Realtime publishers

The Shortcut Guide™ To

Storage Considerations for Microsoft SharePoint

sponsored by

Wendy Henry

Copyright Statement

© 2009 Realtime Publishers, Inc. All rights reserved. This site contains materials that have been created, developed, or commissioned by, and published with the permission of, Realtime Publishers, Inc. (the "Materials") and this site and any such Materials are protected by international copyright and trademark laws.

THE MATERIALS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. The Materials are subject to change without notice and do not represent a commitment on the part of Realtime Publishers, Inc or its web site sponsors. In no event shall Realtime Publishers, Inc. or its web site sponsors be held liable for technical or editorial errors or omissions contained in the Materials, including without limitation, for any direct, indirect, incidental, special, exemplary or consequential damages whatsoever resulting from the use of any information contained in the Materials.

The Materials (including but not limited to the text, images, audio, and/or video) may not be copied, reproduced, republished, uploaded, posted, transmitted, or distributed in any way, in whole or in part, except that one copy may be downloaded for your personal, non-commercial use on a single computer. In connection with such use, you may not modify or obscure any copyright or other proprietary notice.

The Materials may contain trademarks, services marks and logos that are the property of third parties. You are not permitted to use these trademarks, services marks or logos without prior written consent of such third parties.

Realtime Publishers and the Realtime Publishers logo are registered in the US Patent & Trademark Office. All other product or service names are the property of their respective owners.

If you have any questions about these terms, or if you would like information about licensing materials from Realtime Publishers, please contact us via e-mail at info@realtimepublishers.com.

[**Editor's Note:** This eBook was downloaded from Realtime Nexus—The Digital Library for IT Professionals. All leading technology eBooks and guides from Realtime Publishers can be found at http://nexus.realtimepublishers.com.]

Chapter 3: Best Practices for Deploying SharePoint on iSCSI

If the previous chapter emphasized nothing else, it stressed the benefits of employing a fast and vast disk architecture for the components of SharePoint. Everything from the OS through IIS to SQL Server make good use of multiple disks, and the scaling of SharePoint across multiple servers demands data transfer via speedy, reliable protocols. Settling on a storage solution for SharePoint and its burgeoning SQL Server databases often has less to do with budget and more to do with long-term investment. Those who underestimate storage needs and purchase minimal storage in an effort to reduce the initial cost of implementing SharePoint invariably pay for their decision in the long run when they have to replace insufficient systems that cannot be expanded.

Of the three storage architectures described in the previous chapter (Internal/DAS, NAS, or SAN), the most popular, and perhaps appropriate, strategy for a SharePoint enterprise is SAN. Though by comparison the SAN is the most complex storage strategy to install and manage, its implementation and administration costs are quickly overshadowed by its extensive availability and recovery options, immense scalability, and autonomous configuration flexibility. But saying you will choose a SAN over Internal/DAS or NAS is like saying you are buying an automobile instead of a skateboard or bicycle…the question becomes: what *kind* of automobile?

There are different SAN solutions, each employing their own transport protocols and medium architecture. In this chapter, we will focus on the Internet Small Computer Systems Interface (iSCSI) mass storage/networking protocol that operates over various mediums. The iSCSI protocol is an affordable, popular SAN implementation that lends itself well to Microsoft Windows networks that rely on TCP/IP for a transport protocol. Offering a wide array of disaster recovery and data availability solutions, most iSCSI SAN providers capitalize on both the reliable storage and speedy delivery data needs of SharePoint.

iSCSI Basics

Before examining the reason for the popularity of iSCSI storage for SharePoint enterprises, we should understand the architecture and its purpose in a SAN platform. For instance, how does the server communicate with the external disks? What does the OS consider the disk: internal or external? If we did not choose iSCSI, what would other alternatives be? How do most commercial implementations of iSCSI actually configure their architecture? Let's take a look at these points before planning an iSCSI storage solution.

What Protocols Are Involved?

The iSCSI set of rules that apply to the correct transmission and receipt of information between computers is actually a combination of multiple protocols working together to define data transfer between a server and an independent SAN device. The iSCSI protocol is positioned between the TCP/IP transport protocol and the SCSI I/O protocol of the actual storage device (see Figure 3.1). Recall that an iSCSI conversation occurs between the software requesting the data on the SAN LUN (initiator) and the SAN system supplying the LUN (target). iSCSI handles the transportation of TCP/IP packets across whatever physical medium is in place (twisted pair copper cable, etc.) for the express purpose of delivery to an iSCSI SAN device. RFC 3720 published in 2004 defines iSCSI and dictates that both the TCP/IP and SCSI protocols on either side need not require any modification to support an RFC-compliant iSCSI interface. In other words, proper iSCSI interoperates with any RFC-compliant SCSI and TCP/IP stack.

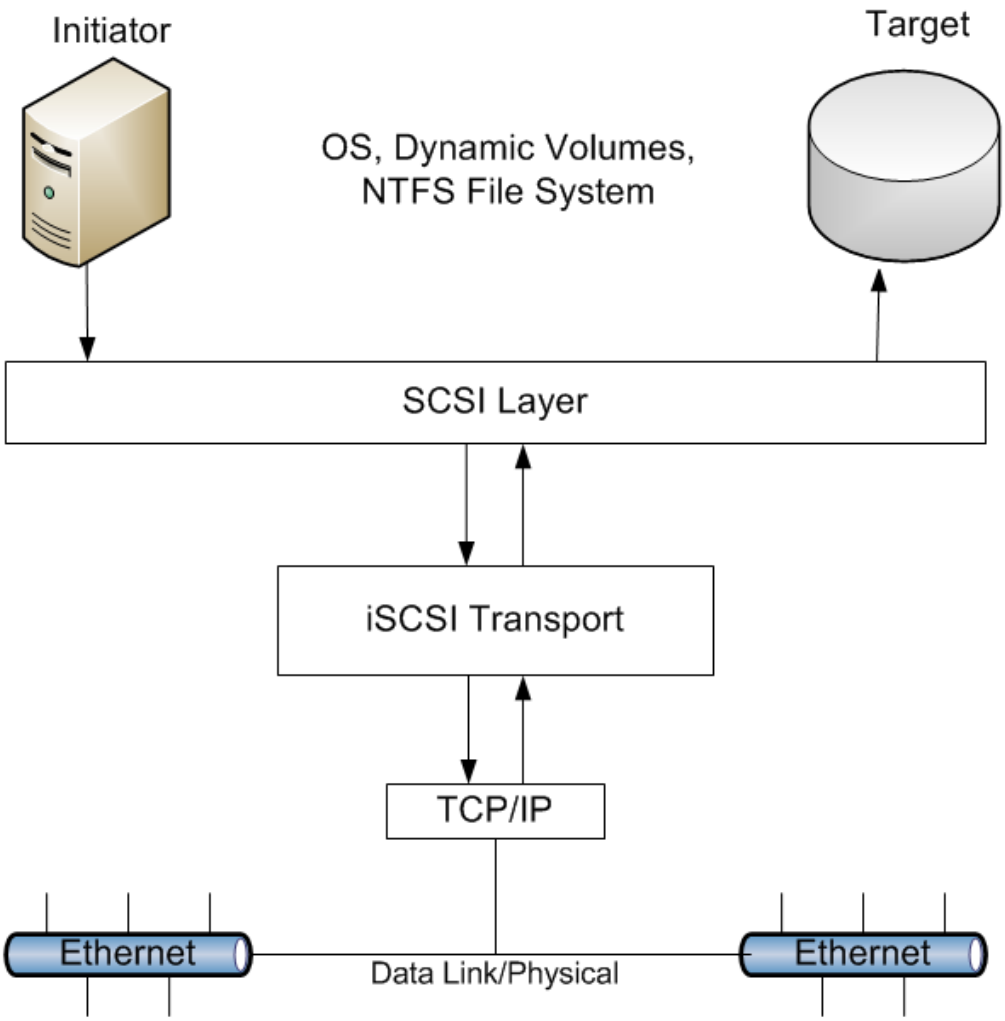

Figure 3.1: iSCSI communication path.

