



Red Hat Performance Briefs

Virtualized Oracle Database Deployments using Red Hat Enterprise Linux with KVM

**End-to-end hardware infrastructure from Dell allows
business continuity with seamless VM migrations**

Sanjay Rao, Tim Wilkinson
Performance Engineering

Version 1.0
September 2012





1801 Varsity Drive™
Raleigh NC 27606-2072 USA
Phone: +1 919 754 3700
Phone: 888 733 4281
Fax: +1 919 754 3701
PO Box 13588
Research Triangle Park NC 27709 USA

Linux is a registered trademark of Linus Torvalds. Red Hat, Red Hat Enterprise Linux and the Red Hat "Shadowman" logo are registered trademarks of Red Hat, Inc. in the United States and other countries.

Intel, the Intel logo and Xeon are registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Dell, the Dell logo, EqualLogic and PowerEdge are trademarks of Dell, Inc.

Oracle is a registered trademark of Oracle.

All other trademarks referenced herein are the property of their respective owners.

© 2012 by Red Hat, Inc. This material may be distributed only subject to the terms and conditions set forth in the Open Publication License, V1.0 or later (the latest version is presently available at <http://www.opencontent.org/openpub/>).

The information contained herein is subject to change without notice. Red Hat, Inc. shall not be liable for technical or editorial errors or omissions contained herein.

Distribution of modified versions of this document is prohibited without the explicit permission of Red Hat Inc.

Distribution of this work or derivative of this work in any standard (paper) book form for commercial purposes is prohibited unless prior permission is obtained from Red Hat Inc.

The GPG fingerprint of the security@redhat.com key is:
CA 20 86 86 2B D6 9D FC 65 F6 EC C4 21 91 80 CD DB 42 A6 0E

Send feedback to refarch-feedback@redhat.com



Table of Contents

1 Executive Summary.....	1
2 Test Configuration.....	2
2.1 Hardware Configuration.....	3
2.2 Software Configuration.....	4
2.2.1 Storage Configuration.....	4
2.2.2 KVM Storage Pool Configuration.....	5
2.3 Configuring the Hosts for Migration.....	5
2.4 Virtual Machine Creation.....	6
2.5 Install Oracle.....	6
3 Testing	6
3.1 Performing the Migration.....	7
4 Tuning and Monitoring Tools.....	10
4.1 RHEL Performance Tuning and Optimization.....	10
4.1.1 Tuned.....	10
4.1.2 Numad.....	10
4.2 RHEL Performance Monitoring	10
4.2.1 Perf.....	10
4.2.2 Numastat.....	11
4.2.3 Tuna.....	11
4.3 Dell EqualLogic Storage Tools	12
5 What Does It All Mean?.....	13
Appendix A: References and Further Information.....	14



1 Executive Summary

Red Hat partnered with Dell to demonstrate deploying an Oracle database in a virtualized environment on an end-to-end Dell hardware infrastructure. The test configuration includes Oracle 11g R2 running on Red Hat Enterprise Linux 6 with the Kernel-based Virtual Machine (KVM) hypervisor on Dell PowerEdge R710 servers with Dell EqualLogic PS6010 virtualized iSCSI storage. The architecture is a set of Dell rack-mount servers connected to the PS Series iSCSI-based storage array over 10 Gigabit Ethernet (10 GbE). This document includes the specifications and instructions for recreating the deployment of virtual machines (VMs), achieving consistent performance, and a high-availability scenario for the Oracle database.

To illustrate that you can make mission-critical Oracle deployments highly available in a virtualized environment, multiple VMs running Oracle are set up on two physical hosts and an online transaction processing (OLTP) workload is executed on each of them. The VMs are then migrated back and forth between hosts while the OLTP workload continues to run uninterrupted.

Migrating VMs between hosts demonstrates one aspect of the enterprise-class service capability of Red Hat Enterprise Linux and inherent scalability and stability of a KVM-based virtualization infrastructure.



2 Test Configuration

The following commercially available, industry-standard hardware and software components are used to build the system under test (SUT) configuration.

Table 1: Hardware Configuration

Server	Dell PowerEdge R710 server 2 Socket – 8 Cores Intel (R) Xeon(R) X56570 @2.93 GHz 64 GB RAM (32 GB per NUMA node) Dual port Intel 82599EB 10-Gigabit HBA
Storage	Dell EqualLogic PS6010 array
Network Switch	Dell PowerConnect 8024F

Table 2: Software Configuration

Operating System (Host)	Red Hat Enterprise Linux 6.3 (2.6.32.279.el6.x86_64) qemu-kvm-0.12.1.2-2.295.el6.x86_64
Operating System (Virtual Machines)	Red Hat Enterprise Linux 6.3 (2.6.32.279.el6.x86_64)
Database	Oracle Database 11g Release 2 (11.2.0.3)
Storage	Dell EqualLogic Host Integration Tools for Linux 1.2



2.1 Hardware Configuration

Figure 1 illustrates the physical connectivity of the hardware used in the demonstration testbed.

