Oracle® Databases on VMware vSphere™ 4

May 2010

ESSENTIAL DEPLOYMENT TIPS

Table of Contents

Introduction

Even the most demanding Oracle® database workloads can now be virtualized with VMware vSphere™ and ESX® 4—with greater than 95 percent of Oracle instances matching native performance. This paper provides the essential tips necessary to successfully deploy Oracle on VMware virtual infrastructure to enable database administrators (DBAs) to meet their performance and availability goals.

The successful deployment of Oracle on VMware virtual infrastructure is not significantly different from deploying Oracle on physical servers. To paraphrase an excerpt from Dr. Burt Scalzo's book, "98 percent of Oracle database physical tuning and optimization is directly applicable to the virtual world." So, it is essentially a myth that DBAs must relearn their skills in order to deploy Oracle on VMware. The fact is that DBAs can fully leverage their current skill set, while delivering all the benefits associated with virtualization.

This paper also takes a proactive approach to addressing performance issues. At VMware greater than 90 percent of the performance issues encountered by our customers were due to configuration errors at the storage tier. For this reason, a significant portion of the paper will deal with the storage tier.

Purpose

The purpose of this document is to provide technical guidance when deploying Oracle databases on VMware vSphere. This document will also show that the same best practices, tuning tips and tricks, and skill sets necessary to deploy Oracle databases in physical environments can be leveraged when deploying Oracle databases in virtual environments. This document assumes a moderate understanding of Oracle databases and a fundamental understanding of VMware virtualization technology.

VMware vSphere 4

Chasing the Database Bottleneck

In order to maintain acceptable performance levels in production databases, DBAs can spend much of their time "chasing the bottleneck". Bottlenecks change not only as the number of users and/or the size of the database grows; they also change with technology. Think how radically different it is when tuning and sizing Oracle System Global Area (SGA) shared memory for 32-bit or 64-bit operating systems. DBAs are limited to 1.75 GB SGA memory for 32-bit operating systems as opposed to SGAs with tens to hundreds of GB memory for 64-bit operating systems.

VMware ESX is no different. VMware vSphere technology has removed the virtualization bottleneck (see Table 1) and advances made to vSphere now make it possible to virtualize the most challenging database workloads. With vSphere and ESX 4, 95 percent of Oracle databases can match native performance, while fully saturated Oracle databases only experience anywhere from 2 to 10 percent overhead.[1]

TIP 1: Upgrade to vSphere ESX 4.

To attain maximum performance it is prudent to upgrade your current ESX deployment to VMware vSphere. Post-upgrade, VMware administrators and database administrators can realize anywhere from a 10 to 20 percent performance boost. By putting the question of performance behind, administrators can focus on introducing vSphere core and advanced features to the enterprise—like vMotion™, VMware Distributed Resource Manager (DRS), VMware HA, and VMware Disaster Recovery.

[1] See performance study "Virtualizing Performance Critical Database Applications with vSphere".

Upgrade Philosophy

When upgrading databases, it is perfectly acceptable for DBAs to deploy new Oracle database features, which may introduce overhead at the expense of performance. Incurring the minimal overhead introduced by vSphere is also a perfectly acceptable architectural tradeoff, especially when considering all the benefits associated with VMware virtualization.

Table 1. ESX Versions

	ESX 2.0	ESX 3.0	ESX 3.5	vSphere ESX 4
Overhead	30% - 60%	20% - 30%	10% - 20%	2% - 10%
CPU	1 vCPU	2 vCPU	4 vCPU	8 vCPU
Memory	< 4GB	16 GB	64 GB	255 GB
Network	380 Mb/Sec	800 Mb/Sec	9 Gb/Sec	30 Gb/Sec
IOPS	< 10,000	20,000	100,000	> 350,000

Purpose-Built Computing Environments

Deploying Oracle on the VMware vSphere platform gives DBAs the ability to create optimized, purpose-built computing environments. The first step in creating such an environment requires a careful examination of BIOS settings, disabling of unnecessary processes and peripherals, and compilation of a monolithic kernel to direct the critical compute resources (which are CPU, memory, network, and I/O) to the databases.

TIP 2: Create a Computing Environment Optimized for vSphere.

BIOS Settings

The BIOS settings listed in Table 2 vary based on chipset family and the motherboard.[2]

Table 2. Chipset BIOS Settings

BIOS Setting	Recommendations	Description
Virtualization Technology	Yes	Necessary to run 64-bit guest operating systems.
Turbo Mode	Yes	Balanced workload over unused cores.
Node Interleaving	No	Will disable NUMA benefits if disabled.
VT-x, AMD-V, EPT,RVI	Yes	Hardware-based virtualization support.
C1E Halt State	No	Disable if performance is more critical than power saving.
Power-Saving	No	Disable if performance is more important than power saving.
Virus Warning	No	Disables warning messages when writing to the master boot record.

