Oracle Databases on VMware

VMware vSphere® 5 RAC Workload Characterization Study (VMware VMFS)

December 2011

© 2011 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. This product is covered by one or more patents listed at http://www.vmware.com/download/patents.html.

VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

VMware, Inc.
3401 Hillview Ave
Palo Alto, CA 94304
www.vmware.com

Contents

© 2011 VMware, Inc. All rights reserved.

© 2011 VMware, Inc. All rights reserved.

1. Executive Summary

An increasing number of datacenters are adopting virtualization-first policies. Historically, Tier 1 workloads generated by applications deployed on Oracle Real Application Clusters (RAC) have been a barrier to this adoption. This is no longer the case because Oracle now provides support for Oracle RAC on VMware, and VMware vSphere® 5.0 can handle the intensive workloads of business-critical applications. This document describes the architecture of a multi-node Oracle RAC deployment in a virtualized VMware environment and shows how it processes a heavy OLTP load with good performance.

The following is demonstrated:

- Configuration of hardware, storage, and network components comprised of:
  - 12-core Cisco UCS blade servers.
  - EMC VNX5500 Unified storage array leveraging EMC Fully Automated Storage Tiering (FAST) and EMC FAST Cache.
- Configuration of vSphere based virtual machines which included:
  - A 4-node Oracle RAC installation across four VMware ESXi™ hosts running vSphere 5.0.
  - Each virtual machine (that is, each Oracle RAC node) was configured with twelve vCPUs and 160GB RAM.
- Storage layout for a 1TB+ database backed by VMware vSphere® VMFS datastores loaded with the Oracle SwingBench large order entry benchmark schema.
- A load test of 1000 OLTP users over a 24-hour period experienced no performance issues or degradation in transaction performance.
- Elimination of application downtime from planned server maintenance (made possible by live migration of Oracle RAC nodes using VMware vSphere® vMotion®).

© 2011 VMware, Inc. All rights reserved.

2. Introduction

This paper describes a workload characterization conducted on a deployment of Oracle 11g R2 RAC on VMware vSphere 5 with Cisco UCS servers and EMC VNX5500 Unified storage. It further describes the architecture and demonstrates how the deployment can sustain a heavy OLTP load without any degradation to transaction performance.

2.1 Oracle Real Application Clusters

Oracle RAC is a cluster database with a shared cache architecture that provides highly scalable and available database solutions for business applications. Oracle RAC is a key component of Oracle enterprise grid architecture and uses Oracle Clusterware for the internode communication required in cluster database environments to enable node interaction. Clusterware is the technology that transforms multiple servers into a cluster.

In a typical Oracle RAC installation, Oracle Automatic Storage Management (ASM) is the underlying, cluster volume manager. Oracle ASM provides the performance of raw I/O with the easy management of a file system, and is the basis for a shared storage pool in Oracle enterprise grid architectures.

2.2 Virtualization with vSphere

vSphere virtualizes and aggregates the underlying physical hardware resources across multiple systems and provides pools of virtual resources to the datacenter. Customers can achieve tangible savings from this consolidation, and realize operational cost savings from reduced datacenter floor space, power, and cooling. The VMware terminology, products, and features used in this document are as follows:

- VMware vSphere – A virtualization platform on which to build and deploy a private cloud that increases control through service-level automation, allowing resources to be pooled to deliver IT as a Service (ITaaS).
- VMware vCenter Server™ – Manages vSphere by providing unified management of all hosts and virtual machines in the datacenter from a single console that monitors aggregate performance of clusters, hosts, and virtual machines. VMware vCenter™ gathers performance metrics that you can view in graphical format with user-friendly charts.
- `esxtop` – A utility included with vSphere that can monitor and collect data for CPU, memory, disk, and network. This utility runs at the VMware ESX® or ESXi hypervisor level and is used for advanced troubleshooting. Most common resource metrics can be obtained from vCenter performance charts.
- VMware vSphere VMFS – Virtual Machine File System which is VMware’s cluster file system used by ESX/ESXi. It stores virtual machine disk images.
- VMware vSphere vMotion – Enables the live migration of running virtual machines from one physical server to another with zero downtime and complete transaction integrity. vMotion enables users to perform hardware maintenance without scheduled downtime and to proactively migrate virtual machines away from failing or underperforming servers.
- VMware vSphere High Availability (HA) – Provides easy to use, cost-effective high availability for applications running in virtual machines. In the event of server failure, affected virtual machines are automatically restarted on other production servers with spare capacity.

© 2011 VMware, Inc. All rights reserved.

2.3 Oracle RAC on VMware Support

In November 2010, Oracle included Oracle RAC in its VMware support statement which is defined in document ID #249212.1, available on MyOracleSupport.com. The VMware support policy on Oracle is defined at http://www.vmware.com/support/policies/oracle-support.html. VMware provides support for their customers running Oracle products including Oracle RAC. VMware will open a support request for all Oracle cases referred to VMware technical support, and will take complete ownership of the issue until resolution.

2.4 Features of Virtualized Oracle RAC

The deployment of Oracle RAC on VMware is similar to physical environments except that each node corresponds to a separate virtual machine that typically resides on a separate ESX/ESXi host. The features of a virtualized Oracle RAC deployment include the following:

- Facilitates consolidation – While each Oracle RAC node virtual machine resides on a separate ESX/ESXi host, spare capacity on each ESX/ESXi host allows hosting of other virtual machines.
- VMware templates can quickly provision new Oracle RAC nodes – The operating system can be pre-installed and patched in a template from which further virtual machine Oracle RAC nodes can be created.
- In training environments Oracle RAC can be installed in multiple virtual machines on the same ESX/ESXi host. This minimizes the need for multiple servers, but still enables Oracle RAC functionality to be tested and deployed in a similar manner to production environments.

2.5 vMotion and VMware HA

vMotion technology can move a live, running Oracle database virtual machine from one ESX/ESXi host to another with no downtime. Because virtual machines running on vSphere are abstracted from the underlying hardware, vMotion can even move virtual machines across hardware from different vendors and between physical machines or servers having different hardware configurations (as long as the CPUs meet compatibility requirements).

vMotion is an invaluable tool for Oracle database administrators because of its following capabilities:

- Avoids planned downtime – You can move Oracle RAC node virtual machines off an ESX/ESXi host that requires downtime (hardware replacement/upgrade, firmware upgrade, and the like) with no loss of service.
- Simplifies server refresh cycles – Server refresh cycles can be challenging as the application and operating system typically need to be re-installed. With vMotion, moving an Oracle RAC node onto new hardware can be done in minutes, with no downtime.
- Troubleshooting – Moving Oracle RAC node virtual machines onto a different ESX/ESXi host can be an effective tool for troubleshooting suspected issues with underlying hardware.

© 2011 VMware, Inc. All rights reserved.

An Oracle RAC node failure in a virtual deployment results in user failover that is the same as physical RAC environments, that is, user sessions fail over to the remaining nodes (assuming configuration of Oracle session failover functionality, TAF). A multi-node Oracle RAC deployment by design is highly available, so the use of VMware HA in this environment is not critical for protection against hardware failure. However, VMware HA can coexist with and complement a virtualized Oracle RAC installation in the following ways:

- While Oracle RAC maintains database availability, VMware HA can automatically restart the failed RAC virtual machine node on another ESX/ESXi host where no other Oracle RAC node exists, to return to full capacity as soon as possible. As of vSphere 4.0 and later, affinity rules can enforce placement of Oracle RAC virtual machines such that they reside on separate ESX/ESXi hosts. Note that the restart of the Oracle RAC node virtual machine on a new ESX/ESXi host after a VMware HA event requires that the target ESX/ESXi host be licensed. For more information see *VMware High Availability (HA): Deployment Best Practices* at http://www.vmware.com/resources/techresources/10166.
- Automatic restart of Oracle RAC virtual machine nodes can be disabled if there is an ESX/ESXi host failure. In this case, impacted user sessions fail over and continue to be processed on the remaining nodes in a degraded state (because there are fewer nodes). This is only temporary until the failed server is repaired and brought back into the cluster, at which point the failed virtual machine node can be manually restarted.

2.6 Large Scale Order Entry Benchmark Kit – SwingBench

The SwingBench kit is comprised of scripts and load drivers that generate business transactions that simulate I/O-intensive large scale order entry OLTP loads. It is similar to a TPC-C workload generator that includes a data generator tool which was used to create larger schemas that generate much higher levels of I/O (larger index lookups). This workload has a read/write ratio of 60/40.

More information about the benchmark kit is available at http://www.dominicgiles.com/largesoe.html.

© 2011 VMware, Inc. All rights reserved.

3. Architecture

This section describes the physical and logical architecture, hardware and software used, network configuration, storage layout and workload test conducted with four-node Oracle RAC virtual machines on vSphere.

Table 1 summarizes the hardware and software used in the tests.

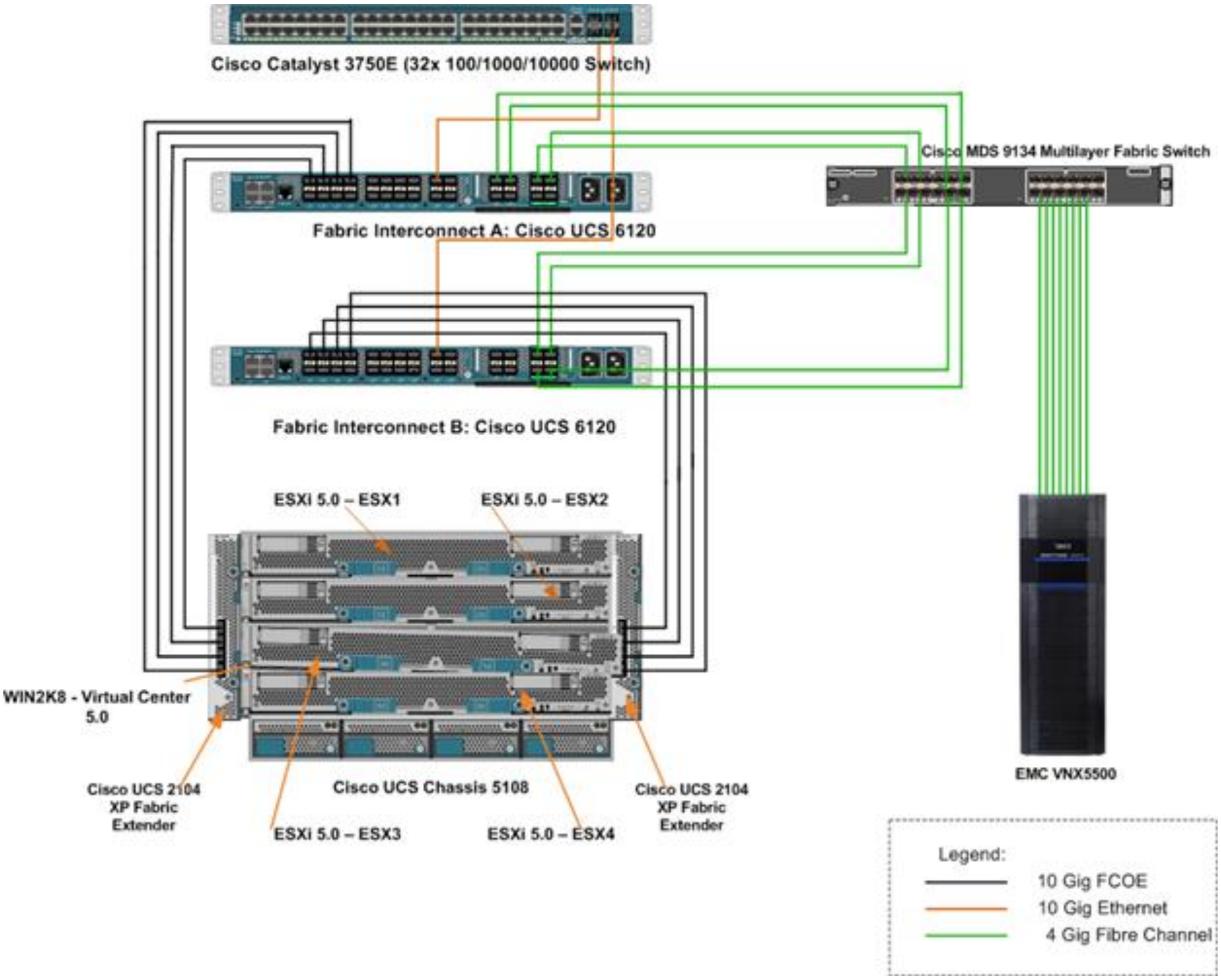
Table 1. Hardware and Software Used

Item	Description
Hardware	• 4x Cisco UCS B-250 blade server: 2 socket, 6 cores each, Intel Xeon X5670 @ 2.93GHz 192GB RAM (running ESXi 5.0) • 1x B200 blade server for SwingBench load generator virtual machine • 1x B200 blade server running Windows Server 2008 and vCenter • Network: 10Gb Ethernet • Shared storage: EMC VNX5500 Unified storage, running: VNX OE for block: level 05.31.000.5.008 VNX OE for file: level 7.0.13
VMware software	• ESXi 5.0 • vCenter Server 5.0 • vCenter Client 5.0
Guest operating system, Oracle database software in virtual machine and multipathing software	• Oracle Enterprise Linux 5.5 x86_64 • Oracle Database 11g R2 (with Oracle RAC and Oracle Grid Infrastructure) Enterprise Edition • EMC PowerPath/VE Version 5.7
Oracle OLTP workload	• Oracle SwingBench 2.4 Load Generator – Large scale order entry benchmark

© 2011 VMware, Inc. All rights reserved.

Figure 1 shows the physical architecture used for the tests.

Figure 1. Physical Architecture

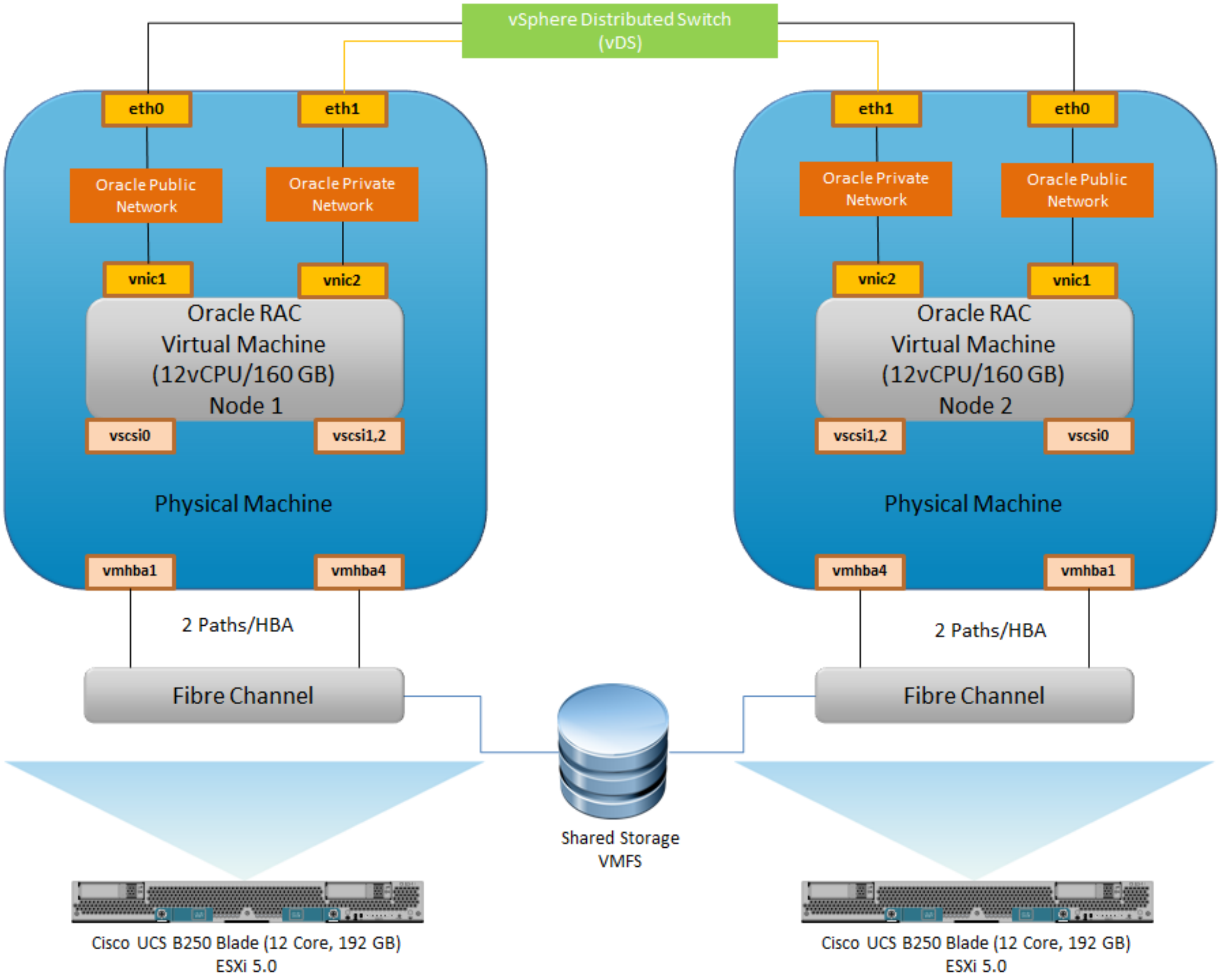

Cisco Catalyst 3750E (32x 100/1000/10000 Switch)
Cisco MDS 9134 Multilayer Fabric Switch
Fabric Interconnect A: Cisco UCS 6120
Fabric Interconnect B: Cisco UCS 6120
ESXi 5.0 – ESX1
ESXi 5.0 – ESX2
WIN2K8 - Virtual Center 5.0
Cisco UCS 2104 XP Fabric Extender
Cisco UCS Chassis 5108
Cisco UCS 2104 XP Fabric Extender
ESXi 5.0 – ESX3
ESXi 5.0 – ESX4
EMC VNX5500
Legend:
10 Gig FCOE
10 Gig Ethernet
4 Gig Fibre Channel

The server and network hardware is based on the Cisco Unified Computing System which provides compute, network, virtualization, and storage access resources that are centrally controlled and managed. These resources are stateless and are provisioned dynamically by the Cisco UCS Manager which handles every aspect of system configuration, from a server's firmware and identity settings to the network connections that connect storage traffic to the destination storage system.

