The Complete Reference

Storage Networks

Chapter 2

The Battle for Size and Access

Chapter 2 will begin to define the general concepts of the storage networking model and the differences between it and the client/server storage model. We'll explore the details of how application design and system implementation influences storage infrastructures. With these discoveries, we'll begin to demonstrate the benefits of storage networking and how it meets the challenges posed by traditional client/server storage implementations as discussed in Chapter 1. By understanding the characteristics of business and support applications, the inherent value in putting storage on the network begins to take shape. Using standard IT planning activities and the upfront knowledge of application-oriented issues, integrating storage networking design benefits prior to system implementations will greatly increase an IT organization's ability to develop a successful storage infrastructure.

Client/server storage continues to use an I/O bus connectivity standard to communicate with storage devices, both from an internal and external perspective to the server. The I/O bus standard has evolved from original ISA bus configurations (standard to early PC servers) to today's Peripheral Component Interconnect or PCI. The server uses the I/O bus to connect with a storage device; however, it then communicates with the device through a device-specific language or protocol.

These range from ATA and ASA device protocol standards to SCSI, the Small Computer Storage Interface protocol. SCSI is the most popular among device vendors and customers for use with externally connected storage devices because of their capability to support multiple devices and more sophisticated data protection strategies such as RAID (Redundant Array of Inexpensive Disks). It is important to distinguish between the bus and device communications standards and methods because device communication protocols, like SCSI, and data protection architectures, like RAID, will continue to be used within storage networking solutions. However, as we'll discover, the I/O bus connectivity model will give way to network connectivity architecture. (More detailed information on the I/O bus, bus standards, and device protocols can be found in Chapters 6 and 7.)

As indicated in Figure 2-1, the storage devices were directly connected to the server, regardless of the type of protocol. This required all communications with the data stored on the storage devices to be handled by the server. As we have seen from examples of data warehousing and web site data movement, this places a tremendous amount of overhead on the server. Rendering it, as we have seen, to further specialization just to handle the overhead of dealing with application data.

Storage networks changed this entire model by creating unique network-based storage architectures that enhance existing client/server computing topologies. By allowing storage devices to be connected directly to an existing network, or through its own (storage) network, creates two important changes to traditional storage configurations:

- By providing more direct access paths between storage devices, servers, and the clients, the user transactions can go more directly to the data, bypassing much of the overhead of I/O operations and unnecessary access to and through the server.
- Storage networking grants business applications greater access to data with increased efficiency. In other words, storage networking makes it easier for applications to share data, providing servers the capability to connect to larger amounts of data.

Storage networking technologies have evolved into two distinct models—Network Attached Storage, commonly referred to as NAS, and Storage Area Networks, referred to as SAN. NAS allows storage to be placed onto an existing client/server network based upon Ethernet standards utilizing the standard TCP/IP network protocols. SANs, on the other hand, create a unique network just for storage and are based on a protocol called Fibre Channel. (More information can be found in Chapter 16.)

Both NAS and SAN are depicted in Figures 2-2 and 2-3, respectively, showing enhancements to the client/server computing model.

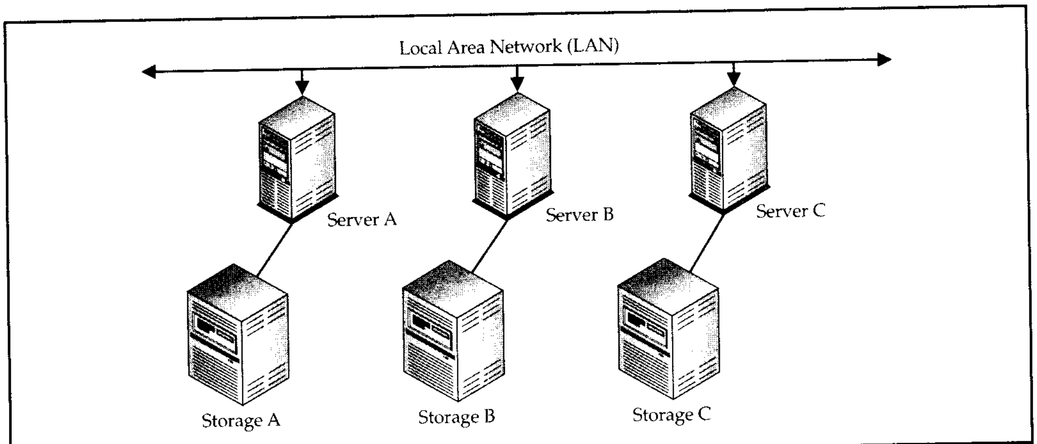

Figure 2-1. *Directly connected storage in traditional client/server environments*

