AberdeenGroup

The Business Case for a Storage Virtualization Engine

An Executive White Paper

September 2000

Aberdeen Group, Inc.
One Boston Place

Boston, Massachusetts 02108 USA
Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

The Business Case for a Storage Virtualization Engine

Executive Summary

Many businesses have adopted storage area network (SAN) technology in order to centralize and consolidate data — their most valuable corporate asset. Now, having experienced the difficulty and high cost of managing the amassed storage, these early adopters, as well as others who have hesitated, are demanding that data storage become a centrally managed *utility*.

A storage utility provides everyone using it with access to data. In order to do that, a storage utility automatically and transparently allocates storage space to users on demand. In the utility model, storage is simply a resource to which a user can "plug in." Management of the resource is not the user's concern.

Most storage users seeking to create a storage utility will follow a SAN to storage utility path. However a storage utility does not require a SAN, and SAN storage may not be a storage utility.

Advancing Toward the Storage Utility

Storage virtualization is a key step toward the storage utility. By separating storage access and storage control functions within a storage architecture, fully functional storage virtualization should allow a storage solution to deal with data that spans disks, arrays, storage types, storage platforms, and storage suppliers *without* partitioning or zoning.

There are quite a few ways to implement storage virtualization. However, some of these are less effective than others at supporting the ultimate goal — the storage utility. Therefore, the IS buyer should assess a storage virtualization solution based not only on its immediate benefits, but also on its support for the oncoming storage utility architecture. By what storage-utility criteria, then, should the IS buyer judge the storage virtualization engine?

- A utility must be robust — that is, reliable under all circumstances. A utility must also perform to user demands and be able to scale to whatever load users place on it. Thus, a storage virtualization engine should provide *reliability* and *scalability* from a single array on up, across multiple existing storage devices, computing platforms, and storage suppliers as appropriate.
- The storage utility should be flexible: It should be independent of any brand of component that is attached to the utility, whether that component is a server of whatever type, a storage solution, or a network device. Just as important, the storage utility should be independent from the choice of network protocol that underlies the SAN. Then, and only then, is an enterprise free to aggregate its storage and consolidate its data both for operational efficiency and to best serve its customers, protecting and preserving its Information Technology (IT) capital investments in the process. In turn, the storage virtualization engine for the storage utility must be *flexible* at the storage-access level.

- Users must be able to manage a storage utility easily and cost-effectively. Because a storage utility will inevitably be the repository for the enterprise's mission-critical — and even business-critical — data, it will require special administrative care. However, as Aberdeen has noted in the past, storage is scaling exponentially, and is likely to continue to do so in the immediate future. If administration is not easy and cost-effective, so will the administrative people costs of the business-critical storage utility. Again, a storage virtualization engine designed for *manageability* can have a major impact on the storage utility's ease-of-use and cost-effectiveness.

What Storage Buyers Need, and What They Want

Faced with a major purchase or how-to-implement decision, storage buyers are forced to set priorities on the relative amounts of storage scalability, availability, robustness, and manageability they are willing to pay for. Peace of mind is likely to override all other concerns, however. The bottom line: No IT manager wants to answer questions about why a particular choice was made after a mission- or business-critical application has gone down.

This paper describes the state of the art in storage virtualization and considers the trade-offs involved with today's varied approaches, concluding with recommendations for prudent IS buyers and implementers.

Storage Virtualization: The Promise

Someday, storage virtualization will enable any end-user anywhere to access and, if necessary, share any file anywhere without regard to — or even knowledge of — the physical nature of the storage or its network, differences between operating platforms, other applications that may share or later use the same data, data security, or protection. To put it another way, storage virtualization aims to make an enterprise's entire storage look like a single common disk.

In order to achieve this "single view," storage virtualization:

- Separates data access (what outside applications, servers, and end-users want to do) from storage control (what administrators want to do); and
- Separates per-disk or per-array "physical" data storage from overall "logical" data storage.

In order to separate out and "virtualize" data access, storage virtualization must virtualize data location as well. In the broadest sense, full storage virtualization must support data that spans not only the disks in an array, but also arrays for different platforms, arrays from different suppliers, and even different storage types.

Note the way that this approach fits into SANs and the upcoming storage utility. While a SAN makes possible having a huge amount of data on-tap, the task of somehow managing all that data also expands. A SAN has any-to-any connectivity that enables virtualization and also enables (and can significantly reduce the cost of) flexible clustering topologies, which in turn can provide the highest possible availability.

