VMware Hybrid Cloud Best Practices Guide for Oracle Workloads

Version 1.0

May 2016

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

Contents

List of Figures

List of Tables

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 vSphere 5.x/6.0 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.

This document also includes information from two white papers, *Performance Best Practices for VMware vSphere 5.5* (https://www.vmware.com/pdf/Perf_Best_Practices_vSphere5.5.pdf) and *Performance Best Practices for VMware vSphere 6.0* (http://www.vmware.com/files/pdf/techpaper/VMware-PerfBest-Practices-vSphere6-0.pdf)

See Section 17, References for a list of other documents that can help you successfully deploy Oracle on vSphere.

2. vSphere

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 the following:

- 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 can be moved from server to server with no downtime using VMware vSphere® vMotion®, which simplifies common operations such as hardware maintenance, and reduces planned downtime.
- Availability – If an unplanned hardware failure occurs, VMware vSphere High Availability (HA) restarts affected virtual machines on another host in a VMware cluster. With VMware HA you can reduce unplanned downtime and provide higher service levels to an application. VMware vSphere 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. VMware Support for Oracle Databases on vSphere

Oracle has a support statement for VMware products (MyOracleSupport 249212.1). While there has been much public discussion about Oracle's perceived position on support for VMware virtualization, experience shows that Oracle Support upholds its commitment to customers, including those using VMware virtualization in conjunction with Oracle products.

VMware is also an Oracle customer. The E-Business Suite and Siebel implementations of VMware IT are virtualized. VMware routinely submits and receives assistance with issues for Oracle running on VMware virtual infrastructure. The MyOracleSupport (MetaLink) Document ID 249212.1 provides the specifics of Oracle's support commitment to VMware. Gartner, IDC, and others also have documents available to their subscribers that specifically address this policy.

The *Understanding Oracle Certification, Support and Licensing for VMware Environments* guide available on the Business Critical Applications landing page (http://www.vmware.com/business-critical-apps/oracle-virtualization/index) provides information pertaining to both Oracle support and Oracle licensing on VMware.

3.1 VMware Oracle Support Policy

The following are some of the key facts about Oracle Support:

- Oracle RAC support is included for the RDBMS version 11.2.0.2 and later (Metalink 249212.1Updated Nov 8, 2010).
- 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 Database 9i, 10g, 11g or another product with a long history, it is likely that Oracle has addressed the presented problem before. If they have 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. This is reasonable, as this is essentially the same policy which is common throughout the industry and applies to all hardware platforms, virtualized or non-virtualized. It is key to look at the history of Oracle Support with regard to new issues. It is important to run all Oracle implementations on a certified operating system (OS).
- Certification –vSphere is a technology that resides under the certified Oracle stack (unlike other virtualization technologies that alter the OS and other elements of the stack). As a result, Oracle does not separately certify VMware virtual infrastructure. However, VMware is no different in this regard from an x86 server—Oracle does not certify Dell, HP, IBM, or Sun x86 servers.
- Remember, Oracle does not certify any third-party infrastructure elements below the operating system. Similarly, in a virtualized environment, Oracle officially certifies only its own Oracle VM platform, but customers should not allow this to limit their options. As in the physical world, customers should choose the virtualization platform that best meets their needs. As long as customers run a certified operating system on VMware, and that operating system is certified by Oracle, customers should feel confident that the most rigorous testing has occurred across the different layers of the stack.

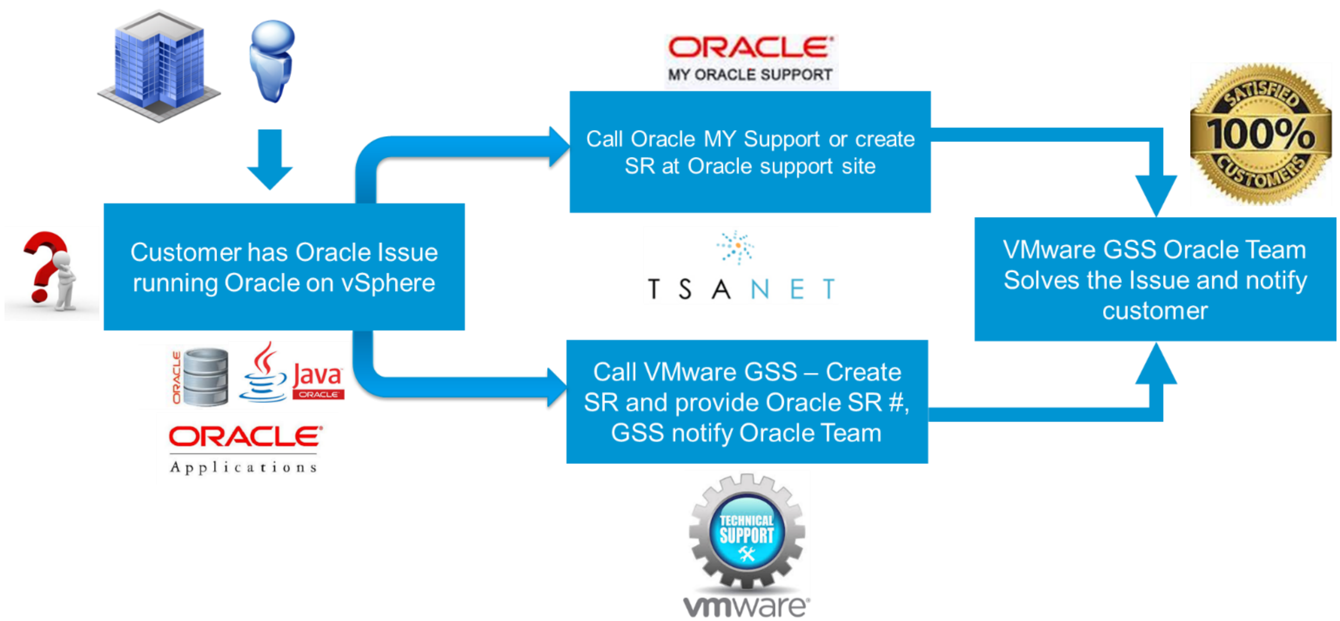
VMware recommends that customers take a logical approach and test Oracle's support statement. Begin with pre-production systems, and as issues are encountered and Service Requests (SRs) are filed, track Oracle's response. Experience shows that customers see no difference in the quality and timeliness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

3.2 VMware Oracle Support Process

VMware support will accept tickets for any Oracle-related issue reported by a customer and will help drive the issue to resolution. To augment Oracle's support document, VMware also has a total ownership policy for customers with Oracle issues as described in the letter at www.vmware.com/files/pdf/solutions/oracle/VMware-Oracle-Support-Affirmation.pdf.

By being accountable, VMware Support will drive the issue to resolution regardless of which vendor (VMware, Oracle or other) is responsible for the resolution. In most cases, reported issues can be resolved through configuration changes, bug fixes, or feature enhancements by one of the involved vendors. VMware is committed to its customers' success and supports their choice to run Oracle software in modern, virtualized environments.For further information, see https://www.vmware.com/support/policies/oracle-support.

Figure 1. VMware vSphere Oracle Support Process



4. Server Guidelines

4.1 General Guidelines

The following table lists general best practices for host systems.

Table 1. Host Systems Best Practices

Recommendation	Justification
Create a computing environment optimized for vSphere.	The VMware ESX® or VMware ESXi™ host BIOS settings can be specifically adjusted to maximize compute resources (such as disabling unnecessary processes and peripherals) to Oracle databases.
Create golden images of optimized operating systems using vSphere cloning technologies.	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.
Upgrade to the latest version of ESXi and vSphere.	VMware and database administrators can realize a significant performance boost after upgrading to the latest vSphere release from prior versions.
Allow vSphere to choose the best virtual machine monitor based on the CPU and guest operating system combination.	Confirm that the virtual machine setting has **Automatic** selected for the **CPU/MMU Virtualization** option.
Verify that all hardware in the system is on the hardware compatibility list for the specific version of VMware software you will be running.	Verify that the hardware meets the minimum configuration supported by the VMware software installed.

4.1.1 BIOS Settings

BIOS settings for an x86 server can be set to disable unnecessary processes and peripherals to maximize performance. The following table describes the optimized settings.

Table 2. BIOS Settings Maximized for Performance

BIOS Setting	Recommendations	Description
Virtualization Technology	Yes	Necessary to run 64-bit guest operating systems.
VT-x, AMD-V, EPT, RVI, VT-d, AMD-Vi	Yes	Hardware-based virtualization support.
Node Interleaving	No	Disables NUMA benefits if disabled.
Turbo Mode	Yes	Balanced workload over unused cores.
Hyperthreading	Yes	For use with some modern Intel processors. Hyperthreading is always recommended with Intel’s newer Core i7 processors such as the Xeon 5500 series.
Power Saving	No	Disable if performance is more important than saving power.
Power Management	OS Controlled Mode	Allow ESXi to control CPU power-saving features.
Power Management Policy	High Performance	Prefer Performance over Power savings.
C-States	No	Disable Processor C-states for performance.
C1E Halt State	No	Disable for performance.
Virus Warning	No	Disables warning messages when writing to the master boot record.
Wake On LAN	Yes	Required for VMware vSphere Distributed Power Management™ feature.
Execute Disable	Yes	Required for VMware vSphere vMotion® and VMware vSphere Distributed Resource Scheduler™ (DRS) features.
Video BIOS Cacheable	No	Not necessary for database virtual machine.
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.

BIOS Setting	Recommendations	Description
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.

4.2 Hardware Assisted Virtualization

Most processors from both Intel and AMD include hardware features to assist virtualization and improve performance. These features include hardware-assisted CPU virtualization, MMU virtualization, and I/O MMU virtualization:

- Hardware-assisted CPU virtualization (HV) which:
  - Includes CPU virtualization (Intel VT-x and AMD-V)
  - Automatically traps sensitive events and instructions, eliminating the software overhead of monitoring all supervisory level code for sensitive instructions
  - Gives the option of using either HV or binary translation (BT) to VM (HV outperforms BT for most workloads)
- Hardware-assisted Memory Management Unit (MMU) virtualization includes the following:
  - Intel Extended Page Tables (EPT) and AMD Rapid Virtualization Indexing (RVI)
  - Intel EPT and AMD RVI / Nested Page Tables (NPT) address the overheads due to MMU virtualization by providing hardware support to virtualize the MMU
  - 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 VMware vSphere Client™ by setting the **CPU/MMU Virtualization** parameter to **Automatic**.
- Hardware-assisted I/O MMU virtualization includes:
  - Intel Virtualization Technology for Directed I/O (VT-d) and AMD-Vi/IOMMU are I/O MMUs that remap I/O DMA transfers and device interrupts, allowing VMs to have direct access to hardware I/O devices, such as network cards, HBAs, and GPUs.

For more information about hardware-assisted virtualization, see *Performance Best Practices for VMware vSphere 6.0* (https://www.vmware.com/pdf/Perf_Best_Practices_vSphere6.0.pdf).

4.3 Upgrading vSphere

vSphere 6.0 includes numerous performance and scalability enhancements that can provide a 10 to 20 percent performance boost compared to previous versions. The following table summarizes the improvements to the hypervisor, including current metrics for vSphere.

Table 2. Performance and Scalability Improvements by vSphere Version

	vSphere 4.0	vSphere 5.0	vSphere 5.1	vSphere 5.5	vSphere 6.0
CPU per VM	8 vCPU	32 vCPU	64 vCPU	64 vCPU	128 vCPU
Memory per VM	255 GB	1 TB	1 TB	1 TB	4 TB
Network	30 Gb/Sec	> 36 Gb/Sec	> 36 Gb/Sec	> 40 Gb/sec	> 80 Gb/sec
IOPS	> 350,000	1,000,000	1,000,000	1,000,000	1,000,000
Virtual Disk	2 TB	2 TB minus 512 bytes	2 TB minus 512 bytes	62 TB	62 TB
Memory per Host		2 TB	2 TB	4 TB	6 TB
Hardware Version				v10	v11

vSphere supports large capacity virtual machines, so it can accommodate larger-sized Oracle databases and SGA footprints. vSphere 6.0 host and virtual machine specifications are as follows:

- Each ESXi host supports up to 6 TB RAM, 1024 virtual machines, and 4096 virtual CPUs.
- Each virtual machine can support up to 64 virtual CPUs and 4 TB RAM.

For more information about the configuration maximum, see the references listed in Appendix C.**Error! Not a valid bookmark self-reference.** Configuration Maximums.

5. Virtual CPU Guidelines

5.1 General Guidelines for Virtual CPU

The following table describes virtual CPU-related best practices.

Table 3. Virtual CPU-Related Best Practices

Recommendation	Justification
Do not over-allocate vCPUs – try to match the exact workload.	If monitoring of the actual workload shows that the Oracle database is not benefitting from the increased virtual CPUs, the excess vCPUs impose scheduling constraints and can degrade overall performance of the virtual machine.
Enable hyperthreading for Intel Core i7 processors.	With the release of Intel Xeon 5500 series processors, enabling hyperthreading is recommended.
Keep default 1 core per socket for vNuma to match physical NUMA topology and try to align VMs with physical NUMA boundaries.	If vNUMA is different from the actual physical NUMA topology, it might result in NUMA imbalance and can degrade overall performance of the virtual machine.
Leave the latency sensitive setting at the default of Normal	Examples of applications that require the setting to be High include VOIP, media player apps, and apps that require frequent access to the mouse or keyboard devices.

5.2 Right-Sizing Virtual CPU

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 multiprocessing) virtual machines.

VMware vSphere Virtual Symmetric Multiprocessing (Virtual SMP) enhances virtual machine performance by enabling a single virtual machine to use multiple physical processors simultaneously. vSphere 6.0 supports up to 64 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 multithreaded and includes multiple processes, which make it a good candidate to take advantage of Virtual SMP.

Though larger virtual machines are possible in vSphere, VMware recommends not over-allocating vCPUs but matching the exact workload based on the actual Oracle workload.

VMware recommends the following practices for allocating CPU to Oracle Business Critical Application (BCA) database virtual machines:

- Start with a thorough understanding of your workload. Database server utilization varies widely by application.
    - If the application is commercial, follow published guidelines where appropriate.
    - If the application is custom-written, work with the application developers to determine resource requirements.
- 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. Allocate multiple vCPUs to a virtual machine only if the anticipated database workload can take advantage of all the vCPUs.
- If unsure of the workload, use hardware vendor recommended Oracle sizing guidelines.
- After the workload is established, vCPU over commitment can be done with caution. Verify that proper monitoring processes and procedures are set in place to monitor the system.
- For Tier 1 production BCA databases, the recommendation is to avoid over commitment of processor resources (maintain 1:1 ratio of physical cores to vCPUs).
- For lower-tiered workloads, reasonable over commitment can increase aggregate throughput and maximize license savings. The consolidation ratio varies depending on workloads.
- When consolidating multiple virtual machines on single ESX / ESXi host, proper hardware sizing is critical for optimal performance. Confirm that cumulative physical CPU resources on a host are adequate to meet the needs of the virtual machines by testing your workload in the planned virtualized environment.
- CPU over commitment should be based upon actual performance data to avoid adversely affecting virtual machine performance.

5.3 Hyperthreading

Hyperthreading enables a single physical processor core to behave like two logical processors, allowing two independent threads to run simultaneously. Unlike having twice as many processor cores—which can roughly double performance—hyperthreading can provide anywhere from a slight to a significant increase (up to 24 percent) in system performance by keeping the processor pipeline busier.

For example, an ESX / ESXi host system enabled for hyperthreading on an 8-core server is aware of 16 threads that appear as 16 logical processors. With the release of Intel Xeon 5500 series processors, enabling hyperthreading is recommended. Prior to the 5500 series, VMware had no uniform recommendation with respect to hyperthreading because the measured performance results were not consistent across applications, run environments, or database workloads.

Avoid CPU affinity on systems with Hyper-threading. Pinning vCPUs from different/single SMP VMs to both logical processors on one core causes poor performance because logical processors share processor resources and that will affect the ability of the NUMA scheduler to rebalance VMs across NUMA nodes for fairness.

5.4 NUMA and Virtual NUMA (vNUMA)

5.4.1 Basics of NUMA and vNUMA

vSphere supports AMD (Opteron, Barcelona, and so on), Intel (Nehalem, Westmere, and so on), and IBM (X-Architecture) non-uniform memory access (NUMA) systems. The BIOS settings for node interleaving (also known as interleaved memory) determine whether the system behaves like a NUMA system or like a uniform memory accessing (UMA) system. If node interleaving is disabled, ESXi detects the system as NUMA and applies NUMA optimizations. If node interleaving is enabled, ESXi does not detect the system as NUMA.

For more information about NUMA and vNUMA, see *Performance Best Practices for VMware vSphere 6.0* (http://www.vmware.com/files/pdf/techpaper/VMware-PerfBest-Practices-vSphere6-0.pdf).

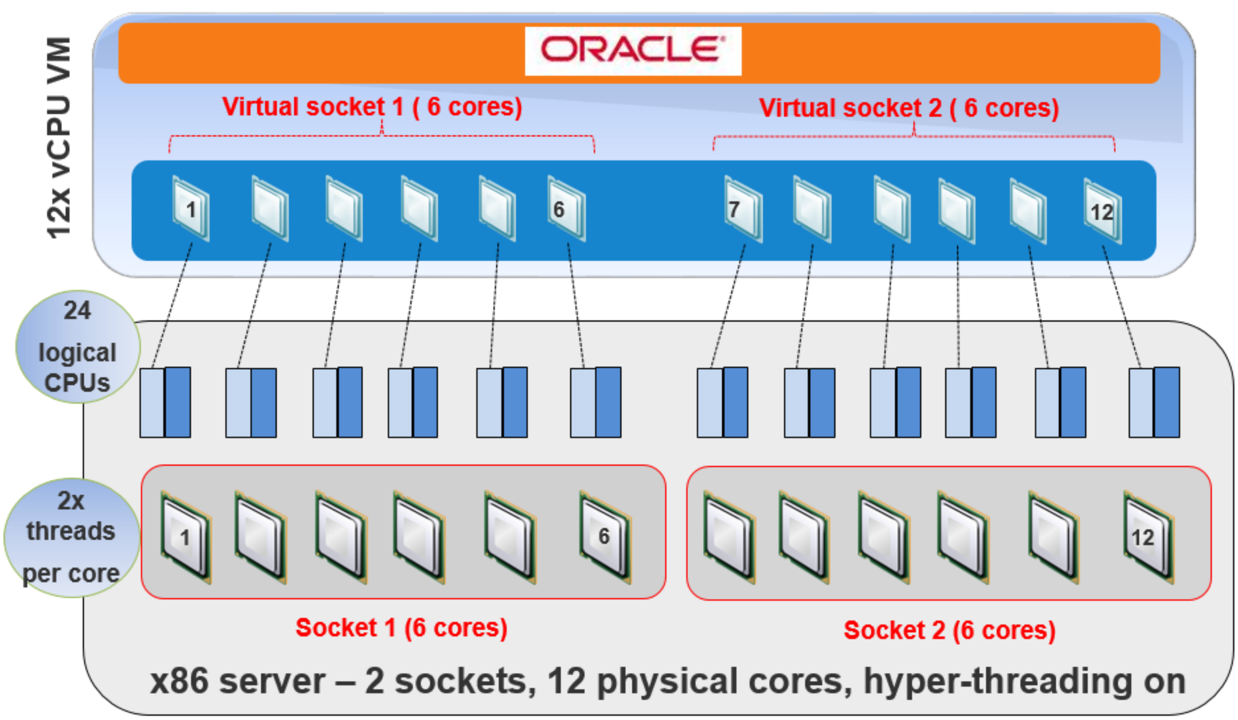
Virtual NUMA (vNUMA), a new feature in ESXi 5.0, exposes NUMA topology to the guest operating system, allowing NUMA-aware guest operating systems and applications to make the most efficient use of the underlying hardware's NUMA architecture.

