



What's New in Performance in VMware vSphere™ 5.0

TECHNICAL WHITE PAPER

Introduction

VMware vSphere™ 5.0 (“vSphere”), the best VMware® solution for building cloud infrastructures, pushes further ahead in performance and scalability. vSphere 5.0 enables higher consolidation ratios with unequaled performance. It supports the build-out of private and hybrid clouds at even lower operational costs than before.

The following are some of the performance highlights:

VMware vCenter™ Server scalability

- Faster high-availability (HA) configuration times
- Faster failover rates – 60% more virtual machines within the same time
- Lower management operational latencies
- Higher management operations throughput (Ops/min)

Compute

- 32-way vCPU scalability
- 1TB memory support

Storage

- vSphere® Storage I/O Control (Storage I/O Control) now supports NFS – Set storage quality of service priorities per virtual machine for better access to storage resources for high-priority applications

Network

- vSphere® Network I/O Control (Network I/O Control) – Gives a higher granularity of network load balancing

vSphere vMotion®

- vSphere vMotion® (vMotion) – Multi-network adaptor enablement that contributes to an even faster vMotion
- vSphere Storage vMotion® (Storage vMotion) – Fast, live storage migration with I/O mirroring

VMware vCenter Server and Scalability Enhancements

Virtual Machines per Cluster

Even with the increased number of supported virtual machines per cluster, vSphere 5.0 delivers a higher throughput (Ops/min) of management operations—up to 120%, depending on the intensity of the operations. Examples of management operations include virtual machine power-on, virtual machine power-off, virtual machine migrate, virtual machine register, virtual machine unregister, create folder, and so on. Operation latency improvements range from 25% to 75%, depending on the type of operation and the background load.

HA (High Availability)

Significant advances have been made in vSphere 5.0 to reduce configuration time and improve failover performance. A 32-host, 3,200-virtual machine, HA-enabled cluster can be configured 9x faster in vSphere 5.0. In the same amount of time as with previous editions, vSphere 5.0 can fail over 60% more virtual machines. The minimal recovery time, from failure to the first virtual machine restarts, has been improved by 44.4%. The average virtual machine failover time has improved by 14%.

In addition, in vSphere 5.0, the default CPU slot size—that is, spare CPU capacity reserved for each HA-protected virtual machine—is smaller in comparison to its value in vSphere 4.1, to enable a higher consolidation ratio. As a result, Distributed Power Management (DPM) might be able to put more hosts into standby mode when the cluster utilization is low, leading to more power savings.