© 2011 VMware, Inc. All rights reserved.

Figure 2 shows two of the four Oracle RAC nodes.

Figure 2. Logical Architecture – Oracle RAC nodes (Only Two Shown)



Each virtual machine Oracle RAC node was configured with 12 vCPU and 160GB of memory. Each ESX host had 12 cores and hyper-threading was enabled so that there were a total of 24 threads/logical CPUs.

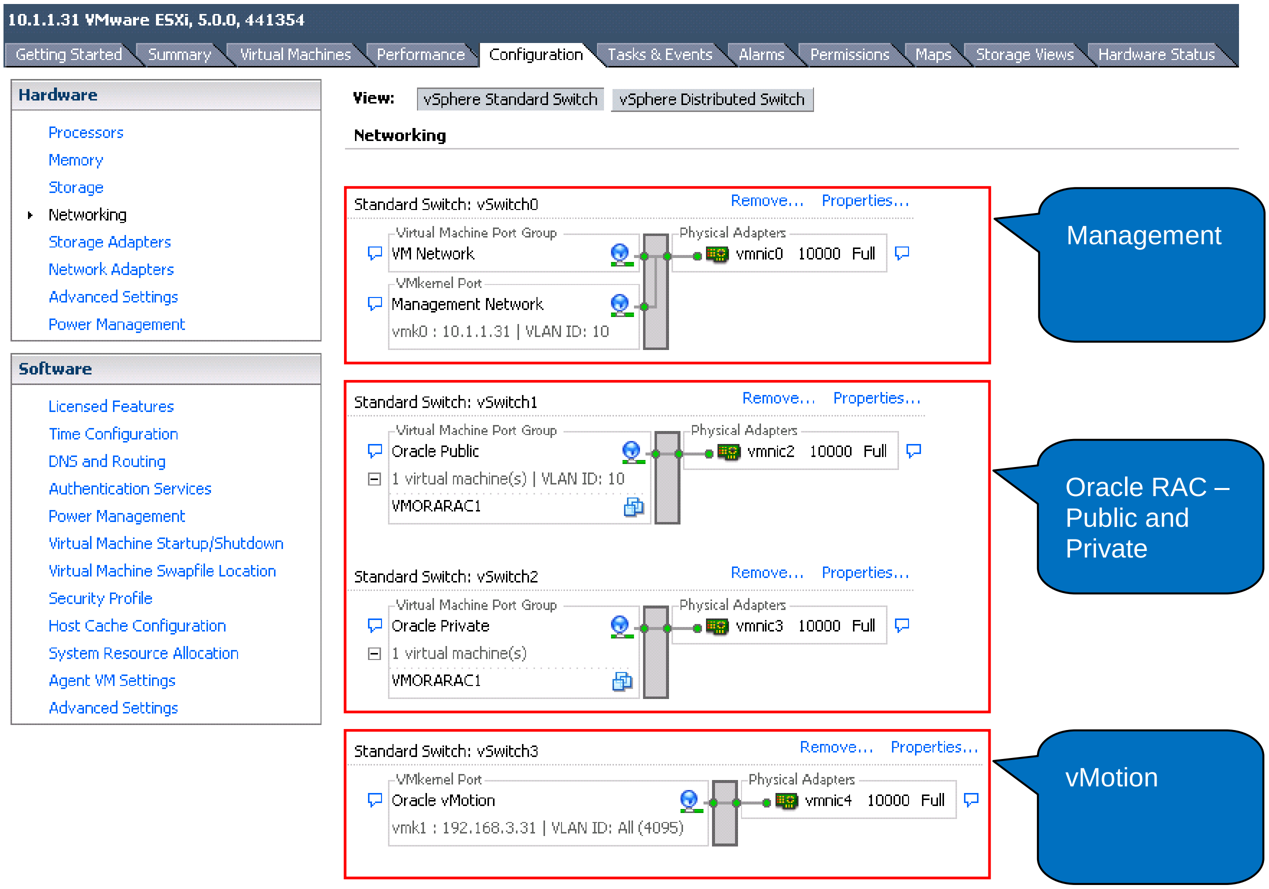
© 2011 VMware, Inc. All rights reserved.

3.1 Network Configuration

In this example, four separate vSwitches (corresponding to four separate physical subnets) were created in each ESXi host for the VMware system console, vMotion operations, public networks, and private networks. VMware recommends vSphere's Distributed Switch (vDS) as it spans many vSphere hosts and aggregates networking to a centralized cluster level administration and monitoring through VMware vCenter.

Figure 3 shows the virtual switch configuration as viewed from vCenter.

Figure 3. Virtual Switch Configuration



10.1.1.31 VMware ESXi, 5.0.0, 441354

Networking

Standard Switch: vSwitch0
Virtual Machine Port Group: VM Network
VMkernel Port: Management Network
vmk0 : 10.1.1.31 | VLAN ID: 10
Physical Adapters: vmnic0 10000 Full

Standard Switch: vSwitch1
Virtual Machine Port Group: Oracle Public
1 virtual machine(s) | VLAN ID: 10
VMORARAC1
Physical Adapters: vmnic2 10000 Full

Standard Switch: vSwitch2
Virtual Machine Port Group: Oracle Private
1 virtual machine(s)
VMORARAC1
Physical Adapters: vmnic3 10000 Full

Standard Switch: vSwitch3
VMkernel Port: Oracle vMotion
vmk1 : 192.168.3.31 | VLAN ID: All (4095)
Physical Adapters: vmnic4 10000 Full

Management

Oracle RAC – Public and Private

vMotion

The ESXi host/UCS blade network adaptors are based on the Cisco UCS Fabric Interconnect, which consisted of a 10Gbps unified fabric that combines storage and network I/O. While NIC redundancy was not factored into this lab design, it is a best practice recommendation for production deployments. Follow the VMware networking best practice guidelines described in *Performance Best Practices for VMware vSphere 5.0* at http://www.vmware.com/pdf/Perf_Best_Practices_vSphere5.0.pdf.

© 2011 VMware, Inc. All rights reserved.

3.2 Storage Configuration

3.2.1 EMC VNX Storage Platforms

The EMC VNX family of storage systems represents EMC's next generation of unified storage optimized for virtualized environments. The massive virtualization and consolidation trend of servers demands a new storage technology that is dynamic and scalable. The EMC VNX series offers several software and hardware features for optimally deploying mission-critical enterprise applications.

A key distinction of this new generation of platforms is support for both block- and file-based external storage access over a variety of access protocols, including Fibre Channel (FC), iSCSI, FCoE, NFS, and CIFS network shared file access. Furthermore, data stored in one of these systems, whether accessed as block- or file-based storage objects, is managed uniformly via Unisphere, a Web-based interface window.

The new VNX storage family from EMC now supports the 2.5-inch SAS drives in a 2U disk array enclosure (DAE) that can hold up to 25 drives, one of the densest offerings in the industry. For example, compared to the older-generation technology of storing 15x 600GB worth of data using the 3.5-inch FC drives in a 3U DAE, the new DAE using 25x 600GB drives in a 2U footprint means an increase by 2.5 times. The power efficiency of the new DAEs also makes it more cost-effective to store the increased data in this much more compact footprint without the need to increase power consumption and cooling.

The data points discussed in this paper were generated on a VNX5500 model.

3.2.2 FAST VP

VNX FAST VP is a policy-based auto-tiering solution for enterprise applications. FAST VP operates at a granularity of 1GB, referred to as a "slice." The goal of FAST VP is to efficiently utilize storage tiers to lower customers' TCO by tiering colder slices of data to high-capacity drives, such as NL-SAS, and to increase performance by keeping hotter slices of data on performance drives, such as flash drives. This occurs automatically and transparently to the host environment. High locality of data is important to realize the benefits of FAST VP. When FAST VP relocates data, it moves the entire slice to the new storage tier. To successfully identify and move the correct slices, FAST VP automatically collects and analyzes statistics prior to relocating data. Customers can initiate the relocation of slices manually or automatically by using a configurable, automated scheduler that can be accessed from the Unisphere management tool. The multitiered storage pool allows FAST VP to fully utilize all three storage tiers: flash, SAS, and NL-SAS. The creation of a storage pool allows for the aggregation of multiple RAID groups, using different storage tiers, into one object. The LUNs created out of the storage pool can be either thickly or thinly provisioned. These "pool LUNs" are no longer bound to a single storage tier. Instead, they can be spread across different storage tiers within the same storage pool. If you create a storage pool with one tier (flash, SAS, or NL-SAS) then FAST VP has no impact on the performance of the system. To operate FAST VP, you need at least two tiers.

In this lab example, a RAID 5 storage pool was created with 40x 600GB SAS drives and 10x 100GB flash drives.

3.2.3 FAST Cache

In traditional storage arrays, the DRAM caches are too small to maintain the hot data for long periods of time. Very few storage arrays give an option to non-disruptively expand DRAM cache, even if they support DRAM cache expansion. FAST Cache extends the cache available to customers by up to 2TB using flash drives. FAST Cache tracks the data activity temperature at a 64KB chunk size and copies the chunks to the flash drives after its temperature reaches a certain threshold. After a data chunk gets copied to FAST Cache, the subsequent accesses to that chunk of data are served at flash latencies. Eventually, when the data temperature cools down, the data chunks get evicted from FAST Cache and are replaced by newer hot data. FAST Cache uses a simple Least Recently Used (LRU) mechanism to evict the data chunks.

FAST Cache is built on the premise that the overall applications' latencies can improve when most frequently accessed data is maintained on a relatively smaller sized, but faster storage medium, like flash drives. FAST Cache identifies the most frequently accessed data that is temporal in nature and copies it to flash drives automatically and non-disruptively. The data movement is completely transparent to applications, thereby making this technology application-agnostic and management-free. For example, FAST Cache can be enabled or disabled on any storage pool simply by selecting/clearing the "FAST Cache" storage pool property in advanced settings.

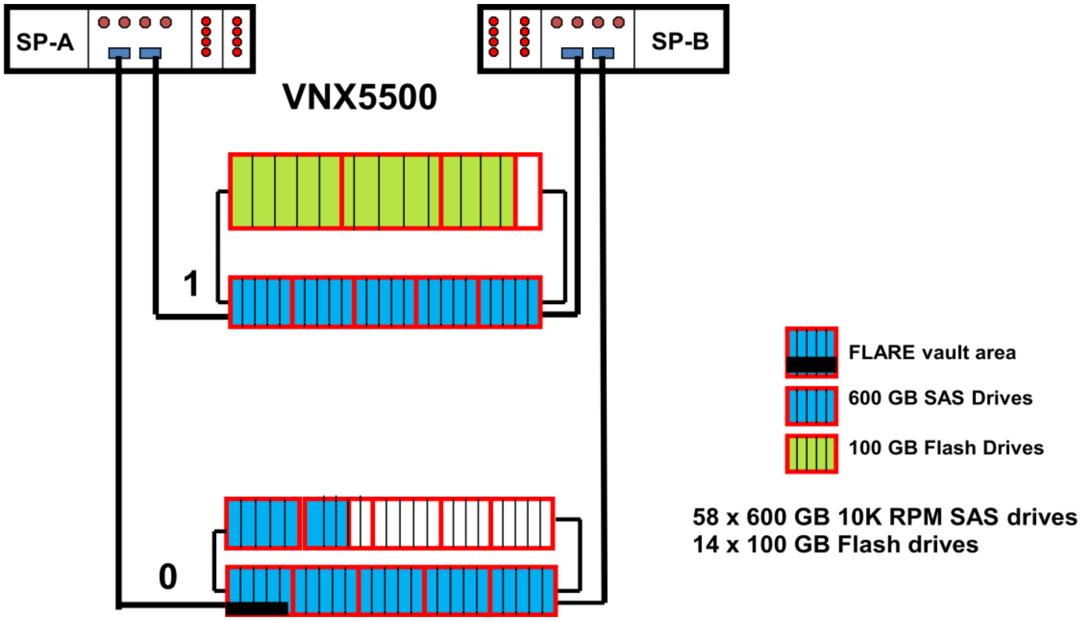
FAST Cache can be selectively enabled on a few or all storage pools within a storage array, depending on application performance requirements and SLAs.

There are several distinctions to EMC FAST Cache:

- It can be configured in read/write mode, which allows the data to be maintained on a faster medium for longer periods, irrespective of application read-to-write mix and data rewrite rate.
- FAST Cache is created on a persistent medium like flash drives, which can be accessed by both storage processors. In the event of a storage processor failure, the surviving storage processor can simply reload the cache rather than repopulating it from scratch by observing the data access patterns again, which is a differentiator.
- Enabling FAST Cache is completely non-disruptive. It is as simple as selecting the Flash drives that are part of FAST Cache and does not require any array disruption or downtime.
- Because FAST Cache is created on external flash drives, adding FAST Cache does not consume any extra PCI-E slots inside the storage processor.

In this lab example, 4x 100GB flash drives were allocated for FAST Cache.

Figure 4. Backend Storage Connectivity



The LUNs were configured as shown in Table 2.

© 2011 VMware, Inc. All rights reserved.