What Storage Virtualization Does

In SANs, storage virtualization provides a bridge between the various data access mechanisms on each storage node in the network. It also allows the SAN provider to build tools that monitor data access and data location across the multi-platform or multi-supplier SAN much more easily. In effect, storage virtualization abstracts storage, separating out logical data access from physical access, no matter how the storage may appear to different operation systems of heterogeneous servers on a network. And storage virtualization works both unseen and dynamically, applying physical resources to create the logical storage users need, as they need it, and freeing resources to be available to other users as necessary.

Storage virtualization also simplifies life for storage managers by dynamically creating space for backup and by giving administrators a high-level view of the logical-to-physical resource mapping. The easier storage virtualization can make storage administration, the better off are storage buyers, because, as Aberdeen research shows, the growing cost of administering storage can easily negate the declining price of the storage. Also, Aberdeen research suggests that the explosion in content growth, and hence storage use, can outpace the more gradual growth in the amount of storage that can be managed per administrator. In addition, Aberdeen is finding that qualified storage administrators can be difficult to recruit and/or retain, even with top-dollar salaries. In a non-virtualized system, precious, scarce administrator time can be wasted in managing unallocated storage!

Storage virtualization therefore provides a strong foundation for overall SAN storage administration. And by providing a strong foundation for the SAN, storage virtualization implements a cornerstone of the storage utility. By acting as a repository for data-location information, storage virtualization will be the means by which the storage utility determines what data is being accessed and, therefore, how to account for its use.

Recognizing Storage Virtualization

Because storage virtualization is so promising, it will be promised — a lot. Therefore it is important to recognize what storage virtualization is *not*. Redundant array of independent disk (RAID) disks is not storage virtualization, although the mirroring or striping RAID provided by array controllers is transparent to the user. Dynamic storage partitioning is not storage virtualization — it is to virtualization as writing disk calls to memory overlays in a program is to virtual memory. Also, there are schemes that enable a SAN — which handles I/O on a block basis — to recognize files by using a metadata controller. This technique allows cross-platform file sharing on a SAN, but it is not virtualization.

Consolidating physical storage that had been separate and attached to, or within, individual servers into fewer storage units or a single unit provides some efficiencies and savings, but is also not virtualization, because the storage in the consolidated unit must be zoned or partitioned into storage that is only available to particular operating systems or even particular processors.

Storage Virtualization: The Reality

Storage virtualization performs three tasks:

1. It creates an abstraction layer that separates physical storage from logical storage;
2. It masks complexities to simplify storage management; and
3. It enables storage-resource optimization via pooling.

Although storage virtualization is a new concept, virtualization as such is not. Virtual memory is main memory backed by disk, and virtual tape is tape front-ended by a disk cache. Although virtualized storage does not have to belong to a SAN, SAN technology has placed storage virtualization in the spotlight.

At present, it appears that storage virtualization is just beginning to transition into reality. Storage virtualization suppliers are just beginning to create the software necessary to bridge multiple platforms, with multiple suppliers and storage types.

There is a direct analogy between storage virtualization and virtual memory: virtual memory makes memory plus disk storage in a system look like one storage pool, while storage virtualization makes non-memory storage — disk, solid-state disk and, quite possibly, tape and optical storage, too — across a network look like one storage pool.

Likewise, storage virtualization, like virtual memory, is engendering much discussion as to how to manage it, how to adjust legacy code to fit the new storage paradigm, and how to achieve optimum performance. However, we live in an era functioning in “Internet time,” and can expect storage virtualization’s incubation period to be shorter than that of virtual memory.

Where Storage Virtualization Should Be Implemented

A storage area network can be considered to have server and storage “edges” and a network “core.” Applications reside at the server edge, and physical storage devices sit at the storage edge. Thus, storage and server suppliers typically approach SAN virtualization from their edge of the network. Other companies, and some server/storage companies, too, are approaching storage virtualization from within the core of the network, where the devices that interconnect networked storage are found. Each approach offers the benefits of its focus — for example, network-focused storage virtualization is likely to allow especially good storage-network administration.

All other things — performance, scalability, and robustness, for example — being equal (and they may not be), equipment buyers may benefit most in the medium term from a network-centric approach, because a network-centric approach presents them with the greatest degree of vendor and technology independence. Of course, mileage may vary, and it is likely that over the very long term, these approaches will blend into each other.

