Performance Study of Oracle RAC on VMware vSphere® 5.0

A Performance Study by VMware and EMC

TECHNICAL WHITE PAPER

vmware EMC²

Table of Contents

Introduction

IT organizations that have implemented Oracle® Real Application Clusters (RAC) often use RAC to support critical functions of their business. Performance of these database clusters is key to enabling IT organizations to meet the requirements of their businesses, customers, and shareholders.

VMware vSphere® provides a high performance virtualization platform that is capable of hosting the most critical portions of infrastructure, including Oracle RAC databases. vSphere 5 has increased these capabilities, including support for virtual machines with up to 32 vCPUs and 1TB of RAM. This support for larger virtual machines (VMs) combined with vSphere's high performance capabilities makes vSphere a great platform for running large Oracle RAC database clusters.

EMC® IT has implemented vSphere throughout many parts of their infrastructure and is now in the process of moving the largest and most critical applications to vSphere. EMC and VMware worked together to evaluate the performance of one of EMC's largest Oracle RAC databases on vSphere 5. A copy of this database was used to conduct a series of tests to compare physical performance versus virtual performance. This testing, which was done with an Oracle RAC cluster made up of six 32-CPU-based nodes, four of them virtual and two of them physical, found that virtual performed within 12% of native and was acceptable to be used for the production instance. EMC IT plans to adopt this virtual configuration in its next refresh cycle.

Test Environment

An environment was set up that was as similar to the existing production Oracle RAC database as possible to enable the best evaluation.

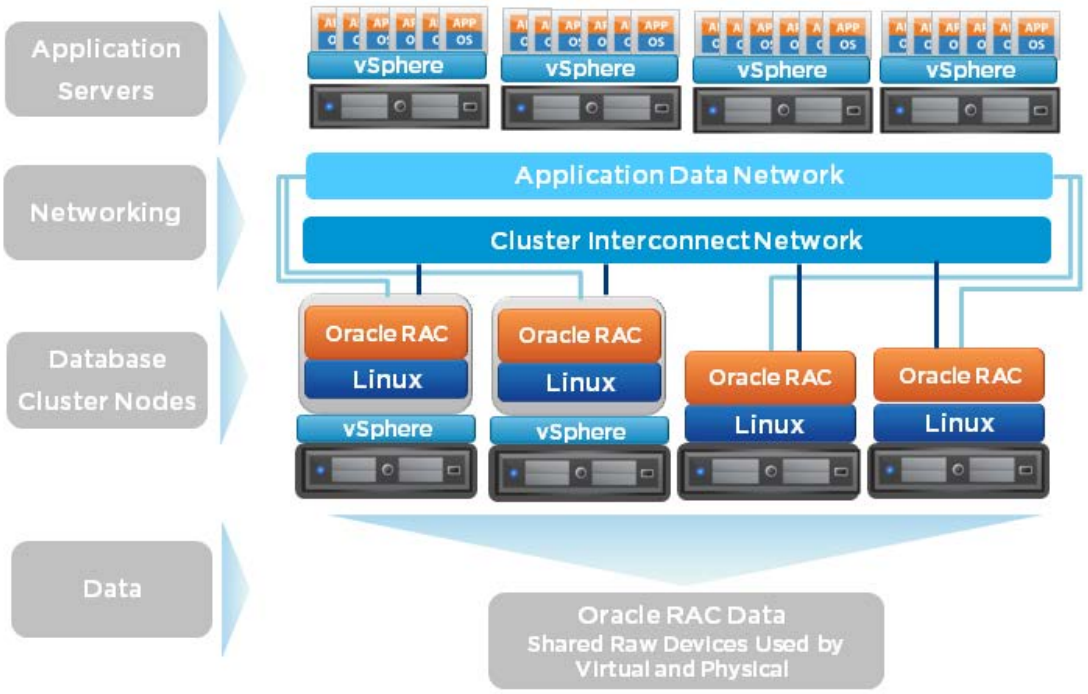

Figure 1. Test bed for performance study models production Oracle RAC database setup

Hardware

The servers were Cisco UCS servers, each with 4 Intel Nehalem 8-core processors and 256GB of RAM. Storage was an EMC Symmetrix VMAX. The database VMs were created with 32 vCPUs, 128GB of RAM, and 2 VMXNET3 virtual NICs. This configuration allowed for a direct physical and virtual comparison with the same number of CPUs in both systems.