Table 2. Storage Layout Configuration

RAID Group and Type	LUN	Virtual Device	LUN/ VMDK Size	VMFS Datastore	Purpose	Storage Processor Owner
RAID Group 0, RAID 5 4+1	LUN 1	N/A	20GB	N/A	ESXi 5 boot LUN	SP-B
RAID Group 0, RAID 5 4+1	LUN 2	N/A	20GB	N/A	ESXi 5 boot LUN	SP-A
RAID Group 0, RAID 5 4+1	LUN 3	N/A	20GB	N/A	ESXi 5 boot LUN	SP-B
RAID Group 0, RAID 5 4+1	LUN 4	N/A	20GB	N/A	ESXi 5 boot LUN	SP-A
RAID Group 0, RAID 5 4+1	LUN 5	N/A	20GB	N/A	ESXi 5 boot LUN	SP-B
RAID Group 1, RAID 1+0	LUN 6	SCSI 0:0	1TB	VMDATASTORE	Virtual machines and Oracle binary	SP-A
Pool 0, RAID 5	LUN 7	SCSI 1:0	20GB	CRS1	OCR and Voting – Oracle ASM	SP-A
Pool 0, RAID 5	LUN 8	SCSI 1:1	20GB	CRS2	OCR and Voting – Oracle ASM	SP-B
Pool 0, RAID 5	LUN 9	SCSI 1:2	20GB	CRS3	OCR and Voting – Oracle ASM	SP-A
Pool 0, RAID 5	LUN 10	SCSI 1:3	300GB	DATA1	Data disks – Oracle ASM	SP-A
Pool 0, RAID 5	LUN 11	SCSI 1:4	300GB	DATA2	Data disks – Oracle ASM	SP-B
Pool 0, RAID 5	LUN 12	SCSI 1:5	300GB	DATA3	Data disks – Oracle ASM	SP-A
Pool 0, RAID 5	LUN 13	SCSI 1:6	300GB	DATA4	Data disks – Oracle ASM	SP-B
Pool 0, RAID 5	LUN 14	SCSI 1:8	300GB	DATA5	Data disks – Oracle ASM	SP-A
Pool 0, RAID 5	LUN 15	SCSI 1:9	300GB	DATA6	Data disks – Oracle ASM	SP-B

© 2011 VMware, Inc. All rights reserved.

RAID Group and Type	LUN	Virtual Device	LUN/ VMDK Size	VMFS Datastore	Purpose	Storage Processor Owner
Pool 0, RAID 5	LUN 16	SCSI 1:10	300GB	DATA7	Data disks – Oracle ASM	SP-A
Pool 0, RAID 5	LUN 17	SCSI 1:11	300GB	DATA8	Data disks – Oracle ASM	SP-B
Pool 0, RAID 5	LUN 18	SCSI 1:12	300GB	DATA9	Data disks – Oracle ASM	SP-A
Pool 0, RAID 5	LUN 19	SCSI 1:13	300GB	DATA10	Data disks – Oracle ASM	SP-B
RAID Group 1, RAID 1+0	LUN 20	SCSI 2:0	64GB	REDO1	Redo disks – Oracle ASM	SP-A
RAID Group 1, RAID 1+0	LUN 21	SCSI 2:1	64GB	REDO2	Redo disks – Oracle ASM	SP-B
RAID Group 1, RAID 1+0	LUN 22	SCSI 2:2	64GB	REDO3	Redo disks – Oracle ASM	SP-A
RAID Group 1, RAID 1+0	LUN 23	SCSI 2:3	64GB	REDO4	Redo disks – Oracle ASM	SP-B

In this lab example, RAID 1+0 was used for the Oracle redo disks and RAID 5 for the Oracle data disks. Customer deployments can differ in their RAID choices and they should follow the database best practices recommended by their specific storage array vendor.

For each LUN, a storage processor was chosen as the default owner so that the storage array service processors were evenly balanced.

3.3 Multipathing

3.3.1 PowerPath/VE

PowerPath/VE is host-based software that provides multipath capability to implement quality of service for vSphere users by delivering business continuity and availability as well as performance to meet service-level agreements. PowerPath/VE automates data path utilization in dynamic, VMware virtualized environments to provide predictable and consistent information access while delivering investment protection with support for heterogeneous servers, operating systems, and storage.

Increasingly, deployments leverage virtualization for consolidation and to enable scale-out of mission-critical applications. PowerPath/VE manages the complexity of large virtual environments which could contain hundreds or thousands of independent virtual machines running intensive I/O applications. To manually configure this type of scenario, ensuring all the virtual machines get the I/O response time needed is very difficult and time-consuming. If other variables, such as vMotion and the need for HA in the VMware environment are a requirement, any assumption about which I/O streams share which channels is invalidated. PowerPath/VE manages this complexity, adjusting the I/O path usage to changes in I/O loads coming from the virtual machines. Assigning all devices to all paths allows PowerPath/VE to do the work, optimizing the overall I/O performance for the virtual environment.

© 2011 VMware, Inc. All rights reserved.

Key benefits realized when using PowerPath/VE in a vSphere 5 environment include the ability to manage these large environments, drive increased performance by ensuring optimum use of resources, providing for high availability, automating I/O path failover, and enabling recovery in the event of a path failure. Additional information on EMC PowerPath/VE is available at http://www.emc.com/collateral/software/data-sheet/l751-powerpath-ve-multipathing-ds.pdf.

3.4 Workload Test

The workload was based on the SwingBench large scale order entry benchmark which was implemented as follows:

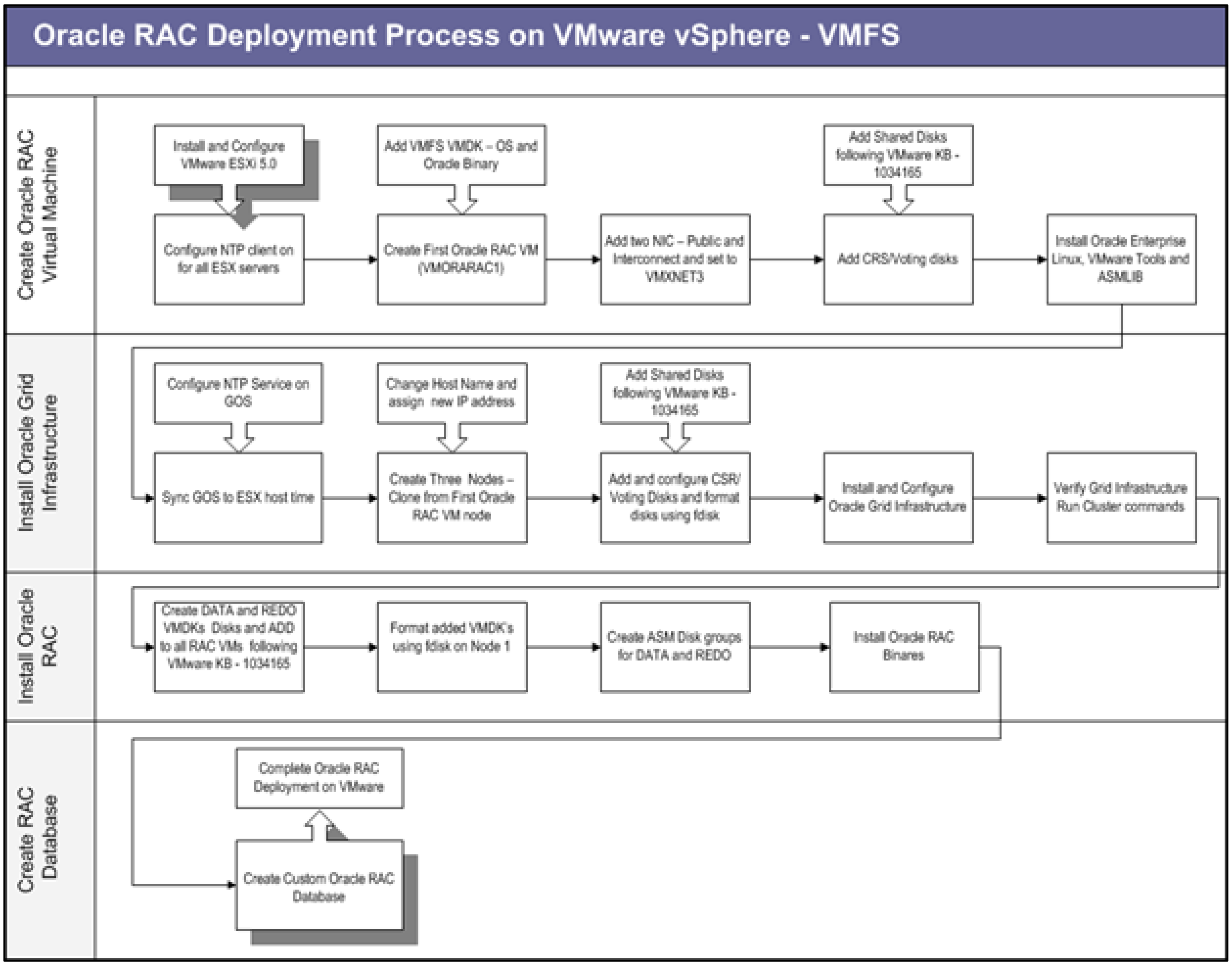
- A single SwingBench load generator was used that was load balanced to spread user load evenly across Oracle RAC nodes.
- Minimum and maximum think times were 4ms and 10ms respectively.
- Performance data was collected using VMware `esxtop` and Oracle AWR (Automatic Workload Repository) reports.
- 1000 users were loaded and run continuously over a 24-hour period.

© 2011 VMware, Inc. All rights reserved.

4. Oracle RAC Installation Overview

Figure 5 shows a flow diagram of the high level steps to install Oracle RAC on the virtualized platform.

Figure 5. Oracle RAC Installation on VMware VMFS


Oracle RAC Deployment Process on VMware vSphere - VMFS
Create Oracle RAC Virtual Machine
Install and Configure VMware ESXi 5.0
Configure NTP client on for all ESX servers
Add VMFS VMDK – OS and Oracle Binary
Create First Oracle RAC VM (VMORARAC1)
Add two NIC – Public and Interconnect and set to VMXNET3
Add Shared Disks following VMware KB - 1034165
Add CRS/Voting disks
Install Oracle Enterprise Linux, VMware Tools and ASMLIB
Install Oracle Grid Infrastructure
Configure NTP Service on GOS
Sync GOS to ESX host time
Change Host Name and assign new IP address
Create Three Nodes – Clone from First Oracle RAC VM node
Add Shared Disks following VMware KB - 1034165
Add and configure CSR/ Voting Disks and format disks using fdisk
Install and Configure Oracle Grid Infrastructure
Verify Grid Infrastructure Run Cluster commands
Install Oracle RAC
Create DATA and REDO VMDKs Disks and ADD to all RAC VMs following VMware KB - 1034165
Format added VMDK's using fdisk on Node 1
Create ASM Disk groups for DATA and REDO
Install Oracle RAC Binares
Create RAC Database
Complete Oracle RAC Deployment on VMware
Create Custom Oracle RAC Database

© 2011 VMware, Inc. All rights reserved.

5. 24-Hour Workload Test

This section describes the results of the 24-hour workload test. Performance results were obtained from the following:

- Performance data captured by the VMware esxtop utility, viewed in Windows Performance Monitor.
- Output data from the Oracle SwingBench generator.
- Oracle AWR report based on snapshots taken at the beginning and end of the run.

See Appendix A for detailed results. Table 3 summarizes some of the key findings.

Table 3. 24-Hour Workload Test Results

Result	Value
Total runtime.	24 hours.
SwingBench average transactions per second.	12,787.
SwingBench maximum transactions rate.	804,191.
SwingBench average response time.	35ms.
CPU utilization of each Oracle RAC node (during last 12 hours).	90–93% approximately (source: vCenter performance charts).
ESXi host core utilization (during last 12 hours).	88–92 % approximately.
Total IOPS (all four Oracle RAC nodes).	32,000–36,000 approximately.

The results show the following:

- Workload stabilized in the last 12–15 hours of the run such that all the performance counters leveled out.
- The workload was evenly distributed across the four Oracle RAC nodes. Each node showed similar CPU utilization (90%+) and generated similar IOPS (0800-9000 approximately).
- The redo LUN latency was less than 10 milliseconds, after the workload leveled out. Values below the 10-millisecond threshold indicate good I/O performance (see *Using esxtop to identify storage performance issues* at http://kb.vmware.com/kb/1008205). The 10-millisecond threshold is a guideline and it is possible for applications to experience satisfactory response times with higher disk latencies.
- The approximate interconnect traffic measured by the Oracle AWR report was 242MB/sec. The network counters from esxtop reflected similar numbers.
- The AWR top five wait events for the 24-hour period demonstrates that latency due to the Oracle RAC interconnect was less than 5%.
- Overall throughput of SwingBench transactions did not degrade and there were zero transaction errors.

© 2011 VMware, Inc. All rights reserved.

6. Oracle RAC Node vMotion Test

This section describes the results of an Oracle RAC node vMotion test. It demonstrates the elimination of application downtime due to planned server maintenance. The use case assumes that live migration is initiated at a time when the workload is not at peak so that one ESX/ESXi host can temporarily host two Oracle RAC nodes (there is vCPU overcommitment where the total vCPUs exceeds the number of cores on a host). This consolidation is possible without impact to transaction performance as long as the CPU utilization of two Oracle RAC nodes can be sustained on one ESX/ESXi host.

All the results were obtained from the following:

- Performance data captured by the VMware `esxtop` utility.
- Output data from the Oracle SwingBench generator.
- Oracle cluster logs and database alert logs.

See Appendix B for detailed results. Table 4 shows the configuration of four Oracle RAC nodes.

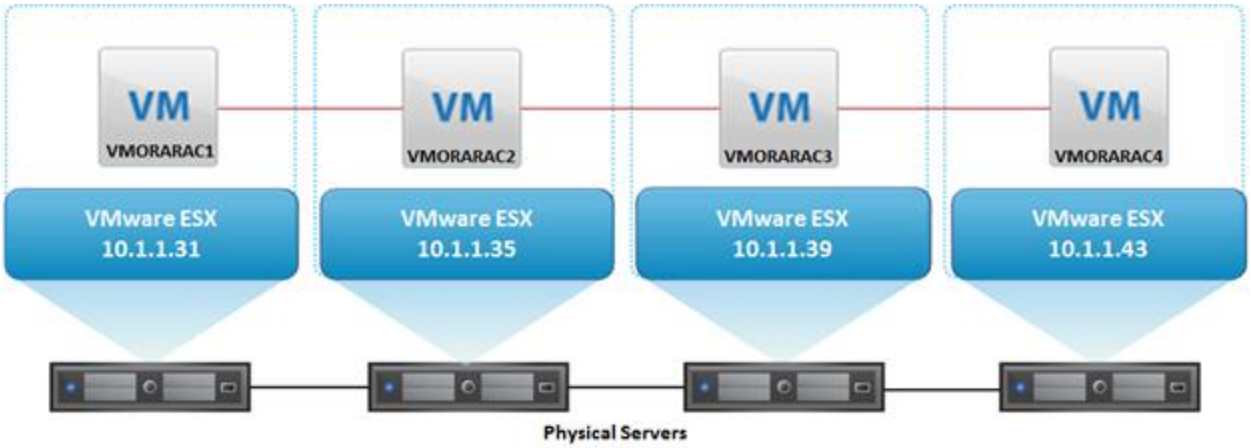
Table 4. Oracle RAC Node vMotion Test Configuration

ESX Host	RAC Node VM	# of vCPUs	Memory (GB)	Oracle SGA (GB)
ESX1 – 12 Core, 192GB RAM	VMORARAC1	12	64	50
ESX2 – 12 Core, 192GB RAM	VMORARAC2	12	64	50
ESX3 – 12 Core, 192GB RAM	VMORARAC3	12	64	50
ESX4 – 12 Core, 192GB RAM	VMORARAC4	12	64	50

The following is the procedure for testing Oracle RAC node vMotion migration.