By default, ESXi NUMA scheduling and related optimizations are enabled only on systems with a total of at least four CPU cores and with at least two CPU cores per NUMA node.

Wide VMs are defined as VMs with more vCPUs than the number of cores in each physical NUMA node are referred to as wide VMs. Analytics focused Tier 1 database VMs typically fall into this category.

Because vCPUs in these wide VMs might need to access memory outside their own NUMA node, they might experience higher average memory access latencies than virtual machines that fit entirely within a NUMA node.

Figure 2. Example of a 12 vCPU Wide Virtual Machine



The recommendation is to keep the default of one core per socket (with the number of virtual sockets therefore equal to the number of vCPUs) for vNUMA to match physical NUMA topology. Try to align VMs with physical NUMA boundaries.

Factors to keep in mind about vNUMA include the following:

- vNUMA requires virtual hardware version 8 or later.
- vNUMA for VM is enabled with vCPUs greater than 8 (set `numa.vcpu.min` to a lower value if there is a need to expose vNUMA to a guest with less vCPUs).
- Hot add CPU

  When configuring vNUMA with vCPU hot plug settings, the virtual machine will be started without virtual NUMA enabled. Instead it will use Uniform Memory Access with interleaved memory access. See vNUMA is disabled if VCPU hotplug is enabled (http://kb.vmware.com/kb/2040375).
- Use ESXTOP to monitor NUMA performance at vSphere

  The following table lists a few rESXTOP metrics to consider. NMIG is the key metric to examine for NUMA imbalances.

Table 4. rESXTOP NUMA Metrics

Metric	Explanation
NHN	Current home node for virtual machine.
NMIG	Number of NUMA migrations between two snapshots. It includes balance migration, inter-mode VM swaps performed for locality balancing, and load balancing.
NRMEM (MB)	Current amount of remote memory being accessed by VM.
NLMEM (MB)	Current amount of local memory being accessed by VM.
N%L	Current percentage memory being accessed by VM that is local.
GST_NDx (MB)	The guest memory being allocated for VM on NUMA node x. “x” is the node number.
OVD_NDx (MB)	The VMM overhead memory being allocated for VM on NUMA node x.

For additional details, refer to “ESXi CPU Considerations” in *Performance Best Practices for VMware vSphere 6.0* (http://www.vmware.com/pdf/Perf_Best_Practices_vSphere6.0.pdf).

5.4.2 Oracle and NUMA

Oracle NUMA support is enabled by default in Oracle 10g. Oracle NUMA support is disabled by default for 11g and above (see Oracle MySupport Doc ID: 864633.1).

The two key NUMA init.ora parameters in 10g and 11g are _enable_numa_optimization (10g) and _enable_numa_support(11g). VMware recommends not changing parameters in the `InitSID.ora` file that begin with an “_” because these are “hidden” and should only be changed when receiving specific instructions from Oracle support.

In some circumstances, enabling Oracle NUMA support might improve performance, and the Oracle documentation suggests that it be tested in a test environment before deciding to use it with production system.

VMware recommends keeping NUMA enabled in server hardware BIOS and at the guest operating system level which should also be the default settings for NUMA support with most servers and guest operating systems.

6. Memory Guidelines

6.1 General Guidelines for Memory

The following tables lists memory-related best practices.

Table 5. Memory-Related Best Practices

Recommendation	Justification
Set memory reservations equal to the sum total of the size of the Oracle SGA, the Oracle PGA , the Oracle Background processes stack space and Operating System Used Memory.	The memory reservation should be large enough to avoid kernel swapping between ESX and the guest OS because Oracle databases can be memory-intensive.
Use large memory pages.	Large page support is enabled by default in ESX 3.5 and later, and is supported from Oracle 9i R2 for Linux operating systems and 10g R2 for Windows. Enable large pages in the guest OS to improve the performance of Oracle databases on vSphere.
Do not turn off ESXI memory management mechanisms unless directed by VMware support.	ESXI uses various memory management mechanisms to reclaim virtual machine memory when under memory duress. Disabling it will lead to performance degradation and a possible crash.

Appendix A provides a description of virtual machine memory settings that are discussed in this section. For further background on VMware memory management concepts, see *vSphere Resource Management* (https://pubs.vmware.com/vsphere-60/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-60-resource-management-guide.pdf).

6.2 ESXI Memory Management Mechanism

ESXi uses several innovative techniques to reclaim virtual machine memory:

- Transparent page sharing (TPS) – Reclaims memory by removing redundant pages with identical content.
- Ballooning – Reclaims memory by artificially increasing the memory pressure inside the guest.
- Memory compression – Reclaims memory by compressing the pages that need to be swapped out.
- Hypervisor swapping – Reclaims memory by having ESXi directly swap out the virtual machine's memory.
    - Swap to host cache – ESXi will next forcibly reclaim memory using host-level swapping to a host cache which is a special swap cache on SSD storage reducing access latency.
    - Regular host-level swapping – If the host cache is full or not configured, ESXi will next reclaim memory from the virtual machine by swapping out pages to a regular swap file.

When consolidating Oracle database instances, Transparent Page Sharing technique presents the opportunity to share memory across virtual machines that are running the identical operating systems, applications, or components. TPS also allows DBAs to overcommit memory while minimizing performance degradation.

In production environments, carefully consider the impact of overcommitting memory and overcommit only 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.

The general recommendation is to not disable any of the above memory management mechanisms including balloon driver, Transparent Page Sharing (TPS) and memory compression.

Starting with vSphere 6.0, hot-adding memory to a virtual machine will distribute the new memory regions across the vNUMA nodes.

6.3 Memory Reservations

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

- Set the virtual machine memory reservation equal to the sum total of the size of the Oracle SGA, the Oracle PGA , the Oracle background processes and the operating system used memory.
- 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 can limit vSphere vMotion operations. A virtual machine can be live migrated only if the target ESX / ESXi host has free physical memory equal to or greater than the size of the reservation.
- 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 sum total of the size of the Oracle SGA, the Oracle PGA, the Oracle background processes stack space, and operating system used memory in production environments, it is acceptable to overcommit more aggressively in non-production environments such as development, test, or QA.

So for Oracle, set memory reservation = SGA + PGA + BG process sharable memory + operating system used memory

where

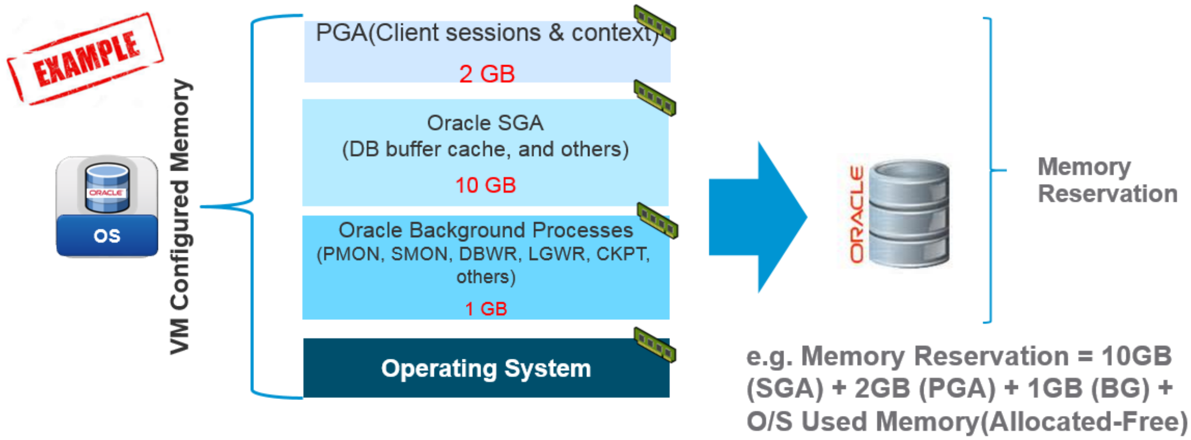
SGA – System Global Area

PGA - Program Global Area

BG – Background Processes

The following illustration shows more details.

Figure 2. Oracle Memory Reservation Best Practices



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.

6.4 Huge Pages

vSphere supports huge pages in the guest operating system. The use of huge pages results in reduced memory management overhead and can increase hypervisor performance.

Oracle supports the use of huge memory pages in Oracle 9i R2 for Linux operating systems and in 10g R2 for Windows. The following MetaLink Notes are relevant when setting huge 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*

In addition to the usual 4 KB memory pages, ESXi also provides 2 MB memory pages. These huge pages improve performance by significantly reducing TLB misses (applications with large active memory working sets).

ESXi will use large pages to back the guest operating system memory pages even if the guest operating system does not make use of large memory pages. The full benefit of huge pages occurs when the guest operating system uses them as well as ESXi does. ESXi does not share large pages unless under memory pressure. See *Transparent Page Sharing (TPS) in hardware MMU systems* at https://kb.vmware.com/kb/1021095 and *Use of large pages can cause memory to be fully allocated* at https://kb.vmware.com/kb/1021896).

VMware recommends using guest operating system level huge memory pages.

Huge pages are not compatible with the Oracle 11g Automatic Memory Management (AMM) model. The recommendation is to use huge pages with 10g Automatic Shared Memory Management model (ASMM).

See Metalink Notes 401749.1 *Shell Script to Calculate Values Recommended Huge Pages/Huge TLB.*

Oracle recommends disabling transparent huge pages because it causes performance issues with Oracle databases and causes node reboots in RAC. See Metalink Note 1557478.1.

6.5 Oracle 12c Multitenant Feature

Oracle 12c multitenant feature delivers a new architecture that allows a multitenant container database to hold many pluggable databases. An existing database can simply be adopted with no application changes required.

The many pluggable databases in a single multitenant container database share its memory and background processes, letting you operate many more pluggable databases on a particular platform than you can single databases that use the old architecture. See the Oracle 12c multitenant architecture document at http://www.oracle.com/technetwork/database/multitenant/overview/index.html.

The best practices for setting up Oracle 12c are no different from setting up Oracle 11g. The key point to consider is when deploying a multitenant architecture, memory, CPU, and IOPS requirements will increase based on the individual tenant requirements, which must be provided for individually and adequately.

7. Storage Guidelines

7.1 General Guidelines

The following table describes storage-related best practices.

Table 6. Recommended Storage Best Practices

Recommendation	Justification
Enable jumbo frames for IP-based storage using iSCSI and NFS.	Jumbo frames enable Ethernet frames to have a larger payload, allowing for improved performance.
Create dedicated datastores to service database workloads.	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.
Enable jumbo frames for IP-based storage using iSCSI and NFS.	Jumbo frames enable Ethernet frames to have a larger payload, allowing for improved performance.
Create dedicated datastores to service database workloads.	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.
Use VMware vSphere VMFS for Oracle database deployments.	To balance performance and manageability in a virtual environment, deploy Oracle using VMFS.
Align VMFS properly.	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.
Use Oracle Automatic Storage Management.	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.
Use your storage vendor’s best practices documentation when laying out the Oracle database.	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.
Avoid silos when designing the storage architecture.	At a minimum, designing the optimized architecture should involve the database administrator, storage administrator, network administrator, VMware administrator, and application owner.

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

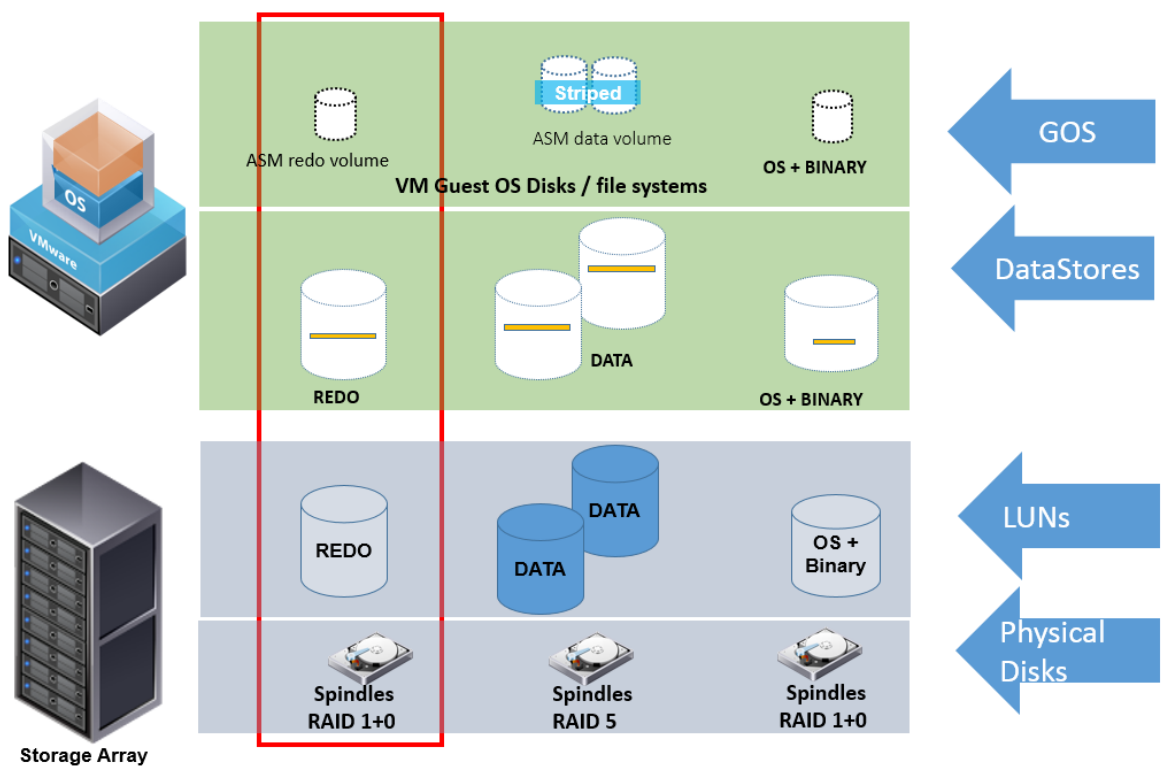
Storage configuration is essential for any successful database deployment, especially in virtual environments where you can consolidate many different Oracle database workloads on a single ESX / ESXi host. Your storage subsystem should provide sufficient I/O throughput as well as storage capacity to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

7.2 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 / ESXi hosts 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.

Figure 3. Different Layers of Storage Technology



7.2.1 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 / ESXi instances to access shared virtual machine storage concurrently. It also enables virtualization-based distributed infrastructure services such as vSphere vMotion, DRS, and vSphere HA to operate across a cluster of ESX / ESXi hosts.

7.2.2 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 be used only 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.

7.2.3 VMFS and RDMs Trade-Off

The decision to place Oracle database server data on VMFS as opposed to RDM is no longer related to the performance requirements.

VMFS has been proven to provide and in certain cases exceed native performance. The results for VMFS, RDM (virtual), and RDM (physical) are all almost identical in their results. See *Performance Characterization of VMFS and RDM Using a SAN* (http://www.vmware.com/files/pdf/performance_char_vmfs_rdm.pdf).

VMware generally recommends VMFS, but there might be situations where RDMs are required. The following table summarizes some of the options and tradeoffs between VMFS and RDM.

For a more complete discussion, see *VMware vSphere Storage* at http://pubs.vmware.com/vsphere-50/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-50-storage-guide.pdf.

Table 7. VMFS and Raw Disk Mapping Trade-Offs

VMFS	RDM
• Volume can host many virtual machines (or can be dedicated to one virtual machine). • Increases storage utilization, provides better flexibility, easier administration and management. • Can potentially support clustering software that does not issue SCSI reservations, such as Oracle Clusterware. To configure, follow the procedures given in *Disabling simultaneous write protection provided by VMFS using the multi-writer flag* (http://kb.vmware.com/kb/1034165). • Oracle RAC node live migration.	• Maps a single LUN to one virtual machine, so only one virtual machine is possible per LUN. • More LUNs are required, so it is easier to reach the LUN limit of 256 that can be presented to an ESX/ESXi host. • RDM might be required to leverage third party storage array-level backup and replication tools. • 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.

7.3 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 price, performance, and manageability. In addition, the protocols available with vSphere—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 bandwidth is the limiting factor for I/O throughput when comparing the storage protocols. VMware ESX / ESXi can reach the full bandwidth capacity in a single virtual machine environment, and also maintain the throughput for 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 (http://www.vmware.com/files/pdf/perf_vsphere_storage_protocols.pdf). Fibre Channel might provide maximum I/O throughput, but iSCSI and NFS can offer a better price-to-performance ratio.

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

Table 5. Storage Protocol Capabilities

Type	Boot VM	Boot vSphere	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 as well as 10 GbE connectivity are recommended for IP-based storage using iSCSI and NFS. Jumbo frames must be enabled for each virtual switch through the VMware vSphere Client. Also, if you use an ESX / ESXi host, you must create a VMkernel network interface with jumbo frames enabled. It is also necessary to enable jumbo frames on the hardware as well, including the network switches and storage arrays. When employing jumbo frames, every network hop in the path should be enabled.

See *Enabling Jumbo Frames on virtual distributed switches* at https://kb.vmware.com/kb/1038827.

7.4 Partition Alignment

Aligning file system partitions is a well-known storage best practice for database workloads. Partition alignment on both physical machines and VMFS partitions prevents performance I/O degradation caused by I/O crossing track boundaries.

VMware test results show that aligning VMFS partitions to 1 MB track boundaries results in reduced latency and increased throughput. vCenter 5.x/ 6.0 aligns VMFS 5 partitions along the 1 MB boundary, while vCenter 4.x aligns VMFS 3 partitions along a 64K boundary.

VMFS partitions created using vCenter are aligned on 1 MB boundaries as recommended by storage and operating system vendors.

Align the data disks within the guest operating system also using disk partitioning utilities. Partition mis-alignment can add significant latency to high-end workloads due to a single I/O having to cross physical boundaries. Partition alignment reduces the I/Os sent to disk by the controller thus reducing latency.

It is considered a best practice to do the following:

- Create VMFS partitions from within vCenter because they are aligned by default.
- Align the data disk for heavy I/O workloads using `diskpart`, `fdisk` or `parted` depending on the operating system.
- Consult with the storage vendor for alignment recommendations on their hardware.

