vmware® | **SOLUTION READINESS**

Oracle Databases on VMware
Best Practices Guide

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

Contents

1. Introduction

This *Oracle Databases on VMware Best Practices Guide* provides best practice guidelines for deploying Oracle databases on VMware vSphere. The recommendations in this guide are not specific to any particular set of hardware, or size and scope of any particular Oracle database implementation. The examples and considerations provide guidance, but do not represent strict design requirements.

The successful deployment of Oracle on VMware vSphere™ 4 is not significantly different from deploying Oracle on physical servers. DBAs can fully leverage their current skill set while also delivering the benefits associated with virtualization.

In addition to this guide, VMware has created separate best practice documents for storage, networking and performance. There is also a white paper, *Oracle on VMware vSphere Essential Database Deployment Tips*, information from which has been included in this *Best Practices Guide*. See the, References section (Section 10) for a list of other documents that can help you successfully deploy Oracle on VMware vSphere.

2. VMware vSphere 4

VMware virtualization solutions provide numerous benefits to DBA administrators. VMware virtualization creates a layer of abstraction between the resources required by an application and operating system, and the underlying hardware that provides those resources. This abstraction layer provides value for:

- Consolidation – VMware technology allows multiple application servers to be consolidated onto one physical server, with little or no decrease in overall performance.

- Ease of Provisioning – VMware virtualization encapsulates an application into an image that can be duplicated or moved, greatly reducing the cost of application provisioning and deployment.

- Manageability – Virtual machines may be moved from server to server with no downtime using VMware vMotion™, which simplifies common operations such as hardware maintenance and reduces planned downtime.

- Availability – If an unplanned hardware failure occurs, VMware High Availability (HA) restarts affected virtual machines on another host in a VMware cluster. With HA you can reduce unplanned downtime and provide higher service levels to an application. VMware Fault Tolerance (FT) features zero downtime and zero data loss, providing continuous availability in the face of server hardware failures for any application running in a virtual machine.

.

3. Production Support for Oracle Databases on vSphere

Oracle has a support statement for VMware products that is honored around the world. While there has been much public discussion about Oracle's perceived position on support for VMware virtualization, our experience is that Oracle Support upholds its commitment to customers, including those using VMware virtualization 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 are provided by the MyOracleSupport Metalink document ID #249212.1. Gartner, IDC, and others also have documents for their subscribers that specifically address this policy. Though prohibited from reproducing this document here, the following highlights some of the key facts about Oracle Support:

- Oracle RAC support is now included for 11.2.0.2 and above.

- Known issues – Oracle Support will accept customer support requests for Oracle products running on VMware virtual infrastructure if the reported problem is 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 Oracle has seen your problem 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 as this is essentially the same policy that other ISVs use to some degree. It is key to look at the history of Oracle Support with regard to new issues.

- Certification – VMware vSphere is a technology that lives under the certified Oracle stack (unlike other virtualization technologies that alter the OS and other elements of the stack). As a result, Oracle cannot certify VMware virtual infrastructure. However, VMware is no different in this regard from an x86 server—Oracle doesn't certify Dell, HP, IBM, or Sun x86 servers.

VMware recommends that customers take a logical approach and test Oracle's support statement. Begin with pre-production systems, and as issues are encountered and SRs are filed, track Oracle's response. Our experience is that customers see no difference in the quality and timeliness of Oracle Support's response.

4.   ESX Host Guidelines

4.1   General

The following are general best practices for host systems.

Item	Comments
Recommendation	Create a computing environment optimized for vSphere.
Justification	ESX host BIOS settings can be specifically adjusted to maximize compute resources (such as disabling unnecessary processes and peripherals) to Oracle databases.

Item	Comments
Recommendation	Create golden images of optimized operating systems using vSphere cloning technologies.
Justification	After the operating system has been prepared with the appropriate patches and kernel settings Oracle can be installed in a virtual machine the same way it is installed on a physical system. This speeds up the installation of a new database.

Item	Comments
Recommendation	Upgrade to vSphere ESX 4.
Justification	VMware and database administrators can realize a 10- to 20-percent performance boost after upgrading to the latest vSphere release from prior 3.X versions.

Item	Comments
Recommendation	Allow vSphere to choose the best virtual machine monitor based on the CPU and guest operating system combination.
Justification	Make sure the virtual machine setting has Automatic selected for the CPU/MMU Virtualization option.