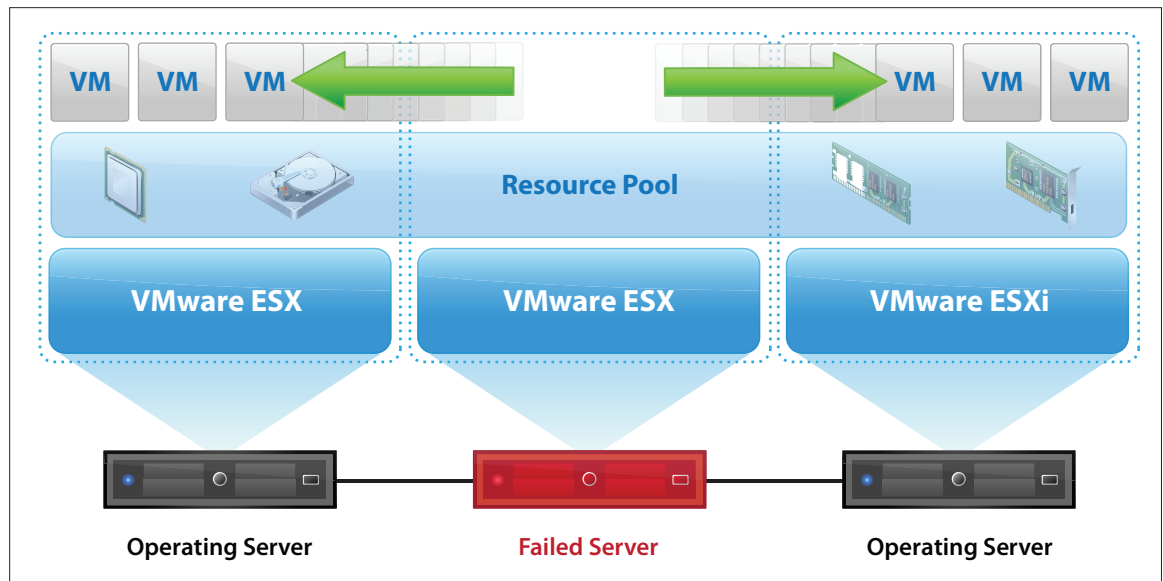


Figure 1. vSphere 5.0 High Availability

CPU Enhancements

32-vCPU performance and scalability

VMware vSphere 5.0 can now support up to 32 vCPUs. Using a large number of vCPUs in a virtual machine can potentially help large-scale, mission-critical, Tier 1 applications achieve higher performance and throughput and/or faster runtime for certain applications.

The VMware Performance Engineering lab conducted a variety of experiments, including commercial Tier 1 and HPC (high-performance computing) applications. Performance was observed close to native, 92-97%, as virtualized applications scaled to 32 vCPUs.

Intel SMT-related CPU scheduler enhancements

Intel SMT architecture exposes two hardware contexts from a single core. The benefit of utilizing two hardware contexts ranges from 10% to 30% in improved application performance, depending on the workloads. In vSphere 5.0, the VMware® ESXi™ CPU scheduler's policy is tuned for this type of architecture to balance between maximum throughput and fairness between virtual machines. In addition to the number of performance optimizations made around SMT CPU scheduling in vSphere 4.1 (running the VMware ESXi hypervisor architecture), we have further enhanced the SMT scheduler in VMware ESXi 5.0 to ensure high efficiency and performance for mission-critical applications.

Virtual NUMA (vNUMA)

Virtual NUMA (vNUMA) exposes host NUMA topology to the guest operating system, enabling NUMA-aware guest operating systems and applications to make the most efficient use of the underlying hardware's NUMA architecture.

Virtual NUMA, which requires virtual hardware version 8, can provide significant performance benefits for virtualized operating systems and applications that feature NUMA optimization.

Storage Enhancements

NFS support for Storage I/O Control

vSphere 5.0 Storage I/O Control now supports NFS and network-attached storage (NAS)-based shares on shared datastores. It now provides the following benefits for NFS:

- Dynamically regulates multiple virtual machines' access to shared I/O resources, based on disk shares assigned to the virtual machines.
- Helps isolate performance of latency-sensitive applications that employ smaller (<8KB) random I/O requests. This has been shown to increase performance by as much as 20%.
- Redistributes unutilized resources to those virtual machines that need them in proportion to the virtual machines' disk shares. This results in a fair allocation of storage resources without any loss in utilization.
- Limits performance fluctuations of a critical workload to a small range during periods of I/O congestion. This results in up to an 11% performance benefit compared to that in an unmanaged scenario without Storage I/O Control.

Storage I/O Control provides a dynamic control mechanism for managing virtual machines' access to I/O resources, Fibre Channel (FC), iSCSI or NFS, in a cluster. It delivers the same performance benefits to NFS datastores as it does to already supported FC or iSCSI datastores. Tests have shown that with the right balance of workloads, Storage I/O Control might improve performance of critical applications by as much as 10%, with a latency improvement per I/O operation of as much as 33%.