For more information, see the white paper *Performance Best Practices for VMware vSphere 6.0* (http://www.vmware.com/pdf/Perf_Best_Practices_vSphere6.0.pdf).

7.5 Deduplication and Compression Function on Storage Array

Most storage venders provide data deduplication and compression features at the array level. The storage offloads the deduplication and compression CPU overhead from the server to the storage.

Good candidates for deduplication and compression are the root disks of Windows /Linux Guest OS which has the same data pattern.

Oracle databases can also be placed on datastores with deduplication and compression capabilities.

See the guidelines of storage vendors' best practices for deduplication and compression features.

7.6 Choosing Right Path Policy Connect to Storage Array

Most of the storage vendors provide best practices for optimal storage path connectivity to the ESXi hosts. VMware provides various storage path policies which include MRU (default), fixed, and round-robin. In addition, the storage vendors also provide third-party multipathing vendor-specific solutions.

VMware recommends increasing the aggregate I/O performance is to change the round-robin I/O operations limit to a lower value from the default value of 1000. This is also a recommendation from the storage vendors. For example, EMC Powerpath/VE provides HBA tuning for optimizing effective I/O path selection.

VMware defers the optimal I/O path selection policy to the specific storage vendor.

The following articles detail the best practices for choosing the correct and optimal path policy:

- *Best Practices for Running VMware vSphere on iSCSI*

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

- *Tuning the VMware Native Multipathing Performance of VMware vSphere Host Connected to EMC Symmetrix Storage*

  http://www.emc.com/collateral/hardware/white-papers/h8119-tuning-vmware-symmetrix-wp.pdf

7.7 Minimum Hub Between HBA/iSCSI to Storage Array

Recommendation is to separate the storage network from the production network. Reducing the hub distance between the HBA/iSCSI cards and the storage array leads to reduction in I/O latency.

Routing is not supported when iSCSI port binding is deployed. The best practice recommendation is to keep the HBA/iSCSI and storage array connections in same switch.

Refer to the following article for best practices regarding optimal hub distance:

- *Best Practices for Running VMware vSphere on iSCSI*
  http://www.vmware.com/files/pdf/iSCSI_design_deploy.pdf

7.8 VMDK Creation and vSphere Storage APIs - Array Integration

There are three ways a VMDK can be created:

- Eager zeroed thick
- Lazy zeroed thick
- Thin provisioned

The default VMDK allocation policy is thick provisioned lazy zeroed where 1 MB VMFS blocks are zeroed out on the first write. There is a write penalty on an untouched VMDK which could affect database operations like write operations and read operations that use temporary tablespace / tempdb extensively and bulk load/index maintenance.

For best performance, format VMDK as eager zeroed thick for VMDKs from a storage array that is not VMware vSphere Storage APIs - Array Integration aware.

For storage arrays with vSphere Storage APIs - Array Integration capability, the zero offload capability improves the zeroing process, so the VMDKs can be thin provisioned instead of thick provisioned.

VMware recommends choosing a storage that supports vSphere Storage APIs - Array Integration.

VMware vSphere Storage APIs - Array Integration delivers the following capabilities:

- Hardware-accelerated cloning (that is, Clone Blocks/Full Copy/XCOPY)
- Scalable lock management (that is, Atomic Test and Set [ATS])
- Zero blocks/write same used to zero-out disk regions
- Using thin provision UNMAP to allow the array to reuse no-longer needed blocks
- Block Delete allows for space to be reclaimed using the SCSI UNMAP feature

See *Frequently Asked Questions for vStorage APIs for Array Integration* (http://kb.vmware.com/kb/1021976).

When deploying Oracle RAC on vSphere, the shared disks have to be eager zero thick to use the multi-writer capabilities.

See *Enabling or disabling simultaneous write protection provided by VMFS using the multi-writer flag* (http://kb.vmware.com/kb/1034165).

7.9 VMware vSphere Flash Read Cache Creation

vSphere 5.5 introduced a new functionality to leverage flash storage devices on an ESXi host. The vSphere flash infrastructure layer is part of the ESXi storage stack for managing flash storage devices that are locally connected to the server. These devices can be of multiple types (primarily PCIe flash cards and SAS/SATA SSD drives) and the vSphere flash infrastructure layer is used to aggregate these flash devices into a unified flash resource if needed.

The flash resource created by the vSphere flash infrastructure layer can be used for two purposes:

- Read caching of virtual machine I/O requests (VMware vSphere Flash Read Cache™)
- Storing the host swap file

Flash Read Cache is a feature in vSphere 5.5 that utilizes the vSphere flash infrastructure layer to provide a host-level caching functionality for virtual machine I/Os using flash storage. The goal of introducing the Flash Read Cache feature is to enhance performance of certain I/O workloads that exhibit characteristics suitable for caching.

Flash Read Cache is a volatile write-through cache that is enabled on a per-VMDK basis with a cache block size (4 KB to 1 MB). See the *VMware vSphere Flash Read Cache1.0* FAQ at www.vmware.com/files/pdf/techpaper/VMware-vSphere-Flash-Read-Cache-FAQ.pdf.

VMware recommends deploying Flash Read Cache for single instance database. Currently, there is no support for Oracle RAC. See *Enabling or disabling simultaneous write protection provided by VMFS using the multi-writer flag* at https://kb.vmware.com/kb/1034165. The following figure is a snapshot of a table from this Knowledge Base article.

Figure 4. VMFS Multi-Writer Flag Supported/Unsupported Features

Actions or Features	Supported	Unsupported	Notes
Power on, off, restart virtual machine	√		
Suspend VM		×	
Hot add virtual disks	√		Only to existing adapters
Hot remove devices	√		
Hot extend virtual disk		×	
Connect and disconnect devices	√		
Snapshots		×	Virtual backup solutions leverage snapshots through the vStorage APIs; for example, VMware Data Recovery, vSphere Data Protection. These are also not supported.
Snapshots of VMs with independent-persistent disks	√		Supported in vSphere 5.1 update2 and later versions
Cloning		×	
Storage vMotion		×	Neither shared nor non-shared disks can be migrated using Storage vMotion due to the virtual machine stun required to initiate the storage migration.
Changed Block Tracking (CBT)		×	
vSphere Flash Read Cache (vFRC)		×	Stale writes can lead to data loss and/or corruption
vMotion	√		Supported for ORAC only and limited to 8 ESX/ESXi hosts

7.10 VMware vSphere Virtual Volumes

There are many common concerns about virtualizing multi-terabyte business critical databases that inhibit and delay implementation, including the following:

- Business critical virtualized databases must meet strict SLAs for performance and storage is usually the slowest component
- The size of the databases are growing, while at the same time there is an increasing need to reduce backup windows, impacting system performance.
- There is a regular need to clone and refresh databases from production to QA and other environments. The size of the modern databases makes it harder to clone and refresh data from production to other environments.
- Databases of different levels of criticality require different storage performance characteristics and capabilities.

- Deficiencies in earlier versions of VMFS sometimes led to the use of RDMs, because they were a superior virtual storage option at the time. This is no longer true.

It is a big challenge to back up multi-terabyte databases due to minimal and restricted backup windows. The data churn itself can also be substantial. It is not feasible to make full backups of multi-terabyte databases in the allotted backup windows.

There are three different levels of triggering backups of Oracle databases on vSphere:

- Application (for example, using Oracle RMAN)
- vSphere (for example, using VMware snapshots)
- Storage (for example, storage-based snapshots, disk sync/split, and so on)

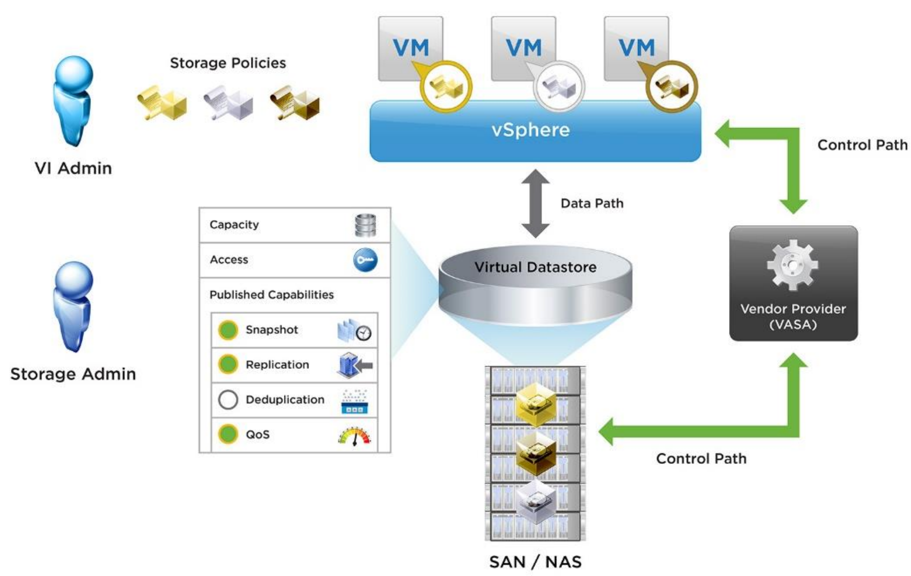
Backup solutions like Oracle RMAN and SQL backup provide fine-level granularity for database backups, but often do not constitute the fastest solution. Virtual machine snapshots containing the Oracle database as backups would be ideal but as described in *A snapshot removal can stop a virtual machine for long time* (http://kb.vmware.com/kb/1002836), the brief stun moment of the VM can potentially cause performance issues. Storage-based snapshots would be the fastest of all the options, but unfortunately the datastore/LUN level snapshot is not the same as the VM level. Therefore, there is no VMDK-level granularity with storage-level snapshots.

An ideal backup solution combines the capabilities at the storage level with the granularity of a VM level snapshot:

- The solution should be able to trigger backups and clones with VMDK granularity at the same time from a virtual machine level.
- The solution should match the speed of the storage based snapshots.

VMware vSphere Virtual Volumes™ is the technology that provides the solution because it has the capability of achieving both objectives—VMDK level granularity with storage level snapshot capability.

Figure 5. vSphere Virtual Volumes Architecture

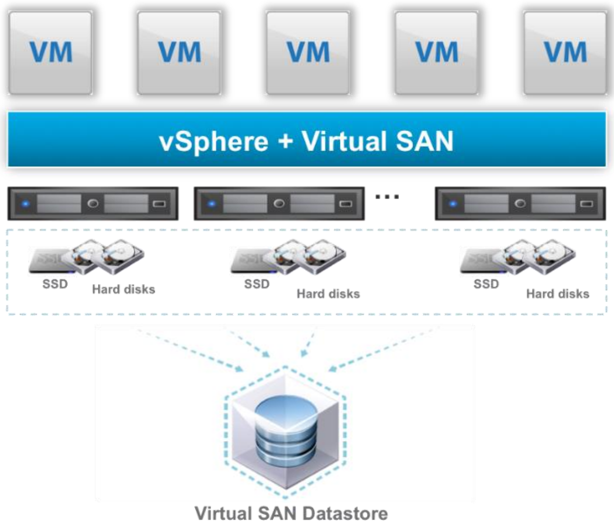


For a detailed description of VMware vSphere Virtual Volumes, see the *VMware vSphere Virtual Volumes* white paper at http://www.vmware.com/files/pdf/products/vvol/vmware-oracle-on-virtual-volumes.pdf.

7.11 VMware Virtual SAN

VMware Virtual SAN™ is the VMware software-defined storage solution for hyper-converged infrastructure, a software-driven architecture that delivers tightly integrated computing, networking, and shared storage from a single virtualized x86 server. Virtual SAN delivers high performance and highly resilient shared storage by clustering server-attached flash devices and hard disks (HDDs). Virtual SAN delivers enterprise-class storage services for virtualized production environments along with predictable scalability and all-flash performance. Virtual SAN can be configured as all-flash or hybrid storage.

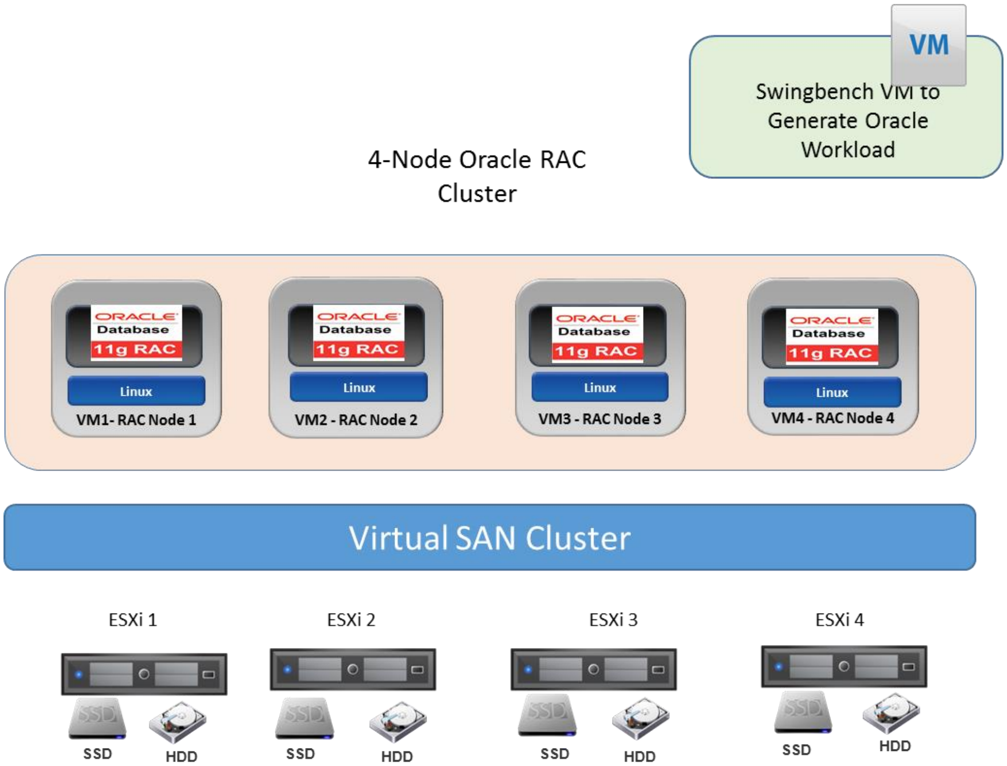
Figure 5. Virtual SAN Cluster Datastore



Virtual SAN supports a hybrid disk architecture that leverages flash-based devices for performance and magnetic disks for capacity and persistent data storage. In addition, Virtual SAN can use flash-based devices for both caching and persistent storage. It is a distributed object storage system that leverages the vSphere Storage Policy-Based Management (SPBM) feature to deliver centrally managed, application-centric storage services and capabilities. Administrators can specify storage attributes, such as capacity, performance, and availability, as a policy on a per-VMDK level. The policies dynamically self-tune and load balance the system so that each virtual machine has the appropriate level of resources

Oracle single-instance and Oracle Real Application Clusters (RAC) databases can be deployed on Virtual SAN.

Figure 6. Virtual SAN Cluster for Oracle RAC



For more details, see the *Oracle Real Application Clusters on VMware Virtual SAN* reference architecture (http://www.vmware.com/files/pdf/products/vsan/vmware-oracle-real-application-clusters-on-vmware-virtual-san-reference-architecture.pdf).

8. Networking Guidelines

8.1 General Networking Guidelines

The following table describes networking-related best practices.

Table 8. Recommended Networking-Related Best Practices

Recommendation	Justification
Use the VMXNET family of paravirtualized network adapters	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.
Separate infrastructure traffic from virtual machine traffic for security and isolation.	Virtual machines should not see infrastructure traffic (security violation) and should not be impacted by infrastructure traffic bursts (for example, vSphere vMotion operations).
Use NIC teaming for availability and load balancing.	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.
Take advantage of Network I/O Control to converge network and storage traffic onto 10 GbE.	This can reduce cabling requirements, simplify management, and reduce cost.
Enable jumbo frames for Oracle interconnect traffic.	This can reduce cache fusion traffic thereby enhancing performance.

8.2 Networking Best Practices

The standard VMware networking best practices apply to running Oracle databases on vSphere. For details, see *VMware vSphere Networking* (http://pubs.vmware.com/vsphere-50/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-50-networking-guide.pdf).

This guide includes designs, which will efficiently manage multiple networks, and redundancy of network adaptors on ESX / ESXi 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 virtual switch to form a team.
- Take advantage of Network I/O Control to converge network and storage traffic onto 10 GbE. Network I/O 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, and vSphere vMotion.

- Use the VMXNET3 network adapter in vSphere. 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 more information about network adaptors and compatibility with the ESX / ESXi release and supported guest operating system, see *Choosing a network adapter for your virtual machine* (http://kb.vmware.com/kb/1001805).
- Use VMXNET3 paravirtualized adapter drivers for Oracle RAC private interconnect traffic.
- Allocate separate NICs for vSphere vMotion, FT logging traffic, ESXi console access management, and Oracle RAC interconnect.
- vSphere 5.0 supports the use of more than one NIC for vSphere vMotion, allowing more simultaneous vSphere vMotion instances; added specifically for memory-intensive applications like databases.

Figure 7. Example Network Layout of Oracle RAC Database on VMware

8.3 Virtual Switches

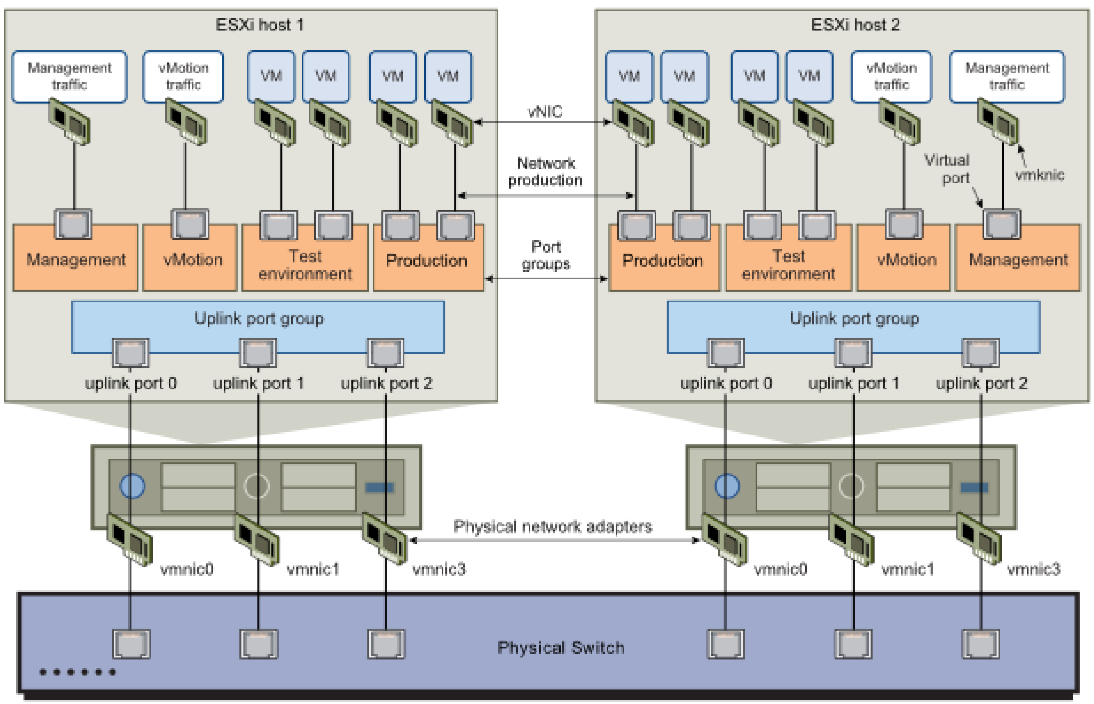
8.3.1 Standard Switches

An administrator can create abstracted network devices called vSphere standard switches. A standard switch will bridge traffic internally between virtual machines in the same port group and link to external networks.

