BUILDING AN ORACLE® GRID
WITH ORACLE VM ON DELL™ BLADE SERVERS AND DELL EQUALLOGIC™ iSCSI STORAGE

A Dell Technical White Paper

Dell™ Database Solutions Engineering
Kai Yu and David Mar
Dell Product Group

October 2009
Abstract ........................................................................................................................................... 5
Audience and Scope ........................................................................................................................ 5
Section I – The Grid architecture and the components within the Grid ......................................... 6
   Oracle RAC: The original Grid layer ............................................................................................. 6
   Virtualization Layer: The latest addition to the Grid infrastructure........................................... 6
   A new Grid model: Grid with Server Virtualization ................................................................. 7
   Server Layer: Blade Servers as the Grid platform ................................................................. 10
   Storage Layer: EqualLogic iSCSI SAN storage ........................................................................ 10
SECTION II - Grid Reference Configuration POC Project ............................................................... 11
   Hardware Configuration for the Grid architecture ............................................................... 12
   Components used for the Grid infrastructure ................................................................... 12
   Blades server network connections and Storage IO connections .......................................... 15
   Physical Grid Design .............................................................................................................. 17
   Virtual Grid Design ................................................................................................................. 18
      Guest VMs ............................................................................................................................. 19
   EqualLogic Storage provides the storage virtualization for the Grid ....................................... 20
   Enterprise Manager Grid control as the unified management solution for the Grid ............... 21
Section III - Grid Implementation .................................................................................................. 21
   Grid Control Management Infrastructure Configuration ..................................................... 21
   EqualLogic Shared Storage Configuration ............................................................................. 22
   Physical Grid Configuration .................................................................................................... 24
   Network configuration: ......................................................................................................... 24
      iSCSI storage connection configuration: ............................................................................. 24
   Configure 11g RAC Database Infrastructure on the physical Grid ....................................... 27
      11g RAC clusterware configuration ..................................................................................... 27
      11g RAC software installation ......................................................................................... 27
   Consolidating the multiple databases on the Physical 11g RAC infrastructure ...................... 28
   Virtual Grid Configuration .................................................................................................. 31
   Oracle VM Server Installation ............................................................................................ 31
   Oracle VM Server Network Configuration .......................................................................... 32
   Oracle VM Server Storage Configuration on EqualLogic iSCSI Storage ............................... 37
Connect the VM servers to Enterprise Manager Grid Control.................................................. 40
Create guest VMs using a VM template...................................................................................... 43
Add additional disks to the guest VM ........................................................................................ 47
Configure guest VMs for Oracle 11g RAC database .................................................................. 50
Section IV – Case Studies of the Grid Hosting Applications ..................................................... 51
Oracle E-business suite R12 on RAC/Virtual Machines................................................................. 52
Oracle E-Business suite R12 running on Virtual Machines........................................................... 52
Banner University ERP Applications suite .................................................................................. 53
Provisioning a two node 11g RAC database on VMs................................................................. 53
Section IV – Summary................................................................................................................ 53
References:.................................................................................................................................. 54
Abstract

Oracle Grid consolidates compute resources into an on-demand sharing resource pool. Oracle™ VM provides the server virtualization that enables the flexible Oracle Grid based on a pool of virtual machines. Blade servers and iSCSI storage are ideal hardware platforms for the Grid. This paper discusses a Grid reference configuration based on Dell PowerEdge™ blade servers and Dell EqualLogic™ iSCSI storage. This Grid implementation is comprised of the physical Grid based on physical servers and the virtual Grid based upon Oracle virtual machines. This paper will examine the physical Grid using Oracle 11g RAC, the virtual Grid using Oracle VM technology and Oracle Enterprise Manager Grid control as the unified management solution for the Grid. This paper will also explore the various options and considerations to consolidate multi-tier enterprise applications and database services on this Grid infrastructure.

Audience and Scope

This paper is designed to utilize a top down view of the Grid infrastructure, starting with a high level discussion and the Grid infrastructure design, then iteratively covering the details by diving deeper into detailed implementation aspects. Audiences interested in a high level overview may be interested in content presented at the beginning of the paper; while those interested in implementation and best practices may read about the details in the latter sections of this document.

To accommodate this, this paper has been divided into four major sections:

- High-level overview on the Grid infrastructure and recommended components
- Grid Reference Configuration POC
- Grid Implementation details and setup and configuration settings
- Considerations and case studies of consolidating enterprise applications on the Grid
Section I – The Grid architecture and the components within the Grid

Oracle RAC: The original Grid layer

Grid computing enables groups of systems to be pooled and provisioned in order to accommodate changing needs of business problems. Rather than using dedicated servers for specific tasks, each application may make use of shared computing resources. This approach results in better scalability and high availability; enabling different applications to consolidate infrastructure, save on cost and enhance power efficiency.

Dell believes that scale-out methodologies based on open solutions are more cost effective and reliable over proprietary, big-iron systems. Older business models rely on sizing for the worst case, requiring customers to buy oversized systems for peak periods and uncertain growth. Since such large machines may be in use for several years, forecasting outlook may be difficult and expensive to determine and often leads to gross underutilization.

Building a commodity grid infrastructure enables IT datacenters to achieve higher resource utilization through higher degrees of consolidation. It is flexible enough to add future resources when necessary so that capacity may be added dynamically and on demand. Additionally, smaller workloads may be shifted towards underutilized servers enabling higher resource utilization and lower overall costs. Smaller building blocks based upon open architectures and standards may enhance scalability and improve reliability, however, grid infrastructure has been largely confined to the database and implemented by Oracle RAC.

Virtualization Layer: The latest addition to the Grid infrastructure

Introducing virtualization to the Grid infrastructure environment allows for greater flexibility, enabling consolidation beyond the database grid by adding virtualization to the Grid infrastructure and databases. This allows multiple application servers or database servers to run together on one physical server. It expands the database grid to the virtual servers, providing an ideal location for test and development for the Oracle RAC infrastructure.

With the commodity hardware’s advance in higher CPU speeds, growing x86 core density and servers capable of holding ever growing amounts of memory, virtualization coupled with previous grid implementations adds flexibility and quicker deployment capabilities.

Virtualization drives resource utilization higher through “scale-up” methodologies. Within a grid infrastructure this means that the application layer may run multiple guest VMs on one single machine.

Additionally, virtualization enables quicker and more efficient provisioning and migration of both hardware and software. While the original DB grid had these capabilities, adding and
deleting a node involved many hardware deployment components, and usually did not address the operating system layers. Consequently, virtualization abstraction of hardware, drivers and other machine specific components brings to this next generation grid greater flexibility. If hardware components of the primary server running virtual machines were to fail, moving the guest environment to a different physical server is simple, even if that hardware is different than the original host.

A new Grid model: Grid with Server Virtualization

The follow diagram shows the logical view of the Grid infrastructure that consists of the physical server Grid based on Oracle RAC technology and the virtual Grid based on the Oracle VM. The Grid is based on the commodity servers and SAN shared storage.

![Logical View of the Grid Infrastructure](image)

Figure 1: Logical View of the Grid Infrastructure

The following figure is a physical view of the Grid infrastructure which represents the combination of Oracle RAC, and Oracle VM and management components. These software components scale out well on x86 building blocks, and virtualized iSCSI storage.
A grid infrastructure built on hardware that can is designed for future growth will allow greater flexibility. Scaling the infrastructure also means that the functions being performed on one unit of hardware should easily be migrated to another.

The figure above shows management by Oracle Enterprise manager. It’s recommended to use EM, outside of the blade chassis because if there is an issue with the blade infrastructure EM will be able to report failures. Additionally, it’s a best practice to have 2 nodes for high availability. On the figure above, the OVM Server layer may easily be grown by scale-out methodologies by adding blades as the VM servers to the Virtual Grid. Each additional blade is added to the VM server as the resource that allows guest virtual machines on the VM server pool to be easily migrated as the load becomes high. The additional VM servers added to into the Grid also increases the Grid capacities to consolidate applications guest VM on the Grid infrastructure.

As shown in the physical Grid for the DB layer, Oracle RAC technology enables both scale-up and scale-out paradigms to coexist. The figure above demonstrates that multiple databases can either run on one server to maximize utilization or a single database can scale-out to span multiple servers. These principles have existed since the introduction of Oracle RAC and provide the original cornerstone of the grid infrastructure.

