vmware SOLUTION READINESS

Microsoft SharePoint 2010 on VMware Best Practices Guide

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

Contents

1. Overview 7
1.1 Purpose 8
1.2 Target Audience 8
1.3 Scope 8
2. VMware ESX/ESXi Host Best Practices for SharePoint 2010 9
2.1 CPU Configuration Guidelines 9
2.2 Memory Configuration Guidelines 12
2.3 Storage Guidelines 14
2.4 Networking Guidelines 17
3. SharePoint Performance on vSphere 19
3.1 Overview 19
3.2 vSphere Performance Enhancements 19
3.3 Performance Testing 21
3.4 Ongoing Performance Monitoring and Tuning 25
4. Capacity Management Concepts and Reference 29
4.1 Capacity Management Overview 29
4.2 Workload Modeling and Architectural Design 29
4.3 SQL Server Capacity and Performance 33
4.4 Minimum Resource Requirements for SharePoint Server Roles 37
4.5 vSphere Limitations 38
4.6 Sample Architecture Deployment to vSphere 38
5. vSphere Enhancements for Deployment and Operations 43
5.1 Responding to Changing Workloads 43
5.2 Monitoring Performance with vCenter AppSpeed 46
5.3 Intelligently Monitor Capacity with vCenter CapacityIQ 47

List of Figures

Figure 1. Virtual Machine Memory Settings 12
Figure 2. VMware Storage Virtualization Stack 15
Figure 3. Virtual Networking in vSphere 17
Figure 4. Percentage of Guest vCPUs Utilized in the Three Virtual Machine Configuration 22
Figure 5. Percentage of Guest vCPUs Utilized in the Four Virtual Machine Configuration 22
Figure 6. Percentage of Guest vCPUs Utilized in the Five Virtual Machine Configuration 23
Figure 7. Number of Heavy Users at 1% Concurrency 23
Figure 8. SharePoint 2010 Medium Topology Example 31
Figure 9. Enterprise Intranet Collaboration Environment Technical Case Study 32
Figure 10. Departmental Collaboration Environment Technical Case Study 39
Figure 11. Initial Virtual Machine Placement 42
Figure 12. Dynamic Scaling on VMware 43
Figure 13. Using VMware Hot-Add with an SQL Server 2008 Guest 44
Figure 14. Virtual Machine Templates 45
Figure 15. Monitor Multi-Tiered Application Performance with VMware AppSpeed 46
Figure 16. Right-Size Applications with vCenter CapacityIQ 47

List of Tables

Table 1. Esxtop CPU Performance Metrics 11
Table 2. Esxtop Memory Counters 14
Table 3. vSphere Performance Enhancements 19
Table 4. Virtual Machine Configurations for Each Use Case 21
Table 5. Partner Testing Summary 24
Table 6. Performance Counters of Interest to VMware and SharePoint Administrators 26
Table 7. SharePoint 2010 Performance Counters from TechNet (partial) 27
Table 8. Enterprise Intranet Collaboration Environment Technical Case Study Example 30
Table 9. SharePoint 2007 User Loads from Microsoft TechNet 30
Table 10. Sample Content Database Sizing Criteria 34
Table 11. SQL Server Memory Recommendations 36
Table 12. Web, App, and Single Server Minimum Requirements from Microsoft TechNet 37
Table 13. Database Server Minimum Requirements from Microsoft TechNet 37
Table 14. SharePoint Resource Requirements by Server Role 39
Table 15. Database Server Minimum Requirements from Microsoft TechNet 41
Table 16. Database Server Minimum Requirements from Microsoft TechNet 41
Table 17. Departmental Collaboration Case Study Processor Utilization 42

1. Overview

A Microsoft SharePoint infrastructure is frequently deployed first as a collaboration and document sharing tool. SharePoint makes workgroups more efficient and reduces the cost of coordinated collaboration, while allowing the flexibility needed by workgroups with diverse needs and goals. This flexibility and utility often leads to rapid growth in demand for both capacity and bandwidth as more users leverage these tools to coordinate workflows and manage more documents. More sophisticated SharePoint deployments can optimize business workflow and communications, and can quickly become critical components in everyday commercial operations.

While SharePoint offers great benefits to organizations, these benefits are only realized when the underlying systems are available and perform at an acceptable level. Rapid growth and high availability are difficult features to manage in a traditional IT environment. Accommodating both often requires the high cost of over-designing and over-building at the earliest stages of deployment.

Because SharePoint encourages rapid growth and "viral" proliferation, user goals may conflict with the ability of the IT staff to deliver the desired services when needed within budgetary and manpower constraints. Flexibility is extremely valuable during this early period. If rapid growth and evolution can be supported at realistic costs, SharePoint can become an important tool to rapidly increase everyday productivity. VMware vSphere® can facilitate this capability, allowing organizations to leverage the benefits of SharePoint on a pay-as-you-go basis. Because high availability features are inherent to the vSphere platform, these can be leveraged on demand. By virtualizing SharePoint, the common problems of deploying a complex, high-growth IT service are alleviated, allowing resources to be spent on maximizing the value of the tool in everyday business practice.

Contrast the benefits of a virtual infrastructure with the limitations of a traditional deployment. Using conventional physical infrastructure typically leads to over-provisioning. This creates significant resource underutilization and high system power, cooling, and operating costs. In addition, for complex architectures such as SharePoint 2010, using physical servers and infrastructure may have other limitations such as the following:

- Application delivery is traditionally gated by the need for manual configuration and provisioning for each new application or configuration change on a specific hardware platform. This can be slow, and in an existing infrastructure, lead to excessive downtime. It can also constrain growth to the organization's ability to purchase new hardware. Virtual deployments typically take minutes, can share currently deployed hardware, and can be adjusted "on the fly" when more resources are required.
- Application architectures, such as those provided by SharePoint 2010, are rapidly evolving towards highly distributed, loosely-coupled applications. The conventional x86 computing model, in which applications are tightly coupled to physical servers, is too static and restrictive to efficiently support these complex applications. With a virtual deployment, the architecture can be as modular as is appropriate, without expanding the hardware footprint. The dynamic nature of virtual machines means that the design can grow and adapt as required, and the need for a "perfect" initial design is eliminated.
- Availability becomes a critical factor. In a physical environment, the cost and complexity of server clustering is required to increase availability. In a highly distributed environment, VMware vSphere can increase application availability at a much lower cost than using traditional HA strategies.
- Rising datacenter costs (for power, cooling, floor space, and so on.), even while some server computing resources go under-utilized. It is well understood that server consolidation through virtualization is a significant factor in reducing cost. The higher density available from vSphere versus other products can provide significant savings on hardware and Microsoft licenses.

This paper demonstrates that virtualization with VMware vSphere can minimize the challenges in a SharePoint deployment so that maximum commercial value is realized from an investment in SharePoint systems.

1.1 Purpose

This guide provides best practice guidelines for deploying SharePoint 2010 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 SharePoint implementation. The examples and considerations in this document provide guidance only, and do not represent strict design requirements, because the flexibility of both SharePoint and vSphere allows for a wide variety of valid configurations.

1.2 Target Audience

This guide assumes a basic knowledge and understanding of vSphere and SharePoint 2010.

- Architectural staff can reference this document to gain an understanding of how the system works as a whole as they design and implement various components.
- Infrastructure engineers and administrators can use this document as a catalog of technical capabilities.
- Application owners and administrators can reference this document to understand how SharePoint might fit into a virtualized infrastructure/private cloud environment.
- 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 ESX®/ESXi™ Host Best Practices for SharePoint – This section provides best practice guidelines for properly preparing the vSphere platform for running SharePoint 2010. This section includes guidance in the areas of CPU, memory, storage, and networking.
- SharePoint Performance on vSphere – This section provides background information on SharePoint performance in a virtual machine. It also provides information on official VMware partner testing and guidelines for conducting and measuring internal performance tests.
- SharePoint 2010 Capacity Management Concepts and Reference – Sizing SharePoint 2010 to run in a virtual machine follows many of the same best practices as sizing on physical servers. This section walks through the high points of the capacity management process with references to deeper dive material on Microsoft TechNet.
- vSphere Enhancements for Deployment and Operations – This section provides a brief look at vSphere features and add-ons that enhance deployment and management of SharePoint 2010.

The following topic is out of scope for this document, but may be addressed in other documentation in this solution kit: 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 the *Microsoft SharePoint 2010 on VMware Availability and Recovery Options* document.

This and other guides in this solution kit are limited in focus to deploying SharePoint on vSphere. SharePoint deployments cover a wide subject area, and SharePoint-specific design principles must *always* follow Microsoft guidelines for best results.

2. VMware ESX/ESXi Host Best Practices for SharePoint 2010

After the SharePoint architecture is established, those workloads can be mapped to the vSphere environment. Careful planning of the ESX/ESXi host platform to support SharePoint 2010 is crucial to a successful implementation. In the following sections ESX/ESXi host best practices relating to SharePoint 2010 workloads are discussed in detail.

2.1 CPU Configuration Guidelines

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 x86/x64-based processor cores. Virtual machines with more than one virtual CPU are also called SMP (symmetric multiprocessing) virtual machines. The virtual machine monitor (VMM), or hypervisor, is responsible for virtualizing the CPUs. When a virtual machine starts running, control transfers to the VMM, which virtualizes the guest OS instructions.

2.1.2 Virtual SMP

VMware Virtual Symmetric Multiprocessing (Virtual SMP) enhances virtual machine performance by enabling a single virtual machine to use multiple physical processor cores simultaneously. vSphere supports the use of up to eight virtual CPUs per virtual machine. The biggest advantage of an SMP system is the ability to use multiple processors to execute multiple tasks concurrently, thereby increasing throughput (for example, the number of transactions per second). Only workloads that support parallelization (including multiple processes or multiple threads that can run in parallel) can really benefit from SMP.

The virtual processors from SMP-enabled virtual machines are co-scheduled. That is, if physical processor cores are available, the virtual processors are mapped one-to-one onto physical processors and are then run simultaneously. In other words, if one vCPU in the virtual machine is running, a second vCPU is co-scheduled so that they execute nearly synchronously. Consider the following points when using multiple vCPUs:

- Simplistically, if multiple, idle physical CPUs are not available when the virtual machine wants to run, the virtual machine remains in a special wait state. The time a virtual machine spends in this wait state is called *ready time*.
- Even idle processors perform a limited amount of work in an operating system. In addition to this minimal amount, the ESX host manages these “idle” processors, resulting in some additional work by the hypervisor. These low-utilization vCPUs compete with other vCPUs for system resources.

In VMware ESX 4 and ESXi, the CPU scheduler underwent several improvements to provide better performance and scalability; for details, see the paper *VMware vSphere 4: The CPU Scheduler in VMware ESX 4* at http://www.vmware.com/files/pdf/perf-vsphere-cpu_scheduler.pdf. For example, in ESX 4, the relaxed co-scheduling algorithm was refined so that scheduling constraints due to co-scheduling requirements are further reduced. These improvements resulted in better linear scalability and performance of SMP virtual machines. Consequently, in vSphere, the larger 4-way and 8-way virtual machines used for SharePoint 2010 can be deployed without concern for scalability limitations.

VMware recommends the following practices when considering the allocation of vCPUs for SharePoint 2010:

- Allocate the minimum requirement for production virtual machines based on Microsoft guidelines, the role of the virtual machine, and the size of the environment. Additional vCPUs can be added later if necessary.
- Test, development, and proof-of-concept environments can get along with fewer vCPUs allocated to virtual machines. These environments typically require a fraction of the resources needed to satisfy user demand in production.
- When overcommitting CPU resources (number of vCPUs allocated to running virtual machines is greater than the number of physical cores on a host), monitor the responsiveness of SharePoint to understand the level of overcommitment which can be provided while still performing at an acceptable level.

SharePoint 2010 minimum processor requirements recommended by Microsoft may be excessive in some environments. For this reason, VMware recommends reducing the number of virtual CPUs if monitoring of the actual workload shows that the virtual machine is not benefitting from the increased virtual CPUs. Having virtual CPUs allocated but sitting idle reduces the consolidation level and efficiency of the ESX/ESXi host. For more background, see the "ESX CPU Considerations" section in the white paper *Performance Best Practices for VMware vSphere 4* at http://www.vmware.com/pdf/Perf_Best_Practices_vSphere4.0.pdf.

2.1.3 Overcommitment

VMware 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. Because the performance degradation is "graceful," any virtual CPU overcommitment can be effectively managed by using VMware DRS and VMware vSphere® VMotion® to move virtual machines to other ESX/ESXi hosts to obtain more processing power. By intelligently implementing CPU overcommitment, consolidation ratios of SharePoint Web front-end and application servers can be driven higher while maintaining acceptable performance. If it is chosen that a virtual machine not participate in overcommitment, setting a CPU reservation provides a guaranteed CPU allocation for the virtual machine. This practice is generally not recommended because the reserved resources are not available to other virtual machines and flexibility is often required to manage changing workloads. However, SLAs and multi-tenancy may require a guaranteed amount of compute resources to be available. In these cases, reservations make sure that these requirements are met.