4.1.1 BIOS Settings

x-86 server BIOS settings can be set to disable unnecessary processes and peripherals to maximize performance. Table 1 describes the optimized settings.

Table 1. BIOS Settings Maximized for Performance

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	Disables 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 saving power .
Power-Saving	No	Disable if performance is more important than saving power .
Virus Warning	No	Disables warning messages when writing to the master boot record.
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 and	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.

4.1.2 Operating System Host Processes

VMware recommends disabling unnecessary foreground and background processes within the guest operating system.

- Examples of unnecessary Linux processes are: anacron, apmd, atd, autofs, cups, cupsconfig, gpm, isdn, iptables, kudzu, netfs, and portmap.

- 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.

- For Linux installs, the database administrator (DBA) should request that the system administrator compile a monolithic kernel that will only load the necessary features. Whether you intend to run Windows or Linux as the final optimized operating system, these host installs should be cloned by the VMware administrator for reuse.

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

4.1.3 Upgrade to vSphere

vSphere includes numerous performance and scalability enhancements that provide a 10- to 20-percent performance boost compared to previous 3.X versions. The following table summarizes the improvements to the hypervisor including current metrics for vSphere.

Table 2. Performance and Scalability Improvements by ESX Version

	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

VMware vSphere supports large capacity virtual machines, so it can support larger sized Oracle databases and SGA footprints. vSphere host and virtual machine specifications are:

- Each VMware ESX™ host supports up to 1TB RAM, 64 logical CPUs and 512 virtual CPUs.

- Each virtual machine can support up to eight virtual CPUs and 255GB of RAM.

4.1.4 Hardware-assisted Memory Management Unit

For best performance VMware recommends using servers with the latest chip generations that support a hardware-assisted Memory Management Unit (MMU). Hardware-assisted MMU refers to hardware support for memory management unit virtualization. Features that provide the sport are available from Intel and AMD and are called EPT and RVI, respectively. Support consists of an additional level of page tables implemented in hardware. These page tables contain guest physical to machine physical memory address translations.

On processors that support it, vSphere by default uses hardware-assisted MMU virtualization for virtual machines. This default behavior is configured in the "virtual machine settings" using the vSphere Client, by setting the **CPU/MMU Virtualization** parameter to **Automatic** (which is default).

4.2 Memory

The following are memory-related best practices.

Item	Comments
Recommendation	Set memory reservations equal to the size to the Oracle SGA.
Justification	To avoid kernel swapping between ESX and the guest OS as Oracle databases can be memory-intensive.

Item	Comments
Recommendation	Use large memory pages.
Justification	Large page support is enabled by default in ESX versions 3.5 and later, and is supported from version 9iR2 for Linux operating systems and version 10gR2 for Windows. Enable large pages in the guest OS to improve the performance of Oracle databases on vSphere.

Appendix A provides a description of virtual machine memory settings that are discussed in this section. For further background on VMware memory management concepts, refer to the vSphere Resource Management Guide.

When consolidating Oracle database instances, vSphere presents the opportunity to share 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 on a physical machine. Transparent page sharing also allows DBAs to overcommit memory without any performance degradation. In production environments, carefully consider the impact of overcommiting memory and only overcommit after collecting data to determine the amount of overcommitment possible. To determine the effectiveness of memory sharing and the degree of acceptable overcommitment for a given database, run the workload, and use resxtop or esxtop to observe the actual savings.

Because Oracle databases can be memory-intensive, and to account for situations where performance is a key factor (and avoid kernel swapping between ESX and the guest OS in mission critical production environments), VMware recommends:

- Set the memory reservation equal to the size of the Oracle SGA.

- Where the Oracle database is part of a third-party commercial enterprise application (ERP), follow virtualization guidelines from the ERP vendor.

- Note that setting reservations may limit vMotion. A virtual machine can only be live migrated if the target ESX host has free physical memory equal to or greater than the size of the reservation.

- Do *not* disable the balloon driver.

- The guest operating system within the virtual machine still needs its own separate swap/page file. Follow the same swap space guidelines given for physical environments.

Though VMware recommends setting memory reservations equal to the size of the Oracle SGA in production environments, it is acceptable overcommit more aggressively in non-production environments such as development, test, or QA. In these environments, a DBA can introduce memory overcommitment to take advantage of VMware memory reclamation features and techniques. Even in these environments, the type and number of databases that can be deployed using overcommitment largely depend on their actual workload.

