realtimepublishers.com™

The Definitive Guide™ To

Windows 2003 Storage Resource Management

VERITAS™

Evan Morris

Copyright Statement

Chapter 8: SRM and Storage Futures

In this final chapter, we will look at the future of storage and SRM, including both hardware technology and software changes. First, we'll look at the immediate future, at changes that are happening all around us that you may be wise to learn about and consider. Then I'll take a more predictive look into the future and attempt to divine what the predominant or surviving technologies and standards will be. Much of this chapter will focus on networked storage, as clearly that is where the most improvement and increases in adoption will occur.

In the area of hardware, we'll look at changes in "speeds and feeds" as we get faster pipes and possibly even greater distances. Some changes will be more about being able to achieve greater performance over greater distance; but what we are really interested in for this chapter is how these changes will affect storage management. Will they make it better or worse? One of the pending changes that will definitely improve storage management is Directory Enabled Networks (DEN) or, more precisely, directory-enabled storage networks.

One of the upcoming changes is in virtualization of devices and storage, which we touched upon earlier. In this chapter, we will also look at what virtualization means from a storage-management perspective. We will look at the server-side of storage networks, changes in host bus adapters (HBAs), booting from the SAN, and multi-path I/O and what it means for performance and fault tolerance. No discussion would be complete without covering disaster recovery, so we will look at distance mirroring, cloning and snapshots, and serverless backup. Some of these technologies exist today, albeit in their infancy, so we will look at where they will need to go to speed adoption.

One thing to keep in mind during this chapter is the idea of principles over protocols—that is, keep the business value in mind whenever you are investigating new technology. To give you a concrete example, some new technologies promise to give you the ability to bridge storage islands (which may be defined or isolated as such by specific protocols or cabling). Perhaps this technology is of interest to you, but what will be the benefit to the business? Do the applications running in each of these islands really need to be bridged or are they better off in the safety of isolation?

Immediate Future

This section covers changes to your environment that you may face immediately or over the next 6 months. These are not wild predictions; instead, they offer guidance on where you can define your storage strategy. Many of these technologies are currently available to you.

The Future of SRM

With storage capacities and storage consumption growing at phenomenal paces, SRM will play an increasingly important role in the organization. Storage administrators will need to manage complex environments without regard to whether the storage is SAN or NAS or whether the storage is connected to server X, Y, or Z. The desire is for *single seat administration* using a single set of management tools; however, that dream is a long, long way from reality.

In the near future, the best that we can hope for are changes in the management capabilities at the lower levels of the storage architecture (from the disks, controllers, switches, host adapters, OS, and applications) that enable information to roll up to the directory services. The first phase will be enabling the view of the entire enterprise storage infrastructure, followed by the ability to apply storage resource policies to every level of configuration detail.

Storage-Management Utilities

The following Win2K and WS2K3 utilities will make disk management easier for you. WS2K3 ships with additional command-line tools and utilities that were previously available only in the resource kits. Many of these are disk- and storage-related (for example, Freedisk, which lets a command run only if a specified percentage of disk space is available; and TakeOwn, which lets administrators take ownership of orphaned files).

DiskPart

DiskPart lets you manage disks, for example, by extending a disk volume while the storage is online to the OS. DiskPart is fully scriptable, using the syntax

```
Diskpart /s <script>
```

Figure 8.1 shows the commands for using DiskPart, which is also useful for rescanning the server to detect any devices that have been presented from a SAN. For example, after breaking off a Business Continuance Volume (BCV—such as a clone) and presenting it to the host, we use DiskPart to detect the new drive and mount it.

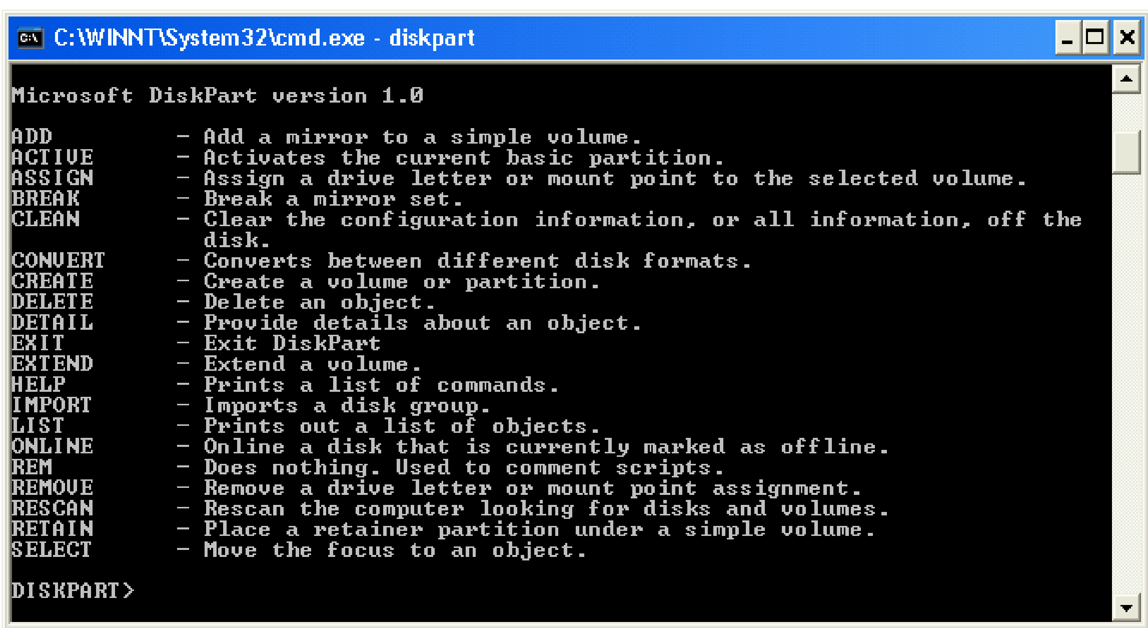

Figure 8.1: DiskPart commands for managing a disk volume (Windows XP version).

DiskPart is available for Win2K by download or as part of the Recovery Console (I discuss the installation instructions later) as well as in the default installation of Windows XP and WS2K3. Be careful and make sure that you are using the appropriate OS version, as there can be differences in how they operate (see the following note). In Win2K, the DiskPart command is only available when you are using the Recovery Console, so most of the benefit in a production environment for changing disks will be to WS2K3 systems.

> Microsoft previously released an earlier version of DiskPart on the resource kit Web site. To verify that you have the correct version of DiskPart, check the properties. Earlier versions of the DiskPart.exe file have a file version of either 0.52 or 1.0 and the later version has the following properties:
>
> Created: September 21, 2001
> Size: 146,432 bytes
> File version: 5.1.3553

> To install the Recovery Console as a startup option in Win2K, insert the Win2K CD-ROM, and hold down the Shift key to prevent the CD-ROM auto-run feature from running or wait for the auto-run feature to bring up the installation options. Close the installation wizard, run a command prompt, and type the following

```
x:\i386\winnt32.exe /cmdcons
```

> where *x* is your CD-ROM drive letter. If you have the bits copied to disk, you can run the installation directly from the hard drive. Answer Yes to the prompt that Figure 8.2 shows, and installation will begin.