- **SCSI Layer—**[Initiator]=command descriptor blocks (CDBs) are constructed to perform the SCSI disk access requested by upper-level applications. [Target]=the mass storage operation request is extracted from the CDBs received.
- **iSCSI Layer—**[Initiator]=protocol data units (PDUs) are built and include the CDBs from the upper-level SCSI layer along with iSCSI headers that carry session and node information, among other things. [Target]=the iSCSI layer extrapolates the SCSI CDB from the PDU and forwards it to the SCSI layer.
- **TCP/IP Layer—**The TCP/IP layer performs normal packet building and deconstruction of TCP segments and IP datagrams as on any TCP/IP node. There is no direct correlation between iSCSI PDUs and TCP segments. A single PDU may traverse multiple TCP segments and a single TCP segment may contain less than one or multiple PDUs (see Figure 3.2).
- **Ethernet Layer—**The DataLink and Physical layers handle frame encapsulation and extrapolation just like any other network node.

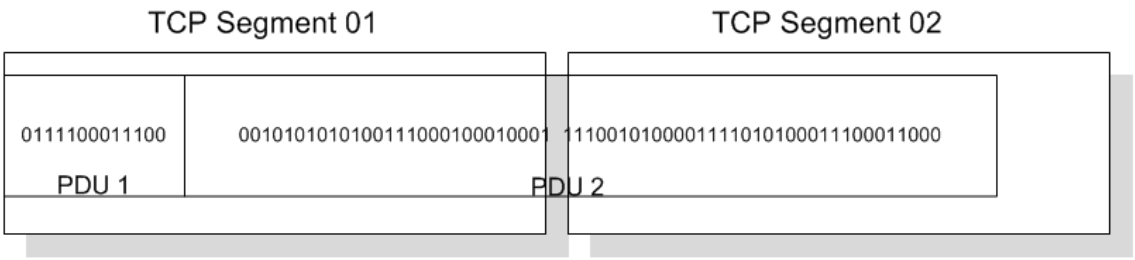

Figure 3.2: Abstract of PDU to TCP Segment layout.

Warning

TCP/IP protocol analyzers or "packet sniffers" are not all created equal. When obtaining an application for analyzing iSCSI SAN traffic, be sure to select a product that supports PDU dissection. Remember, though, there are many freeware and shareware protocol analyzers out there; sometimes it's worth getting what you paid for.

The entire iSCSI communication process integrates mass storage (SCSI) with networking (TCP/IP and Ethernet) using iSCSI encapsulation as the liaison between the two technologies. Because iSCSI operates independently of the transport protocol TCP/IP, any protocol processing hardware useful in a TCP/IP network is just as beneficial on an iSCSI platform. Devices such as offload engines and host bus adapters (HBAs) can be employed on the initiator and target side of the conversation to speed data transportation. And because the iSCSI PDU is enveloped well inside of the TCP/IP packet information, iSCSI packets do not impact other TCP/IP activity on the network. Lastly, on a Windows Server, the OS interprets the storage media as a local SCSI disk and need not be aware that the data will actually be traversing a network! Keep in mind that iSCSI SAN target LUNs can be physical disks, removable media devices (tape drives, CD/DVD jukeboxes), or any other physical storage device managed by the SAN system.

Why to Choose iSCSI for SharePoint

Of the two current fabrics for SAN connectivity, iSCSI is the relative newcomer to a game that has been dominated by the more mature Fibre Channel (FC) technology for more than a decade. Unlike iSCSI, which was sanctioned by the Internet Engineering Task Force (IETF) in 2003, FC architecture uses the Fibre Channel Protocol (FCP), a serial bus protocol, to structure data for transport and is designed to move massive amounts of data on a frequent basis. Boasting ANSI standards that date back to 1994 and multiple RFCs, the first of which published in 2000, FC is still a valid and stable communication technology for SAN devices. But is it the best choice for SharePoint?

Note

There are actually two protocols for transporting Fibre Channel over IP: FCIP and iFCP. FCIP uses tunneling while iFCP uses routing to encapsulate FC frames over an IP network through gateways.

The differences between FCIP and iSCSI make iSCSI the more appropriate candidate for SharePoint. Importantly, iSCSI eliminates the need for specialized hardware (see Figure 3.3) relying instead on existing IP infrastructure and human resource expertise. In terms of packet protocol layering, iSCSI requires less overhead than FCIP to transmit original SCSI I/O requests inside IP packets. iSCSI is the ideal SAN connectivity fabric to implement for your SharePoint enterprise because it offers:

- More security protocols
- Lower implementation costs/use of existing hardware
- Lower administration costs
- More disaster recovery options
- Comparable performance to FC

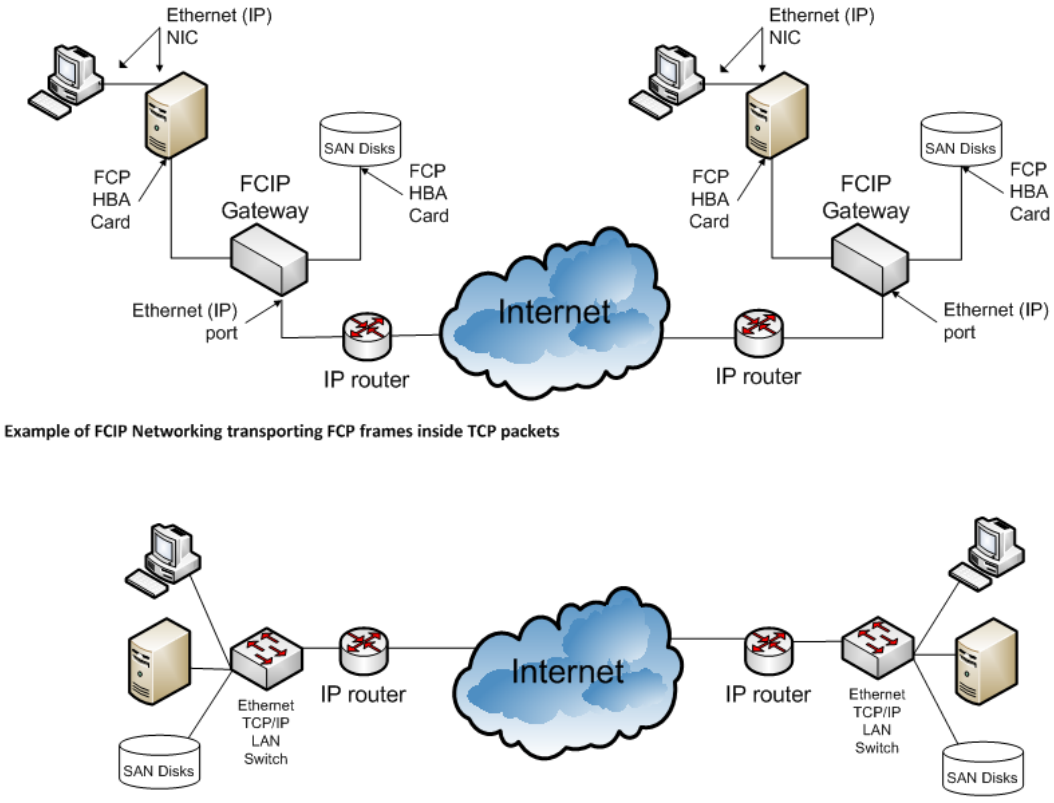

Figure 3.3: Comparison of typical network topologies for FCIP vs. iSCSI.

Because iSCSI uses familiar Ethernet devices and offers user-friendly administration utilities, the cost of skills needed to implement and maintain an iSCSI SAN are less than most FCIP implementations. By using TCP/IP security mechanisms, data is no less secure on iSCSI fabric than it would be on a mature FCP fabric. Considering that Windows SharePoint Services 3.0 for Windows Server 2003/2008 can be downloaded and installed for free, the total cost of purchasing and maintaining a reliable SharePoint enterprise on iSCSI SAN storage can now be afforded by both small and large companies alike.

Common Implementations of iSCSI

In recent years, many myths about iSCSI being less secure or unable to handle large data loads have been debunked in large enterprises. SharePoint investors prefer the automated workflows, collaboration tools, and management capabilities they can employ when placing their mission-critical data in WSSv3 but are not willing to risk inconsistent data delivery. SharePoint takes advantage of the flexibility, scalability, and reliability of iSCSI SAN storage to optimize data delivery performance and the availability of important company information. But exactly how are iSCSI storage solutions being implemented to best serve WSS?