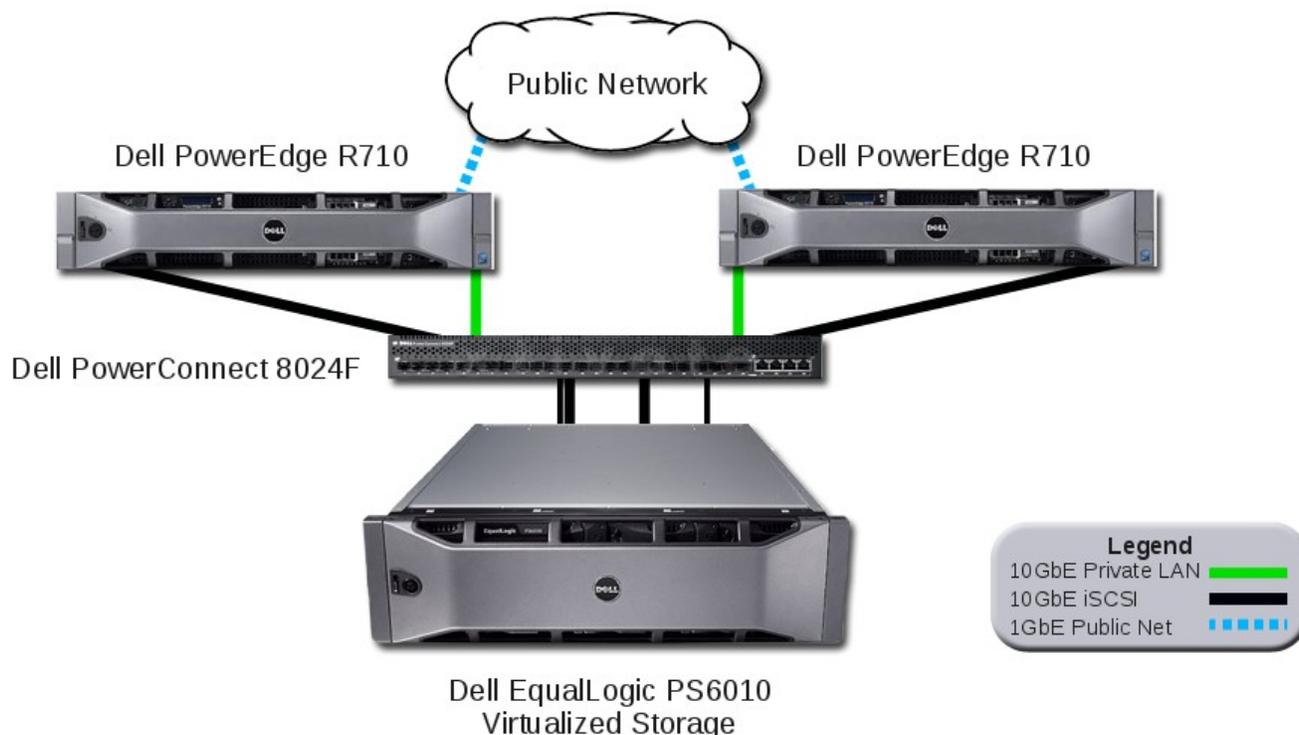


Figure 1: Hardware Connectivity

A dual-port Intel 82599EB 10GbE host bus adapter (HBA) is installed in each server and configured into two subnets to separate network traffic from data traffic. Both ports on each card are physically connected to the Dell PowerConnect switch.

The first port is for data connection to the Dell EqualLogic PS6010 storage array. While only a single port is configured for iSCSI traffic in this particular case, the best practice would be to configure multiple I/O ports for redundancy and bandwidth.

The second port is configured using a different subnet as a private LAN connection between the two hosts, used for networking traffic during the VM migrations.

The Dell EqualLogic PS6010 storage array is connected to the Dell PowerConnect switch and configured via the dedicated subnet used for iSCSI traffic.



2.2 Software Configuration

The Red Hat Enterprise Linux 6 operating system is installed on two identically configured Dell PowerEdge R710 servers by selecting the **Virtual Host** optional package group during the installation. This option installs the kernel as well as the KVM and Virtual Machine Manager tools required to create a host for VMs.