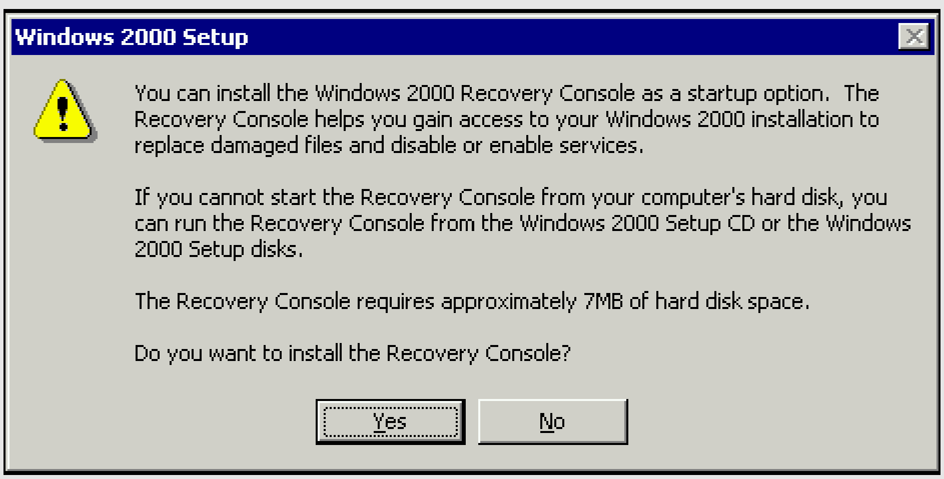

Figure 8.2: Installing the Recovery Console.

> The installation won't prompt you to reboot your system, but the Recovery Console will be available as a boot option the next time you reboot your system. The installation did not prompt me for the SP2 source location, so I recommend running the installation from a Win2K source that has had SP2 slipstreamed in (by running

```
\i386\update>update –s <dir>
```

> where <dir> is the location of your Win2K source files).

DiskPart can also add or break mirrors, assign or remove a disk's drive letter, create or delete partitions and volumes, convert basic disks to dynamic disks, import disks and bring offline disks and volumes online, and convert master boot record (MBR) disks to GUID Partition Table (GPT) disks. The options under CONVERT for DiskPart are as follows:

- BASIC—Converts a disk from dynamic to basic
- DYNAMIC—Converts a disk from basic to dynamic
- GPT—Converts a disk from MBR to GPT
- MBR—Converts a disk from GPT to MBR

For information about GPT disks, see the "GPT Disks" section later in this chapter.

> Just because you can run it from a command line or script does not mean that it will not destroy your data! Always test your backup before you perform these types of disk operations!

Fsutil

Fsutil is another command-line utility that an administrator can use to perform file–system–related tasks, such as managing reparse points, managing sparse files, dismounting a volume, or extending a volume. Figure 8.3 shows sample Fsutil commands and usage for the quota command.

> You must be logged on as an administrator or a member of the Administrators group to use fsutil.

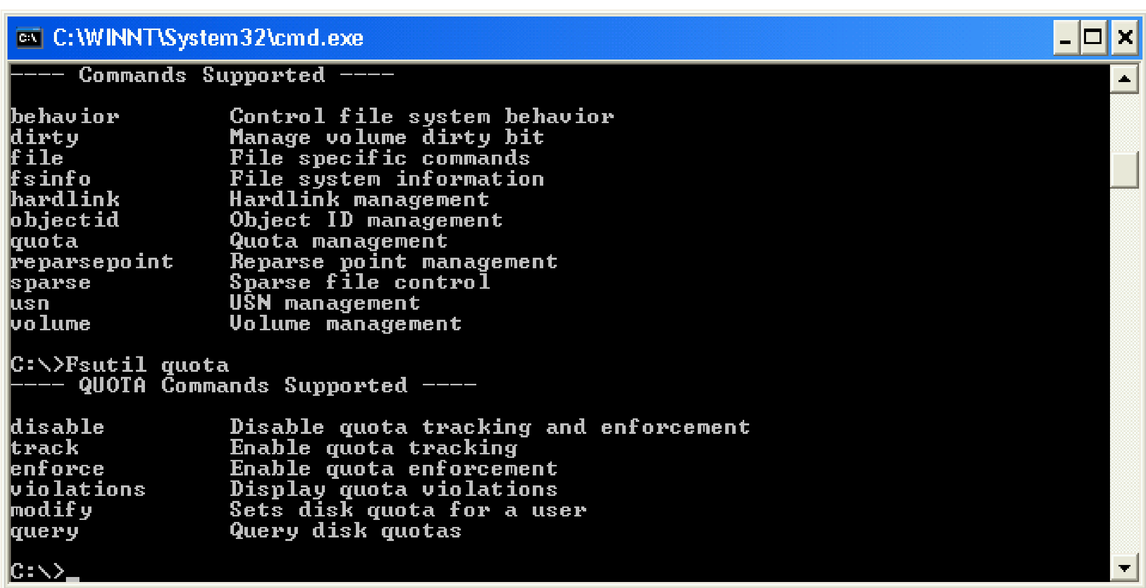

Figure 8.3: Fsutil quota commands.

Enhanced Device Support

Although SAN support has been available for quite some time (from back in the NT days), there are a great number of improvements yet to be made; how the OS handles new devices on the SAN, for example. From what I have seen and heard, performance of the storage subsystem will be capable of using much more of the bandwidth available in new devices such as HBAs. Vendors will need to provide new miniport drivers, however. You may not be able to upgrade the underlying OS without first getting new storage device drivers.

GPT Disks

In the 32-bit Intel world, we are used to using MBR disks, but the 64-bit Intel world brings a new type of disk known as a GPT disk. Starting with the Intel Itanium processor, the new 64-bit servers use the Extensible Firmware Interface (EFI) between the computer's firmware, hardware, and the OS instead (of a BIOS).

Disk partitions on basic GPT disks are created either by using the EFI firmware utility, diskpart.efi, the diskpart.exe command-line utility, or the Disk Management MMC in the 64-bit WS2K3. You will be able to manage both MBR disks and GPT disks in the Disk Management MMC. However, on a 32-bit machine, the GPT disk appears as a basic MBR disk with a single partition, and the data cannot be accessed, as there is no translation to the MBR disk format that the 32-bit machine can understand. Understandably, you cannot easily move GPT and MBR disks between machines. The 64-bit servers can certainly house MBR disks, but they must boot to a GPT disk. Interestingly, you can combine MBR and GPT disks in dynamic disk groups, but I find it highly unlikely that you would want to do so. Converting between the two, MBR and GPT, is data destructive, so it should only be done on a disk that is considered empty.