First and foremost, the most storage-dependent component of a SharePoint enterprise is the SQL Server database system. Therefore, most iSCSI SAN investments are initially procured for the SQL Server of a SharePoint farm. In the event of budget constraints, SQL may be the only component of the SharePoint farm residing on the SAN. But another likely candidate for SAN storage in a large SharePoint farm would be a dedicated SharePoint index server for the purpose of housing the large full-text index catalog. Dedicated index servers are a rare topology usually reserved for very large SharePoint environments and not commonly seen in small or medium-sized businesses that store less data and have limited budgets. Lastly, the Web front-end servers of a SharePoint farm can also use bootable SAN LUNs for their OS or just for the Web site virtual directories if space is an issue. A multi-server Web farm can be installed on a cluster solution that employs the SAN as its shared storage component. It is even possible to have all the SharePoint farm servers supporting each of the three farm roles utilizing separate LUNs on a single SAN system (see Figure 3.4) to get the most out of your storage dollar.

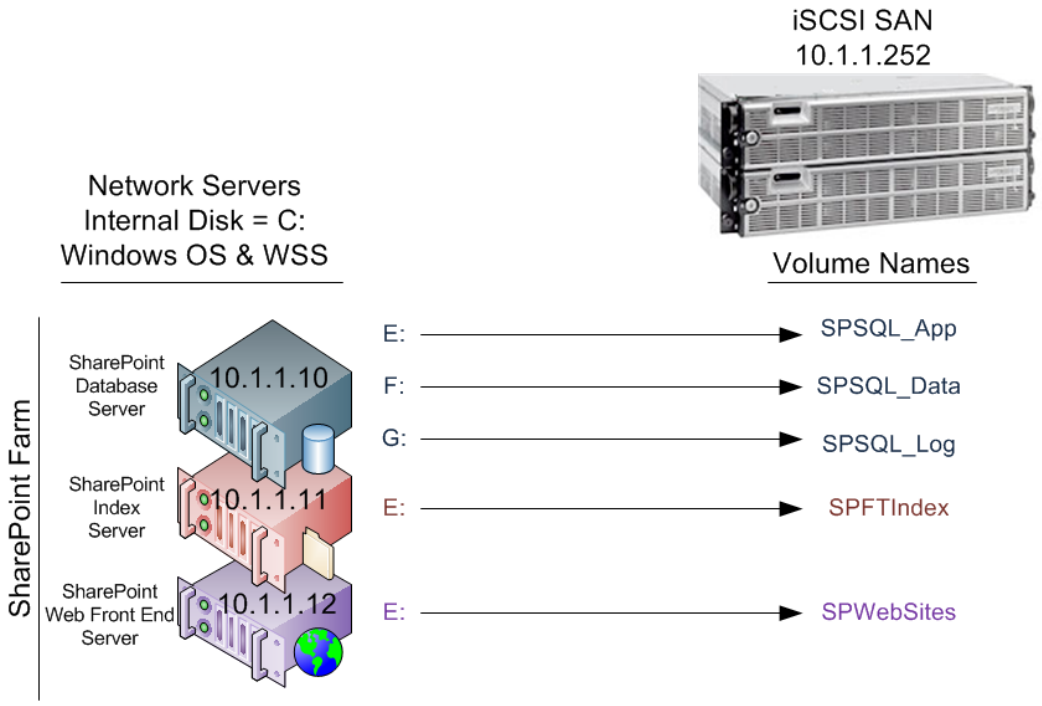

Figure 3.4 Example of multiple WSS servers using a single SAN

Another strategy gaining favor in the IT industry is the virtualization of SharePoint and SQL Server. Virtualization is the art of creating multiple, autonomous software systems on a single piece of hardware to take full advantage of your hardware investment (see Figure 3.5). Purchasing a huge server but only loading a small OS and one network application on it may be a waste of potential resources. By implementing a virtual server application, multiple OSs can make use of the robust hardware simultaneously, thereby stretching your hardware budget. Installing WSSv3 onto a "virtual" Windows Server OS uses the same setup program and in fact can sometimes perform better than a local host OS installation. If the virtual OS resides on an iSCSI SAN LUN, the WSS instance gains all the performance and reliability benefits of the SAN alongside other mission-critical virtual servers, making the most of your SAN investment.

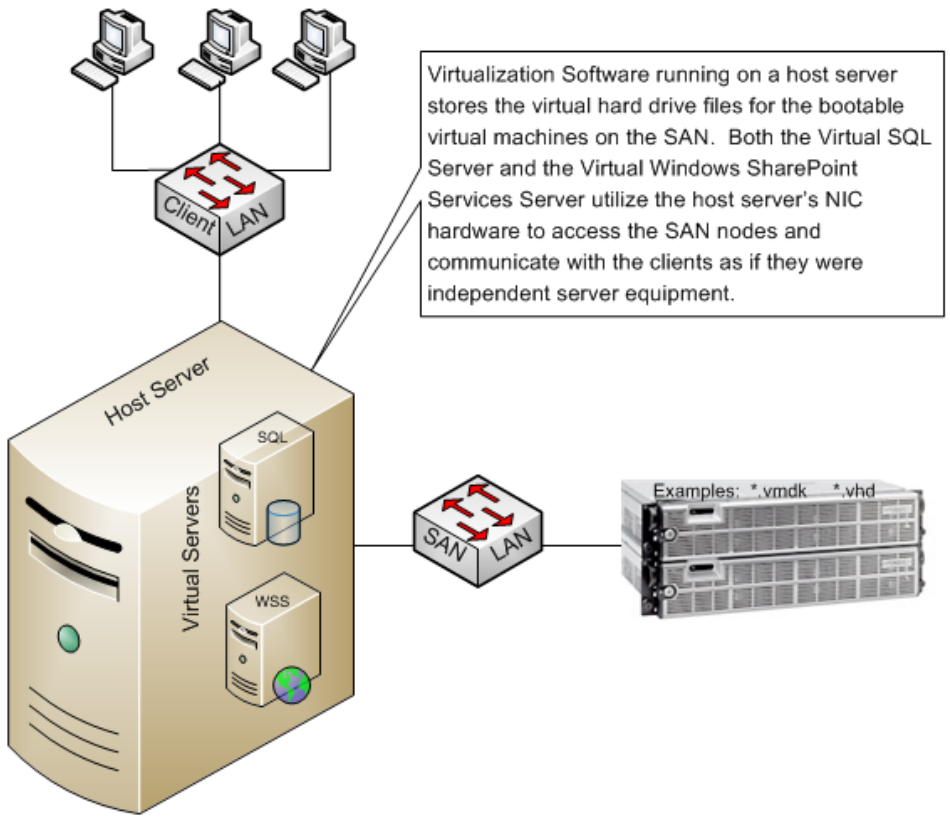

Figure 3.5: Typical virtualized WSS and SQL Server on SAN.

Flexibility by Abstraction

When it comes to SharePoint storage, the key trait is *flexibility*. Remember that SharePoint, as an organic system, will need to grow and shrink according to use. Being able to expand SharePoint's storage space without wasting hardware investment is paramount to retaining data at a reasonable cost. Luckily, iSCSI SAN systems offer a wide range of adaptable configurations to most effectively support a unique SharePoint environment. Though planning the SAN is always a good idea, having tools available to quickly extend disk functionality and space will save your SharePoint data if even the best of plans are laid asunder.

One thing a well-built iSCSI SAN should do is isolate storage configuration and management from the servers that use the SAN. By abstracting actual disk management from the Windows Server OS running on a SharePoint or SQL Server, the SAN is free to better manipulate efficient data placement and take care of its own high-availability and disaster recovery configurations without requiring any special settings on the SAN clients (namely the Windows servers). With such abstraction comes opportunities for duplicate efforts, such as backing up data. The sophistication level of the tools offered by your SAN manufacturer will determine whether you should perform a data administration task on the network server or the SAN.

Planning Expansion

Any discussion about SAN disk management should focus on the organization of the disks. Different SAN manufacturers use various names for their disk allocations, such as *groups*, *collections*, *LUNs*, *volumes,* or *containers*, but the bottom line is: when you are planning expansion of a SAN, you should plan allocations rather than impulsively purchase more disks for the array. To begin, determine whether there is unused space on existing disks. If so, it may be possible to statically construct additional LUNs for iSCSI initiator connections by merely reconfiguring the disk allocation containers. Beware that not all SAN systems allow online configuration changes and some may even require reconfiguration of the disk array itself. Be sure to purchase a SAN that provides software gracious enough to mitigate service interruption during expansion. You might need to extend your allocations often for WSS!