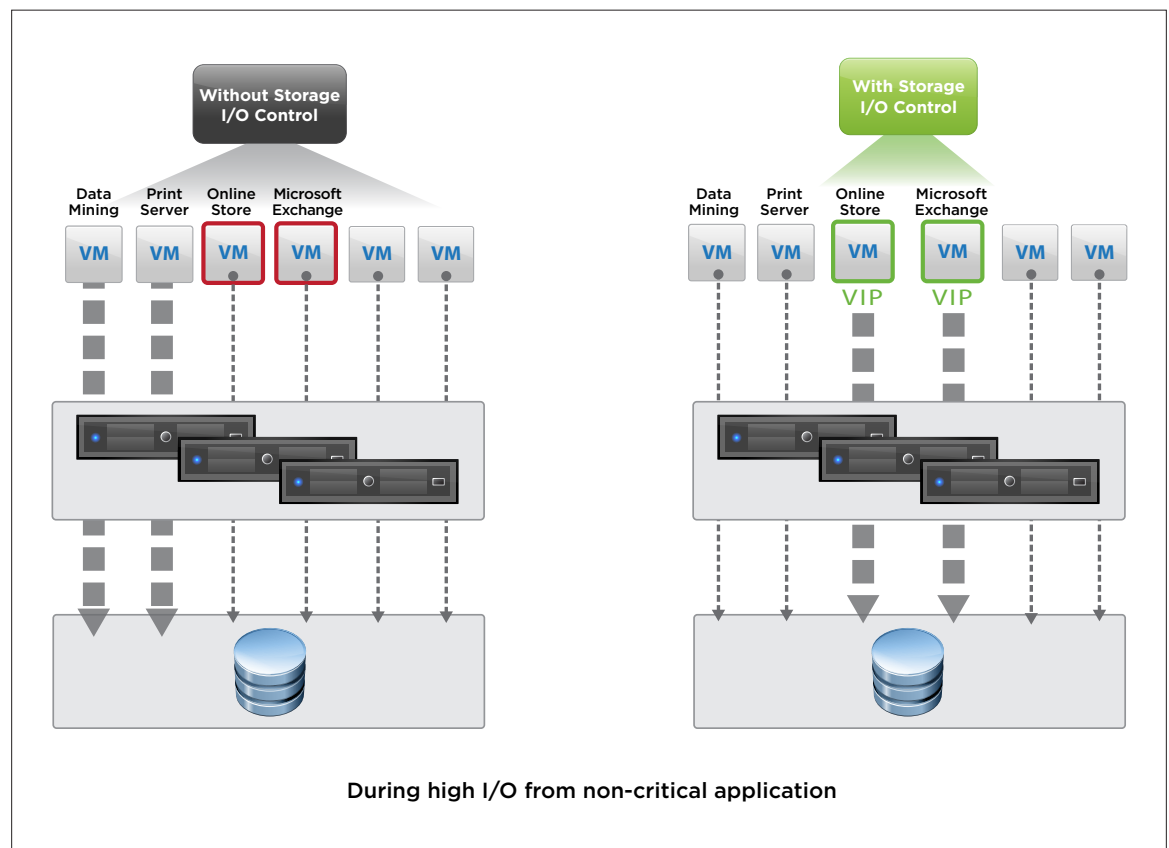


Figure 2. vSphere 5.0 Storage I/O Control

Memory Management Enhancements

1TB vMem support

In vSphere 5.0, virtual machines can now support vMem up to 1TB. With the new 1TB vMem support, Tier 1 applications consuming large amounts of memory (> 255GB), such as in-memory databases, can now be virtualized. Experimental measurements with memory intensive workloads for virtual machines with up to 1TB vMem demonstrate similar performance when compared to identical physical configurations.

SSD Swap Cache

vSphere 5.0 can be configured to enable VMware ESXi host swapping to a solid-state disk (SSD). In the low host memory-available states (high memory usage), where guest ballooning, transparent page sharing (TPS) and memory compression have not been sufficient to reclaim the needed host memory, hypervisor swapping is used as the last resort to reclaim memory from the virtual machines. vSphere employs three methods to address limitations of hypervisor swapping to improve hypervisor swapping performance:

- Randomly selecting the virtual machine physical pages to be swapped out. This helps mitigate the impact of VMware ESXi pathologically interacting with the guest operating system's memory management heuristics. This has been enabled from early releases of VMware ESX®.
- Memory compression of the virtual machine pages that are targeted by VMware ESXi to be swapped out. This feature, introduced in vSphere 4.1, reduces the number of memory pages that must be swapped out to the disk, while reclaiming host memory effectively at the same time and thereby benefiting application performance.
- vSphere 5.0 now enables users to choose to configure a swap cache on the SSD. VMware ESXi 5.0 will then use this swap cache to store the swapped-out pages instead of sending them to the regular and slower hypervisor swap file on the disk. Upon the next access to a page in the swap cache, the page will be retrieved quickly from the cache and then removed from the swap cache to free up space. Because SSD read latencies are an order of magnitude faster than typical disk access latencies, this significantly reduces the swap-in latencies and greatly improves the application performance in high memory over commitment scenarios.