MBR disks support volumes as large as 2 terabytes (TB) with as many as four primary partitions per disk or three primary partitions and one extended partition with unlimited logical drives. GPT disks support volumes as large as 18 exabytes in size and as many as 128 partitions per disk. Also, GPT disks keep redundant primary and backup partition tables for fault resilience.

Cluster Cluster

This subhead illustrates the redundancy of clustering. However, clustering in Win2K and WS2K3 has been a mixed success. At first, it looked rosy, especially compared with the difficulty of clustering in NT. In the application arena, SQL Server 2000 has done well, but Exchange 2000 Server has not, with severe limits being placed on users per server in active-active clusters. Other services such as file and print and DHCP have enjoyed fair success, though they aren't as easily cost-justified in clustering as more business-critical applications.

WS2K3 clustering has increased the number of nodes in a cluster, up to eight nodes in WS2K3 Enterprise Edition, for example. Another new clustering enhancement in WS2K3 is in the area of geographically dispersed clusters. With four nodes or more, you can have a pair of failover partners on either side of the geographically dispersed locations (called a *majority node set* cluster). The chances of failover as a result of server hardware fault or application crash are greatest, and you don't want to fail over to the remote cluster unless a catastrophic failure causes loss of the data center.

A potential improvement I'd like to see for a future version of Windows Server is the inclusion of clustering support in the base OS. With Win2K and WS2K3, you add the clustering component and reboot to load the driver. If you are having trouble with a node, you must evict the node and reboot to remove the driver. Microsoft may realize that this requirement is less than desirable, especially in an environment in which the customer has chosen to spend extra money on clustering to gain higher availability. In addition, when the clustering driver is assumed to be installed by default, running compatibility tests, such as for filter drivers in antivirus applications, is easier. Clustering is, of course, now supported in the WS2K3 64-bit environment, but all nodes need to be 64 bit, as it is not possible to run mixed 64-bit and 32-bit applications.

Yet another change to look for down the road is in how the quorum resource is handled. One of the prerequisites for installing a cluster is to have access to a device on the shared storage system. In Win2K, if the cluster installation wizard does not detect an external device, it will not allow cluster installation. By changing to the concept of a majority node set in WS2K3, the quorum can be kept on a local disk volume and other cluster members look to that source to be kept in sync. This setup allows cluster installation before the node is on the SAN, and allows for a greater number of nodes in a geographically dispersed cluster. However, not all WS2K3 clusters are majority node set clusters; the restrictive external storage device requirement for traditional clusters remains, even under WS2K3.

> Although I use the phrase *shared storage system*, this terminology can cause some confusion as the Microsoft Cluster Services (MSCS) use a shared-nothing model. Perhaps a better phrase would be networked storage.

Another badly needed change in Microsoft clustering is a change in the cluster arbitration mechanism. Current Windows clustering technology uses a bus reset, which is extremely disruptive in a SAN environment and should only be used as an absolute last resort.

Finally, one other area that needs improvement in clustering is the ability to break past the 22 drive-letter limitation when configuring clusters. No, that is not a typo—I know that our alphabet contains 26 letters, but 4 letters are usually consumed by the floppy drive, boot drive, CD-ROM, and the quorum, leaving 22 letters. You might rarely encounter this 22-letter limitation, but I have run into it in a few very large database environments. The solution to this problem is to allow volume mount points in clusters so that a database volume can be mounted as a folder on an existing drive letter.

SAN Boot

Although technically possible in Win2K and WS2K3, placing the boot drive (usually C) on a SAN device has been recognized by Microsoft as a feature that needs to be officially supported in upcoming Windows Server releases. I remember the first time that I built a Win2K Server without any internal disk drives. At the time, the ability to SAN boot had just become available in the latest firmware of the Fibre-Channel HBAs. It was a strange sight to see a rack full of servers running without internal disks. However, it worked without a hitch, and we were able to fulfill our primary mission, which was to boot the servers to disks on either side of a remotely mirrored storage network. This capability is the type of flexibility offered by SAN boot—the ability to place an OS image on the SAN, enable access to a server, and bring the server online. There is potential for the large data center in being able to manage an OS image library—bringing new servers online in a matter of minutes to meet growing or shifting demand—rather than using the traditional setup build processes. I'll give you more information later about why this ability is important from a storage manager's standpoint.

Multipath I/O

When placing multiple storage adapters in a server, it is necessary to use some form of vendor-provided software to manage the multiple data paths. Typically, software such as Compaq Secure Path and EMC PowerPath is used to ensure that failover performs without any interruption to the data flow. But the greater ability of using multiple adapters is beyond fault tolerance—the greater ability is to combine both paths into one wide data path. This type of functionality could actually be provided at the OS level, detecting multiple paths and managing the failover.

Volume Mounting

One method of finding out how much others on your SAN know about proper volume security (LUN masking or selective storage presentation) is to bring up an NT or Win2K server on the SAN. This method is not the best, and I do not recommend doing it if you have any doubts about the safety of other hosts' disks. Windows servers are known for claiming any disks that they can see as their own, attempting to mount them, and even requesting that a disk signature be written. The issue is that NT was developed back when the majority of storage was direct attached, and the implications of SAN storage were not fully realized.

Another feature that would be nice in the next- generation Windows server would be more administrative control over whether Windows attempts to mount any disk that it can see. Granted, if the disk can be seen, proper volume security (LUN masking or selective storage presentation) may not be in place; however, you should still have control over whether the disk belongs to that particular host and should be mounted.

DAS vs. SAN vs. NAS

Why consider networked storage (SAN or NAS)? Perhaps you just know that your company or department would benefit from the technology, but you need to further explain it to higher-level decision makers in your organization. Networked storage technology must meet the requirements of your business application or it is just expensive, complex technology, which in the wrong hands can put your business information at risk. The following list provides enterprise organizations' common business requirements for a separate storage network:

- Do you copy or move large files over the network? Is this usage of the network adversely affecting other business functions?
- Do you back up or restore a large amount of data over the network? Does the restore take longer than acceptable SLAs?
- Is the network a source of congestion for these operations (especially for recovery operations, which may mean failure to meet SLAs)?
- Do you need to share data amongst multiple servers, such as in a clustered environment?
- Is data (or more appropriately, information) so critical to your business function that it needs to be replicated to another location (in case of fire, flood, or earthquake damage at the primary site)?
- Would the business benefit from the ability to present the logical disk units to multiple hosts with few configuration changes (for example, in a disaster recovery situation)?
- Is the recovery window so critical that the cost justification of BCVs (cloning or snapshotting data volumes) also sound business justification? The cost of BCVs includes additional disk space used, such as triple-mirroring or a virtual drive pool; administrative overhead to develop, test, and manage the solution; as well as any additional hardware or software required to support the solution.

