Microsoft Exchange 2013 on VMware
Best Practices Guide

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

Contents

List of Tables

List of Figures

1. Introduction

The communication vehicles available to an organization have increased greatly over the past few years. Many organizations have adopted the approach of social networking to encourage conversation between employees and provide a more interactive experience. Although this new way of communicating is becoming more popular, email is still used as the main mode of communication. The term “email” has come to mean much more than sending only electronic mail between user mailboxes. Today, when someone says, “I need to get into my email,” it can mean that they must check their calendar, look for a document, obtain contact information, or even make a call. This stems from the way collaboration tools have become integrated and the proliferation of Microsoft Exchange.

Exchange is the most widely used email system in the world. The integration between Exchange and Microsoft Office, SharePoint, and Lync makes it a strong contender for new organizations looking for a core to their email and collaboration strategy. In established organizations, Exchange has been the communications engine for many versions, and it has met the requirements. With each new version of Exchange, enhancements to availability, performance, and scalability become compelling reasons to explore migration. Exchange 2013 continues the tradition with a new architecture, enhanced availability features, and further optimized storage I/O operations.

Even with the improvements, however, Exchange 2013 it is still susceptible to the shortcomings inherent in most applications running directly on physical hardware, such as hardware platform dependence, underutilization of server computing resources, lack of flexibility to respond to changing workloads, and heavy costs associated with maintaining disaster recovery, test, and development environments. The architectural improvements in Exchange Server 2013 cannot fully address these limitations.

The ideal platform for Exchange adapts easily to changing workloads, provides flexibility to accommodate changing demands on an organization’s IT infrastructure, remains reliable and resilient despite system outages, and improves both staff and infrastructure hardware effectiveness. A new operational platform based on VMware® vSphere® can accomplish these goals.

1.1 Purpose

This guide provides best practice guidelines for deploying Exchange Server 2013 on vSphere. The recommendations in this guide are not specific to any particular set of hardware or to the size and scope of any particular Exchange implementation. The examples and considerations in this document provide guidance but do not represent strict design requirements because the flexibility of Exchange Server 2013 on vSphere allows for a wide variety of valid configurations.

1.2 Target Audience

This guide assumes a basic knowledge and understanding of vSphere and Exchange Server 2013.

- Architectural staff can use this document to gain an understanding of how the system will work as a whole as they design and implement various components.
- Engineers and administrators can use this document as a catalog of technical capabilities.
- Messaging staff can use this document to gain an understanding of how Exchange might fit into a virtual infrastructure.
- Management staff and process owners can use this document to help model business processes to take advantage of the savings and operational efficiencies achieved with virtualization.

1.3 Scope

The scope of this document is limited to the following topics:

- VMware ESXi™ Host Best Practices for Exchange – This section provides best practice guidelines for preparing the vSphere platform for running Exchange Server 2013. Guidance is included for CPU, memory, storage, and networking.
- Using VMware vSphere vMotion®, VMware vSphere Distributed Resource Scheduler™, and VMware vSphere High Availability (HA) with Exchange 2013 – This section provides an overview of vSphere vMotion, vSphere HA, and DRS, and guidance for usage of these vSphere features with Exchange 2013 virtual machines.
- Exchange Performance on vSphere – This section provides background information on Exchange Server performance in a virtual machine. It also provides information on official VMware partner testing and guidelines for conducting and measuring internal performance tests.
- VMware Enhancements for Deployment and Operations – This section provides a brief look at vSphere features and add-ons that enhance deployment and management of Exchange 2013.

The following topics are out of scope for this document but are addressed in other documentation in the *Microsoft Exchange 2013 on VMware Solution Sales Enablement Toolkit*.

- Design and Sizing Guidance – This information is available in *Microsoft Exchange 2013 on VMware Design and Sizing Guide.* This document details the capacity planning process and provides sizing examples for split-role, multi-role and real-world customer configurations.
- Availability and Recovery Options – Although this document briefly covers VMware features that can enhance availability and recovery, a more in-depth discussion of this subject is covered in *Microsoft Exchange 2013 on VMware Availability and Recovery Options*.

It is important to note that this and other guides are limited in focus to deploying Exchange on vSphere. Exchange deployments cover a wide subject area, and Exchange-specific design principles should always follow Microsoft guidelines for best results.

2. ESXi Host Best Practices for Exchange

A well-designed ESXi host platform is crucial to the successful implementation of enterprise applications such as Exchange. The following sections outline general best practices for designing vSphere for Exchange 2013.

2.1 CPU Configuration Guidelines

The latest release of vSphere has dramatically increased the scalability of virtual machines, enabling configurations of up to 64 virtual processors in a single virtual machine. With that many resources available it seems that the answer to performance is to do nothing more than to create larger virtual machines. There is much more that should go into deciding how much processing power goes into a virtual machine. This section reviews various features that are available in vSphere with regard to virtualizing CPUs. Where relevant, this document discusses the impact of those features to Exchange 2013 and the recommended practices for using those features.

2.1.1 Physical and Virtual CPUs

VMware uses the terms virtual CPU (vCPU) and physical CPU to distinguish between the processors within the virtual machine and the underlying physical processor cores. Virtual machines with more than one virtual CPU are also called symmetric multiprocessing (SMP) virtual machines. The virtual machine monitor (VMM) is responsible for virtualizing the CPUs. When a virtual machine begins running, control transfers to the VMM, which is responsible for virtualizing guest operating system instructions.

2.1.2 vSphere Virtual Symmetric Multiprocessing

VMware vSphere Virtual Symmetric Multiprocessing enhances virtual machine performance by enabling a single virtual machine to use multiple physical processor cores simultaneously. vSphere supports allocating 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 ESXi scheduler uses a mechanism called *relaxed co-scheduling* to schedule processors. *Strict co-scheduling* required all vCPUs to be scheduled on physical cores simultaneously, but relaxed co-scheduling monitors time skew between vCPUs to make scheduling or co-stopping decisions. A leading vCPU might decide to co-stop itself to allow for a lagging vCPU to catch up. Consider the following points when using multiple vCPUs:

- Virtual machines with multiple vCPUs perform very well in the latest versions of vSphere, as compared to older versions where strict co-scheduling was used.
- Regardless of relaxed co-scheduling, the ESXi scheduler prefers to schedule vCPUs together when possible to keep them in synch. Deploying virtual machines with multiple vCPUs that are not used wastes resources and might result in reduced performance of other virtual machines.

For detailed information regarding the CPU scheduler and changes made in vSphere 5.1, refer to *The CPU Scheduler in VMware vSphere 5.1* (http://www.vmware.com/files/pdf/techpaper/VMware-vSphere-CPU-Sched-Perf.pdf).

VMware recommends the following practices:

- Allocate multiple vCPUs to a virtual machine only if the anticipated Exchange workload can truly take advantage of all the vCPUs.
- If the exact workload is not known, size the virtual machine with a smaller number of vCPUs initially, and increase the number later if necessary.
- For performance-critical Exchange virtual machines (production systems), the total number of vCPUs assigned to all the virtual machines should be equal to or less than the total number of *physical* cores on the ESXi host machine, not *hyperthreaded* cores.

Although larger virtual machines are possible in vSphere, VMware recommends reducing the number of virtual CPUs if monitoring of the actual workload shows that the Exchange application is not benefitting from the increased virtual CPUs. Exchange sizing tools tend to provide conservative recommendations for CPU sizing. As a result, virtual machines sized for a specific number of mailboxes might be underutilized.

2.1.3 CPU Reservations

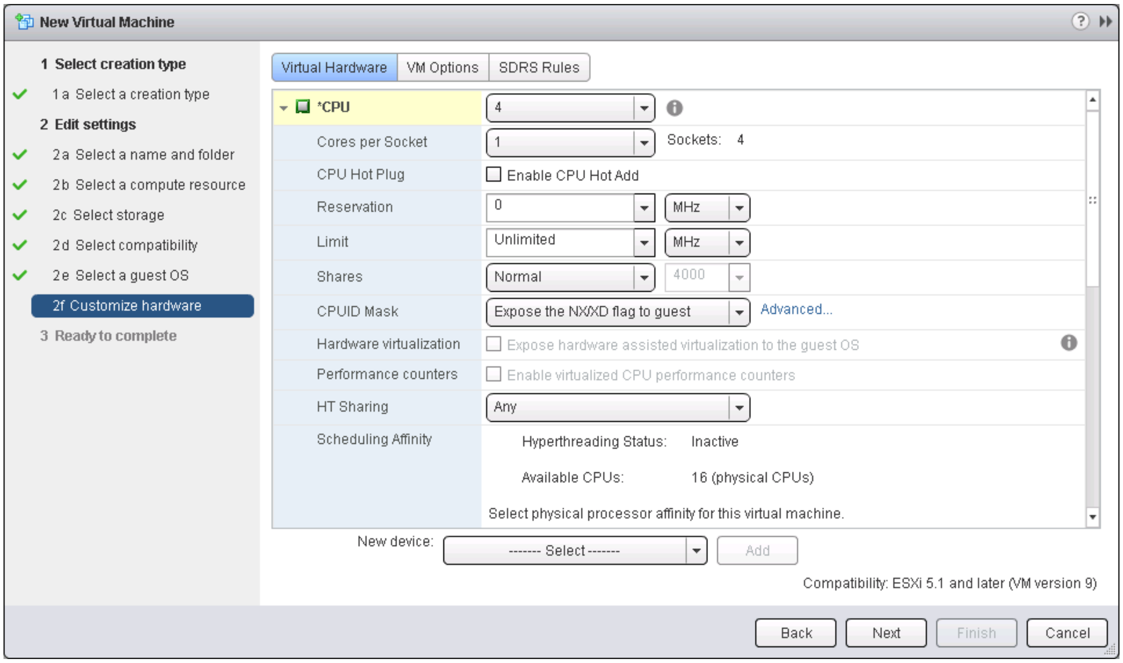
Setting a CPU reservation sets a guaranteed CPU allocation for the virtual machine. This practice is generally not recommended because after used, the reserved resources are not available to other virtual machines, and flexibility is often required to manage changing workloads. However, SLAs and multitenancy might require a guaranteed amount of compute resource to be available. In these cases, reservations can be used to meet these requirements. VMware has conducted tests on virtual CPU overcommitment with SAP and SQL, showing that the performance degradation inside the virtual machines is linearly reciprocal to the overcommitment. As the performance degradation is *graceful*, any virtual CPU overcommitment can be effectively managed by using DRS and vSphere vMotion to move virtual machines to other ESXi hosts to obtain more processing power.

2.1.4 Virtual Cores and Virtual Sockets

In vSphere 5, configuring the number of virtual cores per virtual socket is exposed in the GUI. This feature provides two functions. When used with virtual Non-Uniform Memory Access (vNUMA)-enabled virtual machines, this setting can be used to present specific NUMA topologies to the guest operating system. More commonly, this feature allows a guest operating system to utilize all of its assigned vCPUs in the case of an operating system that is limited to a certain number of CPUs.

As an example, Windows Server 2008 R2 Standard Edition is limited to seeing four CPUs. In a physical environment, physical servers with more than four sockets utilize only the total cores in four of those sockets. In previous versions of vSphere, configuring a Windows Server 2008 R2 Standard virtual machine with more than four vCPUs resulted in the virtual machine seeing only four vCPUs. This was configurable within the virtual machine VMX file and is now configurable in the GUI. By configuring multiple cores per socket, the guest operating system can see and utilize all configured vCPUs.

Figure 1. New Virtual Machine CPU Configuration



New Virtual Machine

1 Select creation type
✓ 1a Select a creation type
2 Edit settings
✓ 2a Select a name and folder
✓ 2b Select a compute resource
✓ 2c Select storage
✓ 2d Select compatibility
✓ 2e Select a guest OS
2f Customize hardware
3 Ready to complete

*CPU	4
Cores per Socket	1 Sockets: 4
CPU Hot Plug	□ Enable CPU Hot Add
Reservation	0 MHz
Limit	Unlimited MHz
Shares	Normal 4000
CPUID Mask	Expose the NX/XD flag to guest Advanced...
Hardware virtualization	□ Expose hardware assisted virtualization to the guest OS
Performance counters	□ Enable virtualized CPU performance counters
HT Sharing	Any
Scheduling Affinity	Hyperthreading Status: Inactive Available CPUs: 16 (physical CPUs) Select physical processor affinity for this virtual machine.

