Realtime publishers

The Shortcut Guide™ To

Storage Considerations for Microsoft SharePoint

sponsored by

Wendy Henry

Introduction to Realtime Publishers

by Don Jones, Series Editor

For several years, now, Realtime has produced dozens and dozens of high-quality books that just happen to be delivered in electronic format—at no cost to you, the reader. We've made this unique publishing model work through the generous support and cooperation of our sponsors, who agree to bear each book's production expenses for the benefit of our readers.

Although we've always offered our publications to you for free, don't think for a moment that quality is anything less than our top priority. My job is to make sure that our books are as good as—and in most cases better than—any printed book that would cost you $40 or more. Our electronic publishing model offers several advantages over printed books: You receive chapters literally as fast as our authors produce them (hence the "realtime" aspect of our model), and we can update chapters to reflect the latest changes in technology.

I want to point out that our books are by no means paid advertisements or white papers. We're an independent publishing company, and an important aspect of my job is to make sure that our authors are free to voice their expertise and opinions without reservation or restriction. We maintain complete editorial control of our publications, and I'm proud that we've produced so many quality books over the past years.

I want to extend an invitation to visit us at http://nexus.realtimepublishers.com, especially if you've received this publication from a friend or colleague. We have a wide variety of additional books on a range of topics, and you're sure to find something that's of interest to you—and it won't cost you a thing. We hope you'll continue to come to Realtime for your educational needs far into the future.

Until then, enjoy.

Don Jones

Copyright Statement

Chapter 1: SharePoint Capacity Requirements

In March 2008, Bill Gates revealed that SharePoint had joined the ranks of Exchange and SQL Server by becoming one of Microsoft's fastest growing network applications, surpassing 100 million licenses sold and anticipating 2008 sales in excess of one billion US dollars. The seed idea of using Microsoft's collaboration platform as a viable alternative to traditional file storage systems has germinated among enterprise network administrators thanks to SharePoint's improved SQL Server repository and mature document management features added in 2007. Integrated content management capabilities have caught the fancy of Web masters looking for a simple, flexible alternative to heavy ASP.NET development on traditional IIS hosting. In fact, some enterprise network implementations of SharePoint include such diverse custom programming that the application has become a platform unto itself, supporting not only data delivery but client software interface delivery as well.

With so many reasons to prefer SharePoint as an all-in-one solution for disseminating information, it stands to reason that an integral factor in any solid SharePoint deployment plan should be storage. And yet, selecting an appropriate storage system solution is often overshadowed by elaborate governance and logical design planning. Perhaps many procurement engineers simply rely on the old "more free space is always better" adage by purchasing big hard drives in the server hardware. Or maybe the SharePoint planning committee is too overwhelmed by the unknown aspect of user content contributions to accurately choose an appropriate storage system for SharePoint at the onset of deployment. Whatever the reason, SharePoint suffers one of the highest content migration rates within the same version when compared with other Microsoft enterprise server applications.

SharePoint, by its very nature, is a user-empowering platform that lives and breathes according to the whims of those contributing to it. SharePoint employs IIS to deliver multiple Web sites worth of pages containing anything from flat text content to streaming audio/video, all of which can be contributed by end users. Multiple SharePoint Web sites are organized into hierarchical Site Collections for administrative ease, and these collections reside together or separately in SharePoint Web Applications. Each Web Application object in SharePoint has a 1:1 relationship with a Web site object in IIS as well as a one-to-one relationship with a SharePoint Content Database in SQL Server (see Figure 1.1). SQL Server is rounded out by several configuration, search, and administrative content databases to provide the entire SharePoint environment. More about IIS and the databases will appear later in this guide.