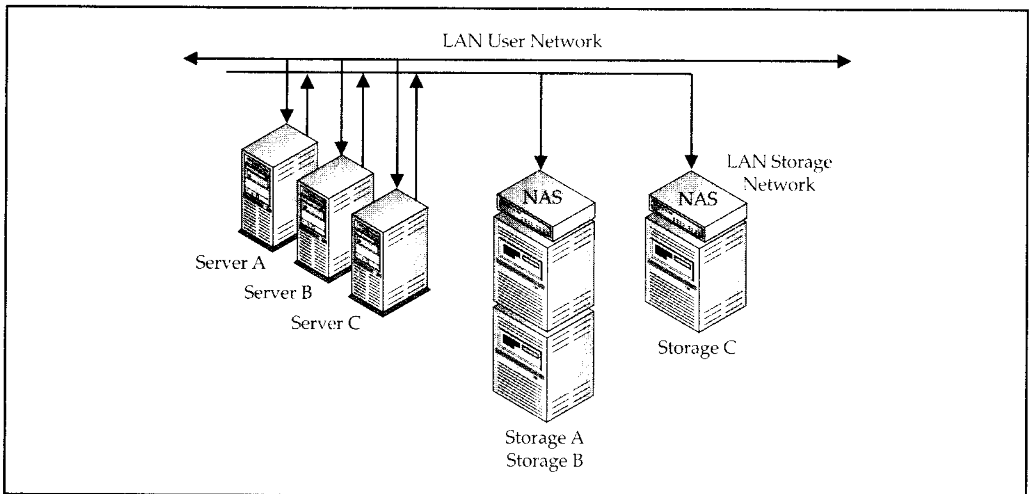

Figure 2-2. *NAS example showing enhancements to data access*

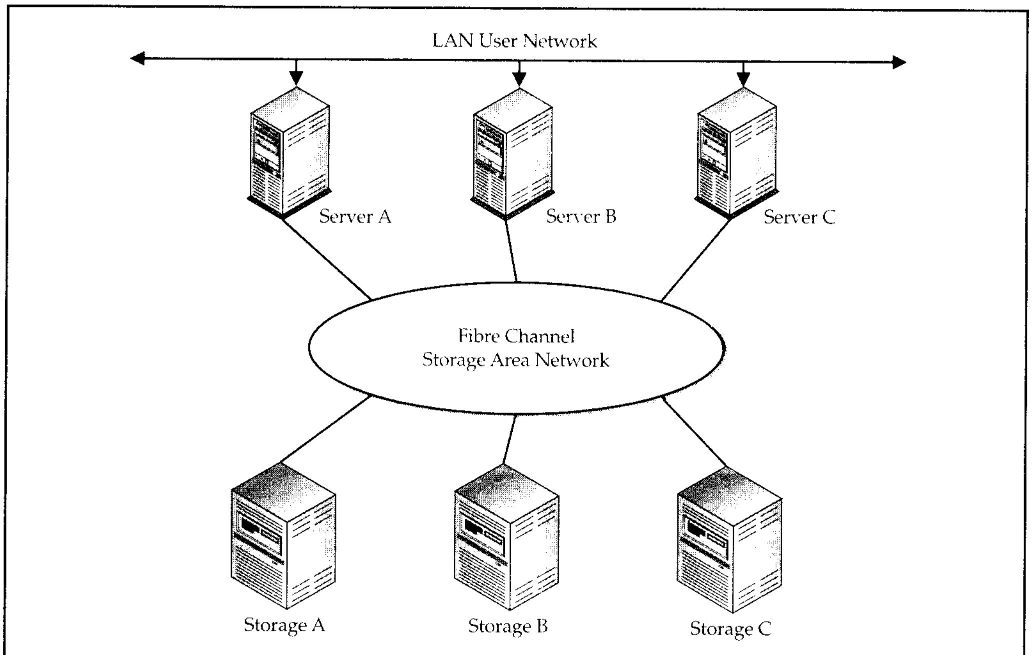

Figure 2-3. *SAN example showing enhancement to data size*

Metro Data Areas—A Look to the Future of Storage Infrastructures

In the future, distributed computer systems will increasingly take on characteristics of our urban and transportation infrastructures. Our data is stored in large server farms that produce circumstances and problems similar to those in a large metropolitan area, managing traffic flow in and out, policing and monitoring performance and security, and handling overcrowding and sprawl and the occasional accident.

Fueled by powerful servers connected to larger and larger storage systems, the Internet will continue to evolve into a complex array of data highways that literally crisscross the globe. This complex network of highways allows users to access data regardless of location, application logic, or specific storage system. These conditions will accelerate the growth of concentrated data areas that serve both public and private data highways (e.g., Internet and intranet networks).

Metro Data Areas will be located in corporate data centers, Internet service provider infrastructures, government and public institutions, and small business computer configurations. Research indicates that Metro Data Area evolution is being driven by an almost 100 percent annual growth in online storage. The Metro Data Area will be driven into the next generation of storage technology in an attempt to deal with this congestion and scalability dilemma and to construct a data transportation system that will handle the tremendous amount of data moving in and out of these areas.

Be prepared for the rush hour commute. How long will it eventually take to get your data?

Why Storage Networking Makes Sense

Many IT support organizations comment on their applications, saying, "We have bigger data, more sources, and need a single distribution strategy." In the eyes of IT, this translates as:

- **Bigger data** The data stored online and transferred between storage and server has increased in size, and the amount of data transmitted between server and client PC is much larger, driven by structured data (text and numeric data) combined with unstructured data (images, audio, and video).
- **More sources** The application must work on several sources of data to satisfy the client transaction. This means there are several online storage units that the server must connect to process the application.
- **Single distribution strategy** The results of the application need to be placed in a central location for access. Generally, this means one thing: Internet accessibility.