Additional details and screen captures of the installation procedures using the **Virtual Host** option are documented in Chapter 16.19 of the Red Hat Enterprise Linux 6 installation guide located at https://access.redhat.com/knowledge/docs/en-US/Red_Hat_Enterprise_Linux/6/html/Installation_Guide/index.html.

2.2.1 Storage Configuration

Volumes on the array are configured as described in Table 3.

Table 3: Dell EqualLogic Storage Array Configuration

Volume Name	Size	Purpose
kvm1	20GB	System disk for VM1
kvm2	20GB	System disk for VM2
kvm3	20GB	System disk for VM3
kvm4	20GB	System disk for VM4
G1data1	100GB	Data disk 1 for VM1
G1data2	80GB	Data disk 2 for VM1
G2data1	100GB	Data disk 1 for VM2
G2data2	80GB	Data disk 2 for VM2
G3data1	100GB	Data disk 1 for VM3
G3data2	80GB	Data disk 2 for VM3
G4data1	100GB	Data disk 1 for VM4
G4data2	80GB	Data disk 2 for VM4

The volumes detailed were created by logging in to the storage array, creating the volumes, and granting the required iSCSI volume access. Alternatively the Dell EqualLogic Host Scripting Tools could have been used to automate this procedure within a perl or python script.



2.2.2 KVM Storage Pool Configuration

To migrate the VMs from one physical host to the other, both hosts require access to the system disks of each VM. To accomplish this, storage pools are created on each host using `virt-manager`:

1. From the `virt-manager` Edit menu, select Connection Details.
2. In the Storage tab, click the “+” sign in the lower left of the screen.
3. Specify a **Name** for the pool and select the “iscsi: iSCSI Target” option from the **Type** pull-down menu.
4. Do not change default **Target Path** “/dev/disk/by-path”.
5. Set **Host Name** to the IP address of the iSCSI group.
6. Enter the **Source Path** using the IQN of the target device.
7. Click **Apply**.
8. Repeat steps 2 through 7 on both hosts for each of the volumes created as a VM system disk. In this case there are four system disks (see Table 3).

2.3 Configuring the Hosts for Migration

A private LAN subnet is created exclusively for the VM migration and to minimize the impact of additional network traffic on the public LAN. The private LAN is created using 10 GbE to provide adequate bandwidth required for the VM migration. The time required to migrate a VM from one physical host to another is directly proportional to the size of VM’s memory and the amount of memory activity occurring at the time of migration.

Two network bridges are created to provide each VM access to the public network and the iSCSI storage. One bridge (`br0`, using interface `em1`) is used for public network traffic and the other (`br2`, using interface `p2p1`) for the iSCSI storage traffic. The iSCSI storage traffic interface is configured to use jumbo frames.

Add firewall rules to allow SSH and libvirt traffic.

```
# iptables -A INPUT -p tcp --dport 22 -j ACCEPT
# iptables -A OUTPUT -p tcp --sport 22 -j ACCEPT
# iptables -I INPUT -p tcp --dport 49152:49261 -j ACCEPT
```

Set the maximum migration bandwidth to take advantage of the 10GbE interconnect.

```
# virsh migrate-setspeed vm1 100000
# virsh migrate-setspeed vm2 100000
# virsh migrate-setspeed vm3 100000
# virsh migrate-setspeed vm4 100000
```

Configure passwordless SSH between hosts and VMs.

```
# ssh-keygen -t rsa -P ""
# for i in 1 2 3 4 ; do ssh-copy-id -i ~/.ssh/id_rsa.pub vm$i ; done
```



2.4 Virtual Machine Creation

Four VMs are created using the four volumes in the storage pool as system disks. Two virtual network interfaces are created on each VM, one using br0 and the other using br2. Red Hat Enterprise Linux 6 is then installed on each VM.

2.5 Install Oracle

After the VMs are installed and configured, install the RPMs for the Oracle database on each. The database software and OLTP workload kit are then installed. Each VM accesses and mounts the data disks created for it on the storage array.

3 Testing

Figure 3-1 illustrates how both Host 1 and Host 2 are able to access the storage array. VM1 and VM2 are initially running on Host 1 while VM3 and VM4 are running on Host 2.