The administrator can use standard switches to combine the bandwidth of multiple network adapters and balance communications traffic among them. You can also configure a standard switch to handle physical NIC failover.

A vSphere standard switch models a physical Ethernet switch. The default number of logical ports for a standard switch is 120. You can connect one network adapter of a virtual machine to each port. Each uplink adapter associated with a standard switch uses one port. Each logical port on the standard switch is a member of a single port group. Each standard switch can also have one or more port groups assigned to it.

Figure 8. Standard Switch

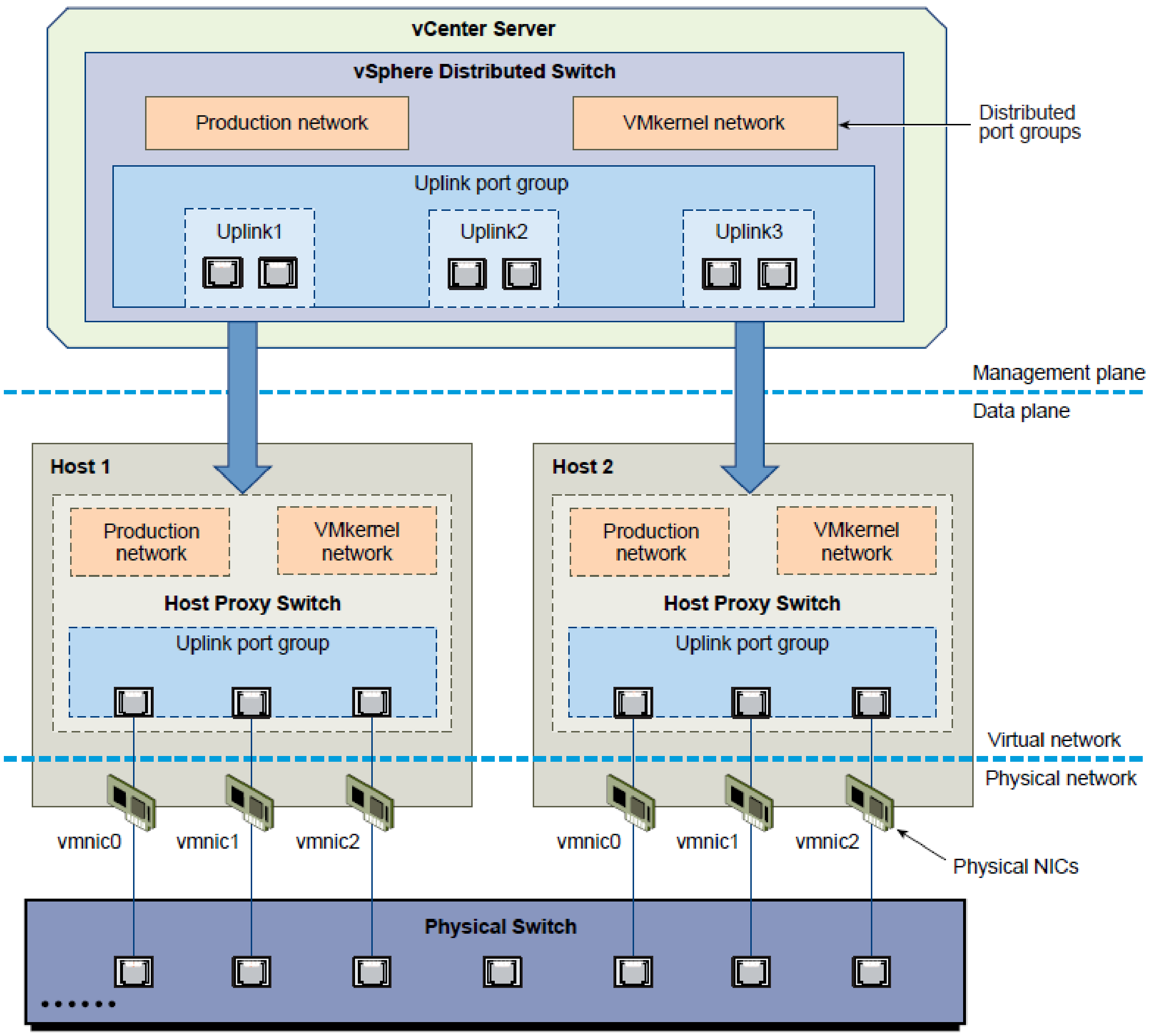


8.3.2 Distributed Switches

A VMware vSphere Distributed Switch™ provides centralized management and monitoring of the networking configuration of all hosts that are associated with the switch. When an administrator uses a distributed switch setup on a VMware vCenter Server® system, its settings are propagated to all hosts that are associated with the switch.

A network switch in vSphere consists of two logical sections which are the data plane and the management plane. The data plane implements the package switching, filtering, tagging, and so on. The management plane is the control structure that you use to configure the data plane functionality. A vSphere standard switch contains both data and management planes, and you configure and maintain each standard switch individually. A vSphere Distributed Switch separates the data plane and the management plane. The management functionality of the distributed switch resides on the vCenter Server system that lets you administer the networking configuration of your environment on a data center level. The data plane remains locally on every host that is associated with the distributed switch. The data plane section of the distributed switch is called a host proxy switch. The networking configuration that you create on vCenter Server (the management plane) is automatically pushed down to all host proxy switches (the data plane).

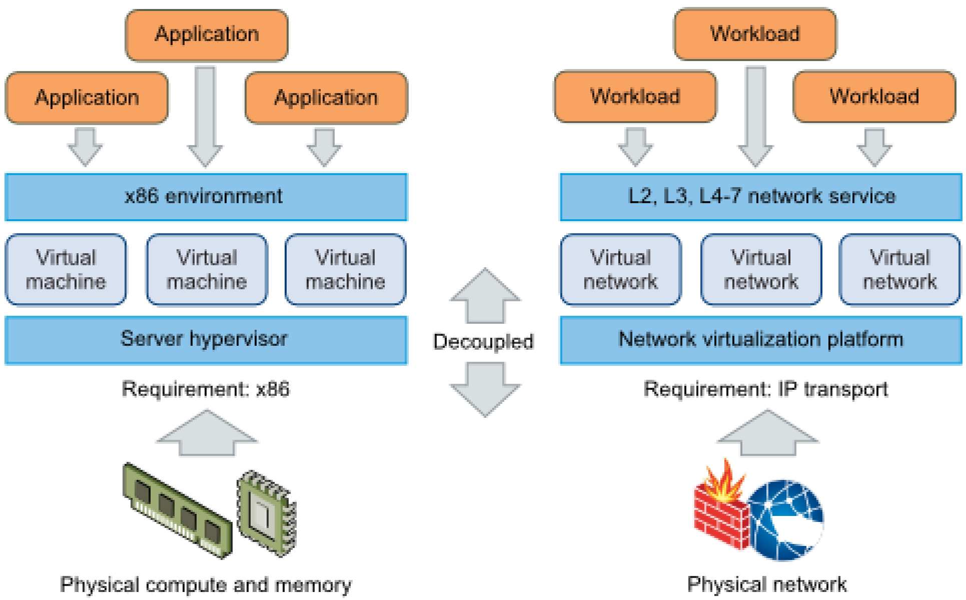
Figure 9. Distributed Switch



8.4 VMware NSX

The VMware software-defined data center (SDDC) architecture extends virtualization technologies across the entire physical data center infrastructure. VMware NSX®, the network virtualization platform, is a key product in the SDDC architecture. With NSX, virtualization delivers for networking what has previously been delivered for compute and storage. In much the same way that server virtualization programmatically creates, snapshots, deletes and restores software-based virtual machines (VMs), VMware NSX network virtualization programmatically creates, snapshots, deletes, and restores software-based virtual networks. The result is a completely transformative approach to networking that not only enables data center managers to achieve orders of magnitude better agility and economics, but also allows for a vastly simplified operational model for the underlying physical network. With the ability to be deployed on any IP network, including both existing traditional networking models and next-generation fabric architectures from any vendor, VMware NSX is a completely non-disruptive solution. In fact, with VMware NSX, the physical network infrastructure you already have is all you need to deploy a software-defined data center.

Figure 10. Architecture Deploying VMware NSX



8.5 Networking for Oracle RAC

8.5.1 Interrupt Coalescing

VMXNET 3 by default also supports an adaptive interrupt coalescing algorithm, for the same reasons that physical NICs implement interrupt moderation. This virtual interrupt coalescing helps drive high throughputs to VMs with multiple vCPUs with parallelized workloads (for example, multiple threads), while at the same time striving to minimize the latency of virtual interrupt delivery.

However, if your workload is extremely sensitive to latency, we recommend you disable virtual interrupt coalescing for VMXNET 3 virtual NICs. To do so, through the VMware vSphere Web Client, go to **VM Settings** > **Options** tab > **Advanced General** > **Configuration Parameters** and add an entry for **ethernetX.coalescingScheme** with the value of disabled.

Note that this new configuration option is only available in ESXi 5.0 and later. An alternative way to disable virtual interrupt coalescing for all virtual NICs on the host, which affects all VMs, not just the latency-sensitive ones, is to set the advanced networking performance option (**Configuration** > **Advanced Settings** > **Net**) **CoalesceDefaultOn** to **0** (disabled).

For Oracle RAC Interconnect ,VMware recommends disabling virtual interrupt coalescing for VMXNET3 Private NIC by setting *ethernetX*.coalescingScheme=disabled where *ethernetX* is the vmnic of the Oracle RAC interconnect.

8.5.2 Jumbo Frames

A single default Ethernet frame is 1500 bytes. A jumbo frame MTU (maximum transmission unit) can be increased from the default 1500 bytes to 9000 bytes. The larger payload provides for increased efficiency and higher throughput. For general information, see *Jumbo frame* at http://en.wikipedia.org/wiki/Jumbo_frame.

All storage networks with 10 GbE topology can benefit from jumbo frames which make the storage network much more efficient by sending larger payloads. The cost of jumbo frames occurs if a packet must be retransmitted because it was dropped or malformed upon receiving it. However, LANs and in particular storage networks, do not usually experience many dropped packets. By enabling jumbo frames on a 10-GbE network, an 85 percent performance improvement has been recorded.

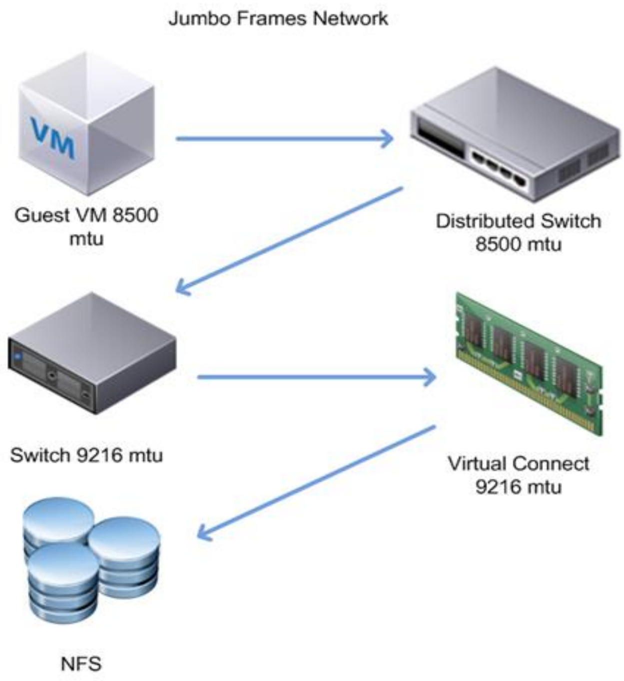
Oracle, by default, uses an 8 KB block size to store data in the database. Increasing the TCP/IP payload to 9000+ bytes allows an entire Oracle block to be transmitted as one jumbo frame instead of dividing it up into six TCP/IP packets.

Only enhanced VMXNET3 adapters can be used with jumbo frames. See *Virtual network adapters that support jumbo frames* at http://kb.vmware.com/kb/1015556.

Implementing jumbo frames requires that it be enabled throughout the network topology. This can include the guest virtual machine, virtual switch, physical switch or port, and storage NICs if applicable. See *iSCSI and Jumbo Frames configuration on ESX 3.x and ESX 4.x* at http://kb.vmware.com/kb/1007654.

For RAC interconnect VMware recommends using jumbo frames. To enable jumbo frames, follow the instructions described in *Enabling and verifying IOAT and Jumbo frames* at http://kb.vmware.com/kb/1003712.

Figure 11. Jumbo Frame Network



9. Guest Operating System Guidelines

9.1 Host Processes

The following table provides a recommendation for host processes.

Table 9. Recommendation for Host Processes

Recommendation	Justification
Disable unnecessary foreground and background processes within the guest operating system to save CPU cycles.	The impact of unnecessary foreground and background processes in the guest operating systems can lead to unnecessary CPU wastage.

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 to load only 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 as installing for a physical environment. Use the recommended kernel parameters listed in the Oracle Installation guide. It is a good practice to check with Oracle Support for the latest settings to use prior to beginning the installation process.

9.2 Oracle Installation Prerequisite Checklist

VMware recommends referring to the specific Oracle documentation for hardware and operating system requirements to verify that all system and storage preparation and configuration tasks are completed before starting an Oracle installation.

For details on hardware and operating system specific requirements, see the Oracle Installation chapter at https://docs.oracle.com/en/database/database.html.

9.3 Timekeeping in Virtual Machines

The following table provides a timekeeping best practice for virtual machines.

Table 10. Virtual Machine Timekeeping Best Practice Recommendation

Recommendation	Justification
To minimize time drift in virtual machines, follow guidelines in relevant VMware Knowledge Base articles.	The impact of high timer-interrupt rates 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-interrupt rates can lead to time synchronization errors.

To address timekeeping issues when running Oracle databases, follow the guidelines in the following VMware Knowledge Base articles:

- *Timekeeping best practices for Linux guests*
  http://kb.vmware.com/kb/1006427
- *Timekeeping best practices for Windows, including NTP*
  http://kb.vmware.com/kb/1318

For Oracle RAC environments, follow the guidelines in the following VMware Knowledge Base articles:

- *Disabling Time Synchronization*
  http://kb.vmware.com/kb/1189

In the VMware Tools control panel, the time synchronization check box is unselected, but you might experience these symptoms:

- When you suspend a virtual machine, the next time you resume that virtual machine it synchronizes the time to adjust it to the host.
- Time is resynchronized when you migrate the virtual machine using vSphere vMotion, take a snapshot, restore to a snapshot, shrink the virtual disk, or restart the VMware Tools service in the virtual machine (including rebooting the virtual machine).

10. Database Guidelines

10.1 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.

Virtual machine virtual disks can be provisioned on Virtual Machine File System (VMFS) or Raw Device Mapping (RDM) as already explained in in Section 7.2, Storage Virtualization Concepts.

10.1.1 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 / ESXi instances to access shared virtual machine storage concurrently. It also enables virtualization-based distributed infrastructure services such as vSphere vMotion, DRS, and vSphere HA to operate across a cluster of ESX / ESXi hosts.

10.1.2 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 be used only 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.

The tradeoffs between VMFS and RDMs is covered in Section 7.2.3, VMFS and RDMs Trade-OffVMFS and RDMs Trade-Off. VMware generally recommends VMFS, but there might be situations where RDMs are required.

10.1.3 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, observe the following practices:

- 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 another for data files, which are random in nature.

Consider configuring the ASM disk groups with external redundancy if the storage is capable of providing protection through hardware mirrors / copies for the primary LUN. When using external redundancy, disk failures are transparent to the database and consume no additional database CPU cycles because this is offloaded to the storage processors.

VMware recommends not using Oracle ASM failure groups. Oracle failure groups consume additional CPU cycles and can operate unpredictably after suffering a disk failure.

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.

10.1.4 Oracle Clustered File System

The Oracle Clustered File System (OCFS) 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.

Starting with Oracle Database 12c, OCFS is desupported on Windows. Support and distribution of OCFS on Linux (OCFS and OCFS2) remain unaffected. Databases currently using OCFS on Windows to host either the Oracle cluster files (Oracle Cluster Registry and voting files) or database files or both need to have these files migrated off OCFS before upgrading to Oracle Database 12c.

For more information, see https://docs.oracle.com/database/121/UPGRD/deprecated.htm#BABFEJDB.

10.1.5 PVSCSI Controllers

VMware highly recommends using multiple paravirtual SCSI (PVSCSI) controllers for the database virtual machines or virtual machines with high I/O load. The use of multiple paravirtual SCSI controllers allows the execution of several parallel I/O operations inside the guest operating system.

PVSCSI drivers are installed when VMware Tools is installed on the virtual machine.

VMware also highly recommends separating the Redo/Log I/O traffic from the data file I/O traffic through separate virtual SCSI controllers.

See the “Add a Paravirtual SCSI Controller” section in the *VMware vSphere 6.0 Documentation Center* for information on how to add additional paravirtual SCSI controllers (http://pubs.vmware.com/vsphere-60/index.jsp#com.vmware.vsphere.hostclient.doc/GUID-C5A081C0-CFF9-4D54-87A8-54FE7BA8D853.html?resultof=%2522%2550%2556%2553%2543%2553%2549%2522%2520%2522%2570%2576%2573%2563%2573%2569%2522%2520%2522%2543%256f%256e%2574%2572%256f%256c%256c%2565%2572%2573%2522%2520%2522%2563%256f%256e%2574%2572%256f%256c%2522%2520).

Follow guidelines in these VMware Knowledge Base articles:

- *Configuring disks to use VMware Paravirtual SCSI (PVSCSI) adapters* at http://kb.vmware.com/kb/1010398
- *Do I choose the PVSCSI or LSI Logic virtual adapter on ESX\ESXi 4.0 for non-IO intensive workloads?* at http://kb.vmware.com/kb/1017652

10.1.6 `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.

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. 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.

10.2 Example of Oracle Database Storage Layout on 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 data files in separate LUNs). Follow similar guidelines when virtualized.

An example layout is shown in the following figure.