4.2.1 Large Pages

vSphere supports large pages in the guest operating system. The use of large pages results in reduced memory management overhead and can increase hypervisor performance. Oracle supports the use of large memory pages in version 9iR2 for Linux operating systems and in version 10gR2 for Windows. The following Metalink Notes are relevant when setting large pages:

- 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*

- Note 46001.1 – *Oracle Database and the Windows NT memory architecture, Technical Bulletin*

- Note 46053.1 – *Windows NT Memory Architecture Overview*

4.3 Virtual CPU

The following are virtual CPU-related best practices.

Item	Comments
Recommendation	Use as few virtual CPUs (vCPUs) as possible.
Justification	If monitoring of the actual workload shows that the Oracle database is not benefitting from the increased virtual CPUs, the excess vCPUs imposes scheduling constraints and can degrade overall performance of the virtual machine.

Item	Comments
Recommendation	Enable hyper-threading for Intel Core i7 processors.
Justification	With the release of Intel Xeon 5500 series processors, enabling Hyper-threading is recommended.

VMware uses the terms *virtual CPU* (vCPU) and *physical CPU* to distinguish between the processors within the virtual machine and the underlying physical x86-based processors. Virtual machines with more than one virtual CPU are also called SMP (symmetric multi-processing) virtual machines.

VMware Virtual Symmetric Multi-Processing (Virtual SMP) enhances virtual machine performance by enabling a single virtual machine to use multiple physical processors simultaneously. vSphere supports use of up to eight virtual CPUs per virtual machine. The biggest advantage of an SMP system is the ability to use multiple processors to execute multiple tasks concurrently, thereby increasing throughput (for example, the number of transactions per second). Only workloads that support parallelization (including multiple processes or multiple threads that can run in parallel) can really benefit from SMP. The Oracle architecture is multi-threaded and includes multiple processes which makes it a good candidate to take advantage of Virtual SMP.

In ESX 4, the CPU scheduler has undergone several improvements to provide better performance and scalability; for details, see *VMware vSphere: The CPU Scheduler in VMware ESX 4.1*. Though larger virtual machines are possible in vSphere, VMware recommends reducing the number of virtual CPUs if monitoring of the actual workload shows that the Oracle database is not benefitting from the increased number of virtual CPUs. For more background, please see "ESX CPU Considerations" in *Performance Best Practices for VMware vSphere 4*.

Hyper-threading technology enables a single physical processor core to behave like two logical processors, essentially allowing two independent threads to run simultaneously. Unlike having twice as many processor cores—which 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. For example, an ESX host system enabled for hyper-threading on an 8-core server sees 16 threads that appear as 16 logical 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 because the measured performance results were not consistent across applications, run environments, or database workloads.

VMware recommends the following practices for allocating CPU to Oracle database virtual machines:

- Start with a thorough understanding of your workload. Database server utilization varies widely by application. If the application is commercial, be sure to follow published guidelines where appropriate. If the application is custom-written, work with the application developers to determine resource requirements. VMware Capacity Planner can analyze your current environment and provide resource utilization metrics that can aid in the sizing process.

- If the exact workload is not known, start with fewer virtual CPUs and increase the number later if necessary. Only allocate multiple vCPUs to a virtual machine if the anticipated database workload can truly take advantage of all the vCPUs.

- When consolidating multiple virtual machines on single ESX host, proper hardware sizing is critical for optimal performance. Make sure that cumulative physical CPU resources on a host are adequate to meet the needs of the VMs by testing your workload in the planned virtualized environment. CPU overcommitment should be based upon actual performance data to avoid adversely affecting VM performance.

5. Storage Guidelines

The following are storage-related best practices.

Item	Comments
Recommendation	For IP-based storage iSCSI and NFS, enable jumbo frames.
Justification	Jumbo frames enable Ethernet frames to have larger "payload," allowing for improved performance.

Item	Comments
Recommendation	Create dedicated datastores to service database workloads.
Justification	The creation of dedicated datastores for I/O-intensive databases is analogous to provisioning dedicated LUNs in the physical world. This is a typical design for a mission-critical enterprise workload.

Item	Comments
Recommendation	Use vSphere VMFS for single instance Oracle database deployments.
Justification	To balance performance and manageability in a virtual environment, deploy Oracle using VMFS.