Now let's turn to NAS. There have been several recent developments that will impact the future of NAS. Fibre-Channel SANs have had a head start, and NAS has been playing catch up. To some degree, Fibre Channel is synonymous with SAN. NAS has excelled at offering cheaper storage, as long as it is file-level access that you need. This functionality is fine for file sharing but not for relational databases. What NAS has lacked, it is beginning to provide—block-level access to data. There are NAS devices that provide a block-level driver to ensure write-ordering on the storage devices, which is usually not possible using the TCP/IP network redirector.

Recent market studies have shown greater adoption of both SAN and NAS as opposed to DAS. Although I have long been a fan of DAS for its initial low cost and the high performance of newer DAS devices, the high cost of DAS' ongoing management will steer many companies to NAS or SAN instead of DAS. In addition, the differences between SAN and NAS will begin to lessen as we see greater performance NAS and easier-to-configure-and-manage SAN environments. Although we will eventually see convergence of SAN and NAS, it won't take place for some time, and in the meanwhile, we must consider interoperability as the primary mission. Table 8.1 compares common storage obstacles for DAS, SAN, and NAS.

Storage Issue	DAS	SAN	NAS
Storage Network	Isolated storage may mean over-utilization or under-utilization.	Requires separate Fibre-Channel network; capacity of 1Gbps and 2Gbps networks will soon be surpassed by 10Gbps Ethernet (if it can move the same amount of data).	Must redesign your IP networks to ensure that any problems, such as security and saturation, do not affect storage.
Data Mode	Block Mode, but unable to share devices.	Block Mode—Not designed to share data across applications (unless managed by a cluster service).	File Mode—Not appropriate for some applications such as high-end database servers. Vendors are starting to provide block-mode filter drivers.
Difficulty	Easiest to set up, except when large amounts of storage are needed. Becomes most difficult to manage in the long run.	Most difficult to set up, requires specialized knowledge (expert roles from many vendors, especially to interoperate).	Main difficulty is in getting NAS to work with applications and environments that may demand high performance and near-zero latency.

Table 8.1: Common storage obstacles to overcome.

The Demise of NAS?

The demise of NAS is a difficult scenario to envision; however, there is some industry speculation that NAS will fill only a niche market as storage is dominated by SAN. After the previous explanation of NAS becoming more SAN-like by providing block-level access to data, I offer this counterpoint: The big changes in SANs that will lessen the ability of NAS to compete are virtualization and volume shadow copies. Virtualization in the SAN may allow for more efficient utilization of those expensive SAN resources. And volume shadow copies allow the data to be replicated over distance and protected from data center disasters. I find it hard to believe that SAN could dominate NAS, but I bet you will find someone arguing just this point in a storage planning meeting over the next few years.

Interoperability

Although each storage vendor has its own interests centered on a line of products, your interests as the customer is for the storage to interoperate, or work well together. Multi-vendor interoperability is a key future direction, and some vendors have a definite interest in ensuring compliance and are investing resources in that effort. For example, the Supported Solutions Forum is a multi-vendor group sponsored by the Storage Networking Industry Association (SNIA). The forum is designed to bring together competitors such as EMC, IBM, and Compaq to create solutions involving servers, HBAs, and storage components such as switches and backup devices.

SAN and NAS Interoperability

Traditionally, deployments of storage networks have focused on SAN devices and excluded NAS devices; however, the technological reasons for this gap are disappearing and will soon be leaving only organizational and political reasons. There are existing devices designed to connect SAN and NAS networks. Future gateway devices will allow protocol bridging or routing between SCSI, Fibre-Channel, and enterprise system connection devices, in addition to the newly emerging iSCSI for storage over IP. What will emerge is the greater ability to classify and organize storage based on qualities such as

- Cost
- Performance
- Availability and fault tolerance (through RAID levels and redundant hardware, which are intrinsically aligned with cost)

The ability to present multiple classes of storage to a server is highly beneficial. The direct benefit to you will be that you can prioritize and provision storage based on how critical the data is to business and the ROI to the organization.

SAN Management API

A recent trend is that storage-product vendors are exchanging information on managing storage devices (not necessarily by giving the information away, often charging for the information). Did you ever think that you would see the day when rivals EMC and Compaq would be sharing information? Such was the case recently when the two companies decided that the need for storage management (for either company's hardware) was a higher call that requires their cooperation. The results of the joint effort will be that Compaq will add an Element Manager to its SAN appliance to manage the EMC Symmetrix initially and the EMC CLARiiON later. On the other side, EMC will use both the Compaq APIs and the SANworks Command Scripter (which provides a command-line interface at the host) in the EMC AutoIS SAN-management software. The other development to watch is a common model, such as the SAN Management API version 2.0, and the work groups at the SNIA.

> For more information about storage-group developments, go to http://www.snia.org/ and http://www.T11.org/. There are two excellent white papers (although the information may be getting old by now, the overall framework still applies) at http://www.snia.org/English/Collaterals/Whitepapers/Shared_Storage_Model.pdf and http://www.snia.org/English/Collaterals/Whitepapers/SANWP2.PDF

Hardware Technology's Future

In this section, we will look at some pretty sure bets in the future of hardware technology, and how these developments will impact storage management.

Speeds and Feeds

Let's look at some of the imminent changes in getting data to and from storage. We will see some changes in the hard wiring of storage networks. But will we see wireless SANs? My first reaction is to say doubtful, but we have recently seen wireless speeds increase, especially in burst mode over short distances, which could eventually be used for SANs. Perhaps I'll look back one day at my doubt and laugh, just as I laugh at my doubt in 1991, when I read that the typical personal computer in 1999 would have more than 100MB of memory.

2Gbps Fibre Channel and Beyond

2Gbps Fibre Channel is essentially here, just as Gigabit Ethernet to the desktop may be here, but that doesn't mean that you and I have it! The change to 2Gbps Fibre Channel will require upgrades to our HBAs and switches, and our Gigabit Interface Connectors (GBICs—the converter between the electrical and optical signals) will need to be changed to Small Form-Factor Pluggable (SFP) transceivers, which require a new connector on the fiber cables. Unlike Ethernet equipment, a Fibre-Channel fabric will fall back to the lowest speed of a device communicating on the fabric, so if you do not upgrade all devices in the data path, you will not benefit.

Some people may wait for the next big jump in speed (for example, 10Gbps) before upgrading from 1Gbps. As we covered in the previous chapter about storage performance, your applications may not even benefit from the upgrade. It is analogous to upgrading from Ultra2 SCSI to Ultra3 SCSI and increasing bandwidth from 80MBps to 160MBps: This upgrade provides an increase in throughput, but your application may be constrained by disk I/O operations or even disk response times. In fact, in testing the new 2Gbps hardware, it is very difficult to push the performance test hard enough to demonstrate the throughput capabilities of 2Gbps—it takes such a huge array of disks to fill a 200MBps pipeline, that the only real way to max out the throughput is to use a large cache so that you are reading the bits from solid state memory instead of spinning magnetic disks.