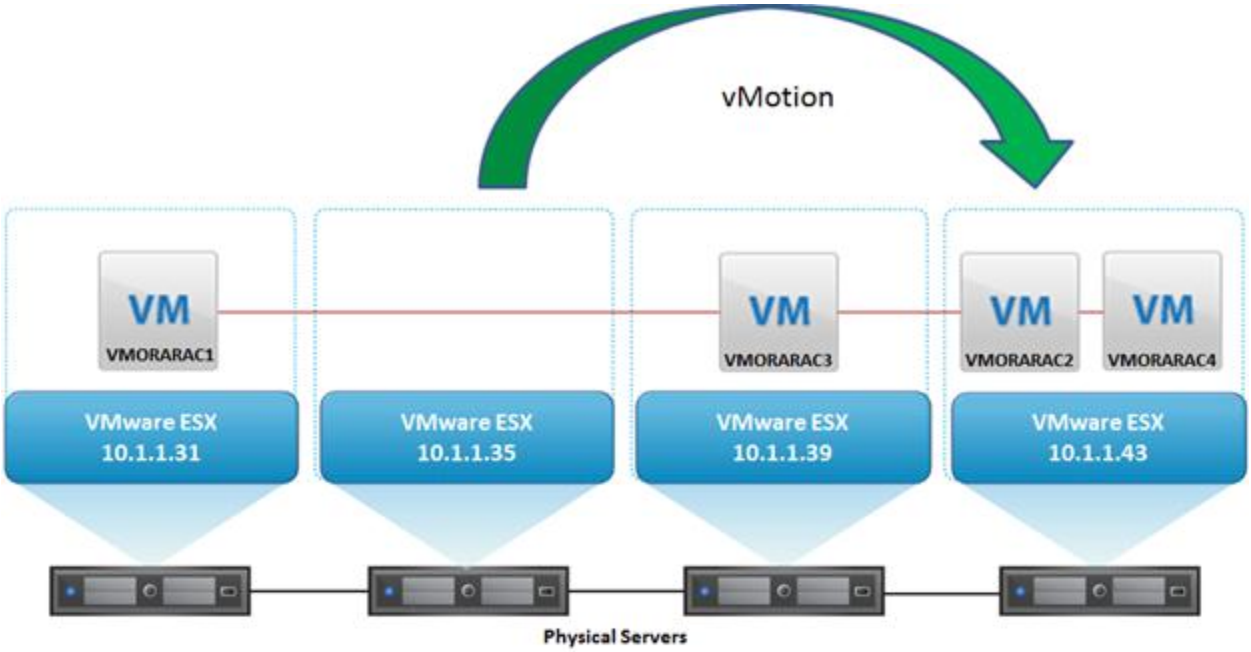
To test Oracle RAC node vMotion migration

1. All four nodes are running with their respective ESX hosts. The SwingBench large scale order entry benchmark was run against all four nodes with 400 users. ESX2 must be taken down for hardware maintenance (such as a firmware upgrade).

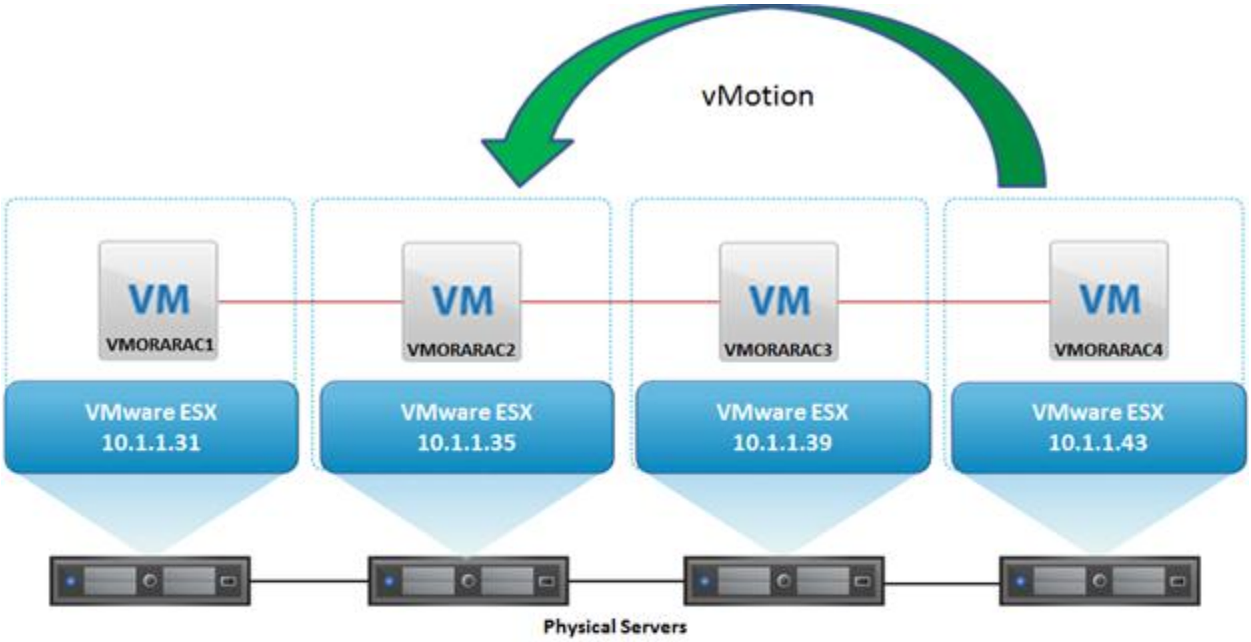


© 2011 VMware, Inc. All rights reserved.

2. Migrate the Oracle RAC node VMORARAC2 from ESX2 (10.1.1.35) to ESX4 (10.1.1.43) so that ESX2 can be taken down for a firmware upgrade.



3. After the hardware maintenance is completed on ESX2, migrate VMORARAC2 from ESX4 (10.1.1.43) back to ESX2 (10.1.1.35).



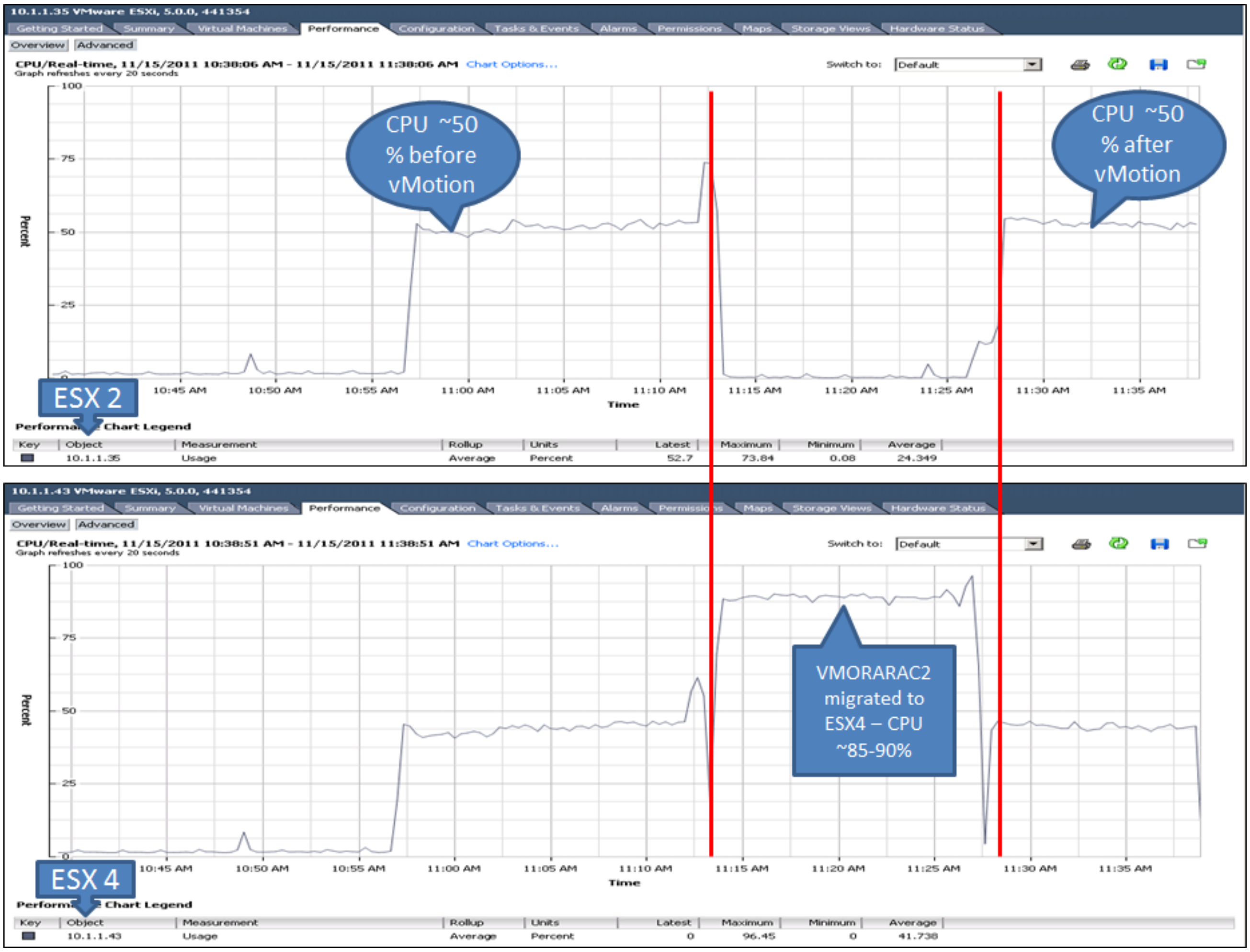
© 2011 VMware, Inc. All rights reserved.

The results show the following:

- CPU utilization on all four ESX hosts was approximately 45–51%.
- After migration of VMORARAC2 (RAC node 2) from ESX2 to ESX4, ESX4 had a CPU utilization of 85–91%. The CPU of this host was not saturated which shows that it was able to handle the loads of both RAC nodes.
- Transactions per minute (TPM) are similar before and after vMotion migration.
- Oracle database logs and cluster logs did not have errors during vMotion migration.

The following chart shows the CPU utilization of both the ESX2 (10.1.1.35) and ESX4 (10.1.1.43) hosts during vMotion test. CPU utilization for ESX4 is approximately 85–91%, which is the sum of the two Oracle RAC nodes (VMORARAC2, approximately 50% and VMORARAC4, approximately 45%).

Figure 6. CPU Utilization during vMotion Test – vCenter Chart



CPU/Real-time, 11/15/2011 10:38:06 AM - 11/15/2011 11:38:06 AM
Graph refreshes every 20 seconds

Performance Chart Legend

Key	Object	Measurement	Rollup	Units	Latest	Maximum	Minimum	Average
■	10.1.1.35	Usage	Average	Percent	52.7	73.84	0.08	24.349

CPU/Real-time, 11/15/2011 10:38:51 AM - 11/15/2011 11:38:51 AM
Graph refreshes every 20 seconds

Performance Chart Legend

Key	Object	Measurement	Rollup	Units	Latest	Maximum	Minimum	Average
■	10.1.1.43	Usage	Average	Percent	0	96.45	0	41.738

7. Best Practices

VMware recommends the following best practices and guidelines when deploying Oracle RAC on vSphere:

- Use the multi-writer option to allow VMFS-backed disks to be shared by multiple virtual machines. Follow *Disabling simultaneous write protection provided by VMFS using the multi-writer flag* (http://kb.vmware.com/kb/1034165).
- Installation of Oracle RAC on VMware is the same as a physical installation, so leverage existing Oracle RAC skills and expertise.
- Oracle NUMA support is disabled by default for 11g and above (see Oracle MySupport Doc ID: 864633.1). In some circumstances enabling Oracle NUMA support may improve performance and the Oracle doc suggests that it be tested in a test environment before deciding to use it with production system. VMware recommends keeping NUMA enabled in server hardware BIOS and at the guest operating system level which should also be the default settings for NUMA support with most servers and guest operating systems.
- For business-critical applications on Oracle RAC, set memory reservations to the size of the virtual machine. Correctly size the virtual machine to avoid wasting memory.
- Use the latest servers with chips that have second-generation hardware-assisted virtualization features such as AMD's Rapid Virtualization Indexing (RVI) and Intel's Extended Page Tables (EPT).
- For the RAC interconnect:
  - Use jumbo frames – To enable jumbo frames, follow *iSCSI and Jumbo Frames configuration on ESX 3.x and ESX 4.x* (http://kb.vmware.com/kb/1007654).
  - Separate the interconnect network to isolate it from other traffic.
- For VMware HA follow *vSphere High Availability Deployment Best Practices* (http://www.vmware.com/files/pdf/techpaper/vmw-vsphere-high-availability.pdf).
- Follow setting up information and best practices in *vSphere Networking* (http://pubs.vmware.com/vsphere-50/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-50-networking-guide.pdf). If available, use converged network and storage I/O onto 10Gb to reduce cabling requirements.
- Follow EMC's best practice guidelines for the storage layout of the Oracle database and SAN array configuration:
  - Guidelines used in the physical world apply equally when virtualized. An example - use of different RAID types for Oracle data and redo LUNs.
  - Balance workloads across storage processors.
  - Using flash drives with EMC's FAST Cache and Fully Automated Storage Tiering (FAST VP) software can enhance I/O performance in a set it and forget it model. Follow the best practice guidelines available at http://www.emc.com.
- Use VMFS for Oracle clustering, data, redo, and root drive, and leverage VMware templates and cloning to quickly provision new Oracle RAC nodes.
- Use VMware monitoring tools to measure and monitor performance in the virtual environment. Follow guidelines in *vSphere Troubleshooting* (http://pubs.vmware.com/vsphere-50/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-50-troubleshooting-guide.pdf).
  - Use vCenter performance charts to monitor performance in the virtual environment.
  - For more advanced troubleshooting, work with a VMware administrator to use the `esxtop` utility.

© 2011 VMware, Inc. All rights reserved.

- AWR can be used in the same way as in physical environments.
- Use Oracle ASM as per Oracle best practices—no changes are required for the virtual environment.
- Use huge/large pages per Oracle My Support Document ID: 361468.1. HugePages is a feature of the Linux kernel which allows larger pages to manage memory as the alternative to the small 4KB page size.
- To avoid time synchronization errors when using a Linux guest operating system, follow *Timekeeping best practices for Linux guests* (http://kb.vmware.com/kb/1006427).
- Use jumbo frames for vMotion migrations. To enable jumbo frames, follow *Enabling IOAT and Jumbo frames* (http://kb.vmware.com/kb/1003712). See also *VMware vSphere vMotion Architecture, Performance and Best Practices in VMware vSphere 5* (http://www.vmware.com/files/pdf/vmotion-perf-vsphere5.pdf).
- VMware recommends using multiple virtual SCSI controllers for database virtual machines. The use of multiple virtual SCSI controllers allows the execution of several parallel I/O operations inside the guest operating system. It is recommended to use one controller for OS and swap, one controller for redo, and one or more controllers for data files.

© 2011 VMware, Inc. All rights reserved.

8. Summary

This paper described a workload characterization conducted against a 4-node Oracle RAC deployment on VMware vSphere 5. The architecture was tested by subjecting it to a sustained heavy OLTP workload over a 24-hour period and results showed substantial transaction throughput without any degradation in performance. This performance can be attributed to the overall system design and architecture components which included the following:

- No overcommitment of memory resources. Spare capacity was available on the ESXi hosts such that more virtual machines could be hosted.
- The latest high performing hypervisor in vSphere 5.0 with the enhanced scheduler running on high performing UCS server blades with chips that have second-generation hardware-assisted virtualization features.
- Cisco UCS Fabric Interconnect with 10Gbps unified fabric connectivity.
- Use of jumbo frames for the Oracle RAC cluster interconnects and for vMotion.
- EMC VNX5500 Unified Storage Array configured with mixed drives—(SAS drives and flash drives) and EMC FAST Suite array-based software.
- Following SAN array best practices for database storage layout and array configuration (such as multipathing and evenly balanced service processors).

By following best practices, customers can virtualize their Oracle RAC deployments and run business-critical applications with satisfactory performance. Because disk I/O is a major factor in Oracle database performance, follow guidelines from the storage array vendor. Many of these best practices are similar to physical deployments.

© 2011 VMware, Inc. All rights reserved.

9. References

The following are resources and references for Oracle and VMware vSphere.

Performance Best Practices for VMware vSphere 5.0
http://www.vmware.com/pdf/Perf_Best_Practices_vSphere5.0.pdf

vSphere Storage
http://pubs.vmware.com/vsphere-50/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-50-storage-guide.pdf

Oracle Database Installation and Configuration Guides
http://www.oracle.use one's com/pls/db111/portal.portal_db?selected=11

Interpreting esxtop Statistics
http://communities.vmware.com/docs/DOC-9279

vSphere High Availability Deployment Best Practices
http://www.vmware.com/files/pdf/techpaper/vmw-vsphere-high-availability.pdf

vSphere Networking
http://pubs.vmware.com/vsphere-50/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-50-networking-guide.pdf

Timekeeping best practices for Linux guests
http://kb.vmware.com/kb/1006427

Using esxtop to identify storage performance issues
http://kb.vmware.com/kb/1008205

Disabling simultaneous write protection provided by VMFS using the multi-writer flag
http://kb.vmware.com/kb/1034165

Using EMC VNX Storage with VMware vSphere
http://www.emc.com/collateral/hardware/technical-documentation/h8229-vnx-vmware-tb.pdf

EMC PowerPath/VE for VMware vSphere: Best Practices Planning
http://www.emc.com/collateral/software/white-papers/h6340-powerpath-ve-for-vmware-vsphere-wp.pdf

© 2011 VMware, Inc. All rights reserved.