The Problem: Size

IT organizations also discuss ways in which they will address these problems. A typical snippet of conversation might sound like this: "We must have wider bandwidth for data transfers from storage. The problem is size. The databases we work with are exceeding the limitations of the server's storage connectivity." This translates as:

- *Wider bandwidth is needed.* The connection between the server and storage unit requires a faster data transfer rate. The client/server storage model uses bus technology to connect and a device protocol to communicate, limiting the data transfer to about 10MB per second (maybe 40MB per second, tops).
- *The problem is size.* The database and supporting online storage currently installed has exceeded its limitations, resulting in lagging requests for data and subsequent unresponsive applications. You may be able to physically store 500GB on the storage devices; however, it's unlikely the single server will provide sufficient connectivity to service application requests for data in a timely fashion—thereby bringing on the non-linear performance window quite rapidly.

Solution Storage networking enables faster data transfers, as well as the capability for servers to access larger data stores through applications and systems that share storage devices and data.

The Problem: Access

Others in IT may argue the point, saying, "Actually, the real problem is access. We don't have sufficient resources to access the application server. This will only get worse if we go to a single distribution strategy." This translates as:

- *The problem is access.* There are too many users for the supported configuration. The network cannot deliver the user transactions into the server nor respond in a timely manner. Given the server cannot handle the number of transactions submitted, the storage and server components are grid locked in attempting to satisfy requests for data to be read or written to storage.
- *The single distribution strategy needs revisiting.* A single distribution strategy can create an information bottleneck at the disembarkation point. We will explore this later in Parts III and IV of this book where application of SAN and NAS solutions are discussed. It's important to note, however, that a single distribution strategy is only a logical term for placing user data where it is most effectively accessed. It doesn't necessarily mean they are placed in a single physical location.

Solution With storage networking, user transactions can access data more directly, bypassing the overhead of I/O operations and unnecessary data movement operations to and through the server.