2Gbps Fibre Channel may be the standard for future implementations, especially if you're setting up your first SAN and are buying new equipment. However, upgrading from 1Gbps to 2Gbps does not make sense unless you know that your existing Fibre-Channel fabric is the bottleneck. It will be an estimated 2 to 3 years before you need to look into 10Gbps Fibre Channel.

Fibre-Channel Topology

In the near future, Fibre-Channel switches will drive down prices into the current range of Fibre-Channel hubs, slowing the adoption rate of looped environments or driving them to specialized tasks such as isolated clusters. Of course, some of them may end up under your desk! Arbitrated loop environments may be fine for learning the ropes of Fibre-Channel, but it will not get you into fabrics, zoning, and multi-host environments. In addition, port densities on Fibre-Channel switches are increasing to the range of 256 ports (albeit at a price premium).

10GB Ethernet

Storage networks have been leapfrogging each other in maximum speed, with 2Gbps Fibre Channel surpassing Gigabit Ethernet. Meanwhile, work is being done to push Ethernet networks to 10Gbps Ethernet. Even though the capacity of 1Gbps and 2Gbps networks will soon be surpassed by 10Gbps Ethernet, the determination will be whether it can move the same amount of data. Theoretically, a 10Gbps link can fill even the largest of today's hard drives in a matter of minutes! However, Ethernet-based storage typically has much higher processing overhead than Fibre Channel, including packet re-sends. Much of this processing is moved to a processor onboard the NIC, yet we still see some server CPU overhead remain, and the impact of this overhead for 10Gbps or 1.25GBps has yet to be seen. So, the end result in data throughput may be that 2Gbps Fibre Channel is still a viable competitor for 10Gbps Ethernet for a few more years.

Volume Management

Changes in how disk volumes are secured and how host access is determined may be forthcoming. Arguments are being made in the industry that the old method of securing access by HBA MAC address may not work in a directory-enabled storage network. Some of the access may be determined by use of a SAN appliance on the Fibre-Channel fabric.

The Role of HBAs

What role will HBAs play in the future of storage management? Any device that is connected to a storage network can play a pivotal role in the gathering of information about other attached devices. Already, we have seen changes in HBA abilities have an impact, for example, in providing SAN boot capability. In the next section, we will also see how HBAs can play a role in virtualization.

Virtualization

Probably the biggest change in hardware technology is that of storage virtualization. This change will impact everything from performance, how storage units are created, and how storage is managed. I opted to include virtualization in this section rather than the earlier section about the immediate future because virtualization is such a drastic shift (in essence one of the few places where it would be appropriate to use the phase “paradigm shift” and not be in a Dilbert cartoon). Virtualization is really in its infancy, and we will see much to come in this area. I can understand virtualization when it comes to disks and have worked with new storage systems that abstract or virtualize some disk array creation, but imagine what is possible when entire pools of storage including disk, optical, and tape are virtualized and categorized to support an HSM system.

The central concept of virtualization is to create a large pool of storage resources and control host access, presenting only what is needed or requested. Virtualization requires upgraded capabilities in the SAN, either in the form of new storage systems and a SAN appliance or new drivers for the OS and HBAs. There are several different methods of implementing storage virtualization, usually either in-band or out-of-band.

In-Band

The first method of virtualization, in-band, uses a SAN appliance or software directly in the data path to control access to storage devices at the block level. When a block request is made for data, the SAN appliance or software is responsible for locating and retrieving the data. The main advantage of this approach is that there is no access to data on the SAN without the "permission" of the virtualization manager, which helps to prevent unwanted access to disk resources. If you have ever had someone plug a Windows server into the wrong Fibre-Channel switch, you know the reasons why this advantage is important—Windows has had a tendency to think that any devices it can see are local devices and that it should own those devices, including writing new disk signatures.

The disadvantage of the in-band approach is that the SAN appliance or software can become a limiting factor in SAN scalability, as it must control all I/O, both to and from the servers and storage. In the event that there is a problem with the SAN appliance, it is critical that the appliance hardware be fully fault tolerant, including redundant components and paths, otherwise you could lose your entire SAN.

Out-of-Band

Another implementation of storage virtualization is out-of-band, in which the SAN appliance (or even memory on the HBAs) acts as an intermediary, storing the location of data blocks requested by the host. A table in memory must be kept updated by a virtualization manager and any new devices added to the SAN must comply with the virtualization scheme. The main advantage of out-of-band virtualization is that there is no central route for data access, as each server can transfer data directly to storage once the location is known.

Distance Mirroring

The replication of storage data to a second location, also known by names such as *distance mirroring* and *remote copy sets*, is not new to SAN technology and has long been available. The differences that we will see in the immediate future are the ability to use a wider variety of transports and to increase the distance between storage systems. Currently single-mode Fibre Channel reaches 10 kilometers, and other transports can be used for greater distances (100km and beyond with distance-enabling technology), albeit at a certain latency penalty. The primary criteria for determining distance capability are the application's ability to tolerate latency and keep I/O synchronous. As latency increases, the ability to perform synchronous I/O becomes increasingly difficult, and the replication must be done asynchronously if the application can tolerate it.

The transport choices for Fibre-Channel distance mirroring include native (dark fiber), Synchronous Optical Network (SONET), asynchronous transfer mode (ATM), and IP. In addition, as we see 10Gigabit Ethernet networks hit the market, Fibre Channel over IP (FCIP) and iSCSI will play a greater role in distance replication, extending the potential distances while lowering the cost. The complexity and cost of these solutions will decrease to the extent that more businesses will look to distance mirroring as a disaster-tolerant solution.

SAN and WAN Converters

There are really two categories to pay attention to when it comes to the transport of SAN protocols over WANs. The first are the boxes that act as protocol bridges or routers and perform the protocol translation. As you may well know, these are expensive boxes and are often connected to expensive runs of fiber. Fibre-Channel bridges to FCIP, and in the future perhaps iSCSI, enable wide-area connectivity over IP or even ATM.

The second category to watch are the companies who purchase this expensive capital equipment and sell it by the connection or channel. Quite a bit of investment has been made recently in metropolitan area optical networks, and companies such as CNT act as a communications provider for data.

Native Fibre Channel to Disk

The ability to use native Fibre Channel all the way from the HBA to the disk devices has been available for a number of years, but the technology has been prohibitively expensive except for high-end data centers, which usually do not include Windows servers. We will see that change, as pure Fibre Channel becomes more affordable. Yet it must still be justified from a cost-performance standpoint, and the primary advantage is that there is no SCSI translation at the disk.