When choosing to overcommit CPU resources, monitor vSphere and SharePoint to be sure responsiveness is maintained at an acceptable level. The following table lists counters that can be monitored to help drive consolidation numbers higher while maintaining performance.

Table 1. Esxtop CPU Performance Metrics

Esxtop Metric	Description	Implication
%RDY	Percentage of time a vCPU in a run queue is waiting for the CPU scheduler to let it run on a physical CPU.	A high %RDY time (use 20% as a starting point) may indicate the virtual machine is under resource contention. Monitor this—if application speed is OK, a higher threshold may be tolerated.
%MLMTD	Percentage of time a vCPU was ready to run but was deliberately not scheduled due to CPU limits.	A high %MLMTD time may indicate a CPU limit is holding the VM in a ready to run state. If the application is running slow consider increasing or removing the CPU limit.
%CSTP	Percentage of time a vCPU spent in read, co-descheduled state. Only meaningful for SMP virtual machines.	A high %CSTP time usually means that vCPUs are not being used in a balanced fashion. Evaluate the necessity for multiple vCPUs.

2.1.4 Hyper-threading

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

2.1.5 Non-Uniform Memory Access (NUMA)

Non-Uniform Memory Access (NUMA) compatible systems contain multiple nodes that consist of a set of processors and memory. The access to memory in the same node is local, while access to the other node is remote. Remote access can take longer because it involves a multihop operation. In NUMA-aware applications, there is an attempt to keep threads local to improve performance.

ESX/ESXi provides load-balancing on NUMA systems. To achieve the best performance, it is recommended that NUMA be enabled on compatible systems. On a NUMA-enabled ESX/ESXi host, virtual machines are assigned a home node from which the virtual machine's memory is allocated. Because it is rare for a virtual machine to migrate away from the home node, memory access is mostly kept local.

In applications that scale out well, such as SharePoint, it is beneficial to size the virtual machines with the NUMA node size in mind. For example, in a system with two hexa-core processors and 64GB of memory, sizing the virtual machine to six virtual CPUs and 32GB or less means that the virtual machine does not have to span multiple nodes.

2.2 Memory Configuration Guidelines

This section provides guidelines for allocating memory to SharePoint virtual machines. The guidelines outlined here take into account vSphere memory overhead and the virtual machine memory settings.

2.2.1 ESX/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:

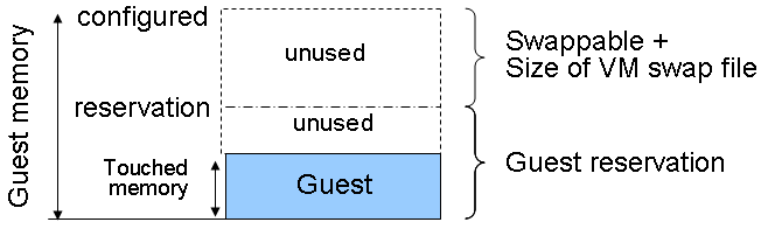
- Memory sharing across virtual machines that have similar data (that is, same guest operating systems).
- Memory overcommitment, which means allocating more memory to virtual machines than is physically available on the ESX/ESXi host.
- A memory balloon technique whereby virtual machines that do not need all the memory they were allocated give memory to virtual machines that require additional allocated memory.

For more information about vSphere memory management concepts, see the *VMware vSphere Resource Management Guide* at http://www.vmware.com/pdf/vsphere4/r40_u1/vsp_40_u1_resource_mgmt.pdf.

2.2.2 Virtual Machine Memory Concepts

Figure 1 illustrates the use of memory settings parameters in the virtual machine.

Figure 1. 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 allocates only 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 (that is, 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 Virtual Machine File System (VMFS) file system (VSWP file).
- 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 SharePoint 2010 Virtual Machines

The proper sizing of memory for a SharePoint 2010 virtual machine is based on many factors. With the number of application services and use cases available determining a suitable configuration for an environment requires creating a baseline configuration, testing, and making adjustments, as discussed later in this paper. Regardless of how much memory virtual machines require there are best practices to consider when planning for the underlying virtual infrastructure to support SharePoint. The following are descriptions of recommended best practices:

- Account for memory overhead – Virtual machines require memory beyond the amount allocated, and this memory overhead is per-virtual machine. Memory overhead includes space reserved for virtual machine devices, such as SVGA frame buffers and internal data structures. The amount of overhead required depends on the number of vCPUs, configured memory, and whether the guest operating system is 32-bit or 64-bit. As an example, a running virtual machine with one virtual CPU and two gigabytes of memory may consume about 100 megabytes of memory overhead, where a virtual machine with two virtual CPUs and 32 gigabytes of memory may consume approximately 500 megabytes of memory overhead. This memory overhead is in addition to the memory allocated to the virtual machine and must be available on the ESX host.
- ”Right-size” memory allocations – Over-allocating memory to virtual machines can waste memory unnecessarily, but it can also increase the amount of memory overhead required to run the virtual machine, thus reducing the overall memory available for other virtual machines. Fine-tuning the memory for a virtual machine is done easily and quickly by adjusting the virtual machine properties. In most cases, hot-adding of memory is supported and can provide instant access to the additional memory if needed.
- Intelligently overcommit – Memory management features in vSphere allow for overcommitment of physical resources without severely impacting performance. Many workloads can participate in this type of resource sharing while continuing to provide the responsiveness users require of the application. When looking to scale beyond the underlying physical resources, consider the following:
    - Establish a baseline before overcommitting. Note the performance characteristics of the application before and after. Some applications are consistent in how they utilize resources and may not perform as expected when vSphere memory management techniques take control. Others, such as Web servers, have periods where resources can be reclaimed and are perfect candidates for higher levels of consolidation.
    - Use the default balloon driver settings. The balloon driver is installed as part of the VMware Tools suite and is used by ESX/ESXi if physical memory comes under contention. Performance tests show that the balloon driver allows ESX/ESXi to reclaim memory, if required, with little to no impact to performance. Disabling the balloon driver forces ESX/ESXi to use host-swapping to make up for the lack of available physical memory which adversely affects performance.
    - Set a memory reservation for virtual machines that require dedicated resources. Virtual machines running Search or SQL services consume more memory resources than other application and Web front-end virtual machines. In these cases, memory reservations can guarantee that those services have the resources they require while still allowing high consolidation of other virtual machines.
    - As with overcommitting CPU resources, proactive monitoring is a requirement. The following table lists counters that can be monitored to avoid performance issues resulting from memory overcommitment.

Table 2. Esxtop Memory Counters

Esxtop Metric	Description	Implication
SWAP /MB: r/s, w/s	The rate at which machine memory is swapped in and out of disk.	High rates of swapping affect guest performance. If free memory is low, consider moving virtual machines to other hosts. If free memory is OK, check resource limits on the virtual machines.
MCTLSZ	The amount of guest physical memory reclaimed by the balloon driver.	If the guest working set is smaller than guest physical memory after ballooning, no performance degradation is observed. However, investigate the cause for ballooning. It could be due to low host memory or a memory limit on the virtual machine.

2.3 Storage Guidelines

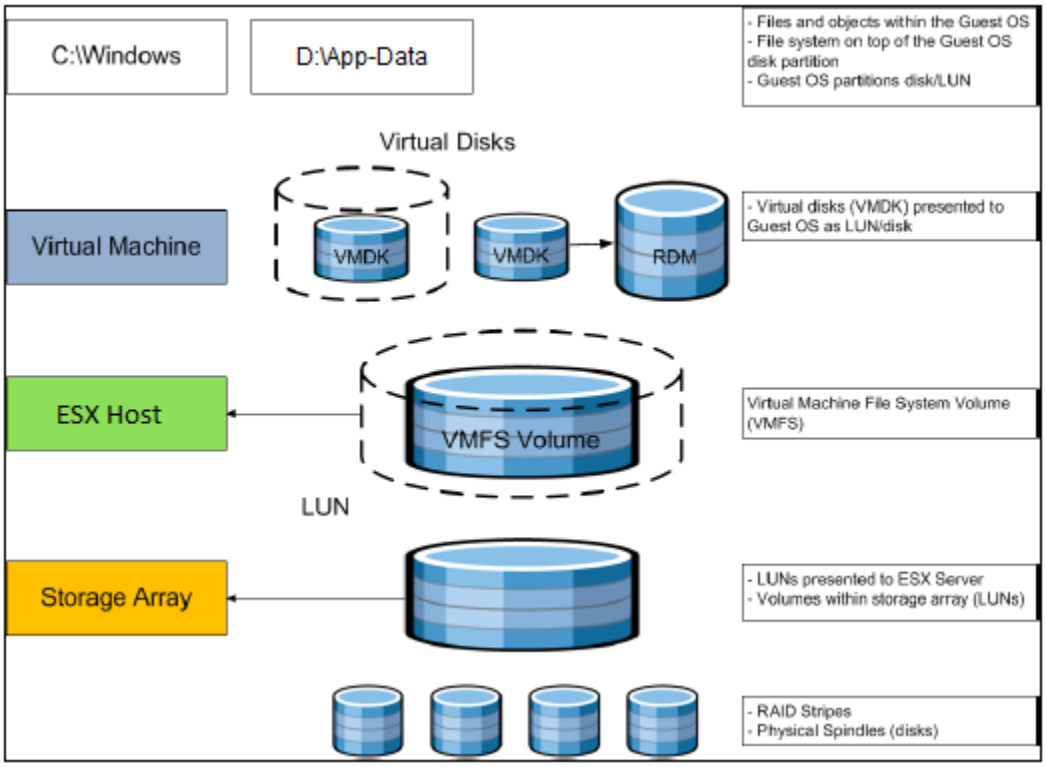
vSphere provides many features that take advantage of commonly used storage technologies such as storage area networks and storage replication. Features such as VMware vMotion, VMware HA, and VMware Distributed Resource Scheduler (DRS) use these storage technologies to provide high availability, resource balancing, and uninterrupted workload migration.

2.3.1 Storage Virtualization Concepts

Figure 2 shows that VMware storage virtualization can be categorized into three layers of storage technology:

- The Storage array is the bottom layer, consisting of physical disks presented as logical disks (storage array volumes or LUNs) to the layer above, with the virtual environment occupied by vSphere.
- Storage array LUNs that are formatted as VMFS datastores that provide storage for virtual disks.
- Virtual disks that are presented to the virtual machine and guest operating system as SCSI attached disks that can be partitioned and used in file systems.

Figure 2. VMware Storage Virtualization Stack



2.3.1.1. VMFS File System

The VMFS file system was created by VMware to allow multiple vSphere hosts to read and write to the same storage concurrently. VMFS is a clustered file system that allows you to simplify virtual machine provisioning and administration by consolidating virtual machines into smaller units of storage. Unique virtualization-based capabilities provided by the VMFS file system include live migration using vMotion, and increased availability using VMware HA.

Virtual machines are stored on VMFS datastores as a unique set of encapsulated files, including configuration files and virtual disks (VMDK files). VMFS is supported on both iSCSI and Fibre Channel attached storage.

2.3.1.2. Raw Device Mapping

For instances where isolation or direct access to the underlying storage subsystem is required a raw device mapping can be used in place of virtual disks. Raw device mappings use a mapping file that is located on a VMFS datastore to point to a physical LUN. The physical LUN is accessible to the virtual machine in its raw form and must be formatted from within the virtual machine. Unlike VMFS, a raw device mapping is typically only assigned to a single virtual machine; however, RDMs can be shared, for example, in a Microsoft Cluster configuration where multiple nodes use SCSI reservations to handle arbitration. RDMs cannot provide all of the features available with VMFS and should be limited to use only when technically required.

2.3.2 Storage Protocol Capabilities

VMware vSphere provides vSphere and storage administrators with the flexibility to use the storage protocol that meets the requirements of the business. This can be a single protocol datacenter wide, such as iSCSI, or multiple protocols for tiered scenarios such as using Fibre Channel for high-throughput storage pools and NFS for high-capacity storage pools.

For SharePoint 2010 on vSphere there is no single option that is considered superior to another. It is recommended that this decision be made based on your established storage management practices within the virtualized environment.

For more information, see the VMware whitepaper *Comparison of Storage Protocol Performance in VMware vSphere 4* at http://www.vmware.com/files/pdf/perf_vsphere_storage_protocols.pdf.

2.3.3 Storage Best Practices

The following are vSphere storage best practices.