Item	Comments
Recommendation	Make sure VMFS is properly aligned.
Justification	Like other disk-based file systems, VMFS suffers a penalty when the partition is unaligned. Use VMware vCenter to create VMFS partitions because it automatically aligns the partitions.

Item	Comments
Recommendation	Use Oracle automatic storage management.
Justification	Oracle ASM provides integrated clustered file system and volume management capabilities for managing Oracle database files. ASM simplifies database file creation while delivering near-raw device file system performance.

Item	Comments
Recommendation	Use your storage vendors best practices documentation when laying out the Oracle database.
Justification	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.

Item	Comments
Recommendation	Avoid silos when designing the storage architecture.
Justification	At a minimum, designing the optimized architecture should involve the database administrator, storage administrator, network administrator, VMware administrator, and application owner.

Item	Comments
Recommendation	Use paravirtualized SCSI adapters for Oracle datafiles with demanding workloads.
Justification	The combination of the new paravirtualized SCSI driver (pvscsi) and additional ESX kernel-level storage stack optimizations dramatically improves storage I/O performance.

Storage configuration is essential for any successful database deployment, especially in virtual environments where you may consolidate many different Oracle database workloads on a single ESX host. Your storage subsystem should provide sufficient I/O throughput as well as storage capacity to accommodate the cumulative needs of all virtual machines running on your ESX hosts.

5.1   Storage Virtualization Concepts

VMware storage virtualization can be categorized into three layers of storage technology.

- The storage array is the bottom layer, consisting of physical disks presented as logical disks (storage array volumes or LUNs) to the layer above.

- The next layer is the virtual environment occupied by vSphere. Storage array LUNs are presented to ESX servers as datastores and are formatted as VMFS volumes.

- Virtual machines consist of virtual disks that are created in the datastores and presented to the guest operating system as disks that can be partitioned and used in file systems.

5.1.1 VMware Virtual Machine File System (VMFS)

VMFS is a cluster file system that provides storage virtualization optimized for virtual machines. Each virtual machine is encapsulated in a set of files and VMFS is the default storage system for these files on physical SCSI disks and partitions. 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.

5.1.2 Raw Device Mapping

VMware also supports Raw Device Mapping (RDM). RDM allows a virtual machine to directly access a volume on the physical storage subsystem, and can only be used with Fibre Channel or iSCSI. RDM can be thought of as providing a symbolic link from a VMFS volume to a raw volume. The mapping makes volumes appear as files in a VMFS volume. The mapping file, not the raw volume, is referenced in the virtual machine configuration.

5.2   Storage Protocol Capabilities

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 are, 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. During its testing, VMware has found that wire speed is the limiting factor for I/O throughput when comparing the storage protocols. VMware ESX can reach the link speeds in single VM environment, and also maintain the throughput up to 32 concurrent virtual machines for each storage connection option supported. For details, refer to the *Comparison of Storage Protocol Performance in VMware vSphere 4* white paper. Fibre Channel may provide maximum I/O throughput, but iSCSI and NFS may offer a better price-performance ratio.

When selecting networked storage systems and protocols, it is critical to understand which vSphere features are supported. The table below 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	SRM
Fibre Channel	Yes	Yes	Yes	Yes	Yes	Yes
iSCSI	Yes	Yes	Yes	Yes	Yes	Yes
NAS	Yes	No	Yes	No	No	No
Local Storage	Yes	Yes	No	Yes	No	No

Jumbo frames are recommended for IP based storage iSCSI and NFS. 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.

5.3 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.

5.3.1 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.

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. 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 is 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. Refer to your Oracle installation guide to create ASM disk groups.

5.3.2 Oracle Clustered File System (OCFS)

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

5.3.3 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 log or database files), without isolating the physical storage layer as well, does not have the desired effect on performance.

5.3.3.1. Example of Oracle Database Storage Layout on VMware vSphere

For mission-critical databases it is common practice in physical environments to spread the database over multiple LUNs to maximize I/O performance (for example placing log and datafiles in separate LUNs). Similar guidelines should be followed when virtualized. An example layout is shown in the following figure.