SAN Boot

The ability to place the OS boot drive on the SAN is in ever-increasing demand these days. Part of the demand is the result of the increasing density of servers (especially blade designs)—environments in which servers may have no internal storage, instead relying on the internal HBA to gain access to the boot device on the SAN. In addition, having the OS boot drive on the SAN aids in recoverability, as a standby server can be brought in to replace a failed server and access the boot drive (assuming the volume security has been changed to allow the second host access).

The future potential of SAN boot also includes the possibility of extremely rapid server deployment—imagine a scenario in which another Web server or application server is needed to satisfy increasing load. Either through administrator intervention or dynamically through software, a boot image is placed on a SAN disk and a server is brought online, fully configured and ready to perform.

New Device Classes

Of course the future will bring entirely new types of devices, but we have little idea what those devices will be (other than the obvious extension of existing capabilities). It seems that the weakest link in the data center lately is in the area of backup media. Disk capabilities have grown at a rate that is difficult to keep up with, and tape technologies have lagged behind. Some companies are building backup systems using inexpensive IDE drives (these can be rotated offsite and treated very similarly to disposable tape media), and we may see these become common in the data center if they prove to be a viable substitute.

In the area of extending our existing capabilities, we will see technological innovation create higher performance designs. For example, with memory cost relative to density dropping so low, it is surprising that we do not see greater adoption of solid state disk (SSD) technology. SSDs are still expensive, but there are uses for this technology where RAM (memory) is just not a suitable replacement (for example, in high-speed transaction logging disks that must be non-volatile, surviving a power interruption). So perhaps we will see a hybridization of this technology in which greater amounts of solid state memory are pooled across an array of disks, similar to the way that caching is used currently but accessible to the host as a disk volume for when it needs to make temporary disk swaps or transactions.

Bus Architecture

For quite some time, we have been hearing about a new bus architecture, InfiniBand, designed to eventually replace the PCI bus. If InfiniBand is slow in coming, it is alright by me—changes in the bus architecture of servers are pretty disruptive. They *can* be beneficial in relieving one source of potential data bottlenecks, but you may be just now implementing PCI-X as the next generation of server bus architecture and aren't quite ready for another drastic change.

Software Technology Futures

In this section, we will look at the direction of storage futures beyond the immediate future. These are extremely speculative, as the current next release of Windows Server—codenamed "Longhorn Server"—is still under definition by Microsoft and scheduled for a 2006 or 2007 rollout. Even that release is intended only to add support for the Longhorn Client release of that same timeframe; the next major release of Windows Server, codenamed "Blackcomb," is even further out. So take this with a grain of salt.

Perhaps you are the type of storage administrator or systems administrator who has a personal wish list of future OS features. You can easily perform what is called a gap analysis, by taking a look within your organization at where you spend what you consider to be your most unfulfilling or least productive time. At times, I bet it is performing frustrating hours of troubleshooting only to find a simple but overlooked cause, such as a device that is intermittent or a configuration change that had a larger impact than the operator knew about. At other times, I bet it is the process of adding or expanding existing storage and recreating the information to manage it. No doubt the features that you most desire are the ability to provision and expand storage systems while you also control the configuration of your environment.

WinFS

Windows Future Storage (WinFS) is a new file system built upon the existing NTFS file system. WinFS combines NTFS and SQL Server to provide a number of enhanced features and capabilities. For example, objects within the file system—such as files and folders—can have complex relationship with one another, and can be searched by a powerful full-text search engine more rapidly than today's Indexing Service permits. However, there's no current information on whether WinFS will make it into the initial release of Longhorn, or be scheduled for release later. When released, WinFS should provide benefits primarily in the area of data and storage management, making it easier to find out what data is stored where, and to make that data more readily accessible to users and applications.

DEN Enhancements

The goal of DENs, or directory-enabled storage networks, is to be able to provision (configure new devices) and manage your entire storage infrastructure from a centralized point with as little repetition of duties or tasks as possible. Similar efforts are being made in the server and network infrastructure toward the goal of plugging a server blade or switch blade into a rack, having it poll the master directory for its role, and configuration taking place with little human intervention. During the rest of its production life cycle, the hardware, whether server, network, or storage, may be re-deployed dynamically to meet shifting demand in Web transactions, transaction processing, or any other current needs.

One of the milestones on the road to directory-enabled storage networks is a centralized repository and a common set of protocols. The Distributed Management Task Force (DMTF) is working on this task—defining a common information model, CIM 2.5, and mapping the schema from CIM to an X.500 LDAP directory. The design goal is to deliver network services based on pre-defined business criteria. For example, letting multiple priorities and classes of services assist in the provisioning of storage hardware. But first, all the devices that make up enterprise storage infrastructure must fit into the data model, and the storage software must understand the rules that we define.

Think about the explosion of the Internet when it seemed to hit critical mass (although we know that it will continue to grow and evolve). The main factors were bandwidth at a reasonable price (initially dial-up), a standardized set of protocols (HTTP over TCP/IP), and the means to create a virtually centralized (but distributed) repository in the form of the World Wide Web. Even search engines have evolved their techniques from Gopher to Webcrawler to Alta Vista to Google and beyond. A research associate and I once discussed this topic at length, and he added that it also took a robust user experience, initially provided by the Mosaic browser and later IE and Netscape. I remember my Lynx browser that worked fine until I started getting more and more graphics placeholders—but my point is to find parallel patterns in the growth or evolution of storage management. From an administrative standpoint, you will need a centralized management starting point, even if the configuration information is stored in distributed devices. As long as the devices speak the same language or use common interchangeable protocols, you can essentially hyperlink or browse them. The rich UI is, of course, very helpful when dealing with everything from disk spindles to enterprise data centers in one console.

So it may be some time before we are fully directory-enabled, but in the meantime, you can keep a watchful eye in the industry press or at storage conferences for the emerging standards. You can't expect to pick the winner in every detail of directory-enabled storage networks (unless of course you were just sure of the winner in Super Bowl XXXVI), but expect the vendors that you use for your storage networks to participate in the creation of an open DEN standard.

> For more information about DENs, see http://www.dmtf.org/standards/standard_den.php.

Dynamic Volume Management

The rules that determine adding space to an existing drive (dynamic volume management) have typically been very limiting. If you are most familiar with NT storage, you are probably used to rebuilding RAID sets manually. With WS2K3 and the addition of third-party utilities such as VERITAS Volume Manager, it is possible to extend a disk size without having to back up, rebuild, and restore data. In the not-too-distant future, these limitations will seem like a bad memory and will be hard to explain to the storage newcomer. We will pull from a pool of storage resources to dynamically grow any disk that requires additional capacity. Obviously, this possibility will take some re-work of the Windows OS.