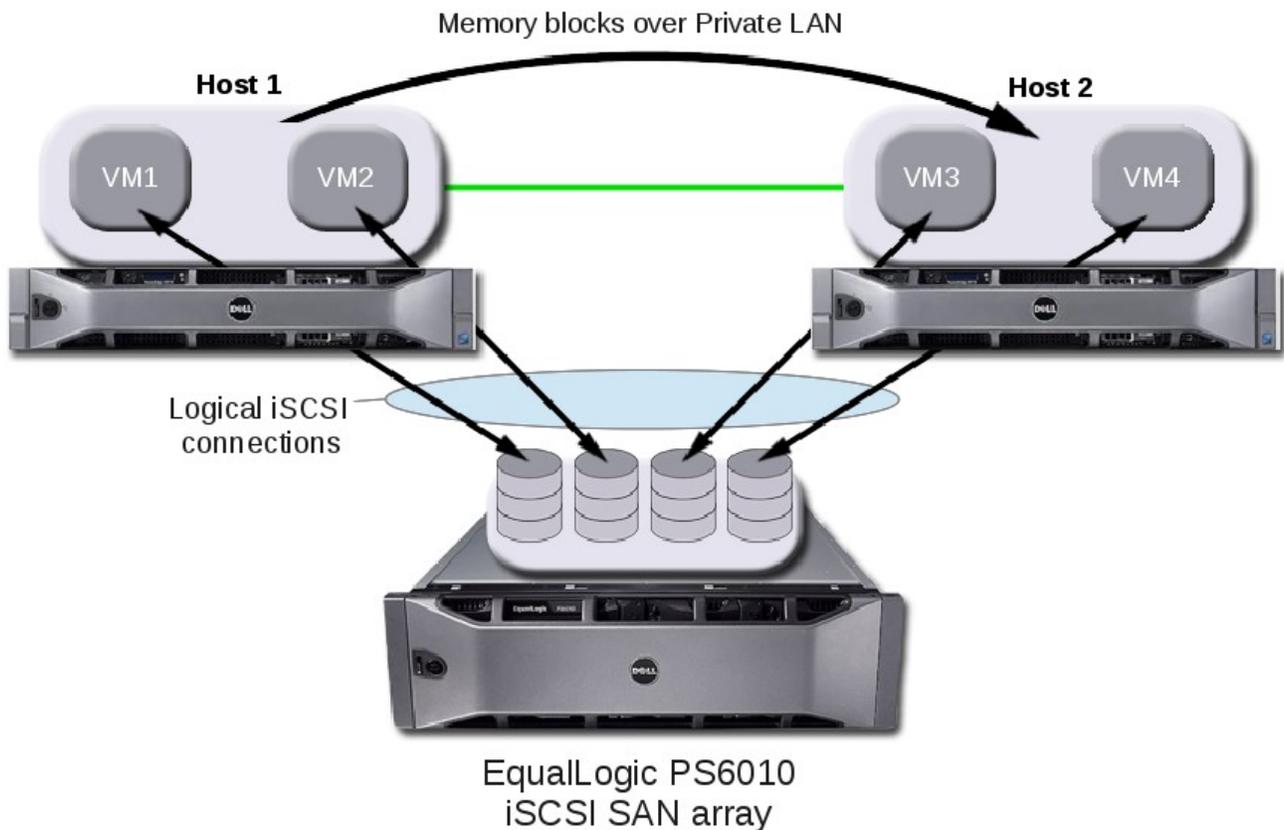


Figure 3-1: VM Migration Activity



3.1 Performing the Migration

Both hosts having physical paths to the storage array allows VMs access to their respective data from either host so that during the migration, only the memory content of the VM needs to be moved from one host to the other. Upon a significant event such as component failure or service outage on Host 1, VM1 and VM2 are instructed to migrate to Host 2 where all four VMs will reside.

```
# virsh migrate --live --verbose vm1 qemu+ssh://root@host2/system  
# virsh migrate --live --verbose vm2 qemu+ssh://root@host2/system
```

As soon as the migration command is issued to the VMs residing on Host 1, the respective VM is initiated on Host 2 and its memory blocks are moved over the private LAN. The VMs continue to operate while the migration activity takes place in the background. Any changes that occur to the memory pages during migration will be copied again.

After all memory is successfully moved, the VM shuts down on Host 1 and becomes fully operational on Host 2. During this procedure, the database OLTP workload in the VMs continues without interruption.