- Host multipathing – Having a redundant set of paths to the storage area network is critical to protecting the availability of your environment. This redundancy can be in the form of dual host-bus adapters connected to separate fabric switches, or a set of teamed network interface cards for iSCSI and NFS.
- Partition alignment – Partition misalignment can lead to severe performance degradation due to I/O operations having to cross track boundaries. Partition alignment is important both at the VMFS level as well as within the guest operating system. Use the vSphere Client when creating VMFS datastores to be sure they are created aligned. When formatting volumes within the guest, Windows 2008 aligns NTFS partitions on a 1024KB offset by default.
- Use shared storage – In a vSphere environment, many of the features that provide the flexibility in management and operational agility come from the use of shared storage. Features such as VMware HA, DRS, and vMotion take advantage of the ability to migrate workloads from one host to another host while reducing or eliminating the downtime required to do so.
- Calculate your total virtual machine size requirements – Each virtual machine requires more space than that used by its virtual disks. Consider a virtual machine with a 20GB OS virtual disk and 16GB of memory allocated. This virtual machine will require 20GB for the virtual disk, 16GB for the virtual machine swap file (size of allocated memory), and 100MB for log files (total virtual disk size + configured memory + 100MB) or 36.1GB total.
- Understand I/O Requirements – Under-provisioned storage can significantly slow responsiveness and performance for SharePoint. As a multitiered application, you can expect each tier of SharePoint to have different I/O requirements. These requirements are discussed in further detail as they pertain to SharePoint in the performance and capacity planning sections of this document. However, as a general recommendation, pay close attention to the amount of virtual machine disk files hosted on a single VMFS volume. Over-subscription of the I/O resources can go unnoticed at first and slowly begin to degrade performance if not monitored proactively.

2.4 Networking Guidelines

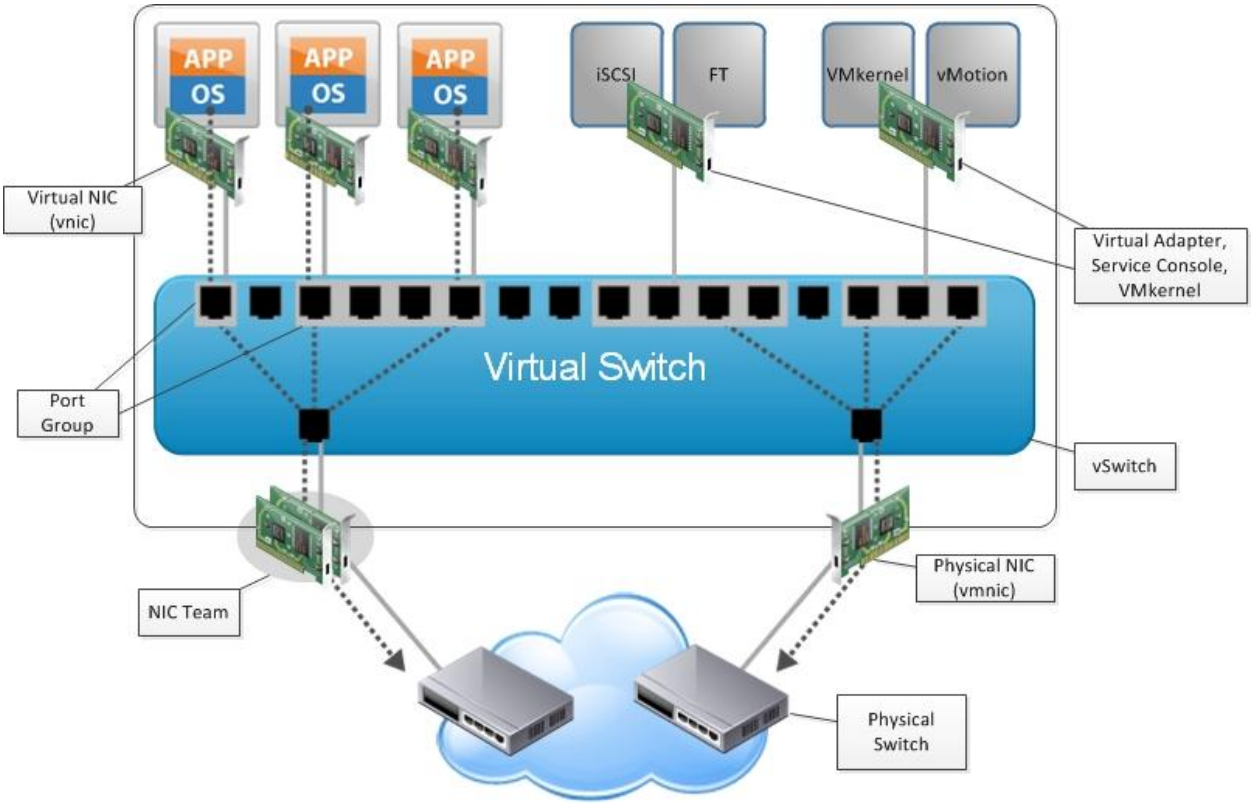
Networking in the virtual world follows the same concepts as in the physical world, but these concepts are applied in software instead of using physical cables and switches. Many of the best practices that apply in the physical world continue to apply in the virtual world, but there are additional considerations for traffic segmentation, availability, and making sure that the throughput required by services hosted on a single server can be fairly distributed.

2.4.1 Virtual Networking Concepts

Figure 3. Virtual Networking in vSphere

provides a visual overview of the components that make up the virtual network.

Figure 3. Virtual Networking in vSphere



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

- Physical switch – vSphere host-facing edge of the local area network.
- Physical network interface (vmnic) – Provides connectivity between the ESX host and the local area network.
- vSwitch – 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 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, vMotion, and FT logging.
- Service Console (vswif) – Interface for the service console present in ESX Classic. Not present in ESXi.
- Virtual Adapter – Provides Management, vMotion, and FT Logging when connected to a vNetwork Distributed Switch.
- NIC Team – Group of physical NICs connected to the same physical/logical networks providing redundancy.

2.4.2 Virtual Networking Best Practices

The following are vSphere networking best practices:

- Separate virtual machine and infrastructure traffic – Keep virtual machine and VMkernel or service console traffic separate. This can be accomplished physically using separate virtual switches that uplink to separate physical NICs, or virtually using VLAN segmentation.
- Use NIC Teaming – Use two physical NICs per vSwitch, and if possible, uplink the physical NICs to separate physical switches. Teaming provides redundancy against NIC failure and, if connected to separate physical switches, against switch failures. NIC teaming does not necessarily provide higher throughput.
- Enable PortFast on ESX/ESXi host uplinks – Failover events can cause spanning tree protocol recalculations that can set switch ports into a forwarding or blocked state to prevent a network loop. This process can cause temporary network disconnects. To prevent this situation, set the switch ports connected to ESX/ESXi hosts to PortFast, which immediately sets the port back to the forwarding state and prevents link state changes on ESX/ESXi hosts from affecting the STP topology. Loops are not possible in virtual switches.
- Converge Network and Storage I/O with 10Gbps Ethernet – When possible consolidating storage and network traffic can provide simplified cabling and management over having to maintain separate switching infrastructures.

3. SharePoint Performance on vSphere

3.1 Overview

Among the drivers for customers to deploy new or upgrade to SharePoint 2010 are the added functionality and performance optimizations. For many customers with existing SharePoint 2007 environments this is seen as a logical next step, and one that can allow them to further scale their environments to grow with the business. However, a subject that often comes up is performance characteristics and how to design a vSphere environment to support the upgrade to SharePoint 2010.

Most of the performance optimizations in SharePoint 2010 are specific to the application, regardless of whether the deployment is on a physical or virtual infrastructure. Improvements to the SQL side include optimizations to the content database to allow for better scaling and reduced SQL blocking through the use of list throttling, both of which affect overall SharePoint performance. At the Web front-end, SharePoint 2010 provides cleaner and leaner pages, XHTML compliance, and a 12% `core.js` file size reduction, all of which are meant to produce a more responsive end user experience. SharePoint Search was improved and made more scalable by the addition of distributed indexing and index mirroring, as well as the ability to have multiple crawlers. With the increased scalability of SharePoint 2010 and the distribution of server roles, leveraging vSphere as the underlying infrastructure can provide the desired levels of consolidation while maintaining the performance required by end users.

3.2 vSphere Performance Enhancements

With every release of vSphere the overhead of running an application on the vSphere virtualized platform is reduced by way of new performance improving features. Typical virtualization overhead for applications, such as SharePoint 2010, is less than 10%. Many of these features not only improve performance of the virtualized application itself, but also allow for higher consolidation ratios. Understanding these features and taking advantage of them in your SharePoint 2010 environment helps guarantee the highest level of success in your virtualized deployment.

Table 3. vSphere Performance Enhancements

Esxtop Metric	Description	Implication
NUMA support	ESX/ESXi uses a NUMA load-balancer to assign a home node to a virtual machine. Because memory for the virtual machine is allocated from the home node, memory access is local and provides the best performance possible. Even applications that do not directly support NUMA benefit from this feature.	See *VMware vSphere 4: The CPU Scheduler in VMware ESX 4* at http://www.vmware.com/files/pdf/perf-vsphere-cpu_scheduler.pdf
Transparent page sharing	Virtual machines running similar operating systems and applications typically have identical sets of memory content. Page sharing allows the hypervisor to reclaim the redundant copies and keep only one copy, which frees up the total host memory consumption. Because most SharePoint virtual machines run the same operating system and application binaries, total memory usage can be reduced to increase consolidation ratios.	See *Understanding Memory Resource Management in VMware ESX 4.1* at http://www.vmware.com/files/pdf/techpaper/vsp_41_perf_memory_mgmt.pdf

Esxtop Metric	Description	Implication
Memory ballooning	By using a balloon driver loaded in the guest operating system, the hypervisor can reclaim host physical memory if memory resources are under contention. This is done with little to no impact to the performance of the application.	See *Understanding Memory Resource Management in VMware ESX 4.1* at http://www.vmware.com/files/pdf/techpaper/vsp_41_perf_memory_mgmt.pdf
Memory compression	Before a virtual machine resorts to host swapping, due to memory overcommitment the pages elected to be swapped attempt to be compressed. If the pages can be compressed and stored in a compression cache, located in main memory, the next access to the page causes a page decompression as opposed to a disk swap out operation, which can be an order of magnitude faster.	See *Understanding Memory Resource Management in VMware ESX 4.1* at http://www.vmware.com/files/pdf/techpaper/vsp_41_perf_memory_mgmt.pdf
Large memory page support	An application that can benefit from large pages on native systems, such as MS SQL, can potentially achieve a similar performance improvement on a virtual machine backed with large memory pages. Enabling large pages increases the memory page size from 4KB to 2MB.	See *Performance Best Practices for VMware vSphere 4.0* at http://www.vmware.com/pdf/Perf_Best_Practices_vSphere4.0.pdf and *Performance and Scalability of Microsoft SQL Server on VMware vSphere 4* at http://www.vmware.com/files/pdf/perf_vsphere_sql_scalability.pdf
Para-virtualized network and storage controllers	The PVSCSI virtual storage and VMXNet3 virtual network adapters, which are available to the guest operating system after installation of VMware Tools, are high-performance virtual I/O adapters that can provide greater throughput while requiring lower CPU utilization.	See *PVSCSI Storage Performance* at http://www.vmware.com/pdf/vsp_4_pvscsi_perf.pdf and the VMware Blog at: http://blogs.vmware.com/performance/2009/09/vsphere-40-introduces-a-new-para-virtualized-network-device---vmxnet3-we-recently-published-a-paper-demonstrating-its-perfo.html
Distributed Resource Scheduler (DRS) and vMotion	DRS dynamically balances computing capacity across a vSphere cluster, creating an aggregate of compute power. As resource utilization fluctuates within a vSphere cluster workloads are migrated with no impact to performance or uptime.	See the VMware vSphere product overview at http://www.vmware.com/products/drs/overview.html

3.3 Performance Testing

Chances are high that almost every deployment of SharePoint has different performance characteristics. Variables, such as the amount of content, user concurrency, document size, query freshness requirements, among others, vary from environment to environment. Because of this, modeling SharePoint is a difficult task, and the general consensus is that to truly understand the requirements and characteristics there must be a lot of testing and piloting in the target environment, followed by adjustments and continued monitoring.

Given the level of ambiguity in SharePoint workloads, it is become a common practice to standardize on a common workload, or set of tasks, when attempting to examine how SharePoint performs on a given platform. VMware and its partners have used the Microsoft Visual Studio Team System (VSTS) to simulate requests from SharePoint users. Using VSTS a workload distribution is created to mix the reads, searches, and modification transactions to emulate that of the target environment. VSTS reports the number of transactions per second along with the average response time. To create a continuous workload, the think-time (time usually spent by a user in between requests) is set to zero. By using the reported requests per second, the maximum number of users supported by the configuration can be computed. In the following sections, both VMware and partner testing is outlined and provided for reference. You can use these documents as part of the planning process, as prescribed by Microsoft and detailed in the Capacity Management Concepts and Reference section of this paper.