[2] See Web-Based Compatibility Guide.

BIOS Setting	Recommendations	Description
Hyper-Threading	Yes	For use with some Intel processors. Hyper-Threading is always recommended with Intel's newer Core i7 processors such as the Xeon 5500 series.
Video BIOS Cacheable	No	Not necessary for database virtual machine.
Wake On LAN	Yes	Required for vSphere Distributed Power Management feature.
Execute Disable	Yes	Required for vMotion and Distributed Resource Scheduler features.
Video BIOS Shadowable	No	Not necessary for database virtual machine.
Video RAM Cacheable	No	Not necessary for database virtual machine.
On Board Audio	No	Not necessary for database virtual machine.
On Board Modem	No	Not necessary for database virtual machine.
On Board Firewire	No	Not necessary for database virtual machine.
On Board Serial Ports	No	Not necessary for database virtual machine.
On Board Parallel Ports	No	Not necessary for database virtual machine.
On Board Game Port	No	Not necessary for database virtual machine.

Operating System Installation

In planning an operating system install, do not install operating system components that are not necessary for an optimized compute environment. Note that disabling the peripheral components in the BIOS does not guarantee these components will be fully disabled. In addition to disabling these components in the BIOS, make sure they are also not part of the operating system installation process.

Examples of software components that should not be part of the operating system install are:

- Office Productivity Suites
- Graphics, sound, and video programs
- Instant Messaging services

Operating System Host Processes

After the operating system has been successfully installed, the next step is to disable unnecessary foreground and background processes.

Linux Processes

Examples of unnecessary Linux processes are:

- anacron, apmd, atd, autofs, cups, cupsconfig, gpm, isdn, iptables, kudzu, netfs, and portmap.

Windows Processes

Examples of unnecessary Windows processes are:

- alerter, automatic updates, clip book, error reporting, help and support, indexing, messenger, netmeeting, remote desktop, and system restore services.

Optimized Operating Systems

When creating an optimized operating system installation for Oracle on vSphere deployment, do not limit your review of components just to the BIOS settings, software, or operating system processes described in the previous sections. That said, depending on your IT organizational structure, reviewing such details may not solely be the database administrator’s responsibilities. It may and should require the inputs from other network, storage, and system administrators to formulate the most optimal system environment and make the best decisions.

As an example, large page tables should be used if they are supported by the operating system as well as the database. Further details about this are covered in the “Memory Considerations” section, provided later in this document.

Lastly, for Linux installs, the database administrator (DBA) should request that the system administrator compile a monolithic kernel, which will only load the necessary features. Whether you intend on running Windows or Linux as the final optimized operating system, these host installs should be cloned by the VMware administrator for reuse.[3]

TIP 3: Create Golden Images of Optimized Operating Systems using vSphere Cloning Technologies.

Once the operating system has been prepared, Oracle can be installed the same way as you would normally install the database for a physical environment. Use the recommended kernel parameters listed in the appropriate Oracle Installation guide. Also, it is always a good practice to check with Oracle Support for the latest settings to use, prior to beginning the installation process.

[3] Work with your VMware administrator when creating clones.

CPU Considerations

Oracle databases are not usually heavy CPU consumers and therefore are not characterized as CPU-bound applications. This makes Oracle databases excellent candidates for virtualization because unused CPU cycles are available to allow for consolidation and advanced virtualization features. For this reason, the vast majority of virtualized Oracle databases will exhibit throughput similar to that of native implementations.

Virtual CPUs

When configuring Oracle database virtual machines, the total CPU resources needed by the virtual machines running on the system should not exceed the CPU capacity of the host. It is good practice to actually under-commit CPU resources on the host because, if the host CPU capacity is overloaded, the performance of your virtual database may degrade.

However when using VMware vSphere advanced workload management features such as vMotion and VMware DRS, the database is freed from the resource limitations of a single host. VMware vMotion enables DBAs to move running Oracle virtual machines from one physical ESX server to another, to balance available resources with little impact to end users. VMware DRS dynamically allocates and balances computing resources by continually monitoring the utilization of resource pools associated with virtual machines in a VMware cluster.

Performance should normally be monitored through vSphere vCenter. However, it is a good practice to periodically collect additional statistical measures of the host CPU usage. This can be done through the vSphere Client, or by using esxtop or resxtop. CPU usage tips are listed below. Work with your VMware administrator to interpret esxtop data:

- If the load average listed on the first line of the esxtop CPU Panel is equal to or greater than the number of physical processors in the system, this indicates that the system is overloaded.
- The usage percentage of physical CPUs on the PCPU line can be another indication of a possibly overloaded condition.