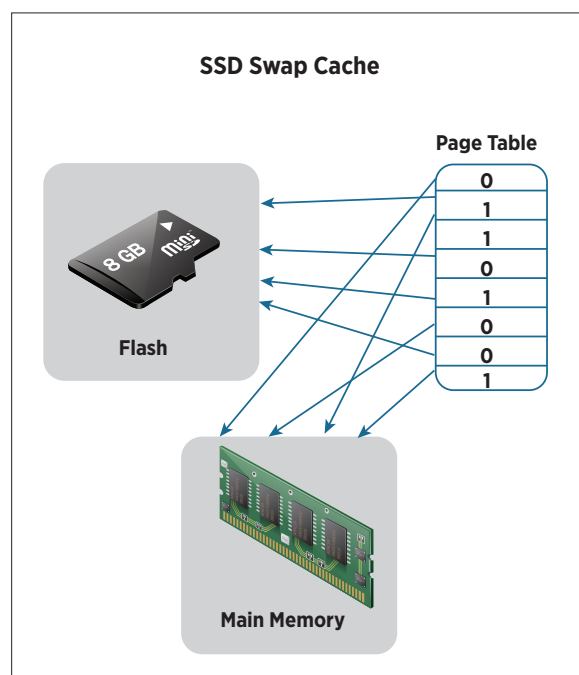


Figure 3. SSD Swap Cache

Network Enhancements

Network I/O Control

Network I/O Control (NIOC) enables the allocation of network bandwidth to resource pools. In vSphere 5.0, users can create new resource pools to associate with port groups and specify 802.1p tags, whereas in vSphere 4.1, only predefined resource pools could be used for bandwidth allocation. So different virtual machines can now be in different resource pools and be given a higher or lower share of bandwidth than the others. The predefined resource groups include vMotion, NFS, iSCSI, vSphere® Fault Tolerance (Fault Tolerance), virtual machine, and management.

splitRxMode

vSphere 5.0 introduces a new way of doing network packet receive processing in the VMkernel. In previous releases of vSphere, all receive packet processing for a network queue was done in a single context, which might be shared among various virtual machines. In cases where there is a higher density of virtual machines per network queue, it is possible for the context to become resource constrained. Multicast is one such workload where multiple receiver virtual machines share one network queue. vSphere 5.0 includes a new mechanism to split the cost of receive packet processing to multiple contexts. SplitRxMode can specify for each virtual machine whether receive packet processing should be in the network queue context or a separate context.

We called this mode splitRxMode, and it can be enabled on a per-vNIC basis by editing the VMX file of the virtual machine with `ethernetX.emuRxMode = "1"` for the Ethernet device.

NOTE: This configuration is available only with vmxnet3. Once the option is enabled, the packet processing for the virtual machine is handled by a separate context. This improves the multicast performance significantly. However, there is an overhead associated with splitting the cost among various PCPUs. This cost can be attributed to higher CPU utilization per packet processed, so VMware recommends enabling this option for only those workloads, such as multicast, that are known to benefit from this feature.

VMware performance labs conducted a performance study for multiple receivers with the sender transmitting multicast packets at 16,000 packets per second. As each additional receiver was added, there was an increase in packets handled by the networking context. After the load of the context reached its maximum, there were significant packet losses. By enabling splitRxMode on these virtual machines, the load on the context was decreased and thereby increasing the number of multicast receivers handled without significant drops. Figure 4 shows that there is no noticeable difference in loss rate for the setup until there are 24 virtual machines powered on. After there are more than 24 virtual machines powered on, there is an almost 10–25% packet loss. Enabling splitRxMode for the virtual machines reduces the loss rate to less than 0.01%.