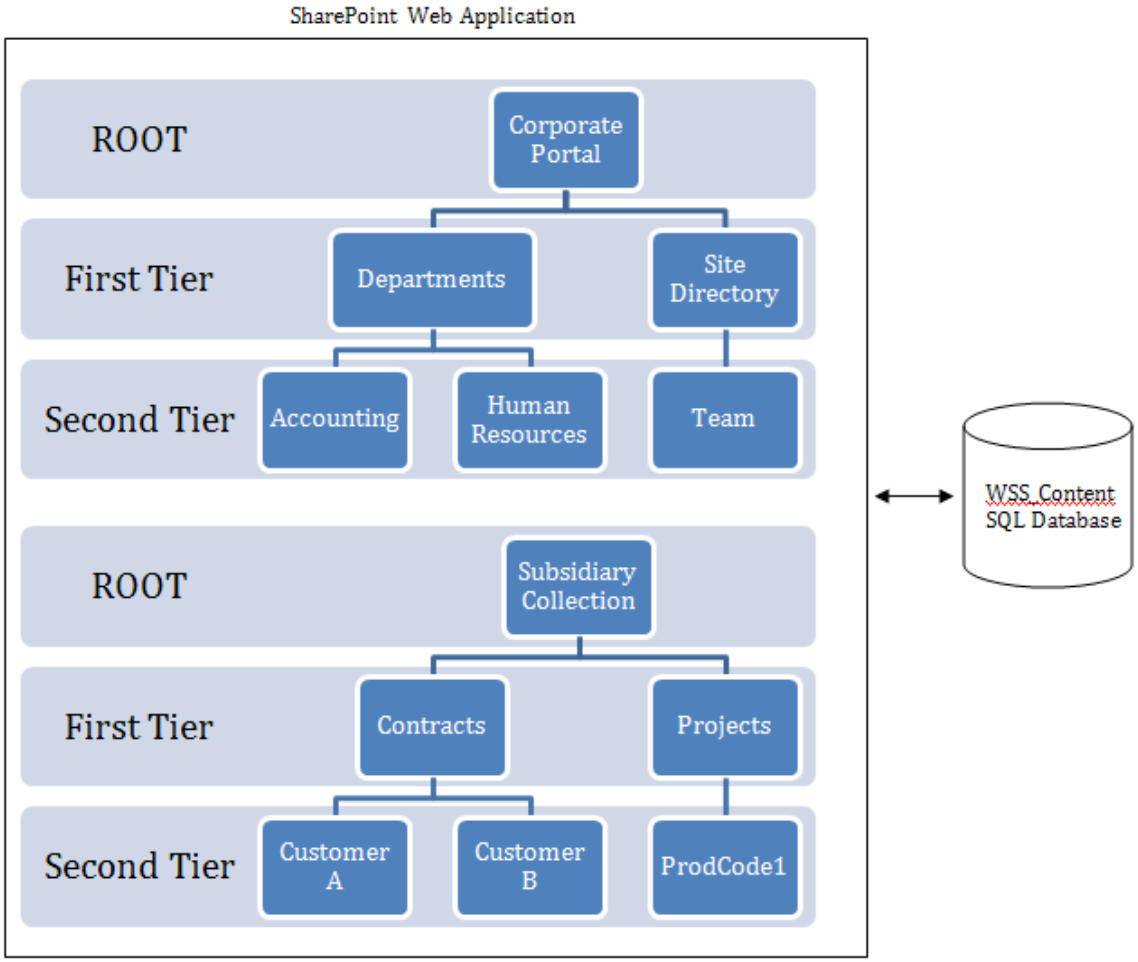

Figure 1.1: Two site collections in a single Web application.

This guide will examine the storage capacity requirements of SharePoint by dissecting the SharePoint components and storage options for them. Specifically, the following topics will be discussed in detail:

- Organic growth in collaboration systems and why a flexible storage system is key to success
- A breakdown of SharePoint components and SQL databases and their individual storage needs
- Storage system options for SharePoint and best practices for iSCSI solutions
- Assuredness considerations for SharePoint including high availability and disaster recovery

Why Capacity Planning Is Important

As you may be starting to realize, planning SharePoint storage strictly by the Microsoft Guidelines for user content storage needs would be a huge mistake. There are far more space requirements to consider beyond the current total file sizes on the User Home Folder file server you plan to migrate into SharePoint. And not all the components of a SharePoint realm will benefit from the same storage solution. For example, SQL Server thrives on fast network storage whereas IIS prefers robust local storage due to its residency within the Windows Server operating system (OS). Attempting to force fit all of SharePoint into a single storage design at best limits future expansion and at worst may detrimentally affect performance and functionality.

Collaboration = Organic Growth of Content

Any honest discourse about SharePoint should assume an understanding of the product's intended use, and this guide is no exception. Microsoft advertises SharePoint as a collaboration platform and while the term *collaboration* seems to be a buzzword bantered about excessively in IT, within the context of this discussion it has great importance. Collaborative solutions are meant to provide communication tools that encourage free thought, brainstorming, and creativity. The data derived from such collaborative communication will tend to shrink and grow without logical forethought or predictability. Any system providing the "space" for such communication and its derived content therefore must be flexible enough to breathe with the human element. Forecasting models and prediction castings may prove ineffectual while planning a collaborative environment.

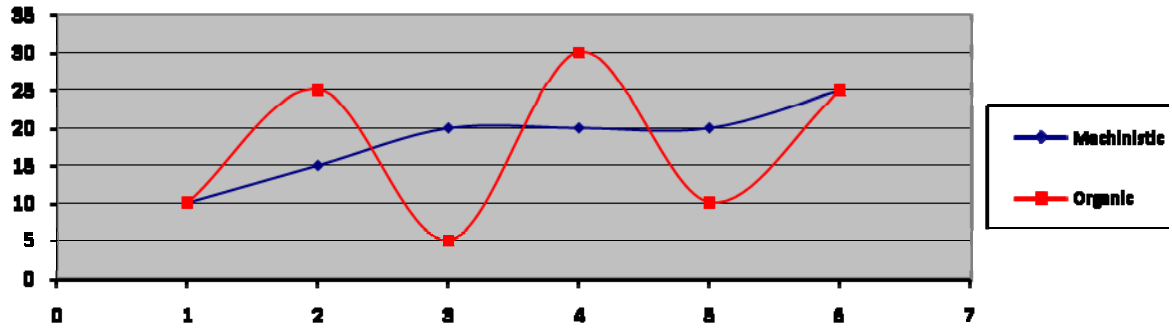

Figure 1.2: Data size fluctuation along known intervals—mechanistic vs. organic.

Planning end user activity in a loosely coupled system such as SharePoint is problematic at best. Figure 1.2 displays the difference between a well-managed, linear pattern of data growth in mechanistic systems such as traditional relational databases and a dynamic, chaotic pattern of data growth in organic systems such as collaborative applications. Often, engineers responsible for choosing SharePoint storage will approach planning the same as they would for a traditional SQL Server production database need (one with fixed data types, defined constraints, and managed triggers). The danger in such an approach is that SQL Server databases built for the purpose of SharePoint should never be manually controlled by a database administrator in the same fashion as traditional production databases. In fact, Microsoft best practices for SharePoint discourage manipulating any SharePoint databases through the SQL Server management utilities. SharePoint automatically builds normalized databases during installation that are then managed strictly by SharePoint processes and that allow end-user defined construct choices such as column types on a list that dictate data types within the database tables.

Cross-Reference

For a detailed comparison of mechanistic and organic systems, check out "Theory of Mechanistic and Organic Systems" by Tom Burns and G.M. Stalker at http://www.valuebasedmanagement.net/methods_burns_mechanistic_organic_systems.html.