10. Disclaimers

All data is based on in-lab results with vSphere 5.0. Our workload was a fair-use implementation of the TPC-C business model. These results are not TPC-C compliant and are not comparable to official TPC-E results. TPC Benchmark and TPC-C are trademarks of the Transaction Processing Performance Council.

The throughput here is not meant to indicate the absolute performance of Oracle Database 11g R2, or to compare its performance to another database management system. Oracle Database was used to place a DBMS workload on VMware ESX and observe and optimize the performance of ESX.

The goal of the experiment was to show virtualization of Oracle RAC on VMware, and its ability to handle heavy database workloads. It was not meant to measure the absolute performance of the hardware and software components used in the study.

The throughput of the workload used does not constitute a TPC benchmark result.

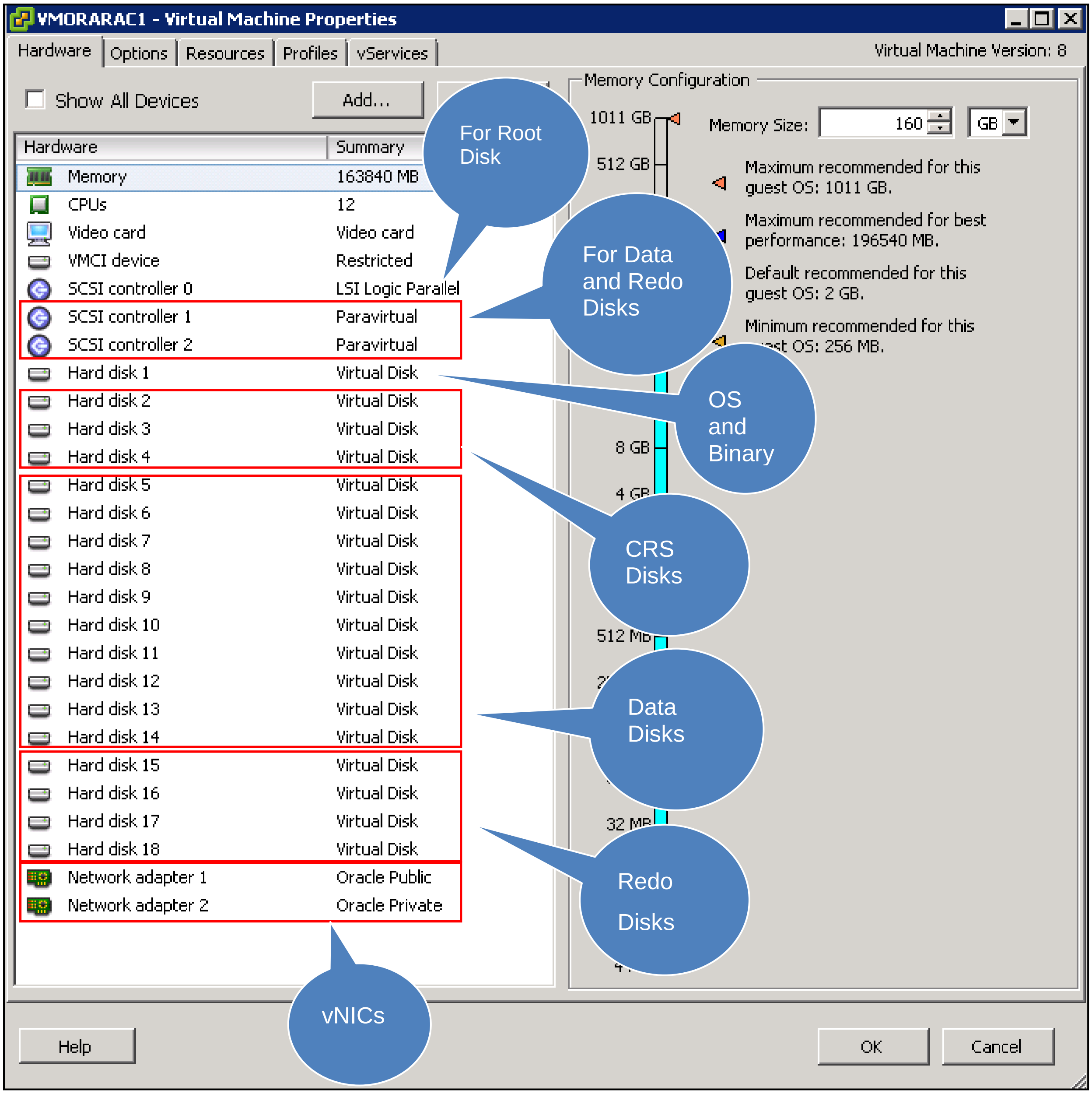
Appendix A: 24-Hour Workload Test Results

This appendix outlines some of the results from the 24-hour workload test. It includes charts obtained from the VMware `esxtop` performance monitoring utility. Background on `esxtop` performance counters is available in *Interpreting esxtop Statistics* (http://communities.vmware.com/docs/DOC-9279). Data from `esxtop` was loaded into and viewed in Windows Performance System Monitor (Perfmon).

Oracle RAC Node VM Configuration

The following shows the configuration and properties of the first RAC Node VM – VMORARAC1.

Figure 7. VMORARAC1 – VM Configuration



Virtual Machine Version: 8

☐ Show All Devices

Hardware	Summary
Memory	163840 MB
CPUs	12
Video card	Video card
VMCI device	Restricted
SCSI controller 0	LSI Logic Parallel
SCSI controller 1	Paravirtual
SCSI controller 2	Paravirtual
Hard disk 1	Virtual Disk
Hard disk 2	Virtual Disk
Hard disk 3	Virtual Disk
Hard disk 4	Virtual Disk
Hard disk 5	Virtual Disk
Hard disk 6	Virtual Disk
Hard disk 7	Virtual Disk
Hard disk 8	Virtual Disk
Hard disk 9	Virtual Disk
Hard disk 10	Virtual Disk
Hard disk 11	Virtual Disk
Hard disk 12	Virtual Disk
Hard disk 13	Virtual Disk
Hard disk 14	Virtual Disk
Hard disk 15	Virtual Disk
Hard disk 16	Virtual Disk
Hard disk 17	Virtual Disk
Hard disk 18	Virtual Disk
Network adapter 1	Oracle Public
Network adapter 2	Oracle Private

Memory Configuration

Memory Size: 160 GB

Maximum recommended for this guest OS: 1011 GB.

Maximum recommended for best performance: 196540 MB.

Default recommended for this guest OS: 2 GB.

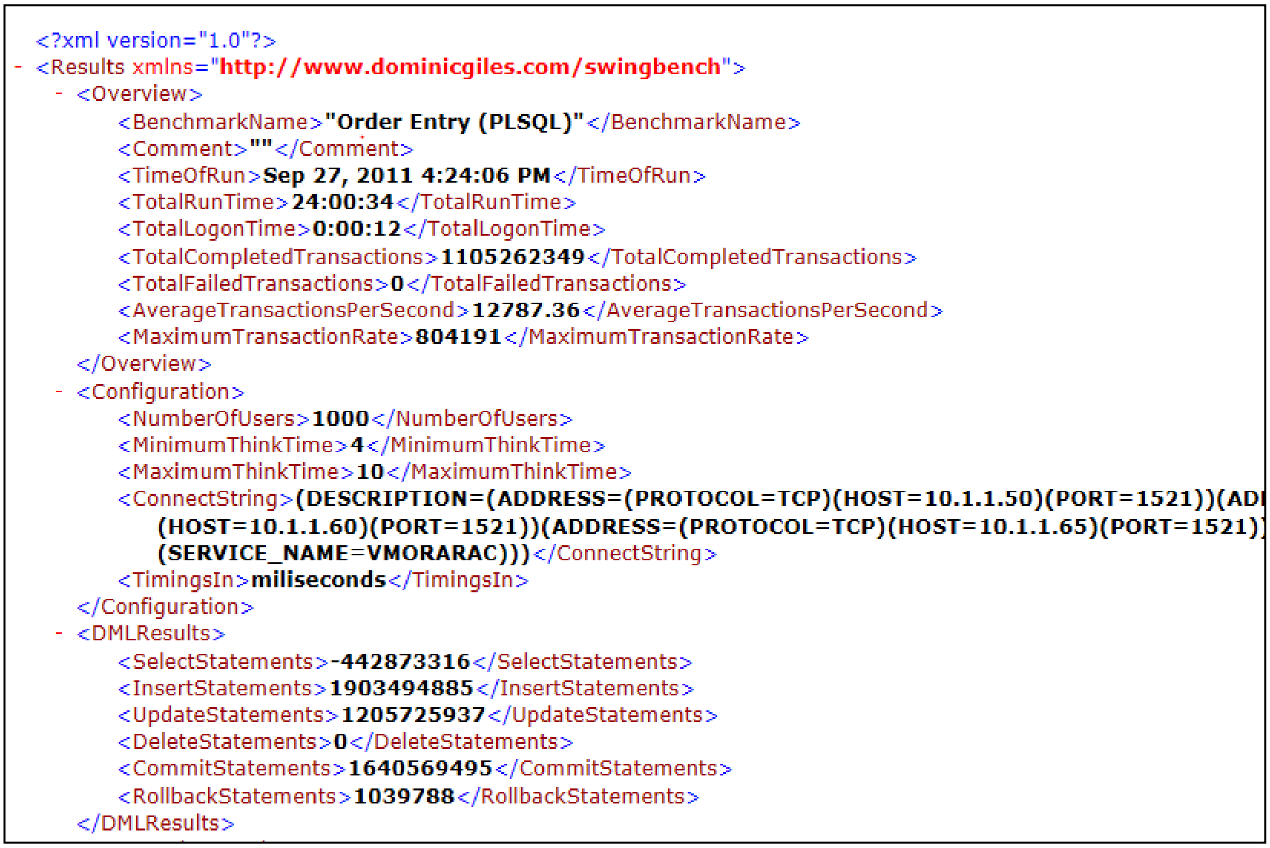
Minimum recommended for this guest OS: 256 MB.

© 2011 VMware, Inc. All rights reserved.

SwingBench Results

The following is the beginning of the output report generated by SwingBench after the 24-hour run.

Figure 8. Head of SwingBench Output Report



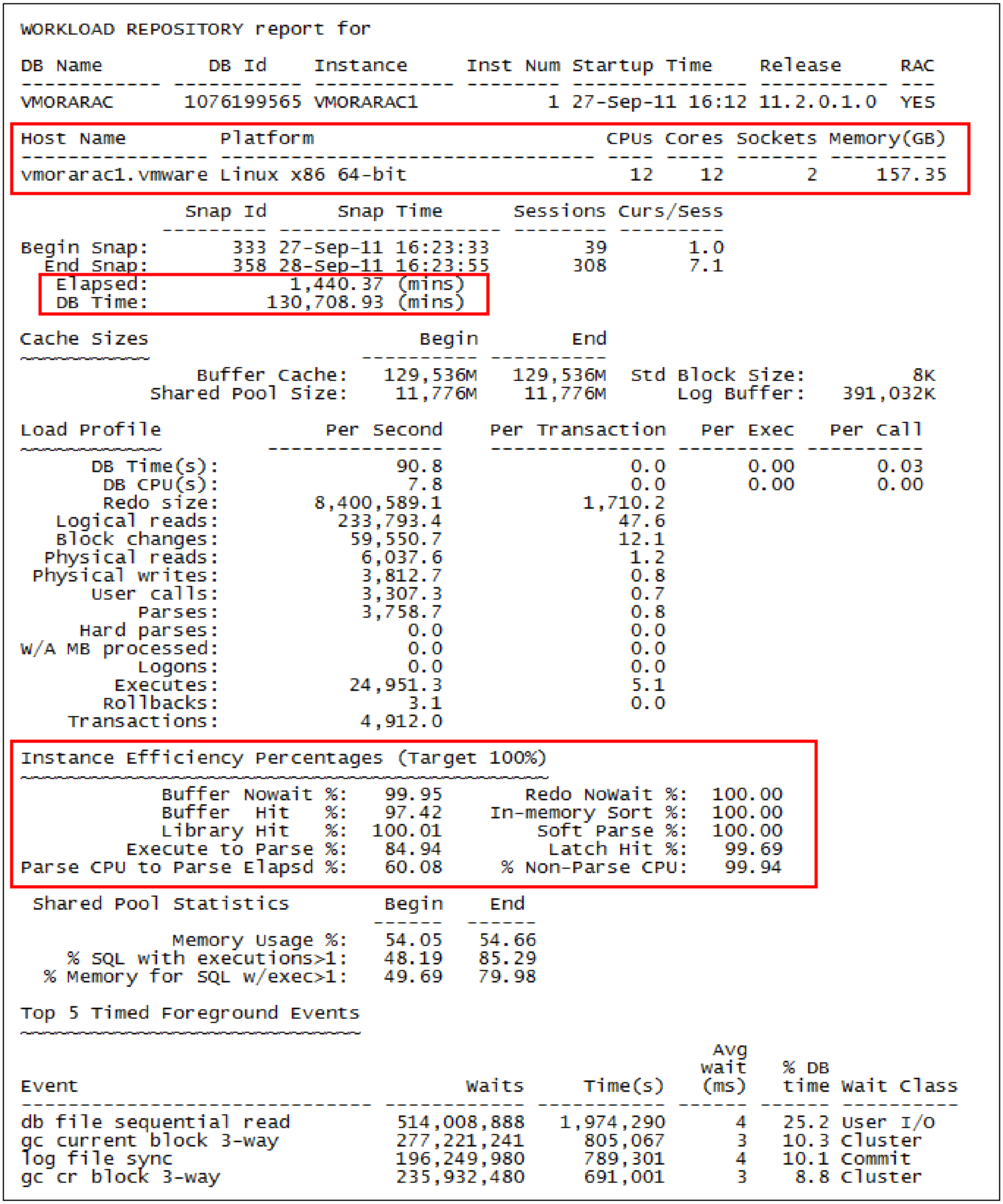
```
  <?xml version="1.0"?>
- <Results xmlns="http://www.dominicgiles.com/swingbench">
  - <Overview>
        <BenchmarkName>"Order Entry (PLSQL)"</BenchmarkName>
        <Comment>""</Comment>
        <TimeOfRun>Sep 27, 2011 4:24:06 PM</TimeOfRun>
        <TotalRunTime>24:00:34</TotalRunTime>
        <TotalLogonTime>0:00:12</TotalLogonTime>
        <TotalCompletedTransactions>1105262349</TotalCompletedTransactions>
        <TotalFailedTransactions>0</TotalFailedTransactions>
        <AverageTransactionsPerSecond>12787.36</AverageTransactionsPerSecond>
        <MaximumTransactionRate>804191</MaximumTransactionRate>
    </Overview>
  - <Configuration>
        <NumberOfUsers>1000</NumberOfUsers>
        <MinimumThinkTime>4</MinimumThinkTime>
        <MaximumThinkTime>10</MaximumThinkTime>
        <ConnectString>(DESCRIPTION=(ADDRESS=(PROTOCOL=TCP)(HOST=10.1.1.50)(PORT=1521))(ADI
            (HOST=10.1.1.60)(PORT=1521))(ADDRESS=(PROTOCOL=TCP)(HOST=10.1.1.65)(PORT=1521))
            (SERVICE_NAME=VMORARAC)))</ConnectString>
        <TimingsIn>miliseconds</TimingsIn>
    </Configuration>
  - <DMLResults>
        <SelectStatements>-442873316</SelectStatements>
        <InsertStatements>1903494885</InsertStatements>
        <UpdateStatements>1205725937</UpdateStatements>
        <DeleteStatements>0</DeleteStatements>
        <CommitStatements>1640569495</CommitStatements>
        <RollbackStatements>1039788</RollbackStatements>
    </DMLResults>
```

© 2011 VMware, Inc. All rights reserved.

AWR Report for Oracle RAC Node 1

The following shows the beginning of the AWR report taken on one Oracle RAC node and is based on snapshots taken at the start and end of the 24-hour run.