As we have demonstrated, albeit at a very high level, storage networking strategies can address each of these issues and make application strategies like single distribution a successful reality. However, this book's definition of storage infrastructures encompasses the entire IT experience, including both business and overhead applications. To understand storage networking technology and its benefits, it's important we define a few terms, especially those regarding applications that have evolved to drive the storage configurations.

Business Applications Defined for Storage

Business applications, or, more specifically, *enterprise* business applications, will be most effected by and find value in storage networking. In order to completely understand this phenomenon, we must first define an enterprise business application, as well as differentiate it from maintenance applications and support programs processed within IT organizations.

Microsoft defines enterprise business applications as complex, scalable, distributed, component-based, and mission-critical applications that may be deployed on a variety of platforms across corporate networks, intranets, or the Internet.[1] Enterprise business applications are datacentric and user friendly, but must meet stringent requirements for security, administration, and maintenance. Beyond these common qualities, Microsoft further characterizes enterprise business applications by highlighting three specific attributes, which include:

- **Large** A long-lived, multiuser, multideveloper, multimachine, multicomponent application that can manipulate massive data and utilize extensive parallel processing, network distributed resources, and complex logic, as well as be deployed across multiple platforms and interoperate with many other applications.
- **Business oriented** To meet specific business requirements, the application encodes business policies, processes, rules, and entities, is developed in a business organization, and is deployed in a manner responsive to business needs.
- **Mission critical** An enterprise business application must be robust enough to sustain continuous operations, be extremely flexible for scalability and deployment, and allow for efficient maintenance, monitoring, and administration.

Further noted is the difference in scope between enterprise applications—for example, multiuser, interactive, highly available architectures, and personal applications intended for single-user interactions. Although many PC applications function as the client in client/server architectures, and active code is commonly distributed to our client through Java or Java-like applications, they lack the distributed nature and high availability requirements previously stated.

---

1 Reference Microsoft Publication, Visual Studio - Developing for the Enterprise, Document No. DTX03-55298-0598

Maintenance and Support Applications

Supporting large, highly available business applications and their infrastructures are numerous maintenance and support applications. These are implemented within support infrastructure areas. IT management must use these applications, often called tools and utilities, to maintain the mission-critical characteristics of enterprise applications. (More information on management practices and processes can be found in Chapter 21.)

It is important to differentiate between these two distinct types of applications, especially given the often confusing and misleading global use of the term *application* among software and hardware vendors. This difference is important due to IT organizational boundaries that focus IT specialties between the programming and design personnel, the people who develop and implement the business applications, and systems support and administration personnel—the people who support the infrastructure.

The maintenance and support applications have evolved to demonstrate the same non-linear performance that business applications do (see Chapter 1 for a refresher on this). Therefore, they can perform just as poorly and be just as degrading to the operations of the data center as their business counterparts. The support applications have the following distinct attributes:

- Their results are focused upon supporting and maintaining server platforms and components that create a continuous operational infrastructure.
- They enhance IT's abilities to maintain levels of reliability, availability, and serviceability for the data center and evolving Metro Data Areas.
- The usage of support applications provides necessary information to maintain a scalable and manageable data center and Metro Data Area infrastructure.

The tools and products within these areas defined as storage management applications can be some of the most datacentric within the data center. They are defined by the following broad categories:

- **Archive/Backup/Recovery** Provides IT with the tools to initiate data protection policies that copy data, provide recovery mechanisms, and move data to offline storage media.
- **Storage Resource Management** Enhances the usage of resources that make up the storage infrastructure. These include utilities such as monitoring the activity of stored data, mechanisms, and the enforcement of online storage space quotas, as well as the collection of historical I/O system information and status.

- **Storage Data Management** Complements the storage infrastructure with enhancements to file/systems, volume management strategies, fault tolerant mechanisms, and utilities such as online space data maintenance activities, defragmentation, optimization, and compression.
- **Storage Security** A new category that has extended into the storage area from network user authentication and resource access technologies.