Throwing Money Down an I/O Black Hole

Yet incorporating the idea of collaboration with a tangible data system requires a certain degree of prediction. Otherwise, how would you know what specifications to request on your new server hardware for SharePoint? Despite your best guess, eventually SharePoint users will fully populate the initial storage and the system will require a hardware upgrade or migration. In SharePoint, this growth is compounded by the fact that working space isn't enough, you must also consider high-availability and disaster recovery space requirements. It would be simple to assume that adding hard drive free space to a SharePoint server would resolve the issue. However, repeatedly purchasing additional hard drives, whether installed on the local system or a network storage device, only postpones the inevitable (and spends a large amount of your hardware budget in the process). There will come a point when the bell curve peaks and the overhead associated with managing multiple devices begins to outweigh the benefit of offering more free space to SharePoint. Purchasing additional hard drives after this point is like throwing money down an I/O subsystem black hole. The system would be better served by consolidating onto larger, faster media solutions.

There are many factors of an I/O subsystem that may cause inefficient use of a multiple drive solution. First and foremost, if RAID striping is not being employed, each drive is treated as a single entity without the benefit of multiple mechanical devices being utilized simultaneously. Additionally, rotation and seek speeds are often overlooked during purchase in favor of connection interface and total size. Connection interface choice is routinely based on cost, unfortunately, rather than best transfer rate. Occasionally, engineers may believe that adding external storage rather than internal drives will alleviate some of the controller and bus congestion. However, introducing drivers and paths for the external devices can create challenges of their own, not to mention the increase in human administration overhead.

When it comes to the Windows OS software, simply throwing additional hardware at a storage inadequacy can cause instruction bottlenecks, virtual memory inefficiencies, and wasted unused space on the hard drive. The file system requirement for SharePoint NTFS invokes a dynamic cluster definition based on multiples of the hardware's size. The enhanced stability of NTFS is courtesy of logged transactional writing and other overhead data that can utilize more than 4% of each volume. Due to its relatively tight default block size, NTFS becomes fragmented easily and may require frequent or scheduled defragmentation maintenance. These limitations are compounded by SharePoint's repository application, SQL Server. Microsoft SQL Server uses a managed 64KB extent to store row data that requires overhead data bits not only for every extent but also on every page within an extent.

Point of Diminishing Return in Performance

Diminishing returns on your I/O subsystem performance may not appear immediately upon additional hard drive installation but rather over time. The best way to track performance of a hard drive or storage facility is to use an object-oriented trace application that queries the hardware itself to determine activity details. Windows Server 2003 ships with a free performance monitoring tool called System Monitor. For-purchase monitoring utilities are also available from third-party manufacturers. Despite your choice of tool, you should develop a baseline of acceptable counter values and continually monitor your hard disk activity to be on the lookout for a negative change in course. It is virtually impossible to know what a server looks like sick unless you know what it looks like when it is healthy and operating within acceptable performance ranges. Measuring performance counters only *after* a perceptible degradation in performance without the benefit of comparison statistics from prior to the hard drive additions will be of little proof.

Prior to SharePoint deployment, monitoring the I/O subsystem of your current file servers will give you an idea of the capacity and performance needs your SharePoint server equipment will face. Similarly, upon adding hard disks to a SharePoint storage system, monitoring these same object counters will expose a point of diminishing return. Common measurements can be found in Table 1.1.

Object	Counter	Description	Preferable Values
Physical Disk	Avg. Disk Read Queue Length	Number of read instructions waiting on controller	<2
Physical Disk	Avg. Disk Write Queue Length	Number of write instructions waiting on controller	<2
Physical Disk	% Disk Time	Percentage of time disk was busy processing instructions during query interval	<50%
Physical Disk	Disk sec/Transfer	Duration of average transfer	<4
Logical Disk	% Free Space	Percentage of volume that is empty	>20

Table 1.1: Common disk performance measures.

The preferable values suggested in Table 1.1 are merely a baseline starting point. In a production SharePoint environment, you might find that values beyond those in the table are still acceptable in your environment. You can make such a determination by examining these counters within the context of your SLA, user feedback, network traffic density, and other overall network performance indicators. Once you determine the acceptable ranges for these counters, monitor them on a routine basis to catch fluctuations that may indicate your storage may be approaching a point of diminishing return. This can be accomplished by using the Performance Logs & Alerts in Windows Server 2003 to schedule routine counter value collection and save the log to a text file or SQL table for further analysis (see Figure 1.3). System monitor counters and performance logs deliver the most accurate results when initiated from a remote computer. Launching performance monitoring software from a remote system and configuring it to query the server of interest will alleviate inaccurate readings that may include the resources utilized by the monitoring application itself.

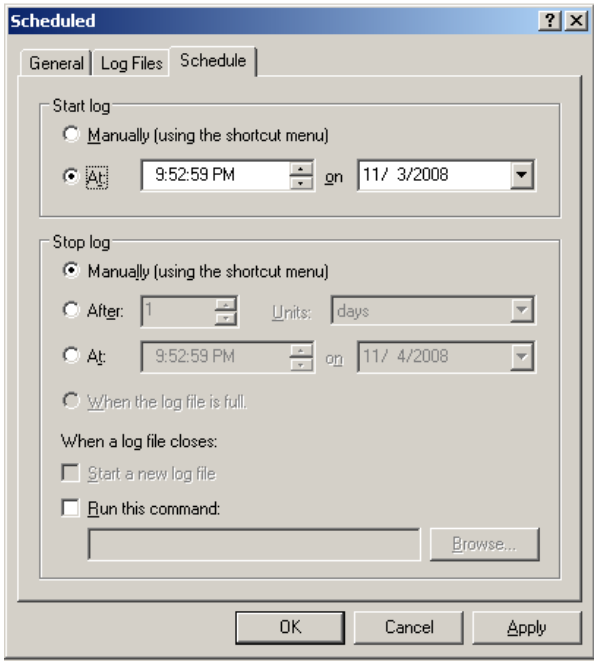

Figure 1.3: Scheduling performance counters in Windows performance logs.