3.3.1 Internal Performance Testing

The performance team at VMware conducted a set of experiments using vSphere 4.1 and SharePoint Server 2007 to demonstrate the performance and scalability that can be achieved when virtualizing SharePoint. The experiments, which supported up to 171,600 heavy SharePoint users on a single physical server, highlight the benefits gained by being able to easily deploy additional SharePoint virtual machines to satisfy changing demands.

In the tests conducted by VMware Performance Engineering, the goal was to understand the performance of SharePoint in a virtualized and consolidated environment. The experiments consisted of three use cases with virtual machine configurations as shown in the following table.

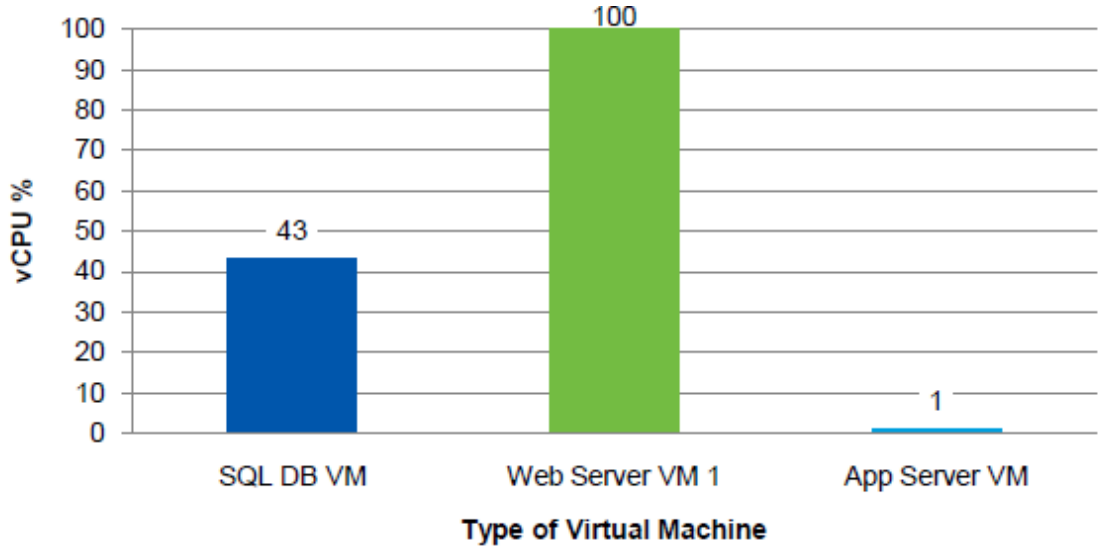
Table 4. Virtual Machine Configurations for Each Use Case

Number of Virtual Machines in Test	Configured as Number of Servers
Three virtual machines	1 Web server, 1 application server, 1 SQL Server
Four virtual machines	2 Web server, 1 application server, 1 SQL Server
Five virtual machines	3 Web server, 1 application server, 1 SQL Server

In each use case, the Web server front-ends were configured with the Web Server role as well as the query server role using Windows Network Load Balancing to distribute the load. The application server virtual machine was used as a dedicated indexing server, and the SQL Server virtual machine held the 200GB of SharePoint user data. Each virtual machine was configured with two virtual CPUs and 4GB of memory, the SQL Server virtual machine was allocated 16GB of memory.

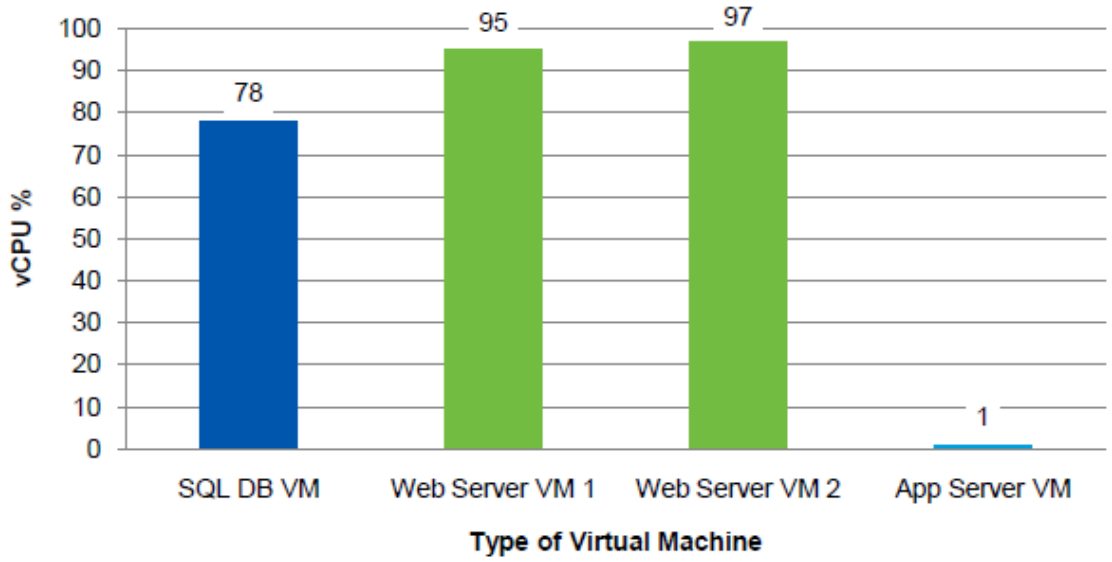
During the first round of testing the single Web server is the bottleneck with 100% vCPU utilization.

Figure 4. Percentage of Guest vCPUs Utilized in the Three Virtual Machine Configuration



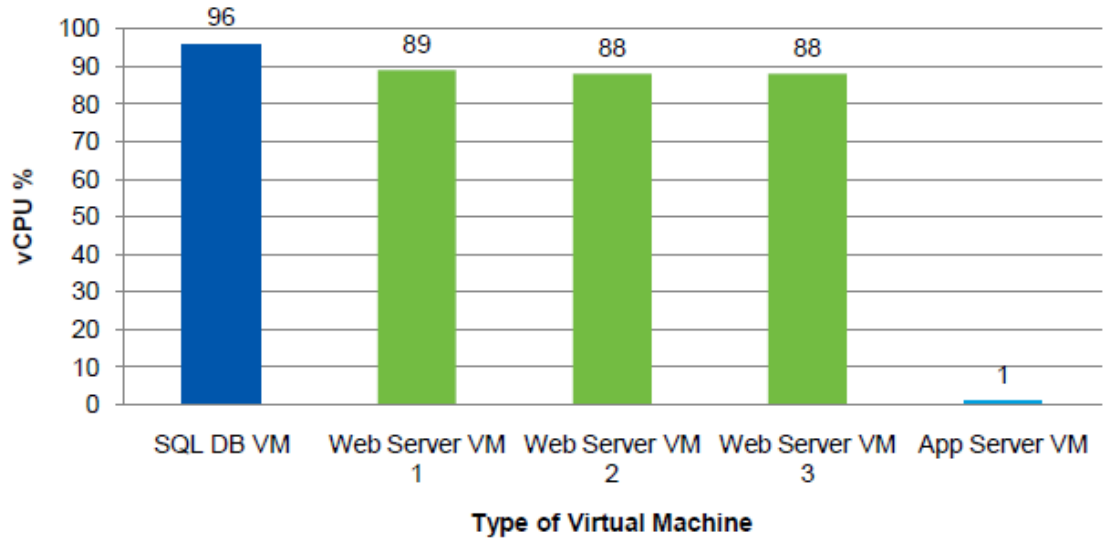
The second test consisted of two Web server virtual machines, which were also the bottlenecks at 95% and 97% vCPU utilization. The SQL Server utilization increased to 78%.

Figure 5. Percentage of Guest vCPUs Utilized in the Four Virtual Machine Configuration



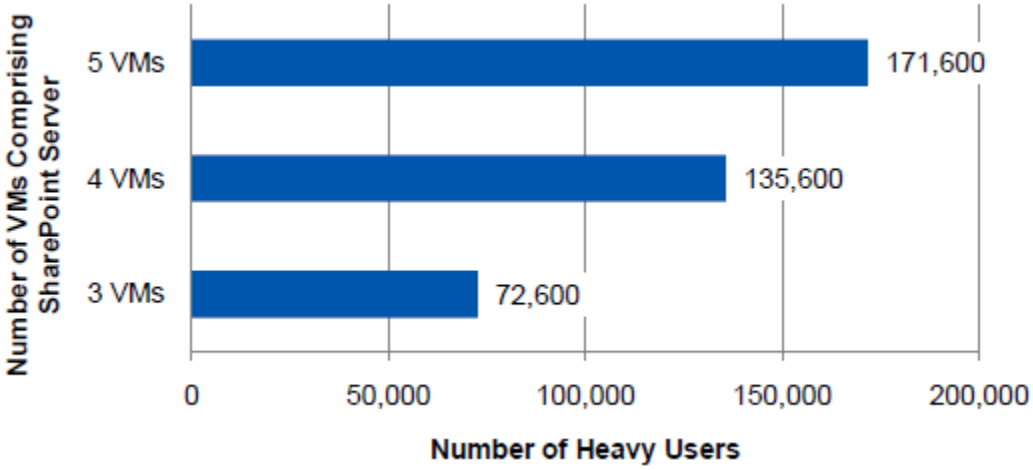
In the third test, a third Web server virtual machine was added and the Web servers were no longer the bottleneck, however, the SQL Server was close to saturation at 96% vCPU utilization.

Figure 6. Percentage of Guest vCPUs Utilized in the Five Virtual Machine Configuration



Between the first and last experiment, the number of supported heavy SharePoint users increased from 72,600 users to 171,600 users, a 136% increase, by simply scaling out the Web server virtual machines on the same physical server.

Figure 7. Number of Heavy Users at 1% Concurrency



The testing concluded that vSphere could provide very good results (response times less than three seconds) while providing consolidation of a SharePoint environment. Additionally, vSphere features, such as cloning and templates, enable the quick deployment of additional virtual machines to meet increased demand and increase server utilization. See *Microsoft Office SharePoint Server 2007 Performance on VMware vSphere 4.1* at: http://www.vmware.com/resources/techresources/10130.

3.3.2 Partner Testing

VMware OEM partners conduct testing to show the value of virtualizing SharePoint along with some of their proprietary features and products. The following table summarizes testing performed by VMware partners for both SharePoint 2007 and 2010 on vSphere.

Table 5. Partner Testing Summary

Partner	Type	Summary	Link
Cisco	Design Guide	*Microsoft SharePoint 2010 on FlexPod for VMware* Description: This design guide demonstrates how enterprises can apply best practices for VMware vSphere, VMware vCenter™, Cisco Unified Computing System (UCS), Cisco Nexus family switches, and NetApp FAS. Design validation was completed using Microsoft Visual Studio 2010 Ultimate workloads and Visual Studio Load Test agents simulating a realistic 50,000 user load across SharePoint 2010 Enterprise features. Virtual user connections from branch offices, remote offices, and mobile users were optimized using Cisco Wide Area Application Services (WAAS). Cisco Application Control Engine (ACE) was implemented for Web and application server load balancing within the SharePoint 2010 architecture.	http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/App_Networking/SharePoint_FlexPod.html
EMC	White Paper	*EMC Continuous Protection for Virtualized SharePoint 2010 Farms: A Detailed Review* Description: A 2011 guide to protecting a virtualized SharePoint 2010 environment with storage replication and VMware vCenter Site Recovery Manager™.	http://www.emc.com/collateral/software/white-papers/h8139-protection-virtualized-sharepoint-wp.pdf
EMC	Reference Architecture	*EMC Virtual Infrastructure for Microsoft SharePoint Server 2010 Enabled by EMC Clariion and VMware vSphere 4.* Description: A solution for virtualizing SharePoint 2010 with EMC Clariion Storage.	http://www.emc.com/collateral/software/white-papers/h8024-virtual-sharepoint-clariion-vsphere-wp.pdf

Partner	Type	Summary	Link
EMC	White Paper	*EMC Unified Storage for Microsoft SharePoint Server 2010: BLOB Externalization Enabled by EMC Celerra and Metalogix StoragePoint.* Description: This solution demonstrates the functional and performance aspects of the external BLOB storage solution for Microsoft SharePoint Server 2010 built on a virtual infrastructure that uses VMware vSphere 4.	http://www.emc.com/collateral/hardware/technical-documentation/h8112-unified-sharepoint-blob-psg.pdf
HP	White Paper	*Best practices for deploying Microsoft Office SharePoint Server 2007 with VMware on HP ProLiant servers.* Description: In this white paper, HP outlines the best practices for using VMware ESXi 4.0 Server for deploying Microsoft Office SharePoint Server 2007.	http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA2-9968ENW.pdf

3.4 Ongoing Performance Monitoring and Tuning

SharePoint Server performance monitoring leverages the Microsoft Windows performance monitor tool PerfMon to collect statistics. PerfMon provides familiar counters that indicate system performance. As with all in-guest measurement tools, time-based performance measurements are subject to error based on the load being placed on the underlying ESX host. Generally it is safe to assume the results are no more than 10% in error if CPU utilization stays below 80%.