Multipath I/O

As mentioned in the previous section, multipath I/O will eventually become integral to the core OS. However, there will still be a market for third-party vendors to extend multipath to include the ability to dynamically load balance across a variety of adapters (even from multiple vendors) and even to multiple devices. New devices classes will include optical, and tape, in addition to disk. We will see support for larger numbers of paths beyond the usual two.

Security

As we watch the evolution of storage networks—especially as we get close to bridging our SAN with our LAN/WAN infrastructure—we will see analogous attention given to security. Recently, security has been driven to the forefront of IT in a manner that I only hope we can avoid by being proactive in our storage networks. For example, LUN masking or selective storage presentation methods may vary across vendors and hosts, and access to physical devices is of utmost importance.

Shared File Systems

In a shared file system, multiple hosts have access to data on the same disks at the block level. Contrast this setup to the Microsoft approach to clustering (which is known as a *shared-nothing* approach), which uses the Cluster Manager to limit disk access to one host at a time. The advantage of sharing the file system is that multiple hosts can respond to client requests, such as allowing a large bank of read-only analysis servers and several data-input or transaction-update servers. This setup requires the use of a Distributed Lock Manager (DLM) to coordinate the updates to disk, ensure that there are no write collisions, and ensure that the read-only servers see the same data. There are currently third-party shared–file–system solutions for Windows available from companies such as VERITAS and IBM Tivoli, and we may see one of these vendor's solutions adopted into Windows' core similar to how we have seen a subset of the VERITAS Volume Manager functionality licensed for the core OS Logical Disk Manager (LDM).

Storage Protocols

Much of the focus on storage protocols is on storage internetworking, as each type of storage network (SAN or NAS) has its set of strengths and its proponents. What it boils down to is Fibre Channel versus IP. Most SANs are Fibre-Channel protocol, but IP networks have great appeal because of the existing cabling and routing infrastructure, wide choices of standardized hardware, and the highly skilled workforce available to support them. In addition, Ethernet IP networks have been more open and inter-connected and have struggled through some of the security issues that Fibre-Channel networks have yet to face.

FCIP

FCIP is a protocol standard developed by the IPS Working Group of the Internet Engineering Task Force (IETF). FCIP is designed for connecting geographically dispersed Fibre-Channel SANs over IP networks, with a key distinction being that all Fibre-Channel devices are unaware of the presence of the IP network.

> For more information about FCIP, see http://www.snia.org/English/Forums/IP_Storage/IP_Storage_FCIP.html and http://search.ietf.org/internet-drafts/draft-ietf-ips-fcovertcpip-09.txt (note that this document name will change, so you can find information about FCIP by searching at http://search.ietf.org/search/brokers/internet-drafts/query.html).

Storage over IP and iSCSI

iSCSI is one of the potential protocols for storage over IP, enabling more efficient transport than the network protocol. iSCSI is intended to provide block-level access to storage, which would include application databases in addition to the traditional file-based storage. Although there are other contenders, such as Storage over IP (SoIP), iSCSI is positioned to be the dominant protocol, and the one I choose to watch.

In addition to using existing network infrastructure, iSCSI may be an enabler for distance replication beyond the 10 kilometer limit of single-mode Fibre Channel. However, iSCSI will need to overcome the early momentum being gained by FCIP enabled by Fibre-Channel bridges. In fact, I would bet that iSCSI is a greater market threat to DAS than Fibre-Channel SANs because iSCSI allows much easier interconnects and networking of low-end storage systems that would be traditionally SCSI-cabled.

The Direct Access File System

Several large players in the NAS arena, such as Network Appliance and IBM, are also counting on the implementation of the Direct Access File System (DAFS) to give them the performance that they will need to become more widely adopted in the database and application server market. SNIA has created the DAFS Implementers' Forum to carry on the work started by the DAFS Collaborative.

> For more information about DAFS, see http://www.snia.org/English/Forums/DAFS/DAFS_Docs.html or http://www.snia.org/English/Forums/DAFS/DAFS.html

Storage Management

SAN devices will play a key role in SAN management, with current work progressing on providing APIs for managing common device classes. As I previously mentioned, major storage vendors are even sharing information, which is a surprise considering that they often have competing products. But the call to action in the industry is clear—storage management lags far behind the capabilities and importance of the storage devices being implemented.

SAN Devices

Whatever standards emerge for the storage components of DENs, the devices on the storage network will need to provide manageability to be compliant. For example, you may be familiar with network device elements providing a Management Information Base (MIB) to comply with SNMP standards. Over the past few years, we have seen storage and fabric elements provide a MIB, which will change slightly as the Common Information Model (CIM) and Web-Based Enterprise Management (WBEM) shift the format to using XML. Overall, look to the Desktop Management Task Force (http://www.dmtf.org) for definition of the standards.

Policy-Based Management

Just recently, we have seen progression and adoption in the area of policy-based SRM. The primary benefit of policy-based SRM is to define policies at the highest level in the organization and place storage within those policies; thus, reducing the effort and duplication of effort required to administer a typical storage environment. Policy-based SRM will have a vital role in virtualized enterprise storage, as the creation and application of that storage can fit under the same policy (as opposed to having two storage-management systems and hence two sets of policies to maintain). The latest generation of software (for example, InterSAN PATHLINE) helps to automate the entire process of device discovery, provisioning, monitoring, and control, including storage virtualization, as we discussed earlier.

Operations and Procedural Futures

Although constant change can be somewhat unsettling, there is comfort to be had in knowing that storage will continue to play a key role in the Windows strategy and storage resources will need to be managed. The difficult part is knowing that we must drive the rickety old stagecoach for a few more years before the shiny new locomotives appear to whisk us into the future. Just think about how the Internet changed what was possible from an application-development standpoint—everything from Smart Tags (or just hyperlinks) in documents that let you cross-reference a world of information, to peer-to-peer applications that let you create a virtual network of like-minded users. At this point, we don’t really know how changes in storage will transform the underlying applications, other than to say that we should need less duct tape and bailing wire to hold it all together.

The concept of partitioning disks and each application/server owning its little piece of disk real estate will vanish. And the next thing to vanish will be the old concept of each application/server owning its little piece of data, as multiple hosts will be able to pool that as well.

Storage Certifications

SNIA has been developing certification for storage administrators, implementers, and resellers focused on knowledge of Fibre-Channel SAN technology. The SNIA FC-SAN Certification Program certifications are classified by the following levels:

- Professional—Limited technical knowledge required; mainly understanding of terminology, principles, and purposes associated with Fibre-Channel SANs. Targeted at sales and marketing or other support personnel with limited exposure to Fibre-Channel SANs.
- Practitioner—More technical background in features, functions, and underlying technology of Fibre-Channel SANs; required for pre- and post-sales support, consulting, and field-service engineers.
- Specialist—Similar to Practitioner except adds the ability to plan, build, and configure a complex Fibre-Channel SAN. Substantial technical background required for system architects, consultants, and other full-time SAN support personnel.
- Expert—Most comprehensive of the technical levels, requiring the ability to analyze, diagnose, and troubleshoot Fibre-Channel SAN system problems.
- Master—Reserved for future use, awarded to innovators and architects in SAN technology as a recognition for lifetime achievement in storage networking.