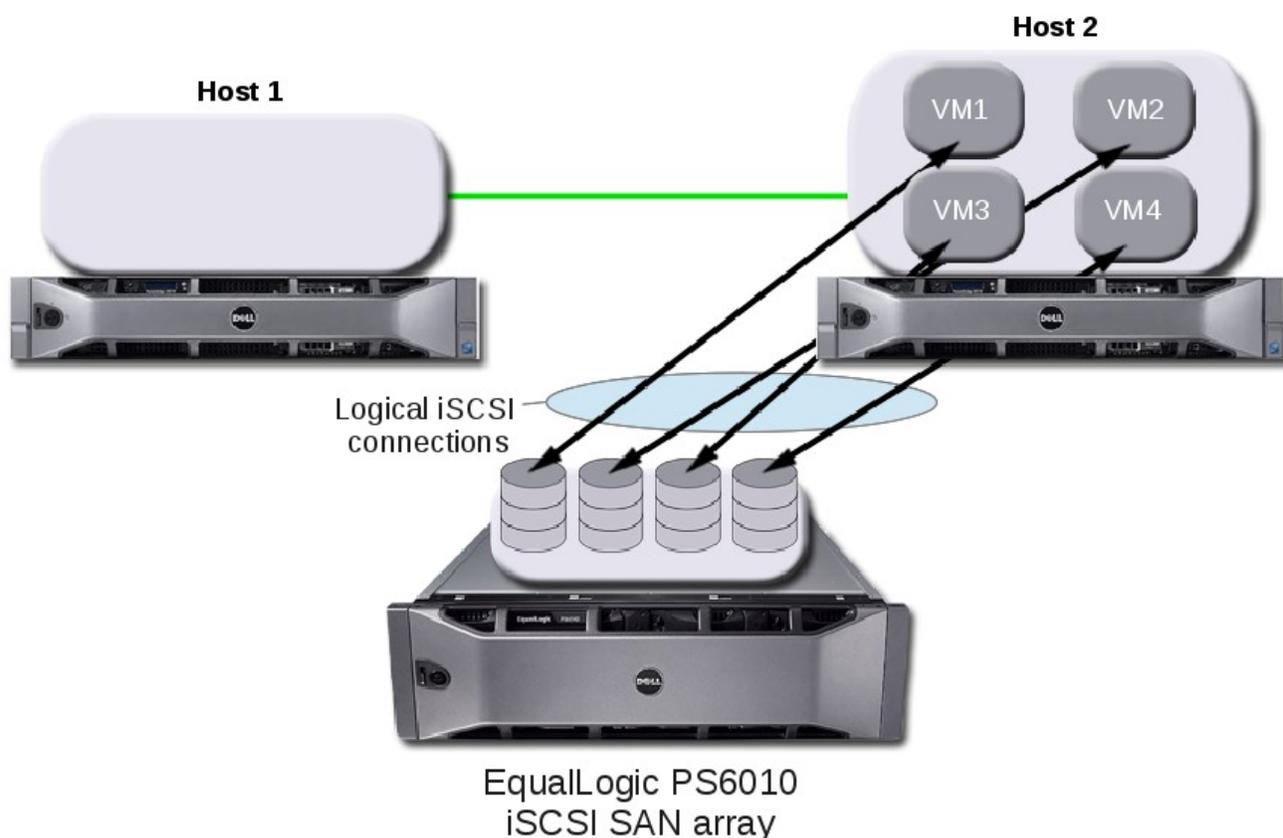


Figure 3.1-1: Post-Migration VM Layout



After the migrations complete and Host 1 becomes available again, the VMs can be spread out across the hosts for optimal balance of resources.

```
# virsh migrate --live --verbose vm1 qemu+ssh://root@host1/system
# virsh migrate --live --verbose vm2 qemu+ssh://root@host1/system
```

After the migrations to Host 1 complete, the VMs return to their preferred residence and again, the database workload in the migrated VMs continues uninterrupted.