VMware and SharePoint administrators should pay close attention to the counters listed in Table 5. Refer to online documentation on *Performance Monitoring and Analysis* (http://communities.vmware.com/docs/DOC-3930) for more information on these counters and their interpretation.

Table 6. Performance Counters of Interest to VMware and SharePoint Administrators

Subsystem	Esxtop Counters	vCenter Counter
CPU	%RDY	Ready (milliseconds in a 20,000 millisecond 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

Windows Performance Monitor counters can be correlated with esxtop counters for compute resources such as CPU, memory, and disk latencies. When used in conjunction with esxtop counters, the PerfMon counters can be evaluated for their accuracy and used to pinpoint any bottlenecks in the system. See Table 7 (more information is available at http://technet.microsoft.com/en-us/library/ff758658.aspx).

Table 7. SharePoint 2010 Performance Counters from TechNet (partial)

Objects and Counters	Descriptions
Processor	
% Processor Time	This shows processor usage over a period of time. If this is consistently too high, you may find performance is adversely affected. Remember to count "Total" in multiprocessor systems. You can measure the utilization on each processor as well, to achieve balanced performance between cores.
Disk	
Avg. Disk Queue Length	This shows the average number of read and write requests that were queued for the selected disk during the sample interval. A bigger disk queue length may not be a problem as long as disk reads/writes are not suffering and the system is working in a steady state without expanding queuing.
Avg. Disk Read Queue Length	The average number of read requests that are queued.
Avg. Disk Write Queue Length	The average number of write requests that are queued.
Disk Reads/sec	The number of reads to disk per second.
Disk Writes/sec	The number of writes to disk per second.
Memory	
Available Mbytes	This shows the amount of physical memory available for allocation. Insufficient memory leads to excessive use of the page file and an increase in the number of page faults per second.
Cache Faults/sec	This counter shows the rate at which faults occur when a page is sought in the file system cache and is not found. This may be a soft fault, when the page is found in memory, or a hard fault, when the page is on disk. The effective use of the cache for read and write operations can have a significant impact on server performance. You must monitor for increased cache failures, indicated by a reduction in the Async Fast Reads/sec or Read Aheads/sec.
Pages/sec	This counter shows the rate at which pages are read from or written to disk to resolve hard page faults. If this rises, it indicates system-wide performance problems.

Objects and Counters	Descriptions
Paging File	
% Used and % Used Peak	The server paging file, sometimes called the swap file, holds "virtual" memory addresses on disk. Page faults occur when a process has to stop and wait while required "virtual" resources are retrieved from disk into memory. These are more frequent if the physical memory is inadequate.
NIC	
Total Bytes/sec	This is the rate at which data is sent and received by way of the network interface card. You may need to investigate further if this rate is over 40–50 percent network capacity. To fine-tune your investigation, monitor **Bytes Received/sec** and **Bytes Sent/sec**.

4. Capacity Management Concepts and Reference

4.1 Capacity Management Overview

Unlike some applications that have a fairly consistent, well-defined workload (for example. Exchange or SAP), SharePoint workloads can vary greatly depending on how the application is deployed and used within the organization. SharePoint services can be deployed in a wide variety of combinations to accommodate very specific application use cases. Even within a specific application use case, usage patterns can vary greatly depending on frequency of user access, time of day, document reads/writes, and document sizes.

Another important factor to consider is the changeable nature of a SharePoint environment. In many cases, SharePoint deployments start small, providing limited functionality to a small subset of users. Over time, user adoption becomes more widespread across the organization and more services are deployed to support specialized job functions. Eventually the increase in usage results in a performance impact to the SharePoint environment, forcing the administrator to adjust the architecture to suit the new requirements.

Capacity Management, in the context of SharePoint, refers to a continual cycle of designing, deploying, monitoring, and adjusting the SharePoint environment to match the most current usage requirements. In some cases adjustment may require "scaling up" from standard guidelines by adding more resources (such as processor and memory) to bottlenecked servers. In other cases, adjustment may involve "scaling out" by deploying additional server instances to better balance the workload. VMware features, such as Hot-Add and vMotion, can help you make these types of changes to your SharePoint environment very quickly, allowing a near real-time response to changing business requirements.

4.2 Workload Modeling and Architectural Design

4.2.1 Estimating User Activity

The first step in modeling a production workload is to estimate the amount of activity you expect the users to generate. If you're deploying SharePoint for the first time in your organization, you must do a bit of educated guesswork on the expected user activity.

If you're upgrading from SharePoint 2007, you can analyze the existing environment to estimate the expected load in SharePoint 2010. To do this, you can either mine IIS logs or use information gathered by a third-party monitoring solution. Having current usage information greatly simplifies the capacity planning process and Microsoft provides the capability to use exiting IIS logs for load testing. To use an existing environment as a baseline requires the following from Microsoft:

- SharePoint 2010 Load Testing Kit available at http://technet.microsoft.com/en-us/library/ff823736.aspx.
- Visual Studio 2008 Team System available at http://www.microsoft.com/downloads/en/details.aspx?FamilyID=d95598d7-aa6e-4f24-82e3-81570c5384cb&DisplayLang=en.
- Visual Studio 2008 Service Pack 1 available at http://www.microsoft.com/download/en/details.aspx?displaylang=en&id=10986.

Whether you are deploying a new environment or upgrading from SharePoint 2007, you must define the expected user workload by using the criteria in the table below. Again, if you have an existing environment, this information can be gathered from the IIS logs; if you're starting from scratch, you'll need to estimate these values. Microsoft has published several case studies (available at http://technet.microsoft.com/en-us/library/cc261716.aspx) profiling different types of user load to help with your estimation process. Table 8 is from Microsoft's *Enterprise intranet collaboration environment*

technical case study example (available at http://technet.microsoft.com/en-us/library/ff758650.aspx), which services over 69,000 unique users per day.

Table 8. Enterprise Intranet Collaboration Environment Technical Case Study Example

Workload Characteristics	Value
Average daily RPS	157
Average RPS at peak time	350
Total number of unique users per day	69,702
Average daily concurrent users	420
Peak concurrent users at peak time	1,433
Total number of requests per day	18,866,527

- Requests per second (RPS) – The demand on the server farm expressed in the number of requests processed by the farm per second, but with no differentiation between the type and size of requests. RPS is highly dependent on an organization's unique usage characteristics.
- Concurrent users – The number of distinct users generating requests in a given time frame.
- Total daily users – The actual number of unique users in a 24-hour period, not the total number of employees in the organization.
- Total daily requests – All requests (except authentication handshake requests) over a 24-hour period.

At this time there are no definitive guidelines on what might be considered “normal” usage for SharePoint 2010 users; however, there is some information in the SharePoint 2007 TechNet documentation (http://technet.microsoft.com/en-us/library/cc261795%28office.12%29.aspx) that may give you an idea on what to expect in terms of user load. You can use the **Requests Per Second Per User** value to calculate the overall RPS that you must support. For example, 5,000 concurrent, “heavy” users each at 60 requests per hour would translate into .017 requests per second per user or 85 requests per second (RPS) for the system as a whole.

Table 9. SharePoint 2007 User Loads from Microsoft TechNet

User Load	Request Rate	Requests Per Second Per User
Light	20 requests per hour. An active user generates a request every 180 seconds.	.006
Typical	36 requests per hour. An active user generates a request every 100 seconds.	.010
Heavy	60 requests per hour. An active user generates a request every 60 seconds.	.017
Extreme	120 requests per hour. An active user generates a request every 30 seconds.	.034

4.2.2 Understanding Workload Distribution

Sometimes it can be useful to understand the distribution of the requests based on the client applications that are interacting with the server farm. Newer clients, such as Office 2010, offer new capabilities that can increase the overall load on the system.

In the *Capacity Planning for SharePoint Server 2010* TechNet page (at http://technet.microsoft.com/en-us/library/ff758645.aspx), there is a chart that shows a snapshot of a live internal Microsoft environment serving a typical social solution. In the chart, you can see that the majority of the load is generated by the search crawler and typical end user web browsing. You can also observe that there is significant load introduced by the new Outlook Social Connector feature.

4.2.3 Understanding Concurrency and Peak Usage

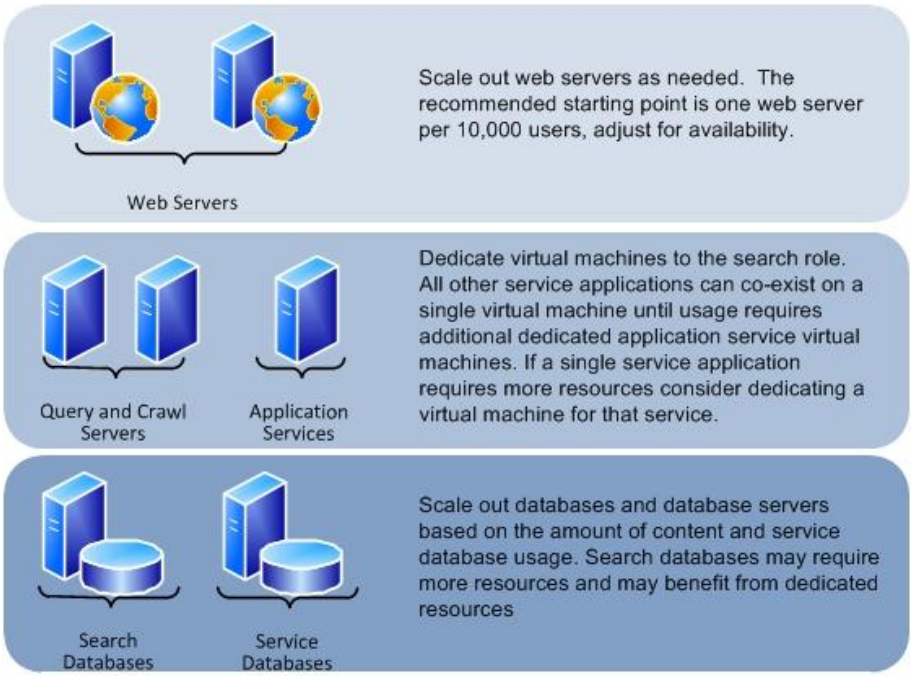
Unlike Exchange, where usage is fairly ubiquitous and users remain connected continually, SharePoint usage is much more sporadic from an end-user perspective. When thinking about long-term usage patterns, there may be users or groups of users in your organization that have no need to access SharePoint; therefore the system should be designed with a certain number of *active users* in mind. Out of these active users, you should also consider how many users will be accessing the system simultaneously at any given time; referred to as *concurrent users*. Finally, SharePoint concurrent usage might vary significantly over the course of any given 24-hour period. The *peak usage period* is the point in time where the maximum number of concurrent users are accessing the SharePoint environment.

4.2.4 Selecting a Starting Point Architecture

4.2.4.1. SharePoint 2010 Topologies

Microsoft has published a set of topologies ranging from small deployments to large enterprise farm implementations (see http://technet.microsoft.com/en-us/library/cc263044.aspx) that you can use as a starting point for your architecture. Once you've chosen a topology, you can use the recommended server role requirements to plan resource allocation for your SharePoint virtual machines.