New device: ------- Select ------- Add

Compatibility: ESXi 5.1 and later (VM version 9)

Virtual machines, including those running Exchange 2013, should be configured with multiple virtual sockets/CPUs. Only use the **Cores per Socket** option if the guest operating system requires the change to make all vCPUs visible. Virtual machines using vNUMA might benefit from this option, but the recommendation for these virtual machines is generally to use virtual sockets (CPUs in the web client). Exchange 2013 is not a NUMA-aware application, and performance tests have not shown any significant performance improvements by enabling vNUMA.

2.1.5 Hyperthreading

Hyperthreading technology—recent versions are called symmetric multithreading, or SMT—allows a single physical processor core to behave like two logical processors, essentially allowing two independent threads to run simultaneously. Unlike having twice as many processor cores that can roughly double performance, hyperthreading can provide anywhere from a slight to a significant increase in system performance by keeping the processor pipeline busier. For example, an ESXi host system enabled for SMT on an 8-core server sees 16 threads that appear as 16 logical processors.

Guidance provided by Microsoft regarding Exchange sizing and the use of hyperthreading has led to some confusion among those looking at virtualizing Exchange. The Exchange 2010 Microsoft TechNet article *Understanding Processor Configurations and Exchange Performance* (http://technet.microsoft.com/en-us/library/dd346699.aspx) states the following:

Hyperthreading causes capacity planning and monitoring challenges, and as a result, the expected gain in CPU overhead is likely not justified. Hyperthreading should be disabled by default for production Exchange servers and only enabled if absolutely necessary as a temporary measure to increase CPU capacity until additional hardware can be obtained.

vSphere uses hyperthreads to provide more scheduling choices for the hypervisor. Hyperthreads provide additional targets for *worlds*, a schedulable CPU context that can include a virtual CPU or hypervisor management process. Additionally, for workloads that are not CPU bound, scheduling multiple vCPUs onto a physical core's logical cores can provide increased throughput by increasing the work in the pipeline. It should be noted that the CPU scheduler schedules to a whole core over a hyperthread, or partial core, if CPU time is lost due to hyperthread contention.

When designing for a virtualized Exchange implementation, sizing should be conducted with physical cores in mind. Although Microsoft supports a maximum virtual CPU to physical CPU overcommitment ratio of 2:1, the recommended practice is to keep this as close to 1:1 as possible. For example, in the case of an ESXi host with eight physical cores, the total number of vCPUs across all Exchange virtual machines on that host should not exceed eight vCPUs. After hyperthreading is enabled on the ESXi host, the hypervisor has 16 logical processors where it can schedule worlds to run.

2.1.6 Non-Uniform Memory Access

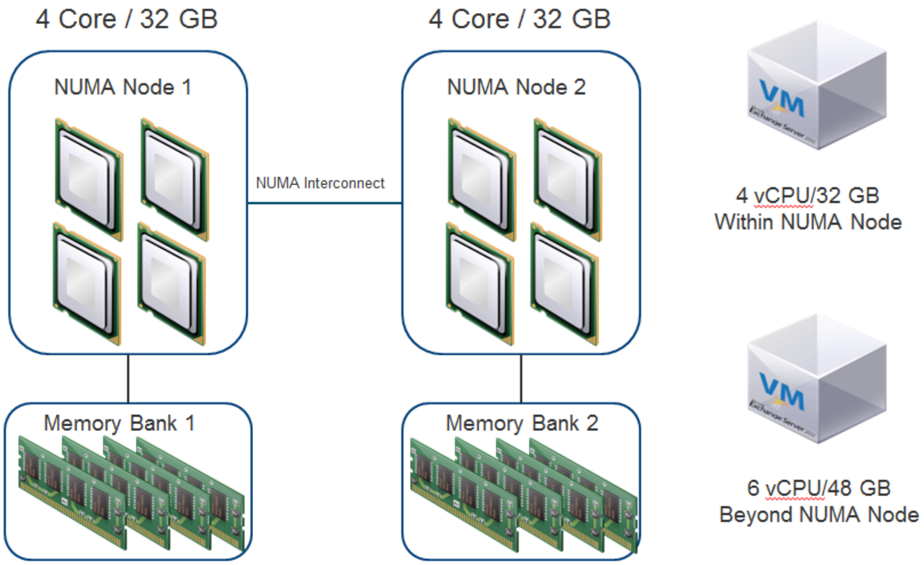
In NUMA systems, a processor or set of processor cores have memory that they can access with very little latency. The memory and its associated processor or processor cores are referred to as a NUMA node. Operating systems and applications designed to be NUMA-aware can make decisions as to where a process might run, relative to the NUMA architecture. This allows processes to access memory local to the NUMA node rather than having to traverse an interconnect, incurring additional latency. Exchange 2013 is not NUMA-aware, but ESXi is.

vSphere ESXi provides mechanisms for letting virtual machines take advantage of NUMA. The first mechanism is transparently managed by ESXi while it schedules a virtual machine's virtual CPUs on NUMA nodes. By attempting to keep all of a virtual machine's virtual CPUs scheduled on a single NUMA node, memory access can remain local. For this to work effectively, the virtual machine should be sized to fit within a single NUMA node. This placement is not a guarantee because the scheduler migrates a virtual machine between NUMA nodes based on the demand.

The second mechanism for providing virtual machines with NUMA capabilities is vNUMA. When enabled for vNUMA, a virtual machine is presented with the NUMA architecture of the underlying hardware. This allows NUMA-aware operating systems and applications to make intelligent decisions based on the underlying host's capabilities. vNUMA is enabled for virtual machines with nine or more vCPUs. Because Exchange 2013 is not NUMA aware, enabling vNUMA for an Exchange virtual machine does not provide any additional performance benefit.

Consider sizing Exchange 2013 virtual machines to fit within the size of the physical NUMA node for best performance. The following figure depicts an ESXi host with two NUMA nodes, each comprising four physical cores and 32GB of memory. The virtual machine allocated with four vCPUs and 32GB of memory can be scheduled by ESXi onto a single NUMA node. The virtual machine allocated with six vCPUs and 64GB of memory must span NUMA nodes and might incur some memory access latency. Large environments might choose to test each configuration to determine whether the additional latency warrants creating additional, smaller virtual machines.

Figure 2. NUMA Architecture Sizing



Verify that all ESXi hosts have NUMA enabled in the system BIOS. In some systems NUMA is enabled by disabling node interleaving.

2.2 Memory Configuration Guidelines

This section provides guidelines for memory allocation to Exchange virtual machines. These guidelines consider vSphere memory overhead and the virtual machine memory settings.

2.2.1 ESXi Memory Management Concepts

vSphere virtualizes guest physical memory by adding an extra level of address translation. Shadow page tables make it possible to provide this additional translation with little or no overhead. Managing memory in the hypervisor enables the following:

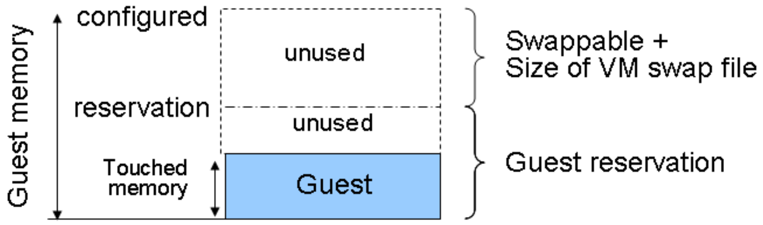
- Memory sharing across virtual machines that have similar data – Same guest operating systems.
- Memory overcommitment – Allocating more memory to virtual machines than is physically available on the ESXi host.
- A memory balloon technique – Virtual machines that do not need all the memory they have been allocated give memory to virtual machines that require additional allocated memory.

For more details about vSphere memory management concepts, consult *Understanding Memory Resource Management in VMware vSphere 5.0* (http://www.vmware.com/files/pdf/mem_mgmt_perf_vsphere5.pdf).

2.2.2 Virtual Machine Memory Concepts

The following figure illustrates the use of memory settings parameters in the virtual machine.

Figure 3. Virtual Machine Memory Settings



The vSphere memory settings for a virtual machine include the following parameters:

- Configured memory – Memory size of virtual machine assigned at creation.
- Touched memory – Memory actually used by the virtual machine. vSphere only allocates guest operating system memory on demand.
- Swappable – Virtual machine memory that can be reclaimed by the balloon driver or by vSphere swapping. Ballooning occurs before vSphere swapping. If this memory is in use by the virtual machine (touched and in use), the balloon driver causes the guest operating system to swap. Also, this value is the size of the per virtual machine swap file that is created on the VMware vSphere Virtual Machine File System (VMFS) file system.
- If the balloon driver is unable to reclaim memory quickly enough or is disabled or not installed, vSphere forcibly reclaims memory from the virtual machine using the VMkernel swap file.

2.2.3 Allocating Memory to Exchange Virtual Machines

Microsoft has developed a thorough sizing methodology for Exchange Server that has matured with recent versions of Exchange. VMware recommends using the memory sizing guidelines set by Microsoft. Simplistically the amount of memory required for an Exchange Server is driven by its role and, if it is a Mailbox server, the number of mailboxes on that server.

As Exchange Servers are memory-intensive, and performance is a key factor, such as in production environments, VMware recommends the following practices:

- Do not overcommit memory on ESXi hosts running Exchange workloads. For production systems, it is possible to enforce this policy by setting a memory reservation to the configured size of the virtual machine. Also note that:
    - Setting memory reservations might limit vSphere vMotion. A virtual machine can only be migrated if the target ESXi host has free physical memory equal to or greater than the size of the reservation.
    - Setting the memory reservation to the configured size of the virtual machine results in a per-virtual machine VMkernel swap file of zero bytes that consumes less storage and eliminates ESXi host-level swapping. The guest operating system within the virtual machine still requires its own page file.
    - Reservations are generally only recommended when it is possible that memory might become overcommitted on hosts running Exchange virtual machines, when SLAs dictate that memory be “guaranteed,” or when there is a desire to reclaim space used by a virtual machine swap file.
- It is important to right-size the configured memory of a virtual machine. This might be difficult to determine in an Exchange environment because the Exchange JET cache is allocated based on memory present during service start-up. Understand the expected mailbox profile and recommended mailbox cache allocation to determine the best starting point for memory allocation.
- Do not disable the balloon driver (which is installed with VMware Tools™).
- Enable DRS to balance workloads in the ESXi host cluster. DRS and reservations can give critical workloads the resources they require to operate optimally. More recommendations for using DRS with Exchange 2013 are available in Section 3, Using vSphere Technologies with Exchange 2013.

2.2.4 Memory Oversubscription and Dynamic Memory

In Exchange 2013 JET cache is allocated based on the amount of memory available to the operating system at the time of service start-up. After allocated, JET cache is distributed among active and passive databases. With this model of memory pre-allocation for use by Exchange databases, adding memory to a running Exchange virtual machine provides no additional benefit unless the virtual machine was rebooted or Exchange services were restarted. In contrast, removing memory that JET had allocated for database consumption impacts performance of the store worker and indexing processes by increasing processing and storage I/O.

Microsoft support for the virtualization of Exchange 2013 states that the oversubscription and dynamic allocation of memory for Exchange virtual machines is not supported. In vSphere, memory oversubscription or overcommitment is made possible with transparent page sharing, memory ballooning, memory compression, and host swapping. Although these features have been shown to allow for some level of memory overcommitment for Exchange with no impact to performance, the recommendation is to follow Microsoft guidance and not allow any level of memory overcommitment for Exchange virtual machines.