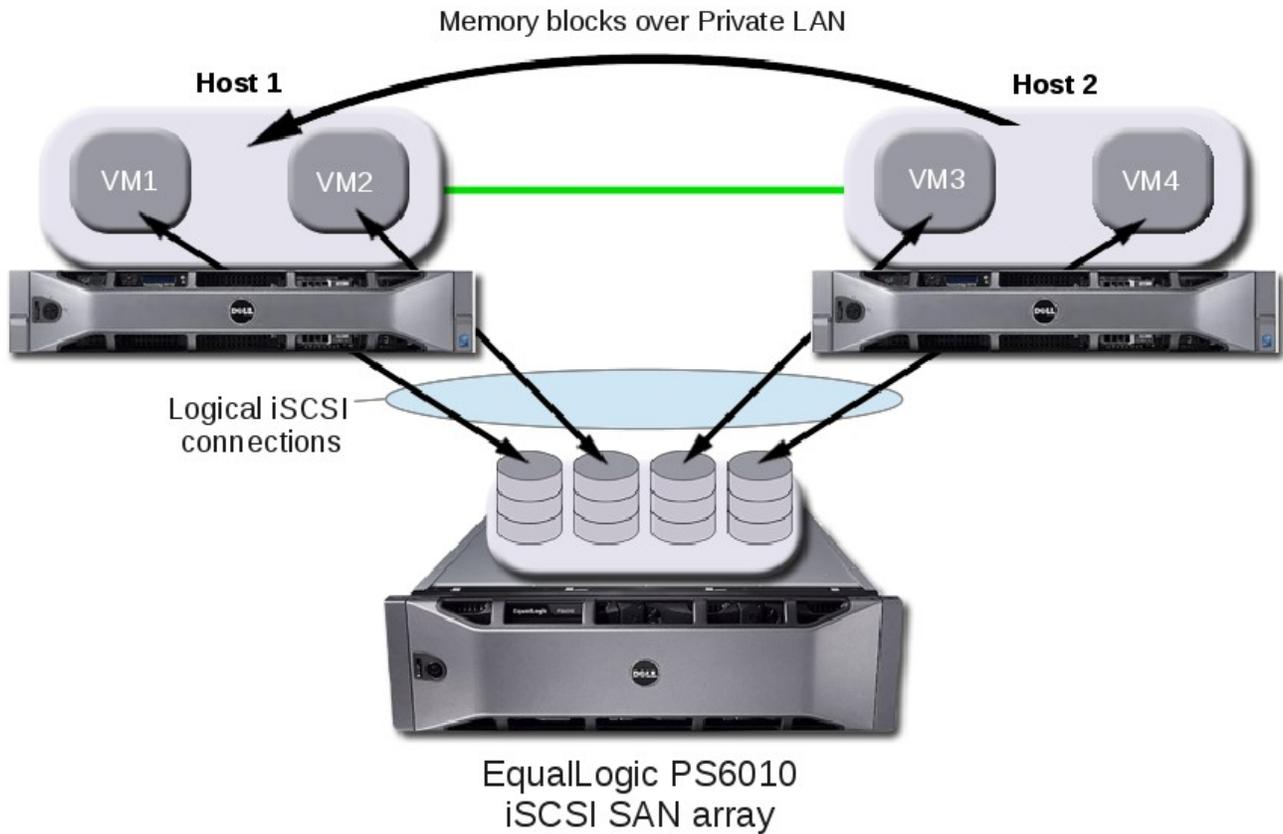
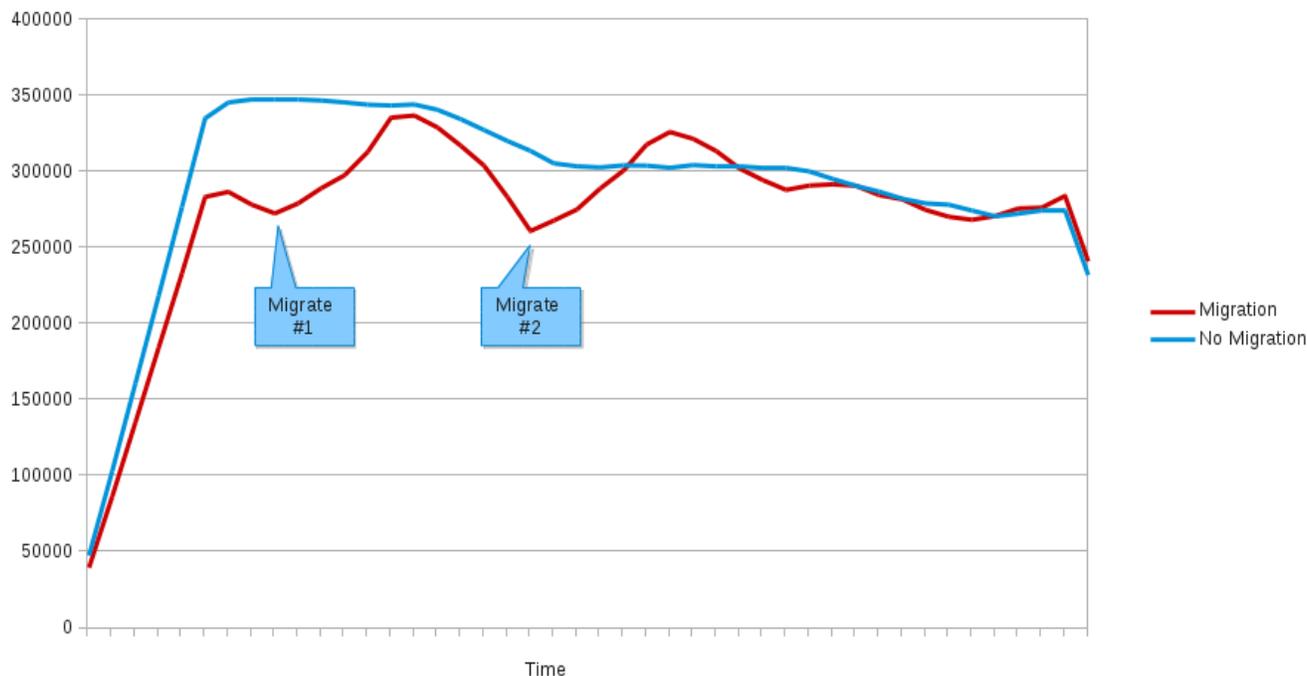