With Equal Logic storage, a volume may be transferred to any member array to grow capacity or a volume may be moved to perform maintenance or in the event of down-time. With EqualLogic virtualization, storage is no longer tied to any particular physical enclosure or storage disk.
Deploying a virtual environment enables customers to decouple the physical server from the database, making it easier to upgrade or migrate the application to a completely different server. Virtualization makes it easier to deploy solutions that are typically difficult and error prone to setup. The problem is that virtualization by itself is limited to scale-up methodologies where the potential of the application is limited to the biggest physical server that is available.

Combining both Virtualization and GRID technologies enables customers to make use of the best of both worlds. Applications may consolidate as necessary and make use of scale-up and consolidation methodologies, as well as run several virtual instances and span several servers with limited performance impacts, enabling them to grow far beyond the capabilities of any one server. Virtualization complements Oracle RAC by enabling quick provisioning and hardware abstraction. When coupled with GRID technology, this enables growth far beyond the bounds of a single server, enabling the solution to grow linearly as more physical servers are added.

As illustrated in the following diagram, the top row is the physical grid, the bottom row is the virtual grid. Add one more OVM servers and form the share pool on node1, node 2 or node 3, and specify this pool can be expanded to have more OVM servers.

Figure 3 – Scalable Grid Infrastructure
Server Layer: Blade Servers as the Grid platform

With the ever growing requirements of IT datacenters, space is always a valuable commodity. Dell PowerEdge blades servers offer a dense solution to maximize space constraints. Dell’s M1000e blade enclosure offers the ability to enclose 16 blades in a 10U-sized enclosure of space.

Building an infrastructure that centralizes the database environment reduces complexity across the IT organization. It also enables more efficient uses of the resources by sharing resources and eliminating redundancy.

Easily removable expansion is key to building a successful grid. Dell blades offer easy rip and replace functionality so that administrators can easily deploy a replacement blade with the same configuration as a previous blade.

The server components used in this grid infrastructure have the advantage of continuously being scaled up as newer and faster components are brought to the market. Continuously adding higher core densities and speeds will ensure that a grid infrastructure today can be upgraded well into the future.

![M1000e Blade chasis front and back](image)

For more information on Dell blade servers, see [www.dell.com/blades](http://www.dell.com/blades).

Storage Layer: EqualLogic iSCSI SAN storage

With EqualLogic peer architecture, storage arrays work together to share resources and distribute load and optimize performance. In a similar manner a grid infrastructure seeks to provide building blocks that can incrementally grow to meet the needs of the datacenter. An EqualLogic iSCSI SAN can be set up quickly, and can seamlessly grow online storage as more
capacity is needed. Many arduous setup tasks common with traditional storage products are eliminated. EqualLogic’s linear performance improvements mean that capacity may be increased when the business mandates.

EqualLogic improves Oracle database performance and high availability by automatically load balancing across multiple storage resources. Each array is designed for high availability including hot-swappable components, multiple RAID types and hot spare disks. For more information on EqualLogic’s complementary performance when used in conjunction with Oracle ASM review the whitepapers located at http://www.dell.com/oracle.

Additionally, software built into every member allows administrators to create consistent local and remote copies of the database for test, development and recovery. Included in each array are advanced data protection features such as Auto Replication and Auto-Snapshot Manager. EqualLogic storage allows administrators to meet growing capacity demands by adding storage without database down time.

For more information on Dell EqualLogic storage, see www.dell.com/equallogic.

Figure 3 – Dell EqualLogic Storage

Figure 5: EqualLogic storage

SECTION II - Grid Reference Configuration POC Project

This paper is about the Poof-of-Concept project was worked on within the Dell | Oracle Solutions Engineering team to establish a Grid reference configuration. The reference Grid infrastructure is composed of the physical grid, virtual grid, shared storage layer and Oracle Enterprise Manager Grid Control. As each host of the physical grid is a physical server, the physical grid is intended to provide the database grid infrastructure to provide database services. The virtual grid is server based virtualization in which each host of the grid is essentially a virtual host or a guest virtual machine. These virtual guests are allocated from server pools that are based on Oracle VM technology. The virtual grid is intended to host application tier, middleware tier or the test/development databases which are not required to run on the physical servers. The applications or middleware running on the virtual grid may connect to their databases running on the physical grid or the virtual grid through the public network link.
This section draws on a number of reference documents to create a specific instantiation of Oracle Grid. Although other components may be used, this section of the paper documents a Poof-of-concept implementation of Grid with Oracle VM on Dell Blades and the EqualLogic platform.

Our goal was to expand the Grid infrastructure so that it was comprised of both the physical Grid provided by Oracle RAC and the virtual grid components provided by Oracle VM. This Grid is designed to provide a pre-built scalable infrastructure. It is comprised of a pool of system resources that enable consolidation of enterprise applications as well as their databases. Using this methodology it creates a way that applications and database services may share and obtain the resources within the Grid on demand. Additionally, the project was also designed to establish Grid control as the centralized management aspect of the infrastructure.

This document outlines the design and implementation of this Grid. This section of the document covers:

- Component hardware used in the proof of concept
- The Physical Grid – for Oracle DB
- The Virtual Grid – for the Test DB and Apps
- The Storage Grid – for the DB and Virtual Machines
- Enterprise Manager Grid Control - for management.

**Hardware Configuration for the Grid architecture**

This reference configuration is built upon Dell M610 blade servers and Dell EqualLogic PS6000 storage. While building this reference configuration the following components were used:

- Dell M1000e Blade chassis
- Eight M610 blade servers for Oracle 11g RAC / physical grid
- Four M610 blade server for the virtual Grid
- One Oracle Enterprise Manager Grid control server outside of Blade chassis.
- Two EqualLogic storage PS6000XV arrays as the shared storage
- Two Dell™ PowerConnect™ 6200 switches.

**Components used for the Grid infrastructure**

The following components were used to build the grid infrastructure proof of concept. For in-depth details on configuration please refer to the Tested and Validated configurations at http://www.dell.com/oracle.
<table>
<thead>
<tr>
<th>Tested Hardware/Software Requirements (For details, see below)</th>
<th>Validated Component(s)</th>
<th>Tested Oracle RAC Config</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PowerEdge™ Nodes</strong></td>
<td>M1000e (11xM610)</td>
<td>8 for physical grid (M610)</td>
</tr>
<tr>
<td></td>
<td>2xR710</td>
<td>3 for virtual grid (M610)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 for Enterprise Manager (M600)</td>
</tr>
<tr>
<td><strong>Memory</strong></td>
<td>All valid PowerEdge memory configurations</td>
<td>24Gig (per node)</td>
</tr>
<tr>
<td><strong>Dell EqualLogic iSCSI Storage Array</strong></td>
<td>6000X/XV</td>
<td>2</td>
</tr>
<tr>
<td><strong>Ethernet Ports</strong></td>
<td>Intel or Broadcom Gigabit NICs</td>
<td>6 per node in blade infrastructure</td>
</tr>
<tr>
<td><strong>Ethernet Switches for iSCSI Storage Area Network</strong></td>
<td>Dell PowerConnect 62xx Gigabit Switches, Dell PowerConnect</td>
<td>2 dedicated iSCSI switch for separated and multi-pathed iSCSI network.</td>
</tr>
<tr>
<td><strong>Ethernet Switches (For Private Interconnect)</strong></td>
<td>Gigabit-only Switches</td>
<td>2 dedicated switches for RAC interconnect</td>
</tr>
<tr>
<td><strong>Public Network</strong></td>
<td>Gigabit-only Switches</td>
<td>2</td>
</tr>
<tr>
<td><strong>Raid Controllers (Used for internal storage only)</strong></td>
<td>PERC 4e /Di, PERC 5/i, PERC 6i, CERC6, SAS 6/iR</td>
<td>1 (Per Node)</td>
</tr>
<tr>
<td><strong>Internal Drive</strong></td>
<td>All valid PowerEdge internal storage configurations</td>
<td>73 Gig/node</td>
</tr>
<tr>
<td><strong>Oracle Software &amp; Licenses</strong></td>
<td>Oracle 11g R1 11.1.0.6 Enterprise Edition (Base) + Oracle Patchset 11.1.0.7</td>
<td>RAC</td>
</tr>
<tr>
<td><strong>Oracle Patch(es)</strong></td>
<td>Oracle 11g R1 11.1.0.7 Critical Security Patches and Recommended Patches</td>
<td></td>
</tr>
<tr>
<td><strong>Operating System</strong></td>
<td>Red Hat Enterprise Linux AS 5 Update 2 with kernel 2.6.18-92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oracle Enterprise Linux 5 Update 3 with kernel 2.6.18-92.0.0.0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oracle VM server 2.1.2</td>
<td></td>
</tr>
<tr>
<td>Validated Servers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dell PowerEdge™ Servers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>BIOS¹</td>
<td>ESM - BMC Firmware¹</td>
</tr>
<tr>
<td>PE M610, PE M710</td>
<td>1.1.4</td>
<td>v2.00 (iDRAC)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Internal Disks RAID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PERC 4e/Di Driver</strong></td>
</tr>
<tr>
<td><strong>PERC 5/i</strong></td>
</tr>
<tr>
<td><strong>PERC 6/i</strong></td>
</tr>
<tr>
<td><strong>CERC6 (PE Blades only)</strong></td>
</tr>
<tr>
<td><strong>SAS 6/ir</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network Interconnect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intel Proset Gigabit Family Of Adapters</strong></td>
</tr>
<tr>
<td><strong>Intel Proset PCIE Single/Dual Port Gigabit Family Of Adapters</strong></td>
</tr>
<tr>
<td><strong>Intel Proset PCIE Quad Port Gigabit Family Of Adapters</strong></td>
</tr>
<tr>
<td><strong>Intel Proset PCIE 10 Gigabit Family Of Adapters</strong></td>
</tr>
<tr>
<td><strong>Broadcom NetXtreme II</strong></td>
</tr>
<tr>
<td><strong>Broadcom NetXtreme</strong></td>
</tr>
<tr>
<td><strong>NIC Teaming (Not available yet for TOE NICs)²</strong></td>
</tr>
<tr>
<td>Native teaming driver</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>iSCSI SAN Switches</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dell PowerConnect 54xx Gigabit Ethernet Switches</strong></td>
</tr>
<tr>
<td><strong>Dell PowerConnect 62xx Gigabit</strong></td>
</tr>
</tbody>
</table>
Blades server network connections and Storage IO connections