Importantly, a network-centric approach to storage virtualization may not necessarily require that the storage being virtualized be part of a SAN.

The Business Benefits of Today's Approaches

Today's initial steps in storage virtualization can yield major business benefits if effectively implemented. The clear view of logical storage that storage virtualization can provide will reduce storage administrators' workloads and make their tasks simpler, with more predictable results. Therefore, storage virtualization will bring real cost savings to businesses whose IT organizations use it.

Storage virtualization also loosens the constraints on administrators in the first SAN generation. For example, storage virtualization solutions currently available greatly ease the task of SAN storage capacity planning and allow easy storage space reallocation without rebooting, or even halting an application that is executing. Storage virtualization solutions now in the field also allow an administrator to create backup and mirror space on disks configured nearly anywhere in a system.

The Long-Term Benefits of Storage Virtualization

Aberdeen notes that the growth in businesses' storage requirements is not being matched by a concomitant rise in employee headcount — and probably never will be. Moreover, Aberdeen research suggests that, over the long haul, the cost of managing data storage can significantly outweigh the cost of acquiring and installing storage in the first place. Storage virtualization aims to ameliorate this storage-to-personnel imbalance by simplifying storage management.

The sheer volume of storage managed per person has been growing at a rate that has made it nearly impossible for many people and organizations to cope. Simplified storage management can translate directly into less personnel staffing pressure.

Storage virtualization can also free users from becoming captive to a supplier or a technology, because virtualized storage is "agnostic" with respect to both. When a user can choose freely among suppliers, both prices and products offered tend to improve. The same can be said with respect to technology, when the "you can't do that because…" syndrome is eliminated.

A virtualized system of storage resources can be a part of a self-healing infrastructure that assigns storage and path-to-storage resources to applications on demand. Storage virtualization can provide end-to-end manageability, allowing storage systems that can be available constantly and scale to meet future needs.

First Principles for Storage Virtualization

Virtual storage should enable diverse computing platforms to make effective shared use of diverse storage devices. Storage management under virtual storage principles should give

Telephone: 617 723 7890
One Boston Place
Fax: 617 723 7897
Boston, Massachusetts 02108
www.aberdeen.com

administrators flexibility to decide and control how they will reconfigure the networked storage to meet dynamically shifting demands, and how they will distribute the data among the available devices. A storage utility can be created only if these capabilities — and reasonable tools for administrators to use these capabilities — are available.

A storage virtualization engine must accomplish the following separations:

- How storage is used from where it is. A storage virtualization engine should allow all types of storage (e.g., disk, tape, optical) to connect to the network and be shared among all systems on the network.
- Physical storage from logical storage. File systems and databases — i.e., storage objects — cannot be encumbered by physical limitations as to how they may be distributed.
- User availability from administrative housekeeping. Replicating and/or relocating a storage object are housekeeping tasks that should be unseen to users.

Today's Approaches to Storage Virtualization

At present, the approaches to storage virtualization being taken fall into three main categories:

1. Host-based — with the ability to couple tightly with an application (say, an online transaction database) and/or drive additional availability features such as multi-path or failover software;
2. Storage-based — often with the benefit of unified support from a single vendor;
3. Network-centric — with possibilities for lowering virtualization's cost, working with the greatest range of server or storage hardware, and affording the easiest way to change or replace network components.

Most advantages, however, involve tradeoffs, as the following descriptions of approaches to storage virtualization explain.

The Host-Based Approach

A number of companies that offer both servers and storage in their product lines take a host-based approach to storage virtualization, supported by their own software and/or software from third-party vendors (Figure 1).

This approach can be a safe and generally works well, as long as the buyer follows the supplier's prescription on how to configure and use the SAN. However, there may be limited — or no — interoperability with hardware or software components that are not defined by the vendor as part of the "end-to-end SAN solution."