Figure 3.1-2: VM Migration Activity



- Four Virtual Machines migrated between the 2 hosts back and forth
- Workload continues uninterrupted.
- Brief drop in performance, returns back to expected levels once migration is complete

Figure 3.1-3: Migration Performance over Time

Figure 3.1-3 graphs the aggregate transactions per minute (TPM) of four VMs running the database workload twice. This was done once leaving all VMs balanced across hosts and again with the VMs migrated back and forth between hosts. The two performance dips in the results that include migrations correspond to the actual migration of all four VMs from host to host and back again. All database workloads continue uninterrupted and the brief drop in performance returns to expected levels after migrations complete.



4 Tuning and Monitoring Tools

4.1 RHEL Performance Tuning and Optimization

This section describes the tools used for optimizing performance.

4.1.1 Tuned

Tuned is a daemon that configures the system for various performance profiles. It monitors the use of system components and dynamically tunes system settings based on that information. Dynamic tuning accounts for the way that various system components are used differently throughout the uptime for any given system. For example, the hard drive is used heavily during startup and login, but is barely used later when a user might mainly work with applications like OpenOffice or email clients. Similarly, the CPU and network devices are used differently at different times. *Tuned* monitors the activity of these components and reacts to changes in their use. This testing used `tuned -adm` to apply the *virtual-host* and *virtual-guest* profile accordingly.

```
# yum -y install tuned*
# service tuned start
# chkconfig tuned on
# tuned-adm profile virtual-host      # executed on the host servers
# tuned-adm profile virtual-guest    # executed on the VMS
```

4.1.2 Numad

The *numad* package provides a user-level daemon for Non-Uniform Memory Architecture (NUMA) systems that monitors available system resources on a per-node basis and assigns processes to aligned resources for optimum NUMA performance. As an alternative to manual static CPU pinning and memory assignment, *numad* provides dynamic adjustment to minimize memory latency on an ongoing basis. The package also provides an interface that can be used to query the *numad* daemon for the best manual placement of an application and was used to bind KVM VMs optimally on multi-socket x86_64 servers.

4.2 RHEL Performance Monitoring

4.2.1 Perf

`perf` is an easy to use statistical profiling tool that ships with Red Hat Enterprise Linux 6. It provides a number of useful performance counters that let the user assess the impact of commands on their system and is useful in locating system resource bottlenecks. It can report live profiling or record a profile over a length of time and can report on the saved data later. Further information on the `perf` tool can be found at https://access.redhat.com/knowledge/docs/en-US/Red_Hat_Enterprise_Linux/6/html/Developer_Guide/perf.html.



4.2.2 Numastat

`numastat` displays per-node NUMA hit and miss system statistics and can display per-node memory allocation information for the specified pattern provided.

This example shows the memory pages of all four VMs spread across both NUMA nodes on the host not running `numad`.

```
# numastat kvm
Per-node process memory usage (in Mbs):

      PID                Node 0                Node 1
-----
15222 (qemu-kvm)         6478.90            3075.53
15284 (qemu-kvm)         4030.89            5540.93
15353 (qemu-kvm)         3437.39            6127.42
15420 (qemu-kvm)         6878.36            2666.66
```

The following example shows the memory page locales of each VM on the host running the `numad` service. In this case, VM memory no longer crosses NUMA node boundaries.

```
# numastat kvm
Per-node process memory usage (in Mbs):

      PID                Node 0                Node 1
-----
 439 (qemu-kvm)          9045.89              0.03
28731 (qemu-kvm)          4.75                9073.39
29864 (qemu-kvm)          9054.27              0.03
31170 (qemu-kvm)          4.74                9044.47
```

4.2.3 Tuna

Tuna can be used to modify thread attributes (processor affinity, scheduling policy, and scheduler priority) and interrupts (processor affinity). It is designed to be used on a running system where changes take place immediately. This allows any application-specific measurement tools to see and analyze system performance immediately after the changes have been made.