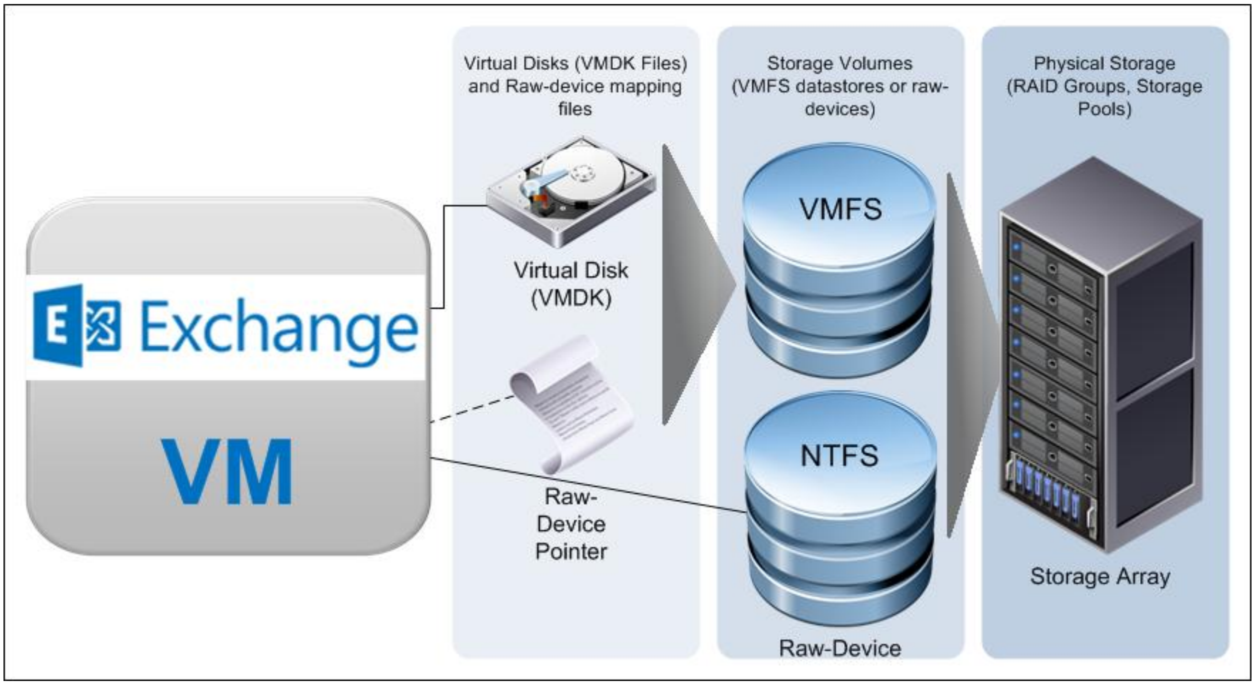
2.3 Storage Virtualization

VMFS is a cluster file system that provides storage virtualization optimized for virtual machines. Each virtual machine is encapsulated in a small set of files, and VMFS is the default storage system for these files on physical SCSI disks and partitions. VMware supports Fibre Channel, iSCSI, and network-attached storage (NAS) shared-storage protocols.

It is preferable to deploy virtual machine files on shared storage to take advantage of vSphere vMotion, vSphere HA, and DRS. This is considered a best practice for mission-critical Exchange deployments that are often installed on third-party, shared-storage management solutions.

VMware storage virtualization can be categorized into three pillars of storage technology, as illustrated in the following figure. The storage array is the physical storage pillar, comprising physical disks presented as logical storage volumes in the next pillar. Storage volumes, presented from physical storage, are formatted as VMFS datastores or with native file systems when mounted as raw device mappings. Virtual machines consist of virtual disks or raw device mappings that are presented to the guest operating system as SCSI disks that can be partitioned and formatted using any supported file system.

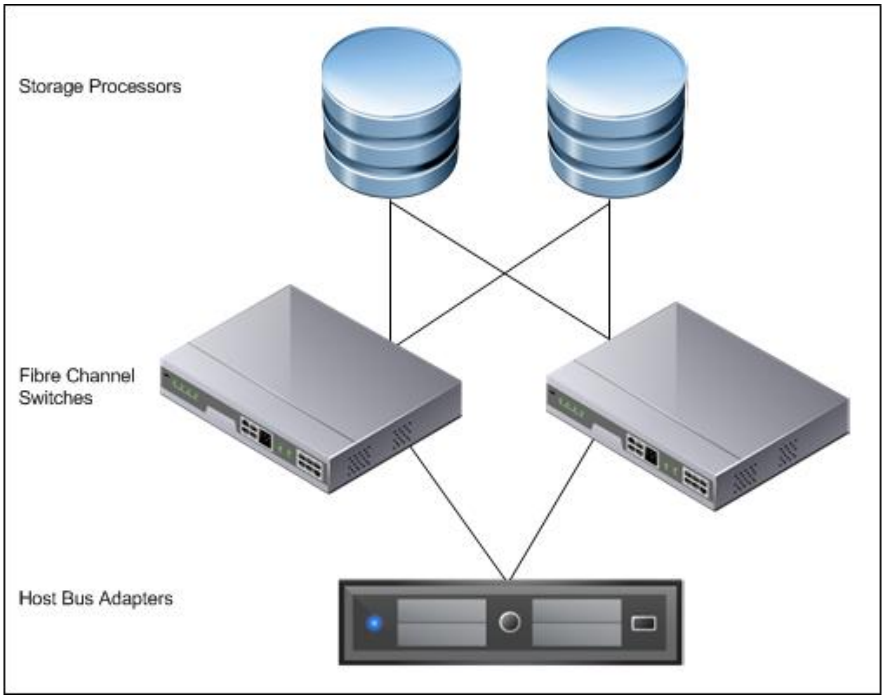
Figure 4. VMware Storage Virtualization


Virtual Disks (VMDK Files) and Raw-device mapping files
Storage Volumes (VMFS datastores or raw-devices)
Physical Storage (RAID Groups, Storage Pools)
Exchange
VM
Virtual Disk (VMDK)
Raw-Device Pointer
VMFS
NTFS
Raw-Device
Storage Array

2.3.1 Storage Multipathing

VMware recommends you set up a minimum of four paths from an ESXi host to a storage array. To accomplish this, the host requires at least two host bus adapter (HBA) ports.

Figure 5. Storage Multipathing Requirements for vSphere



The terms used in the preceding figure are:

- HBA – A device that connects one or more peripheral units to a computer and manages data storage and I/O processing.
- Fibre Channel (FC) – A gigabit-speed networking technology used to build storage area networks (SANs) and to transmit data.
- Storage Processor (SP) – A SAN component that processes HBA requests routed through an FC switch and handles the RAID/volume functionality of the disk array.

2.3.2 Raw Device Mapping

VMFS also supports Raw Device Mapping (RDM). RDM allows a virtual machine to directly access a volume on the physical storage subsystem and can only be used with Fibre Channel or iSCSI. RDM can be thought of as providing a symbolic link or mount point 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. Connectivity from the virtual machine to the raw volume is direct, and all data is stored using the native file system, NTFS. In the case of a failure at the VMFS datastore holding the RDM mapping file, a new mapping file can be created, access to the raw volume and its data is restored, and no data loss occurs.

The decision to use VMFS or RDM for Exchange data should be based on technical requirements. The following table provides considerations when making a decision between the two.

Table 1. VMFS and Raw Disk Mapping Considerations for Exchange 2013

VMFS	RDM
• Volume can contain many virtual machine disk files, reducing management overhead. • Increases storage utilization and provides better flexibility and easier administration and management. • Supports existing and future vSphere storage virtualization features. • Fully supports VMware vCenter™ Site Recovery Manager™. • Supports the use of vSphere vMotion, vSphere HA, and DRS. • Supports VMFS volumes up to 64TB and virtual disks/VMDK files up to 2TB. • Has open virtual disk capacity of up to 25TB per host (8TB by default) – More information is available at http://kb.vmware.com/kb/1004424.	• Ideal if disks must be dedicated to a single virtual machine. • Requires more LUNs, making it is easier to reach the limit of 256 LUNs that might be presented to an ESXi host. • Might be required to leverage array-level backup and replication tools (VSS) integrated with Exchange databases. • Facilitates data migration between physical and virtual machines using the LUN swing method. • Fully supports vCenter Site Recovery Manager. • Supports vSphere vMotion, vSphere HA, and DRS. • Supports presenting volumes of up to 64TB to the guest operating system – Physical-mode RDM only.

2.3.3 Virtual SCSI Adapters

VMware provides two commonly used virtual SCSI adapters for Windows Server 2008 R2 and Windows Server 2012, LSI Logic SAS and VMware Paravirtual SCSI (PVSCSI). The default adapter when creating new virtual machines with either of these two operating systems is LSI Logic SAS, and this adapter can satisfy the requirements of most workloads. The fact that it is the default and requires no additional drivers has made it the default vSCSI adapter for many organizations.

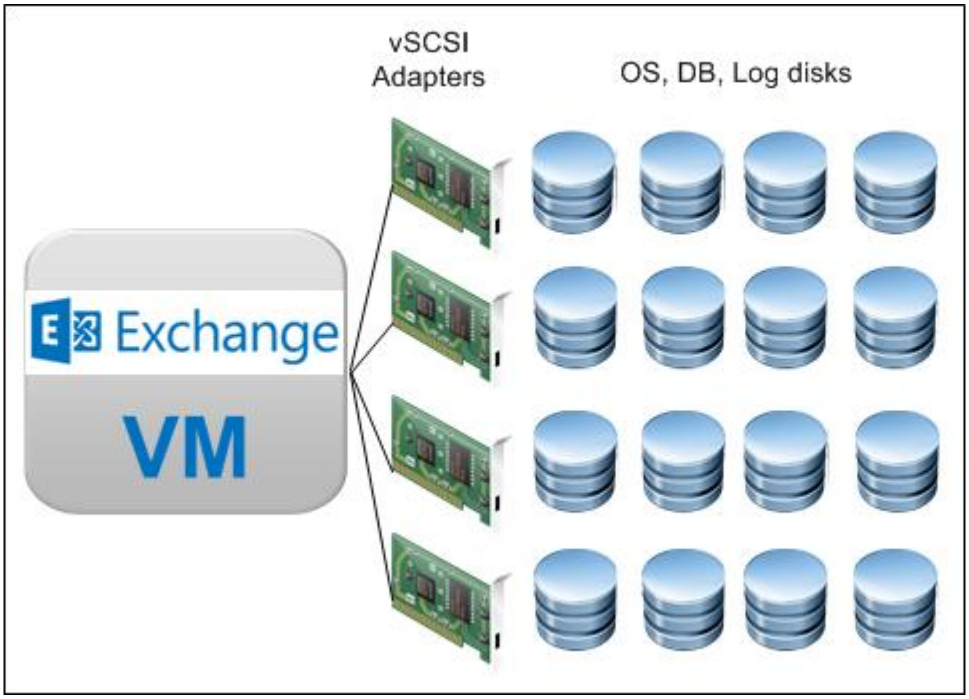
The Paravirtual SCSI adapter is a high-performance vSCSI adapter developed by VMware to provide optimal performance for virtualized business critical applications. The advantage of the PVSCSI adapter is that the added performance is delivered while minimizing the use of hypervisor CPU resources. This leads to less hypervisor overhead required to run storage I/O-intensive applications.

Note For environments running ESXi versions prior to 5.0 Update 1 considering using PVSCSI. Refer to *Windows 2008 R2 virtual machine using a paravirtual SCSI adapter reports the error: Operating system error 1117 encountered* (http://kb.vmware.com/kb/2004578).

Exchange 2013 has greatly reduced the amount of I/O generated to access mailbox data, however storage latency is still a factor. In environments supporting thousands of users per Mailbox server, PVSCSI might prove beneficial. The decision on whether to use LSI Logic SAS or PVSCSI should be made based on Jetstress testing of the predicted workload using both adapters. Additionally, organizations must consider any management overhead an implementation of PVSCSI might introduce. Because many organizations have standardized on LSI Logic SAS, if the latency and throughput difference is negligible with the proposed configuration, the best option might be the one with the least impact to the current environment.