A welcome alternative to static LUN management on an iSCSI SAN is storage virtualization. Most iSCSI SAN software is capable of abstracting the logical allocation space from the physical disk by maintaining metadata about such *virtualized storage* in a mapping table. SharePoint's requests for storage space can be dynamically mapped to a specific virtual disk (*vdisk*) from a bevy of available physical locations. The data is then placed onto the corresponding vdisk via I/O redirection while SharePoint is completely oblivious. In this manner, multiple heterogeneous physical storage systems can be pooled to optimize disk usage while providing transparent allocation growth or shrinkage operations. Similarly, data migration and replication activities will not disrupt SharePoint's access to the logical storage units, so they can be conducted at the administrator's discretion without negative impact to users.

There are two possible architectures for providing storage virtualization. As depicted in Figure 3.6, in-band devices are placed directly between the SharePoint server and the SAN system along the data path. The SharePoint server targets the storage virtualization device rather than the SAN itself and the device maps the I/O request to the appropriate underlying storage solution. This is known as symmetric virtualization in that the SharePoint data flows straight through the virtualization device en route to the SAN system or back. Symmetric virtualization can be accomplished by employing dedicated storage virtualization appliances or by exploiting sophisticated physical switch hardware. By handling the data, a symmetric appliance or switch provides an opportunity for data caching to improve delivery performance according to accepted latency margins.

The alternative to symmetric virtualization is to implement a storage virtualization metadata server as an out of band auxiliary. Labeled asymmetric virtualization, in this design, a separate host maintains the storage virtualization mappings. SharePoint's I/O requests on the SharePoint server are interrupted at a lower level in the OS while a lookup query is sent to the metadata server to determine which physical target should be used. Upon receiving a response, the SharePoint I/O request is directed to the physical location dictated by the metadata server. The data therefore flows between the SharePoint server and the physical storage only, never passing through the metadata server itself. Asymmetric virtualization offers no opportunity for data caching because the metadata server never actually handles the SharePoint data.

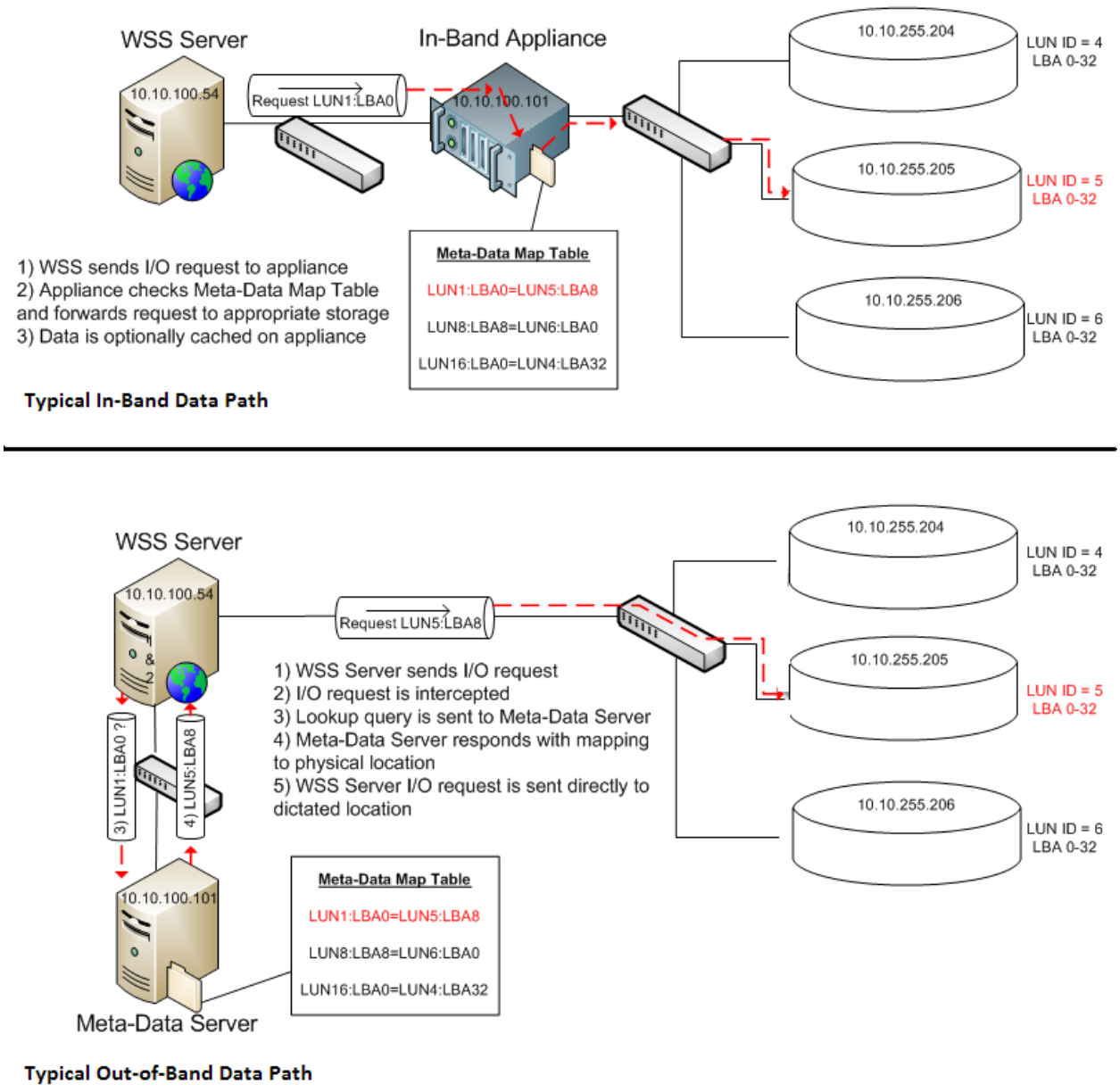

Figure 3.6: Comparison of in-band (symmetric) vs. out-of-band (asymmetric) storage virtualization data paths.

Statically Redefining Storage

Realizing you need more storage space from your SAN does not, necessarily, mean you must immediately break out the checkbook. In fact, many manufacturers today including DELL, HP, and others are producing SAN software that allows construction of logical entities without taking the SAN offline. Knowing your structure will help you determine whether additional physical disks are truly necessary. A generally accepted rule of 20% free space per disk gives opportunity for proactive disk purchases in response to unforeseen data growth spurts.

For example, say you own a SAN with 10 disks of 500GB each. I know, the size is small by today's standards but the math is simple, so work with me here. Regardless of the array, it is possible to distinguish areas of space independently from the disk the space is coming from. For instance, say you determine that your network servers are going to require the following areas of space on the SAN:

- WSS web front-end/indexing/query server needs 500GB
- SQL Server supporting WSS needs 2000GB
- Exchange Server needs 2000GB

You could construct a single logical container representing the entire available space on all 10 disks, then generate three LUNs of 500GB, 2000GB, and 2000GB, respectively, leaving 500GB of the single container free.

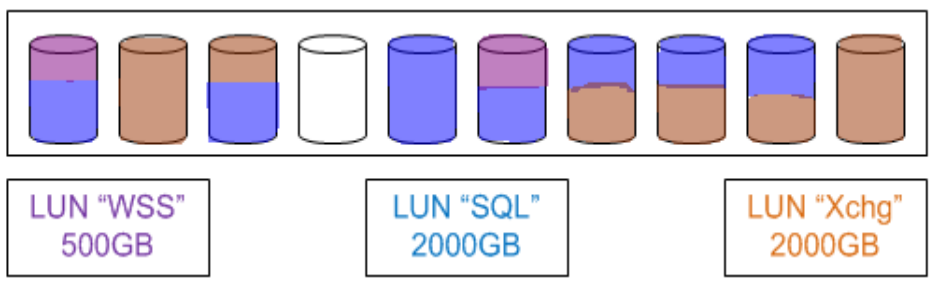

Figure 3.7: 10 disk array in single container with 3 distributed LUNs.