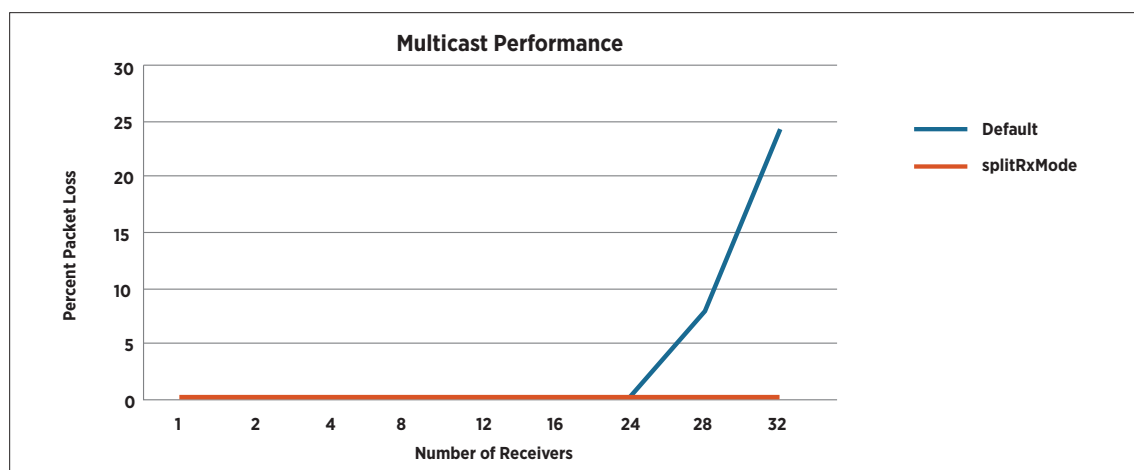


Figure 4. Multicast Performance

vSphere DirectPath I/O

vSphere® DirectPath I/O (DirectPath I/O) enables guest operating systems to directly access network devices. DirectPath I/O for vSphere 5.0 has been enhanced to allow the vMotion of a virtual machine containing DirectPath I/O network adaptors on the Cisco Unified Computing System (UCS) platform. On such platforms, with proper setup and ports available for passthrough, vSphere 5.0 can transition the device from one that is paravirtualized to one that is directly accessed, and the other way around. While both modes can sustain high throughput (beyond 10Gbps), direct access can additionally save CPU cycles in workloads with high packet rates (for example, approximately greater than 50kps).

Because DirectPath I/O does not support some vSphere features—memory overcommitment and NIOC—VMware recommends using DirectPath I/O only for workloads with very high packet rates, where saving CPU cycles can be critical to achieving the desired performance.

Host Power Management Enhancements

VMware® Host Power Management (VMware HPM) in vSphere 5.0 provides power savings working at the host level. With vSphere 5.0, the default power management policy is “balanced.” The balanced policy uses an algorithm that exploits a processor’s P-states. Workloads operating under low CPU loads can expect to see some power savings (depending entirely on the workload CPU usage characteristics) with minimal performance loss (dependent on application CPU usage characteristics). VMware strongly advises performance-sensitive users to conduct controlled experiments to observe the power savings/performance of their application with vSphere 5.0. If the performance loss is unacceptable, switch VMware ESX to the “high performance” mode. Most of the workloads operating under low loads can see up to 5% in power savings when using balanced mode.

VMware vMotion

Multi-network adaptor enablement for vMotion

The new performance enhancements in vSphere 5.0 enable vMotion to effectively saturate a 10GbE network adaptor bandwidth during the migration, significantly reducing vMotion transfer times. In addition, VMware ESX 5.0 adds a multi-network adaptor feature that enables users to employ multiple network adaptors for vMotion. The VMkernel will transparently load-balance the vMotion traffic over all the vMotion-enabled vmknics in an effort to saturate all of the connections. In fact, even when there is a single vMotion, VMkernel uses all the available network adaptors to spread the vMotion traffic.