Figure 1. Example Storage Layout of Oracle OLTP Database on VMware
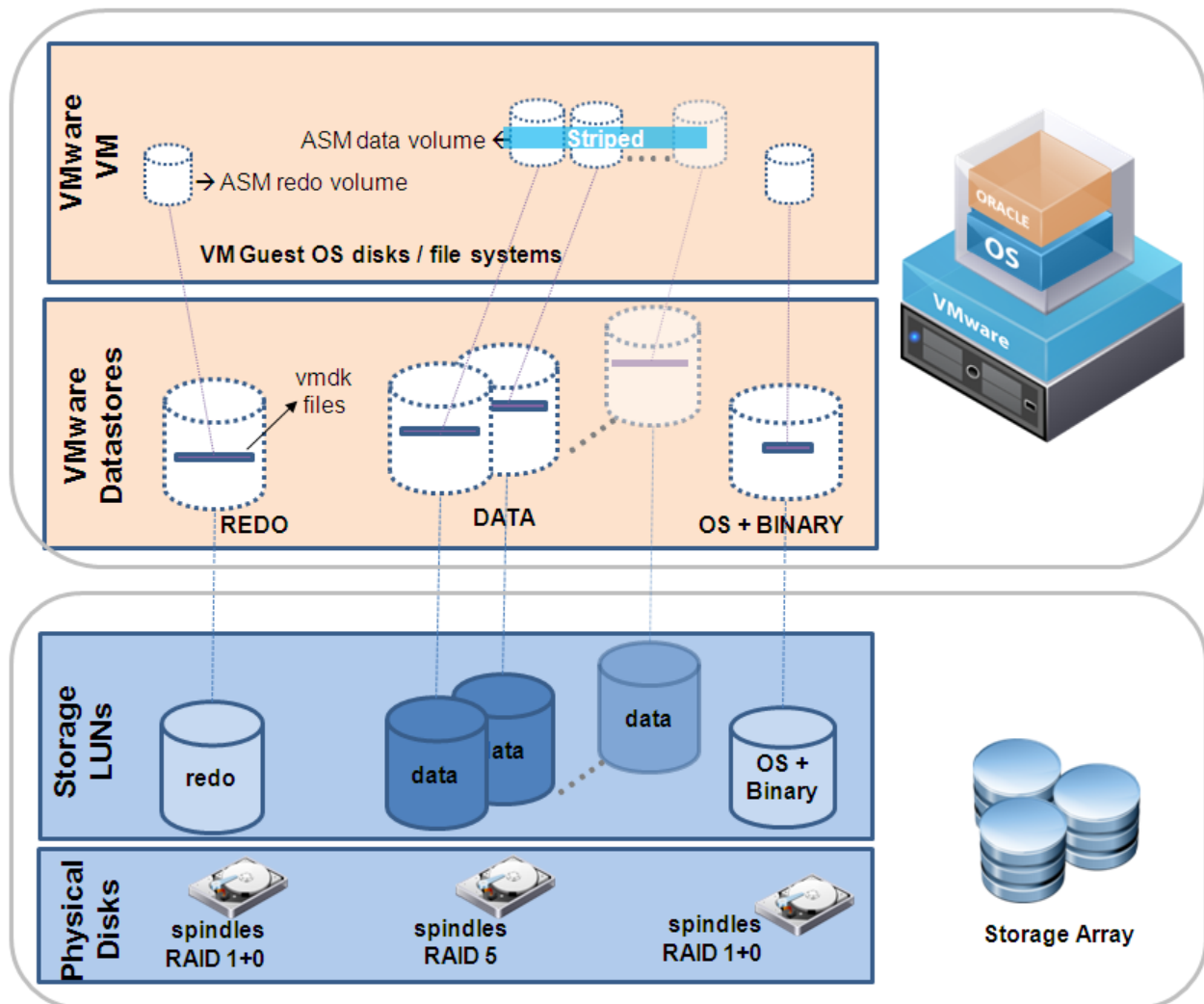

Figure 1 represents an example storage design for a virtualized Oracle OLTP database. The design is based on the following principles:

- At a minimum an optimized architecture requires joint collaboration between the database, VMware and storage administrators.

- Follow storage vendor best practices for database layout on their arrays (as is done in the physical world).

Note that Figure 1 is only an example and actual configurations for customer deployments may differ.

5.3.4 Virtual SCSI Controllers

We highly recommend using multiple virtual SCSI Controllers for the database virtual machines or virtual machines with high I/O load. The use of multiple virtual SCSI Controllers allows the execution of several parallel I/O operations inside the Guest OS. We also highly recommend separating the Redo/Log I/O traffic from the Datafile I/O traffic through separate virtual SCSI Controllers. As a Best Practice, you can use one Controller for OS and Swap, one Controller for DB Log, and one or more Controllers for DB Datafiles (depending on the amount and size of DB Datafiles). Please refer to the VMware Administration Guide on how to add additional virtual SCSI Controllers

5.4 VMFS versus RDM

5.4.1 Performance

VMware is often asked which offers better performance, VMFS or RDM? Both VMFS and RDM volumes can provide similar transaction throughput. For more details, refer to the *Performance Characterization of VMFS and RDM Using a SAN*.