Virtual machines can be deployed with up to four virtual SCSI adapters. Each vSCSI adapter can accommodate up to 15 storage devices for a total of 60 storage devices per virtual machine. During allocation of storage devices, each device is assigned to a vSCSI adapter in sequential order. Not until a vSCSI adapter reaches its 15th device will a new vSCSI adapter be created. To provide better parallelism of storage I/O, equally distribute storage devices among the four available vSCSI adapters, as shown in the following figure.

Figure 6. Storage Distribution with Multiple vSCSI Adapters



2.3.4 In-Guest iSCSI and Network-Attached Storage

Similar to RDM, in-guest iSCSI initiator-attached LUNs provide dedicated storage to a virtual machine. Storage presented using in-guest iSCSI is formatted natively using NTFS within the Windows guest operating system and bypasses the storage management of the ESXi host. Presenting storage in this way requires that additional attention be provided to the networking infrastructure and configuration both at the vSphere level as well as at the physical level. However, some of the benefits from using in-guest iSCSI attached storage include the ability to allocate more than 256 LUNs to virtual machines on a single ESXi host and retaining the ability to use array-level backup and replication tools.

Although VMware testing has shown that NAS-attached virtual disks perform very well for Exchange workloads, Microsoft does not currently support accessing Exchange data (mailbox databases, transport queue, and logs) stored on network-attached storage. This includes accessing Exchange data using a UNC path from within the guest operating system, as well as virtual machines with VMDK files located on NFS-attached storage.

2.4 Networking Configuration Guidelines

This section covers design guidelines for the virtual networking environment and provides configuration examples at the ESXi host level for Exchange Server 2013 installations.

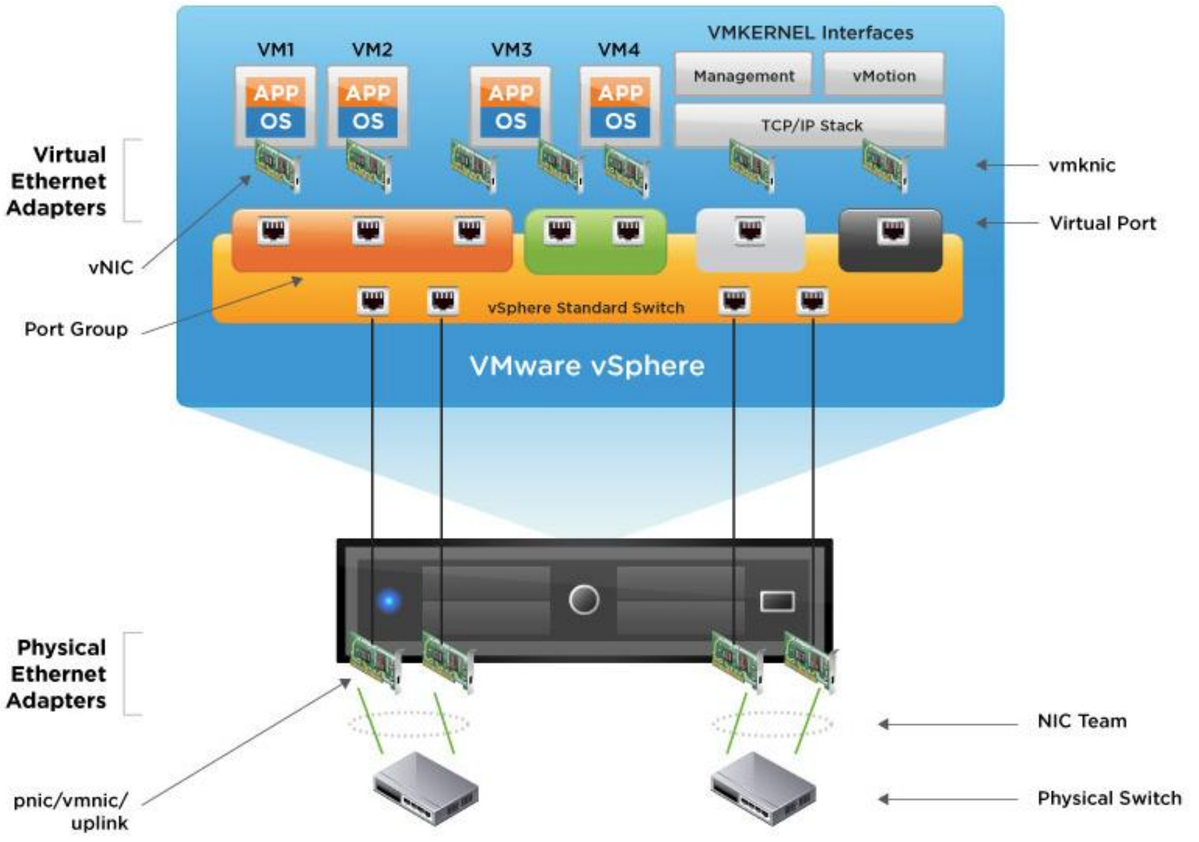
Note The examples do not reflect design requirements and do not cover all possible Exchange network design scenarios.

2.4.1 Virtual Networking Concepts

The virtual networking layer comprises the virtual network devices through which virtual machines and the ESXi host interface with the rest of the network and users. In addition, ESXi hosts use the virtual networking layer to communicate with iSCSI SANs and NAS storage.

The virtual networking layer includes virtual network adapters and the virtual switches. Virtual switches are the key networking components in vSphere. The following figure provides an overview of virtual networking in vSphere.

Figure 7. vSphere Virtual Networking Overview



As shown in the preceding figure, the following components make up the virtual network:

- Physical switch – vSphere host-facing edge of the physical local area network.
- NIC team – Group of physical NICs connected to the same physical/logical networks to provide redundancy.
- Physical network interface (pnic/vmnic/uplink) – Provides connectivity between the ESXi host and the local area network.
- vSphere switch (standard and distributed) – The virtual switch is created in software and provides connectivity between virtual machines. Virtual switches must uplink to a physical NIC (also known as vmnic) to provide virtual machines with connectivity to the LAN, otherwise virtual machine traffic is contained within the virtual switch.

- Port group – Used to create a logical boundary within a virtual switch. This boundary can provide VLAN segmentation when 802.1q trunking is passed from the physical switch, or it can create a boundary for policy settings.
- Virtual NIC (vNIC) – Provides connectivity between the virtual machine and the virtual switch.
- VMkernel (vmknic) – Interface for hypervisor functions, such as connectivity for NFS, iSCSI, vSphere vMotion, and VMware vSphere Fault Tolerance logging.
- Virtual port – Provides connectivity between a vmknic and a virtual switch.

2.4.2 Virtual Networking Best Practices

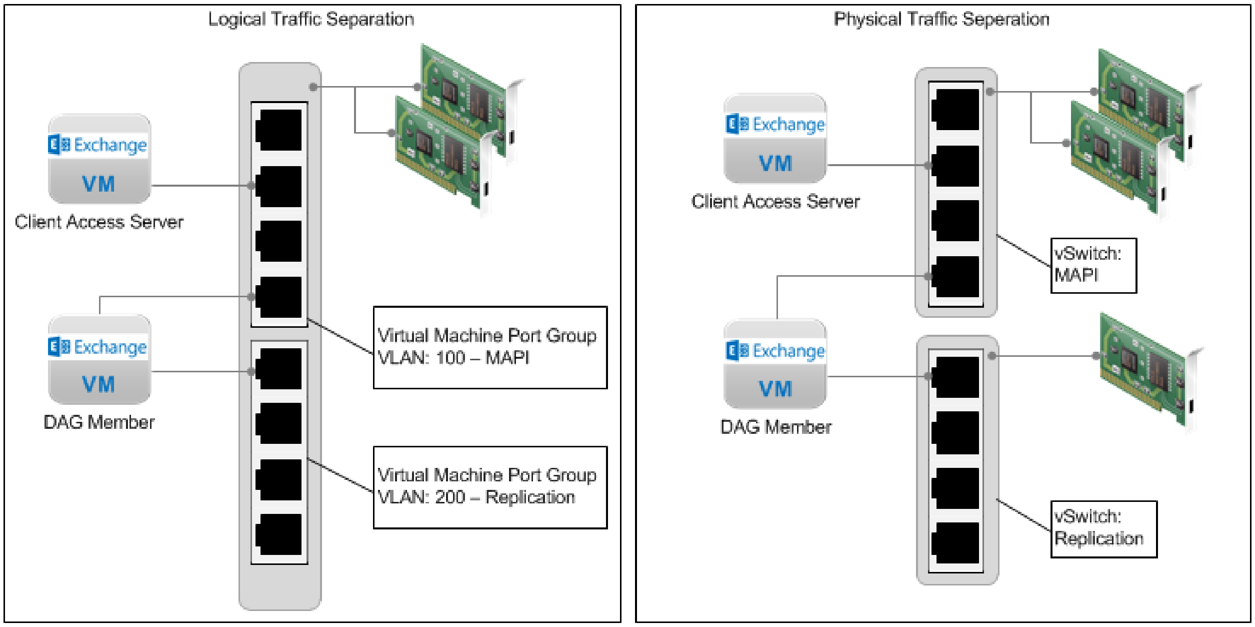
The standard VMware networking best practices apply to running Exchange on vSphere:

- The choice between standard and distributed switches should be made outside of the Exchange design. Standard switches provide a straightforward configuration on a per-host level. For reduced management overhead and increased functionality, the distributed virtual switch should be considered. Both virtual switch types provide the functionality needed by Exchange 2013.
- Traffic types should be separated to keep like traffic contained to designated networks. vSphere can use separate interfaces for management, vSphere vMotion, and network-based storage traffic. Additional interfaces can be used for virtual machine traffic. Within virtual machines, different interfaces can be used to keep certain traffic separated. Use 802.1q VLAN tagging and virtual switch port groups to logically separate traffic. Use separate physical interfaces and dedicated port groups or virtual switches to physically separate traffic. This is shown in Figure 8.
- Leverage network interface teaming capabilities to provide redundant uplinks for virtual switches. To use this capability, assign at least two physical network interfaces per virtual switch.
- Use the VMXNET3 network adapter – This is a paravirtualized network device that provides better throughput with reduced hypervisor CPU utilization.
- For Exchange 2013 virtual machines participating in a database availability group (DAG), configure at least two virtual network interfaces, connected to different VLANs or networks. These interfaces provide access for Messaging Application Programming Interface (MAPI) and replication traffic.
- Follow the guidelines in the *Hardware Networking Considerations* and *Guest Operating Systems* sections of *Performance Best Practices for VMware vSphere 5.0* (http://www.vmware.com/pdf/Perf_Best_Practices_vSphere5.0.pdf).

2.4.3 Sample Exchange Virtual Network Configuration

Because of the flexibility of virtual networking, the topology can take many different forms. There is no single recommended practice because each provides its own sets of benefits. The following figure shows two examples of host-level configuration based on the most common configurations, which are single and multiple virtual switches.

Figure 8. Sample Virtual Network Configuration



Although Exchange 2013 DAG members can be deployed with a single network interface for both MAPI and replication traffic, Microsoft recommends using separate network interfaces for each traffic type. The use of at least two network interfaces allows DAG members to distinguish between system and network failures.

In a vSphere environment traffic separation can be established using either virtual or physical networks. The preceding figure provides examples of these two scenarios.

- The scenario on the left depicts an ESXi host with two network interfaces, teamed for redundancy and using virtual networks and port groups to provide the traffic separation for MAPI and replication traffic. This scenario can also utilize VMware vSphere Network I/O Control for dynamic traffic prioritization.
- The scenario on the right depicts an ESXi host with multiple network interfaces. Physical traffic separation is accomplished by allocating two vmnics on one network to a virtual switch. These vmnics are teamed and dedicated to MAPI network traffic. Replication traffic uses a third vmnic on a separate virtual switch.

In both scenarios the DAG member virtual machine is connected to both networks, according to best practice.

3. Using vSphere Technologies with Exchange 2013

The *Microsoft Exchange 2013 Availability and Recovery Options* guide takes an in-depth look at the options available for building a highly available and site-resilient Exchange 2013 environment on vSphere. This section explores the technologies that make those options possible, mainly vSphere HA, vSphere Distributed Resource Scheduler, and vSphere vMotion. This also includes proven best practices for using these technologies with critical applications such as Exchange 2013.