Planning for SharePoint storage is not an exact science. However, it is necessary in order to procure the most appropriate storage system you can afford. Use performance monitoring and Microsoft guidelines to begin anticipating the storage you need. Be sure to purchase a flexible system that can adapt quickly to expanding capacity needs as SharePoint will easily become one of the most frequently modified platforms among your server roles.

Primer for SharePoint Storage Requirements

Identifying the storage requirements for SharePoint considers not only the SharePoint application but its dependency applications, SQL and IIS, as well. Breaking each down into minimum requirements will give you a starting point from which you can then increase demands according to the nature of how SharePoint will be used. The beauty of SharePoint is its versatility. Some companies may choose to employ only the document features as a replacement for an obsolete document management system. Others may choose the enterprise search features to improve user productivity. There is no single, correct way to use SharePoint. The more you use, the larger its storage needs. Although archive and retention policies can help alleviate congestion on expensive disks intended for frequently accessed data, a true storage plan should calculate both current and historic data sizes as well as the appropriate storage solution for each.

> Archiving and data maintenance will be discussed in a later chapter of this book.

SharePoint is a .NET generation application from Microsoft that follows a compartmentalized build strategy. Like pegs on a peg tool board, you insert the code chunks (pegs) you want to use where you want to use them. These isolated components of code are called *Features* and yes, that is a capital F. Features can be installed using the SharePoint stsadm.exe utility at the Command Line Interface (CLI, formerly known as the Command Prompt). Once installed, the Features can then be enabled or disabled throughout the Site Collection site by site by site depending on site template. Planning the use of Features may influence your logical SharePoint design, which might then impact your SQL Server storage needs. Remember, additional SharePoint web applications will generate additional SQL Server content databases.

Feature Management in SharePoint

As if SharePoint weren't versatile enough, some implementations may require additional functionality that simply does not ship with the product. In addition to system-supplied Features that can be easily enabled or disabled, there are third-party sources of custom Features that add pizzazz—and storage space requirements—to your SharePoint environment (such as www.codeplex.com by Scot Hillier, a SharePoint MVP).

Some Features, depending on their author, may include a simple batch file or self-executable that will perform the installation of the Feature into SharePoint for you. If not, the CLI utility stsadm.exe will install a feature via the following command:

```
Stsadm.exe -o installfeature
   -filename {path to new feature's .xml file}
   -name {name of folder to be created to host new feature}
```

The –o, -filename, and –name are required parameters. An optional switch called –force can be used to force the reinstallation of an already existing Feature. The –filename path must be relative to SharePoint's 12\Template\Features directory and must already contain the .xml file of the new Feature.

Don't forget to include custom Features in your storage plans!

Are Microsoft Deployment Guidelines Realistic?

Microsoft has definite opinions when it comes to deploying their network applications such as SQL Server and SharePoint. Deployment guideline documents are available for virtually every Microsoft product but are not easily found in a single resource. Some deployment guidelines outline strict minimum and maximum ranges for all facets of the product, including storage. Others, such as the SharePoint guides, offer a set of tested maximums that breach acceptable performance measurements when installed on finite hardware specifications. The vague nature of the SharePoint guide is a direct result of the product's fluidity and the many configurations in which it can be installed. Links to the Microsoft deployment guides appear in Table 1.2.

Product	Microsoft Deployment Guide Link
IIS 6.0	http://www.microsoft.com/technet/prodtechnol/WindowsServer2003/Library/IIS
SQL Server 2005	http://technet.microsoft.com/en-us/sqlserver/bb671416.aspx
Windows SharePoint Services 3.0	http://technet.microsoft.com/en-us/library/cc288426.aspx
Microsoft Office SharePoint Server 2007	http://technet.microsoft.com/en-us/library/cc262957.aspx

Table 1.2: Links to Microsoft published deployment guides.

First we begin with the IIS 6.0 deployment guide, which can be found in the Microsoft TechNet Library. This document discusses the minimum hard drive space necessary to install the Windows Server 2003 OS (1.5GB for x86 systems; SharePoint does not support Itanium systems). However, it does not take into account an IIS 6.0 installation because the security directive of Windows Server 2003 does not install IIS by default. To calculate an accurate minimum hard drive space requirement for an IIS 6.0 server, you will need to add on the space used for a basic installation of IIS 6.0 (26.2MB). You must then also tack on optional IIS 6.0 components, such as .NET Web server extensions (146.9MB), SMTP for SharePoint's outgoing email features (1.2MB), and remote management features such as Web-based administration (6.1MB). And although most of the SharePoint content delivered by IIS will be stored in SQL Server databases, there are still a few flat files that will be located in the Web sites' virtual directories that you should also account for.

Next on our list is SQL Server 2005. Unfortunately, there is not one all-inclusive deployment guide document to assist in SQL Server installation like there is for IIS 6.0. In fact, the Internet is rife with installation specifications, upgrade recommendations, and installation tips all published under the Microsoft moniker. Merely assembling all the documents pertaining to SQL Server 2005 installation would take days. All the resources agree that the minimum hard disk space requirement for a single instance of the SQL Server 2005 Database Engine is 280MB. If you plan to employ SQL Server 2005 Reporting Services and incorporate the reports into SharePoint using the Microsoft SSRS Add-On for SharePoint, your SQL Server will need at least another 120MB of free disk space. However, the Database Engine minimum only accommodates system databases and does not account for any SharePoint databases (considered *user-defined* by SQL Server though they are generated by SharePoint processes and not human users directly).