In general, 80 percent usage is a reasonable ceiling in production environments, and 90 percent should be used as an alert to the VMware administrator that the CPUs are approaching an overloaded condition, which should be addressed. However, decisions concerning usage levels should actually be made based on the criticality of the Oracle database being virtualized, regarding the desired load percentage.

When using esxtop, three critical statistics to interpret are:

- **%RUN** – The percentage of total time the “world” [4] is running on the processor; if %RUN is high, it does not necessarily mean that the virtual machine is resource-constrained. (See description of %RDY below.)
- **%RDY** – The percentage of time the world was ready to run but is not scheduled to a core. A world in a run queue is waiting for CPU scheduler to let it run on a PCPU. If %RDY is greater than 10 percent, then this could be an indication of resource contention.
- **%CSTP** – The percentage of time the world is stopped from running to allow other vCPUs in the virtual machine to catch up, co-deschedule state. If %CSTP is greater than 5 percent, this usually means the virtual machine workload is not using VCPUs in a balanced fashion.

By using esxtop, DBAs can gain additional performance insight with respect to CPU resource contention. DBAs should also work with their VMware administrator to fully understand and interpret esxtop statistics (beyond the scope of this paper).

TIP 4: Use as Few Virtual CPUs (vCPUs) as Possible.

[4] Esxtop uses worlds and groups as the entities to show CPU usage. A **world** is an ESX VMkernel schedulable entity, similar to a process or thread in other operating systems. A **group** contains multiple worlds.

Even if some vCPUs are not used, configuring virtual Oracle database with excess vCPUs can imposes some small resource requirements on vSphere due to the fact that unused vCPUs still consume timer interrupts. vSphere attempts to co-schedule multiple vCPUs of a virtual machine, trying to run vCPUs in parallel as much as possible. Having unused vCPUs imposes scheduling constraints on the vCPU being used and can degrade its performance.

Hyper-Threading Technology

Hyper-threading technology allows a single physical processor core to behave like two logical processors, essentially allowing two independent threads to run simultaneously on a single core. Unlike having twice as many processor cores that can roughly double performance, hyper-threading can provide anywhere from a slight to a significant increase in system performance by keeping the processor pipeline busier.

TIP 5: Enable Hyper-Threading for Intel Core i7 Processors.

With the release of Intel Xeon 5500 series processors, enabling Hyper-threading is recommended. Prior to the 5500 series, VMware had no uniform recommendation with respect to Hyper-Threading since the performance results measured were not consistent across applications, run environments, or database workloads.

Memory Considerations

Virtual Memory

One of the primary concerns for DBAs is maintaining consistent and repeatable database performance in order to comply with stringent service level agreements (SLAs) established with application owners.[5]

TIP 6: Set Memory Reservations Equal to the Size of the Oracle SGA.

When consolidating Oracle database instances, vSphere presents the opportunity for sharing memory across virtual machines that may be running the same operating systems, applications, or components. In this case, vSphere uses a proprietary transparent page sharing technique to reclaim memory, which allows databases to run with less memory than physical. Transparent page sharing also allows DBAs to over-commit memory, without any performance degradation.

In production environments, careful consideration should be taken when over-committing memory and should only be introduced after collecting data to determine the amount of over-commitment possible. To determine the effectiveness of memory sharing and the degree of acceptable over-commitment for a given database, run the workload, and use resxtop or esxtop to observe the actual savings.

While VMware recommends setting memory reservations equal to the size of the Oracle SGA in production environments, it is perfectly acceptable to introduce more aggressive over-commitment in non-production environments such as development, test, or QA. In these environments, a DBA can introduce memory over-commitment to take advantage of VMware's memory reclamation features and techniques. Even in these environments, the type and number of databases that can be deployed using over-commitment will be largely dependent on their usage characteristics and their criticality to the business.

[5] Refer to "VMware vSphere Resource Management Guide" for concepts discussed in these sections.

Hardware-Assisted Memory Virtualization

Some recent processors include a new feature that addresses the overhead due to memory management unit (MMU) virtualization by providing hardware support to virtualize the MMU. VMware ESX 4 supports this feature in both AMD and Intel processors. AMD dubs this technology Rapid Virtualization Indexing (RVI) or Nested Page Tables (NPT) and Intel calls it Extended Page Tables (EPT). Without hardware-assisted MMU virtualization, ESX maintains "shadow page tables" that directly map guest virtual memory to host physical memory addresses.

These shadow page tables are maintained for use by the processor and are kept consistent with the guest page tables. This allows ordinary memory references to execute without additional overhead (since the hardware translation look-aside buffer (TLB) will cache direct guest virtual memory to host physical memory address translations read from the shadow page tables). However, extra work is required to maintain the shadow page tables.