The primary servers used in this implementation are blade servers. These servers were chosen for ease of scalability using the blade infrastructure. To configure the blade chassis there are six IO modules for network connections or other IO connections: A1, A2, B1, B2, C1, C2. In this implementation, iSCSI was used so all IO fabrics made use of network modules.

The following diagram illustrates the connections between NIC cards in the blade server and the IO modules for M610.

1 Minimum BIOS and ESM/BMC/iDRAC versions. For latest updates please go to http://support.dell.com.
Figure 6 – NIC to IO Module Connections

In this configuration, each blade server will have 6 Ethernet ports, each which corresponds to one IO modules. Our configuration utilized the following layout.

<table>
<thead>
<tr>
<th>IO module</th>
<th>NIC interface</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>eth0</td>
<td>Public network</td>
</tr>
<tr>
<td>A2</td>
<td>eth1</td>
<td>Public</td>
</tr>
<tr>
<td>B1</td>
<td>eth2</td>
<td>iSCSI connection</td>
</tr>
<tr>
<td>B1</td>
<td>eth3</td>
<td>iSCSI connection</td>
</tr>
<tr>
<td>C1</td>
<td>eth4</td>
<td>Private interconnect</td>
</tr>
<tr>
<td>C2</td>
<td>eth5</td>
<td>Private interconnect</td>
</tr>
</tbody>
</table>

Figure 8: Storage connections between Blade classis and EqualLogic storage.
Physical Grid Design

The Physical Grid provides a consolidated 11g RAC database infrastructure to host multiple databases. Each of databases on this Grid infrastructure can be configured to run on any number of database instances depending on the load. This infrastructure provides each database with the capability to be dynamically migrated to different Grid nodes depending on the workload of the database instances.

The physical Grid may also be scaled out by adding additional nodes to the Grid to meet the demand of workloads. The empty slots of the blade chassis provide the capability to add additional blade servers, and Grid control allows them to be joined to the existing physical Grid. The physical Grid uses ASM to provide storage virtualization for the databases on the cluster.
Virtual Grid Design

The virtual Grid is based on virtual machines which are hosted within the VM server pool. Within the reference design we started with two servers, the minimal to provide some level of high availability. The initial pool was comprised of two M610 blade servers each with 24 GB of memory and 2 XEON 5500 series quad core processors. With blades it is very easy to scale this structure by adding more M610 blade servers to scale out the VM Server pool.
Guest VMs

Virtual Grid Architecture Based on Oracle VM

Figure 11 – Virtual Grid architecture ased on Oracle VM

The figure above illustrates the virtual grid architecture. The nodes of the virtual grid are guest virtual machines instead of physical servers within the physical grid. A guest virtual machine has the operating system running on a set of configurable resources including memory, virtual CPUs, network devices and disk devices. As a best practice the total number of virtual CPUs on a virtual server should be less than or equal to $2 \times$ total number of CPU cores that exist with the VM Server. In our example reference design each OVM server has 8 CPU cores, we are limited to a total 16 virtual CPUs for all the guest VMs that are running on each VM server. The total amount of virtual memory available for all of the Virtual Machines running on a physical server should be less than or equal to the total amount of RAM installed. Each server in our reference design includes 24GB of memory, so the aggregate memory of our guests running on an OVM server should not be larger than 24GB.
Application software may be configured and run on the virtual machines as with physical machines. A guest virtual machine is associated with a VM server pool from which the resources are assigned to the guest virtual machine. A VM server pool contains one or more VM servers running on physical servers and, possibly, shared storage. With shared storage configured in the VM server pool, a guest virtual machine associated with the VM server pool may be started and run on any VM server within the pool that is available and has the most resource free. With HA (high available) enabled on the server pool level and the guest virtual machine level, the virtual machine is migrated or restarted on another available VM server if the VM server that runs the guest virtual machine is shutdown or fails.

To complement the VM manager, the native management solution for Oracle VM environment, Oracle has released the Oracle enterprise Manager 10g R5 with Oracle VM management Pack which provides the complementary management solution to manage the virtual server infrastructure, see [1] for details.

This solution includes the management and provisioning for the components of the virtual Grid such as virtual server pool, VM servers, guest VMs, and the resources allocations to the guest VMs such as CPUs, disks, networks, and memory. The figure above shows the virtual grid infrastructure managed by the enterprise manager Grid control 10g R5 with the Oracle VM management Pack.

**EqualLogic Storage provides the storage virtualization for the Grid**

As shown in the figure above, the Blades servers in the Grid connect to Dell EqualLogic storage PS6000 XV through Gigabit Ethernet IP Storage Area Network (SAN). The Ethernet configuration includes the two Blade switches on the IO modules B1 and B2 and two external Dell PowerConnect 6200 switches. This fully redundant configuration provides high availability and scalability and maximum IO bandwidth for the storage IOs.

As illustrated in figure above, one or more EqualLogic storage array, called members, can be grouped into a storage group which is assigned a DNS name or IP address. Each storage group can be segregated into pools and, at any given time, a member may be assigned to one pool. The member may be moved between pools while keeping the storage online. For this POC project, two members, oracle-member01 and oracle-member02, are assigned to pool Raid10 with the Raid 10 disk array configuration. In order to provide the usage storage for Grid, physical disks are grouped by a usable component called volume. A volume represents a portion of the storage pool that can be spread across multiple disks and group members and is seen as an iSCSI target visible to storage clients. To client OS, it is a logic disk.
Enterprise Manager Grid control as the unified management solution for the Grid

Enterprise Manager Grid control is configured to manage and provision the Grid infrastructure. In addition to the standard Grid control functionalities which manage the hosts, RAC database and various applications on the physical server, Enterprise Manager Grid control adds Oracle VM Management Pack, which provides the complete management of virtual server infrastructure and related applications.

Additionally, Enterprise Manager Grid control provides the unified management solution which spans the entire lifecycle of applications on both the physical and virtual infrastructure. This environment allows customers to manage both physical and virtual environments and the applications running on them. Within this proof of concept, we implemented the enterprise management Grid control 10.2.0.5 R5 with the Virtual management pack as the sole management solution for the Grid.