When anticipating the SharePoint database sizes, Microsoft planning guides offer a formula for calculating the space needed by the content databases that serve SharePoint Web applications. The formula will be discussed further into this section. However, the

configuration and administration databases are not discussed in those SharePoint guides, perhaps because Microsoft best practices discourage customizing them. To provide a frame of reference, a default installation of Microsoft Office SharePoint Server 2007 (MOSS 2007) will build the following configuration and administration databases; be sure to keep their sizes in mind when calculating space requirements for your SharePoint SQL Server:

- Admin_Content
- SSP
- SSP Search
- Configuration
- SSO

} 10MB initial size each (based on default Model DB)

Microsoft SharePoint Deployment Guides endorse the idea of planning storage based on historical user behavior and data growth patterns. Yet this philosophy would seem to serve better the design of a rigid hierarchically mandated system because it would likely be based on antiquated vertical solutions such as file servers or traditional intranet Web servers. Use caution when planning storage space for content databases in SharePoint using Microsoft's formula published in the deployment guides as follows:

```
Z + (Z/3) + 600MB = Minimum SharePoint Content Database Size
where
Z = Nx2 (recommended safety factor)
where
N = number of documents x number of versions kept x average document size
```

Although the Microsoft formula offers a valid starting point, it is unfortunately based on known constants that will likely prove inadequate. The document size of a file stored on an NTFS file system is not equal to the space required for hosting said document in a SharePoint content database. Why? SharePoint applies more metadata to a document than the file attributes maintained by NTFS (thus the safety factor in the formula). Furthermore, the formula employs an *average document size* that can be difficult to calculate accurately. Lastly, the formula does not account for upgrade considerations. When upgrading from SharePoint Portal Server 2003 into MOSS 2007 using the gradual method, the transaction log files of the content database can grow exponentially. The safety factor in the aforementioned formula should be increased, perhaps even doubled, to account for these limitations.

SharePoint as an Organic System

The uniqueness of SharePoint's storage requirements lies in the collaborative nature of the platform. Unlike most network data entry applications that tend to be more mechanistic, SharePoint fosters user creativity to increase productivity and innovation. By definition, it is designed to be an organic system of horizontal management responsibilities with community-defined standards and concerns. In layman's terms, organic systems are meant to be built, expanded, and governed by their users with occasional advice from management or IT. SharePoint can be initially constructed and forever manipulated in just about any way, shape, or form that a user base can dream up thanks to extensible programming support. How can one expect to create a perfect storage plan for that?!

Microsoft defines a three-tier administration model for SharePoint as follows:

- **Tier 1**: Farm-level administration—Managing and monitoring entire farm, creating Web applications and site collections, assigning server roles to extend farm topology
- **Tier 2**: Shared Service administration—Managing search, index, usage reporting, business data and business intelligence features for the SSP
- **Tier 3**: Site administration—Web part, content, and access security management for a single site or site hierarchy

In the administration model, tiers one and two are usually performed by IT personnel. However, Tier-3 duties are often delegated to those who know the purpose of the SharePoint content best—the business users. In fact, though site collection creation is considered a Tier-1 responsibility, MOSS 2007 actually offers a method of allowing business users to create new site collections through Self-Service Site Creation. Regardless of how new sites are organized, the addition of sites and their content will increase the storage requirements of the SharePoint content database servicing the hosting Web application and will do so at the whim of the user.

As an organic system, SharePoint offers a very easy, user-friendly method of producing content into a SharePoint Web site. Web parts add functionality to a Web page, while lists and libraries generate additional Web pages and potential for input data to be stored in SQL Server. Customizing the look and content of a SharePoint site can be accomplished through a client browser application (Microsoft recommends the Level 1 browsers IE6 or IE7 for a full administrative experience). Additionally, custom programming in the form of user-defined Web parts and workflows can significantly change the space needs of a SharePoint Web application. Customization can add quickly to the size of the content database in SQL Server and should be accounted for when selecting a storage system for SharePoint.

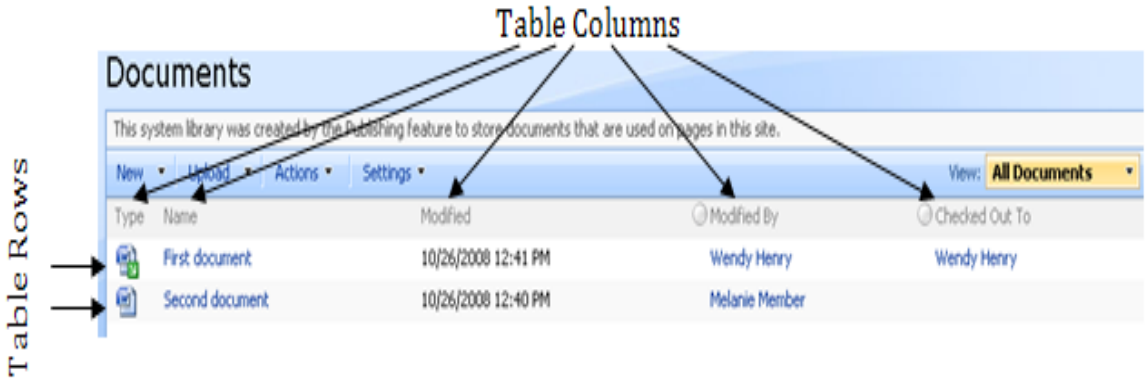

Figure 1.4: SP library to SQL table correlation.