Although all of the Exchange Server roles have been capable of taking advantage of these advanced vSphere features, support for use with DAG members was not available until Exchange 2010 SP1. Exchange 2013 was released with the same support for these features as Exchange 2010 SP1, validating the continued effort by both VMware and Microsoft to provide support for the features customers believe are valuable for virtualized Exchange environments.

3.1 Overview of vSphere Technologies

vSphere vMotion technology enables the migration of virtual machines from one physical server to another without service interruption. This migration allows you to move Exchange virtual machines from a heavily loaded server to one that is lightly loaded or to offload them to allow for hardware maintenance without any downtime.

DRS takes vSphere vMotion a step further by adding an intelligent scheduler. DRS allows you to set resource assignment policies that reflect business needs. DRS does the calculations and automatically handles the details of physical resource assignments. It dynamically monitors the workload of the running virtual machines and the resource utilization of the physical servers within a vSphere cluster.

vSphere vMotion and DRS perform best under the following conditions:

- The source and target ESXi hosts must be connected to the same gigabit network and the same shared storage.
- A dedicated gigabit (or higher) network for vSphere vMotion is recommended.
- The destination host must have enough resources.
- The virtual machine must not use physical devices such as a CD-ROM or floppy disk.
- The source and destination hosts must have compatible CPU models, otherwise migration with vSphere vMotion fails.
- Virtual machines with smaller memory sizes are better candidates for migration than larger ones.

With vSphere HA, Exchange virtual machines on a failed ESXi host can be restarted on another ESXi host. This feature provides a cost-effective failover alternative to third-party clustering and replication solutions. If you use vSphere HA, be aware that:

- vSphere HA handles ESXi host hardware failure and does not monitor the status of the Exchange services. These must be monitored separately.
- A vSphere HA heartbeat is sent using the vSphere VMkernel network, so redundancy in this network is recommended.
- Allowing two nodes from the same DAG to run on the same ESXi host for an extended period is not recommended when using symmetrical mailbox database distribution. DRS anti-affinity rules can be used to mitigate the risk of running active and passive mailbox databases on the same ESXi host.

3.2 vSphere vMotion

Using vSphere vMotion with Exchange virtual machines is not a new concept. Support for its use with Exchange DAG members has existed since early 2011 when Exchange 2010 SP1 was released. A well-designed and purpose-built vSphere infrastructure can provide seamless migration of running workloads with no special configuration. In the case of Exchange 2013 DAG members, this is essential for utilizing vSphere vMotion with no interference to the cluster services. Even with ideal conditions, the heavy load of Exchange workloads and memory usage can cause a vSphere vMotion operation to trigger a database failover. Database failovers are not necessarily a problem if the environment is designed to properly distribute the load and in fact can help to validate the cluster health by activating databases that might normally go for weeks or months without accepting a user load. However, many administrators prefer that database activations be a planned activity or only done in the case of a failure. For this reason VMware has studied the effect of vSphere vMotion on Exchange DAG members and provided the following best practice recommendations.

3.2.1 Cluster Heartbeat Settings

During a vSphere vMotion operation, memory pages are copied from the source ESXi host to the destination. These pages are copied while the virtual machine is running. During the transition of the virtual machine from running on the source to the destination host, a very slight disruption in network connectivity might occur, typically characterized by a single dropped ping. In most cases this is not a concern, however in highly active environments, this disruption might be enough to trigger the cluster to evict the DAG node temporarily, causing database failover. To mitigate cluster node eviction, the cluster heartbeat interval can be adjusted.

The Windows Failover Cluster parameter `samesubnetdelay` can be modified to help mitigate database failovers during vSphere vMotion of DAG members. This parameter controls how often cluster heartbeat communication is transmitted. The default interval is 1000ms (1 second). The default threshold for missed packets is 5, after which the cluster service determines that the node has failed. Testing has shown that by increasing the transmission interval to 2000ms (2 seconds) and keeping the threshold at 5 intervals, vSphere vMotion migrations can be performed with reduced occurrences of database failover.

Note Microsoft recommends using a maximum value of 10 seconds for the cluster heartbeat timeout. In this configuration the maximum recommended value is used by configuring a heartbeat interval of 2 seconds (2000 milliseconds) and a threshold of 5 (default).

Using `cluster.exe` to view and modify the `samesubnetdelay` parameter:

- To view the value currently assigned to the `samesubnetdelay` parameter using `cluster.exe` for a DAG named `dag-name`:
  ```
  C:\> C:\cluster.exe /cluster:dag-name /prop
  ```
- To configure the `samesubnetdelay` parameter using `cluster.exe` for a DAG named `dag-name` and for the maximum recommended value of 2000ms:
  ```
  C:\> C:\cluster.exe /cluster:dag-name /prop samesubnetdelay=2000
  ```

Using PowerShell to view and modify the `samesubnetdelay` parameter:

- To view the value currently assigned to the `samesubnetdelay` parameter using PowerShell for a DAG named `dag-name`:
  ```
  PS C:\> (get-cluster dag-name).SameSubnetDelay
  ```
- To configure the `samesubnetdelay` parameter using PowerShell for a DAG named `dag-name` and for the maximum recommended value of 2000ms:
  ```
  PS C:\> $cluster = get-cluster dag-name; $cluster.SameSubnetDelay = 2000
  ```

3.2.2 Multiple vSphere vMotion Interfaces

Database failover due to vSphere vMotion operations can be mitigated by using multiple dedicated vSphere vMotion network interfaces. In most cases, the interfaces that are used for vSphere vMotion are also used for management traffic. Because management traffic is relatively light, this does not add significant overhead.

vSphere provides the ability to use multiple vmnic interfaces for vSphere vMotion traffic to effectively load balance the vSphere vMotion traffic. Testing has shown up to a 25% increase in throughput achieved when multiple vSphere vMotion interfaces are used. In the test case with two 2Gbps interfaces configured for vSphere vMotion and no cluster heartbeat modifications, vSphere vMotion operations succeeded with no database failover.

Enabling multiple interfaces for vSphere vMotion requires configuring multiple VMkernel ports on different port groups. Each port group is assigned multiple vmnic interfaces as either active or standby. Refer to *Multiple-NIC vMotion in vSphere 5* (http://kb.vmware.com/kb/2007467) for detailed configuration procedures.

3.2.3 Enable Jumbo Frames for vSphere vMotion Interfaces

Standard Ethernet frames are limited to a length of approximately 1500 bytes. Jumbo frames can contain a payload of up to 9000 bytes. Support for jumbo frames on VMkernel ports was added to vSphere 4.0 for both ESX and ESXi. This added feature means that large frames can be used for all VMkernel traffic, including vSphere vMotion.

Using jumbo frames reduces the processing overhead to provide the best possible performance by reducing the number of frames that must be generated and transmitted by the system. During testing, VMware had an opportunity to test vSphere vMotion migration of DAG nodes with and without jumbo frames enabled. Results showed that with jumbo frames enabled for all VMkernel ports and on the VMware vNetwork Distributed Switch, vSphere vMotion migrations of DAG member virtual machines completed successfully. During these migrations, no database failovers occurred, and there was no need to modify the cluster heartbeat setting.

The use of jumbo frames requires that all network hops between the vSphere hosts support the larger frame size. This includes the systems and all network equipment in between. Switches that do not support, or are not configured to accept, large frames will drop them. Routers and Layer 3 switches might fragment the large frames into smaller frames that must then be reassembled and can cause performance degradation.

3.3 vSphere Distributed Resource Scheduler

Distributed resource scheduling provides active load balancing of virtual machine workloads within a vSphere cluster. Aside from the active monitoring and load balancing functions, DRS provides the following features:

- Virtual machine placement during power-on, based on resource requirements and availability.
- Virtual machine evacuation during ESXi host maintenance mode.
- Virtual machine and host groups for grouping like objects.
- Rules to keep virtual machines together or apart and on or off of a set of hosts.

DRS helps make a virtualized Exchange 2013 environment more agile. The following sections provide recommendations for using DRS with Exchange 2013.

3.3.1 Enable DRS in Fully Automated Mode

DRS provides three levels of automation:

- Manual – Migration recommendations are provided by DRS. No migrations are performed by DRS.
- Partially automated – Virtual machines are automatically placed on hosts during power-on, migration recommendations are provided by DRS, and no migrations are performed by DRS.
- Fully automated – Virtual machines are automatically placed on hosts during power-on and are automatically migrated between hosts to optimize resource usage.

When designed according to VMware recommendations, vSphere clusters that have been purpose-built for Exchange 2013 possess sufficient resources and do not incur many DRS migrations. However, when an ESXi host is placed in maintenance mode, DRS makes recommendations on placement of virtual machines running on that host. To leverage automatic host evacuation, the DRS automation level must be set to `Fully Automated`.

If the vSphere cluster hosting Exchange 2013 is a shared environment, DRS fully automated mode helps to maintain resource optimization among the multiple workloads.

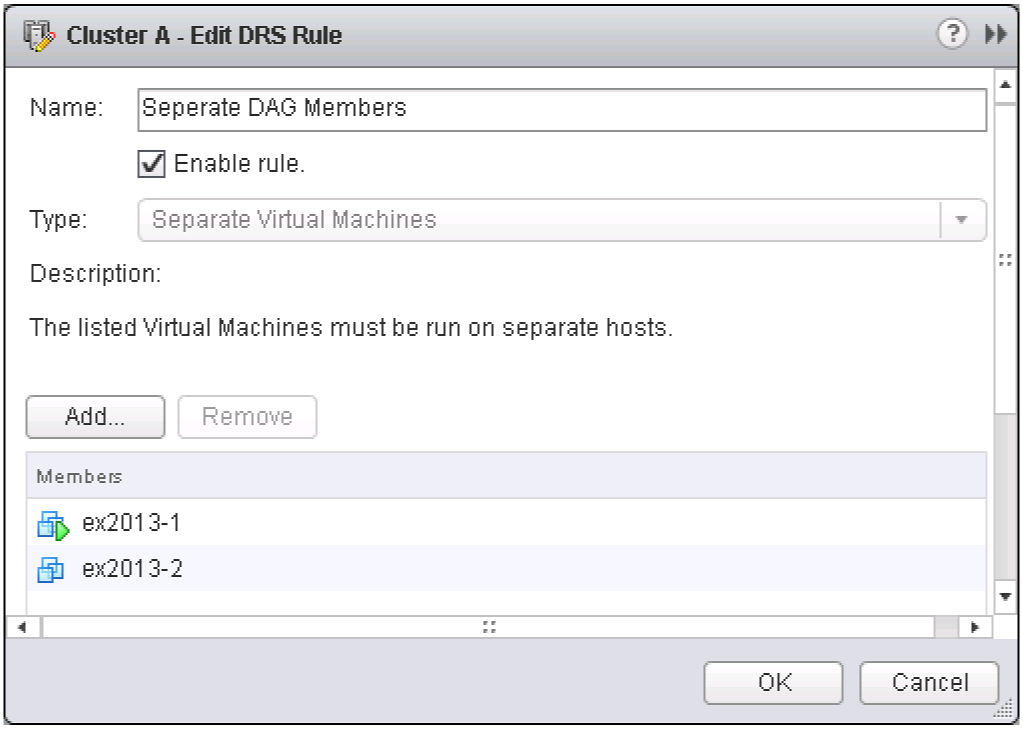
3.3.2 Use Anti-Affinity Rules for Exchange Virtual Machines

DRS provides rules for keeping virtual machines apart or together on the same ESXi host or group of hosts. In an Exchange environment, the common use case for anti-affinity rules is to keep Exchange virtual machines with the same roles installed apart from each other. Client Access servers in a CAS array can run on the same ESXi host, but DRS rules should be used to prevent all CAS virtual machines from running on a single ESXi host.

Microsoft recommends symmetrically distributing mailbox databases among DAG members. Unlike traditional active/passive configurations, this design allows all DAG members to support active users as well as reserve a portion of compute power for failover capacity. In the case of failure of a single DAG member, all remaining members might take part in supporting the failed databases. Because of this, it is recommended that no two members of the same DAG run on the same ESXi host for an extended period of time.