This may work fine if you don't care which disk each network server is placing its data onto. However, for the sake of argument, let's say you do care. Now a single logical group containing all 10 disks isn't appropriate.

If you create separate logical containers, each representing only the space required by each of the network servers, you may be painting yourself into the proverbial corner as well. Let's say your first container named WSS groups only the first 500GB disk from the array bearing a single LUN of 500GB. If the SharePoint server does not end up filling in all 500GB, you may have wasted allocated space in the container. However, an even greater dilemma will be when SharePoint grows beyond its current 500GB need.

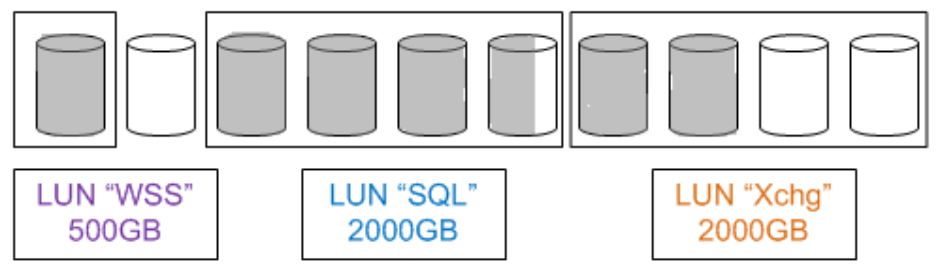

Figure 3.8: 10 disk array in 3 containers with 3 distributed LUNs.

Figure 3.8 uses gray shading to represent data stored and you can see the new multiple container structure. And although the WSS container's disk has run out of available space, there is 1000GB free in the Xchg container and even an entire 500GB empty disk currently not configured in any container. Normally, the LUNs would not map directly to each container but would rather be assigned to different nodes. Here, they are drawn as if related simply to illustrate data density.

Now you have to decide: do you create a container for the last remaining unutilized disk and create a new LUN for WSS or do you redefine the Xchg container to release the unused disk space thereby making it available for another container? Your decision may hinge on how accommodating your SAN software is to reconfiguration. If releasing the two disks from the Xchg container will cause service interruption to the Exchange Server or worse, destroy the existing Exchange data, then it wouldn't be worth it. But if your SAN software offers flexible online reconfiguration of containers and/or LUNs, it might be worth doing.

For a SQL Server supporting SharePoint, plan at least one logical allocation pair for each WSS_Content database that you will be generating. Each content database should have access to two areas so that the log files can be separated from the data files. For the configuration databases, feel free to plan only one individual area for each configuration database's data files and an additional "community" area for all the configuration databases' log files (since configuration databases aren't heavily written to). In a default MOSS2007 environment, a minimum of nine areas would support the default databases. Separating SharePoint databases across containers gives you more data availability options such as granular snapshots and efficient replication strategies. When sizing SAN volumes for SharePoint, be aware of backup and restore durations to avoid violating SLA downtime allowances.

Reducing Allocation Footprints with Thin Provisioning

One of the most aggravating results to see in any network monitoring is nonuse of valuable resources, such as SAN disks. After budgeting, planning, and implementing a robust storage system, it is disheartening to watch it go to waste. Imagine that you originally create a volume on the SAN by allocating the largest space you anticipate your SharePoint SQL Server ever needing: 200TB. You did so to avoid expanding the SAN volume every time your SQL Server outgrows its current data footprint. Ah, but if you have misjudged too high, it could be a costly waste of space. What if the SQL Server never grows larger than 75TB?

There are two possible remedies to over-allocating your storage space, and both are variations on a theme. First, you could initially allocate a small footprint for your volume and use the SAN software volume expansion feature to manually grow the volume footprint as needed. In fact, you might even be able to enhance this solution by introducing automation offered in the volume expansion utility's scheduling tools. If your SAN software does not support expanding existing volumes, you will be forced to create a new volume and migrate the data to it. This operation can be very time-consuming. On-demand volume expansion pales in comparison to the second remedy.

The second antidote to over-allocation takes volume growth automation to the extreme, in a good way. It is called *thin provisioning* and it gives you the best of both worlds. Initially, you can allocate the largest space you anticipate your server ever needing without the worry that some of that footprint might remain empty. Thanks to sophisticated SAN software, only the volume space in use by the iSCSI initiator is reserved. As the Windows server requires more of its already promised footprint, the SAN will expand the reservation to accommodate. Meanwhile, any unused space in the allocation is, in effect, available to be offered out to another reservation.

Many challengers of thin provisioning are blogging that it is akin to gambling with your storage space. Essentially these diatribes will claim that shuffling storage free space around could eventually lead to insufficient space available to fulfill all promises and the SAN will crash if all the clients suddenly want to use all their promised space. Although this may be a possible, yet highly improbable outcome, it is certainly not inevitable. Most SAN manufacturers who offer thin provisioning also offer a bevy of utilities for monitoring reservation space, allowing you to be proactive with additional disk purchases. Use them!

> **Note**
>
> Although thin provisioning is highly recommended for those volumes storing SQL Server database data files, most iSCSI SAN best practices discourage implementing it on volumes storing SQL Server database transaction log files.

By using thin provisioning, you get the most use of your storage hardware while enjoying automated volume expansion that doesn't require human intervention or a great deal of processing overhead. Moreover, the volume expansion is transparent to the initiator, so there is no data delivery interruption. And thanks to the dynamic nature of thin provisioning, SharePoint capacity planning need not be an exact science. SharePoint administrators can feel free to size initial storage requirements including the best practice minimum growth percentage of 50% projected content without unnecessarily dooming disk space to idleness. Does it sound too good to be true? Well, there are a few small things to watch out for. Namely, you might want to set up a notification system so that you can at least be made aware when reservations are grown—not to run panic stricken to the server making sure there is enough available disk space left but rather to collect audit information about how your SAN is being used by its clients. You might also need to keep an eye on which reservations are growing often and by how much so that you can anticipate future disk purchases. Even though thin provisioning setup entails many control settings, including maximum size per reservation, it still encourages open competition among SAN clients for disk space and can inadvertently overpopulate your existing drives. If the SAN runs out of available space and a reservation expansion request arrives from an initiator, things will get bad in a hurry.

> **Note**
>
> Several iSCSI SAN manufacturers offer thin provisioning on their platforms; be sure to choose a platform listed on the Microsoft Windows Server Hardware Compatibility List (HCL) as having passed WHQL testing. Also be aware that to date there are no regulatory standards for thin provisioning, so expect interoperability challenges if employing storage virtualization across heterogeneous storage facilities.

The opposite side of the volume expansion coin is the fact that your servers often legitimately delete data from their LUNs through normal data moving and scrubbing or natural activity on volatile data sets such as SharePoint content. Although thin provisioning is very helpful for growing volumes in response to requests for more space, shrinking volumes in response to requests for less space is not the tool's forte. To shrink the size of a LUN, you must either create a new LUN of smaller size and migrate the data or utilize your SAN software's volume shrink tool. Be sure to monitor LUN density as well as footprint to determine whether the LUN needs to be downsized. But don't get overzealous; remember that a SharePoint content database will grow and shrink repeatedly within a given day due to user activity. Do not configure your SAN to shrink LUNs hosting WSS content databases unnecessarily because doing so often will cause unacceptable overhead. Monitor your SharePoint databases as discussed in the previous chapter before deciding to shrink the supporting LUN.

Compartmentalizing Data Sets via Snapshots

One important reason to choose an iSCSI SAN as the storage solution for a mission-critical SharePoint environment is to protect data availability, and not just from loss. Although disaster recovery of lost data usually tops the priority list, it is just as important to maintain data lineage with the ability to return data values to a past version in the event of corruption, malformed code, or just plain human error. Recovering the previous valid data value in SharePoint can mean the difference between a SharePoint resource being available or unavailable. A side benefit of maintaining data lineage is the ability to query past data values from the data set in its entirety or merely a sub-section. Returning historical data is useful for auditing, troubleshooting, and records management.