Figure 9. AWR Report for Oracle RAC Node 1



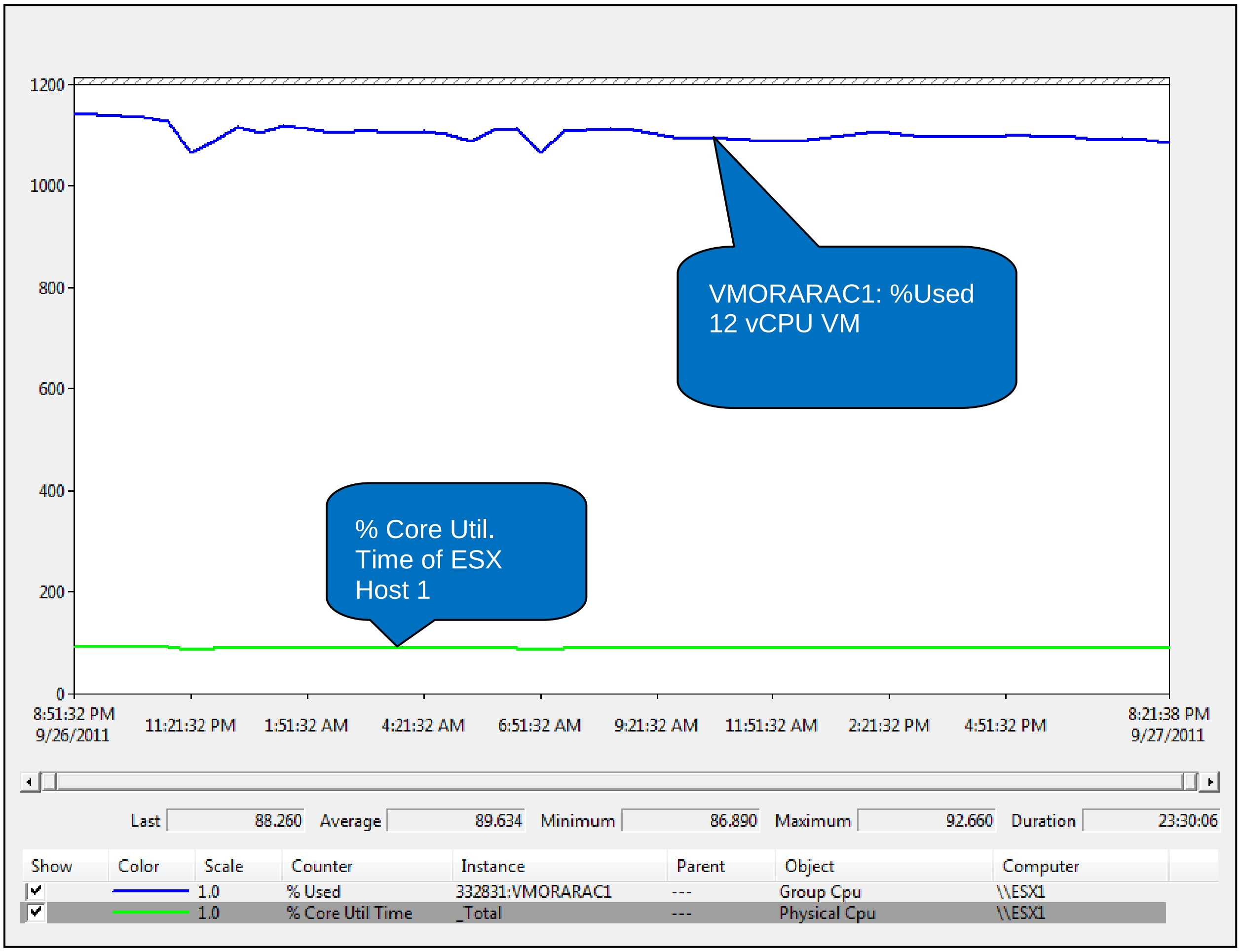
```
WORKLOAD REPOSITORY report for

DB Name         DB Id    Instance     Inst Num Startup Time    Release     RAC
------------ ----------- ------------ -------- --------------- ----------- ---
VMORARAC      1076199565 VMORARAC1           1 27-Sep-11 16:12 11.2.0.1.0  YES

Host Name        Platform                         CPUs Cores Sockets Memory(GB)
---------------- -------------------------------- ---- ----- ------- ----------
vmorarac1.vmware Linux x86 64-bit                   12    12       2     157.35

              Snap Id      Snap Time      Sessions Curs/Sess
            --------- ------------------- -------- ---------
Begin Snap:       333 27-Sep-11 16:23:33        39       1.0
  End Snap:       358 28-Sep-11 16:23:55       308       7.1
   Elapsed:             1,440.37 (mins)
   DB Time:           130,708.93 (mins)

Cache Sizes                       Begin        End
~~~~~~~~~~~                  ---------- ----------
               Buffer Cache:   129,536M   129,536M  Std Block Size:         8K
           Shared Pool Size:    11,776M    11,776M      Log Buffer:   391,032K

Load Profile              Per Second    Per Transaction   Per Exec   Per Call
~~~~~~~~~~~~         ---------------    --------------- ---------- ----------
      DB Time(s):               90.8                0.0       0.00       0.03
       DB CPU(s):                7.8                0.0       0.00       0.00
       Redo size:        8,400,589.1            1,710.2
   Logical reads:          233,793.4               47.6
   Block changes:           59,550.7               12.1
  Physical reads:            6,037.6                1.2
 Physical writes:            3,812.7                0.8
      User calls:            3,307.3                0.7
          Parses:            3,758.7                0.8
     Hard parses:                0.0                0.0
W/A MB processed:                0.0                0.0
          Logons:                0.0                0.0
        Executes:           24,951.3                5.1
       Rollbacks:                3.1                0.0
    Transactions:            4,912.0

Instance Efficiency Percentages (Target 100%)
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
            Buffer Nowait %:   99.95       Redo NoWait %:  100.00
            Buffer  Hit   %:   97.42    In-memory Sort %:  100.00
            Library Hit   %:  100.01        Soft Parse %:  100.00
         Execute to Parse %:   84.94         Latch Hit %:   99.69
Parse CPU to Parse Elapsd %:   60.08     % Non-Parse CPU:   99.94

 Shared Pool Statistics        Begin    End
                              ------  ------
             Memory Usage %:   54.05   54.66
    % SQL with executions>1:   48.19   85.29
  % Memory for SQL w/exec>1:   49.69   79.98

Top 5 Timed Foreground Events
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                           Avg
                                                          wait   % DB
Event                                 Waits     Time(s)   (ms)   time Wait Class
------------------------------ ------------ ----------- ------ ------ ----------
db file sequential read         514,008,888   1,974,290      4   25.2 User I/O
gc current block 3-way          277,221,241     805,067      3   10.3 Cluster
log file sync                   196,249,980     789,301      4   10.1 Commit
gc cr block 3-way               235,932,480     691,001      3    8.8 Cluster
```

© 2011 VMware, Inc. All rights reserved.

CPU Utilization Oracle RAC Node 1

The following chart shows the CPU utilization of the virtual machine RAC node 1 (counter = %Used) and of the ESX host (counter = % Core Util Time). The other Oracle RAC nodes and ESX hosts exhibited similar performance.

Figure 10. CPU Utilization Oracle RAC Node 1



Last 88.260 Average 89.634 Minimum 86.890 Maximum 92.660 Duration 23:30:06

Show	Color	Scale	Counter	Instance	Parent	Object	Computer
☑	🟦	1.0	% Used	332831:VMORARAC1	---	Group Cpu	\\ESX1
☑	🟩	1.0	% Core Util Time	_Total	---	Physical Cpu	\\ESX1

% Core utilization (which is displayed only when hyper-threading is used) is the percentage of CPU cycles averaged over all the cores when at least one of the threads in the core is not halted (that is, it is executing instructions).

The `esxtop` tool uses worlds and groups as the entities to show CPU usage. A *world* is a schedulable entity on the VMware hypervisor, similar to a process or thread in other operating systems. A *group* contains multiple worlds—a virtual machine is classified as a group. The %USED counter shows values between 100% and 1200%. The maximum %USED is the number of worlds in the group multiplied by 100%. Typically, worlds other than vCPU worlds are waiting for events most of time, which does not cost too many CPU cycles. Among all the worlds, vCPU worlds represent the activity of the guest operating system. Therefore, %USED for a virtual machine group usually does not exceed the number of vCPUs assigned to the virtual machine, multiplied by 100%.

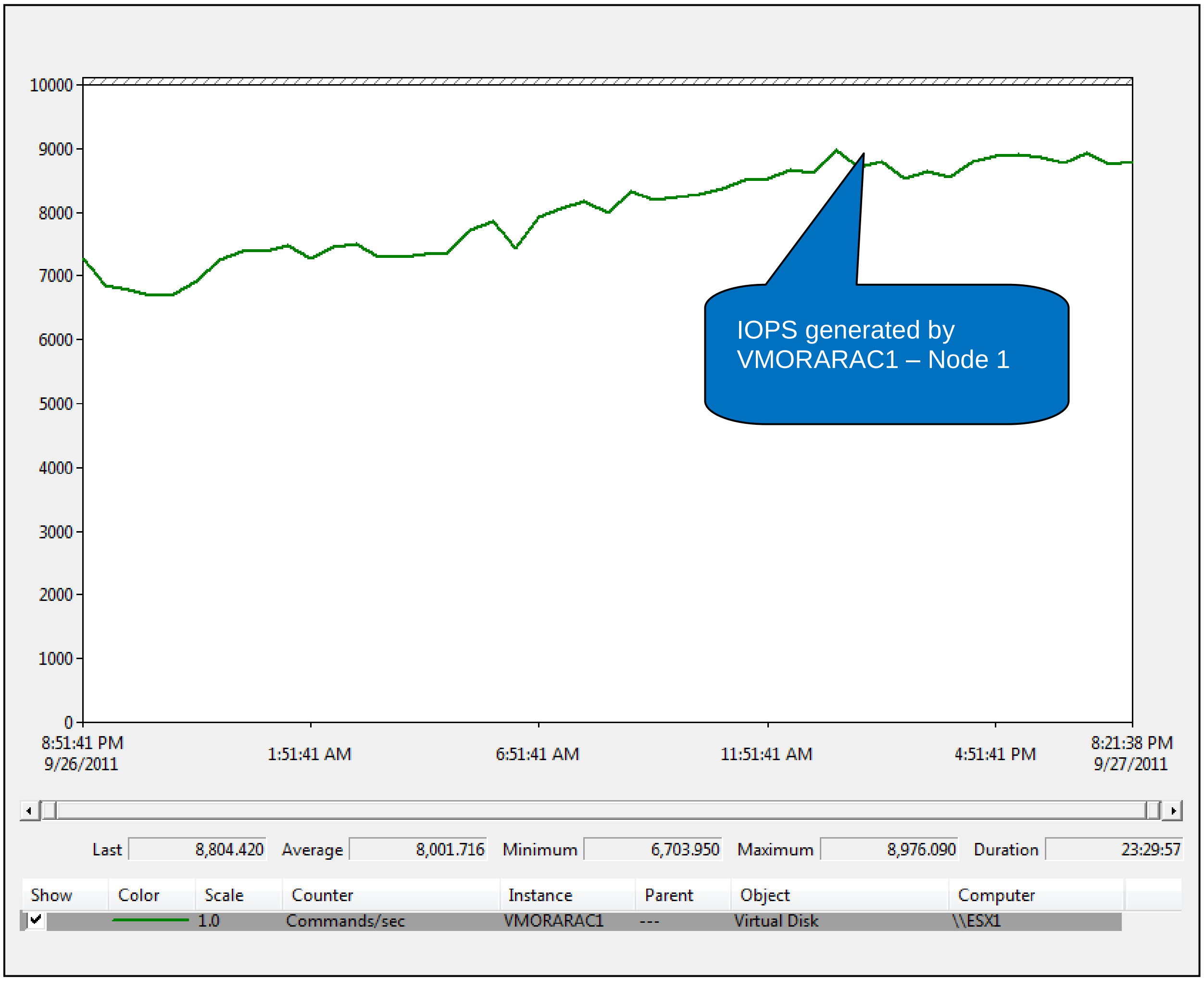
vCenter performance charts monitors the virtual machine's CPU utilization (counter "usage") and reports it between zero and 100%. In this scenario, it was measured at approximately 90–93%.

© 2011 VMware, Inc. All rights reserved.

IOPS – For Each Oracle RAC Node

The following charts show the ESX performance counter "commands/sec" over the 24-hour run, generated by each of the four virtual machine Oracle RAC nodes. This is equivalent to the number of IOPS (input/output operations per second) being sent to or coming from the virtual machine. Refer to *Using esxtop to identify storage performance issues* (http://kb.vmware.com/kb/1008205).

Figure 11. IOPS for Oracle RAC – Node 1 (VMORARAC1)

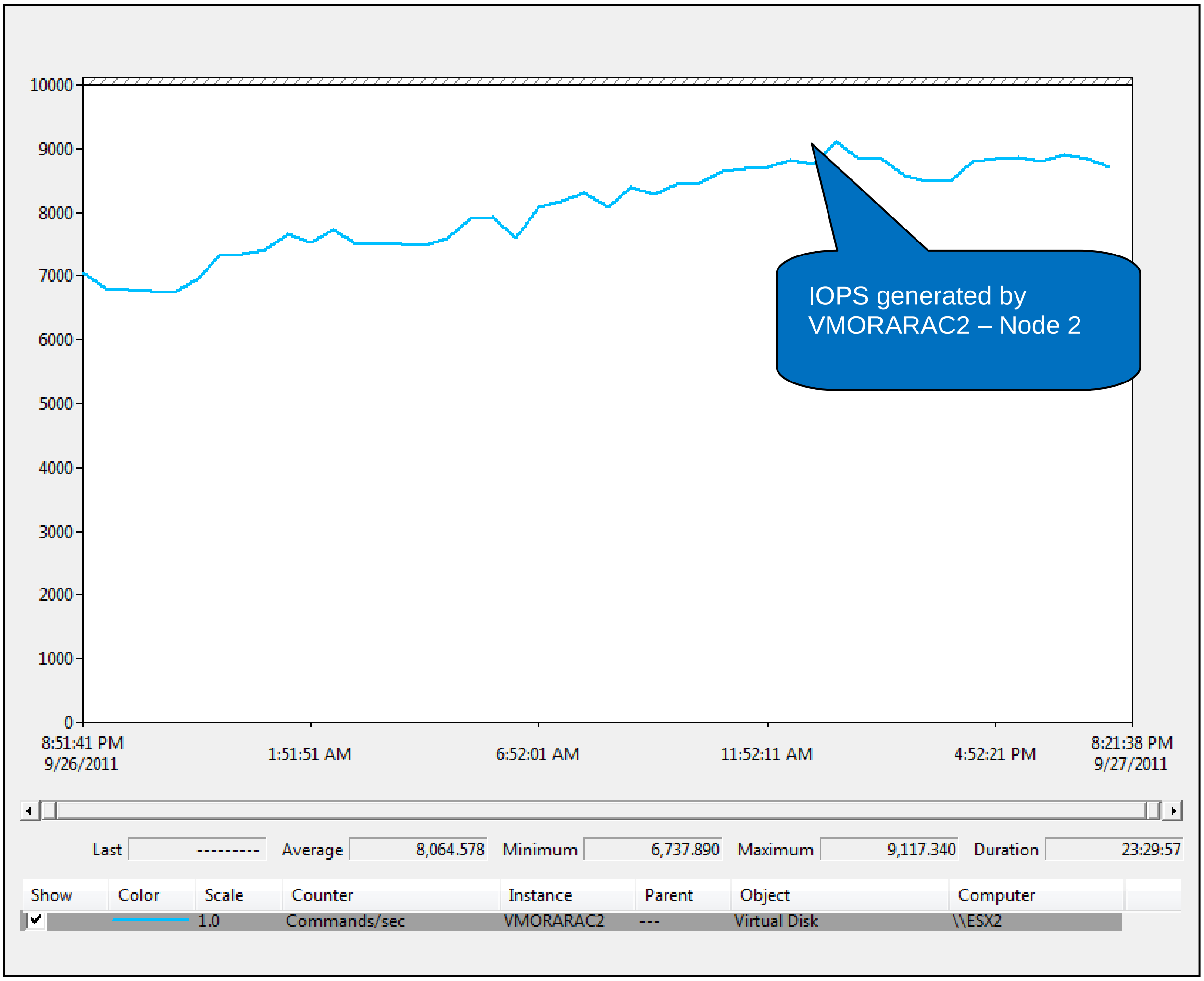


Last 8,804.420 Average 8,001.716 Minimum 6,703.950 Maximum 8,976.090 Duration 23:29:57

Show	Color	Scale	Counter	Instance	Parent	Object	Computer
✓	—	1.0	Commands/sec	VMORARAC1	---	Virtual Disk	\\ESX1

© 2011 VMware, Inc. All rights reserved.