Figure 8. SharePoint 2010 Medium Topology Example



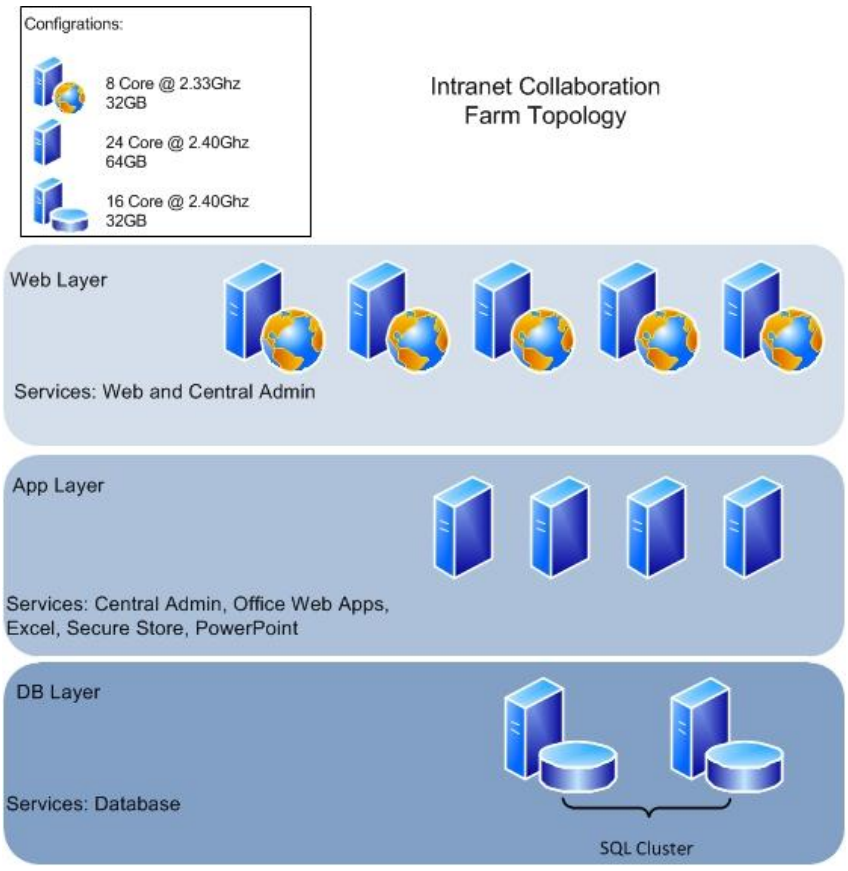
4.2.4.2. SharePoint Server 2010 Technical Case Studies

Microsoft has also published a number of technical case studies to demonstrate several different use cases based on existing production environments (http://technet.microsoft.com/en-us/library/cc261716.aspx). When planning your Starting Point architecture, select the case study that is most applicable to your organizations expected usage patterns:

- *Enterprise intranet publishing environment technical case study (SharePoint Server 2010)* available at http://technet.microsoft.com/en-us/library/ff758652.aspx.
- *Enterprise intranet collaboration environment technical case study (SharePoint Server 2010)* available at http://technet.microsoft.com/en-us/library/ff758650.aspx.
- *Enterprise intranet collaboration environment lab study (SharePoint Server 2010)* available at. http://technet.microsoft.com/en-us/library/ff758657.aspx.
- *Departmental collaboration environment technical case study (SharePoint Server 2010)* available at http://technet.microsoft.com/en-us/library/ff758649.aspx.
- *Divisional portal environment lab study (SharePoint Server 2010)* available at http://technet.microsoft.com/en-us/library/ff758649.aspx.
- *Social environment technical case study (SharePoint Server 2010)* available at http://technet.microsoft.com/en-us/library/ff758654.aspx.
- *Microsoft SharePoint Server 2010 social environment: Lab study* available at http://technet.microsoft.com/en-us/library/gg593613.aspx.

Figure 9 is from http://technet.microsoft.com/en-us/library/ff758650.aspx.

Figure 9. Enterprise Intranet Collaboration Environment Technical Case Study



4.3 SQL Server Capacity and Performance

SharePoint relies on SQL Server databases to share and store content. Architect the database layer carefully to give adequate capacity and disk throughput to the SharePoint application. Planning for SQL server capacity and performance in support of SharePoint is an extremely complex topic and designs can vary widely depending on your deployment specifics and the proliferation of service applications in your environment.

4.3.1 Configuration and Central Administration Database Sizing and Performance

The Configuration and Central Administration databases tend to be relatively small. Microsoft recommends that you allocate:

- 2GB for the configuration database – The configuration database can eventually grow beyond 1GB, but it grows at a very slow rate; approximately 40MB for each 50,000 site collections.
- 1GB for the central administration database.

Because Configuration database transaction logs can be quite large, Microsoft recommends that you change the recovery model for the database from full to simple. In addition, if using SQL Server database mirroring to provide availability for the Configuration database, Microsoft recommends that you use the full recovery model.

Disk throughput requirements for the Configuration and Central Administration databases are minimal.

4.3.2 Content Database Sizing and Performance

Every SharePoint environment is unique; therefore, estimating the disk capacity and throughput required for content databases is not a precise activity. The Microsoft guidelines below can help estimate the initial size of your deployment. Over time, as you monitor the production environment, you can easily adjust to changing disk capacity and throughput requirements using VMware features such as VMware Storage vMotion™.

4.3.2.1. Initial Content Database Sizing

The formula for estimating the size of your content databases is:

```
Database size = ((D × V) × S) + (10KB × (L + (V × D)))
```

- D = *the number of documents* you expect to host in the content database. This number should include both user documents and documents generated by automated processes. If you are migrating from a previous version of SharePoint, these statistics can be gathered from the current environment. If you are installing SharePoint for the first time, you can estimate the number of documents based on existing file shares or third-party document repositories.
- S = *the average size of each document* that will be stored in the content database. If average file size differs significantly between groups of users, it may be useful to estimate averages for different types or groups of sites. For example, Marketing users may have to store larger media files whereas Human Resources users might only need to store relatively small, static documents.
- L = *the number of list items in the environment*. List items are more difficult to estimate than documents. If you are migrating from a previous version of SharePoint, these statistics can be gathered from the current environment. If you are installing SharePoint from scratch, Microsoft uses an estimate of three times the number of documents (D), but this estimate may need adjustment based on actual production workload.
- V = t*he approximate number of document versions*. This value is usually much lower than the maximum allowed number of versions. For the purposes of the formula, the value should be greater than zero.

- 10KB = *a rough estimate for amount of metadata required by SharePoint Server 2010*. If you expect that your system will use a significant amount of metadata, you may want to increase this constant.

Microsoft provides an example that you can use to apply the formula (see *Storage and SQL Server capacity planning and configuration* (SharePoint Server 2010) at http://technet.microsoft.com/en-us/library/cc298801.aspx). Consider the data inputs and values shown in the following table.

Table 10. Sample Content Database Sizing Criteria

Formula Input	Value
Number of Documents (D)	200,000 Calculated by assuming 10,000 users times 20 documents
Average Size of Documents s	250KB
Number of List Items (L)	600,000
Number of non-current versions (V)	2 Assuming that the maximum versions allowed is 10

If you plug these values into the formula you get a value of 105GB for the content database.

```
Database size = (((200,000 x 2)) × 250) + ((10KB × (600,000 + (200,000 x 2))) =
110,000,000KB or 105GB
```

4.3.2.2. SharePoint Features That Impact Content Database Size

Now that there is a basic size for our SharePoint content, you must consider the impact of any SharePoint 2010 features that might significantly affect the size of the content database:

- Recycle bin contents – For information about the impact of recycle bins, see *Configure Recycle Bin settings (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/cc263125.aspx.
- Auditing data – For information about the impact of auditing, see *Storage and SQL Server capacity planning and configuration (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/cc298801.aspx.
- Office Web Apps – By default, the Office Web Apps cache is configured to be 100GB. For information about the size of the Office Web apps cache, see *Manage the Office Web Apps cache* at http://technet.microsoft.com/en-us/library/ee837422.aspx.

4.3.2.3. Content Database Size Guidelines

Limiting the content database size to the following guidelines improves manageability and performance of the SharePoint environment.

- Microsoft strongly recommends that you *limit the size of content databases to 200GB*.
- *A site collection should not exceed 100GB* unless it is the only site collection in the database.
- *Content database sizes up to 1TB* are supported only for large, single-site repositories and archives in which data remains reasonably static.

4.3.2.4. Disk Throughput Requirements for Content Databases

Disk throughput requirements for content databases can vary significantly between implementations. Microsoft recommends that you try to match your expected workload to one of their tested solutions in the TechNet article *Performance and capacity test results and recommendations (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/ff608068.aspx.

4.3.3 Service Application Storage Requirements

The following service applications utilize SQL Server for one or more databases. See the Microsoft TechNet for specific information on how to design these databases for capacity and performance at http://technet.microsoft.com/en-us/library/cc298801.aspx.

- Search
- User profile
- Managed metadata
- Web analytics
- Secure store
- State
- Word automation service
- Performance point

For detailed information about how to estimate capacity and disk throughput requirements for each of these service applications, see *Performance and capacity test results and recommendations (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/ff608068.aspx.

4.3.4 SQL Server Memory Requirements

The memory required for SQL Server to support SharePoint Server 2010 is impacted by the following:

- The size of content databases.
- The addition of service applications or features into the environment.
- The use of SQL Server mirroring.
- The frequent use of files larger than 15MB.

The following table gives Microsoft guidelines for the amount of memory to allocate to SQL Server. References to deployment size are based on *Topologies for SharePoint Server 2010: Model* at http://technet.microsoft.com/en-us/library/cc263044.aspx. Table 11 shows the SQL Server recommendations (derived from *Estimate performance and capacity requirements for Windows SharePoint Services collaboration environments (Office SharePoint Server)* at http://technet.microsoft.com/en-us/library/cc298801.aspx).

Table 11. SQL Server Memory Recommendations

Combined Size of Content Databases	Recommended SQL Server Memory
Minimum for small production deployments	8GB
Minimum for medium production deployments	16GB
Up to 2TB	32GB

Combined Size of Content Databases	Recommended SQL Server Memory
2TB to a maximum of 5TB	64GB

4.3.5 Network Topology Requirements

The following list provides a few best practices for the networking environment that will support SharePoint:

- Be sure to establish LAN-class bandwidth and latency between Web servers, application servers, and the SQL Servers. *Network latency should be <= 1ms.*
- Understand WAN bandwidth requirements if you are planning to use SQL Server mirroring or log shipping to keep a remote site up-to-date.
- Microsoft recommends that Web and application servers have two network adapters (or vNICs): one for end-user traffic and the other for communication with the servers running SQL Server. From a VMware perspective, placing Web or application servers on the same ESX/ESXi host as the SQL Server can improve performance by using internal virtual switches.

4.3.6 Guidelines for Scaling Out SQL Server

If you are running into performance issues where the database is the bottleneck, consider scaling out your databases across more instances of SQL Server.

The following is a summary of Microsoft's guidance on when to deploy an additional SQL Server. These are not necessarily requirements for support as Microsoft supports configurations that do not follow this guidance.

- Add an additional database server when you have *more than four Web servers that are running at full capacity*.
- Add an additional database server when your *content databases exceed 5TB*.
- Add an additional database server to promote *secure credential storage* when you are running the *Secure Store* service application.

4.4 Minimum Resource Requirements for SharePoint Server Roles

Note the new SharePoint 2010 minimum requirements, for example, all server roles are now required to have at least four 64-bit processor cores (or vCPUs).

4.4.1 Web Servers, Application Servers, and Single Server Installations

Table 12 shows the server requirements from *Hardware and software requirements (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/cc262485.aspx.

Table 12. Web, App, and Single Server Minimum Requirements from Microsoft TechNet

Component	Minimum Requirement
Processor	64-bit, four cores.
RAM	4GB for developer or evaluation use. 8GB for production use in a single server or multiple server farm.
Hard disk	80GB for system drive. You must have sufficient space for the base installation and for diagnostics such as logging, debugging, creating memory dumps, and so on. For production use, you also need additional free disk space for day-to-day operations. Maintain twice as much free space as you have RAM for production environments. For more information, see *Capacity management and sizing for SharePoint Server 2010* at http://technet.microsoft.com/en-us/library/cc261700.aspx.

4.4.2 Database Servers

Table 13 shows the database server requirements from *Hardware and software requirements (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/cc262485.aspx.

Table 13. Database Server Minimum Requirements from Microsoft TechNet

Component	Minimum Requirement
Processor	• 64-bit, four cores for small deployments. • 64-bit, eight cores for medium deployments.
RAM	• 8GB for small deployments. • 16GB for medium deployments. For large deployments, see the "Estimate memory requirements" section in *Storage and SQL Server capacity planning and configuration (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/cc298801.aspx.

Component	Minimum Requirement
Hard disk	80GB for system drive. Hard disk space is dependent on the size of your SharePoint content. For information about estimating the size of content and other databases for your deployment, see *Storage and SQL Server capacity planning and configuration (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/cc298801.aspx.

4.5 vSphere Limitations

Although vSphere can support very large virtual machines, there are some limits to be aware of when planning for capacity. Below are some of the most commonly referenced limitations. Complete documentation is available in *vSphere Configuration Maximums* at http://www.vmware.com/pdf/vsphere4/r41/vsp_41_config_max.pdf.

- vSphere virtual machines are limited to 8 vCPUs and 255GB of RAM.
- Each ESX/ESXi host can accommodate only up to 255 LUNs.
- Each vSphere LUN is limited to 2TB (without SAN extents).

Any sizing examples in this document take these limitations into account. For example, be sure to limit virtual machine configurations to 8 vCPUs based on the vSphere maximum.