Section III - Grid Implementation

This session will discuss the configuration and implementation methods of the Grid infrastructure that was discussed in session II. The Grid implementation will cover the following major components and the association among them:

- Grid control management infrastructure
- EqualLogic shared storage
- Physical Grid
- Virtual Grid

Grid Control Management Infrastructure Configuration

As a part of the Grid infrastructure, Enterprise Manager Grid Control 10.2.0.5 is configured to manage the entire Grid. By adding the virtual management pack, Enterprise Manager Grid control 10.2.0.5 is capable of providing a unified management solution for both the physical Grid and the virtual Grid with components such as virtual server pools, virtual servers and virtual machines.

The Enterprise manager grid control configuration steps include:

1. Oracle® Enterprise Linux installation and configuration:
   - OEL 4.7 was configured for this EM Grid control.
2. Install Oracle Enterprise manager Grid control 10.2.3
   - This step installs Oracle management service (OMS) R2(10.2.0.3), the Oracle management repository database with the repository database schema sysman, and oracle management agent 10.2.0.3
3. Upgrade EM Grid control from 10.2.0.3 to 10.2.0.5:
   Apply patch p3731593_10205_x86_64.zip, which contains 10.2.0.5 patch set software, to:
   Upgrading OMS 10gR2 (10.2.0.3) to OMS 10gR5 (10.2.0.5)
   Upgrading Oracle® Management Repository (sysman schema)
   Upgrading Oracle Management Agent on the host to 10.2.0.5 where OMS is running
   Refer to the 10.2.0.5 upgrade patch readme [3] for the detailed upgrade steps.

4. Enable Virtual Management pack on EM grid control 10.2.0.5
   Virtual management pack is cooperated to EM grid control as a management pack plug-in.
   Apply EM virtualization patch 8244731 on EM OMS : shutdown OMS, and apply the patch to OMS home using opatch utility, refer to the readme of patch 8244731 for details
   Install TightVnc Java Viewer (download from http://oss.oracle.com/oraclevm/emplugin/tightvnc/VncViewer.jar)
   Copy the jar file into OMS $ORACLE_HOME/j2ee/OC4J_EM/applications/em/em and restart the OMS for changes to take effect.
   Refer to [4] for more details.

### EqualLogic Shared Storage Configuration

In this proof of concept project, a set of storage volumes were created in the EqualLogic PS6000XV storage to provide the shared storage for the physical Grid and the virtual Grid

<table>
<thead>
<tr>
<th>Volume</th>
<th>Size</th>
<th>Raid</th>
<th>Used for</th>
<th>OS Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>blade-ocrs</td>
<td>2GB</td>
<td>10</td>
<td>11g clusterware OCR votingdisk</td>
<td>3 votingdisks 2 OCRs</td>
</tr>
<tr>
<td>blade-data1</td>
<td>100GB</td>
<td>10</td>
<td>Data for DB1</td>
<td>ASM diskgroup1</td>
</tr>
<tr>
<td>blade-data2</td>
<td>100GB</td>
<td>10</td>
<td>Data for DB2</td>
<td>ASM diskgroup2</td>
</tr>
<tr>
<td>blade-data3</td>
<td>150GB</td>
<td>10</td>
<td>Data for DB3</td>
<td>ASM diskgroup3</td>
</tr>
<tr>
<td>blade-data5</td>
<td>150GB</td>
<td>10</td>
<td>Data for DB4</td>
<td>ASM diskgroup4</td>
</tr>
<tr>
<td>Add more volumes as needed ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Storage volumes for the physical Grid
<table>
<thead>
<tr>
<th>Volume</th>
<th>Size</th>
<th>Raid</th>
<th>Used for</th>
<th>OS Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>blade-data4</td>
<td>400GB</td>
<td>10</td>
<td>VM repository</td>
<td>/OVS</td>
</tr>
<tr>
<td>Blade-data6</td>
<td>500GB</td>
<td>10</td>
<td>VM repository</td>
<td>/OVS/9A87460A7EDE43EE92201B8B7989DBA6</td>
</tr>
<tr>
<td>Vmocr1-4</td>
<td>1GB each</td>
<td>10</td>
<td>For 11g clusterware on VMs</td>
<td>3 votingdisks, 2 OCRs</td>
</tr>
<tr>
<td>vmracdb1</td>
<td>50GB</td>
<td>10</td>
<td>Data for 11g RAC DB on VM</td>
<td>ASM diskgroup1</td>
</tr>
<tr>
<td>vmracdb2</td>
<td>50GB</td>
<td>10</td>
<td>Data for RAC DB on VM</td>
<td>ASM diskgroup2</td>
</tr>
<tr>
<td>Add more</td>
<td></td>
<td></td>
<td>volumes as needed</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Storage volumes for the virtual Grid

The storage can be scaled out by adding additional EqualLogic PS6000XV arrays as members to the storage group to gain more disk spindles or more disk space.

Figure 12 – EqualLogic Storage Management Console
The storage group was given IP address 10.16.7.100 which is be used to connect to the storage from servers from Grid.

**Physical Grid Configuration**

The physical Grid is based on a 11g RAC which is designed to consolidate multiple databases on a single eight-node RAC for various applications.

**Network configuration:**

5 of 6 network interfaces which correspond to 6 IO modules are used for different network configuration:

<table>
<thead>
<tr>
<th>Network Interface</th>
<th>IO modules</th>
<th>Connections</th>
<th>IP address</th>
</tr>
</thead>
<tbody>
<tr>
<td>eth0</td>
<td>A1</td>
<td>Public network*</td>
<td>155.16.9.71-78</td>
</tr>
<tr>
<td>eth2</td>
<td>B1</td>
<td>iSCSI connection</td>
<td>10.16.7.241-255(old number)</td>
</tr>
<tr>
<td>eth3</td>
<td>B2</td>
<td>iSCSI connection</td>
<td>10.16.7.240-254(even number)</td>
</tr>
<tr>
<td>eth4</td>
<td>C1</td>
<td>eth4 and eth5 bonded to form bond0</td>
<td>192.168.9.71-78</td>
</tr>
<tr>
<td>eth5</td>
<td>C2</td>
<td>Virtual IP for 11g RAC</td>
<td>155.16.9.171-178</td>
</tr>
<tr>
<td>VIP</td>
<td></td>
<td>Virtual IP</td>
<td></td>
</tr>
</tbody>
</table>

Note *: eth1 can be used for the optional second NIC for public network.

Table 3: Physical Grid Network Configuration

**iSCSI storage connection configuration:**

To configure the blade servers to access the storage volumes in table 1, the open-iSCSI administration utility is used on each blade server of the 8 node cluster to configure the iscsi connection from the blade host to the EqualLogic storage volumes. Please refer to the Dell Oracle EqualLogic white paper [9] for step by step instruction.

Once the EqualLogic storage volumes in table 1 are accessible from blade server, the volumes will show as disk partitions in /proc/partitions. The next step is to establish multipaths for these partitions using Linux™ Device Mapper.

Edit the /etc/multipath.conf file to specify the mapping between the device multipath alias name and the device WWID, for example:
/etc/multipath
multipaths {
multipath {
    wwid 36090a028e093dc7c6099140639aae1c7
    alias ocr-crs
}
multipath {
    wwid 36090a028e093cc896099340639aac104
    alias data1
}
multipath {
    wwid 36090a028e0939c31b499241039aa416e
    alias data2
}
multipath {
    wwid 36090a028e0936c44b499441039aa0160
    alias data3
}
multipath {
    wwid 36090a028e0936c44b499441039aa0160
    alias data5
}
}

Restart the multipath service:

service multipathd restart

Run the command **multipath –ll** to check the alias

And verify the /dev/mapper/* devices are created such as

ls /dev/mapper/*
/dev/mapper/ocr-crsp5 /dev/mapper/data1 /dev/mapper/ocr-crsp6
/dev/mapper/data1p1 /dev/mapper/ocr-crsp7 /dev/mapper/data2
/dev/mapper/data2p1 /dev/mapper/data3 /dev/mapper/data3p1
/dev/mapper/ocfsp1 /dev/mapper/data5 /dev/mapper/ocr-crsp
/dev/mapper/data5p1 /dev/mapper/ocr-crsp1 /dev/mapper/ocr-crsp2
/dev/mapper/ocr-crsp3
Use block devices for 11g clusterware and database ASM diskgroups

Starting with Linux™ kernel 2.6 (RHEL5/OEL5), support for raw devices have been depreciated in favor of block devices. For example, Red Hat Linux 5 no longer offers services for raw devices. So, for a long term solution one should consider moving away from raw devices to block devices. As 11g clusterware fully supports building OCR and voting disk on the block devices and 11g RAC allows the building of ASM disk groups on the block devices, the multipath blocks devices specified in previous session can be used directly for OCRs, voting disks and ASM disk groups by setting the proper permissions in the /etc/rc.local file.