Figure 12. IOPS for Oracle RAC – Node 2 (VMORARAC2)



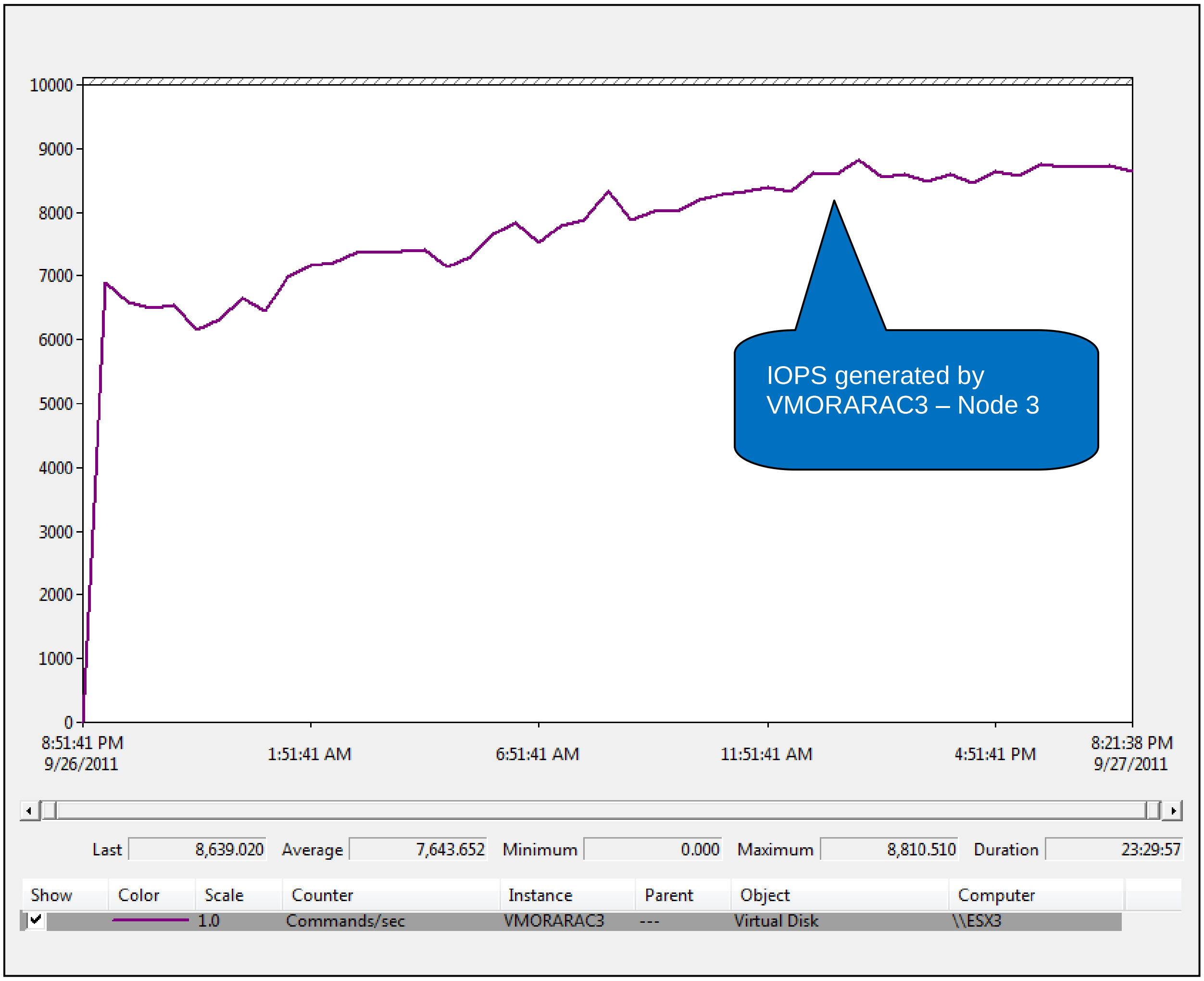
IOPS generated by VMORARAC2 – Node 2

Last --------- Average 8,064.578 Minimum 6,737.890 Maximum 9,117.340 Duration 23:29:57

Show	Color	Scale	Counter	Instance	Parent	Object	Computer
✓	—	1.0	Commands/sec	VMORARAC2	---	Virtual Disk	\\ESX2

© 2011 VMware, Inc. All rights reserved.

Figure 13. IOPS for Oracle RAC – Node 3 (VMORARAC3)



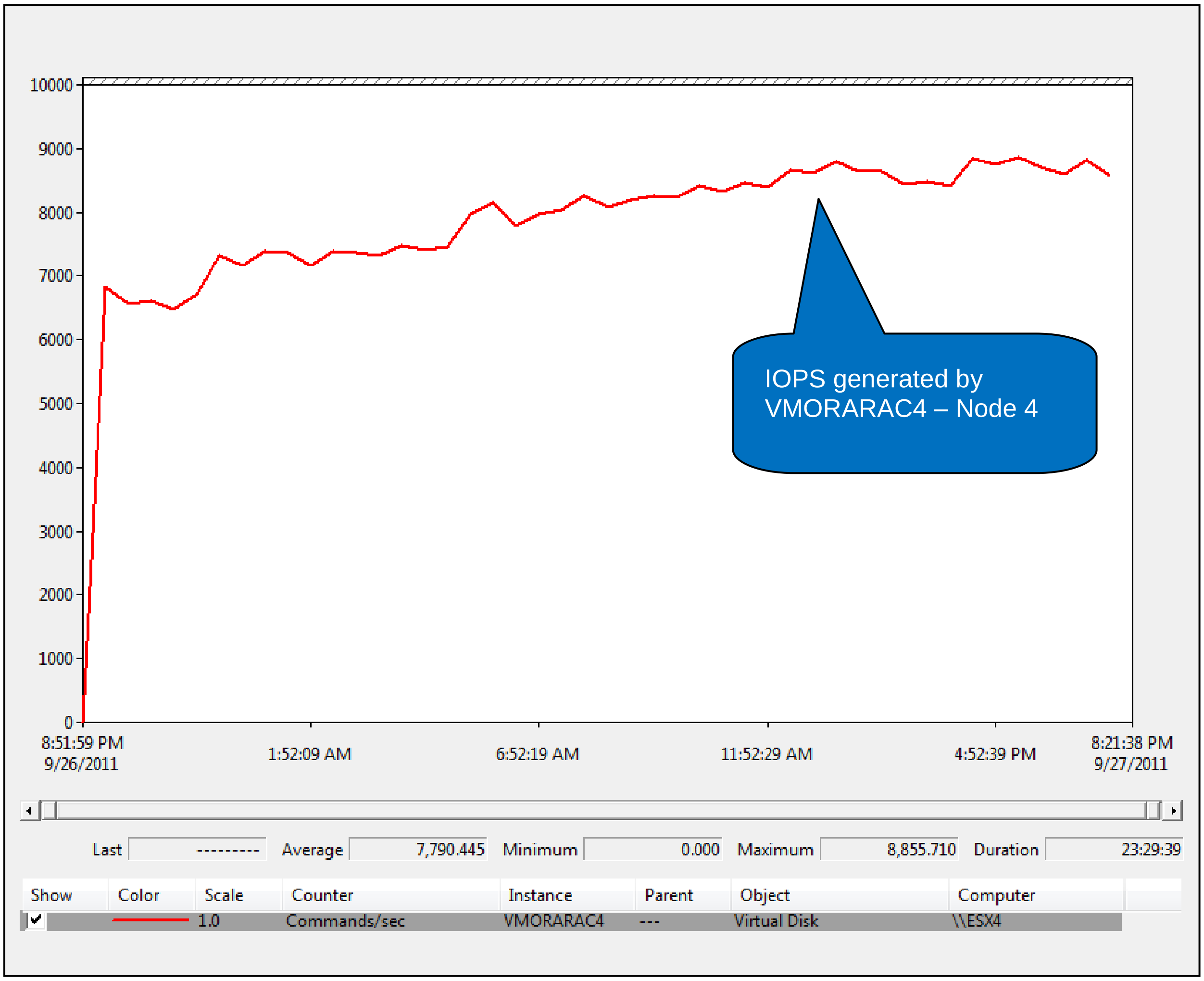
IOPS generated by VMORARAC3 – Node 3

Last	Average	Minimum	Maximum	Duration
8,639.020	7,643.652	0.000	8,810.510	23:29:57

Show	Color	Scale	Counter	Instance	Parent	Object	Computer
☑	—	1.0	Commands/sec	VMORARAC3	---	Virtual Disk	\\ESX3

© 2011 VMware, Inc. All rights reserved.

Figure 14. IOPS for Oracle RAC – Node 4 (VMORARAC4)



Last --------- Average 7,790.445 Minimum 0.000 Maximum 8,855.710 Duration 23:29:39

Show	Color	Scale	Counter	Instance	Parent	Object	Computer
✓	━	1.0	Commands/sec	VMORARAC4	---	Virtual Disk	\\ESX4

© 2011 VMware, Inc. All rights reserved.

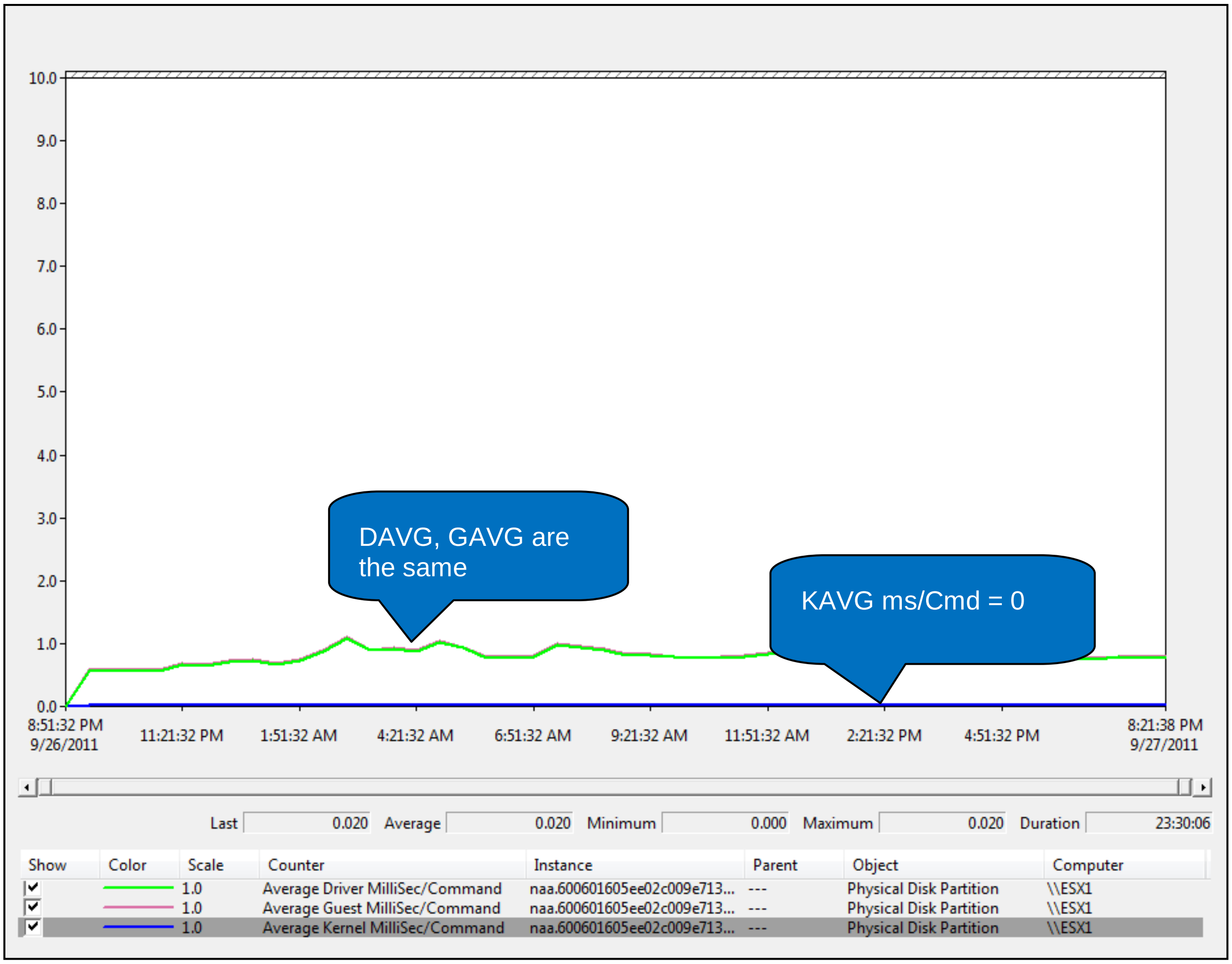
Redo LUN Disk Latency

The following chart shows the disk latencies of one of the Oracle redo LUNs during the workload run. The other redo LUNs exhibited a similar pattern. The ESX performance counters identified here (see the bottom of the chart) are described in *Using esxtop to identify storage performance issues* (http://kb.vmware.com/kb/1008205).

The counters are summarized as follows:

- DAVG/cmd – The average response time in milliseconds per command being sent to the device.
- KAVG/cmd – The time the command spends in the VMkernel.
- GAVG/cmd – The response time as it is perceived by the guest operating system. This number is calculated as DAVG + KAVG = GAVG.