When you use hardware assistance, you eliminate the overhead for software memory virtualization. In particular, hardware assistance eliminates the overhead required to keep shadow page tables in synchronization with guest page tables. However, the TLB miss latency is significantly higher when using hardware assistance. As a result, whether or not a workload benefits by using hardware assistance depends primarily on the overhead the memory virtualization causes when using software memory virtualization. If a workload involves a small amount of page table activity (such as process creation, mapping the memory, or context switches), software virtualization does not cause significant overhead. Conversely, workloads like those from a database, which have a large amount of page table activity, are likely to benefit from hardware assistance.

TIP 7: Allow vSphere to Choose the Best Virtual Machine Monitor based on the CPU and Guest Operating System Combination.

Make sure the virtual machine setting has Automatic selected for the CPU/MMU Virtualization option.[6]

Large Memory Pages

Oracle announced support for the use of large memory pages in version 9iR2 for Linux operating systems and in version 10gR2 for Windows. VMware introduced support for the use of large pages inside virtual machines in ESX version 3.5. The large-page support enables applications like Oracle Database to establish large-page memory regions. The use of large pages can potentially increase TLB access efficiency and thus improve database performance.

TIP 8: Use Large Memory Pages.

The use of large pages can significantly improve the performance of Oracle databases on vSphere, compared to running the workload using small pages.[7] Large page support is enabled by default in ESX versions 3.5 and later. Consult your Oracle Administration Guide to determine SGA memory conversion settings. Also read the following Metalink Notes depending on your operating system.

Linux Huge Pages Metalink Notes:

- ***Note 361323.1*** – "Huge Pages on Linux: What It Is... and What It Is Not..."
- ***Note 361468.1*** – "Huge Pages on 64-bit Linux"
- ***Note 401749.1*** – "Shell Script to Calculate Values Recommended Huge Pages / Huge TLB Configuration"

[6] More detailed instructions can be found in the "Performance Best Practices for VMware vSphere" white paper listed in the References section, provided later in this document.

[7] See the performance study "Large Page Support for ESX Server 3.5 and ESX Server 3i v3.5" listed in the References section.

Windows Server Large Pages Metalink Notes:

- ***Note 46001.1*** – "Oracle Database and the Windows NT memory architecture, Technical Bulletin"
- ***Note 46053.1*** – "Windows NT Memory Architecture Overview"

Network Considerations

VMware vSphere is capable of handling sustained network transfer rates greater than 30Gb per second, however, Oracle is not a large network consumer. In a recent benchmark running a large fully-saturated Oracle database on vSphere, the maximum network bandwidth usage was well below 100Mb per second.[8] As databases grow to terabyte scale, network bandwidth should remain a minor consideration for Oracle database performance, due to advanced compression techniques available in Oracle 11g as well as vSphere's ability to handle throughput rates well beyond the requirements of Oracle itself.[9] So, when deploying Oracle databases, following general networking best practices for vSphere configuration and the guest operating system should be sufficient.

General vSphere Network Guidance

This section provides general guidance for setting up VMware vSphere network topology to match your Oracle system design architecture[10]:

- Use separate virtual switches, with each switch connected to its own physical network adapter to avoid contention between the ESX service console, the VMkernel, and virtual machines (especially virtual machines running heavy networking workloads).
- To establish a network connection between two virtual machines that reside on the same ESX host, connect both virtual machines to the same virtual switch. If the virtual machines are connected to different virtual switches, traffic will go through wire and incur unnecessary CPU and network overhead.

General Guest Operating System Guidance

This section provides general configuration guidance when creating virtual machines for Oracle database deployments:

- The default virtual network adapter emulated inside a guest is either an AMD PCnet32 device (vlance) or an Intel E1000 device (E1000).
- VMware also offers the VMXNET family of paravirtualized network adapters, which can sometimes provide better performance than the default adapters.

TIP 9: Use the VMXNET Family of Paravirtualized Network Adapters.

The VMXNET family contains VMXNET, Enhanced VMXNET (available since ESX 3.5), and VMXNET Generation 3 (VMXNET3; newly added in ESX 4).

The paravirtualized network adapters in the VMXNET family implement an idealized network interface that passes network traffic between the virtual machine and the physical network interface cards with minimal overhead. Drivers for VMXNET-family adapters are available for most guest operating systems supported by ESX.[11]

[8] See the "Virtualizing Performance Critical Database Applications with VMware vSphere" white paper.

[9] The benchmark described here was not performed using Oracle 11g advanced compression features.

[10] For more additional networking configuration details, see the "Performance Best Practices for vSphere 4" white paper.

[11] See "Guest Operating System Installation Guide" for supported drivers and compatibility.