Figure 6. Example Storage Layout of Oracle OLTP Database on VMware

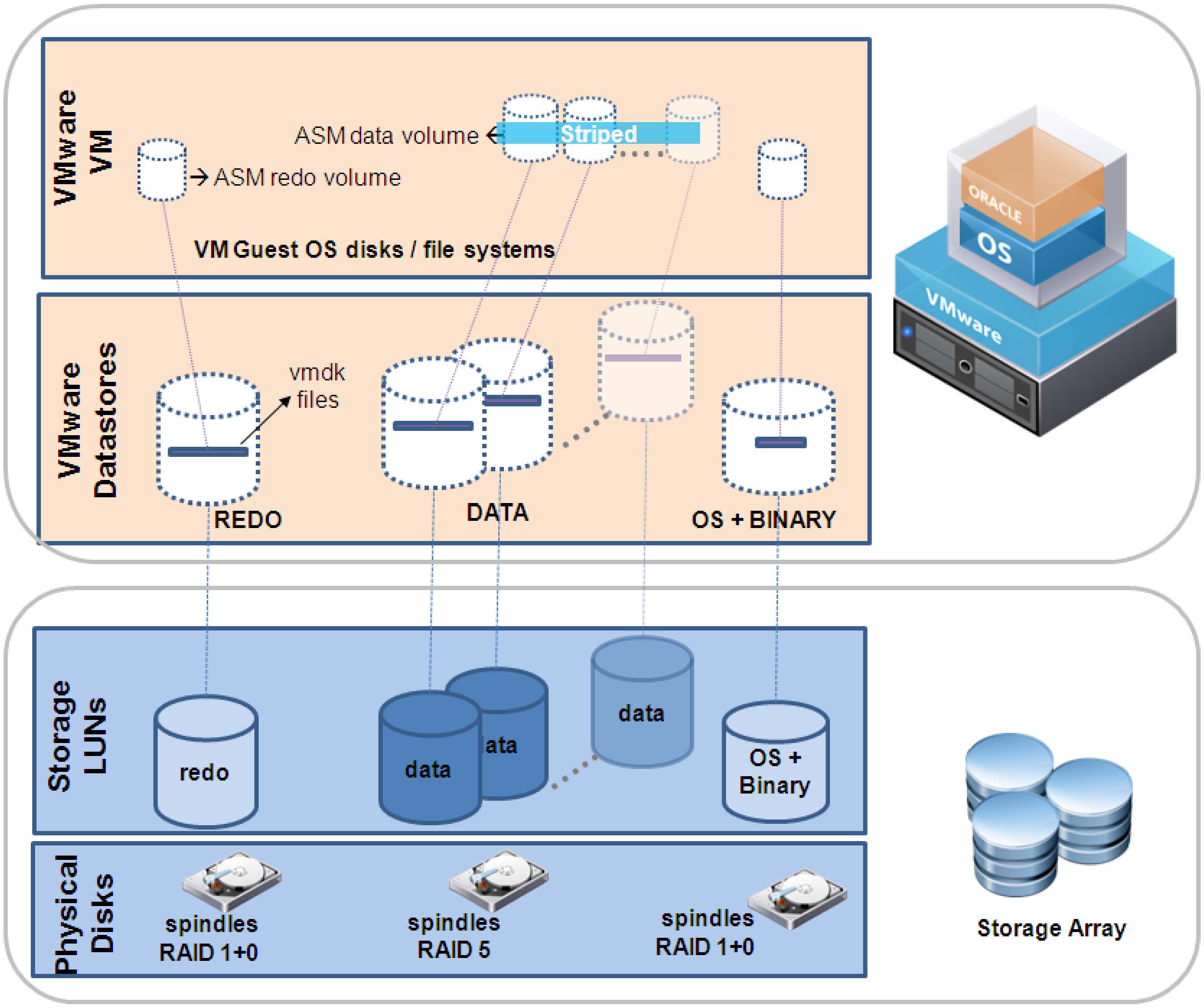


Figure 6This figure represents an example storage design for a virtualized Oracle OLTP database. The design is based on the following principles:

- At a minimum, an optimized architecture requires joint collaboration among 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 this figure illustrates only an example, and actual configurations for customer deployments can differ.

11. Performance Monitoring on vSphere

The following is a performance monitoring best practice.

Table 11. Performance Monitoring Recommendation

Recommendation	Justification
Use vCenter and/or the `esxtop`/`resxtop` utility for performance monitoring in the virtual environment.	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 / ESXi reports.

11.1 ESXTOP/rESXTOP

The resxtop and esxtop command-line utilities provide a detailed look at how ESX / ESXi uses resources in real time. We can start either utility in one of three modes: interactive (default), batch, or replay.

The fundamental difference between resxtop and esxtop is that we can use resxtop remotely (or locally), whereas esxtop can be started only through the service console of a local ESX host.

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 / ESXi reports.

Oracle DBA administrators should pay close attention to the counters listed in the following table.

See *VMware Communities: Interpreting esxtop Statistics* (http://communities.vmware.com/docs/DOC-9279) for a full list of counters.

Table 12. ESX/ESXi Performance Counters

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

Of the CPU counters, the used time indicates system load, and ready time indicates overloaded CPU resources.

A significant swap rate in the memory counters is a clear indication of a shortage of ESX / ESXi 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 whether there is any swapping within the guest operating system, use in the in-guest counters in the same manner as in physical environments.

Relevant performance attributes and measurement methods have been covered under the Virtual CPU Guidelines (Chapter 5), Memory Guidelines (Chapter 6), Storage Guidelines (Chapter 7) and Networking Guidelines (Chapter 8)VMware vRealize Operations and Blue Medora Adapter

VMware vRealize® Operations™ is built on a scale-out, resilient platform designed to deliver intelligent operational insights to simplify and automate management of applications and infrastructure across virtual, physical and cloud environments—from vSphere to Hyper-V, Amazon Web Services (AWS), and more. With vRealize Operations comprehensive visibility across applications and infrastructure in one place, IT organizations of all sizes can improve performance, avoid business disruption, and become more efficient.

vRealize Operations delivers:

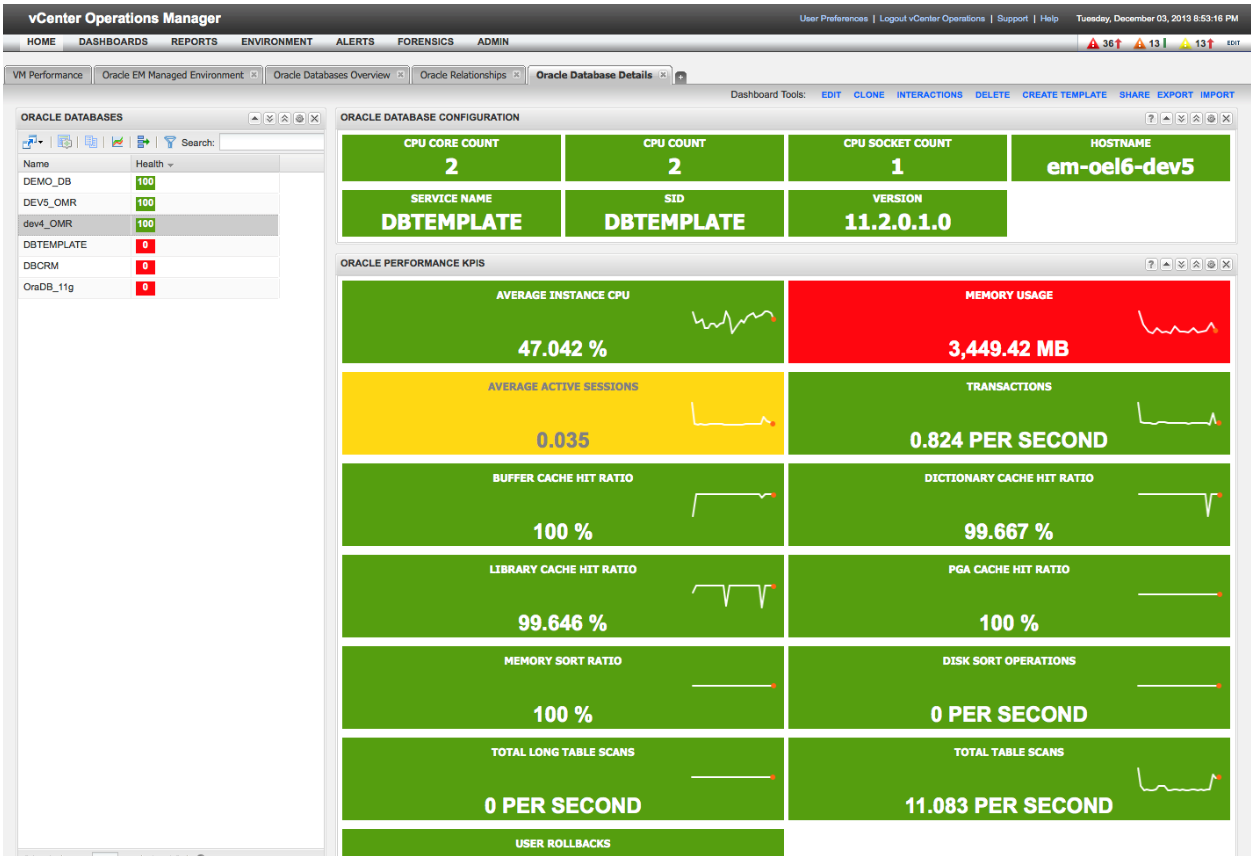
- Intelligent operations – Self-learning tools, predictive analytics, intelligent workload management, and Smart Alerts about application and infrastructure health enable proactive identification and remediation of emerging performance, capacity, and configuration issues.
- Policy-based automation – Out-of-the-box and customizable policies for critical IT operations are associated with Smart Alerts, guided remediation, and compliance standards to deliver recommendations, or trigger actions, that optimize performance and capacity and enforce configuration standards.
- Unified management – An open and extensible platform, supported by third-party management packs for Microsoft, SAP, and others, provides complete visibility in a single console across applications, storage, and network devices.

The Blue Medora - VMware vRealize Operations Management Pack for Oracle Enterprise Manager extends the VMware vRealize Operations Enterprise Edition by integrating with Oracle Enterprise Manager and providing comprehensive visibility and insights into the health, capacity, and performance of Oracle Databases, Oracle Middleware, and Oracle business critical applications. This management package helps to detect capacity and performance issues so they can be corrected before they cause a major impact.

Figure 12. Oracle Database Details (1 of 3)

Oracle Enterprise Manager (OEM) is Oracle's systems management platform and is deployed by many Oracle customers. The latest version is Oracle Enterprise Manager Cloud Control (EM12c). EM12c is used to manage, provision, and patch the Oracle infrastructure landscape including Oracle database, Oracle middleware (WebLogic, Tuxedo, IdM, and so on), various Oracle applications (PeopleSoft, Siebel, JD Edwards, Oracle EBS, Fusion, and so on), and also Oracle hardware (storage, server, Exadata, and so on).

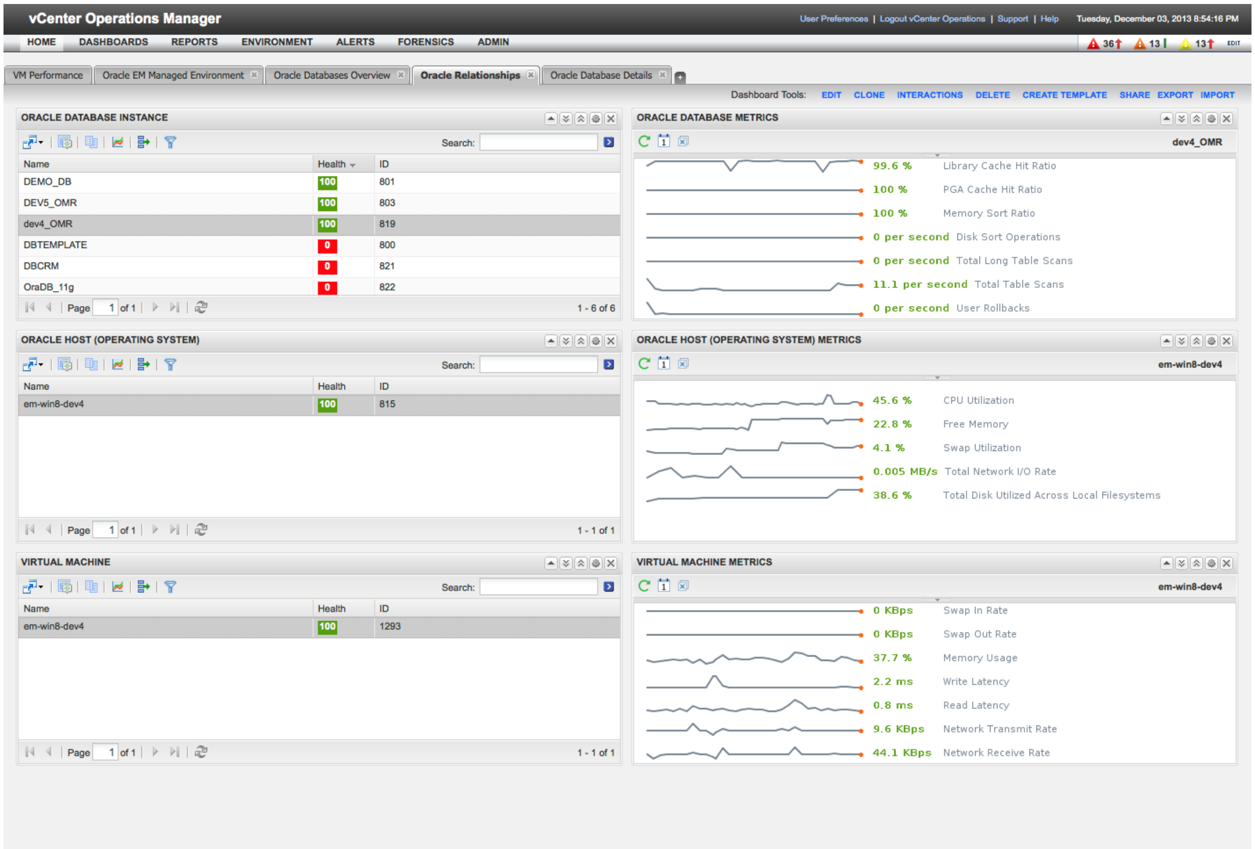
Figure 13. Oracle Database Details (2 of 3)



For each OEM target (for example, a database, an operating system, a WebLogic J2EE instance, and so on), the Oracle Enterprise Manager collects and stores up to hundreds of near-real-time health, performance, availability, and configuration metrics. These are the same metrics that are used within OEM to visualize Oracle environments and generate alerts on issues.

The Blue Medora - VMware vRealize Operations Management Pack for Oracle Enterprise Manager makes all of these Oracle-related metrics available within vRealize Operations. It allows the use of vRealize Operations capabilities to identify performance issues and gain deep insights into the health, risk and efficiency of your virtual and physical Oracle workloads (the underlying operating systems and in particular related VMware infrastructure, if applicable). This management pack helps customers recognize and address symptoms that threaten performance and availability, before the symptoms become issues that affect business users and the business itself.

Figure 14. Oracle Database Details (3 of 3)



For the Blue Medora - vRealize Operations Management Pack for Oracle Enterprise Manager, go to http://www.bluemedora.com/products/vrops-management-pack-oracle-em/.

11.2 Oracle Automatic Workload Repository

The Oracle Automatic Workload Repository (AWR) collects, processes, and maintains performance statistics for problem detection and self-tuning purposes. This data is both in memory and stored in the database. The gathered data can be displayed in both reports and views.

The statistics collected and processed by AWR include:

- Object statistics that determine both access and usage statistics of database segments
- Time model statistics based on time usage for activities, displayed in the V$SYS_TIME_MODEL and V$SESS_TIME_MODEL views
- Some of the system and session statistics collected in the V$SYSSTAT and V$SESSTAT views
- SQL statements that are producing the highest load on the system, based on criteria such as elapsed time and CPU time
- Active Session History (ASH) statistics, representing the history of recent sessions activity

The STATISTICS_LEVEL initialization parameter must be set to the TYPICAL or ALL to enable the Automatic Workload Repository. If the value is set to BASIC, you can manually capture AWR statistics using procedures in the DBMS_WORKLOAD_REPOSITORY package.

Snapshots are sets of historical data for specific time periods that are used for performance comparisons. By default, AWR automatically generates snapshots of the performance data once every hour and retains the statistics in the workload repository for 7 days. The data in the snapshot interval is then analyzed by the Oracle DBAs for performance troubleshooting. AWR compares the difference between snapshots to determine which SQL statements to capture based on the effect on the system load.

When looking at an AWR report, a good place to start is the **Top 5 Timed Foreground Events** section, near the top of the report. This gives you an indication of the bottlenecks in the system during this sample period.

Figure 15. Top 5 Timed Foreground Events Listing

Top 5 Timed Foreground Events

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
DB CPU		819		53.81	
log file sync	11,585	139	12	9.15	Commit
db file sequential read	75,111	59	1	3.85	User I/O
latch free	670	10	15	0.64	Other
latch: cache buffers lru chain	421	7	17	0.46	Other

After you have identified the top events, drill down to see what SQL and PL/SQL are consuming the majority of those resources. In the **Main Report** section, click the **SQL Statistics** link. The Oracle statistics metric indicates the top events and wait time that took place during the workload.

For Oracle statistics gathering and AWR report analysis, see the Oracle documentation at https://docs.oracle.com/database/121/TGDBA/gather_stats.htm#TGDBA167.

11.3 Linux Operating System sosreport

The Linux `sosreport` (RHEL / CentOS) command is a tool that collects configuration and diagnostic information from a Linux system. For instance, the running kernel version, loaded modules, and system and service configuration files. The command also runs external programs to collect further information, and stores this output in the resulting archive.

For any database troubleshooting issue, the Oracle AWR and Linux `sosreport` are collected as part of the diagnosis phase.

To run `sosreport`, the sos package must be installed. The package is part of the default group and will be installed automatically on most systems. After the package has been installed, issue the following command:

```
#sosreport
```

You can also use `sosreport` to enable diagnostics and analytical functions. To assist in troubleshooting a problem, `sosreport` records the information in a compressed file that you can send to a support representative.

For more information, see the Red Hat documentation at https://access.redhat.com/solutions/3592.

See also *Collecting diagnostic information for Linux operating systems* at http://kb.vmware.com/kb/2032614.

12. Backup and Recovery

12.1 Oracle Backup and Recovery Overview

The main purpose of a database backup and recovery strategy is to protect the database against data loss and reconstruct the database after data loss. Typical backup tasks performed by an Oracle DBA would include setting up the database environment for backup and recovery, setting up a backup schedule, monitoring the backup and recovery environment, and troubleshooting backup problems

A backup can be either a physical or a logical backup. Physical backups are physical copies of the database files which include data files, control files, and archive log files. Logical backups contain a logical copy of the data, such as tables, indexes, procedures, functions, and so on. You can use Oracle Data Pump to export logical data to binary files, which you can later import into the database.

There are levels of triggering Oracle database backups within the VMware environment:

- In guest Oracle backup using Oracle Recovery Manager (RMAN)
- VMware level backup using VMware vSphere Data Protection™ / VMware vSphere Data Protection Advanced
- Storage based backup tools
- vSphere Virtual Volumes using vSphere 6.0

vSphere recommends either using Oracle Recovery Manager (RMAN), storage-based backup tools, or vSphere Virtual Volumes using vSphere 6.0.

