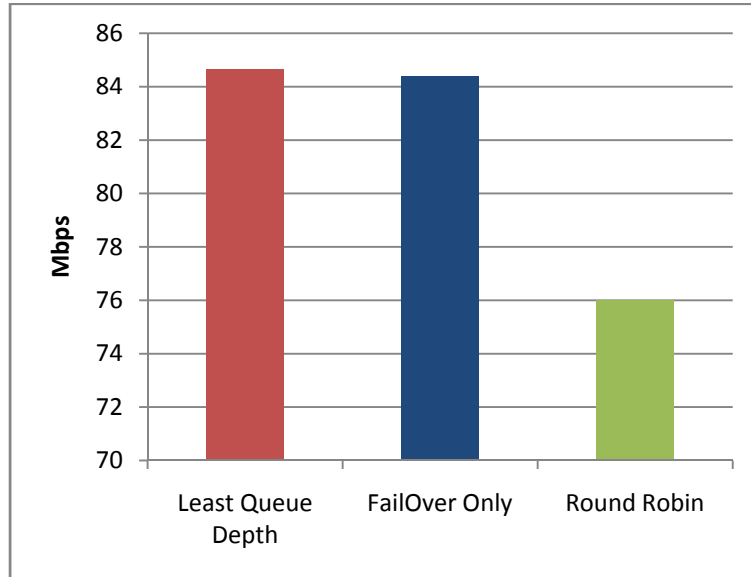
Figure 3: Load Balancing Policies Throughput

Figure 3 shows throughput in Mbps for different load balancing policies. From the above figure we can observe that the "Least Queue Depth" load balancing policy has better throughput than the other load balancing policies. From the above analysis we can conclude that for Windows configurations the "Least Queue Depth" policy provides more throughputs and less latency when compared to the other policies. For Windows the "Least Queue Depth" policy is the recommended configuration.

On a side note, we can observe from Figure 2 and 3 that the "Failover Only" policy provides better performance than the "Round Robin" policy. But when the iSCSI network is saturated, the "Round Robin" policy may outperform the "Failover Only" policy, given that the former uses all available paths where the later uses only half of the available paths.

DM-Multipath for EqualLogic on Linux

With a RHEL 5 Linux configuration, device mapper multipathing (DM-Multipath) is used to configure multiple I/O paths between the server and the EqualLogic PS iSCSI storage arrays. DM-Multipath provides a way of organizing the I/O paths logically, by creating a single multipath device on top of the underlying devices.

Configuration

DM-Multipath devices can be seen in three different places under the `/dev` directory:

- `/dev/mapper/mpathn`: the devices in `/dev/mapper` are created in the boot process. These devices should be used to access the multipath devices.

Best practices for deployment of Oracle ASM with Dell EqualLogic™ PS iSCSI storage System

- `/dev/mpath/mpathn`: the devices in `/dev/mpath` are provided as a convenience, so that all multipath devices can be seen in one directory. These devices are created by the `udev` device manager and may not be available on startup when the system needs to access them.
- `/dev/dm-n` : any devices in the form `/dev/dm-n` are for internal use only .

Each multipath device has a World Wide Identifier (WWID); by default, the name of a multipath device is set to its WWID in the `multipath.conf` file. You can set the `user_friendly_names` option in the multipath configuration file that will set the alias to a node-unique name in the form `mpathn`.

The `path_grouping_policy` parameter in the `multipath.conf` file determines how the different available paths are used by the server while accessing the storage. The different path grouping policies available are failover and multibus. The path grouping policy can be configured either for individual volumes, or for all storage volumes. The performance analysis for the load balancing policies is presented in the following sections.

Test Configuration

To evaluate the Dell EqualLogic™ PS iSCSI storage host-side path grouping policies, a 2-node Oracle 11g R1 (11.1.0.7) RAC cluster was deployed. PS EqualLogic storage was configured with RAID 10 layout. Table 3 below has the hardware and software components details used in the test configuration.

Hardware	Components	
	Server	2xR710
	Storage	4xPS6000XV (1 Storage Pool)
	Network Switch	Power Connect 5448
Software		
	OS	RHEL 5U3
	Oracle RDBMS (with RAC)	EE 11g R1 (11.1.0.7)
	Multipathing	Device-Mapper Multipathing
Database		
	Database Volume Size	500 GB
Benchmark Tools		
	ORION	Version 11.1.0.7

Table 3: Hardware and Software Component List

Test Cases

To evaluate the DM-Multipath path grouping policies (failover and multibus) performance, test runs were carried out for each path grouping policy. The ORION tool was used to simulate the OLAP workload. The OLAP workload processes a lot of data with a large I/O size (1MB), and generates heavy iSCSI traffic on all available paths. For these types of workloads, the throughput measured in megabytes per second is the critical measure for performance.