Most iSCSI SAN manufacturers offer a firmware utility for creating copies of data as it looks at a specific moment in time and freezing that copy so that it can be queried and even reverted to if necessary. Several terms have been used to describe the point in time data replica, the most common of which is *snapshot*. By creating a snapshot of both the database schema (metadata) and the actual row values from a SharePoint configuration database it is possible to return hierarchy of SharePoint back to a previous state. Taking snapshots of WSS content databases protects user-contributed business data while doing the same for MOSS Shared Service Provider databases shelters user profiles, enterprise search settings, Excel Services configuration, and the Business Data Catalog contents. Determine the amount of data to be taken by a snapshot and the frequency at which the data will be recorded to ensure adequate disk purchases.

Not all SAN snapshot utilities are created equal but all demand extra available disk space. Some require complex configuration to interoperate with the Microsoft Volume Shadow Copy Server (aka Volume Snapshot Service or VSS). Still more may not support your backup application of choice for supporting the backing up of open files. SharePoint SQL Server database files are open 24/7, which may cause inaccurate and incomplete backup files. Consider the following before purchasing your iSCSI SAN solution with VSS support:

- Flexibility—Are the snapshots read-only or read/write?
- Mobility—Can the snapshots be easily mounted onto other servers to create test environments or duplicate the data environment for remote users?
- Efficiency—Do the snapshots require preexisting disk reservations? Can the snapshots take advantage of dynamic disk management features such as thin provisioning?
- Interoperability—Do the snapshots integrate with Windows VSS?
- Simplicity—Are the snapshots easy to create and manage? Are the snapshot generation automation tools and recovery utilities intuitive and easy?
- Limitability—Can a maximum be set on snapshots retained or individual snapshot size?

A well-built VSS system on an iSCSI SAN also enhances virtualized SharePoint servers. If the SAN volume containing the virtual machine can have a snapshot generated and mounted to an alternative server or recovered to an alternative volume, it is possible to essentially generate a second virtual machine of the same configuration as the one that has been snapped. Whether for testing, recovery, or migrating a production virtual machine to different storage, duplicating a virtual machine would be easier and quicker than imaging, cloning, or converting to local OS.

You can employ the Windows OS VSS on internal disk volumes of your SharePoint and SQL Server servers in conjunction with using the hardware VSS supplied by your SAN manufacturer on the SAN LUN volumes. Invoking both Windows VSS and SAN VSS on the SAN LUN volumes is overkill. You should use the SAN VSS on the LUN volumes as it will likely perform better and offer more options. To engage Windows OS VSS on internal disk volumes such as the C drive, confirm that the Volume Shadow Copy service is running in the Windows OS, then invoke snapshots manually or using the scheduler available in the Settings of the Shadow Copy properties of a volume (see Figure 3.9).

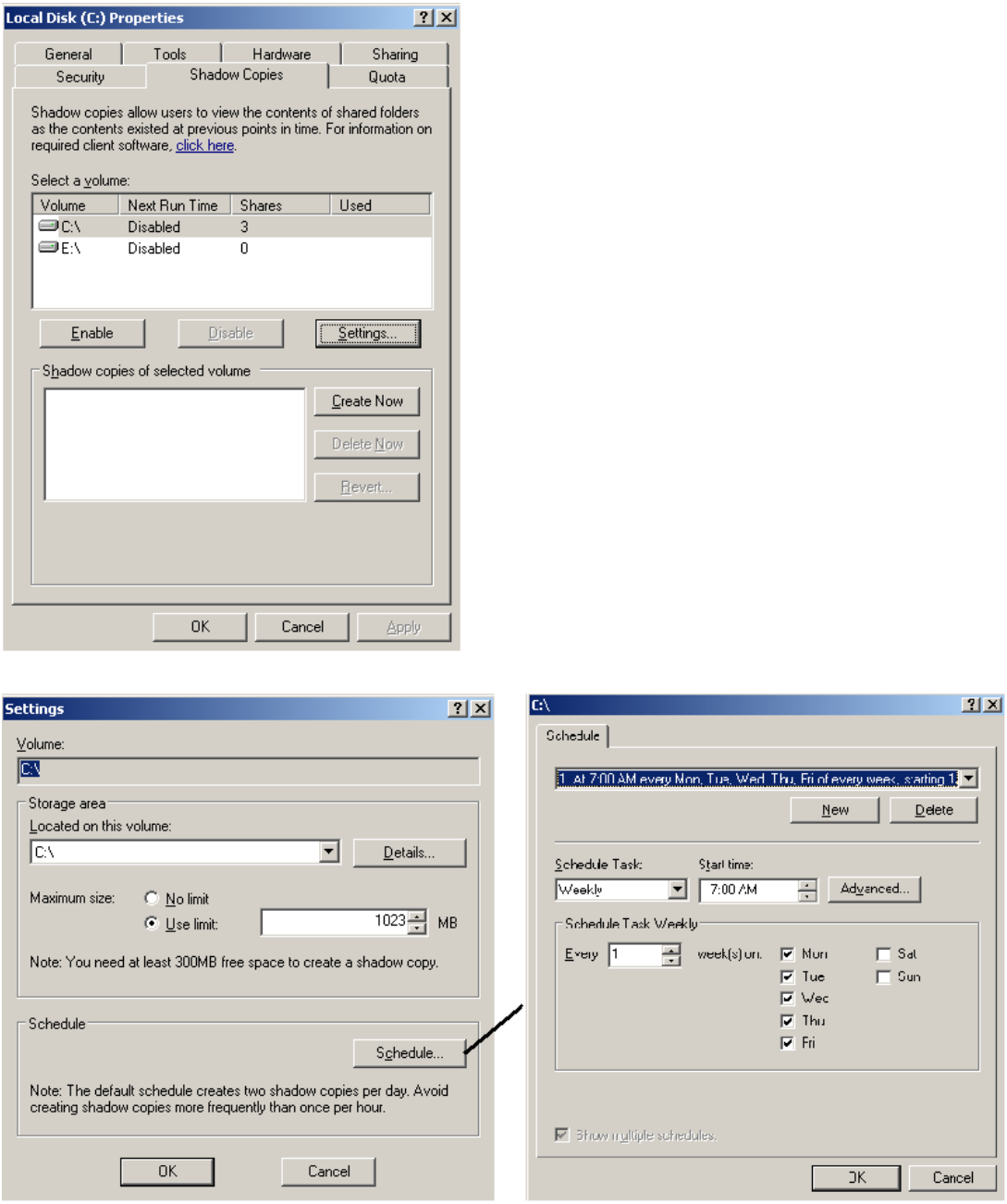

Figure 3.9: Windows Server 2003 volume properties for Shadow Copy settings.

You should save the VSS Shadow Copy snapshots onto a SAN volume to take advantage of faster write performance and less contention on the internal I/O subsystem. And don't be concerned that a snapshot taken in the middle of a transaction being processed within your SharePoint SQL database may leave the historical replica in an inconsistent state. Microsoft wrote SQL Server with VSS-aware plug-ins to accommodate the timing of snapshots and dirty data flushes. Transactional consistency is achieved during *quiesce* when data changes written in buffer but not yet committed to disk (dirty data) are written to the disk prior to taking the snapshot.

Snapshots may also come in handy for meeting regulatory compliance standards regarding data archiving or minimum data replicas. In fact, if you purchase an iSCSI SAN that can mount snapshots onto remote storage such as geographically dispersed SAN disks or mechanical tape libraries, producing snapshots may even assist in meeting offsite data storage requirements. From complete backups to data lineage tracking to disaster recovery, SAN-based snapshots are simply the best strategy.

Best Practices for Implementing iSCSI

As with any technology solution, there are many different ways to incorporate iSCSI into a SharePoint network. Following a few best practices from the field can avert many pitfalls that may detrimentally affect data availability and overall network performance. Above all else, documentation is crucial to any SharePoint across iSCSI implementation including keeping records about the storage system. In fact, it is wise to invest in a formal change control process to avoid miscommunication and conflicting configurations of the network storage settings. Additional best practices that can improve performance of an iSCSI fabric include multiple paths, compression, and proactive monitoring. Data confidentiality best practices include encrypting the traffic between the SharePoint server and the SAN storage unit.