5.4.2 Functionality

VMware generally recommends VMFS, but there may be situations where RDMs are required. Table 4 summarizes some of the options and trade-offs between VMFS and RDM. For a more complete discussion, see the *VMware SAN System Design and Deployment Guide*.

Table 4. VMFS and Raw Disk Mapping Trade-offs

VMFS	RDM
• Volume can host many virtual machines (or can be dedicated to one virtual machine).	• Maps a single LUN to one virtual machine so only one virtual machine is possible per LUN.
• Increases storage utilization, provides better flexibility, easier administration and management.	• More LUNs are required, so it is easier to reach the LUN limit of 256 that can be presented to an ESX host.
• Can potentially support clustering software that does not issue SCSI reservations for example: Oracle Clusterware. To configure follow VMware KB: 1034165	• RDM may be required to leverage third party storage array-level backup and replication tools
• Oracle RAC node Live Migration.	• RDM volumes can help facilitate migrating physical oracle databases to virtual machines. Alternatively enables quick migration to physical in rare Oracle support cases.
	• Required for MSCS quorum disks.

5.5   General

5.5.1 Partition Alignment

Aligning file system partitions is a well-known storage best practice for database workloads. Partition alignment on both physical machines and VMware VMFS partitions prevents performance I/O degradation caused by I/O crossing track boundaries. VMware test results show that aligning VMFS partitions to 64KB track boundaries results in reduced latency and increased throughput. VMFS partitions created using vCenter are aligned on 64KB boundaries as recommended by storage and operating system vendors.

It is considered a best practice to:

- Create VMFS partitions from within vCenter. They are aligned by default.

- Align the data disk for heavy IO workloads using `diskpart`.

- Consult with the storage vendor for alignment recommendations on their hardware.

For more information about this topic see the white paper entitled *Performance Best Practices for VMware vSphere 4.0*.

5.5.2 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.

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.

Results of tests conclude that PVSCSI is not recommended for virtual machines performing less than 2,000 IOPS and issuing greater than four outstanding I/Os. This issue is fixed in vSphere 4.1, so that the PVSCSI virtual adapter can be used with good performance, even under this condition.

Follow guidelines in the following VMware KB articles:

- 1010398 – C*onfiguring disks to use VMware Paravirtual SCSI (PVSCSI) adapters*

- 1017652 – *Do I choose the PVSCSI or LSI Logic virtual adapter on ESX 4.0 for non-IO intensive workloads?*

6. Networking Guidelines

The following are networking-related best practices.

Item	Comments
Recommendation	Use the VMXNET family of paravirtualized network adapters.
Justification	The paravirtualized network adapters in the VMXNET family implement an optimized network interface that passes network traffic between the virtual machine and the physical network interface cards with minimal overhead.

Item	Comments
Recommendation	Separate infrastructure traffic from VM traffic for security and isolation.
Justification	Virtual machines should not see infrastructure traffic (security violation) and should not be impacted by infrastructure traffic bursts (e.g. vMotion).

Item	Comments
Recommendation	Use NIC teaming for availability and load balancing.
Justification	NIC teams can share the load of traffic among some or all of its members, or provide passive failover in the event of a hardware failure or a network outage.

Item	Comments
Recommendation	Take advantage of Net I/O Control to converge network and storage traffic onto 10GE.
Justification	This can reduce cabling requirements, simplify management and reduce cost.

The standard VMware networking best practices apply to running Oracle databases on vSphere. For further details follow the networking design guidelines in VMworld 2010 session TA859 – *Virtual Networking Concepts and Best Practices*. This includes designs to efficiently manage multiple networks and redundancy of network adaptors on ESX hosts. The key best practice guidelines are:

- Separate infrastructure traffic from VM traffic for security and isolation.

- Use NIC teaming for availability and load balancing. NIC teaming occurs when multiple uplink adapters are associated with a single vSwitch to form a team.