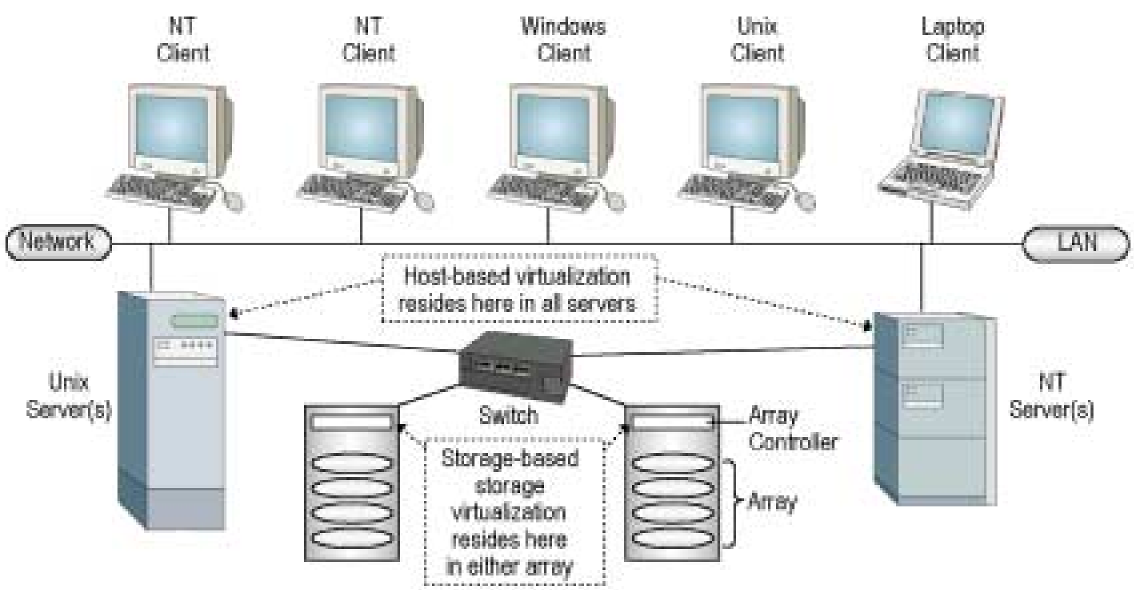
In the host-based approach to storage virtualization, all software must be compatible with the host's operating system. The host operating system typically makes no clear demarca-

Telephone: 617 723 7890
One Boston Place
Fax: 617 723 7897
Boston, Massachusetts 02108
www.aberdeen.com

tion between the host and its storage, and the operating system causes the host to behave only as if it is in communication with a particular device type. The setup will definitely work easiest if all of the host computers are the same. In many host-based approaches, the hosts are responsible for the integrity of the storage SAN-wide. Therefore, centralized management is very important. When hosts sharing SAN storage are managed independently, an inappropriate management action or integrity compromise at one host can impact the integrity of all of the storage on the SAN.

Mainly for these reasons, storage service providers (SSPs) are tending to shy away from the host-based approach to storage virtualization. After all, they are in the storage-as-a-utility business and wish to provide storage to a diverse set of customers whose applications may be just as diversely hosted. The host-based approach

Figure 1: Host- and Storage-Based Approaches to SAN Storage Virtualization



Source: Aberdeen Group, September 2000

to storage virtualization will be acceptable to users determined to remain committed to maintaining homogeneous servers and storage because they may be able to obtain their virtualization soon, with single-source support.

The Storage System-Based Approach

Companies whose product lines consist only of storage tend to take a storage-system-based approach to storage virtualization, supported predominantly by their own software and augmented with software from third-party vendors.

As with the host-based approach, the storage-system-based approach to storage virtualization tends to be a safe course that works well with vendor-defined solution sets. A well-engineered storage-based product may even perform optimally, given particular components. After all, optimization (like a five-speed manual transmission on a sports car), and generalization or simplification (such as a sedan's automatic transmission), are often natural enemies. At the same time, IS buyers should note that scalability for the storage-system-based approach to storage virtualization is tightly defined by how scalable the vendor's product line is.

Also, because of closeness to the actual physical disks, storage virtualization implemented at the physical storage level can make the most of the disks already in racks and protect a user investment in disks by extending their useful life.

However, by far the biggest drawback to the storage-system-based approach to storage virtualization is vendor lock-in. Enterprises must therefore decide whether to trade-off equipment cost or vendor lock-in for relative simplicity and single-source support.