For example, /etc/rc.local appears as follows:

```bash
# permission for two OCR
chown root:oinstall /dev/mapper/ocr-crsp1
chmod 0640 /dev/mapper/ocr-crsp1
chown root:oinstall /dev/mapper/ocr-crsp2
chmod 0640 /dev/mapper/ocr-crsp2

# permission for three votingdisks
chown oracle:oinstall /dev/mapper/ocr-crsp5
chmod 0640 /dev/mapper/ocr-crsp5
chown oracle:oinstall /dev/mapper/ocr-crsp7
chmod 0640 /dev/mapper/ocr-crsp7
chown oracle:oinstall /dev/mapper/ocr-crsp6
chmod 0640 /dev/mapper/ocr-crsp6

# permission for ASM diskgroups
chown oracle:dba /dev/mapper/data1p1
chmod 0660 /dev/mapper/data1p1
chown oracle:dba /dev/mapper/data2p1
chmod 0660 /dev/mapper/data2p1
chown oracle:dba /dev/mapper/data3p1
chmod 0660 /dev/mapper/data3p1
chown oracle:dba /dev/mapper/data5p1
chmod 0660 /dev/mapper/data5p1

# ocr-crsp3 for ASM spfile
chown oracle:dba /dev/mapper/ocr-crsp3
chmod 0660 /dev/mapper/ocr-crsp3
```
Configure 11g RAC Database Infrastructure on the physical Grid

The physical Grid is essentially based on an Oracle 11g RAC infrastructure to host the multiple database services. The Oracle RAC infrastructure consists of four separate Oracle software configurations: Oracle 11g clusterware, Oracle 11g ASM home, Oracle 11g RAC database home and Oracle 10.2.0.5 Grid control agent home.

11g RAC clusterware configuration

The 11g clusterware is installed on ORA_CRS_HOME = /crs/oracle/product/11.1.0/crs

The Private interconnect addresses for the 8 cluster nodes:

- 192.168.9.71             kblade1-priv.us.dell.com       kblade1-priv
- 192.168.9.72             kblade2-priv.us.dell.com       kblade2-priv
- 192.168.9.73             kblade3-priv.us.dell.com       kblade3-priv
- 192.168.9.74             kblade4-priv.us.dell.com       kblade4-priv
- 192.168.9.75             kblade5-priv.us.dell.com       kblade5-priv
- 192.168.9.76             kblade6-priv.us.dell.com       kblade6-priv
- 192.168.9.77             kblade7-priv.us.dell.com       kblade7-priv
- 192.168.9.78             kblade8-priv.us.dell.com       kblade8-priv

Two copies of OCRs: /dev/mapper/ocr-crsp1, /dev/mapper/ocr-crsp2
Three Voting disks: /dev/mapper/ocr-crsp5, /dev/mapper/ocr-crsp6, /dev/mapper/ocr-crsp7

11g ASM configuration

Oracle ASM instance provides the storage virtualization for all the databases on the Grid. Oracle ASM software is installed in Oracle ASM home: /opt/oracle/product/11.1.0/asm on each of the RAC nodes. A single ASM instance was created and runs on each of the RAC nodes to provide the storage virtualization for all databases that are running the Grid.

11g RAC software installation

11g RAC database software is installed in the Oracle database home (/opt/oracle/product/11.1.0/db_1) to provide the cluster database services. The database listener service needs to be created using the netca utility.

Install Oracle Grid control 10.2.0.5 agent

Oracle Enterprise manager agent 10.20.5 is installed in Oracle Agent home (/opt/oracle/product/agent10g). The agent is configured to connect to the Enterprise
Manager OMS server URL: http://www.kblademgr:us.com:1159 to join the Grid management infrastructure.

**Consolidating the multiple databases on the Physical 11g RAC infrastructure**

In this proof of concept project, the physical grid built on this 8-node 11g RAC cluster on 8-node RAC is designed to host multiple database services. Database services can be pre-created as 'container' for future application schemas or can be created on demand whenever a database service is needed to deploy applications.

To create database services, use the following steps:

1. Understand the size and performance requirement of the database such as CPU, IO loads and memory requirement.
2. Determine the how much storage and how many database instances are needed.
3. Based on the workload of the each RAC node, determine which RAC nodes the database will be running.
4. Provision the storage volumes from the shared EqualLogic storage, make it accessible to all the RAC nodes and create mutli-path disk partitions on these storage volumes
5. Create the ASM diskgroups using the disk partitions created in the previous step and mount these ASM diskgroups on the ASM instance on each RAC node.
6. Use the DBCA database creation assistant utility to create the database services based on the RAC nodes determined in steps 4 and 5.
7. For some special ERP applications, such as Oracle E-business suite, where a non-RAC database is pre-installed with the application installation, some special database utility example such as rconfig may be used to convert the pre-configured non-RAC database to the RAC infrastructure.

As examples in this proof of concept project four ASM diskgroups, DG1, DG2, DG3, DG4, were created for data volume data1, data2, data3 and data5. These four ASM diskgroups were used to create four database services. During the initial database creation, each of the four databases was run on 2-3 RAC nodes. Depending on performance requirement and database load changes of each of database, administrators can dynamically add additional database instance or reduce the number of the instance from the databases. The administrator also can move the database to a different RAC node to keep the load balance of the Grid as well as achieve the best performance.

Refer to [7] for the detailed steps of dynamically changing the database instance as well as RAC node allocation of a database service.

**How to scale out the physical database Grid infrastructure**

The Grid infrastructure is designed to scale out on demand. Following are some scale-out scenarios and the steps that may be involved:
To add additional databases to the Grid:

. Allocate additional storage volume and making it accessible to all the RAC nodes
. Create additional diskgroups using the new storage volume
. Create the database using the new diskgroup

To add more RAC nodes to a database:

Use enterprise manager’s adding instance procedure to add a new database instance on the
new node to the database

To add additional storage to a database:

. Create additional volume on the EqualLogic storage group and make it accessible as a
muliplath partition to all of the RAC nodes
. Add this disk partition to the diskgroup of the database

To add additional RAC node to the Grid

Use 11g RAC add node procedure to add new nodes to the cluster and add additional database
instances on the new node.

To add additional EqualLogic storage array (member) to the storage group. In case the
shared storage is running out of space or the Grid needs more storage spindles to
handle more IO loads:

. Connect one of more additional EqualLogic storage members to the EqualLogic storage
. Add the new member to the storage group and the storage pool
. The volumes will be expanded to the new member (array) to take advantage of the
additional spindles and disk space.

After the physical Grid is configured and connected to the Grid control OMS (Oracle
management Server) from the Grid control console, the administrator can manage the Grid
host, the cluster databases and all the database objects.

The followings are the screen shots taken from the Grid control console and show the physical
Grid structure.
Figure 13 - Physical Grid Hosts: kblade1-kblade8

Figure 14 - Example cluster databases on the physical Grid
Virtual Grid Configuration

This session will introduce the virtual grid infrastructure implementation. This implementation includes:

- Virtual servers Installation
- Virtual server network and storage configuration
- Connect VM servers to the Grid Control
- Management VM infrastructure through Grid control
- Create guest VM using VM template
- Manage the resources of guest VMs

Oracle VM Server Installation

The virtual grid configuration starts with the installation of Oracle VM server software on Dell blade server. The following steps are introduced in *How to Add Shared Repositories to Oracle VM Pools with Multiple Servers, Metalink Note # 869430.1* [6].

1. Create a RAID logical volume on the local disk to install the OVM server on the blade servers.
2. Enable virtualization technology for the CPUs on BIOS: during the system startup press F2 to enter the System Setup program; enter CPU information and enable virtualization Technology option on BIOS.