12.2 Oracle Recovery Manager (RMAN)

For implementing an effective Oracle database backup and recovery strategy, Oracle Recovery Manager (RMAN) is typically the preferred Oracle solution.

RMAN provides a common interface for backup tasks across different host operating systems, and offers several backup techniques not available through user-managed methods.

The method of deploying and using RMAN to backup an Oracle database does not change when virtualizing an Oracle database. It is same across both physical and virtualized environments.

For more information on Oracle Recovery Manager, see the Oracle documentation at https://docs.oracle.com/database/121/BRADV/toc.htm.

12.3 vSphere Data Protection

Any virtual machine VMDK can be backed up with VMware snapshot technology as long as it is not set to Independent-Persistent mode.

A virtual machine housing an Oracle database has two types of VMDKs—guest OS VMDK and the VMDKs housing the Oracle data files.

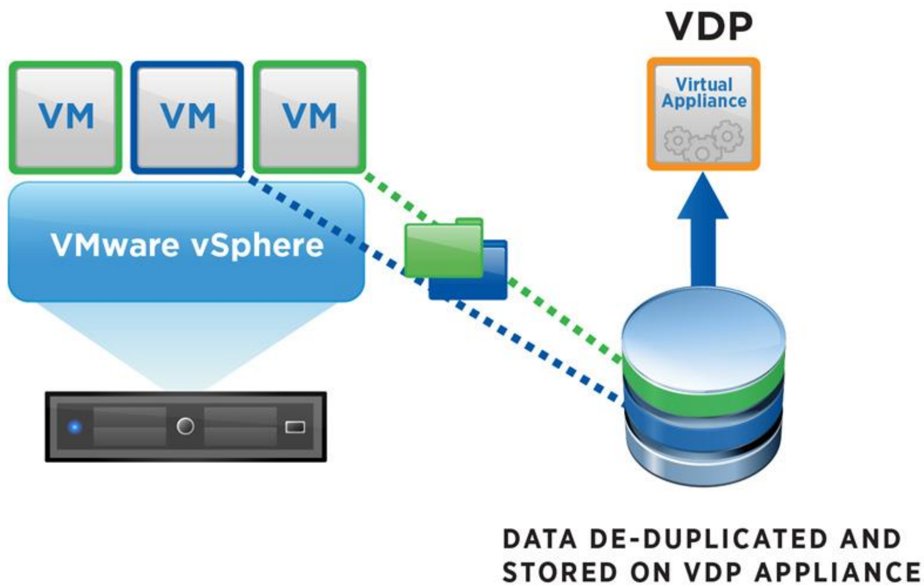
VMware does not recommend that you back up a high transactional, heavy I/O-centric Oracle database using VMware snapshot technology because, during the snapshot removal (consolidation), there is a brief stun moment. No activity is permitted against the virtual machine, which might result in performance issue and service disruptions.

For more information, see *A snapshot removal can stop a virtual machine for long time* (http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1002836).

You can, however, back up Oracle non-production databases (development, test, QA, pre-production, and so on) using VMware snapshot technology.

vSphere Data Protection is a backup and recovery solution and is deployed as a virtual appliance and managed using vSphere Web Client. vSphere Data Protection can back up and restore entire virtual machines. vSphere Data Protection leverages VMware snapshots through the vSphere Storage APIs – Data Protection product to back up virtual machines.

Figure 16. vSphere Data Protection



Before vSphere 6.0 and vSphere Data Protection 6.0, there were two editions of vSphere Data Protection—vSphere Data Protection, included with vSphere, and vSphere Data Protection Advanced, which was sold separately. With the release of vSphere Data Protection 6.0, all vSphere Data Protection Advanced functionality has been consolidated into vSphere Data Protection 6.0 and included with vSphere 6.0 Essentials Plus Kit and higher editions

For more information on vSphere Data Protection, see VMware documentation at https://www.vmware.com/products/vsphere/features/data-protection.html.

12.4 Storage-Based Backup Tools

Database operations, such as backup, data loads, and production refresh, are resource-intensive, time-consuming operations and cause performance issues and service disruption if they are not set up optimally. Above all, they are prone to human error.

To solve this problem, storage vendors have developed automated array-based applications that create copies of primary volumes more easily and quickly. Snapshots and clones are two methods used by the storage vendors for making point-in-time copies, and they use different replication mechanisms and have limits that are independent of each other.

A snapshot copy is a point-in-time file system image with pointers to the actual data blocks on disk. Clones, on the other hand, are a bit-for-bit replica of the primary disk volume. Different storage vendors have different ways of implementing the these technologies. For more details, see storage vendor-specific documentation.

You can back up Oracle databases on vSphere by choosing any of the storage technologies previously mentioned. Keep in mind that the snapshot or clone contains VMDKs of other virtual machines unless the VMware datastore is dedicated to a particular virtual machine.

Some examples of storage-based backup tools include EMC Replication Manager, EMC Avamar, and NetApp SnapManager for Oracle.

12.5 vSphere Virtual Volumes

Starting with vSphere 6.0, vSphere Virtual Volumes capabilities can be harnessed to provide effective and fast backup and recovery for any Business Critical Oracle environment.

Refer to Section 7.10, VMware vSphere Virtual Volumes for more information on VMware vSphere Virtual Volume technology and how it helps with the backup and recovery of Oracle databases.

12.6 Oracle Recovery Overview

As discussed in Section 12.1, Oracle Backup and Recovery Overview, the recommendation for backing up an Oracle database on vSphere is to either use any of the following three technologies:

- Oracle Recovery Manager (RMAN)
- Storage-based backup tools
- vSphere Virtual Volumes using vSphere 6.0

The recovery mechanism depends on the type of technology chosen for Oracle backup.

12.6.1 Recovery Using Oracle Recovery Manager (RMAN)

Using Oracle Recovery Manager (RMAN) to back up an Oracle database requires the database restore to be performed using the same backup mechanism tool.

DBAs should consider using an RMAN recovery catalog, a separate database used to record RMAN activity against one or more target databases. A recovery catalog preserves RMAN repository metadata if the control file is lost, making it much easier to restore and recover following the loss of the control file.

The process of restore and recovery from an RMAN catalog or database control file is the same regardless of physical or virtual environment.

For more information, see the Oracle documentation at https://docs.oracle.com/cd/E11882_01/backup.112/e10642/rcmquick.htm#BRADV89346.

12.6.2 Recovery Using Storage-Based Backup Tools / vSphere Virtual Volumes

With a storage-based recovery mechanism or vSphere Virtual Volumes, the granularity is only at the storage LUN/disk level. That is, the restore is an all or nothing restore.

In a restore scenario, the storage administrator restores the backup of the datastore at a storage level using snapshots, clones or tape-level backup.

In the case of vSphere Virtual Volumes, the VMware administrator triggers the restore of the production virtual volume from a backup snapshot or clone of the virtual volume.

In case of a full database restore, the Oracle instance must be shut down before the storage-based or vSphere Virtual Volume based recovery mechanism is triggered. For a partial database restore, for example, tablespace level, the database can be online.

The Oracle DBA can then proceed to recover the database using the standard Oracle recovery process.

The process of recovering an Oracle database does not change when you virtualize an Oracle database. It is the same in both physical and virtualized environments.

See the Oracle document for user-managed recovery process at https://docs.oracle.com/database/121/BRADV/osadvsce.htm#BRADV018.

13. High Availability

13.1 General HA Considerations

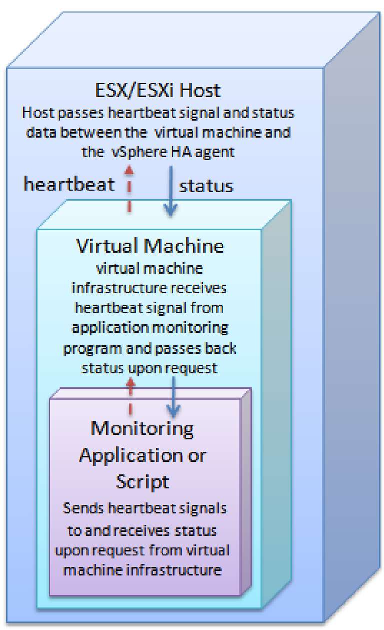
vSphere HA delivers the availability required by most applications running in virtual machines, independent of the operating system and application running in it. HA provides uniform, cost-effective failover protection against hardware and operating system outages within your virtualized IT environment. The following is a list of vSphere HA attributes:

- Monitors vSphere hosts and virtual machines to detect hardware and guest operating system failures.
- Restarts virtual machines on other vSphere hosts in the cluster without manual intervention when a server outage is detected.
- Reduces application downtime by automatically restarting virtual machines upon detection of an operating system failure.

vSphere HA for ESXi hosts in a cluster also provides protection for a guest OS and its applications by restarting the virtual machine if a guest OS or application failure occurs. vSphere HA provides this reset capability through two different mechanisms:

- VM monitoring – Guest OS heartbeats issued by the VMware Tools process.
- Application monitoring – Heartbeats issued by a program that uses the HA Application Monitoring SDK to communicate with the VMware Tools process and the vSphere HA agent. This mechanism involves local monitoring by the program to avoid the overhead of sending messages to and from vCenter Server.

Figure 17. Heartbeat and Status Signals



The following table provides a high availability best practice.

Table 13. High Availability Best Practice Recommendation

Recommendation	Justification
Do not disable vSphere HA.	vSphere HA monitors vSphere hosts and virtual machines to detect hardware and guest operating system failures and restarts virtual machines on other vSphere hosts in the cluster without manual intervention when a server outage is detected.

13.2 Oracle Real Application Cluster

Oracle Real Application Clusters (RAC) is a shared cache clustered database architecture that overcomes the limitations of traditional shared-nothing and shared-disk architectures. The Oracle database with the Oracle Real Application Clusters (RAC) option allows running multiple database instances on different servers in the cluster against a shared set of data files, also known as the database. Deep dive discussion of Oracle Real Application Clusters (RAC) on vSphere is beyond the scope of this section.

The need to deploy Oracle RAC for an application depends on the various requirements of the application being supported by Oracle RAC. Applications can benefit from Oracle RAC if they can identify with at least one of these RAC evaluation criteria:

- There is an explicit requirement in the SLA that the database must not have more than a few minutes of downtime in case of hardware failure.
- Current Oracle database installations take advantage of rolling upgrades offered by RAC. Because there are multiple instances in a RAC cluster, many patches can be completed so that the database does not experience downtime. Examples of patches that are *not* rolling upgrade-eligible include major patch sets that update the database data dictionary and force a downtime for the entire RAC cluster, and Critical Patch Updates (CPUs), which almost always require database downtime.
- Oracle uses Oracle Notifications Services (ONS) to allow application middle tiers to take code branches depending on various RAC cluster notifications. Fast Connection Failover (FCF), which is not connection pooling, is one of the ONS capabilities. This and other capabilities can be leveraged through the insertion of code hooks and dependent logic, as well as through middle-tier configuration in many cases.

Single-instance Oracle on vSphere HA is not equivalent to RAC. With vSphere HA, there is no client side configuration, and no planning for failover services. The modification to an application is similar if it must gracefully handle a loss of database session/connection and automatically retry the impacted logic. In the event of a host failure, vSphere HA can have the guest running again on a surviving member of the cluster within a matter of minutes. Unlike RAC, however, vSphere HA cannot allow in-flight transactions to continue on a surviving node because the entire single-instance database has crashed.

If an enterprise has sufficient business and technical justification to adopt or keep RAC, consider the layering of RAC on vSphere. The underlying vSphere infrastructure is invisible to the RAC stack. Oracle Grid Infrastructure does not use SCSI reservations. Oracle relies on its own software mechanisms to manage the integrity of the shared storage in a RAC configuration. Ultimately, RAC concerns IP, ports, and voting disk I/O. As such, vSphere significantly augments RAC’s functionality while crippling none of it.

For information about setting up shared VMDKs for an Oracle RAC cluster, see *Enabling or 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).

For best practices for deploying Oracle RAC on VMware with respect to compute, memory, networking and storage, see Section 5, Virtual CPU Guidelines, Section 6, Storage Guidelines, Section 7, Storage Guidelines, and Section 8, Networking Guidelines.

See the following Oracle documentation for setting up an optimal Oracle Real Application cluster:

- *Real Application Clusters Administration and Deployment Guide*

  https://docs.oracle.com/database/121/RACAD/admcon.htm#RACAD1111

- *Oracle Databases on VMware RAC Deployment Guide*

  https://www.vmware.com/files/pdf/solutions/oracle/Oracle_Databases_VMware_RAC_Deployment_Guide.pdf.

See also the *Oracle Real Application Clusters on VMware Virtual SAN* reference architecture for more details about deploying Oracle RAC on Virtual SAN (http://www.vmware.com/files/pdf/products/vsan/vmware-oracle-real-application-clusters-on-vmware-virtual-san-reference-architecture.pdf).

13.3 Oracle RAC One Node

The Oracle database with the Oracle Real Application Clusters (RAC) One Node option benefits from the same infrastructure used for Oracle RAC. Unlike Oracle RAC, Oracle RAC One Node normally runs only one instance against a shared set of data files, also known as the database. This database is fully Oracle RAC-enabled, but does not span multiple hardware systems at the same time.

Instead, the Oracle RAC One Node database instance will fail over to another server in the cluster should a server, instance, or a related and monitored component on this server fail. For cases of planned downtime such as operating system or database patching, Oracle RAC One Node provides the Online Database Relocation feature, which allows relocating a database from one server to another without interrupting the database service.

The caveats and best practice recommendations for Oracle RAC One Node are the same as for Oracle RAC.

13.4 Non Oracle Clusterware

Non Oracle Clusterware (for example, Veritas Clustering, Red Hat Cluster Services, HP Service guard Clustering) software that support clustered Oracle databases are also supported on the vSphere platform.

Refer to vendor-specific documentation on support and best practices for deploying the vendor-specific clustering software on vSphere.

Refer to the following documentation for information about vendor-specific implementation of clustering software on the vSphere platform:

- *Virtualization Support for RHEL High Availability and Resilient Storage Clusters*

  https://access.redhat.com/articles/29440

- *Veritas Cluster Server Implementation Guide*

  https://sort.symantec.com/public/documents/sf/5.1/esx/pdf/vcs_implementation.pdf

- *HPE Serviceguard for Linux with VMware virtual machines*

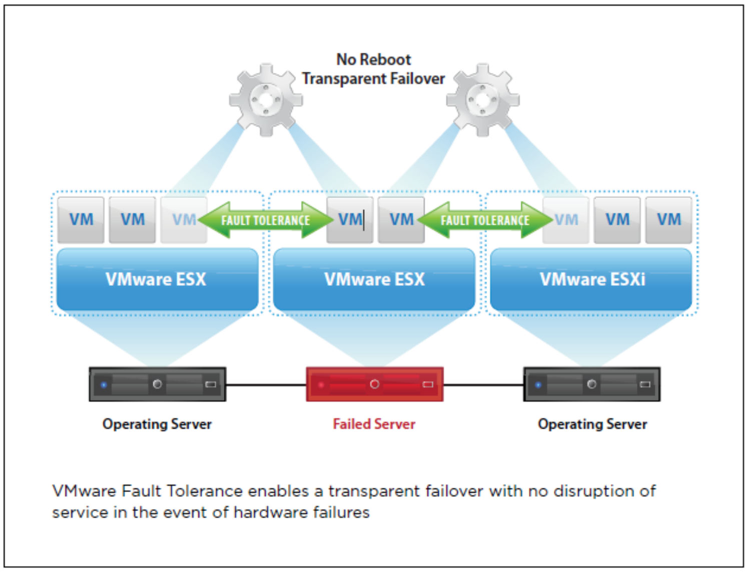
  http://www8.hp.com/h20195/v2/GetPDF.aspx%2F4AA4-2016ENW.pdf

13.5 VMware vSphere Fault Tolerance

VMware vSphere Fault Tolerance (FT) provides continuous availability for applications (with up to four virtual CPUs) in the event of server failures by creating a live shadow instance of a virtual machine that is always up-to-date with the primary virtual machine. In the event of a hardware outage, vSphere FT automatically triggers failover—ensuring zero downtime and preventing data loss.

After failover, vSphere FT automatically creates a new, secondary virtual machine to deliver continuous protection for the application.

Figure 18. vSphere Fault Tolerance



Typical Oracle database virtual machines have more than four vCPUs, and therefore, are not a good fit for the vSphere FT use case. However, see the VMware vSphere 6.0 documentation for best practices if you need to set up vSphere Fault Tolerance for Oracle database virtual machines (http://pubs.vmware.com/vsphere-60/index.jsp).

13.6 VMware vSphere Metro Storage Cluster

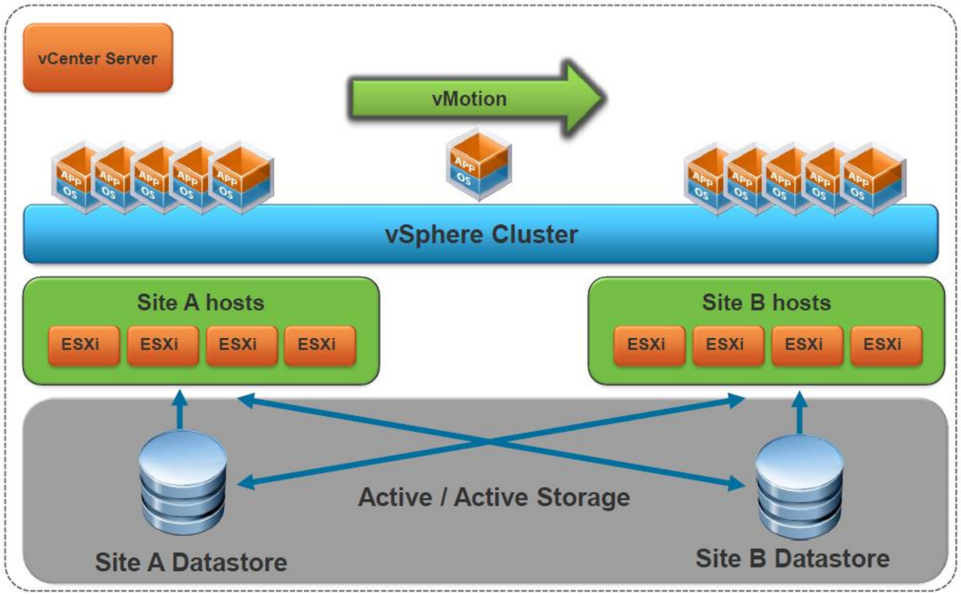
A vSphere *metro storage cluster* configuration is a vSphere certified solution that combines replication with array-based clustering. This solution is typically deployed in environments where the distance between data centers is limited, often metropolitan or campus environments and where disaster avoidance and downtime avoidance is a key requirement.

vSphere metro storage cluster infrastructures are implemented with a goal of reaping the same benefits that high availability clusters provide to a local site, in a geographically dispersed model with two data centers in different locations.