A common misconception about the construction of SharePoint lists and libraries (see Figure 1.4) is that they enjoy a 1:1 relationship with a given table in the SQL Server content database. This inaccurate assumption will sometimes breed complacency when identifying storage requirements because engineers assume that a single list can be easily partitioned in the future using SQL Server 2005 partitioned tables to accommodate growth. Unfortunately, the metadata associated with each list or library item is maintained separately from the item body in the content database (a document library item would have a *body* consisting of the text paragraphs and assorted graphics within the document file itself). There is no single table in the content database to partition when attempting to accommodate growing numbers of list or library items.

Another weak area of analysis during storage planning is the multiple galleries that SharePoint uses to provide redistributable content across many sites. There are template galleries, site column galleries, content type galleries, and more. These galleries ship full of the system-supplied objects and can be expanded to include user-defined items as well. If a high degree of customization will be taking place in your SharePoint environment, these galleries can become quite large. Planning adequate space for galleries is imperative.

One of the selling points of the latest WSS version, 3.0, is the incorporation of workflows courtesy of the .NET Framework Windows Workflow Foundation. Workflows are defined assemblies of code that operate based on manual initiation or automatic initiation resulting from activity on the SharePoint site. The vast purposes that workflows can address make them an attractive solution. However, the more workflows created and their potential complexity can introduce storage concerns, especially if they are archiving large amounts of data. Keep in mind that SharePoint is not appropriate for all types of data; some large files or rarely accessed archived files are better stored on a file server outside of SharePoint. But even these out-of-system files must be accommodated by your storage system plans.

Flexibility Is Key

The point is this: there is no perfect plan for SharePoint storage that can be defined at deployment and hold true throughout the life cycle of a SharePoint environment. Human nature and end user contributions are going to influence modifications and expansions to the SharePoint facilities over time. OS processes and physical hard drive limitations will induce careful selection of appropriate media. Choosing a storage system that is flexible and can easily correct diminishing returns is the key to designing a stable SharePoint enterprise. Because SharePoint is so conditional in its purpose, finding Microsoft white papers outlining exact storage minimums and deployment designs is quite difficult if not impossible. As the manufacturer of the product, Microsoft is understandably hesitant to endorse a specific storage plan when the different manners in which SharePoint can be built and utilized are so diverse.

Flexible storage solutions are not all the same. Some external storage systems require downtime of the host server to be upgraded or to redistribute the volumes. Others may demand significant redefinition of the host server in order to recognize a new drive or volume. Internal storage, even hot swap drives, may necessitate an OS reboot or server hardware power off to accommodate changes or upgrades. The more you can limit the demands of storage maintenance by selecting a forgiving storage product, the less hassle you will experience when frequently growing your SharePoint space.

Planning for Performance as Well as Space

Selecting storage for a SharePoint enterprise should be based not only on sheer volume size but also on the solution's ability to serve critical data to business decision makers efficiently and reliably. Because SharePoint uses SQL Server to store all content, the storage solution employed must be optimized for both applications as much as possible. SQL Server Enterprise Edition benefits greatly from clustered services running off of a shared storage facility. As such, so will a scaled SharePoint enterprise where the database storage services have been offloaded from the Web front end and indexing services.

But performance is not measured strictly by data transfer rates and disk access speeds. Overall, end user satisfaction depends on both quick retrieval of information and reliable access to mission-critical data. In the aforementioned SQL Server cluster example, the fault tolerance of cluster nodes ensure reliable request handling but a single shared storage facility hosting the actual SharePoint database files can easily become a single point of failure. Depending on the SLA between IT and the users, the downtime associated with restoring failed database files or expanding storage volumes may be unacceptable. Designing a SharePoint storage solution must balance both speed and flexibility to ensure continued fast access to business information.

Microsoft Performance Benchmarks for SharePoint

SharePoint performance can be measured by more than physical disk operation. Network latency, storage software delays, and storage architecture congestion can impact overall performance of the SharePoint enterprise. Microsoft offers many tools for monitoring SharePoint behavior and calculating realistic benchmarks to identify performance problems. Selecting a capable storage system for SharePoint can alleviate some of the factors that may contribute to poor performance.

Before you can monitor SharePoint for storage performance gaps, there must be a SharePoint enterprise to monitor. Many of the tools Microsoft offers will need to be run in a development or test SharePoint environment using realistic workloads if you are in the pre-purchase stage of planning. However, if you already have a production SharePoint environment, these tools can be easily engaged to identify storage replacement or expansion opportunities.

> **Cross-Reference**
>
> Microsoft's SharePoint Team Blog comments on a codeplex utility from the SharePoint Toolbox called CopyTimer that will perform random file transfers and report performance. The tool can provide simulated workload to expose overall performance issues that can then be further researched. It can be downloaded from the following URL: http://www.codeplex.com/sptoolbox/Release/ProjectReleases.aspx?ReleaseId=8366.

Microsoft's new System Center Capacity Planner 2007 offers SharePoint models to help plan for hardware needs based on user information and whether your SharePoint server will be exposed to an external network. The tool will validate scenarios up to 100,000 users and 3TB worth of data. Other Microsoft tools for evaluating SharePoint include Windows Server OS utilities such as System Monitor, Network Monitor, and Event Logs. As with any network application, a high number of dropped packets in Network Monitor or application error messages in Event Logs would indicate room for improvement.

SharePoint as a product relies on its dependency upon SQL Server to provide efficient I/O performance. In single-server SharePoint architecture, SQL Server would be on the same server equipment as the SharePoint application itself. However, in a scaled SharePoint Farm architecture, SQL Server is likely placed onto a dedicated server of its own, isolating database file I/O performance and requiring benchmarks to be calculated on multiple systems. If further scaling of the SharePoint architecture is warranted, indexing services and the full text index itself may be moved onto a separate server, adding to the I/O requirements (see Figure 1.5).