3. Install Oracle VM server 2.1.2 software from Oracle VM server CD: at the boot prompt, press Enter and finish the install process.

4. Change the Dom0 memory setting: Login to the VM server and edit /boot/grub/menu.lst. Make sure you have this setting:
   
   ```
   kernel /xen-64bit.gz dom0_mem=1024M
   ```

5. Check the Oracle VM agent status: Oracle VM agent is the agent software that is installed with the Oracle VM server and is used to communicate with the Oracle VM manager or virtual management pack in Grid control. Ensure the OVM agent is running using:
   
   ```
   service ovs-agent status
   ```

---

**Oracle VM Server Network Configuration**

With the updated version of the underlying Xen hypervisor technology, the Oracle VM server also includes a Linux kernel. This Linux kernel runs as dom0 to manage one or more domU guest virtual machines. Oracle VM uses Xen bridge to provide the network for guest VMs running on domU. This bridge functions as a virtual switch presented to the guest VMs. When the VM server is installed in a Dell blade server M610, six network interfaces are shown in the dom0 by default. Each interface is associated with a Xen bridge that can be presented to guest VM in domU. The default configuration can be modified to meet your Grid configuration needs. The figure below shows the customization of the network configuration used in this proof of concept project.
In dom0, six of the network interfaces play similar roles as those in the servers of the physical grid. The difference is that some of these network interfaces need to be presented to the guest VMs through the Xen bridges. The figure above shows the network configuration for the VM servers and how they are presented to the guest VMs.

- **eth0** is for a public network connecting to the IO module A1 (eth1 connecting with A2 can be an optional second public network interface).
- eth2 and eth3 connecting to IO modules B1 and B2 are for the iSCSI connections.
- As all the guest VM’s IO operations are handled by dom0, there is no need to expose eth2 and eth3 to guest VMs, therefore no Xen bridge is needed for eth2 and eth3.
- eth4 and eth5 are bonded as bond0 for the private interconnect. This bonded private interconnection is also used for OCFS2 heartbeat traffic between VM servers in the VM server pool. This OCFS2 cluster file system provides the VM repositories. On other hand, bond0 is also presented to the guest VMs as a virtual network interface eth2 through Xen bridge Xenbr1. This virtual network interface eth2 can be used for the private interconnection between guest VMs which is required for deploying Oracle 11g RAC database on the guest VMs.

The table below illustrated the example network configuration in VM server dom0: how these network interfaces are bonded and bridged and presented to the guest VM as the virtual network interfaces.
The following are the steps used to implement these network configurations by customizing the default Xen bridge configuration and network configurations from the VM server installation:

1. **Shutdown the default Xen bridge configuration:**
   `/etc/xen/scripts/network-bridges stop`

2. **Make sure there is no Xen bridge shown by the command:**
   ```bash
   [root@kblade9 scripts]# brctl show
   bridge name  bridge id    STP enabled  interfaces
   ```

3. **Modify `/etc/xen/xend-config.sxp**
   change the line:
   ```sxp```
   (network-script network-bridges)
   To:
   ```sxp```
   (network-script network-bridges-dummy)

4. **Edit `/etc/xen/scripts/network-bridges-dummy**
   to include only the following two lines::
   ```bash
   #!/bin/sh
   /bin/true
   ```

5. **Configure the network interfaces and bonding and Xen bridges xenbr0 and xenbr1 by editing the following network scripts in** `/etc/sysconfig/network-scripts`:

---

### Table 3: Network configuration in OAM server and guest VMs

<table>
<thead>
<tr>
<th>Network Interface in Dom0</th>
<th>IO modules</th>
<th>Connections</th>
<th>IP address</th>
<th>Xen Bridge</th>
<th>Dom#1</th>
<th>Dome#2</th>
<th>Dom#N</th>
</tr>
</thead>
<tbody>
<tr>
<td>eth0</td>
<td>A1</td>
<td>Public network</td>
<td>155.16.9.82-85</td>
<td>Xenbr0</td>
<td>eth0: 155.16.99.101</td>
<td>eth0: 155.16.99.102</td>
<td>eth0: 155.16.99.10N</td>
</tr>
<tr>
<td>eth2</td>
<td>B1</td>
<td>iSCSI connection</td>
<td>10.16.7.228-234(even)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eth3</td>
<td>B2</td>
<td>iSCSI connection</td>
<td>10.16.7.229-235(odd)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eth4</td>
<td>C1</td>
<td>eth4 and eth5 bonded to form bond0</td>
<td>192.168.9.82-85</td>
<td>Xenbr1</td>
<td>eth1:192.168.9.101</td>
<td>eth1:192.168.9.102</td>
<td>eth1: 155.16.99.10N</td>
</tr>
<tr>
<td>eth5</td>
<td>C2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ifcfg-eth0:
DEVICE=eth0
HWADDR=00:1D:09:FC:B8:10
ONBOOT=yes
BOOTPROTO=none
BRIDGE=xenbr0

ifcfg-eth2:
DEVICE=eth2
BOOTPROTO=none
HWADDR=00:1D:09:72:29:F4
ONBOOT=yes
IPADDR=10.16.7.229
NETMASK=255.255.255.0
USERCTL=no
MTU="9000"

ifcfg-eth3:
DEVICE=eth3
BOOTPROTO=none
HWADDR=00:1D:09:72:29:F6
ONBOOT=yes
IPADDR=10.16.7.230
NETMASK=255.255.255.0
USERCTL=no
MTU="9000"

ifcfg-eth4:
DEVICE=eth4
BOOTPROTO=none
HWADDR=00:1D:09:FC:B8:12
TYPE=Ethernet
MASTER=bond0
SLAVE=yes
USERCTL=no
ONBOOT=yes

ifcfg-eth5:
DEVICE=eth5
ONBOOT=yes
BOOTPROTO=none
HWADDR=00:10:18:3b:37:e0
TYPE=Ethernet
MASTER=bond0
SLAVE=yes
USERCTL=no

ifcfg-bond0:
DEVICE=bond0
ONBOOT=yes
BOOTPROTO=None
BRIDGE=xenbr1

ifcfg-xenbr0:
DEVICE=xenbr0
ONBOOT=yes
TYPE=Bridge
DELAY=0
STP=off
BOOTPROTO=None
IPADDR=155.16.9.82
NETMASK=255.255.0.0

ifcfg-xenbr1
DEVICE=xenbr1
ONBOOT=yes
TYPE=Bridge
DELAY=0
STP=off
BOOTPROTO=None
IPADDR=192.168.9.82
NETMASK=255.255.255.0

Restart the network service:
[root@kblade9 scripts]# Service network restart

And check the Xen bridge configuration:
[root@kblade9 scripts]# brctl show
bridge name  bridge id          STP enabled  interfaces
xenbr0      8000.002219d1ded0     no          eth0
xenbr1      8000.002219d1ded2     no          bond0
Oracle VM Server Storage Configuration on EqualLogic iSCSI Storage

Shared storage volumes for the OVM repositories and virtual disks of guest VMs

All the Oracle VM resources, including guest VMs, Virtual Machine templates (guest seed images), ISO images, shared/non-shared Virtual Disks, etc are stored in Oracle VM repositories. During Oracle VM server installation, it creates a single non-shared, OCFS2 repository (/OVS) on the local storage. As this default configuration does not work with a VM server pool with multiple VM servers required to operate on the shared resources in the repository, on which the virtual Grid is based, the repositories infrastructure of the VM servers in the virtual Grid need to be moved to the shared storage. EqualLogic PS6000XV provides the shared storage for this repository infrastructure.

As indicated in Table 3, the following storage volumes are created for the virtual Grid:

- blade-data4 (400GB) and blade-data6 (500GB) for OVM repositories
- Four 1GB volumes vmorc1, vmorc2, vmorc3, vmorc4 are created as the disk partitions that will be presented to the guest VMs as the virtual disks for OCR/voting disks for the 11g RAC that are running on the guest VMs.
- vmrac1 and vmrac2 are the partitions that will be used as the virtual disks for ASM disk groups of the 11g RAC database running on guest VMs.

Configuring the dom0 host access to the EqualLogic storage volumes

To access these EqualLogic volumes from dom0, two network interfaces eth2 and eth3 shown in table 3 are configured in dom 0 using the open-iSCSI administration utility in the same way as they were for a physical server.