All of the storage used by the Oracle database was presented to both the physical hosts and database VMs as raw devices, which allowed for a mixed virtual and physical cluster to exist. This configuration allowed the database to be quickly switched over from physical to virtual and to use the same storage hardware and configuration for all tests.

Networking and storage were connected to the hosts using 10 gigabit Ethernet adapters. Fibre Channel over Ethernet was used for the storage connection to the server. The VM used the paravirtual SCSI adapter for storage and VMXNET3 adapters for the network. One of the VMXNET3 adapters was dedicated for the interconnect and the other was used for the public network. Coalescing was disabled on the interconnect to achieve the lowest possible latency.

Software

The software used in both the physical and virtual tests was exactly the same except that in the virtual test vSphere 5 was installed on the hosts first, and then a VM was created where Linux and Oracle were installed.

Workload

Testing was designed to allow as direct a comparison as possible to the existing EMC Oracle RAC production environment. In addition to the Oracle RAC database, there was a set of application servers that remained the same for all tests. The workload was directed to these application servers, which in turn used the Oracle RAC-based database.

The workload consisted of over 50 different types of business transactions that were each modeled on activities done against the production system by real users. Load Runner was used to simulate the users with a different script for each business transaction type.

These business transactions range from simple single step operations, to complex workflows of many steps, to requests for the creation of reports. This meant that each transaction had a different profile in terms of the resources needed to complete and the time for completion.

Two different workload levels were simulated: a typical day and a very busy day like the end of quarter or end of year. The typical day was considered to be a 100% load and a very busy day was twice as much, or a 200% load.

Measuring Oracle RAC Performance

Performance was measured from the perspective of the end user which in this testing meant the amount of time it took for each business transaction to complete or the response time. Because the transactions are different and complete in different amounts of time, the average for each transaction was computed separately and then all the average response times were added together resulting in a total response time for physical and a total response time for virtual.

Results

Response times were slightly slower but still within an acceptable range for virtual in both the 100% and 200% workload tests. There was a 12% difference in response time for the 100% test and a 2% difference in response time for the 200% test. This results in an average difference of 7% across the two different workload levels.

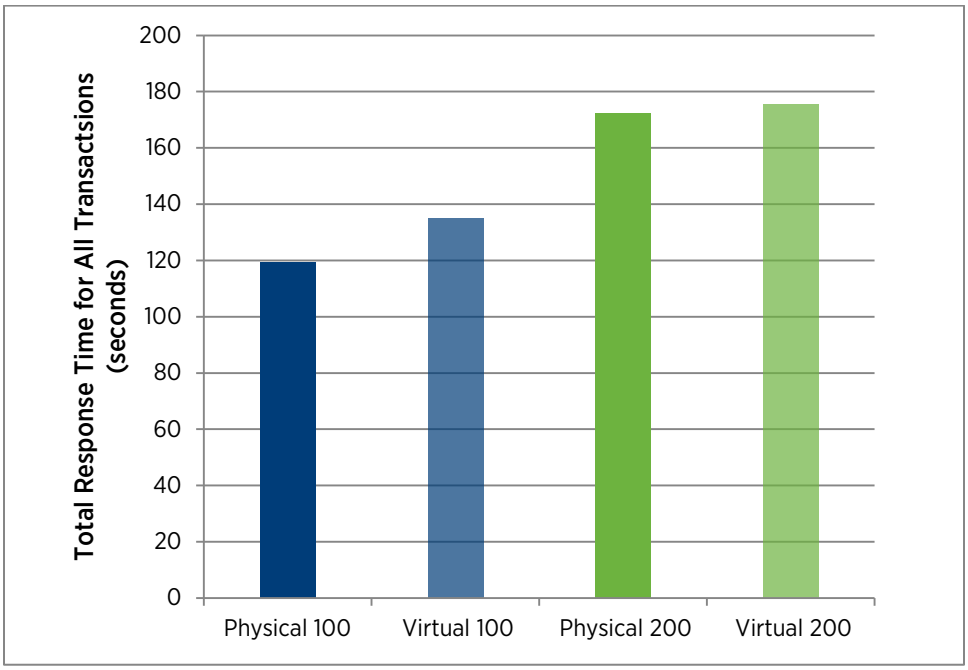

Figure 2. Oracle RAC performance compared on physical and virtual systems at 100% and 200% workload tests

The systems were sized to ensure that CPU utilization stayed below 50% even with the higher workload level so that if a RAC node were to fail or become unavailable, there would still be plenty of capacity available on the remaining nodes to handle the load. This meant that CPU utilization was below 25% in the 100% test case and below 50% in the 200% case.