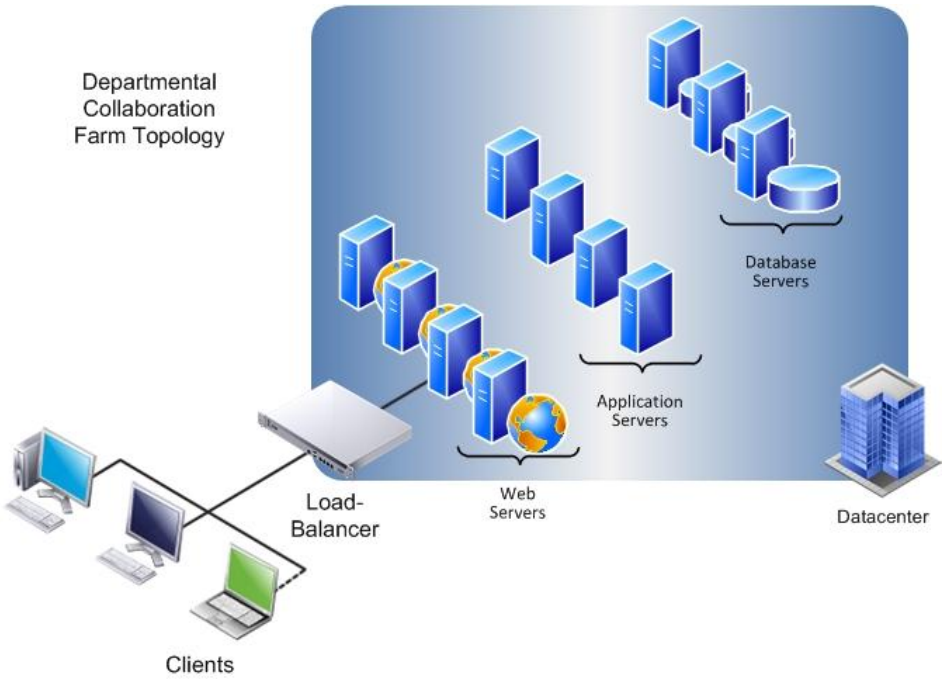
4.6 Sample Architecture Deployment to vSphere

Planning a deployment of SharePoint virtual machines to vSphere is mostly about properly allocating compute resources to maximize application performance. As an example, this refers to Microsoft's *Departmental Collaboration Environment Technical Case Study (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/ff758649.aspx, which is a live, production workload that was documented to aid in the design process. The server role requirements from the case study are used to determine the virtual machine configurations and the appropriate hardware for the ESX/ESXi hosts. Then the initial deployment of the SharePoint virtual machines is mapped to the ESX/ESXi hosts to avoid overcommitment of host resources. To make this architecture work in a virtual environment, some important modifications were made to the case study:

- The SQL Server hosting the content databases calls for 16 processor cores and 32GB of memory. Because this exceeds current vSphere limitations for vCPU allocation (maximum of 8 vCPUs in vSphere 4.1), this large SQL Server was split into two SQL Server virtual machines with 8vCPUs each. The memory was left at 32GB, which is likely much more than is required for this workload, but can be adjusted after you have monitored for actual memory usage.
- Some SQL Server configurations in this case study specify Direct Attached Storage (DAS) for the TempDB databases; however, placing all storage on the SAN enables key VMware features, such as vMotion, DRS, HA, and Storage vMotion. Links to original storage recommendations are included in the resource requirements table below.

This case study is used for illustration purposes only to demonstrate resource allocation and mapping of SharePoint virtual machines to ESX/ESXi hosts. You should always plan your own environment using Microsoft Best Practices and case study examples. Figure 10 is from *Departmental collaboration environment technical case study (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/ff758649.aspx.

Figure 10. Departmental Collaboration Environment Technical Case Study



4.6.1 Resource Requirements by Server Role

With any application, you should plan its resource requirements first, and then the underlying infrastructure requirements to meet the application's needs. Table 14 lists the resource requirements for each SharePoint server role in the Departmental Collaboration Case Study (see http://technet.microsoft.com/en-us/library/ff758649.aspx). Again, note that the SQL Server hosting the content databases was split into two 8 vCPU virtual machines to account for vSphere 4.1 limitations. To avoid resource overcommitment (for now), consider 1 core equal to 1 vCPU for the purposes of this example.

Table 14. SharePoint Resource Requirements by Server Role

SharePoint Role	Physical Resources (per server)
Web Servers 1 & 2 (content)	CPU: 8 cores (2.33 GHz) Memory: 32GB Storage: See the case study at: http://technet.microsoft.com/en-us/library/ff758649.aspx Network: 2 x 1Gbps adapters
Web Server 3 (content) Web Server 4 (search/crawl)	CPU: 8 cores (2.33 GHz) Memory: 16GB Storage: See the case study at: http://technet.microsoft.com/en-us/library/ff758649.aspx Network: 2 x 1Gbps adapters

SharePoint Role	Physical Resources (per server)
Application Server 1 (Central Admin) Application Server 2 and 3 (Office Web Apps)	CPU: 8 cores (2.33 GHz) Memory: 16GB Storage: See the case study at: http://technet.microsoft.com/en-us/library/ff758649.aspx Network: 2 x 1Gbps adapters
Application Server 4 (Crawler)	CPU: 8 cores (2.33 GHz) Memory: 16GB Storage: See the case study at: http://technet.microsoft.com/en-us/library/ff758649.aspx Network: 2 x 1Gbps adapters
Database Server 1 & 2 (Content)	CPU: 8 cores (3.2 GHz) Memory: 32GB Storage: See the case study at: http://technet.microsoft.com/en-us/library/ff758649.aspx Network: 2 x 1Gbps adapters
Database Server 3 (Web Analytics)	CPU: 8 cores (3.2 GHz) Memory: 16GB Storage: See the case study at: http://technet.microsoft.com/en-us/library/ff758649.aspx Network: 2 x 1Gbps adapters
Database Server 4 (Search Databases)	CPU: 8 cores (3.2 GHz) Memory: 32GB Storage: See the case study at: http://technet.microsoft.com/en-us/library/ff758649.aspx Network: 2 x 1Gbps adapters

4.6.2 Planning the Host Hardware Configuration

4.6.2.1. ESX/ESXi Host Specifications

Each ESX/ESXi host must provide sufficient physical hardware resources to accommodate the planned workload and provide some headroom in case there is a VMware HA failover or planned vMotion migration of live virtual machines. Table 15 summarizes the ESX/ESXi host hardware configuration based on the example architecture (see *Hardware and software requirements (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/cc262485.aspx).

Table 15. Database Server Minimum Requirements from Microsoft TechNet

ESX Host	Virtual Machines
All ESX/ESXi hosts	24 cores (4x6)
	96GB RAM (well above requirements to aid in failover)
	2 Fibre Channel HBAs
	4Gbps network adapters

4.6.2.2. Virtual Machine Distribution

In this example, four ESX/ESXi hosts were connected to shared storage to use advanced VMware features such as HA and DRS.

To build infrastructure availability into the architecture, the twelve total virtual machines were distributed across four physical ESX/ESXi hosts. Initial placement of virtual machines is relatively unimportant, especially if you are using DRS. Table 16 shows the configuration (see *Hardware and software requirements (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/cc262485.aspx).

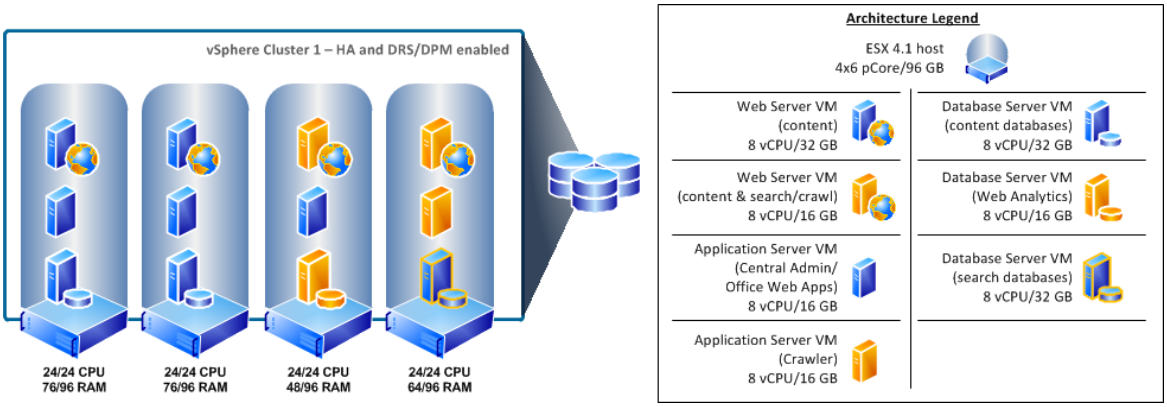
Table 16. Database Server Minimum Requirements from Microsoft TechNet

ESX/ESXi Host	Virtual Machines
ESX/ESXi Host 1	Web Server virtual machine 1 (8 vCPU/32GB RAM)
	Application Server virtual machine 1 (8 vCPU/16GB RAM)
	Database Server virtual machine 1 (8 vCPU/32GB RAM)
ESX/ESXi Host 2	Web Server virtual machine 2 (8 vCPU/32GB RAM)
	Application Server virtual machine 2 (8 vCPU/16GB RAM)
	Database Server virtual machine 2 (8 vCPU/32GB RAM)
ESX/ESXi Host 3	Web Server virtual machine 3 (8 vCPU/16GB RAM)
	Application Server virtual machine 3 (8 vCPU/16GB RAM)
	Database Server virtual machine 3 (8 vCPU/16GB RAM)
ESX/ESXi Host 4	Web Server virtual machine 4 (8 vCPU/16GB RAM)
	Application Server virtual machine 3 (8 vCPU/16GB RAM)
	Database Server virtual machine 3 (8 vCPU/32GB RAM)

4.6.2.3. Initial Virtual Machine Placement

Although the workloads migrate automatically with DRS (including the SQL Servers because there is no clustering), it is often useful to visualize the initial placement of virtual machines to aid in calculating resource allocation and host failover capacity.

Figure 11. Initial Virtual Machine Placement



4.6.3 Improving Consolidation Ratios

According to the case study, actual utilization of memory and processor are fairly low (see Figure 11), offering the opportunity to improve consolidation ratios by making adjustments to the environment. For example, although the Web and database servers appear to have the appropriate processor allocation, the application servers seem to be under-utilized. If this were a real production environment on vSphere, you could reduce the number of vCPUs allocated to the application servers, freeing up more cores and allowing increased virtual machine density on each host. Another approach would be to utilize vCPU shares to throttle the amount of actual processor each virtual machine is consuming without having to change its configuration. See *Hardware and software requirements (SharePoint Server 2010)* at http://technet.microsoft.com/en-us/library/cc262485.aspx).

Table 17. Departmental Collaboration Case Study Processor Utilization

Server Role	Documented Utilization
Front-end Web Server	Typical – 82.82% Peak – 88.12%
Application Server	Typical – 11.88% Peak – 14.32%
Database Server	Typical – 64.06% Peak – 67.97%

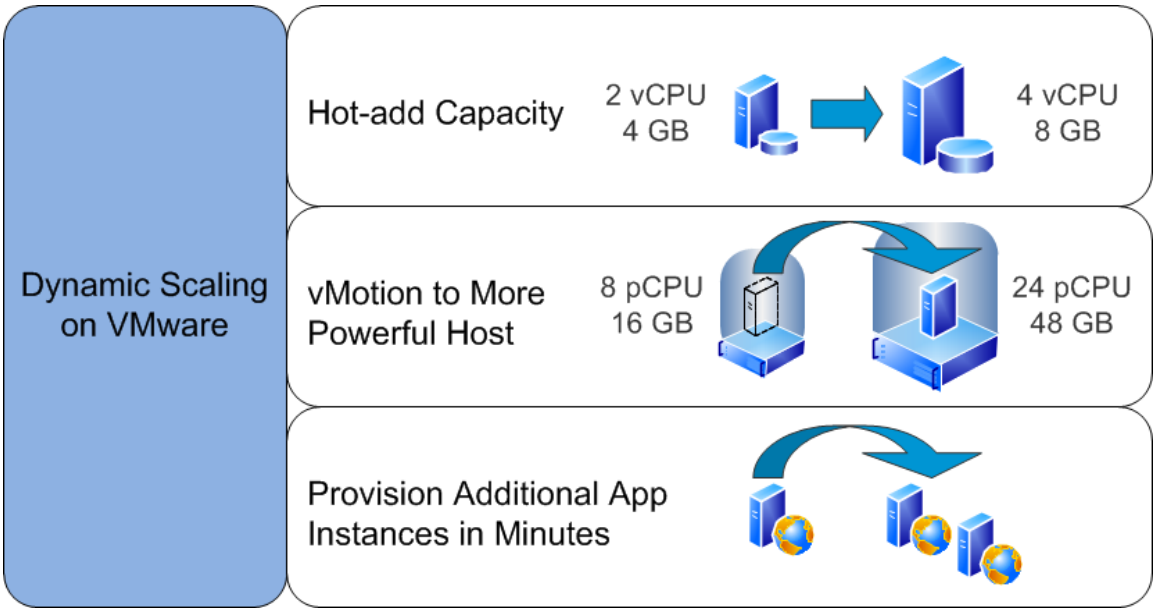
5. vSphere Enhancements for Deployment and Operations

It is very unlikely that a newly-deployed SharePoint environment will remain static. Many SharePoint environments start out small and expand as user demand increases. Even well-established SharePoint environments are subject to change as business needs drive the requirement to support more users or more advanced features. Deploying SharePoint to vSphere offers a flexible environment where resource allocations can be modified quickly and easily, in some cases without interrupting the underlying application. In addition, VMware products, such as VMware AppSpeed™ Server and VMware vCenter CapacityIQ™, help you to detect SharePoint performance bottlenecks and future capacity shortfalls, enabling you to respond proactively to avoid costly end-user impact.