4.3 Dell EqualLogic Storage Tools

Dell EqualLogic PS Series storage implements an all-inclusive software model, with a rich collection of capabilities in host integration, storage management and monitoring. The Host Integration Tools (HIT) for Linux software simplifies Linux host iSCSI configuration and storage administration. HIT/Linux includes eqltune, for automated host analysis and configuration. In addition, HIT/Linux provides intelligent Multipath IO management, optimizing both network and SAN hardware utilization. Additionally, the Host Scripting Tools allow PS Series group management commands to be scripted in perl or python host-side scripts.

The PS Series Group Manager is provided in both GUI and command line versions. The GUI version is a web browser based, portable java application. The Group Manager includes wizard-based storage configuration features, along with a simple and intuitive user interface.

SAN Headquarters is part of the all-inclusive software suite provided with Dell EqualLogic PS Series storage. A full featured SAN performance monitoring and capacity analysis tool, SAN HQ provides both live data capture capabilities and detailed historical reporting. It also can connect to multiple PS Series Groups, providing multi-site monitoring from a single client.



5 What Does It All Mean?

The testing conducted by Red Hat and Dell proves that Red Hat Enterprise Linux with the KVM hypervisor is an ideal platform for building highly-available virtualized Oracle databases on commodity x86 hardware. Built on top of end-to-end Dell hardware infrastructure, this solution represents collaborative efforts in which Red Hat and its industry partners participate to ensure interoperability and performance that directly benefit their mutual customers. The Dell EqualLogic Host Integration Tools simplify configuration and automate iSCSI storage operations. This type of testing removes the burden from customers of having to try out every solution in their own environment or the uncertainty of having to deploy an untested solution.

Testing also demonstrates that customers deploying databases and applications in virtualized environments not only have an opportunity to increase physical server utilization, but also gain an ability to increase availability of their enterprise infrastructure by seamlessly migrating VM instances to different physical hardware without degradation of services provided.

Additionally, customers could rely on VM migrations in the following cases:

- Hardware upgrades – By moving the VMs to newer hardware, users can take advantage of performance improvements without having to change the operating environment of their applications.
- Hardware maintenance – Users do not have to bring down their operating environments for standard maintenance, but instead can migrate to different hardware, complete the maintenance effort on affected systems, and migrate the VMs back to the original hardware.
- Failover – In the event of a physical host failure VMs can be started on another (standby or even active) host. When the outage ends, those VMs can be migrated without interruption back to the original host. While that automated functionality is outside the scope of this document it could be accomplished with Red Hat Cluster Suite (RHCS) software. Refer to list of supporting documentation in Appendix A for further information.

In conclusion, if your organization is looking to ensure business continuity by maintaining access to your mission-critical databases, Red Hat and Dell offer best practices and proven implementation steps for setting up the Oracle Database and storage, and configuring, running, and migrating VMs. The latest features of Red Hat Enterprise Linux allow you to automatically manage memory locality and the footprint of several VMs in multi-tenancy environments.



Appendix A: References and Further Information

1. For details on how to deploy Oracle Database 11g on Red Hat Enterprise Linux 6 for several types of back-end storage, including Fusion-io ioDrives, see “Oracle Database 11g Release 2 on Red Hat Enterprise Linux 6: Deployment Recommendations,” March 2012, <http://www.redhat.com/resourcelibrary/reference-architectures/deploying-oracle-11gr2-on-rhel-6>
2. For more information on Red Hat Cluster Suite deployment and best practices watch this free webinar, Deploying a highly available service with Red Hat Cluster Suite, <https://www.redhat.com/about/events-webinars/webinars/2012-05-08-taste-of-training-deploying-a-highly-available-service-with-red-hat-cluster-suite>
3. For information on Dell EqualLogic PS Series virtualized storage, <http://www.dell.com/equallogic>
4. To download Dell EqualLogic PS Series software (HIT/Linux, Host Scripting Tools, ...), login to the EqualLogic Support site, <https://support.equallogic.com/secure/login.aspx>
5. For information on Dell PowerEdge Servers, <http://www.dell.com/poweredge>
6. For information on Dell PowerConnect Switches, <http://www.dell.com/us/enterprise/p/switch-powerconnect>