Network-Centric Approaches

Network-centric approaches to storage virtualization fall into three classes:

1. Appliance-based, using a commodity server as the intelligent appliance. There are two varieties to the appliance-based approach:
    - Symmetric, whereby the virtualization commands and data use the same path; and
    - Asymmetric, whereby the virtualization commands and data use separate paths.
2. Switch-based, whereby a conventional switch is *not* used, but rather replaced by a super-intelligent switch.
3. Router-based, whereby instead of the simple protocol-translating "routers" used to connect, say, SCSI to Fibre Channel, *intelligent* routers in the storage network handle the logical/physical storage separation and can also be "wire-agnostic" — that is, communication-protocol-independent.

Intelligence — software that can adapt to the particular characteristics of a virtualized storage network — is the critical success factor in all three approaches. Intelligence is what will make storage virtualization happen, and whoever controls intelligence dominates. It is noteworthy that some SSPs are considering or actually using network-centric approaches to storage virtualization.

The Appliance-Based Approach

Under the upcoming SAN appliance, or "SAN metaserver," approach to storage virtualization, access and control are separated because metadata — i.e., data about the data — resides on the SAN appliance, to which all I/O requests from the other servers are initially directed. The SAN appliance responds with metadata that specifies the particular SAN storage resources that will be used for the I/O operation, and initiates the operation.

The SAN appliance approach to storage virtualization has two flavors (Figure 2):

1. Symmetric, whereby data and the metadata used to control data access use the same path — that is, the SAN appliance "sits" in the data path; and
2. Asymmetric, whereby the paths used for data and metadata are separate. Hosts send metadata to the asymmetric SAN appliance via a conventional LAN.

At this time, symmetric SAN appliances are closer to delivery than asymmetric.

It will easily be possible for buyers to confuse the symmetric/asymmetric distinctions in the appliance-based approach to storage virtualization with in-band and out-of-band SAN management. The virtualization metadata control and SAN management messages are different, although the control/data path distinctions are similar.

There are a couple of important points to keep in mind about the appliance-based approach:

- If the disks attached to a SAN using the appliance-based approach to storage virtualization are JBOD (just a bunch of disks — that is, non-RAID), host computers become responsible for providing software RAID data protection if it is desired, and that can impact performance dramatically.
- A symmetric SAN appliance can only virtualize the storage to which it is directly connected. Multiple symmetric SAN appliances create multiple "islands" of virtual storage, not a pool of virtual storage across the entire SAN.

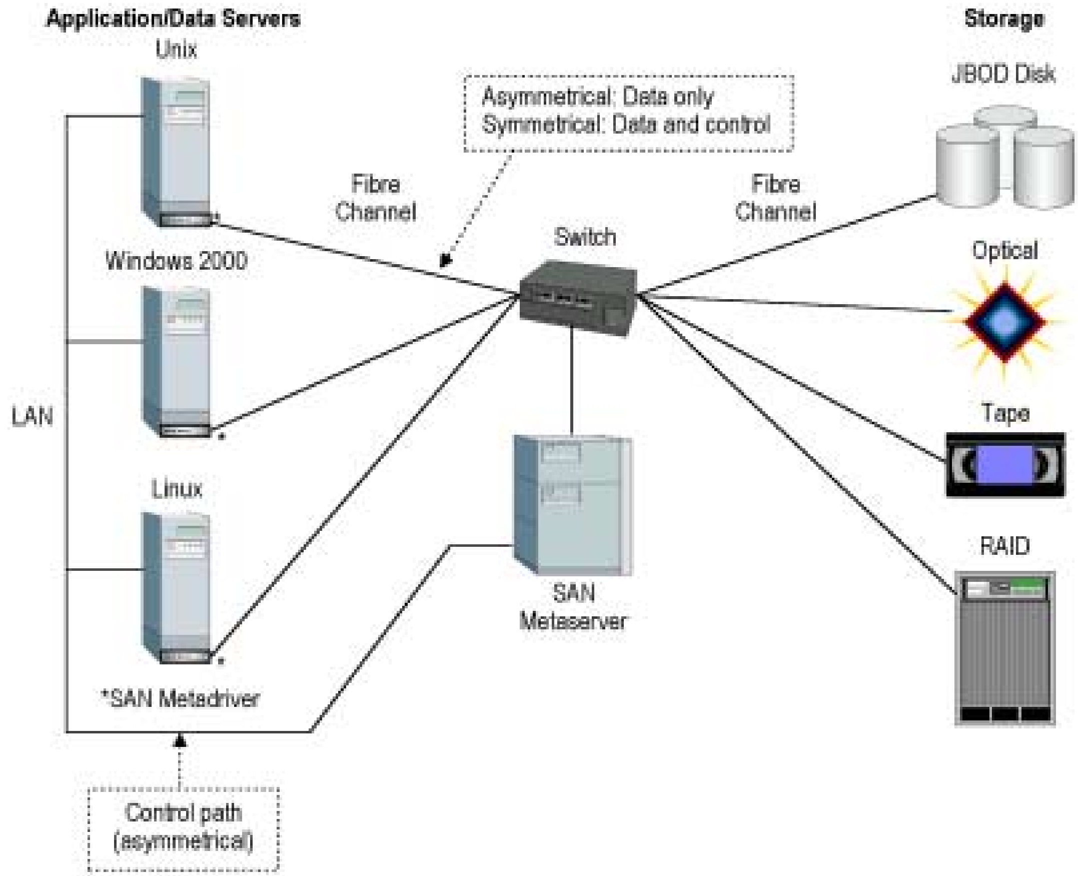
An appliance-based approach to storage virtualization requires an agent to redirect I/O requests to the metaserver resident on every host server on the SAN. The server virtualization provider is responsible for porting agents to hosts or host bus adapters (HBAs) to meet this requirement. Or, the client virtualization agent could be implemented on the SAN HBAs. An HBA in a SAN is analogous to a network interface card (NIC) in a LAN. A NIC connects a host to a LAN, and an HBS connects a host to a SAN. Implementing the client virtualization agent on HBAs would eliminate the need to port the agent to the various host operating systems, but then special HBAs would have to be developed and qualified by the SAN appliance vendor.