- Take advantage of Net I/O Control to converge network and storage traffic onto 10GE. Net IO Control was, released in vSphere 4.1 and enables you to guarantee service levels (bandwidth) for particular vSphere traffic types: VM traffic; FT logging; iSCSI ; NFS ; management; vMotion.

- In vSphere use the VMXNET3 network adapter. This is a paravirtualized device that works only if VMware Tools is installed on the guest operating system. The VMXNET3 adapter is optimized for virtual environments and designed to provide high performance. For further background on network adaptors and compatibility with ESX release and supported guest OS, see the VMware KB article 1001805 – *Choosing a network adapter for your virtual machine*.

- For RAC interconnect use Jumbo frames – To enable jumbo frames follow VMware KB: 103712

7. Performance Monitoring on vSphere

The following is a performance monitoring best practice.

Item	Comments
Recommendation	Use VMware vCenter and/or the esxtop/resxtop utility for performance monitoring in the virtual environment.
Justification	Guest OS counters can be used to get a rough idea of performance within the virtual machine but for example CPU and memory usage reported within the guest OS can be different from what ESX reports.

Always use the VI Client or vSphere Client, esxtop, or resxtop to measure resource utilization. CPU and memory usage reported within the guest OS can be different from what ESX reports.

Oracle DBA administrators should pay close attention to the following counters (see VMware Communities: Interpreting esxtop Statistics for a full list of counters).

Table 5. ESX Performance Counters

Subsystem	esxtop Counters	vCenter Counter
CPU	%RDY	Ready (milliseconds in a 20,000 ms window)
	%USED	Usage
Memory	%ACTV	Active
	SWW/s	Swapin Rate
	SWR/s	Swapout Rate
Storage	ACTV	Commands
	DAVG/cmd	deviceWriteLatency & deviceReadLatency
	KAVG/cmd	kernelWriteLatency & kernelReadLatency
Network	MbRX/s	packetsRx
	MbTX/s	packetsTx

- The table above lists a few key counters that should be added to the list of inspection points for Oracle DBA administrators.

- Of the CPU counters:
  - The total used time indicates system load.
  - Ready time indicates overloaded CPU resources.

- A significant swap rate in the memory counters is a clear indication of a shortage of ESX memory, and high device latencies in the storage section point to an overloaded or misconfigured array.

- Network traffic is not frequently the cause of most database performance problems except when large amounts of iSCSI storage traffic are using a single network line. Check total throughput on the NICs to see if the network is saturated.

- To determine if there is any swapping within the guest OS use in the in-guest counters in the same manner as in physical environments.

8. Timekeeping in Virtual Machines

The following is a timekeeping best practice for virtual machines.

Item	Comments
Recommendation	To minimize time drift in virtual machines follow guidelines in KB articles 1006427(Linux) and 1318(Windows)
Justification	The impact of high timer-interrupts in some operating systems can lead to time synchronization errors.

Most operating systems track the passage of time by configuring the underlying hardware to provide periodic interrupts. The rate at which those interrupts are configured to arrive varies for different operating systems. High timer-interrupt rates can incur overhead that affects a virtual machine's performance. The amount of overhead increases with the number of vCPUs assigned to a virtual machine. The impact of these high timer-interrupts can lead to time synchronization errors.

To address timekeeping issues when running Oracle databases please follow the guidelines in the following VMware KB articles:

- KB 1006427 – *Timekeeping best practices for Linux guests* (http://kb.vmware.com/kb/1006427)

- KB 1318 – *Timekeeping best practices for Windows* (http://kb.vmware.com/kb/1318)

9. Summary

All of the best practices and guidelines discussed in this paper are listed in this section.

Recommendations
Create a computing environment optimized for vSphere.
Create golden images of optimized operating systems using vSphere cloning technologies.
Upgrade to vSphere ESX 4.
Allow vSphere to choose the best virtual machine monitor based on the CPU and guest operating system combination.
Use as few virtual CPUs (vCPUs) as possible.
Enable hyper-threading for Intel Core i7 processors.
For IP-based storage iSCSI and NFS, enable jumbo frames.
Create dedicated datastores to service database workloads.
Use vSphere VMFS for single instance Oracle database deployments.
Make sure VMFS is properly aligned.
Use Oracle automatic storage management.
Use your storage vendors best practices documentation when laying out the Oracle database.
Avoid silos when designing the storage architecture.
Use paravirtualized SCSI adapters for Oracle datafiles with demanding workloads.
Use the VMXNET Family of paravirtualized network adapters.
Separate infrastructure traffic from VM traffic for security and isolation.
Use NIC teaming for availability and load balancing.
Take advantage of Net I/O Control to converge network and storage traffic onto 10GE.
Use VMware vCenter and/or the esxtop/resxtop utility for performance monitoring in the virtual environment.