Figure 15. Disk Latencies of One of the Oracle redo LUNs



DAVG, GAVG are the same

KAVG ms/Cmd = 0

Last 0.020 Average 0.020 Minimum 0.000 Maximum 0.020 Duration 23:30:06

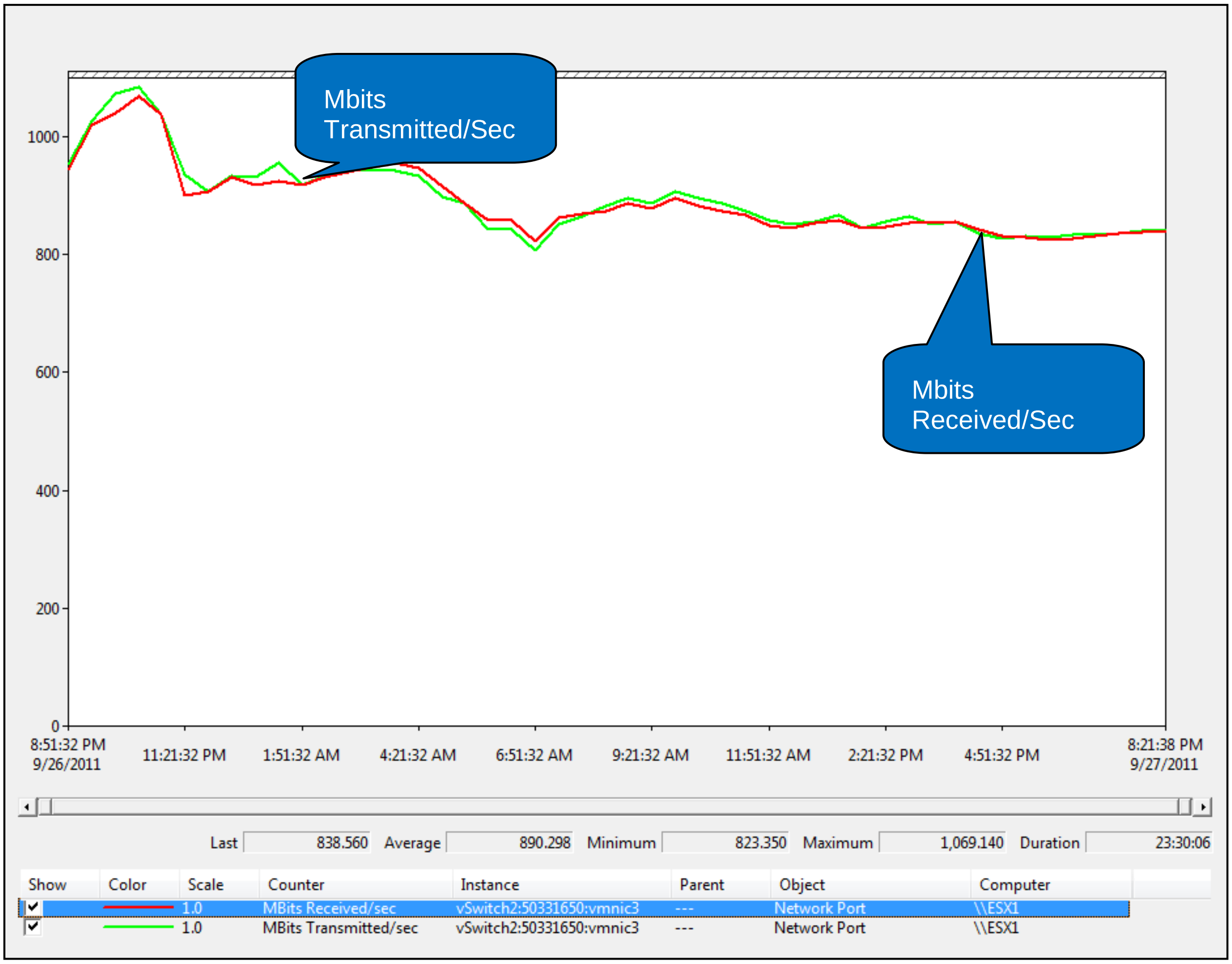
Show	Color	Scale	Counter	Instance	Parent	Object	Computer
☑	🟩	1.0	Average Driver MilliSec/Command	naa.600601605ee02c009e713...	---	Physical Disk Partition	\\ESX1
☑	🟪	1.0	Average Guest MilliSec/Command	naa.600601605ee02c009e713...	---	Physical Disk Partition	\\ESX1
☑	🟦	1.0	Average Kernel MilliSec/Command	naa.600601605ee02c009e713...	---	Physical Disk Partition	\\ESX1

© 2011 VMware, Inc. All rights reserved.

Interconnect Traffic

The following chart shows the ESX counters for megabits received and transmitted per second for the virtual NIC (vmnic3) connected to the private network (the Oracle RAC interconnect).

Figure 16. ESX Counters for Megabits Received and Transmitted per Second for Virtual NIC



Last 838.560 Average 890.298 Minimum 823.350 Maximum 1,069.140 Duration 23:30:06

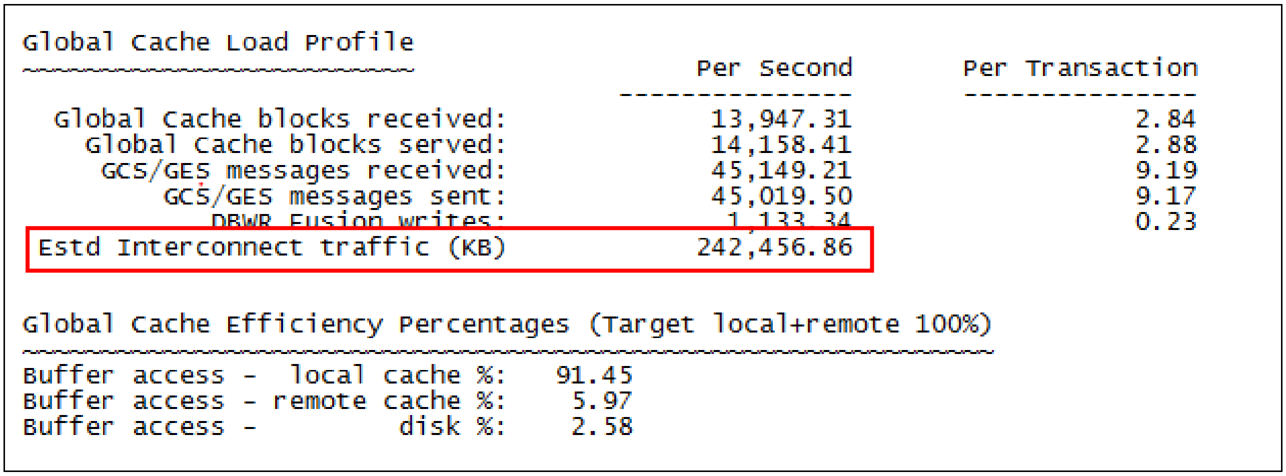
Show	Color	Scale	Counter	Instance	Parent	Object	Computer
☑	🟥	1.0	MBits Received/sec	vSwitch2:50331650:vmnic3	---	Network Port	\\ESX1
☑	🟩	1.0	MBits Transmitted/sec	vSwitch2:50331650:vmnic3	---	Network Port	\\ESX1

The total average received (890) plus transmitted (893) equals approximately 1783Mbps, or approximately 223MB per second.

© 2011 VMware, Inc. All rights reserved.

The following is an extract from the AWR report for Oracle RAC node 1 showing the Oracle RAC interconnect performance. AWR estimates 242MB per second for interconnect traffic.

Figure 17. Section of Report Showing Oracle RAC Interconnect Performance



```
Global Cache Load Profile
~~~~~~~~~~~~~~~~~~~~~~~~~                     Per Second       Per Transaction
                                         ---------------      ---------------
  Global Cache blocks received:                13,947.31                 2.84
    Global Cache blocks served:                14,158.41                 2.88
     GCS/GES messages received:                45,149.21                 9.19
         GCS/GES messages sent:                45,019.50                 9.17
            DBWR Fusion writes:                 1,133.34                 0.23
 Estd Interconnect traffic (KB)               242,456.86

Global Cache Efficiency Percentages (Target local+remote 100%)
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Buffer access -  local cache %:   91.45
Buffer access - remote cache %:    5.97
Buffer access -         disk %:    2.58
```

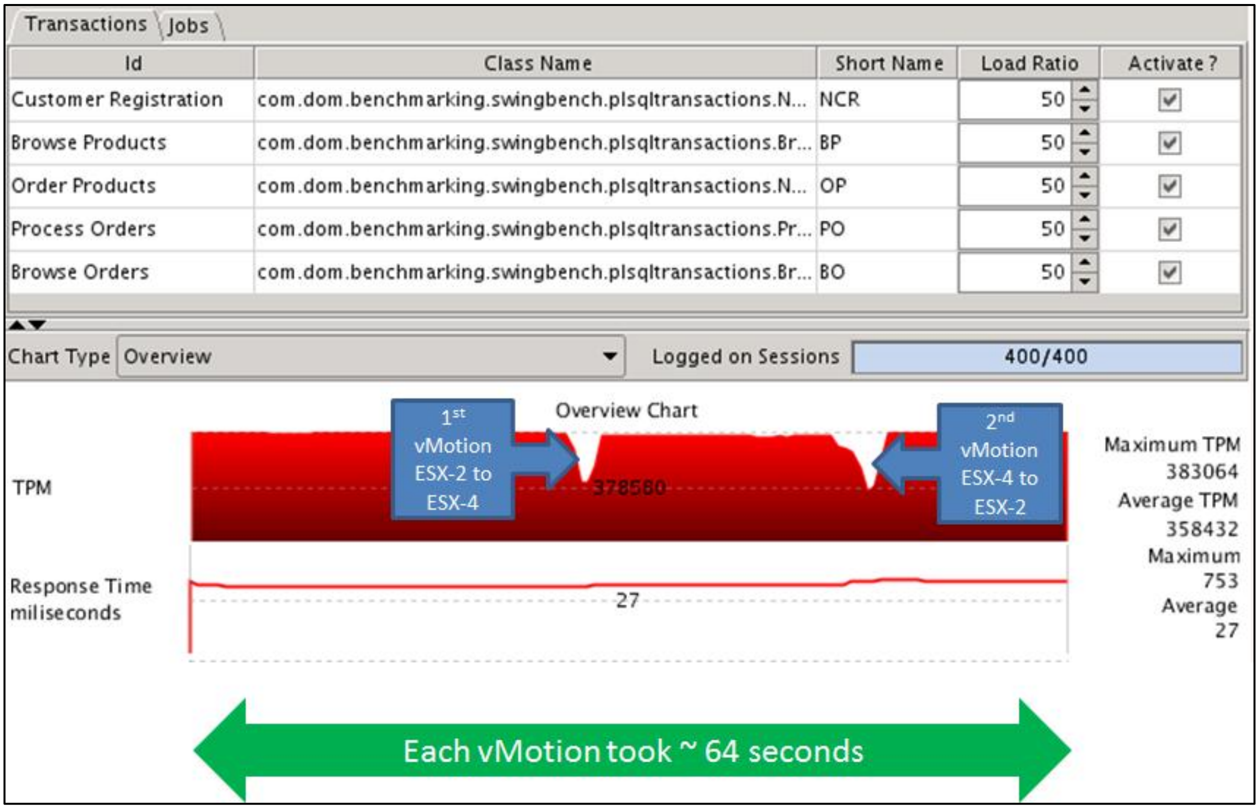
Appendix B: Oracle RAC Node vMotion Test Results

This appendix outlines some of the results from an Oracle RAC node vMotion test. It includes charts obtained from the VMware `esxtop` performance monitoring utility and vCenter.

SwingBench Results

The following SwingBench result for 400 users shows the transactions per minute (TPM) before and after a vMotion migration from ESX2 to ESX4 and back to ESX2. The chart shows TPM drops during the vMotion migration and ramps back after completing the vMotion migration. Similar results were observed during the first vMotion migration from ESX4 to ESX2.

Figure 18. vMotion SwingBench Results



Id	Class Name	Short Name	Load Ratio	Activate ?
Customer Registration	com.dom.benchmarking.swingbench.plsqltransactions.N...	NCR	50	☑
Browse Products	com.dom.benchmarking.swingbench.plsqltransactions.Br...	BP	50	☑
Order Products	com.dom.benchmarking.swingbench.plsqltransactions.N...	OP	50	☑
Process Orders	com.dom.benchmarking.swingbench.plsqltransactions.Pr...	PO	50	☑
Browse Orders	com.dom.benchmarking.swingbench.plsqltransactions.Br...	BO	50	☑

Chart Type: Overview

Logged on Sessions: 400/400

Overview Chart
TPM
378580
Response Time miliseconds
27
Maximum TPM
383064
Average TPM
358432
Maximum
753
Average
27
1st vMotion ESX-2 to ESX-4
2nd vMotion ESX-4 to ESX-2
Each vMotion took ~ 64 seconds

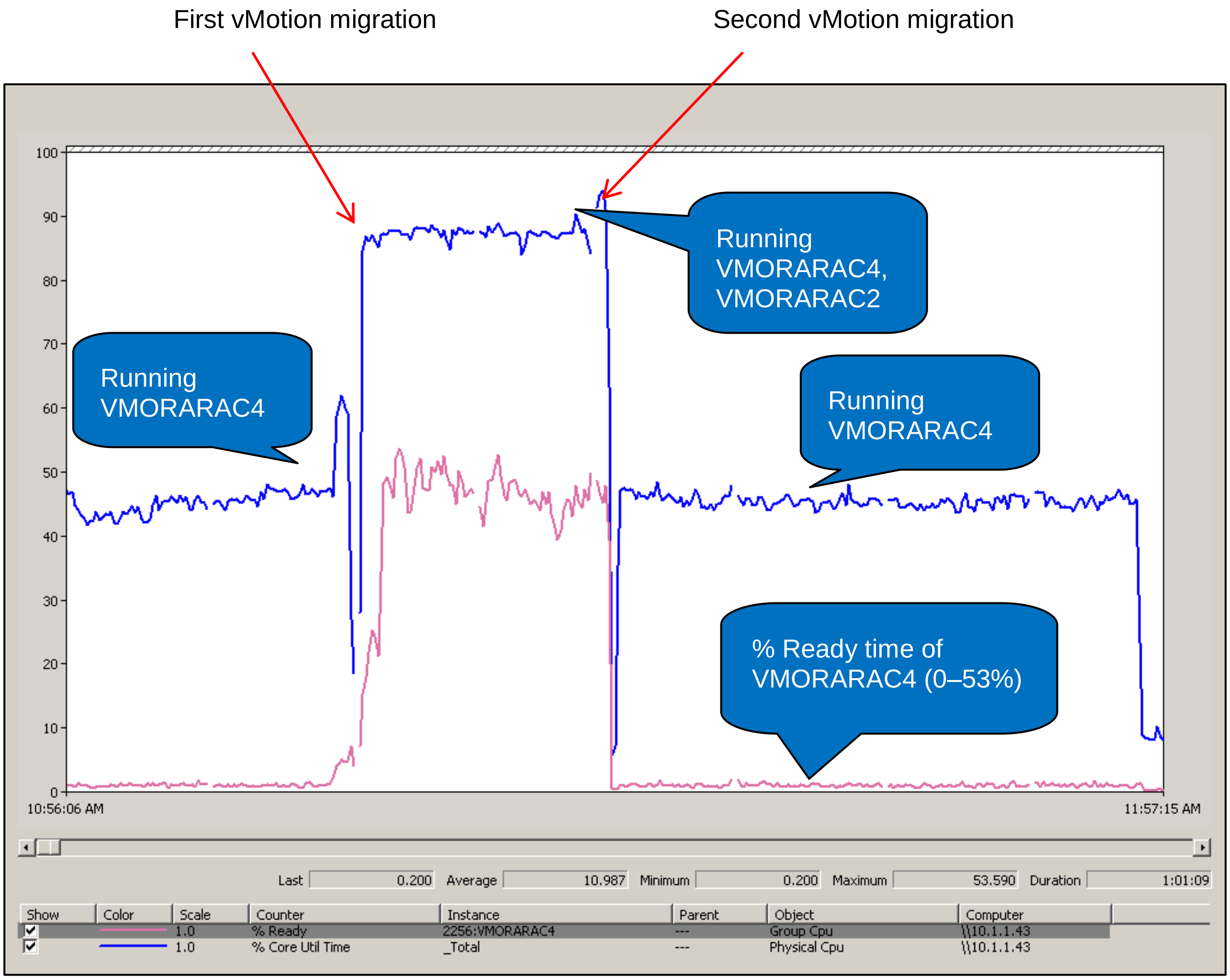
© 2011 VMware, Inc. All rights reserved.

CPU Utilization (ESX4) – vMotion

The following chart shows the percent ready time of the virtual machine VMORARAC4 (Node 4) and the core utilization of ESX host 4.

The core utilization of ESX4 (10.1.1.43) increases to approximately 85% when hosting virtual machines VMORARAC4 (node4) and VMORARAC2 (node 2).

Figure 19. Percent Ready Time and Core Utilization



Last 0.200 Average 10.987 Minimum 0.200 Maximum 53.590 Duration 1:01:09

Show	Color	Scale	Counter	Instance	Parent	Object	Computer
☑	🟪	1.0	% Ready	2256:VMORARAC4	---	Group Cpu	\\10.1.1.43
☑	🟦	1.0	% Core Util Time	_Total	---	Physical Cpu	\\10.1.1.43

Ready time is the time a virtual machine wants to run, but during which it has not been provided CPU execution resources. Typically, high ready times indicate CPU resource contention, that is, too many virtual machines competing for the CPU on the host.

Refer to the VMware Community article *Ready Time* (http://communities.vmware.com/docs/DOC-7390). As a guideline, a %ready time of less than 5% per vCPU is satisfactory. For a 12-way virtual machine this corresponds to 12 x 5% = 60%—the results in the graph are below that threshold. This demonstrates that in this scenario where Oracle RAC node virtual machines were not saturated (that is, running with reduced workload), both Oracle RAC node virtual machines can run on the same ESX host (with vCPU overcommitment, where the total number of vCPUs is greater than the total number of physical cores).

© 2011 VMware, Inc. All rights reserved.