Scalability in appliance-based approaches to storage virtualization varies with type. The symmetric type could turn out to be less scalable than the asymmetric because all data paths and all control paths must route through the appliance.

Performance in appliance-based approaches to storage virtualization also varies with type. The asymmetrical type will often perform slightly better, since the data path need never become backed up, waiting for control messages.

Telephone: 617 723 7890
One Boston Place
Fax: 617 723 7897
Boston, Massachusetts 02108
www.aberdeen.com

Figure 2: SAN Appliance-Based Approaches to Storage Virtualization



Source: Aberdeen Group, September 2000

The Switch-Based Approach

A switch supports multiple simultaneous point-to-point connections in a network. Usually switches are limited in their capability to the switching function. In the switch-based approach to storage virtualization, the switch is enhanced to include storage virtualization capability, creating a costlier switch/storage virtualizer
(Figure 3).

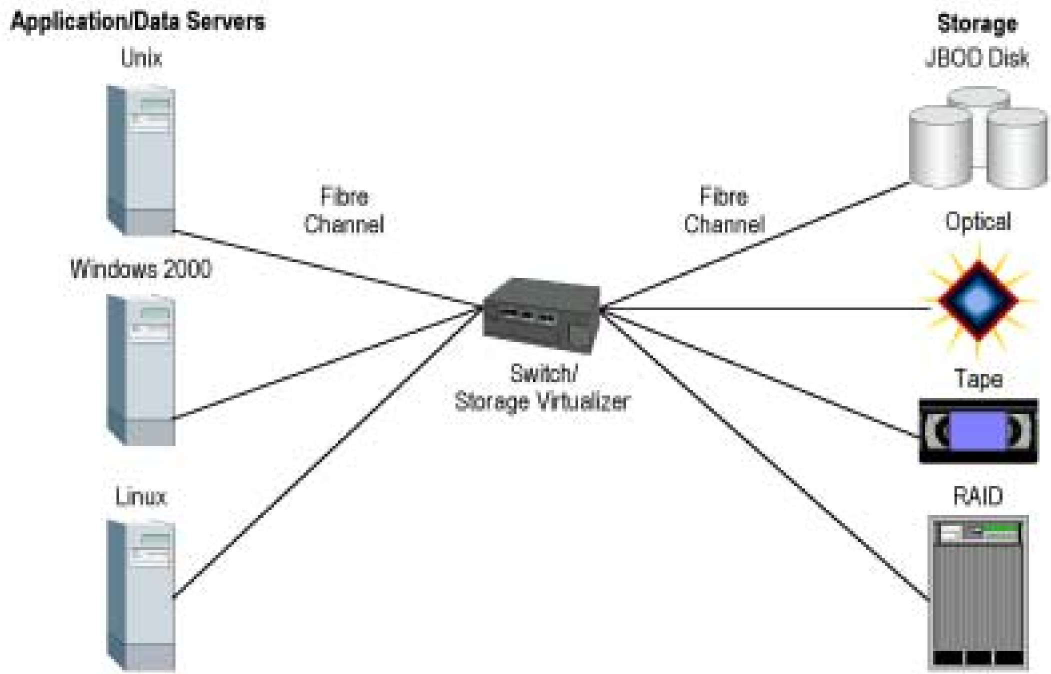
A switch/storage virtualizer can be implemented in software that runs on a SAN switch, either a specialized switch or a dedicated computer. A vendor using the switch-based approach to storage virtualization may sell both the computer and software, or sell only the software and permit the buyer to select his or her own computer.