Networked Storage Systems

When deploying vSphere, the choice of a networked storage system has little to do with virtualization. As with any physical Oracle deployment, the main considerations are still price, performance, and manageability. In addition, the protocols available with vSphere—that is, Fibre Channel, Hardware iSCSI, Software iSCSI, and NFS—are capable of achieving throughput levels that are limited only by the capabilities of the storage array and its connection to vSphere. When considering CPU cost, Fibre Channel and Hardware iSCSI are more efficient than Software iSCSI and NFS. However, when CPU resources are not a bottleneck, Software iSCSI and NFS can also be part of a high-performance solution.[12]

Storage Protocol Capabilities

When selecting networked storage systems and protocols, it is critical to understand which vSphere features are supported. Table 3 describes the capabilities for each of the protocols available in vSphere.

Table 3. Storage Protocol Capabilities

Type	Boot VM	Boot vSphere	VMotion HA/DRS	VMFS	RDM	MSCS Cluster	SRM
Fibre Channel	Yes	Yes	Yes	Yes	Yes	Yes	Yes
iSCSI	Yes	Yes [13]	Yes	Yes	Yes	No	Yes
NAS	Yes	No	Yes	No	No	No	No
Local Storage	Yes	Yes	No	Yes	Yes	No	No

TIP 10: For IP-Based Storage iSCSI and NFS, Enable Jumbo Frames.

Jumbo Frames must be enabled for each vSwitch through the vSphere CLI. Also, if you use an ESX host, you must create a VMkernel network interface enabled with Jumbo Frames. It is also necessary to enable Jumbo Frames on the hardware as well, including the network switches and storage arrays.

Thin Provisioning

For Fibre Channel and iSCSI, thin provisioning is supported in VMFS but not as a default; resizing a datastore is done via extents but is not recommended while the database is in production. An advantage of NFS is that it provides Thin Provisioning as a default for both the datastore and VMDK, which only consumes the actual disk capacity utilized. In addition, the creation of datastores is straightforward. A VMware administrator must first mount the NFS volume to the ESX host before creating datastores to be shared among ESX hosts.[14]

[12] See the performance study " Comparison of Storage Protocol Performance in VMware vSphere 4".

[13] vSphere server boot for iSCSI hardware initiator only.

[14] See the "Dynamic Storage Provisioning: Considerations and Best Practices for Using Virtual Disk Thin Provisioning" white paper.

Datastores

VMware vSphere uses datastores to store virtual disks. Datastores can be thought of as an abstraction of the storage layer that hides all the physical attributes of the storage devices to the virtual machines. VMware administrators can create datastores that can be used as a single consolidated pool of storage, or many datastores, which can be used to isolate various application workloads.

TIP 11: Create Dedicated Datastores to Service Database Workloads.

Consolidated or Dedicated Datastores

It is a generally accepted best practice to create a dedicated datastore if the application has a demanding I/O profile; databases fall into this category. The creation of dedicated datastores allows DBAs to define individual service level guarantees for different applications and is analogous to provisioning dedicated LUNs in the physical world.

It's critical to understand that a datastore is an abstraction of the storage tier and, therefore, it is a logical representation of the storage tier, not a physical representation of the storage tier. So, creating a dedicated datastore to isolate a particular I/O workload (whether that be log or database files), without isolating the physical storage layer as well, will not have the desired effect on performance.

Virtual Machine File System (VMFS)

VMware VMFS is a high performance cluster file system, which provides storage virtualization that is optimized for virtual machines. Each virtual machine is encapsulated in a small set of files; and VMFS is the default storage management interface used to access those files on physical SCSI disks and partitions.

VMFS allows IT organizations to greatly simplify virtual machine provisioning by efficiently storing the entire machine state of a virtual machine in a central location. VMFS allows multiple ESX instances to access shared virtual machine storage concurrently. It also enables virtualization-based distributed infrastructure services such as vMotion, VMware DRS, and VMware HA to operate across a cluster of ESX hosts.

TIP 12: Use vSphere VMFS for Single-Instance Oracle Database Deployments.

To balance performance and manageability in a virtual environment, it is an accepted best practice to deploy Oracle using VMFS. Raw Device Mappings (RDMs) are sometimes selected due to the erroneous belief that they provide increased performance. However, research has shown that the two dominant workloads associated with Oracle databases (random read/write and sequential writes) have nearly identical performance throughput characteristics when deployed on VMFS or using RDM.[15]

Some customers concerned with Oracle's support statement, which gives Oracle the right to ask customers to reproduce an issue in a physical environment (if they suspect VMware is at fault), use RDMs. Since RDMs act as a proxy, they can be directly mapped and mounted to a physical server, in the rare event that Oracle support would make such a request. In addition, Oracle support on VMware is addressed in its entirety later in this document.

[15] See "Performance Characteristics of VMFS and RDM".