Secondly, the device mapper multipath is configured for all the iscsi volumes in the same way as it was in a physical server. Refer to the corresponding steps in the physical Grid for details.

The following devices should be created and available in dom0:

```
[root@kblade10 ~]# ls /dev/mapper/*
/dev/mapper/blade-data6 /dev/mapper/mpath5 /dev/mapper/ovs_data4p1
/dev/mapper/blade-data6p1 /dev/mapper/vmocr-css1 /dev/mapper/vmocr-css1p1
/dev/mapper/vmocr-css1p2 /dev/mapper/vmocr-css1p3 /dev/mapper/vmocr-css2
/dev/mapper/vmocr-css3 /dev/mapper/vmocr-css4 /dev/mapper/vmracdb
/dev/mapper/vmracdb1 /dev/mapper/vmracdb1p1 /dev/mapper/ovs_data4
```

Convert the OVM repository from local storage to the shared storage
• Configure OCFS2 file system for VM repositories:
  Create /etc/ocfs2/cluster.conf file:

node:
  ip_port = 7777
  ip_address = 192.168.9.82
  number = 0
  name = kblade9
  cluster = ocfs2

node:
  ip_port = 7777
  ip_address = 192.168.9.83
  number = 1
  name = kblade10
  cluster = ocfs2

node:
  ip_port = 7777
  ip_address = 192.168.9.85
  number = 2
  name = kblade11
  cluster = ocfs2

cluster:
  node_count = 3
  name = ocfs2

  stop ocfs2 service: service o2cb stop
  Configure oc2b service:

[root@kblade9 ocfs2]# service o2cb configure

Configuring the O2CB driver.
This will configure the on-boot properties of the O2CB driver.
The following questions will determine whether the driver is loaded on boot. The current values will be shown in brackets ('1'). Hitting <ENTER> without typing an answer will keep that current value. Ctrl-C will abort.

Load O2CB driver on boot (y/n) [n]: y
Cluster to start on boot (Enter "none" to clear) [ocfs2]:

38
Specify heartbeat dead threshold (>=7) [31]: 100
Specify network idle timeout in ms (>=5000) [30000]: 60000
Specify network keepalive delay in ms (>=1000) [2000]: 1000
Writing O2CB configuration: OK
Mounting configs filesystem at /sys/kernel/config: OK
Loading module "ocfs2_dlmfs": OK
Creating directory '/dlm': OK
Mounting ocfs2_dlmfs filesystem at /dlm: OK

Make ocfs2 file system on /dev/mapper/ovs_data4p1
[root@kblade9 /]# mkfs.ocfs2 -b 4k -C 64k -L ovs /dev/mapper/ovs_data4p1

Unmount the existing /OVS on local disk:

umount /OVS

Change /etc/fstab to have the shared volume mounted at boot:

### /dev/sda3 /OVS ocfs2 defaults 1 0
/dev/mapper/ovs_data4p1 /OVS ocfs2 _netdev,datavolume,nointr 0 0

mount ocfs2 partitions: mount -a -t ocfs2
create the following directories under /OVS:

mkdir /OVS/iso_pool
mkdir /OVS/proxy
mkdir /OVS/running_pool
mkdir /OVS/seed_pool
mkdir /OVS/ sharedDisk

Repeat the steps above on all the VM servers in the VM server pool.

2. Add a new volume to the repositories:

To increase the size of OVM the repositories, you can add more additional disk partitions to the VM repositories:

Make OCFS2 file system on the new disk partition /dev/mapper/blade-data6p1
mkfs.ocfs2 -b 4k -C 64k -L ovs /dev/mapper/blade-data6p1

Add the new partition to the repositories:
/usr/lib/ovs/ovs-makerepo /dev/mapper/blade-data6p1 1 OVS_repository

Check the new repositories:
df –k | grep /OVS
The 500GB new repository /OVS/9A87460A7EDE43EE92201B8B7989DBA6 is added using the new partition /dev/mapper/blade-data6p1.

Repeat the step 1-3 above on all the VM servers of the VM server pool.

Connect the VM servers to Enterprise Manager Grid Control

The following steps are performed to connect VM servers to the VM management pack of the Enterprise Manager Grid control:

1. Meet the pre-requisites in the VM server:
   - Create Oracle user with group name oinstall in the VM server dom0.
   - Establish Oracle ssh user equivalence between the VM server dom0 and the Enterprise manager Grid control server kblademgr.
2. Create /OVS/proxy directory and set the ownership.
   - chown oracle.oinstall /OVS/proxy
3. Set the sudo privilege for Oracle user in the VM server.
   - visudo -f /etc/sudoers
     - Add line: oracle ALL=(ALL) NOPASSWD: ALL
     - Comment out line: Defaults requiretty
4. Create a virtual server pool:
   - Login to enterprise Manager Grid control console as sysman, go to the Targets → Virtual Servers tab and select:
     - Action: Virtual Pool, Create Virtual Server Pool
Specify the parameters for creating the virtual pool including the first VM server host name, Enterprise manage agent that is used to communicate with the VM server, Oracle user passwords on the VM server and the Grid control host, etc.

VM server pool kblade1 successfully created with the first VM server kblade9 registered as:
Add additional virtual servers by repeating the step 'Meet the pre-requisites in the VM server' on kblade10 and register it to the VM server pool kblade1:

VM server Kblade10 is added to VM server pool kblade1 successfully
The third VM server Klade11 is added to VM server pool kblade1 in the same way.

Create guest VMs using a VM template

The VM server pool and VM servers associated with the VM server pool provide the resources for the virtual Grid. These resources include CPUs, memory, disk storages and network resources and VM templates. As the basic element of the virtual Grid, guest VMs (virtual machines) can be created and run on one of virtual servers with the specific amount of resources such as number of virtual CPUs, memory and disk spaces. The guest VM will contain its own operating system and associated application software.

Depending on the different virtualized method, a virtual machine can be full virtualized (also known as hardware virtualized (HVM)) or paravirtualized. In full virtualized (hardware virtualized), the unmodified operating system runs on the virtual machine, while in the paravirtualized method, the guest operating system is recompiled. The paravirtualized virtual machine runs on near native speed. In this proof of concept, only paravirtualized VMs are created to form the virtual Grid.

A guest VM can be created in several different methods:
   - By a virtual machine template
   - From an installation media
   - From a network bootable (PXE boot) virtual machine.

In this proof of concept project, we only use virtual machine templates to create the guest VM.
Depending on the purpose of the guest VM, a different guest VM template can be used to create the guest VM. Use the following steps to create a guest VM from a template downloaded from Oracle:

1. Download the Oracle Linux OEL5.2 template zip file `OVM_EL5U2_X86_64_11GRAC_PVM.gz` from Oracle VM template download website: [http://edelivery.oracle.com/oraclevm](http://edelivery.oracle.com/oraclevm)

2. Unzip this file to `/OVS/seed_pool` to form the `OVM_EL5U2_X86_64_PVM_10GB` directory which includes three files: README, System.img, vm.cfg

3. Discover the VM template:
   In the VM management pack Virtualization Central tab, select VM server pool `kblade1` and the **Action**: Virtual server Pool, Discover Guest VM

Select undiscovered template:
OVM_EL5U2_X86_64_PVM_10GB template is discovered and is listed as a VM template resource in VM server pool kblade1 and is ready to be used to create a guest VM.

4. Create a guest VM using the template
   In virtual center, select **Action**: guest VM, create guest VM
   Specify the server pool name and preferred server list and select the VM template
   Create the guest VM
Specify the parameters for resource assignment for this guest VM such as virtual CPUs and memory and network configuration, local disks and shared disk.

Guest VM is created and running on VM server kblade 10:
Add additional disks to the guest VM

If the VM template only has one image file, `System.img`, additional disks may need to be added as the local disk partition as well as the shared disks in the guest VM to install the application. For example, if the guest VM is created for Oracle RAC database, an additional local disk needs to be added for Oracle clusterware and Oracle RAC software homes. Shared disks may also need to be added for clusterware OCR and votingdisk and RAC database. The local disk and shared disk can be specified during the guest VM creation process or can be added after the VM creation as follow:

1. **Add local disks:**
   
   In Virtual Central page of VM management Pack: Select the guest VM and select Guest VM: Edit Guest VM from the Action menu and click go to the Edit guest virtual machine page.
   