Microsoft Multi-Path I/O

Before a Microsoft Windows OS can take advantage of an iSCSI SAN disk, the Microsoft iSCSI Initiator application must be installed and configured. Microsoft Windows Server 2008 includes an iSCSI Initiator, but you will need to download the application for Windows Server 2003. In a network that will employ multiple Microsoft iSCSI initiators that will communicate with each other, it would be wise to also implement the optional Microsoft Internet Storage Name Server (iSNS) software to aid in resolution of iSCSI targets and manage iSNS clients (namely all of the iSCSI Initiator servers). Prior to Windows Server 2008, the Microsoft iSCSI Initiator and iSNS applications had to be downloaded from Microsoft and manually installed. However, Windows Server 2008 contains an iSCSI initiator and Windows Storage Server 2008 contains an iSNS feature right out of the box.

The Microsoft iSCSI Initiator application implement the use of redundant network paths to increase data transfer throughput between the iSCSI initiator and the iSCSI target. Data packets that must travel from the SharePoint server to the SAN disks and back are like cars traveling on a highway. If half of the rush hour commuters could be diverted to an alternate route bound for the same destination, both highways would experience faster traffic flow and all the cars would converge on the same destination sooner. Similarly, network throughput can be increased using multi-path I/O (MPIO), giving data packets more "highways" to travel on and providing fault tolerance via redundant paths to ensure data delivery.

Keep in mind that implementing redundant paths means purchasing duplicate network hardware of same or similar capabilities. Practicing MPIO increases budget requirements but the benefits of separating SAN traffic from LAN traffic are surely worth the cost. To implement MPIO with load-balancing and fault-tolerance benefits, consider employing a minimum of three network interface cards (NICs) in each iSCSI initiator, at least three Ethernet switches, and two or more NICs on the SAN (see Figure 3.10).

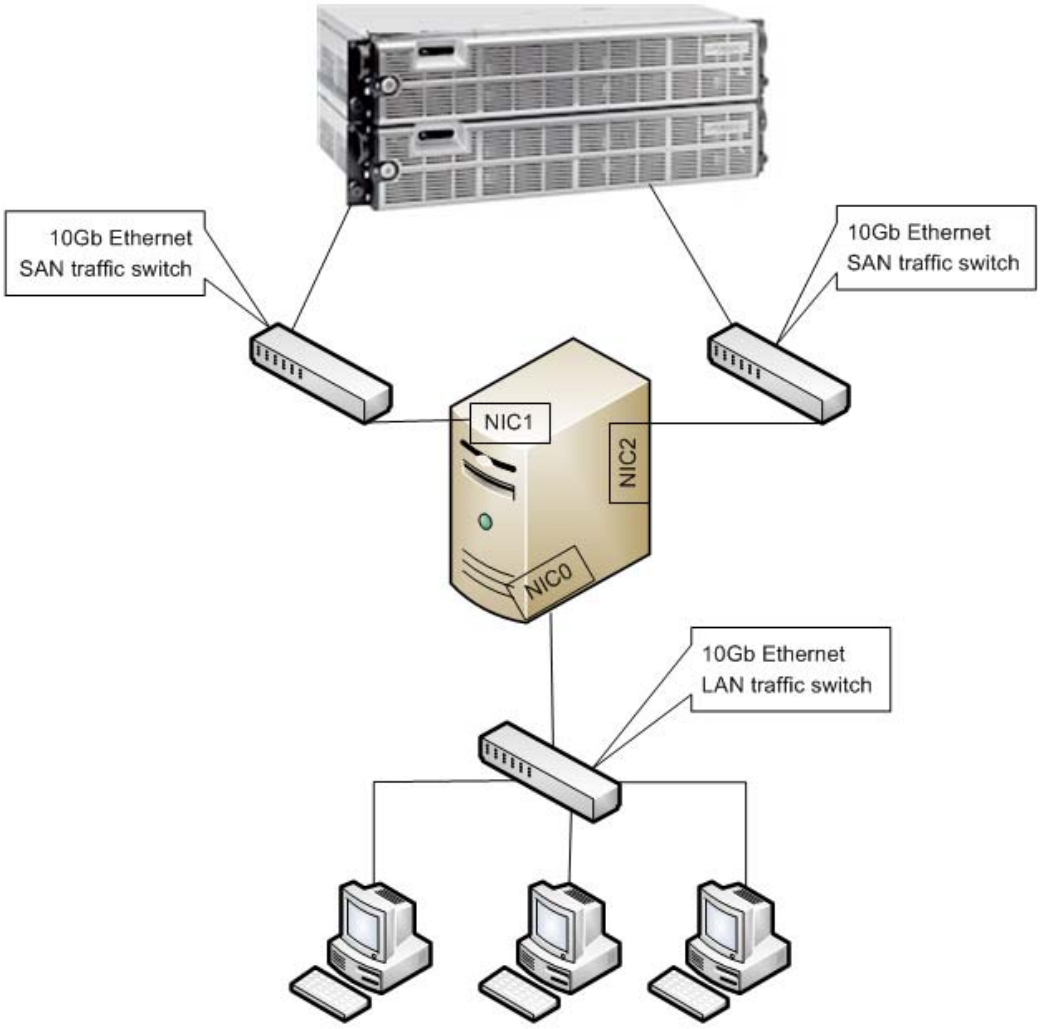

Figure 3.10: Typical MPIO diagram for load balancing.

In a Windows Server OS, the iSCSI Initiator application can be configured to use only a single path for SAN traffic, leaving the alternate path for failover when needed. This choice offers fault tolerance but no load balancing. Conversely, the iSCSI Initiator application can be configured to employ an algorithm to load balance iSCSI traffic across all possible network interfaces. This solution offers load balancing but limited fault tolerance (dropping one NIC causes all traffic to flow across the fewer number of remaining functional interfaces, causing noticeable congestion and decreased data delivery speeds). Lastly, and perhaps the best option, Microsoft iSCSI Initiator can be set to employ an algorithm to load balance across most of the network interfaces yet reserve one or more *standby* paths for fault tolerance. Path selection can also be influenced using weights, queue depth, and path idleness.

Traffic Compression and Encryption

Sensitive data stored into SharePoint may require a safe transport between the iSCSI initiator and target, especially if the SAN is geographically dispersed from the Windows Server and the iSCSI packets will be forced to travel across WAN equipment. Because iSCSI employs a true TCP/IP packet build, industry-standard encryption mechanisms such as IPSec are handy for protecting the PDU and payload of an iSCSI Ethernet frame. But with encryption comes size and potentially excessive bandwidth utilization. To deal with overburdened bandwidth, a compression solution may also be required.

The general rule of thumb is to employ compression first, then encryption. If encryption such as Authentication Header (AH) and Encapsulating Security Payload (ESP) of the IPSec protocol is applied first, the data in the IP datagram is randomized and subsequent attempts to compress the data within the IP datagram will likely corrupt the packet. Depending on your choice of compression solution, following the rule may not be possible. If, for instance, you are using hardware compression on an external coprocessor device, the iSCSI TCP/IP Ethernet frame will have already been constructed and hopefully encrypted at the initiator before being forwarded to the compression device. Such external devices are usually found at the edge of a network to slim the frames heading out to traverse a public medium such as the Internet. Consider carefully how important the SharePoint traffic is before employing such a device that may corrupt the data being delivered to the iSCSI target.

Alternatively, there is a TCP/IP stack protocol for IP datagram compression called IPComp (RFC 3173) that can be employed as an option during IPSec encryption to produce the smallest yet most secure IP datagram for transport. Using IPComp and IPSec technologies together from the initiator will protect the iSCSI PDU as well as the initiator's application data as the technologies are applied at the IP layer of packet construction after the TCP segments containing the iSCSI PDUs have been built. Of course, implementing IPSec can be complex, but it is one of the most secure TCP/IP encryption methodologies available. To use IPSec, be sure to purchase a SAN solution that supports it and do not skimp on the processor power of the SAN controller hardware. Encryption and compression are heavy consumers of processor resources.

Note

To invoke IPSec in Microsoft iSCSI Initiator v2, use the **Advanced** button while adding a new Target Portal on the Discovery tab of the iSCSI Initiator Properties window. The resultant Advanced Settings dialog box offers a General tab for specifying adapter and source IP information and an IPSec tab for configuring encryption. If instead you need to employ IPSec on Windows Server 2008 iSCSI Initiator, simply click the **Set up** button on the General tab of the iSCSI Initiator Properties to configure IPSec encryption.