Best practices for deployment of Oracle ASM with Dell EqualLogic™ PS iSCSI storage System

Analysis

Performance metrics, throughput in Mbps, were captured for the two path grouping policies and a comparison was done as shown in Figure 4.
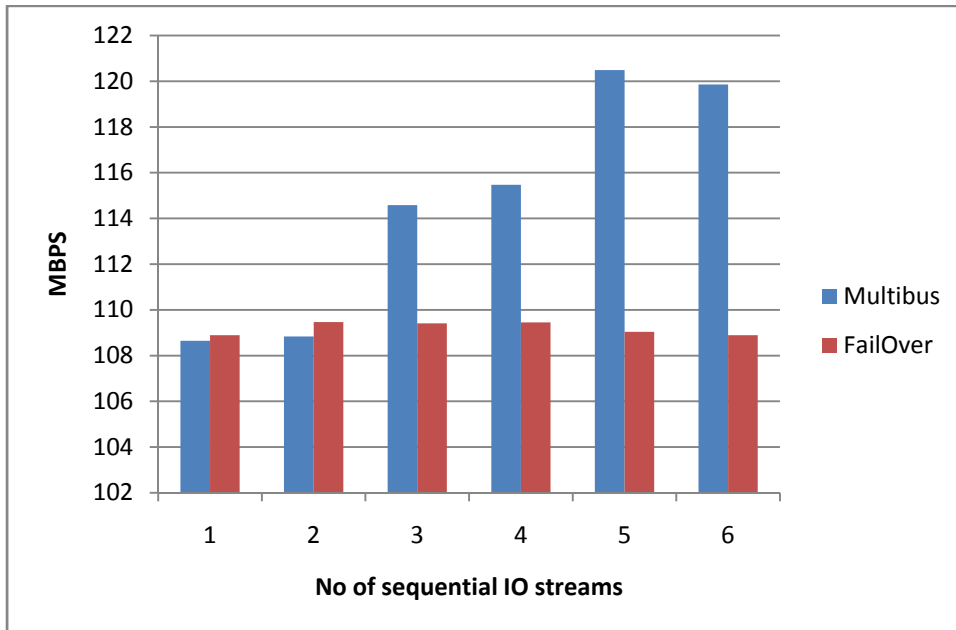

Figure 4: Path Grouping Policies Throughput

In Figure 4, the X-axis represents the number of sequential I/O streams and the Y-axis represents the throughput in Mbps. For the OLAP workload, the multibus path grouping policy delivers higher throughputs when compared to the failover path grouping policy. In case of the multibus path grouping policy, all paths are used for I/O in round robin fashion. Therefore setting the `path_grouping_policy` to `multibus` provides the largest bandwidth, and is the recommended load balancing policy for Linux deployments.

The detailed steps for configuring EqualLogic PS host access is described in the following document: http://support.dell.com/support/edocs/softWARE/Appora11/lin_x86_64/en/stnw/stnw_en.pdf

For Linux, when using ASMLIB with Oracle ASM, the discovery of the multipath disk is ensured by configuring ASMLIB to scan the multipath disk first from storage. For Dell EqualLogic, the scan order for Oracle ASM should be 'DM', and is accomplished by specifying the variables in the Oracle ASM configuration file.

Conclusion

Oracle ASM greatly reduces the administrative tasks associated with managing Oracle database files. It is a fully-integrated, host-level file system and volume manager for Oracle database files. Oracle ASM external redundancy disk groups with volumes protected by Dell EqualLogic storage RAID policies provide reliable, performance-tuned storage for Oracle databases. Oracle ASM provides a simplified storage management interface for DBAs, while Dell EqualLogic PS arrays provide easy-to-use storage provisioning and management. These combined technologies enable simplified provisioning and management of the storage resources for the database.

The following are the recommended best practices for deploying an Oracle database with Oracle ASM and Dell EqualLogic storage arrays:

- For optimal performance, create one data volume per storage pool. Also, use multiple storage pools in order to leverage both EQL and ASM load balancing.
- Configure external redundancy for ASM disk groups.
- Use the thin provisioning feature to reduce initial storage costs, and seamlessly provision storage without downtime.
- Use Microsoft DSM with "Least Queue Depth" load balancing policy on Windows, and "multibus" path grouping policy on Linux, for optimal performance and high availability.

To fully utilize Microsoft's MPIO capabilities, Dell EqualLogic provides MPIO DSM that is complementary to ASM for both high availability and performance.

The test results demonstrated that for Windows configurations the "Least Queue Depth" policy provides more throughputs and less latency when compared to the other policies. Therefore for Windows configurations, it is recommended to use the "Least Queue Depth" load balancing policy. For Linux deployments with the "multibus" path grouping policy, all the I/O paths are used in round robin fashion giving the largest bandwidth. Therefore for Linux configurations, it is recommended to use the "multibus" path grouping policy.

References

1. Automatic Storage Management.
   http://www.oracle.com/technology/products/database/asm/index.html
2. Dell Storage Selection Guidelines for Oracle Databases
   http://www.equallogic.com/resourcecenter/assetview.aspx?id=5117
3. Orion User Guide,
   http://www.oracle.com/technology/software/tech/orion/index.html
4. "Inside the EqualLogic PS Series iSCSI Storage Arrays", John Joseph, Eric Schott and Kevin Wittmer,
   Dell Power Solutions, February, 2008.
   http://www.dell.com/downloads/global/power/ps1q08-20080249-EqualLogic.pdf
5. Deploying Oracle Database on Dell EqualLogic PS5000XV iSCSI Storage
   http://www.equallogic.com/resourcecenter/assetview.aspx?id=5115