Enterprise Backup Strategies

Another significant change to consider from an operational standpoint is disaster recovery in a networked-storage environment. As a result of the critical nature of recovering enterprise storage, we will use more advanced techniques to back up and recover. From a storage-management standpoint, the management tools that you use to ensure that your disaster-recovery procedures are operational will need to be updated. For example, if you are monitoring the Win2K event logs for successful completion of tape backup, you may need to instead watch for successful completion of the SCSI Extended Copy (serverless backup).

BCVs

BCVs are not a new or emerging technology so much as one that will gain wider acceptance and support. Both clones and snapshots are called BCVs, as they both offer the ability to perform much more rapid recovery than traditional restore mechanisms. Some people do not make a clear distinction between snapshots and clones, mainly because both methods reduce the data-recovery window. However, there is a crucial difference—a clone is a triple-mirrored copy of the data on an array of disk drives, so it can be broken off and set aside or used for other production work such as integrity analysis or data warehousing and information analysis.

By contrast, the snapshot is a hybrid of blocks from the original data and the new updated blocks. Setting aside the original files (such as a very large database) requires a substantial amount of disk processing for the host. For example, if you wanted to copy VLDB.dbf from disk E to disk Z on a database server, the action would also impact the ability of the host to respond to client requests. If you take the database offline and want to restore the original, the difference between snapshots and clones becomes clear. With a clone, you can merely swap the logical disk units, but with a snapshot, the original data is a virtual representation and would require swapping out the disk blocks, which is a very slow process when compared with merely replacing the files.

Consider also the difference in backing up the snapshot or clone data. With a clone, the broken off mirror set can be presented to the backup host for backup to a tape device with little or no impact to the database host. A snapshot can be presented, but the original host must service the blocks—again, as it is a virtual representation, a hybrid of the original blocks and the new ones.

In either case, snapshot or clone, the only way that the files on disk are useful for recovery is if the files are consistent to the application writing them. The application, such as a database or file server, could be in the process of writing file data or metadata (about which files are on the disk) from its cache. There are hardware solutions that can break off BCVs without any application knowledge; however, there is little guarantee that the clone or snapshot will be useful in recovery. Enter the introduction of volume snapshot services in.NET Server. The role of the .NET OS in snapshots or clones is to interact with applications and backup processes to ensure that I/O to disk does not interfere and make the data inconsistent.

Serverless Backup

Unless you have tape drives attached to every single application server, it is most likely that you back up your data over some kind of network. If the backup device is attached to a backup server, the data passes from one server to another over the network. In a SAN, the data can be directly transferred from disk device to tape device without involvement of the backup server. Typically, the backup server passes information to an agent running on the application server requesting that it mount a tape device and stream the backup to it. The most typical implementation of this type of backup uses the Network Data Management Protocol (NDMP) to move data from the application server to the backup device over the SAN instead of the network. NDMP was created by Legato Systems and Network Appliance and has been handed over to SNIA for future development.

In the next phase of evolution, true serverless backup, the backup server is not necessary. Some of you early adopters may even be using this new setup now, as the technology is available and will be maturing over the next few years. The SCSI-3 Extended Copy command lets a device execute a series of commands directly to devices. For example, copying data from a source device to a destination device without a server having to read the data from the source into memory and back to the other device. Of course, for this approach to work, the data needs to remain unchanged on disk or the files will be inconsistent between the two sets. Some software intermediary is still needed on the host server to pause I/O to disk during the backup process. In addition, this setup currently will not work in a Windows clustered environment because Windows environments use the shared-nothing storage design and each node owns the devices through disk reservations. I'm sure that Microsoft has considered this fact, and we will just have to wait and see whether the company changes its clustering model or develops another solution to allow SCSI Extended Copy for serverless backup.

Summary

This chapter wraps up *The Definitive Guide to Windows 2003 Storage Resource Management*. We looked into the future of storage and SRM including both hardware technology and software changes. The most immediate changes include technology that is currently available to you, but unless you are an early adopter, it will be a short time before you are using it.

Much of this chapter focused on networked storage, as clearly that is where the most improvement and increases in adoption will occur. In the area of hardware, we looked at changes in speeds and feeds as we get faster pipes and even greater distances. One of the upcoming changes is in virtualization of devices and storage. We looked at what that means from a storage-management perspective. We covered the server-side of storage networks, changes in HBAs, booting from the SAN, and multi-path I/O and what it will mean for performance and fault tolerance. In the area of disaster recovery, we looked at distance mirroring, cloning and snapshots, and serverless backup.

I also covered the next generation of storage technologies, and the features they may provide to enhance our storage management. Finally, we looked at the changes we will need to make from an operational and procedural perspective. All in all, it is both a frustrating and exciting time to be a storage administrator, and I wish you the best in your journey.

Appendix A: SRM Software and Hardware Vendors

Vendor	Focus	URL
Astrum (*now a part of EMC)*	Policy-based object management	http://www.astrumsoftware.com
NTP Software	Policy-based object management	http://www.ntpsoftware.com
BMC Software	Enterprise storage and application storage management	http://www.bmc.com
Brocade	Fibre-Channel switches and management	www.brocade.com
Compaq	Fibre-Channel SAN configuration, virtualization, and device management	http://www.compaq.com/storage
Computer Associates	Application storage management and backup	http://www.computerassociates.com
Dot Hill (SANnet storage solutions)	SAN configuration and management	http://www.dothill.com/products/software/sanpath.htm
EMC	Enterprise storage and device management	http://www.emc.com
Hitachi Data Systems	Fibre-channel SAN devices and management	www.hds.com
HP	SAN configuration and management	http://welcome.hp.com/country/us/en/prodserv/storage.html
IBM	Enterprise storage and device management	http://www.storage.ibm.com http://www.tivoli.com
McData	Fibre-channel switches and management	http://www.mcdata.com/

Vendor	Focus	URL
NetIQ	Application storage management	http://www.netiq.com
Sun Microsystems	Enterprise storage management	http://www.sun.com/storage
VERITAS Software	Fibre-channel SAN management and policy-based object management	http://www.veritas.com

Appendix B: SRM and Storage Web Sites, Portals, and Mailing Lists

Resource	URL
Storage Innovators	http://searchstorage.techtarget.com/tipsIndex/0,289482,sid5_tax287587,00.html
InfoStor	http://www.infostor.com
Search Storage	http://www.searchstorage.com/
Storage Management.org	http://www.stormgt.org/
Enterprise Systems	http://www.esj.com/
Storage Magazine	http://storagemagazine.techtarget.com/
Information Week	www.informationweek.com