Anti-affinity rules enforce virtual machine separation during power-on operations and vSphere vMotion migrations due to a DRS recommendation, including a host entering maintenance mode. If a virtual machine is enabled for vSphere HA and a host experiences a failure, vSphere HA might power-on a virtual machine and, in effect, violate a DRS anti-affinity rule. This is because vSphere HA does not inspect DRS rules during a recovery task. However, during the next DRS evaluation (every 5 minutes), the virtual machine is migrated to fix the violation.

Figure 9. vSphere Distributed Resource Scheduler Anti-Affinity Rule



Cluster A - Edit DRS Rule

Name: Seperate DAG Members

☑ Enable rule.

Type: Separate Virtual Machines

Description:

The listed Virtual Machines must be run on separate hosts.

Members

ex2013-1

ex2013-2

3.3.3 DRS Groups and Group-Based Rules

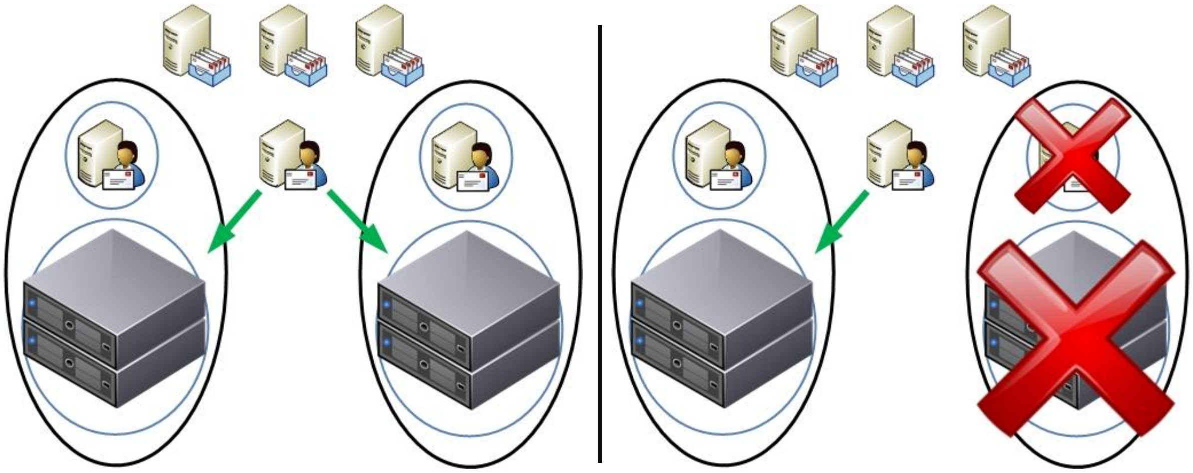
Defining DRS groups helps enforce virtual machine placement that is not possible with only affinity and anti-affinity rules. Host groups are created to group like hosts, such as hosts licensed to run an application or hosts in a blade chassis or rack. Virtual machine groups can be created to group like virtual machines. With groups defined, the *Virtual Machines to Hosts*-type rule is available for use. Virtual Machines to Hosts rules can be created with four variations:

- VM group must run on hosts in group.
- VM group should run on hosts in group.
- VM group must not run on hosts in group.
- VM group should not run on hosts in group.

Must run on rules provide hard enforcement of virtual machine placement. That is, if a rule is created defining that a group of virtual machines must run on a group of ESXi hosts, both DRS and vSphere HA obey these rules. If all hosts in the group are down, the virtual machines are unable to run on any other host in the vSphere cluster.

In the following figure, two virtual machine groups and two host groups are defined. Two *must run on* rules, shown in the solid black ovals, keep the virtual machines in each group running on their respective host group. The virtual machine in the middle is not tied to a group or a rule and might roam. In the case of a failure of all hosts in the group, all virtual machines bound to those hosts by a *must run on* rule stay offline until a host from that group is brought back online.

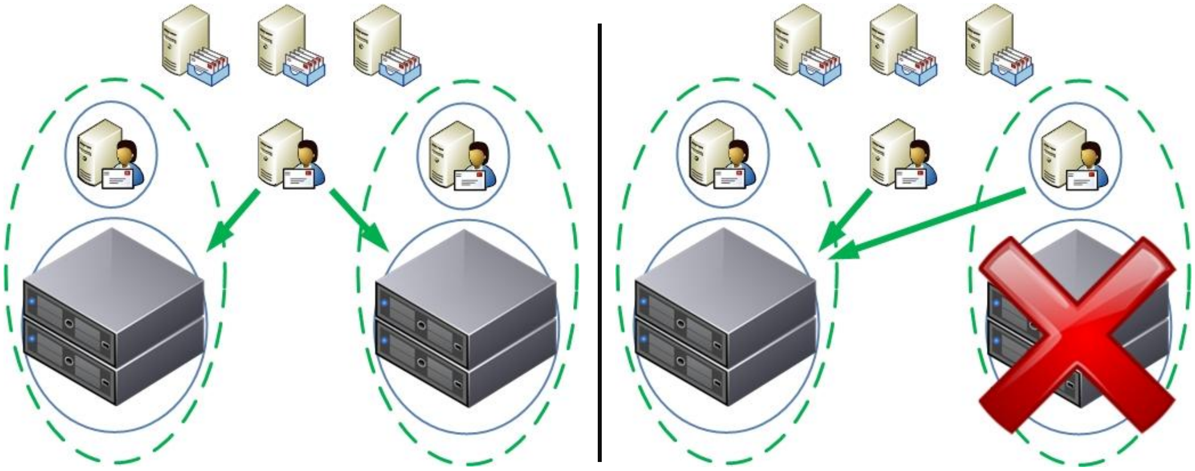
Figure 10. Must Run on Rule Example



Should run on rules provide soft enforcement of virtual machine placement. If a rule is created defining that a group of virtual machines *should run on* a group of ESXi hosts, those virtual machines can still run on other hosts in the vSphere cluster outside of the host group, if needed.

In the following figure two virtual machine groups and two host groups are defined. Two *should run on* rules, shown in the broken green ovals, keep the virtual machines in each group running on their respective host group. The virtual machine in the middle is not tied to a group or a rule and might roam. In the case of a failure of all hosts in the group, virtual machines bound to those hosts by a *should run on* rule can be brought back online by vSphere HA.

Figure 11. Should Run on Rule Example



In an Exchange 2013 environment Virtual Machine to Host rules can be used to provide soft or hard enforcement of virtual machine placement. As an example, consider creating groups of ESXi hosts based on a failure domain, such as a blade chassis or server rack. Create two virtual machine groups with each containing half of the Client Access server virtual machines, and create rules to link each virtual machine group to a host group. In the case of a complete chassis or rack failure, any virtual machines that have failed can be powered back on by vSphere HA.

3.4 vSphere High Availability

By providing a higher level of availability than is possible out-of-the-box for most applications, vSphere HA has become the default HA solution for vSphere virtual machines. Regardless of operating system or application, vSphere HA can provide protection from ESXi host failures, guest operating system failures, and, with the help of third-party add-ons, application failures.

Exchange 2013 environments are built for high availability. Client Access servers are deployed in arrays that are then load-balanced. Mailbox servers are deployed in DAGs for mailbox database high availability. Seemingly this provides all of the availability needed, however in the case of a hardware failure, utilization of the remaining Client Access servers rises as new connections are established, and DAG protection is reduced as passive databases are activated. In a physical deployment, an administrator needs to address the problem quickly to restore availability levels and mitigate any further outages. With a vSphere infrastructure, a hardware failure results in virtual machines powered back on by vSphere HA, restoring availability levels quickly and keeping utilization balanced. The following sections provide recommendations for using vSphere HA with Exchange 2013.

3.4.1 Admission Control

The amount of failover capacity provided by hosts in a vSphere cluster determines how many host failures can be tolerated. For example, in a four host cluster, each host needs to reserve at least 25% of its resources to accommodate the workloads of one host in the event of a failure. Reserving this capacity is a manual task without admission control.

vSphere Admission Control allows administrators to define policies that are used by vCenter to manage failover capacity. The policies determine how failover capacity is reserved and are as follows:

- The number of host failures a cluster can tolerate.
- A percentage of the cluster resources that are reserved as failover spare capacity.
- Dedicated failover hosts.

Surviving a physical host failure without compromising performance or availability is a major driver for virtualizing business critical applications. By configuring admission control, VMware vCenter Server™ monitors utilization and manages failover capacity. To provide sufficient resources in the case of a hardware failure, or even during maintenance, enable and configure admission control.

For more information on admission control and the policies see *vSphere HA Admission Control* in *ESXi and vCenter Server 5 Documentation* (http://pubs.vmware.com/vsphere-50/index.jsp?topic=%2Fcom.vmware.vsphere.avail.doc_50%2FGUID-53F6938C-96E5-4F67-9A6E-479F5A894571.html).

3.4.2 Virtual Machine Monitoring

Along with ESXi host monitoring, vSphere HA can also provide monitoring at the virtual machine level. Virtual machine monitoring can detect guest operating system failures and, with the help of third-party software, application failures. vSphere HA establishes a heartbeat with VMware Tools, installed within the guest operating system, and monitors this heartbeat. If heartbeat communication fails between VMware Tools and vSphere HA, a secondary check is made against network and storage I/O activity. If both network and storage I/O activity has halted, vSphere HA triggers a virtual machine restart.

Enable virtual machine monitoring if there is a desire to reduce downtime due to guest operating system failures that would otherwise require manual intervention.

3.4.3 Using vSphere HA with Database Availability Groups

In a physical environment DAGs are often deployed with three or more database copies to protect from hardware and disk failures. In these environments, when a physical server or storage component fails, the DAG is still protected due to the multiple database copies. This comes at the expense of managing multiple database copies. Exchange environments built on vSphere are typically designed with two database copies and utilize vSphere HA and RAID to protect from hardware and storage failures. vSphere HA restarts a DAG member if the host where it was running experiences a hardware failure, and RAID protects databases from storage failure.

When enabling a vSphere cluster for HA with the intention of protecting DAG members, consider the following:

- Members of the same DAG should not reside on the same vSphere host for an extended period of time when databases are symmetrically distributed between members. Allowing two members to run on the same host for a short period of time (for instance, after a vSphere HA event) allows database replication to resume. DAG members should be separated as soon as the ESXi host has been restored.
- To adequately protect from an extended server outage, vSphere clusters should be designed in an N+1 configuration, where N is the number of DAG members. If a hardware failure occurs causing vSphere HA to power-on a failed DAG member, the DAG maintains the same level of protection as during normal runtime.
- Use anti-affinity rules to keep DAG members separated. vSphere HA might violate this rule during a power-on operation (one caused by a host failure), but DRS fixes the violation during the next interval. To completely eliminate the possibility of DAG members running on the same host (even for a short period of time) *must not run on* virtual machine to host anti-affinity rules must be used.

4. Exchange Performance on vSphere

Since 2006, VMware and its partners have used testing to successfully demonstrate the viability of running Exchange on the VMware Infrastructure platform. This testing has been confirmed by organizations that have deployed Exchange 2003, 2007, and 2010 in virtualized production environments and now benefit from the considerable operational advantages and cost savings. Many customers have virtualized their entire Exchange 2010 environment and have carefully designed their vSphere infrastructure to accommodate application performance, scalability, and availability requirements.

Exchange Server 2013 is an even greater candidate for virtualization than its predecessors. Architectural changes and improvements to the core of Exchange Server, along with advancements in server hardware, make vSphere the default choice for Exchange 2013.

The shift towards running Exchange virtualization as the default design choice is a result of advancements in three key areas:

- The Exchange information store (the *Managed Store*) has been rewritten to further optimize resource consumption. This update to the Managed Store has also led to further reduction in storage I/O requirements.
- Advances in server hardware such as multicore processors, higher memory density, and advances in storage technology are far outpacing the performance requirements for applications, including Exchange 2013. Virtualization becomes an effective way to leverage the full power of these systems.
- The advances in Exchange Server 2013 and server hardware technology have coincided with advances in vSphere. Virtual machines support up to 1TB RAM and 64 vCPUs and are capable of running even the largest Exchange Mailbox servers.