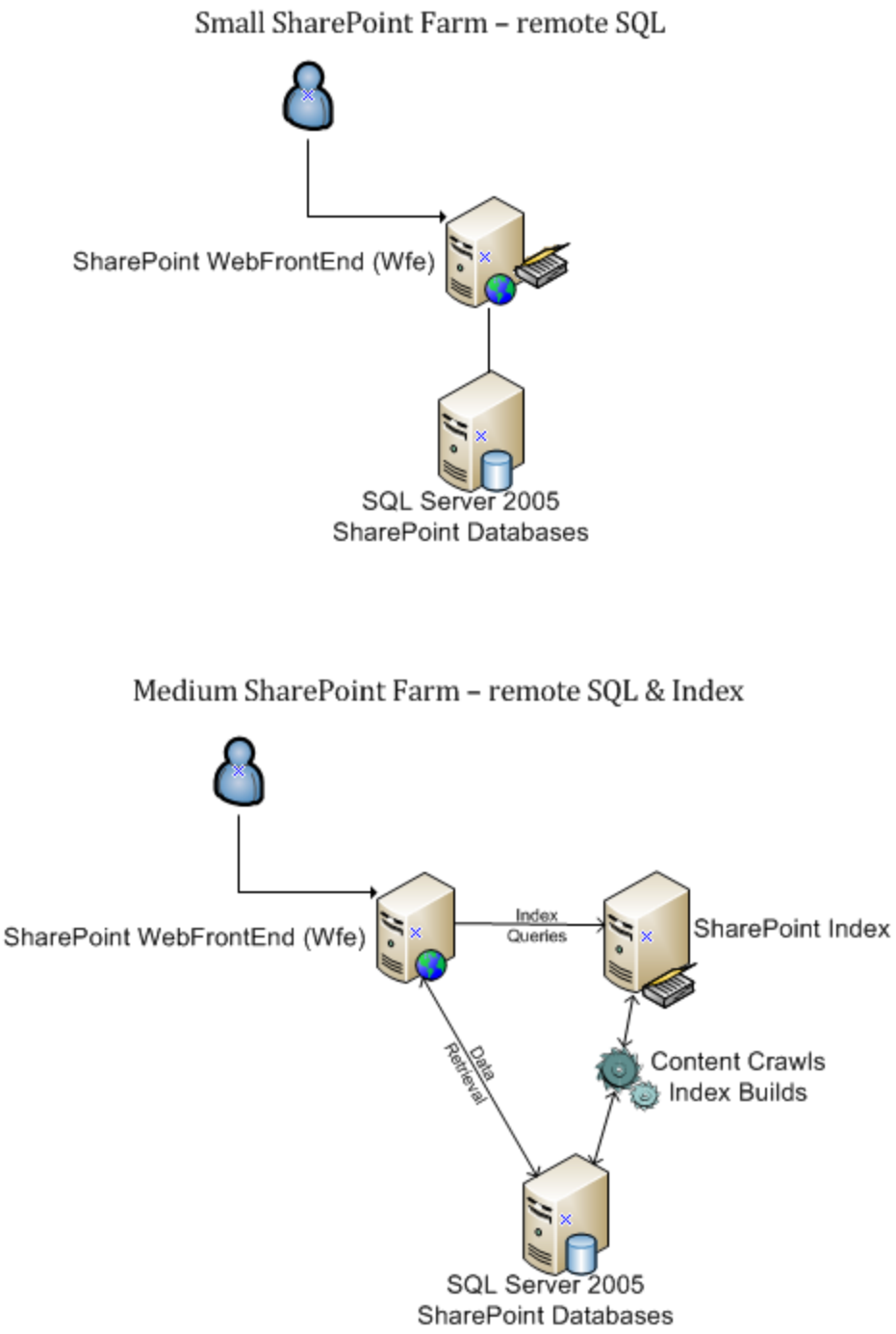

Figure 1.5: Common SharePoint Farm architectures.

Microsoft suggests storage space requirements for SharePoint indexes should begin at 50% of the content database anticipated size and search servers at 25% of same. SharePoint Web front-end servers may also employ Web page caching to improve page delivery and rendering times, which may increase the storage space needs of the Web server. All of these scenarios must factor into your storage solution.

Microsoft Performance Benchmarks for SQL Server

Planning the SQL Server storage facility for a SharePoint deployment is not too unlike planning a SQL Server deployment for any other purpose. Following the SQL Server 2005 minimum requirements and hardware compatibility lists is paramount. However, taking into consideration how critical SQL Server is to SharePoint for configuration as well as content storage and data delivery, you would be wise to pay particular attention to fault tolerance as well as data retrieval performance.

Obviously, volume space is an important consideration when procuring disks for SharePoint's SQL Server. But whether to purchase a single large volume disk or break the total space into multiple disks may not be easily decided. Large volume disks may be tempting but expensive. Additionally, storing all your SharePoint databases on one large disk can cost you lengthy disk seek times. Breaking the total space requirement into multiple, smaller-sized disks can actually improve performance and perhaps introduce fault tolerance (if employing RAID 1 or 5+).