Figure 5 summarizes the vMotion performance enhancements in vSphere 5.0 over vSphere 4.1. The test environment modeled both a single-instance SQL server virtual machine (virtual machine configured with four vCPUs and 16GB memory and running OLTP workload) and multiple-instance SQL server virtual machine (each virtual machine configured with four vCPUs and 16GB memory and running OLTP workload) deployment scenarios for the following configurations:

1. vSphere 4.1: Source/Destination hosts configured with a single 10GbE port for vMotion
2. vSphere 5.0: Source/Destination hosts configured with a single 10GbE port for vMotion
3. vSphere 5.0: Source/Destination hosts configured with two 10GbE ports for vMotion

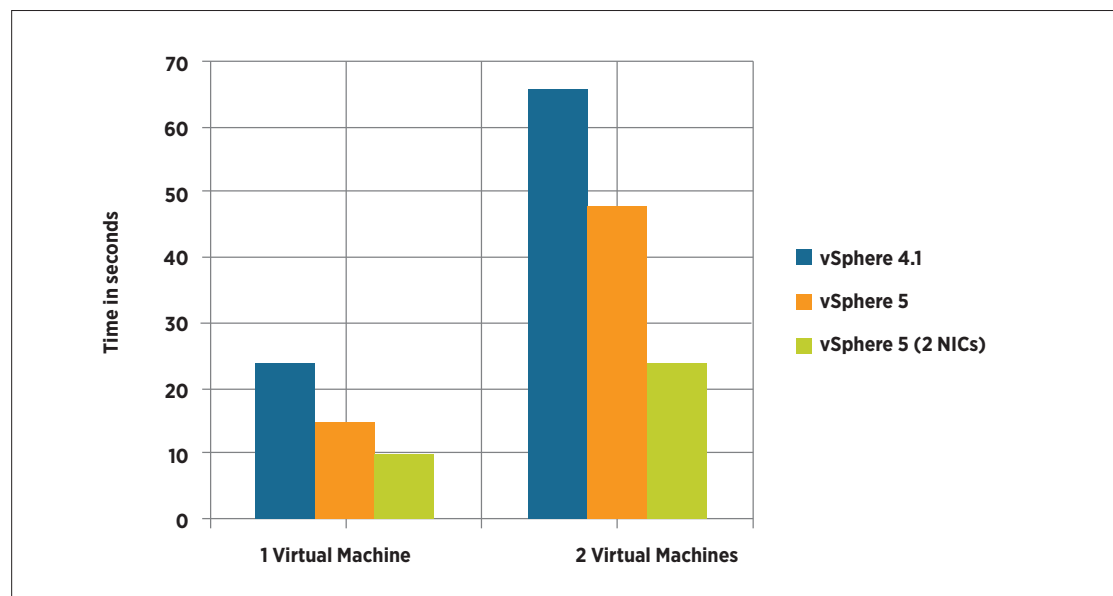


Figure 5. vMotion Performance in vSphere 4.1 and vSphere 5.0

The figure clearly shows the enhancements made in vSphere 5.0 to reduce the total elapsed time in both single-virtual machine and multiple-instance SQL server virtual machine deployment scenarios. The average percentage drop in vMotion transfer time was a little over 30% in vSphere 5.0 when using a single 10GbE network adaptor. When using two 10GbE network adaptors for vMotion (enabled by the new multi-network adaptor feature in vSphere 5.0), the total migration time reduced dramatically: a factor of 2.3x improvement over the vSphere 4.1 result. The figure also illustrates the fact that the multi-network adaptor feature transparently load balances the vMotion traffic onto multiple network adaptors, even in the case where a single virtual machine is subjected to vMotion. This feature can be especially handy when a virtual machine is configured with a large amount of memory.

Metro vMotion

vSphere 5.0 introduces a new latency-aware “Metro vMotion” feature that provides better performance over long latency networks and also increases the round-trip latency limit for vMotion networks from 5 milliseconds to 10 milliseconds. Previously, vMotion was supported only on networks with round-trip latencies of up to 5 milliseconds.

VMware Storage vMotion (Live Migration) Enhancements

vSphere 5.0 now supports live migration for storage with the use of I/O mirroring. In previous releases, live migration moved memory and device state only for a virtual machine, limiting migration to hosts with identical shared storage. Live storage migration overcomes this limitation by enabling the movement of virtual disks spanning volumes and distance. This enables greater virtual machine mobility, zero downtime maintenance and upgrades of storage elements, and automatic storage load balancing.