Third-party testing of Exchange Server 2010 in virtual operation has been completed with the Microsoft Jetstress and Exchange Load Generator (LoadGen) tools, the standard tools for Exchange performance analysis. These tests show that performance for a virtualized Exchange Server is comparable to a non-virtualized server running on the same hardware. This proved to be true for all Exchange Server 2010 server roles.

Although testing utilities for Exchange 2013 are not available at the time of this publication, Microsoft has stated that storage and Exchange role sizing for Exchange 2010 can continue to be used, pending these tools for Exchange 2013. With concerns over relative performance eliminated, many more Exchange administrators are finding the flexibility, enhanced availability, and lower costs associated with virtualization very attractive in supporting an Exchange infrastructure.

4.1 Key Performance Considerations

A variety of factors can affect Exchange Server 2013 performance on vSphere, including processor and memory allocation to the guest virtual machine, storage layout and design, virtual machine placement, and high availability methods. The following are tips for achieving the best possible performance:

- Fully understand your organization’s business and technical requirements for implementing Exchange.
- Fully understand the Exchange workload requirements. Current workloads can be measured using the *Microsoft Exchange Server Profile Analyzer* (http://www.microsoft.com/download/en/details.aspx?displaylang=en&id=10559) for environments running Exchange 2003 and 2007. For environments running Exchange 2010, use current Mailbox server utilization as a baseline.
- Although I/O is reduced in Exchange 2013 over Exchange 2010, there is still a requirement to provide adequate throughput and low latency. Dedicate physical storage for Exchange to avoid compromising I/O by having other workloads running on the same physical disks.
- Use Microsoft sizing and configuration guidelines for the Exchange virtual machines.

- Follow the best practices in Section 2, ESXi Host Best Practices for Exchange, to optimize the ESXi host environment for enterprise applications such as Exchange.

4.2 Performance Testing

Every Exchange environment is different, with varying business and technical requirements, many server and storage options, and requirements for integrating with third-party software solutions such as antivirus, anti-spam, and smartphones. Due to the many variables, it is strongly recommended that each organization test performance on their particular mix of server, storage, and software to determine the best design for their Exchange environment. In addition, several VMware server and storage partners have performed testing to validate Exchange performance on vSphere. Both of these options are discussed in this section.

4.2.1 Internal Performance Testing

Microsoft provides tools to measure the performance of Microsoft Exchange Server architectures. LoadGen is used to measure performance of the entire Exchange Server environment. For storage qualification, Jetstress can be used. Both tools have been written specifically for Exchange 2010. At the time of this publication, the equivalent tools for Exchange 2013 have not yet been released, however Microsoft has stated that Jetstress for Exchange 2010 can be used to validate a storage configuration for Exchange 2013.

Note The reduction in storage I/O in Exchange 2013 might lead to an oversized proposed configuration when using Exchange 2010 tools.

It is important to address a concern with the collection of performance metrics from within virtual machines. Early in the virtualization of high-performance applications, the validity of in-guest performance metrics came into question because of a time skew that can be possible in highly overcommitted environments. With the advancements in hypervisor technology and server hardware, this issue has mostly been addressed, especially when testing is performed on undercommitted hardware. This is validated by Microsoft support for running Jetstress within virtual machines. More information on virtual machine support for Jetstress is available in the Microsoft TechNet article *Microsoft Exchange Server Jetstress 2010* (http://technet.microsoft.com/en-us/library/ff706601(v=exchg.141).aspx).

4.2.2 Partner Performance Testing

VMware and its OEM partners have been working together for years to characterize Exchange performance. This testing helps to understand the performance of Exchange in a virtualized environment, qualify best practice recommendations, and better understand any virtualization overhead impact on Exchange. At this time Exchange 2013 is still a new product, and Microsoft has not yet released LoadGen for Exchange 2013. However, the expectation is that overall, Exchange 2013 will have very similar performance characteristics to Exchange 2010.

The following table summarizes VMware and partner performance testing for Exchange 2010 running on vSphere.

Table 2. Performance Testing Summary

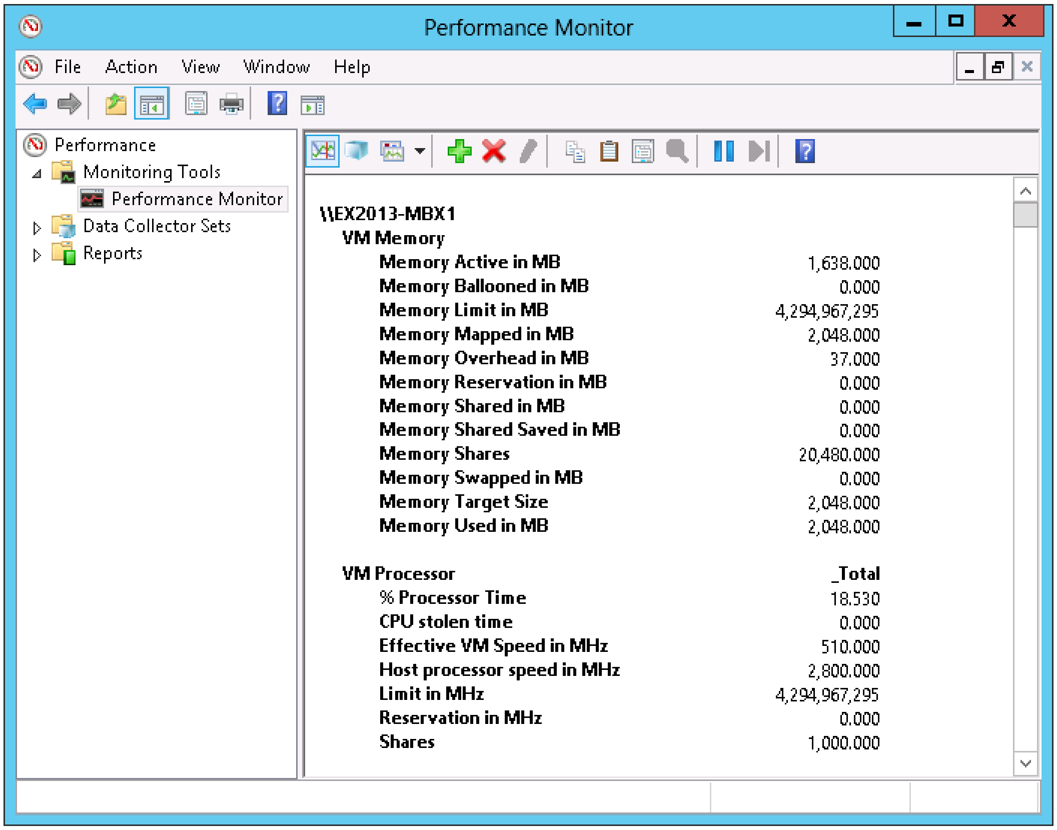
Partner	Type	Summary	Resource
VMware	Performance (2010 on vSphere 5)	This paper examines the performance implications of scaling up an Exchange Server 2010 Mailbox server virtual machine in vSphere 5 in comparison to vSphere 4.1 and scaling out with multiple Exchange Server 2010 virtual machines.	*Microsoft Exchange Server 2010 Performance on VMware vSphere 5* (http://www.vmware.com/files/pdf/exchange-perf-vsphere5.pdf)
HP	Reference Architecture (2010 SP1 on vSphere 5)	This reference architecture describes the benefits and capabilities of HP ProLiant BL460c Gen8 server blades and 3PAR V400 storage solution supporting 15,000 mailbox users running Exchange 2010 in virtualized environments using VMware vSphere 5.	*Implementing HP 3PAR V400 and ProLiant BL460c Gen8 with Microsoft Exchange 2010 running on VMware vSphere 5* (http://h20195.www2.hp.com/V2/GetDocument.aspx?docname=4AA4-3845ENW&cc=us&lc=en)
EMC	Solution Overview (2010 on vSphere 5)	This is a business continuity solution designed for enterprises that require application protection both locally, within a datacenter, and also across multiple datacenters within a metropolitan area or across the globe. The solution offers high availability at every location and automated recovery from a disaster at any location.	*EMC Protection for Microsoft Exchange Server 2010* (http://www.emc.com/collateral/hardware/white-papers/h8891-emc-protection-fr-microsoft-exchng-servr2010.pdf)
Dell	Storage Sizing and Performance (2010 on vSphere 4.1)	This provides server and storage sizing best practices for deploying Exchange 2010 on vSphere and Dell PowerEdge blade servers with a Dell EqualLogic SAN.	*Sizing and Best Practices for Deploying Microsoft Exchange Server 2010 on VMware vSphere and Dell EqualLogic Storage* (http://i.dell.com/sites/content/business/solutions/whitepapers/en/Documents/vmware-vsphere-equallogic.pdf)

EMC	Storage Sizing and Performance (2010 on vSphere 4.1)	This presents methodologies and guidelines for designing a scalable Exchange 2010 environment on EMC VNX5700 series unified storage and vSphere.	*Microsoft Exchange Server 2010 Performance Review Using the EMC VNX5700 Unified Storage Platform* (http://www.emc.com/collateral/hardware/white-papers/h8152-exchange-performance-vnx-wp.pdf)

4.3 Ongoing Performance Monitoring and Tuning

Traditional Exchange Server performance monitoring leverages the Microsoft Windows performance monitor tool Perfmon to collect statistics. Exchange integrates with Perfmon to provide familiar counters that indicate system performance. Exchange administrators should continue to use familiar tools to monitor performance, especially for Exchange-specific counters such as RPC Averaged Latency. In addition to the standard counters familiar to an Exchange administrator, VMware Tools adds two additional Perfmon counters that can be monitored—VM Memory and VM Processor. These counters provide ESXi host-level insight into the resource allocation and usage of the virtual machine.

Figure 12. Virtual Machine Perfmon Counters



```
\\EX2013-MBX1
  VM Memory
     Memory Active in MB                    1,638.000
     Memory Ballooned in MB                     0.000
     Memory Limit in MB                 4,294,967,295
     Memory Mapped in MB                    2,048.000
     Memory Overhead in MB                     37.000
     Memory Reservation in MB                   0.000
     Memory Shared in MB                        0.000
     Memory Shared Saved in MB                  0.000
     Memory Shares                         20,480.000
     Memory Swapped in MB                       0.000
     Memory Target Size                     2,048.000
     Memory Used in MB                      2,048.000

  VM Processor                                 _Total
     % Processor Time                          18.530
     CPU stolen time                            0.000
     Effective VM Speed in MHz                510.000
     Host processor speed in MHz            2,800.000
     Limit in MHz                       4,294,967,295
     Reservation in MHz                         0.000
     Shares                                 1,000.000
```

Many of the counters available can be used to help confirm allocations have been set properly when vCenter Server access is not available or for configuration monitoring. The following table lists counters that can be actively monitored.

Table 3. Virtual Machine Perfmon Counters of Interest

Object	Counter	Description
VM Processor	% Processor Time	Processor usage across all vCPUs.
VM Memory	Memory Ballooned	Amount of memory in MB reclaimed by balloon driver.
	Memory Swapped	Amount of memory in MB forcibly swapped to ESXi host swap.
	Memory Used	Physical memory in use by the virtual machine.

vSphere and Exchange administrators can also use the counters listed in the following table to monitor performance at the ESXi host level. Those metrics can then be correlated with metrics from Exchange virtual machines. Refer to *Performance Monitoring and Analysis* (http://communities.vmware.com/docs/DOC-3930) for more information on these counters and their interpretation.

Table 4. VMware Performance Counters of Interest to Exchange Administrators

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	Device Latency
	KAVG/cmd	Kernel Latency
Network	MbRX/s	packetsRx
	MbTX/s	packetsTx

The preceding table indicates a few key counters that should be added to the list of inspection points for Exchange administrators. Of the CPU counters, the total used time indicates system load. Ready time indicates overloaded CPU resources. A significant swap rate in the memory counters is a clear indication of a shortage of 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 Exchange performance problems, except when large amounts of iSCSI storage traffic are using a single network line. Check total throughput on the NICs to see whether the network is saturated.