Microsoft SQL Server 2005 best practices strongly encourage the use of multiple hard disks organized into specific RAID arrays when implementing their database engine. Benchmarking RAID performance will depend on whether you are employing software RAID at the hands of the Windows Server OS or hardware RAID from your I/O controller hardware manufacturer. Hardware RAID is preferred due to better throughput and more fault tolerance configurations. However, hardware RAID is relatively more expensive than software RAID. The Windows Server OS offers little in the way of RAID performance monitoring, relying instead on the logical disk object counters in System Monitor (see Table 1.3). Conversely, hardware RAID manufacturers usually offer a bevy of monitoring and diagnostic utilities that will audit performance and proactively correct potential storage issues.

Object	Counter	Description	Preferred Value
Logical Disk	Disk Transfers/sec	Overall throughput	n/a – monitor for decreasing trend
Logical Disk	Avg Disk sec/Read	Time required to read data from disk	<10ms
Logical Disk	Avg Disk Byte/Read	Size of data sets being read from disk	n/a – adds context to the Avg Disk sec/read value
Logical Disk	Avg Disk sec/Write	Time required to write data to disk	<10ms
Logical Disk	Avg Disk Byte/Write	Size of data sets being written to disk	n/a – adds context to the Avg Disk sec/write value
Logical Disk	Current Disk Queue Length	Current number of requests waiting to be served by the disk	Fluctuations below 20
Logical Disk	Avg Disk Reads or Writes /sec (2)	Rate of read or write operations on the disk	<85% disk capacity

Table 1.3: Windows OS disk counters for monitoring RAID volumes.

Microsoft best practices dictate that when employing the logical disk counters Average Disk Reads/Sec and Average Disk Writes/sec to monitor a RAID array, the resultant values of these counters should be tempered by the RAID level used:

- RAID 0: (reads + writes)/number of disks
- RAID 1: (reads + (2 x writes))/2
- RAID 5: (reads +(4 x writes))/number of disks
- RAID 10: (reads + (2 x writes))/number of disks

Note

For assistance determining disk capacity for accurate analysis of the result value on a Logical Disk: Average Disk Reads/Sec object counter or on a Logical Disk: Average Disk Writes/Sec object counter, download the Microsoft SQL I/O Disk Subsystem Benchmark Tool from http://www.microsoft.com/downloads/details.aspx?familyid=9A8B005B-84E4-4F24-8D65-CB53442D9E19&displaylang=en.

One of the challenges when benchmarking a SQL Server's I/O performance is determining which portion of the disk usage is actually related to SQL Server as opposed to the OS and other applications. Microsoft System Monitor includes several SQL Server object counters that will expose logical operations occurring within SQL Server. There is also a dynamic management view available in SQL Server 2005 named *sys.dm_io_virtual_file_stats* that can be queried directly to identify I/O activity and latency issues within the SQL Server data files. All these tools combined will assist you in determining storage improvements to be made for SQL Server 2005.

Despite the storage focus of this guide, it would be remiss not to expand on overall SQL Server 2005 performance beyond the I/O. The two most heavily utilized subsystems by Microsoft's relational database management system are processor and memory. Therefore, when benchmarking SQL Server performance, do not forget to include processor object counters such as % Processor Time, memory object counters such as Available Mbytes, and memory manager object counters such as Memory Grants Pending. The overall performance of SQL Server 2005 is paramount to the success of SharePoint and should be carefully considered from all possible angles.

Optimizing SQL Server for SharePoint

Planning SharePoint storage involves anticipating requirements for two separate Microsoft products: SharePoint and SQL Server. The SQL Server Team at Microsoft has published various deployment guides to assist in planning a relational database management system for virtually any purpose. Consider planning the SQL Server 2005 deployment almost separately from the SharePoint deployment to ensure optimal SharePoint database management.

Optimizing SQL Server 2005 for SharePoint is so intricate, the topic could make up a separate book. In fact, Microsoft has published a downloadable document entitled *Planning and Monitoring SQL Server Storage for SharePoint: Performance Recommendations and Best Practices* that dictates guidelines on everything from physical topology to physical storage recommendations. The guide is available at: http://technet.microsoft.com/en-us/library/cc263261.aspx.

Efficient Storage and Compression

SQL Server 2005 ships with enhanced Common Language Runtime (CLR) support that allows for custom objects such as stored procedures to be compiled outside of the SQL Server processes. These CLR objects can be exploited to compress the data being stored into SQL Server at runtime and, conversely, to decompress the data upon read. However, the high rate of data I/O performed by SharePoint against its databases would negate any storage benefit gained from employing custom compression methods. In fact, most of the data SharePoint stores in SQL Server are in the form of binary large object data files (BLOB) which are not compressible. Compressing SharePoint data in SQL Server would mortally wound I/O performance and reap little storage free space.

SharePoint would be better served by fast transfer, multi-channel storage solutions that can partition the data and prioritize frequently accessed content across many disks. Multiple disk solutions require more management but offer better performance and fault tolerance than attempting to stretch a single volume with data compression.

Summary

This chapter introduces the guide and outlines the capacity requirements and considerations for Microsoft SharePoint. It examines organic data growth and why capacity planning is important when deploying SharePoint. It outlines the diversity of SharePoint storage needs and emphasizes the importance of choosing a flexible storage solution for SharePoint. Microsoft best practice guidelines are explored along with utilities for measuring storage performance and forecasting storage needs.

In the remaining chapters, this guide will dissect the individual components of a SharePoint enterprise and examine their storage requirements. Future chapters will also discuss storage options and best practices for various solutions. Additionally, the concept of fault tolerance introduced in this chapter will be further examined to plan for highly available and recoverable storage implementations.

Download Additional eBooks from Realtime Nexus!

Realtime Nexus—The Digital Library provides world-class expert resources that IT professionals depend on to learn about the newest technologies. If you found this eBook to be informative, we encourage you to download more of our industry-leading technology eBooks and video guides at Realtime Nexus. Please visit http://nexus.realtimepublishers.com.