Success stories are available at http://vmware.com/solutions/partners/alliances/oracle-database-customers.html .

10. References

You can find more information about using VMware and Oracle via the links listed below.

- *Oracle on VMware vSphere Essential Database Deployment Tips:*

  http://www.vmware.com/files/pdf/Oracle_Databases_on_vSphere_Deployment_Tips.pdf

- *Virtualizing Performance-Critical Database Applications in VMware vSphere:*

  http://www.vmware.com/pdf/Perf_ESX40_Oracle-eval.pdf

- *Performance Best Practices for VMware vSphere 4:*

  http://www.vmware.com/pdf/Perf_Best_Practices_vSphere4.0.pdf

- *VMware Systems Compatibility Guide:*

  http://www.vmware.com/resources/compatibility/search.php

- *vSphere Resource Management Guide:*

http://www.vmware.com/pdf/vsphere4/r41/vsp_41_resource_mgmt.pdf

- *Understanding Memory Resource Management in VMware ESX 4.1*

  http://www.vmware.com/files/pdf/techpaper/vsp_41_perf_memory_mgmt.pdf

- *VMware vSphere: The CPU Scheduler in VMware ESX 4.1:*

  http://www.vmware.com/files/pdf/techpaper/VMW_vSphere41_cpu_schedule_ESX.pdf

- *VMware Communities: Interpreting esxtop Statistics:*

  http://communities.vmware.com/docs/DOC-9279

- *VMworld 2010 session TA859 Virtual Networking Concepts and Best Practices:*

  http://www.vmworld.com/docs/DOC-5122 (VMworld account required)

- *Comparison of Storage Protocol Performance in VMware vSphere 4:*

  http://www.vmware.com/files/pdf/perf_vsphere_storage_protocols.pdf

- *Performance Characterization of VMFS and RDM Using a SAN:*

  http://www.vmware.com/files/pdf/performance_char_vmfs_rdm.pdf

- *VMware SAN System Design and Deployment Guide:*

  http://www.vmware.com/pdf/vsp_4_san_design_deploy.pdf

- *Disabling simultaneous write protection provided by VMFS using the multi-writer flag:*

  http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1034165

Appendix A. Virtual Machine Memory Settings

The following figure illustrates the memory settings used for a virtual machine.

Figure 2. Virtual Machine Memory Settings.
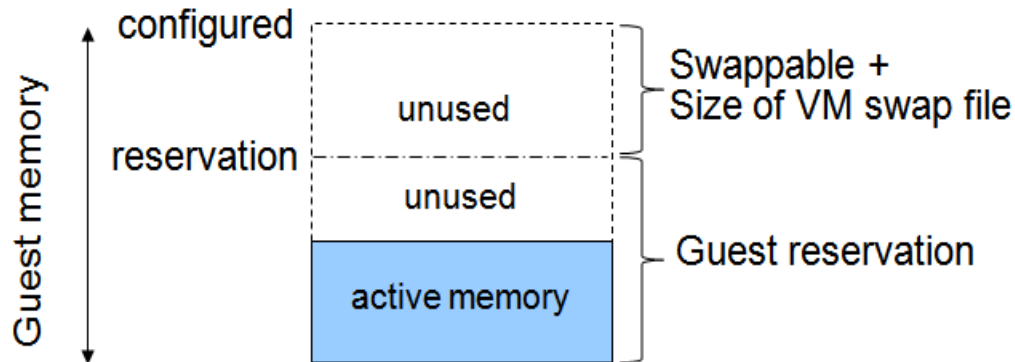
Definition of terms:

- Configured memory – Memory size of virtual machine assigned at creation.

- Active memory – Memory recently accessed by applications in the virtual machine.

- Reservation – Guaranteed lower bound on the amount of memory that the host reserves for the virtual machine, which cannot be reclaimed by VMware ESX for other virtual machines.

- Swappable – Virtual machine memory that can be reclaimed by the balloon driver or, worst case, by ESX swapping. This is the automatic size of the per-virtual-machine swap file that is created on the VMFS file system (.vswp file).

For more information about VMware ESX memory management concepts and the balloon driver, see the VMware Resource Management Guide.