Raw Device Mapping (RDM)

RDM is a mapping file, stored in a VMFS volume that acts as a proxy for a physical device. The RDM file contains metadata used to manage and redirect disk accesses to the physical device. This technique provides advantages of direct access to physical devices, in addition to some of the advantages of virtual disks on VMFS storage. RDMs can be configured in two ways:

- **Virtual compatibility mode**: This mode fully virtualizes the mapped device, which appears to the guest operating system as a virtual disk file on a VMFS volume. Virtual mode provides such VMFS benefits as advanced file locking for data protection and use of snapshots.
- **Physical compatibility mode**: This mode provides minimal SCSI virtualization of the mapped device. VMkernel passes all SCSI commands to the device, with one exception, thereby exposing all the physical characteristics of the underlying hardware.

Table 4. VMFS and RDM Characteristics

	vMotion	VMware DRS	Snapshots	Storage Maximums	Physically Mountable
VMFS	Yes	Yes	Yes	64TB	No
RDM	Yes	Yes	Yes[16]	2TB	Yes

File System Alignment

TIP 13: Make Sure VMFS is Properly Aligned.

As in the physical world, file system misalignment can severely impact performance. File system misalignment not only manifests itself in databases, but with any high I/O workload. VMware makes the following recommendations for VMware VMFS partitions:

- Like other disk-based file systems, VMFS suffers a penalty when the partition is unaligned. Use VMware vCenter to create VMFS partitions, since it automatically aligns the partitions along the 64KB boundary.
- To manually align your VMware VMFS partitions, first check your storage vendor’s recommendations for the partition starting block (for example, EMC CLARiiON/DMX use 128K offsets).

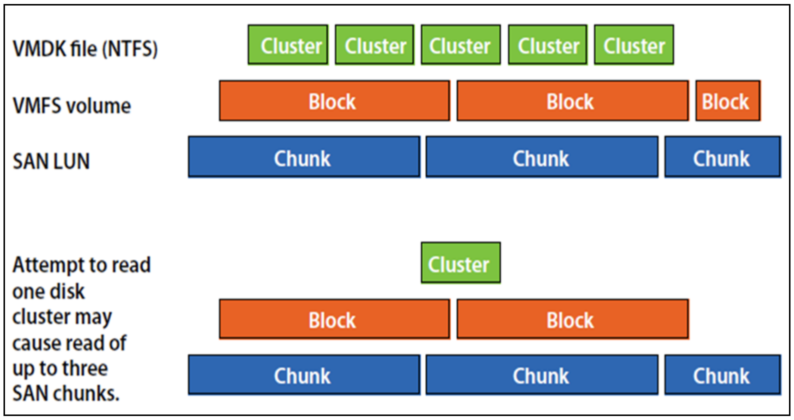

Figure 1. Unaligned Partitions

[16] RDM set to virtual compatibility mode

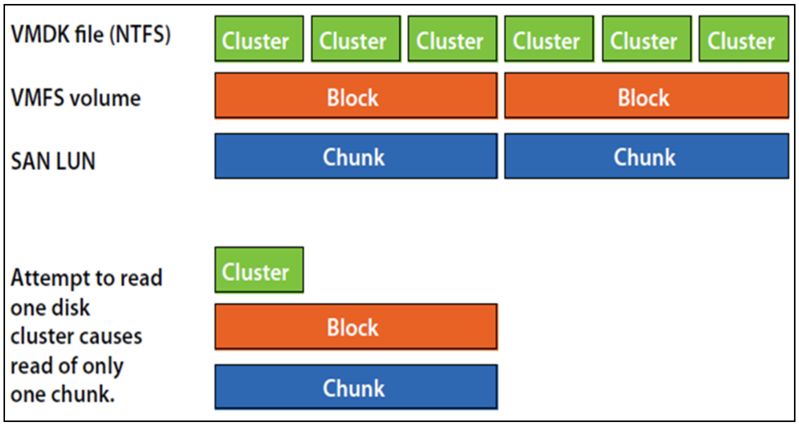

Figure 2. Aligned Partitions

Database Layout Considerations

The Oracle Optimized Flexible Architecture (OFA) is a set of naming standards and best practices to be used when installing and configuring Oracle software. It is a generally accepted best practice to follow the OFA standards for Oracle virtual installations as well. Beginning in 10g, Oracle introduced Automated Storage management, which also conforms to the OFA naming conventions.

TIP 14: Use Oracle Automatic Storage Management.

Automatic Storage Management

Oracle ASM provides integrated clustered file system and volume management capabilities for managing Oracle database files. In addition, ASM simplifies database file creation while delivering near-raw device file system performance.