A switch-based approach to storage virtualization does not necessarily require host-based software agent intelligence or RAID storage. However, if RAID functionality is demanded of JBODs, either the switch/storage virtualizer or the host computers must provide performance-impacting software RAID.

Telephone: 617 723 7890
One Boston Place
Fax: 617 723 7897
Boston, Massachusetts 02108
www.aberdeen.com

One advantage of the switch-based approach to storage virtualization is that it is possible to implement inexpensive caching by using the memory of a switch/ storage virtualizer, which some vendors in this arena do.

Figure 3: A Switch-Based Approach to Storage Virtualization



Source: Aberdeen Group, September 2000

The switch/storage virtualizer runs an operating system that may or may not be embedded and invisible to the user. Introducing a different operating system into an existing configuration tends to raise a red flag. Will the switch/storage virtualizer be as robust and reliable as it is expected to be? This is not a simply answered question. While the commodity operating system may be sufficiently robust itself, there is generally nothing to prevent additional applications being placed on the switch/storage virtualizer, and an additional application or application user can introduce reliability problems.

When users are employing a switch-based approach to storage virtualization, using redundant switch/storage virtualizers can improve robustness. To economize over full “2N” redundancy, wherein every component must be duplicated, the redundancy can be N+1-based.

Note that the switch-based approach to storage virtualization can pose performance problems. A switch/storage virtualizer is basically a gatekeeper. A request to a gatekeeper is handled in one of two ways: Either the gatekeeper handles the request itself and presents the results to the requester, or the gatekeeper authorizes the requester to pass through the gate and access the storage. Either method involves delay, and optimal performance can be difficult to achieve.

The Router-Based Approach

The Storage Networking Industry Association (SNIA) does not define a "router." In usual networking parlance, a router chooses the path a signal may take through a network of switches and is also capable of translating one protocol into another in the process of passing a signal through.

In the lexicon of generalized networking, a "bridge" cannot route a signal, but only perform protocol-to-protocol translation. SNIA defines a "bridge controller" as what a bridge is understood to be in generalized networking. However, some vendors in Fibre Channel markets offer what SNIA defines as a "bridge controller" as a "router."

For this paper, "router" should be taken to mean what it does in generalized networking, plus a bit more. The router in the paper's context is an intelligent device that is also capable of hosting storage virtualization software (Figure 4).

A router-based approach to storage virtualization would include a programmable interface that can accommodate interfacing varied equipment connecting diverse protocols — Fibre Channel (FC) no matter how "standard" or "specially-featured," Fibre Channel Arbitrated Loop (FC-AL), Serial Storage Architecture (SSA), Small Computer Serial Interface (SCSI), and others — such as InfiniBand or IP-based devices that will emerge in the market. The router would translate the protocols used by connected devices into a common network protocol of choice.

A router-based approach to storage virtualization could create a network that interconnects heterogeneous hosts with heterogeneous storage devices — RAID or JBOD disks, SAN-enabled disk subsystems, and tape units. That is, a router can combine a diverse, heterogeneous mix of storage devices and host computing platforms using different communications protocols into a single effective SAN.