IPSec over IPv4 works on the assumption that the data trading partners can agree on a security association (SA). Like a contract, the SA negotiated will dictate the manner in which the data will be authenticated and encrypted. IPSec employs an AH to validate the source and integrity of the data. However, this does not provide confidentiality of the data. IPSec then employs ESP to hide the data from anyone but the intended recipient, providing confidentiality, if so desired. These two IPSec technologies can be employed independently or together. The negotiation of the SA is governed by Internet Key Exchange (IKE) laws that state a secure channel must first be developed, over which one or more SAs will be agreed upon depending on each trading partner's capabilities.

Microsoft Windows Server OS supports only three possible authentication mechanisms and three possible encryption protocols for an IPSec session over IPv4:

Authentication

- Kerberos (K_5)—Industry-standard ticket-based authentication protocol; implemented by Microsoft Active Directory (AD) domains and trust relationships; interoperable with other K_5 realms
- Certificates—Public Key Infrastructure (PKI) certificates issued by trusted Certificate Authorities (CAs) provide valid proof of identity
- Preshared Key—Manually configured string value that must be the same at both trading partners; least secure; last resort if K_5 or Certificates are impossible

Encryption

- Data Encryption Standard (DES) 40-bit—Channel protection without individual data packet encryption; best performance but least security
- DES 56-bit—Enhances DES40 by randomly regenerating keys so that if a key is stolen, only a portion of the data stream has been compromised; appropriate for legacy support
- Triple DES (3DES)—Also uses 56-bit randomly regenerated keys but uses three each time, increasing protection at the cost of performance

Cross Reference

Windows Server 2008 now supports IPSec over IPv6, requiring use of the command-line interface tool IPsec6.exe for set up and configuration. As with IPSec over IPv4, both AH and ESP are present but ESP is limited. Unlike IPSec over IPv4, IKE is not used for SA negotiation but rather the MD5 or SHA-1 keys must be specified during manual creation of SA's and policies. For more information see http://www.microsoft.com/technet/network/ipv6/ipv6faq.mspx.

Protecting traffic between iSCSI initiators and targets is important and does not, necessarily, require complex configurations or special hardware. And although this section focused solely on traffic protection and compression, keep in mind that once the data has arrived at the target, it might require protection and compression during placement onto the SAN disks as well. When shopping for an iSCSI SAN solution, look for manufacturers with mature storage encryption mechanisms, sophisticated permission architectures, and granular packet filtering to control what data will be allowed onto the SAN disks, how it will be stored there, and who can get to it.

Monitoring and Resolving Common Performance Issues

Monitoring an active iSCSI SAN is paramount to maintaining data availability. Over time, data fragmentation on the disks can affect seek times while additional storage requests will require additional disk purchases. A famous Ben Franklin quote, "An ounce of prevention is worth a pound of cure," lends itself well to all challenges a computer network can bring. It is always better to proactively maintain a system than to reactively repair it.

Recall that the Windows Server OS of your SharePoint network servers will recognize the iSCSI SAN LUNs as if they were any other mounted disk. Therefore, all the Windows Server monitoring tools for physical and logical disks such as System Monitor, Performance Logs, and Performance Alerts can be implemented against SAN targets. If the Windows Server belongs to a Simple Network Management Protocol (SNMP) management realm, such as Microsoft SMS, SCCM, or SCOM, then detailed information regarding the Windows Server volumes can be trapped or queried for by the centralized administration software.

Building a notification strategy for having the system alert an administrator allows for proactive maintenance while valuable human resources go about other, more productive daily tasks. All the Microsoft monitoring products offer an alert system for notifying other systems or humans in the event of a performance threshold breach. For example, Performance Alerts in the Windows Server Performance Console can be configured to watch the Physical Disk object's Avg. Disk Sec/Transfer counter to reveal how long each instruction to the SAN LUN is taking (see Figure 3.11) and notify an administrator when the duration becomes unacceptable. Unfortunately, the notification methods of the Windows Server OS Performance tools do not include an email choice but the *Run this program* selection could be configured with an executable that would email or page an administrator.

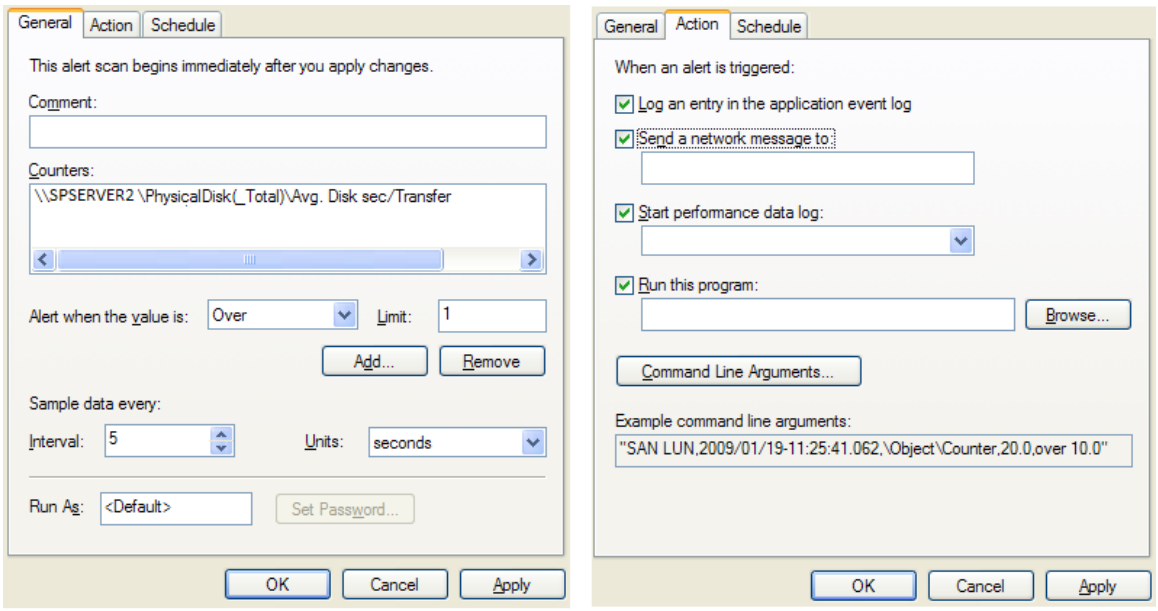

Figure 3.11: Windows Performance Alert configuration settings.

Additionally, a well-commissioned SAN solution should provide on-board monitoring and notification utilities of its own. In fact, these tools should not only perform independently but should also integrate with already established SNMP environment and interoperate with messaging solutions on the network. Be sure to choose an iSCSI SAN manufacturer that offers useful, interoperable monitoring utilities that can provide both real-time information about SAN performance and logged results for trend analysis over time.

Monitoring 101

As with any computer system, monitoring SAN performance accurately demands preliminary effort. First, a baseline must be established by which future counter results will be compared to determine trends. You cannot know what a system looks like sick until you've seen it healthy. Creating an accurate baseline requires that the same counters be taken more than once while the system is performing without known issues. Taking baseline values from a single query could produce skewed results due to irregular activity, and all future trend analysis assumptions would be invalid. If you take baseline counters during issues, the results will be invalid.

Record the same counters during the same time period for at least 3 days. Preferably choose a time period that represents peak *normal user* activity for the SAN. Then take an average of the results, and there is your baseline.

Future queries and reasonable alert thresholds of those same counters can now be set to reveal performance progression over time for trend analysis and to configure notification of failures.

Summary

This chapter examined the fundamentals of iSCSI SAN storage and why iSCSI is gaining popularity in small to large networks over FCIP. We discussed utilizing iSCSI SAN storage for all SharePoint Server roles in a SharePoint farm, and the potential for maximizing disk space investment through the practice of server virtualization. This chapter also revealed advantages an iSCSI solution may provide SharePoint, such as smart allocation through thin provisioning and point-in-time data archiving via snapshots. From a performance viewpoint, we explored the use of MPIO and its implementation on a SharePoint server via Microsoft iSCSI Initiator software as well as enhancements to data transfer via compression and encryption. Lastly, we discussed appropriate monitoring for common disk performance issues to proactively maintain an iSCSI system.

In the next chapter, we will outline various reliability solutions for ensuring data availability and examine common disaster recovery strategies and maintenance tasks to protect your SharePoint investment. Don't leave your finely tuned SAN solution in jeopardy, read the upcoming final chapter full of valuable tips for ensuring data is never lost to those who use it most!

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.