As mentioned earlier, a vSphere datastore is an abstraction of the storage layer; LUNs can be thought of as abstractions of the disks themselves. For this reason, care must be taken before configuring ASM disk groups.[17] When creating ASM disk groups:

- Create ASM disk groups with equal disk types and geometries. An ASM disk group is essentially a grid of disks and the group performance will be limited by its slowest member.
- Create multiple ASM disk groups based on I/O characteristics. At a minimum, create two ASM disk groups; one for log files, which are sequential in nature; and one for datafiles, which are random in nature.
- If using networked storage, configure the ASM disk groups with external redundancy. Do not use Oracle ASM failure groups. Oracle failure groups consume additional CPU cycles and can operate unpredictably after suffering a disk failure. When using external redundancy, disk failures are transparent to the database and consume no additional database CPU cycles, since this is offloaded to the storage processors.

It is extremely important to understand that ASM is not storage-aware; in other words, whatever disks are provisioned to a DBA can be used to create a disk group. Oracle ASM cannot determine the optimal data placement or LUN selection with respect to the underlying storage infrastructure. For that reason, Oracle ASM is not a substitute for close communication between the storage administrator and the database administrator.

[17] Refer to your Oracle installation guide to create ASM disk groups.

TIP 15: Use Your Storage Vendors Best Practices Documentation when Laying Out the Oracle Database.

EMC and Automatic Storage Management

VMware vSphere can leverage all the inherent performance benefits, best practices, tips, and tricks defined by the individual storage vendors. An excellent example is how EMC storage systems interact with Oracle ASM. One of the benefits of ASM is its ability to provide near-raw device file system performance. (By following EMC best practice documentation a database administrator can achieve better than raw device file system performance.) This is due to the interaction of Symmetrix prefetch algorithms when ASM is configured with fine grain striping.[18]

Oracle Clustered File System (OCFS)

The Oracle Clustered File System is a POSIX-compliant shared disk cluster file system for Linux which can be used with Oracle Real Application Clusters. OCFS was the predecessor to Oracle ASM that was introduced in Oracle 10g. (Real Application Clusters is beyond the scope of the paper). ASM is the recommended clustering technology. Also, since ASM can also be used for single instance deployments, it provides an on-ramp to Real Application Clusters.

Paravirtualized SCSI Adapters

A variety of architectural improvements have been made to the storage subsystem of VMware vSphere 4. The combination of the new paravirtualized SCSI driver (pvscsi), and additional ESX kernel-level storage stack optimizations dramatically improves storage I/O performance.

TIP 16: Use Paravirtualized SCSI Adapters for Oracle Datafiles with Demanding Workloads.

VMware recommends that you create a primary adapter for use with a disk that will host the system software (boot disk) and a separate PVSCSI adapter for the disk that will store the Oracle data files. Oracle databases that drive I/O to their virtual disk in excess of 2000 IOPS will benefit from the presence of the pvscsi driver.

Optimizing Performance

The creation of a fully optimized virtual architecture requires coordination and communication of IT personnel at each level of the hardware/software application stack.

Tip 17: Optimized Architectures are Not Designed in Silos.

At a minimum, designing the optimized architecture should involve the database administrator, storage administrator, network administrator, VMware administrator, and application owner.

[18] See “EMC Symmetic DMX-4 for Oracle 10g and Oracle 11g Data Warehouse Layout: Best Practices Planning”.

Tips Summary

The following table provides a complete listing of the best practice tips described in this paper for deploying Oracle on VMware virtual infrastructure.

Table 5. Summary of Oracle Best Practice Deployment Tips

Number	Tip Description
1	Upgrade to vSphere ESX 4.
2	Create a Computing Environment Optimized for vSphere.
3	Create Golden Images of Optimized Operating Systems using vSphere Cloning Technologies.
4	Use as Few Virtual CPUs (vCPUs) as Possible.
5	Enable Hyper-Threading for Intel Core i7 Processors.
6	Set Memory Reservations Equal to the Size to the Oracle SGA.
7	Allow vSphere to Choose the Best Virtual Machine Monitor based on the CPU and Guest Operating System Combination.
8	Use Large Memory Pages.
9	Use the VMXNET Family of Paravirtualized Network Adapters.
10	For IP-Based Storage iSCSI and NFS, Enable Jumbo Frames.
11	Create Dedicated Datastores to Service Database Workloads.
12	Use vSphere VMFS for Single Instance Oracle Database Deployments.
13	Make Sure VMFS is Properly Aligned.
14	Use Oracle Automatic Storage Management.
15	Use Your Storage Vendors Best Practices Documentation when Laying Out the Oracle Database.
16	Use Paravirtualized SCSI Adapters for Oracle Datafiles with Demanding Workloads.
17	Optimized Architectures are Not Designed in Silos.

Oracle Support for VMware Virtualization

Oracle Support has been forthright about its stance toward VMware virtualization for over four years. Oracle has a support statement for VMware products and it is honored around the world. While there has been much public discussion about Oracle's perceived position on support for VMware virtualization, VMware's experience is that Oracle Support upholds its commitment to customers, including those using VMware virtualization to work in conjunction with Oracle products.