This infrastructure is essentially a stretched cluster. The architecture is built on the premise of extending what is defined as “local” in terms of network and storage to enable these subsystems to span geographies, presenting a single and common base infrastructure set of resources to the vSphere cluster at both sites.

A vSphere metro storage cluster stretches storage and the network between sites. The primary benefit of a stretched cluster model is that it enables fully active and workload-balanced data centers to be used to their full potential while gaining the capability to migrate VMs with vSphere vMotion and vSphere Storage vMotion between sites to enable on-demand and nonintrusive mobility of workloads. The capability of a stretched cluster to provide this active balancing of resources should always be the primary design and implementation goal. Although often associated with disaster recovery, vSphere metro storage cluster infrastructures are not recommended as primary solutions for pure disaster recovery.

Figure 19. vSphere Metro Storage Cluster



Stretched cluster solutions offer the following benefits:

- • Workload mobility
- • Cross-site automated load balancing
- • Enhanced downtime avoidance
- • Disaster avoidance

Storage virtualization can be provided by deploying vendor specific products (for example, EMC VPLEX, HP Peer Persistence, Netapp MetroCluster, and IBM SVC). Oracle Automatic Storage Manager (ASM) also provides application-based storage virtualization using ASM mirroring mechanism.

Refer to the following documentation for details on various implementations:

- *Implementing VMware vSphere Metro Storage Cluster (vMSC) using EMC VPLEX*

  http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2007545

- I*mplementing vSphere Metro Storage Cluster (vMSC) using HP 3PAR Peer Persistence*

  http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2055904

- *VMware support with NetApp MetroCluster*

  http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1001783

- *Implementing vSphere Metro Storage Cluster using IBM System Storage SAN Volume Controller*

  http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2032346

- *Oracle RAC and Oracle RAC One Node on Extended Distance (Stretched) Clusters*
  http://www.oracle.com/technetwork/products/clustering/overview/extendedracversion11-435972.pdf

13.7 Virtual SAN

Virtual SAN is the VMware software-defined storage solution for hyper-converged infrastructure, a software-driven architecture that delivers tightly integrated computing, networking, and shared storage from a single virtualized x86 server. Virtual SAN delivers high performance, highly resilient shared storage by clustering server-attached flash devices and hard disks (HDDs).

See Section 7.11, VMware Virtual SAN for more information on VMware Virtual SAN (hybrid and All-Flash).

14. Disaster Recovery

14.1 General DR Considerations

Disaster recovery (DR) and business continuity refers to an organization's ability to recover from a disaster or an unexpected event and resume normal business operations. In such events, organizations have disaster recovery plan in place that outlines how a recovery will be accomplished. The key to successful disaster recovery is to have a plan well before disaster ever strikes. Typically organizations perform a regular audit of their disaster recovery plan to ensure its current validity.

A site is designated as a DR site that an organization can move to after a disaster if the current production facility is unusable. The type of site is often determined by the results of a cost-benefit analysis as well as the needs of the organization.

The type of replication of data from the primary to the DR site often depends on the business service level agreement (SLA), the network bandwidth between the two sites, and the software currently in use to perform the replication.

Replication can be broadly performed on three different levels:

- Application
- vSphere
- Storage

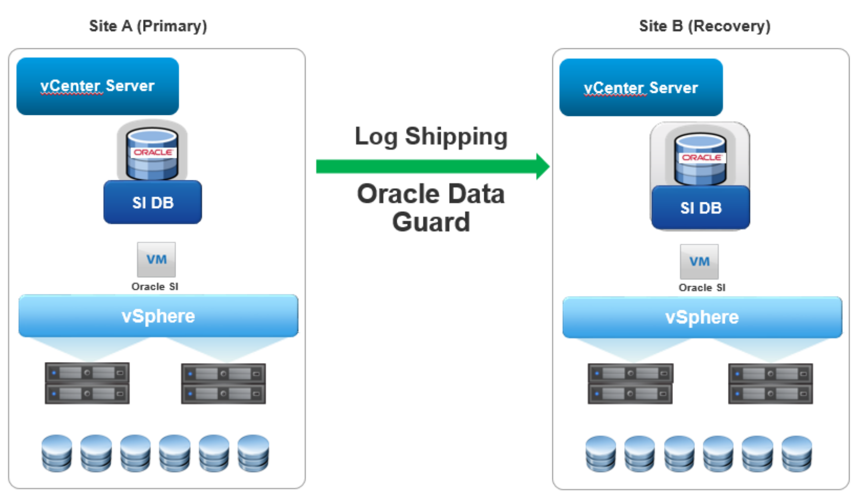
14.2 Application-Based Replication

At the top of the IT stack is application that the end user interfaces with and what is most likely to be down in the event of an outage. Because it is the application which is being protected against a disaster, it makes a sense to leverage any built-in options for data replication on an application level.

14.2.1 Oracle Data Guard

Oracle Data Guard is the premier Oracle data replication tool for disaster recovery. Oracle Data Guard ensures data protection and disaster recovery for Oracle Enterprise data. Oracle Data Guard provides a comprehensive set of services that create, maintain, manage, and monitor one or more standby databases to enable production Oracle databases to survive disasters and data corruptions.

Figure 20. Oracle Data Guard Based Replication of Oracle Databases



Oracle Data Guard maintains these standby databases as copies of the production database. If the production database becomes unavailable because of a planned or an unplanned outage, Oracle Data Guard can switch any standby database to the production role, minimizing the downtime associated with the outage. A standby database can be either a single-instance Oracle database or an Oracle RAC database.

Two types of standby databases are implemented as follows:

- Physical standby database which provides a physically identical copy of the primary database, with on-disk database structures that are identical to the primary database on a block-for-block basis. A physical standby database is kept synchronized with the primary database, through Redo Apply, which recovers the redo data received from the primary database and applies the redo to the physical standby database.
- Logical standby database which contains the same logical information as the production database, although the physical organization and structure of the data can be different. The logical standby database is kept synchronized with the primary database through SQL Apply, which transforms the data in the redo received from the primary database into SQL statements and then executes the SQL statements on the standby database.

Oracle Active Data Guard increases performance, availability and data protection wherever it is used for real-time data protection and availability. An Oracle Active Data Guard standby database can be used to offload a primary database of reporting, ad-hoc queries, data extracts, and backups, making it a very effective way to insulate interactive users and critical business tasks on the production system from the overhead of long-running operations.

Oracle Active Data Guard provides read-only access to a physical standby database while it is synchronized with a primary database, enabling minimal latency between reporting and production data. Oracle Active Data Guard automatically repairs physical corruption on either the primary or standby database, increasing availability and maintaining data protection at all times.

The steps for setting up an Oracle Data Guard on vSphere are the same as setting it up on a physical environment. VMware recommends referring to Oracle best practices as documented in the Oracle documentation at https://docs.oracle.com/database/121/SBYDB/concepts.htm#SBYDB00010.

14.2.2 VMware Continuent

VMware Continuent® provides data replication between relational databases, and to data warehouses and analytics engines. In addition, VMware Continuent provides globally redundant disaster recovery, commercial-grade high availability, and performance scaling.

VMware Continuent allows enterprises running business-critical database applications to affordably achieve commercial-grade high availability (HA), globally redundant disaster recovery (DR), and performance scaling.

VMware Continuent makes it simple to create database clusters in the cloud or in your private data center to keep the data available even when systems fail. In addition, VMware Continuent provides data replication from relational databases to high-performance NoSQL and data analytics engines to derive insight from big data for better business decisions.

Functionality is provided across four primary products:

- VMware Continuent for Clustering – Provides full clustering support, including load balancing, failover, and multimaster, multisite deployments.
- VMware Continuent for Disaster Recovery – Enables replication to another server or site to handle disaster recovery scenarios.
- VMware Continuent for Replication – Provides core replication between MySQL, MariaDB, and Percona servers and replication to and from Oracle.

- VMware Continuent for Analytics and Big Data – Provides replication from MySQL to various Hadoop distributions (including Pivotal HD, MapR, HortonWorks, and Cloudera), HP Vertica, and Amazon Redshift.

Figure 21. VMware Continuent Features

FEATURE	VMWARE CONTINUENT FOR CLUSTERING	VMWARE CONTINUENT FOR DISASTER RECOVERY	VMWARE CONTINUENT FOR REPLICATION	VMWARE CONTINUENT FOR ANALYTICS AND BIG DATA
Flexible MySQL Clustering	Yes			
Zero-Downtime	Yes	Yes		
Automatic Failover	Yes			
Multi-Master, Multi-Site	Yes	Yes		
Disaster Recovery		Yes		
Oracle Replication			Yes	
Replication To Real-Time Analytics And Big Data				Yes
Improved Performance	Yes	Yes		
Supports MySQL, MariaDB and Percona Server	Yes	Yes	Yes	Yes

For more information see the VMware Continuent documentation at https://www.vmware.com/products/continuent.

14.2.3 Third-Party Replication Tools

Apart from replication products like Oracle Data Guard and VMware Continuent, there are other replication products that can be used for replicating Oracle database between sites.

Oracle GoldenGate is a logical data replication tool that does not have any built-in mechanism to facilitate data comparison and address out-of-sync data between source and target database. It requires a certain level of expertise to support the GoldenGate implementations, which increases the resource cost.

Oracle Streams is another data replication tool which is soon to be obsolete, with GoldenGate replacing it.

Dell Shareplex is a data replication tool primarily designed for replicating data across different Oracle databases. It provides a wide range of real-time data replication features. Shareplex replicates only the changes made to the business data, without interrupting the business processing which makes the process very lightweight and fast.

Because Shareplex replicates the changes as they happen; the replication site is always up-to-date and in sync with the business data. The replication sites maintained by Shareplex can be utilized for query offloading as well as for reporting (because the sites are up-to-date). This in turn improves the performance of the OLTP databases. Shareplex data replication supports Oracle Enterprise as well as Standard Edition.

For more information about Dell Shareplex, see http://software.dell.com/products/shareplex/.

14.3 VMware vSphere Replication

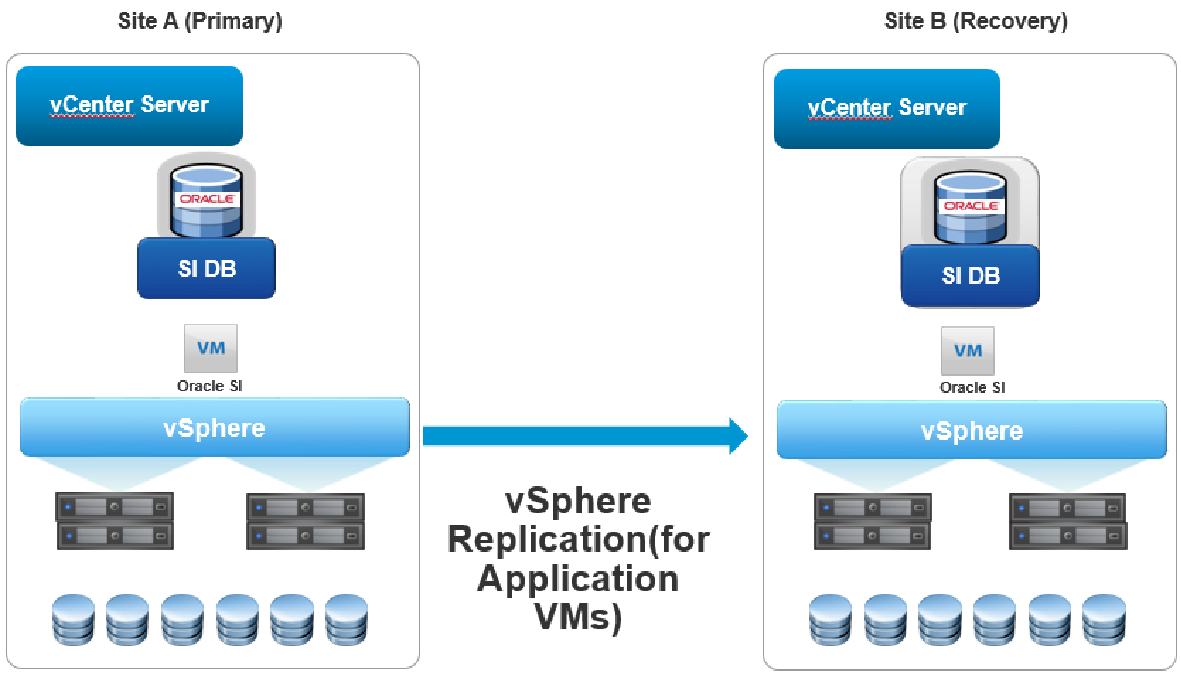
VMware vSphere Replication™ is a virtual machine data protection and disaster recovery solution that is fully integrated with vCenter Server and vSphere Web Client, providing host-based, asynchronous replication of virtual machines.

Because vSphere Replication is host-based replication, it is independent of the underlying storage and it works with a variety of storage types including Virtual SAN, traditional SAN, NAS, and direct-attached storage (DAS), and enables virtual machine replication between same or heterogeneous storage types.

The following are some of vSphere Replication features and benefits:

- Protects nearly any virtual machine regardless of operating system and applications
- Only changes are replicated, which improves efficiency and reduces network utilization
- Recovery point objectives (RPOs) range from 15 minutes to 24 hours and can be configured on a per-virtual machine basis

Figure 22. vSphere Based Replication of Virtual Machine Containing Oracle Databases



vSphere Replication does not use VMware snapshots to perform replication. After replication has been configured for a virtual machine, vSphere Replication begins the initial full synchronization of the source virtual machine to the target location. A copy of the VMDKs to be replicated can be created and shipped to the target location and used as “seeds,” reducing the time and network bandwidth consumed by the initial full synchronization. After the initial full synchronization, changes to the protected virtual machine are tracked and replicated on a regular basis. The transmissions of these changes are referred to as “lightweight delta syncs.” The transmission frequency is determined by the RPO that is configured for the virtual machine. A lower RPO requires more-frequent replication.

vSphere Replication cannot provide the application-level replication that Oracle Data Guard provides because it is not a native Oracle tool.

vSphere Replication can be used to replicate virtual machines housing Oracle databases based on the RPO settings. The Oracle database on the target site can be recovered through an Oracle crash consistent recovery because the database is crash consistent at the point of the replication cycle.

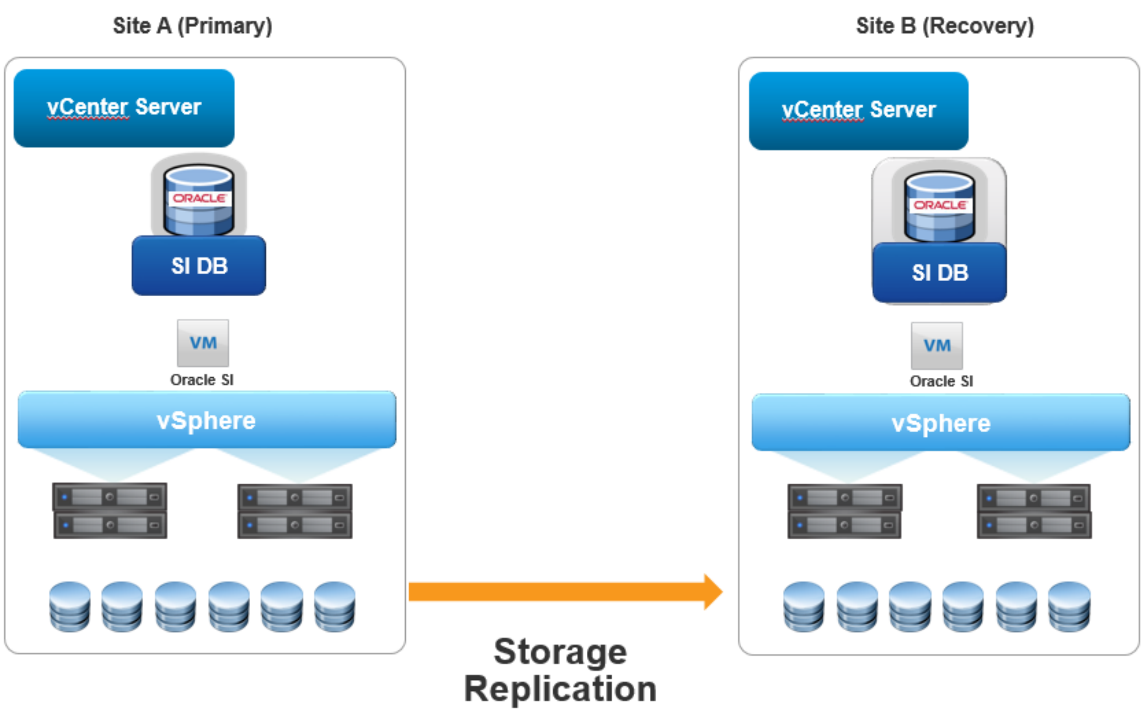
Using vSphere Replication consistency groups and quiescing the Oracle database before every replication cycle is not acceptable in any Oracle production database environment because it introduces database performance issues. This results in business SLA deficiencies and therefore many DBAs choose to use the crash consistent copy of the database on the target site.

14.4 Array-Based Replication

With array-based replication, compatible storage arrays use built-in software to automatically copy data from one storage array to another across sites or within the same site. Array-based replication software can be run synchronously or asynchronously replicating data at the logical unit number (LUN) or volume block level.

Some of the notable storage vendor replication technologies include EMC Recover Point, EMC SRDF, NetApp SnapMirror, HP StorageWorks XP, and IBM Global Mirror.

Figure 23. Storage-Based Replication of Virtual Machine Containing Oracle Databases



With storage-based replication on vSphere, the replication is done at the actual VMware datastore level. The backup might contain VMDKs from other virtual machines that share the same data store, unless the data store is dedicated to an Oracle virtual machine.

Storage-based replication will result in a crash-consistent copy at the target side which can be recovered through an Oracle crash-consistent recovery because the database is crash consistent at the point of the replication cycle.

Using storage-based consistency groups and quiescing the Oracle database before every replication cycle is not acceptable in any Oracle production database environment because it introduces database performance issues. This results in business SLA deficiencies and therefore many DBAs choose to use the crash consistent copy of the database on the target site.