5.1 Responding to Changing Workloads

Out of the box, vSphere offers several capabilities that enable you to quickly respond to changing usage patterns. Allocation of processor and memory resources to virtual machines can be easily changed to suit the most current business requirements and, in the case of Hot-Add, without any interruption to the operating system or application. You can use vMotion to migrate heavily used SharePoint virtual machines to another ESX/ESXi host to alleviate physical resource bottlenecks. Finally, template-based provisioning allows the rapid deployment of new SharePoint virtual machines to satisfy increased load.

Figure 12. Dynamic Scaling on VMware



5.1.1 Controlling Resource Allocation (Scaling-Up)

Compute resources allocated to SharePoint virtual machines running on vSphere can be easily adjusted to accommodate usage fluctuations and business changes. With a quick shutdown/restart, processor and memory resources for any virtual machine can be adjusted as requirements change or as bottlenecks are discovered. For SQL Server 2008 database servers, processor and memory resources can be added without interrupting the application.

5.1.1.1. Adding Resources to Web and Application Servers

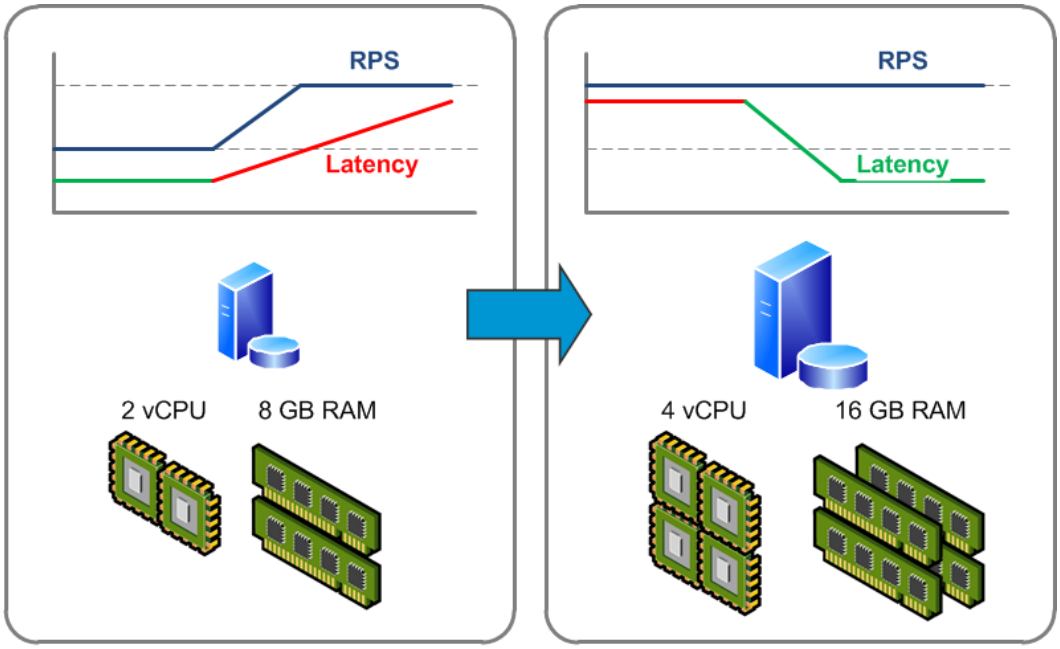
SharePoint Web and application servers unfortunately cannot yet take advantage of the VMware Hot-Add feature; however, processor and memory resources can be adjusted by simply shutting down the virtual machine and changing the virtual machine properties. When the virtual machine restarts, the operating system and application will recognize and begin using the new resources.

5.1.1.2. Hot-Adding Resources to SQL Server

For virtual machines running SQL Server 2008 Enterprise edition, the VMware Hot-Add features allows you to add processor and memory to a running virtual machine without interrupting the operating system or the SQL Server application. SQL Server immediately recognizes and begins using the new resources. If you need to remove processor or memory resources from a virtual machine, the classic approach of shutting down and adjusting properties still applies.

Hot-Add could be used to respond quickly to a detected SQL Server bottleneck to minimize performance impact to SharePoint as shown Figure 13.

Figure 13. Using VMware Hot-Add with an SQL Server 2008 Guest



- **Hot-add capacity with zero application downtime**
- **Minutes to stabilize VM and recover from SLA violation**
- **Other options include VMotion to more powerful host & add instance for fast scale-out**

5.1.2 Scaling Out

In some cases, the best response to changing usage patterns is to scale out the workload by deploying more virtual machine instances or by spreading the workload across more ESX/ESXi hosts. VMware enables you to scale-out your SharePoint environment rapidly, with minimal to no impact on the application.

5.1.2.1. Migration to Additional Hosts with vMotion

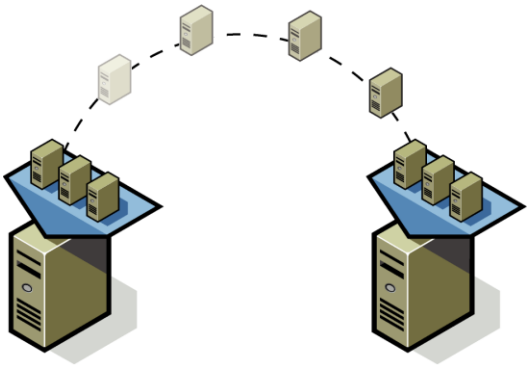
Another approach to scaling out is to run the same number of SharePoint virtual machines on a greater number of ESX/ESXi hosts as the individual virtual machines begin to increase resource requirements. For example, if you have set up an initial environment to support 500 pilot users, over time, more pilot users are added, producing a strain on the system to which you respond by scaling up virtual machine resource allocations. As virtual machine resource requirements start to overwhelm the physical resources on the ESX/ESXi host, you can add more ESX/ESXi hosts to the cluster and use vMotion to seamlessly migrate heavily used virtual machines to the new hosts, alleviating the resource bottlenecks. vMotion is supported for all SharePoint server roles including SQL Server, unless SQL Server is using shared-disk failover clustering (for more information, see *Microsoft Clustering on VMware vSphere: Guidelines for Supported Configurations* at http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1037959.

5.1.2.2. Template-Based Deployments for Rapid Provisioning

In an environment with established procedures, deploying new application servers can be streamlined, but can still take many hours or days to complete. Not only must you complete an OS installation, but downloading and installing service packs and security updates can add a significant amount of time. Applications like IIS and SharePoint require features that are not installed with Windows by default must be installed prior to installing the applications. Inevitably, those features require more security updates and patches. By the time all deployment aspects are considered, more time is spent waiting for downloads and installs than is spent configuring the application.

Virtual machine templates can help speed up this process by eliminating most of these monotonous tasks. By completing the core installation requirements, typically to the point where the application is ready to be installed, you can create a golden image which can be sealed and used as a template for all of your virtual machines. Depending on how granular you want to make a specific template, the time to deployment can be as little as the time it takes to install, configure, and validate the application. You can use PowerShell tools for SharePoint and VMware vSphere PowerCLI to bring the time and manual effort down dramatically.

Figure 14. Virtual Machine Templates

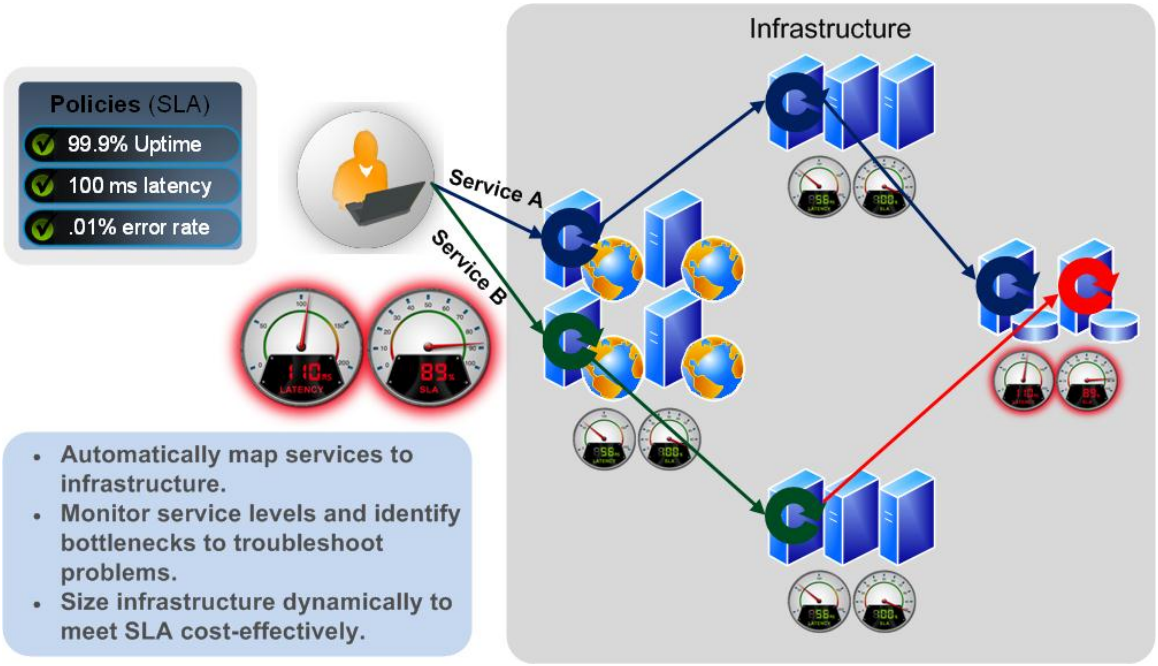


5.2 Monitoring Performance with vCenter AppSpeed

Isolating performance issues in a SharePoint environment can be a tedious task. Having tools that understand the communication patters of the application and which provide insight into the dependencies and how servers communicate can reduce time to troubleshoot performance problems.

VMware vCenter AppSpeed can provide a real-time view of SharePoint performance from an end-user point of view. When integrated with a vSphere environment, AppSpeed discovers and maps the applications and protocols found while traversing the virtual network. You can use the data collected by AppSpeed to measure performance and ensure SLAs are being met. Besides monitoring the data collected is used to map dependencies and help you understand where in the stack the performance degradation is occurring.

Figure 15. Monitor Multi-Tiered Application Performance with VMware AppSpeed

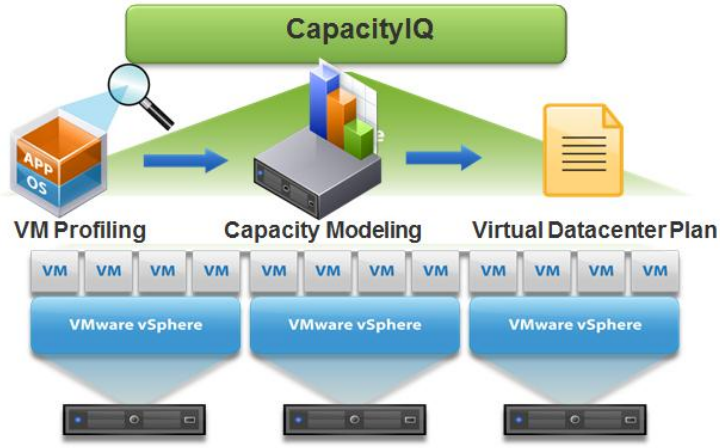


5.3 Intelligently Monitor Capacity with vCenter CapacityIQ

Many organizations deploy tools to monitor SharePoint and the underlying OS and hardware for faults and capacity warnings. Tools from Microsoft and other third-party vendors do a great job of providing application and OS level details. Unfortunately, when virtualized, these tools do very little to monitor the capacity of the underlying environment. Often times SharePoint application and Web servers are built with a fixed amount of memory and CPU, based on the recommendations from Microsoft. While deficiencies in this design can quickly be observed by the users, over-provisioning is usually overlooked. Application owners usually do not complain that their application has more resources than it really requires, however, this can lead to inefficient use of the resources and lower the possible levels of consolidation.

VMware vCenter CapacityIQ brings vSphere-aware capacity monitoring and reporting to vCenter. By monitoring the usage and performance characteristics of virtual machines, CapacityIQ can provide forecasting of resource consumption over time. In many cases, development and test environments are built to mimic production at the request of developers. CapacityIQ can provide usage details to help you determine if these virtual machines are over-provisioned and help reclaim unused capacity.

Figure 16. Right-Size Applications with vCenter CapacityIQ



Forecast timing of capacity shortfalls and needs

Perform "What-If" impact analysis to model effect of capacity changes

Identify and reclaim unused capacity