VMware has been able to greatly reduce the time it takes for the storage to migrate, starting with the use of snapshots in Virtual Infrastructure 3.5, moving to using dirty block tracking in vSphere 4.0, and now to using I/O mirroring in vSphere 5.0.

Live storage migration has the following advantages.

- **Zero downtime maintenance** – Enables customers to move virtual machines on and off storage volumes, upgrade storage arrays, perform file system upgrades, and service hardware without powering down virtual machines.
- **Manual and automatic storage load balancing** – Customers can continue to manually load-balance their vSphere clusters to improve storage performance, or they can take advantage of automatic storage load balancing (vSphere® Storage DRS), which is now available in vSphere 5.0.
- **Live storage migration increases virtual machine mobility** – Virtual machines are no longer pinned to the storage array they are instantiated on. Live migration works by copying the memory and device state of a virtual machine from one host to another with negligible virtual machine downtime.

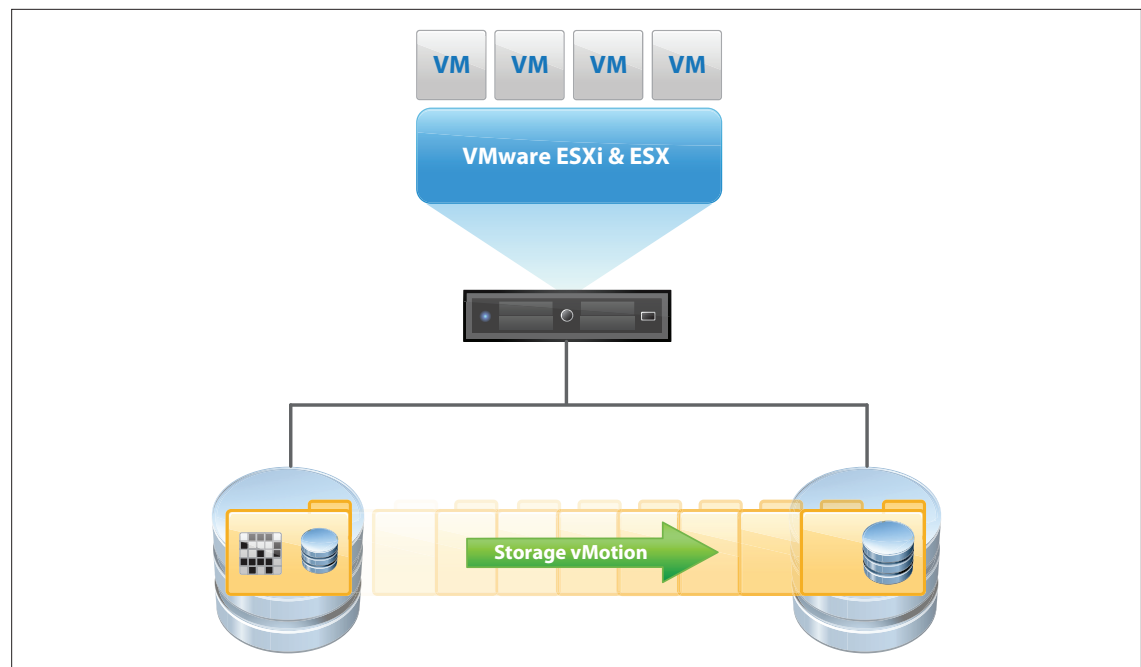


Figure 6. vSphere 5.0 Storage vMotion

Tier 1 Application Performance

Microsoft Exchange Server 2010

Microsoft Exchange Server 2010 significantly benefits from the superior performance and scalability provided by vSphere 5.0. Microsoft Exchange Load Generator 2010 benchmark results indicate that vSphere 5.0 improves Microsoft Exchange transaction response times and vMotion and Storage vMotion migration times when compared with the previous vSphere release.