VMware is also an Oracle customer; our E-Business Suite and Siebel instances are virtualized; and VMware routinely submits and receives assistance with issues for Oracle running on VMware virtual infrastructure. The specifics of Oracle's support commitment to VMware is provided by the MyOracleSupport Metalink document ID #249212.1. While prohibited from reproducing the document, we can highlight a few of the facts that Oracle Support has maintained with regard to this statement for the four-plus years of its existence:

- **Known issues**: Oracle Support will accept customer support requests for Oracle products running on VMware virtual infrastructure if the reported problem was already known to Oracle. This is crucial! If you are running 9i, 10g, or other products with a long history, the odds are in your favor that if you find a problem, Oracle has seen it before. If they've already seen it, they will accept it.
- **New issues**: Oracle Support reserves the right to ask customers to prove that "new issues" attributed to Oracle are not a result of an application being virtualized. We say—fair enough—this is essentially the same as every other ISV, to one degree or another. What is key is to look at the history of Oracle Support with regard to "new issues."
  - From the perspective of MyOracleSupport, in four-plus years of tracking VMware-related issues, you will find essentially no bugs attributed to VMware ESX or vSphere with Oracle.
  - VMware routinely asks reference customers from around the world to share their success stories regarding submission of issues to Oracle and getting appropriate and helpful responses. As noted above, Oracle Support routinely provides support to customers running on VMware virtual infrastructure world-wide.
- **Oracle RAC**: Oracle RAC today is "expressly not supported." Oracle has made it clear that while it is legal to run RAC on VMware infrastructure, they are under no obligation to support it. Yet, even still, we have many customers running RAC on VMware infrastructure; it works fine, and even these customers have been able to submit issues to Oracle for resolution.
- **Certification**: VMware vSphere is a technology that lives under the certified Oracle stack (unlike other virtualization technologies that alter OS and other elements of the stack). As a result, Oracle cannot certify VMware virtual infrastructure. However, VMware is no different in this regard than an x86 server—Oracle also doesn't certify Dell, HP, IBM, or Sun x86 servers.

VMware recommends that customers think logically about Oracle's support position. Test the hypothesis presented above. Begin with pre-production systems; as issues are encountered and SRs are filed, track Oracle's response. VMware's experience is that customers will see no difference in the quality and timeliness of Oracle Support's response.

References

Performance Papers

- Performance Best Practices for VMware vSphere 4.0
- Virtualizing Performance Critical Database Applications in VMware vSphere
- Comparison of Storage Protocol Performance in VMware vSphere 4
- Performance Characterization of VMFS and RDM Using a SAN
- VMware Large Page Performance
- Recommendations for Aligning VMFS Partitions
- Dynamic Storage Provisioning: Considerations and Best Practices for Using Virtual Disk Thin Provisioning

Storage Configuration and Protocols

- iSCSI SAN Configuration Guide for vSphere
- Fibre Channel SAN Configuration Guide for vSphere
- Using VMware vSphere with EMC Symmetrix Storage
- vSphere iSCSI SAN Configuration Guide
- Introduction to Using EMC Celerra with VMware vSphere 4
- NetApp and VMware vSphere Storage Best Practices
- EMC Symmetrix DMX-4 for Oracle 10g and Oracle 11g Data Warehouse Layout

VMware Knowledge Base Articles

- Install and Configure Paravirtualized SCSI Adapters

Oracle Database Customer Success Stories

- Bobst Group
- Canada Interior Health Authority
- HeliVolt Corporation
- JanPak

Oracle Automatic Storage Management

- Using Oracle Database 10g Automatic Storage Management with EMC Storage Technologies

vSphere Compatibility Guide

- Web Based Compatibility Guide

Multi-Media

- YouTube: VMware: IP Storage-iSCSI, NFS, or Fibre Channel?
- YouTube: VMware Distributed Resource Scheduling (DRS) with Oracle demo
- YouTube: Hot adding a CPU to an Oracle database in VMware
- YouTube: VMware on NetApp

Books

- "Oracle on VMware: Expert Tips for Database Virtualization"; Dr. Bert Scalzo

About The Author

Bob Goldsand is a Senior Technical Alliance Manager with the VMware ISV Applications Team. He is responsible for the Data Management segment of the VMware marketplace, which includes transactional and analytic databases, Business Intelligence, and data integration technologies. Bob came to VMware from EMC where he was a Senior Technologist reporting to the Corporate Office of the CTO and was also a member of the EMC/Oracle Global Alliance Team.

Acknowledgements

The author would like to thank the following for their invaluable technical contributions:

Scott Drummonds, Kaushik Banerjee, Chris Rimer, Mel Shum, Mike West, Jeff Freeman, David Korsunsky, Chethan Kumar, Jeffrey Buell, Tim Harris, Bert Scalzo.

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