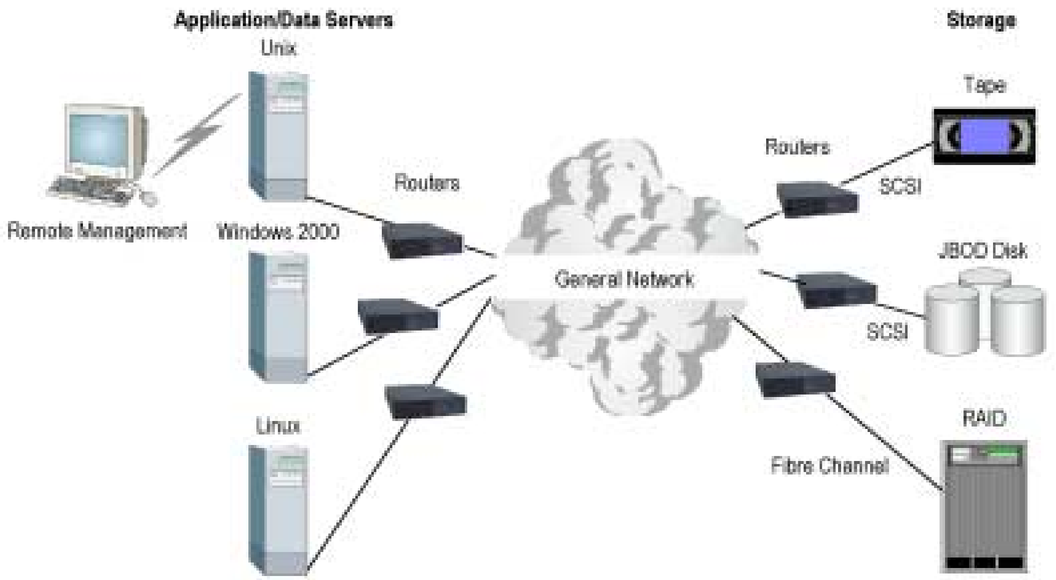
A router-based approach to storage virtualization could also include a SAN-like management system that allows storage administrators to reallocate storage resources using a clear graphical user interface (GUI) run from one convenient location — say, a workstation or laptop on the LAN — or remotely via a TCP/IP connection.

The router-based storage configuration scales simply by just adding routers and nodes. Since each node is connected to the network, the only possible bottleneck is in the network itself. Individual nodes can be price-performance tuned or adjusted separately, and heterogeneity is managed on a node-by-node basis — an individual router can connect multiple heterogeneous servers or storage types having a common protocol. The routers communicate in-band over the network for coordination and centralized management. The solution combines scalable hardware with centralized software.

Telephone: 617 723 7890
One Boston Place
Fax: 617 723 7897
Boston, Massachusetts 02108
www.aberdeen.com

The router-based solution is equally performance scalable, and would lend itself well to, for example, an Enterprise Resource Planning (ERP) application that employs multiple hosts and databases. Host or database storage nodes can be added easily and economically to scale to required performance.

Figure 4: A Router-Based Approach to Storage Virtualization



Source: Aberdeen Group, September 2000

Aberdeen Conclusions

There is no doubt that storage virtualization will yield major benefits, especially in the key area of administrative cost savings. The major remaining question is, What approach should the user take?

Where the intelligence behind storage virtualization belongs — in the host, at the storage, or in the network — is a hotly debated topic. Each of these approaches is appropriate for some users. A user's answer should depend on the stage that the user's storage architecture has reached in its evolution toward a SAN, and on the requirements of the user's solutions (typically they require specific storage environments). Moreover, users should consider that at any given time, some technologies are more mature, and thus safer, than others.

Host-based storage virtualization can be fairly easily configured and generally is well supported, especially when the host computer vendor provides the storage. Storage-based storage virtualization is also relatively straightforward and simple when using storage from one vendor, especially when consolidating. Host- and storage-based solutions that are available now are making storage more open and interoperable than ever, providing greater benefits for a widening circle of users.

Switch-based storage virtualization can integrate non-SAN assets into a storage pool, but a computer used as a switch will probably never achieve the performance of a device designed to be only a switch, and this solution always involves multiple vendors. Appliance-based solutions may offer the safety of single-vendor sourcing and support, but the symmetric flavor may not scale especially easily; and the asymmetric type, while promising good scalability and performance in its design, is not yet being delivered. The router-based solution to storage virtualization is in delivery and should perform the virtualization function well for those users who are prepared to implement a multi-vendor system.

Aberdeen also recommends that where several approaches are viable users choose the most powerful, yet flexible and open, storage virtualization approach, while ensuring that they never jeopardize mission- or business-critical storage.

Aberdeen Group, Inc.
One Boston Place
Boston, Massachusetts
02108
USA

Telephone: 617 723 7890
Fax: 617 723 7897
www.aberdeen.com

September 2000