The following are examples of performance and scalability gains observed in comparison to vSphere 4.1:

- 6% reduction in Send Mail 95th percentile transaction latencies with a four-vCPU Exchange MailBox server virtual machine
- 13% reduction in Send Mail 95th percentile transaction latencies with an eight-vCPU Exchange MailBox server virtual machine
- 33% reduction in vMotion migration time for a four-vCPU Exchange MailBox server virtual machine
- 11% reduction in Storage vMotion migration time for a 350GB Exchange database VMDK

Zimbra Mail Server Performance

The full-featured, robust and open-source Zimbra Collaboration Suite (ZCS) includes email, calendaring, and collaboration software. Zimbra is an important part of the VMware cloud computing portfolio, as software as a service (SaaS) and as a prepackaged virtual appliance for SMBs.

For vSphere 5.0, virtualized ZCS performs within 95% of native performance virtual machine, scaling up to eight vCPUs. Up to eight 4-vCPU ZCS virtual machines supporting 32,000 mail users were scaled out on a single host with just a 10% increase in Sendmail latencies compared to a single virtual machine.

In recent 4-vCPU experiments, ZCS has less than half the CPU consumption of Exchange 2010 with the same number of users, and mail user provision time for ZCS was 10x faster.

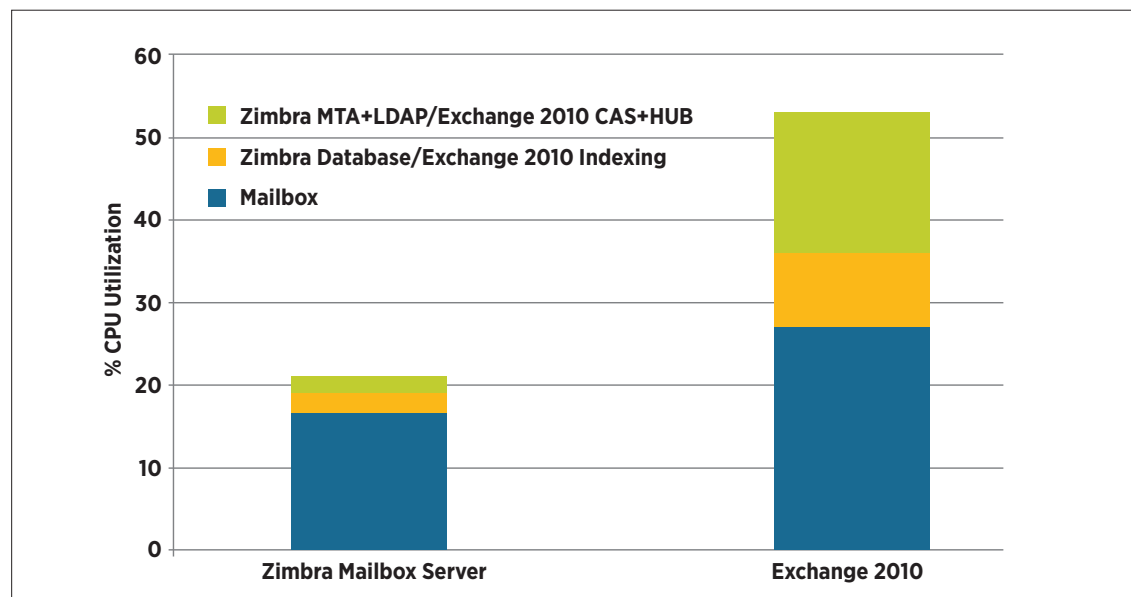


Figure 7. Zimbra CPU Efficiency over Exchange 2010

Summary

VMware innovations continue to ensure that VMware vSphere 5.0 pushes the envelope of performance and scalability. The numerous performance enhancements in vSphere 5.0 enable organizations to get even more out of their virtual infrastructure and further reinforce the role of VMware as industry leader in virtualization.

vSphere 5.0 represents advances in performance, to ensure that even the most resource-intensive applications, such as large databases and Microsoft Exchange email systems, can run on private, hybrid and public clouds powered by vSphere.

References (www.vmware.com)

Understanding Memory Resource Management in VMware vSphere 5

Managing Performance Variance of Applications Using Storage I/O Control in VMware vSphere 5

Performance Best Practices Guide for VMware vSphere 5

Zimbra Mail Performance on VMware vSphere 5

Host Power Management in VMware vSphere 5