5. VMware Enhancements for Deployment and Operations

VMware vSphere provides core virtualization functionality. The extensive software portfolio offered by VMware is designed to help customers to achieve the ultimate goal of 100% virtualization and the *Software-Defined Datacenter*. This section reviews some of the VMware products that can be used in an Exchange 2013 environment virtualized on vSphere.

5.1 VMware vCloud Networking and Security

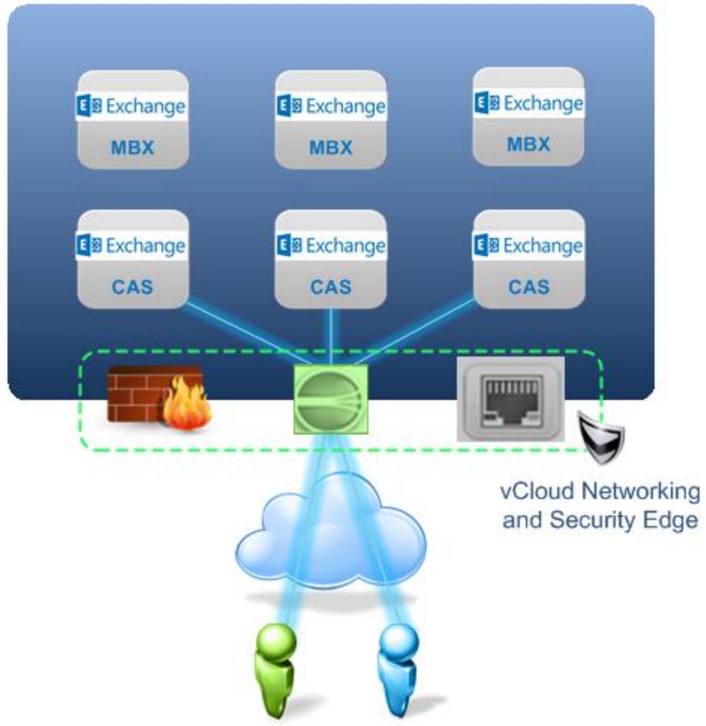
Although virtualization has allowed organizations to optimize their compute and storage investments, the network has mostly remained physical. VMware vCloud® Network and Security™ solves datacenter challenges found in physical network environments by delivering software-defined network and security. Using existing vSphere compute resources, network services can be delivered quickly to respond to business challenges.

5.1.1 VMware vCloud Networking and Security Edge

Client Access servers in a CAS array must be load balanced to provide a highly available and well-performing experience for end users. To provide this functionality, hardware load balancers are deployed in front of a CAS array. If the load balancer solution must be highly available, this can double the hardware investment required. In multisite deployments this can mean up to four hardware load balancers to provide a highly available load balancing solution.

VMware vCloud Networking and Security Edge™ provides load balancing for virtual machines through a virtual appliance. vCloud Networking and Security Edge can be deployed in a high availability pair, providing better protection than hardware load balancing solutions without the additional hardware or management overhead. vCloud Networking and Security Edge supports both Layer 4 (recommended for Exchange 2013) and Layer 7 load balancing of HTTP and HTTPS protocols and supports multiple load balancing methods, such as round robin and least connection.

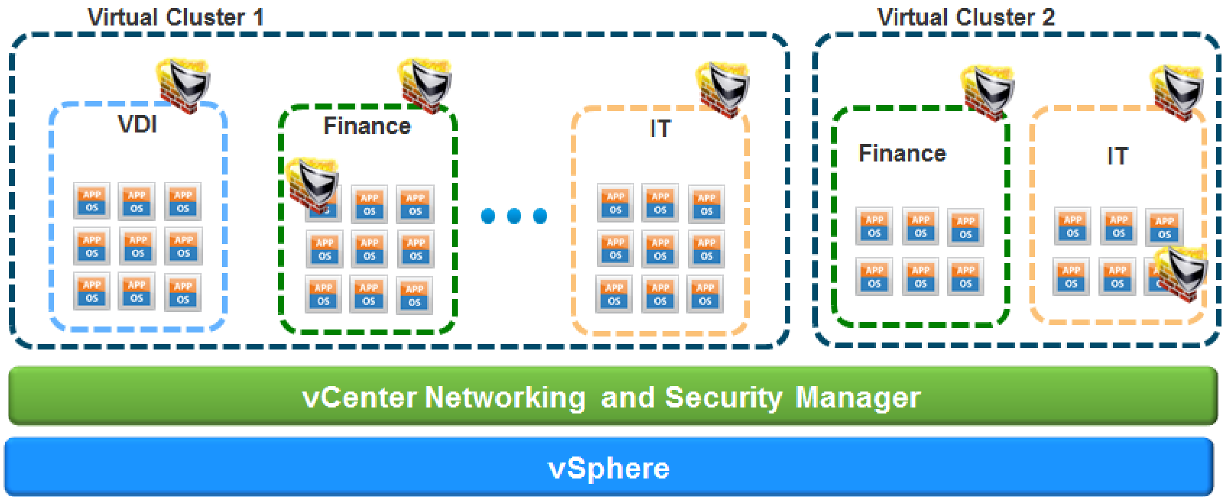
Figure 13. vCloud Networking and Security Edge



5.1.2 VMware vCloud Networking and Security App

Exchange 2013 can leverage VMware vCloud Networking and Security App™ to provide application layer isolation against unauthorized access. Isolation at this level typically requires hardware firewalls and multiple VLANs in a physical networking environment. With vCloud Networking and Security App, this capability is delivered in software through virtual appliances and on the existing vSphere infrastructure.

Figure 14. vCloud Networking and Security App Capability

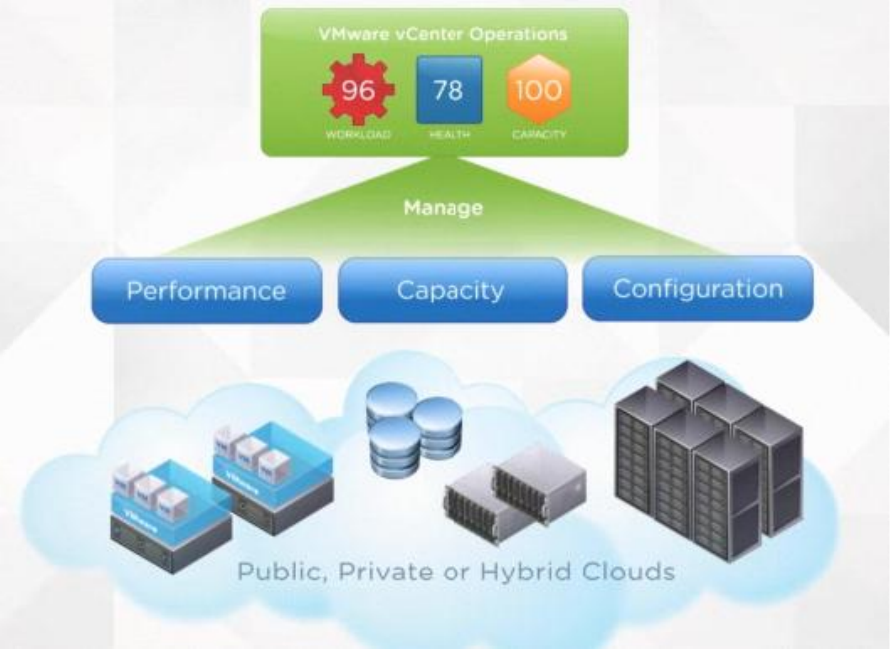


5.2 VMware vCenter Operations Manager

VMware vCenter Operations Manager™ can provide a holistic approach to performance, capacity, and configuration management. By using patented analytics, service levels can be proactively monitored and maintained. When performance or capacity problems arise in your Exchange environment, vCenter Operations Manager is able to analyze metrics from the application all the way through to the infrastructure to provide insight into problematic components, whether they are compute (physical or virtual), storage, networking, OS, or application related. By establishing trends over time, vCenter Operations Manager can minimize false alerts and proactively alert on the potential root cause of increasing performance problems before end users are impacted.

In an Exchange environment constant monitoring is required to maintain acceptable service levels, not only for end users, but also for the Exchange components. vCenter Operations Manager includes patented capacity analytics that can eliminate the need for spreadsheets, scripts, or rules of thumb. Quickly run through “what if” capacity scenarios to understand growth trends and identify upcoming compute power shortages or over-provisioned resources. As an application comprising multiple components, Exchange performance and functionality can be affected by changes made at many levels. vCenter Operations Manager monitors configurations across virtual machines and detects unwanted changes to help maintain continuous compliance with operational best practices.

Figure 15. vCenter Operations



5.3 VMware vCenter Site Recovery Manager

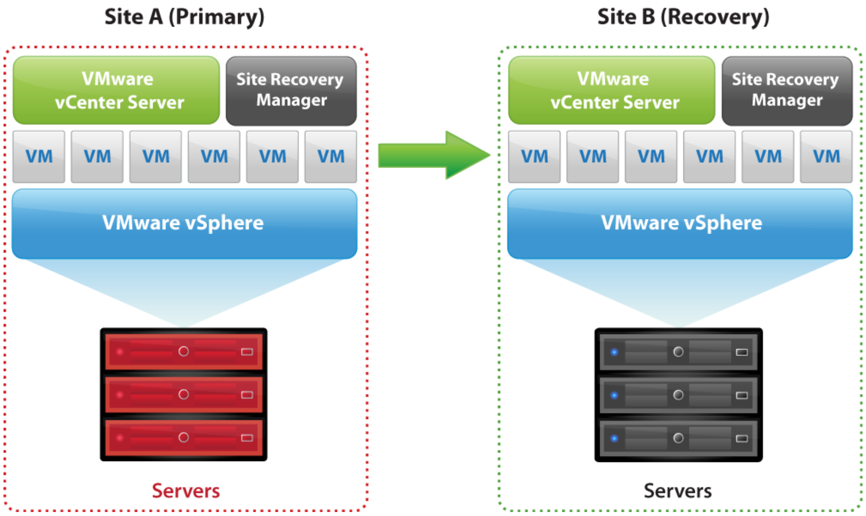
vCenter Site Recovery Manager takes advantage of virtual machine encapsulation to make testing and initiating disaster recovery (DR) failover a straightforward, integrated vCenter process. vCenter Site Recovery Manager runs alongside vCenter Server to provide planning, testing, and automated recovery in the case of a disaster. By using VMware vSphere Replication or storage-based replication technology, vCenter Site Recovery Manager eliminates the manual steps required during a failover scenario to provide consistent and predictable results.

High-level steps that can be performed during a failover test or actual failover

1. Shut down production virtual machines (failover).
2. Promote recovery storage to primary (failover).
3. Take and mount a snapshot of recovery storage in read/write mode (test only).
4. Rescan ESXi hosts to make storage visible.
5. Register recovery virtual machines.
6. Power on virtual machines at the recovery site.
7. Reconfigure IP address settings and update DNS, if required.
8. Verify that VMware Tools starts successfully on recovered virtual machines.
9. Power off recovered virtual machines (test only).
10. Unregister virtual machines (test only).
11. Remove storage snapshot from the recovery side (test only).

Exchange 2013 DAGs can provide high availability by implementing local and remote replication. Although DAG is an excellent choice for datacenter high availability, the application-centric nature of a DAG might not be in line with a company's disaster recovery plans. vCenter Site Recovery Manager is not a replacement for application-aware clustering solutions that might be deployed within the guest operating system. vCenter Site Recovery Manager provides integration of the replication solution, VMware vSphere, and optionally customer-developed scripts to provide a simple, repeatable, and reportable process for disaster recovery of the entire virtual environment, regardless of the application.

Figure 16. vCenter Site Recovery Manager


Site A (Primary)
Site B (Recovery)
VMware vCenter Server
Site Recovery Manager
VM
VM
VM
VM
VM
VM
VMware vSphere
Servers
VMware vCenter Server
Site Recovery Manager
VM
VM
VM
VM
VM
VM
VMware vSphere
Servers