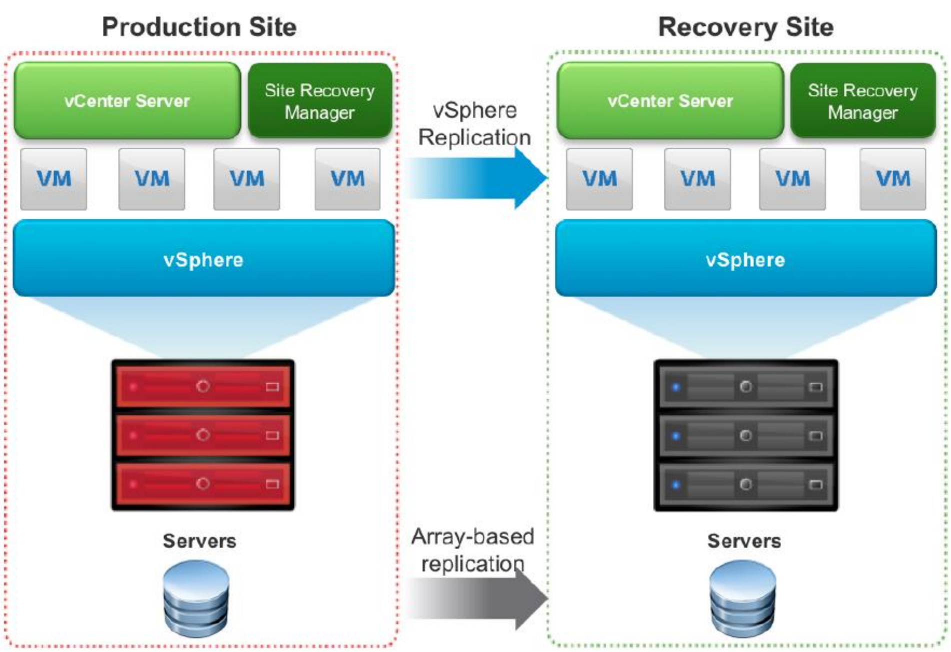
14.5 VMware Site Recovery Manager

VMware Site Recovery Manager™ is the industry-leading solution to enable application availability and mobility across sites in private cloud environments. Taking full advantage of the encapsulation and isolation of virtual machines, Site Recovery Manager enables simplified automation of disaster recovery to meet recovery time objectives (RTOs), reduce costs associated with business continuity plans, and achieve low-risk and predictable results for recovery of a virtual environment.

Site Recovery Manager provides unique capabilities to create, maintain, and non-disruptively test disaster recovery plans without the need for manual runbooks. The ability to automate the disaster recovery planning, maintenance, and testing process enables significant operational efficiencies.

Site Recovery Manager utilizes either vSphere Replication, array-based replication or stretched storage for transferring data between sites and automating the process of migrating, recovering, testing, re-protecting and failing-back virtual machine workloads.

Figure 24. Site Recovery Manager Components



Site Recovery Manager servers coordinate the operations of the vCenter Server at two sites. This is so that as virtual machines at the protected site are shut down, copies of these virtual machines at the recovery site start up. By using the data replicated from the protected site, these virtual machines assume responsibility for providing the same services.

Migration of protected inventory and services from one site to the other is controlled by a recovery plan that specifies the order in which virtual machines are shut down and started up, the resource pools to which they are allocated, and the networks they can access. Site Recovery Manager enables the testing of recovery plans, using a temporary copy of the replicated data, and isolated networks in a way that does not disrupt ongoing operations at either site. Multiple recovery plans can be configured to migrate individual applications and entire sites, providing finer control over which virtual machines are failed over and failed back. This also enables flexible testing schedules.

For more information about VMware Site Recovery Manager, see the documentation at https://www.vmware.com/support/pubs/srm_pubs.html.

In conclusion, the choice of Oracle database replication depends on many factors, often based on the service level agreement (SLA), the level of effort needed to recover in case of a disaster, the cost of such replication technology, and the infrastructure in place

15. VMware Engineered Systems

Hyper-converged infrastructure (HCI) leverages the hypervisor to deliver compute, networking and shared storage from a single x86 server platform. The software-driven architecture allows the convergence of physical storage onto commodity x86 servers, enabling a building block approach with scaleout capabilities. The use of commodity x86 server and storage hardware allows data centers to operate with agility on a highly scalable, cost-effective infrastructure.

VMware EVO represents a new family of “evolutionary” HCI offerings from VMware. VMware EVO is a preconfigured, pre-integrated SDDC stack that is offered by several VMware hardware partners which include Dell, EMC, Fujitsu, Inspur, NetOne, and SuperMicro.

VMware provides the software and specifications to the hardware partners, who will ship and support the VMware EVO solutions. The hardware partners will sell VMware EVO-based hardware solutions. VMware EVO solutions leverage vSphere as well as Virtual SAN and local-attached storage. Currently there are two EVO solutions, VMware EVO:RAIL™ and the tech preview for VMware EVO:RACK™.

15.1 EVO:RAIL

EVO:RAIL combines VMware compute, network, storage, and management resources into a hyper-converged infrastructure appliance to create a simple, easy-to-manage, all-in-one solution for all your virtualized workloads, including tier-1 production and mission-critical applications. Offered by selected partners, EVO:RAIL is backed by a single point of contact for software and hardware support.

EVO:RAIL delivers the following benefits:

- Radically simple: Meet accelerated business demands by simplifying infrastructure design with predictable sizing and scaling, and by streamlining purchase and deployment. With EVO:RAIL, you go from initial power-on to VM deployment in minutes, and easily scale to 16 appliances. Manage your VMs, appliance, and cluster from a single pane of glass.
- Resilient and secure: Designed with four independent hosts, distributed Virtual SAN datastore, built-in VM security policies, and leveraging best-in-class VMware availability capabilities enables zero application downtime during planned maintenance or during disk, network, or host failure.
- Trusted and proven: Powered by the VMware industry-leading infrastructure virtualization technology, including vSphere, vCenter Server and Virtual SAN, EVO:RAIL delivers the first hyper-converged infrastructure appliance designed, powered, and validated by VMware.

EVO:RAIL capabilities are as shown in the following table snapshot.

Figure 25. EVO:RAIL Capabilities

ELEMENT	DESCRIPTION	CAPABILITY
EVO:RAIL Appliance	Nodes per Appliance	4
	Max VM’s per Appliance	800*
EVO:RAIL Cluster	Max Appliances per cluster	16
	Max VM’s per Cluster	12,800*
EVO:RAIL Hardware	Hardware form factor	2U
	Configurations	Flexible hardware configurations with up to 100 options

*Depending on your hardware configuration and VM profile.

EVO:RAIL software is integrated into a pre-specified and optimized 2U/4 node hardware platform.

EVO:RAIL software consists of the EVO:RAIL Engine, vSphere Enterprise Plus, Virtual SAN, vCenter Server, and VMware vRealize Log Insight™.

EVO:RAIL 2U chassis are tightly prescribed by VMware, and are pre-built and tested by EVO:RAIL partners. Each appliance has four servers to ensure resiliency, and is optimized for performance and flexibility.

Figure 26. EVO:RAIL Hardware System Details

EVO:RAIL Hardware	Processor	8 x Intel E5 Processors, Ivy Bridge or Haswell (48, 64, 80, 96 or 144 cores)
	Memory	512, 768, 1024, 1536 or 2048 GB
	Storage	16 or 27.2 TB Hybrid Storage (HDD/SSD)
	Network	• 8 x 10 GbE NIC port (copper or fiber) • 4 x 1 GbE for BMC port

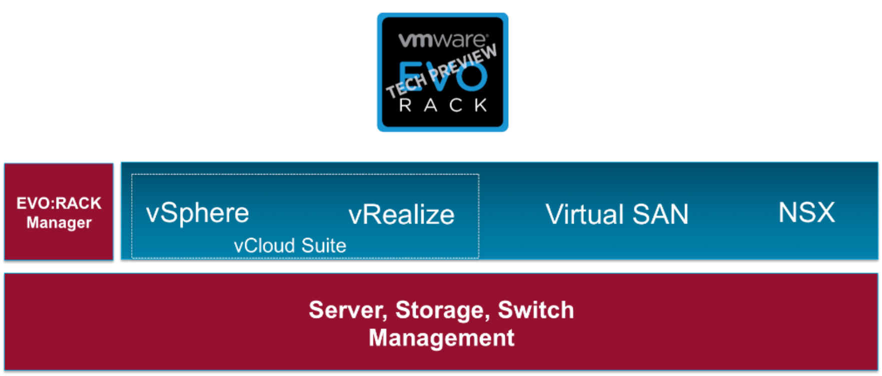
EVO:RAIL is interoperable and optimized for the entire VMware software stack, including VMware NSX®, and vRealize Operations. EVO:RAIL includes vSphere Data Protection and vSphere Data Protection Advanced for backup protection, and vSphere Replication for replication. EVO:RAIL data can also be protected by any third-party back-up product that supports vSphere Enterprise Plus.

15.2 EVO:RACK

EVO:RACK, a hyper-converged infrastructure project, simplifies how companies buy, deploy, and operate software-defined data centers, helping IT organizations rapidly provision applications and services at data center scale.

Designed as a hyper-converged infrastructure solution that can scale to tens of racks; EVO:RACK meets the increasing demands of private clouds at medium-to-large enterprises. It can run on a range of pre-integrated hardware configurations ranging from Open Compute Project-based hardware designs to industry-standard OEM servers and converged infrastructure.

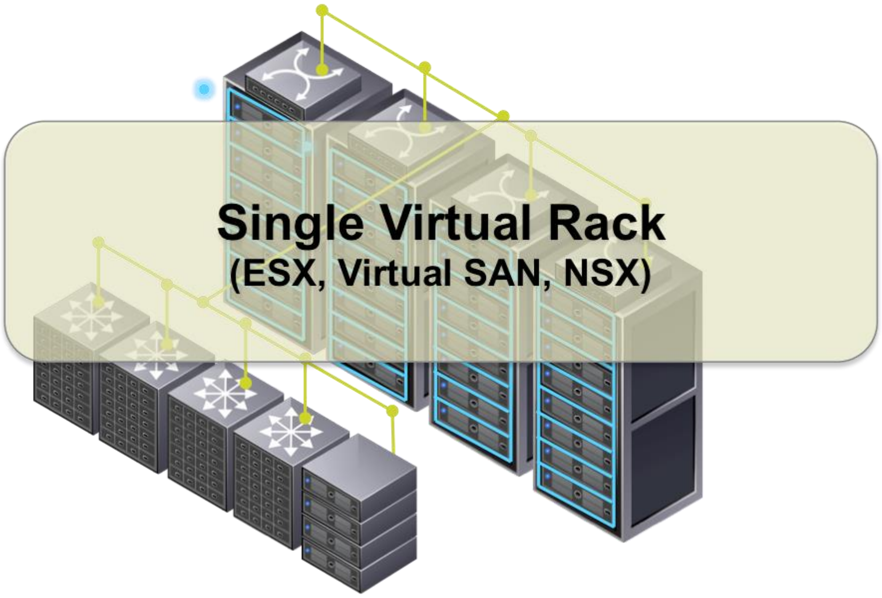
Figure 27. EVO:RACK Configuration



EVO: RACK also leverages the full line-up of VMware software solutions for the SDDC including, VMware vCloud Suite®, Virtual SAN, and VMware NSX. EVO:RACK is designed to support integrated virtual and physical networking, and brings it all together through EVO software specifically developed to simplify the deployment and ongoing lifecycle management of the SDDC software. The EVO software creates a scalable management solution across a distributed, multi-rack hyper-converged cloud infrastructure.

This approach provides a virtual abstraction (across racks and other form factors) that pools compute/storage/network and uses this abstracted resource pool as a unit of SDDC instantiation and operation. The result is a dramatically simplified model for deployment, configuration, and provisioning.

Figure 28. EVO:RACK Model



EVO: RACK provides streamlined and automated lifecycle management of the SDDC components including non-disruptive patching and upgrading of software. The initial target for such an HCI is an environment with workloads spanning IaaS, and VDI with plans to extend it to support PaaS and Big Data workloads in the future.

16. Summary

The best practices and guidelines discussed in this document are summarized in this section.

Recommendations
Create a computing environment optimized for vSphere.
Create golden images of optimized operating systems using vSphere cloning technologies.
Upgrade to the latest version of ESXi and vSphere.
Allow vSphere to choose the best virtual machine monitor based on the CPU and guest operating system combination.
Refer to BIOS settings table for optimal application performance.
Enable hardware assisted CPU/memory virtualization.
Do not over-allocate vCPUs. Try to match the exact workload.
Enable hyperthreading for Intel Core i7 processors.
Keep the default of one core per socket for vNuma to match physical NUMA topology, and try to align VMs with physical NUMA boundaries.
Leave the latency sensitive setting at its default value.
Set memory reservations equal to the sum total of the size of the Oracle SGA, the Oracle PGA , the Oracle background processes and operating system used memory.
Use huge memory pages.
Do not turn off ESXI memory management mechanisms unless directed by VMware support.
Enable jumbo frames for IP-based storage using iSCSI and NFS.
Create dedicated datastores to service database workloads.
Use vSphere VMFS for Oracle database deployments.
Align VMFS properly.
Use Oracle Automatic Storage Management.
Use your storage vendor’s best practices documentation when laying out the Oracle database.
Avoid silos when designing the storage architecture.

Recommendations
Use paravirtualized SCSI adapters for Oracle data files with demanding workloads.
Use the VMXNET family of paravirtualized network adapters.
Separate infrastructure traffic from virtual machine traffic for security and isolation.
Use NIC teaming for availability and load balancing.
Take advantage of Network I/O Control to converge network and storage traffic onto 10 GbE.
Enable jumbo frames for Oracle interconnect traffic.
Disable unnecessary foreground and background processes within the guest operating system to save CPU cycles.
To minimize time drift in virtual machines, follow guidelines in relevant VMware Knowledge Base articles.
Use vCenter and/or the esxtop/resxtop utility for performance monitoring in the virtual environment.

Success stories are available at http://www.vmware.com/company/customers.

17. References

- *VMware Compatibility Guide*

  http://www.vmware.com/resources/compatibility/search.php
- vSphere Performance Best Practices guides

  http://www.vmware.com/files/pdf/techpaper/VMware-PerfBest-Practices-vSphere6-0.pdf

  https://www.vmware.com/pdf/Perf_Best_Practices_vSphere5.5.pdf
- *Deploying Extremely Latency-Sensitive Applications in VMware vSphere 5.5*

  http://www.vmware.com/files/pdf/techpaper/latency-sensitive-perf-vsphere55.pdf
- *Best Practices for Performance Tuning of Latency-Sensitive Workloads in vSphere VMs*

  http://www.vmware.com/files/pdf/techpaper/VMW-Tuning-Latency-Sensitive-Workloads.pdf
- *Virtualizing Performance-Critical Database Applications in VMware vSphere 6.0*

  http://www.vmware.com/files/pdf/products/vsphere/vmware-database-apps-perf-vsphere6.pdf
- *Enabling or disabling simultaneous write protection provided by VMFS using the multi-writer flag*

  http://kb.vmware.com/kb/1034165
- *Oracle Databases on VMware RAC Deployment Guide*

  http://www.vmware.com/files/pdf/solutions/oracle/Oracle_Databases_VMware_RAC_Deployment_Guide.pdf
- *Oracle Real Application Clusters on VMware Virtual SAN*

  http://www.vmware.com/files/pdf/products/vsan/vmware-oracle-real-application-clusters-on-vmware-virtual-san-reference-architecture.pdf
- *VMware vSphere Virtual Volumes*

  http://www.vmware.com/files/pdf/products/vvol/vmware-oracle-on-virtual-volumes.pdf
- *VMware vSphere Resource Management*

  https://pubs.vmware.com/vsphere-60/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-60-resource-management-guide.pdf
- *VMware vSphere Networking*

  https://pubs.vmware.com/vsphere-60/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-60-networking-guide.pdf
- *VMware vSphere Storage*

  https://pubs.vmware.com/vsphere-60/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-60-storage-guide.pdf
- *VMware Communities: Interpreting esxtop Statistics*

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

18. Acknowledgements

Author: Sudhir Balasubramanian, Staff Solutions Architect – Oracle Technologies

Thanks to the following people for their reviews and inputs:

- Scott Salyer – Director, Enterprise Application Architecture
- Vas Mitra – SAP Staff Solutions Architect
- Don Sullivan - Product Line Marketing Manager for Business Critical Applications
- Andrea Couvrey – Senior Technical Writer, GTPS Solutions Tech Writers

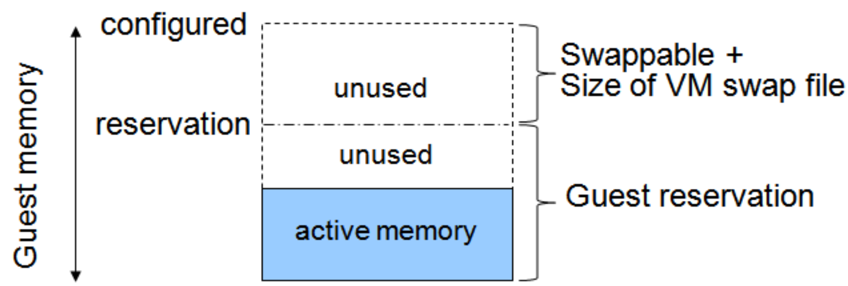
Thanks to the following people for their participation:

- Bob Goldsand - Staff Global Solutions Architect, Solution Engineering
- VMware GSS team

Appendix A: Virtual Machine Memory Settings

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

Figure 29. 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 ESX/ESXi for other virtual machines.
- Swappable – Virtual machine memory that can be reclaimed by the balloon driver or, in the worst case, by ESX/ESXi swapping. This is the automatic size of the swap file that is created for each virtual machine on the VMFS file system (`.vswp`file).

For more information about ESX / ESXi memory management concepts and the balloon driver, see *VMware vSphere Resource Management* at https://pubs.vmware.com/vsphere-60/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-60-resource-management-guide.pdf.

Appendix B: Performance Studies

This appendix provides Oracle on vSphere performance case studies:

- Oracle 12c and vSphere 6.0 Monster VM

  https://www.vmware.com/files/pdf/techpaper/vsphere6-oracle-perf.pdf
- *Accelerating Virtualized Oracle 12c performance with vSphere 5.5 Advanced Features Flash Read Cache and vMotion*

  http://principledtechnologies.com/vmware/vFRC_Oracle_12c_0414.pdf
- Performance Study of Oracle RAC on VMware vSphere 5.0

  www.vmware.com/files/pdf/techpaper/OracleRAC-perf-vSphere5.pdf
- Oracle 12c RAC Performance on vSphere 5 and EMC

  http://www.emc.com/collateral/hardware/white-papers/h23606-emc-vnx-scaling-performance-oracle-12c.pdf
- Oracle RAC and vSphere 5.1 vSphere vMotion

  http://www.principledtechnologies.com/VMware/vMotion_Oracle_RAC_1013.pdf
- *Peeking At the Future with Giant Monster Virtual Machines*

  http://www.vmware.com/files/pdf/business-critical-apps/vmware-peeking-the-future-with-virtual-machines.pdf

Appendix C: Configuration Maximums

For more information about VMware ESXi 5.5 configuration maximums, see the *VMware Configuration Maximums vSphere 5.5* guide (https://www.vmware.com/pdf/vsphere5/r55/vsphere-55-configuration-maximums.pdf).

For more information about VMware ESXi 6.0 configuration maximums, see the *VMware Configuration Maximums vSphere 6.0 guide* (https://www.vmware.com/pdf/vsphere6/r60/vsphere-60-configuration-maximums.pdf).

Appendix D: VMware Compatibility Guide

VMware recommends always checking the VMware Compatibility Guides when embarking on a new virtualization project or platform upgrade initiative to understand compatibility between the hardware on the different layers of the stack and uncover any potential issues that might occur. See https://www.vmware.com/resources/compatibility/search.php.