   Add local disk: Click **Add Disk Dtab** and specify the disk name and size of the local disks.

2. **Create shared disk racdb on server pool and attach to the share disk racdb to the guest VM**
Create shared disk and attach the shared disk to VM

Add shared disk: click the **Attach Share** disk button and specify the name and size of the shared disk to attach.

Now the guest VM has two disks: system for OS, orahome for Oracle software and one share disk oradb for Oracle database.

Two local disks are corresponding two image files in the OVM repository:

```
ls -l /OVS/running_pool/150_big_52vm2/*.img
/OVS/running_pool/150_big_52vm2/orahome.img
/OVS/running_pool/150_big_52vm2/System.img
```

And a shared disk is corresponding the racdb.img in the repository:

```
ls /OVS/sharedDisk/racdb.img
/OVS/sharedDisk/racdb.img.
```

It is also shown as the resources for the guest VM.
In the guest VM OS, the virtual disk partitions can be seen

more /proc/partitions
major minor #blocks name

202 0 13631488 xvda
202 1 104391 xvda1
202 2 13526730 xvda2
16 15360000 xvdb
17 15342043 xvdb1
202 32 21504000 xvdc
202 33 21486906 xvdc1

The virtual disks and the associate image files in the OVM repositories

<table>
<thead>
<tr>
<th>Virtual device Name p in VM</th>
<th>disk name in VM</th>
<th>image file in /OVS</th>
</tr>
</thead>
<tbody>
<tr>
<td>xvda</td>
<td>System</td>
<td>System.img</td>
</tr>
<tr>
<td>xvdb</td>
<td>orahome</td>
<td>orahome.img</td>
</tr>
<tr>
<td>xvdc</td>
<td>racdb</td>
<td>oradb.img</td>
</tr>
</tbody>
</table>

3. Attach disk partition in dom0 to the guest VM:

The disk partition such as the multipath device can be attached to the guest VM by specify the physical disk mapping in the vm.cfg file. The following configuration attached disk partition in dom0 /dev/mapper/vmracdbp1 to the guest VM as the virtual disk /dev/xvdc

vm.cfg:

```
disk = ['phy: /dev/mapper/vmracdbp1, xvdc,w!']
```
This configuration currently can not be configured using Grid control Virtual Management Pack or Vm manager. Manual editing vm.cfg file of the guest VM is required.

For the mapping from the image file in the OVM repository to the virtual device in the guest VM, the following mapping in vm.cfg file was generated by the VM Management Pack:

```plaintext
disk = ['file:/OVS/running_pool/150_big_52vm2/System.img,xvda,w',
       'file:/OVS/running_pool/150_big_52vm2/orahome.img,xvdb,w',
       'file:/OVS/sharedDisk/racdb.img,xvdc,w!'],
```

Configure guest VMs for Oracle 11g RAC database

Oracle 11g RAC database is certified to run on the guest VMs. To configure the guest VMs as the database nodes for Oracle 11g RAC configuration, the guest VMs are configured with the network and shared storage as shown in the following diagram:

![RAC Configure on Guest VMs](image)

Figure 17: guest VMs and OVM server configuration for 11g RAC on VMs. The shared virtual storage should be configured for 11g RAC clusterware as well as the Oracle RAC database. The shared virtual storage can be the shared virtual disks from OVM repositories or the disk partition on OVM server dom0, as illustrated in the previous session. For production database or I/O intensive database, it is recommended to use the disk partition on dom0 for the
shared virtual disk on guest VM.

Section IV – Case Studies of the Grid Hosting Applications

The Grid provides the infrastructure that is designed to consolidate multiple tier applications. The following options are provided:

- High transaction volume database tier can be deployed in the physical Grid to take advantage of the HA and scalability of Oracle 11g RAC database.
- Application tier or middle tier can be deployed in virtual machines in the virtual Grid. By having the virtual machine be dedicated to the application, or middle tier, while multiple virtual machines run on the same physical hardware, the virtual Grid provides the server consolidation as well as OS isolation for the applications so that these applications are running on their own OS and will not affect any other virtual machines.
- The virtual Grid also provides the multiple virtual machines to run the development and test environments which can be provisioned as needed. For example, by developing Oracle 11g RAC on virtual machines, the development and test environments can also simulate the production RAC database environment without actually having multiple physical servers for RAC.

As a part of this Grid proof of concept project, we have built the case studies of consolidating the following applications and databases on this Grid. The following figure shows the Grid structure and the applications of the case studies:
**Oracle E-business suite R12 on RAC/Virtual Machines**

Error! Bookmark not defined.

This Oracle Applications E-Business R12 environment is running on three Applications tier nodes and two database RAC nodes. The three applications nodes are running on three virtual machines on the virtual Grid and the 11g EBS database is running on two physical RAC nodes on the physical Grid. Refer to [11] for details.
Oracle E-Business suite R12 running on Virtual Machines
Both Oracle E-Business Applications tier and database are running on VM as the development and test environment. This enables the fast deployment of the EBS development environment with less physical server needed. Refer to [11] for details.

Banner University ERP Applications suite
The application tier nodes are running on the VMs and the banner database is running on physical grid with the 11g RAC database infrastructure.

Provisioning a two node 11g RAC database on VMs
This case study aims to provision a 11g RAC database on virtual environment using Oracle enterprise manager RAC provisioning pack. Refer to [12] for details.

The following screen shot shows the Grid infrastructure that hosts these applications and/or the databases.

Figure 19: Oracle VM server pool with guest VMs for applications

Section IV – Summary
In this paper, we have discussed the Grid infrastructure which combines the physical Grid of Oracle 11g RAC and the virtual Grid of Oracle VM. As a Poof-of-concept of project, this Grid infrastructure was implemented using the Dell latest Blade server as hardware platform and Equallogic Storage as the shared storage. This paper cover the design and the detailed implementation of such a Grid infrastructure. In the end of the paper, as case studies, we also examined the various options and considerations of consolidating multi-tier enterprise applications such as multiple nodes Oracle E-Business suite applications suit, dev/test Oracle E-Business suite applications instances, and Oracle 11g RAC database on this pre-built Grid infrastructure. The results of the Poof-of-Concept project showed that this Grid infrastructure can consolidate multiple enterprise applications and provide the system resource pool that can be allocated to applications on demand.

We would like to acknowledge the support of following Oracle engineers for this project: Akanksha Sheoran, Rajat Nigam, Daniel Dibbets, Kurt Hackel, Channabasappa Nekar, Premjith Rayaroth, and Dell Engineer: Roger Lopez

References:
2 Oracle® Enterprise Manager Concepts 10g Release 5 (10.2.0.5) Part Number B31949-10
3 Oracle Enterprise Manager Grid Control ReadMe for Linux x86-64 10g Release 5 (10.2.0.5) April 2009
4 How to Enable Oracle VM Management Pack in EM Grid Control 10.2.0.5, Metalink Note #: 781879.1
5 Oracle VM: Converting from a Local to Shared OVS Repository, Metalink note # 756968.1
6 How to Add Shared Repositories to Oracle VM Pools with Multiple Servers, Metalink Note # 869430.1
7 Implementing Oracle Grid: A successful Customer Case Study Kai Yu, Dan Brint, IOUG Select Journal Volume 16, Number 2, Second Quarter 2009
8 Deploying Oracle VM Release 2.1 on Dell PowerEdge servers and Dell/EMC storage Dell white paper
9 Dell Reference Configuration Deploying Oracle® Database on Dell™ EqualLogic™ PS5000XV iSCSI Storage A Dell Technical White Paper, Wendy Chen
10 Building an Oracle Grid with Oracle VM on Dell Blade Servers and EqualLogic iSCSI storage, Kai Yu, David Mar, Oracle OpenWorld 2009 Presentation, Session #S308185
11 Oracle E-Business Suite on Oracle RAC and Oracle VM: Architecture and Implementation, Kai Yu, John Tao, Oracle OpenWorld 2009 Presentation, Session #S310132
12 Provisioning Oracle RAC in a Virtualized Environment, Using Oracle Enterprise Manager, Kai Yu, Rajat, Niagm, Oracle OpenWorld 2009 Presentation, Session #S312109