A bigger difference in performance at lower utilization levels is sometimes seen when comparing physical and virtual. This is because the performance cost of virtualization is bigger in percentage terms at these lower utilization levels. As more work is done, there are more opportunities for optimizations in vSphere to be effective and for performance costs to be spread and averaged out. As the utilization levels increase, there is often a narrowing of the performance difference percentage between physical and virtual.

Best Practices

In the course of testing there were a few best practices developed specific to Oracle RAC to ensure the best performance. They revolved around the network interconnect, NUMA, hyper-threading, and storage. Standard Oracle database and VMware vSphere best practices still apply in addition to the best practices discussed in this section.

The Oracle RAC interconnect network that is used by the cluster for inter-node communication is key to achieving good performance. In order to keep the latency as low as possible, the VMXNET3 adapter type was used for the virtual NIC and coalescing was disabled using `ethernetX.coalescingScheme = disabled`. For details, see the white paper *Best Practices for Performance Tuning of Latency-Sensitive Workloads in vSphere Virtual Machines* at http://www.vmware.com/files/pdf/techpaper/VMW-Tuning-Latency-Sensitive-Workloads.pdf.

NUMA architectures are used in all current x86-based servers. To allow applications running inside VMs to fully take advantage of NUMA, vSphere introduced the vNUMA feature that allows the underlying NUMA architecture to be exposed to the guest. The amount of performance gain possible from NUMA highly depends on the support for it from the operating system and application. Current versions of Linux, including the version used in this testing, are NUMA aware and attempt to make scheduling decisions to optimize performance. Oracle also has support for NUMA, but it is recommended that you test it with your workload to determine if it will improve your performance (see Oracle MySupport Doc ID 864633.1). For the workload used in this testing, using Oracle NUMA support did not improve performance and actually made it slightly worse. This resulted in a configuration with NUMA enabled in the server hardware, for the guest with vNUMA, and in the operating system, but disabled in Oracle. This is the default NUMA configuration for the hardware, operating system, and Oracle database. The vNUMA was configured to be 2 sockets with 16 cores per socket.

Enabling hyperthreading is recommended because it almost always provides some performance boost and has not shown to cause any performance degradation. In addition, it was found that using the PreferHT parameter, which causes hyperthreads to be preferred over cores, further improved performance with this workload because the CPU utilization of the 32 vCPUs was always below 50%, allowing the VM to run comfortably on 16 real cores with hyperthreading enabled for 32 threads. With PreferHT enabled for the VM, all 32 vCPUs were scheduled to run on 16 cores. This is why vNUMA was configured to be 2 sockets with 16 cores per socket. The gain in performance comes from fewer NUMA nodes which results in more local memory accesses. Please see VMware KB article 2003582 at http://kb.vmware.com/kb/2003582 for details on how to set PreferHT.

In addition to PreferHT causing the workload to perform better, it also leaves half of the system available to be used for another Oracle RAC node. This reduces license costs and increases both the utilization and performance at the same time.

Storage is always a big component of the performance for an Oracle database. The paravirtual SCSI adapter was used for the Oracle RAC VMs to provide the lowest latency and highest efficiency.

Conclusion

A large Oracle RAC instance on vSphere 5 was shown to perform within 7% of native physical performance on average across many different business transaction types and two different workload levels. This enabled EMC IT to feel confident about using vSphere to support their production Oracle RAC databases. A few best practices that optimized the performance of the interconnect network, process scheduling, and storage performance were identified and can be easily implemented for vSphere environments that support Oracle RAC databases.

About the Authors

Todd Muirhead is performance engineer at VMware.

Darryl Smith and Dilip Chavan are Oracle architects at EMC.

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com