Although storage management applications are necessary in providing an organized and manageable environment, they are not business applications. The characteristics and base purposes of both are very different. While business applications automate and ultimately provide the process whereby the business runs (as well as directly impact overall business performance), storage management applications are considered "necessary evils" and "maintenance overhead" to business application processing and availability. Make no mistake, the "evils" of support tools and practices are a necessary cost of doing business, and when not managed effectively, they will impact business performance.

Building the Business Application

In trying to bring more and more data into play, application designers typically place things closest to the user. However, as we learned in Chapter 1, if this is done without regard for storage or data infrastructures, the application can be quickly compromised once in production. The application design process (as shown in Figure 2-4) will then give way to the realities of system implementation (as shown in Figure 2-5).

The application designers concentrate on the business requirements and produce a requirements document. This determines what will be provided to end users at the completion of the development process. The document, once finished and signed off by management, will also be used as a basis for an external design document. This is generally comprised of an external view of end-user functions, access points, and outputs. It's possible that sources, types, and quantities of data will be factored into external design discussions. This is followed by internal design activities where the application programmers determine the logic and identify external software development components to be used in building the application system.

It is during the internal design process that application designers and programmers work with database administrators to provide a logical design for the application data. Unfortunately, in many cases the physical attributes of the database requirements remain unaddressed. These design activities are followed by an iterative process of actually building (for example, coding) the application by a team of application programmers.

As components are completed, they are tested on their own—something known as unit testing—and then integrated with other components to form a major function

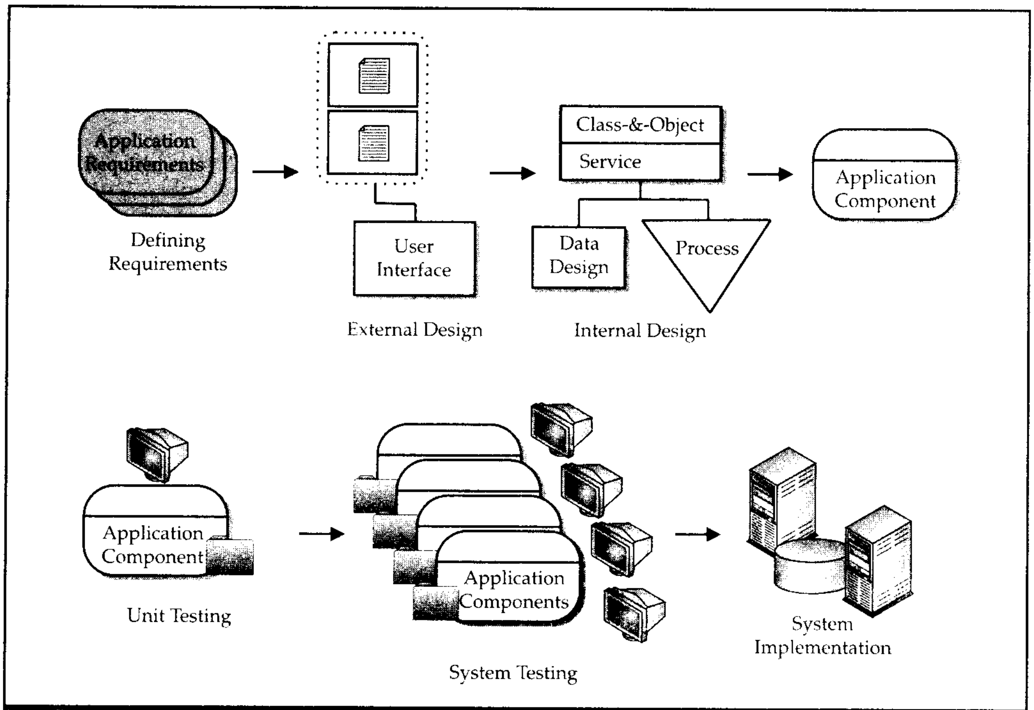

Figure 2-4. *The application development lifecycle*

of the application system. These functions or combined sets of components are integrated and tested in what evolves into a larger process of system testing. As all the components come together, user testing and acceptance begins, where actual end users test the system for accuracy and functionality. The components are modified in an iterative process as functionality errors are encountered. Once the system is completely tested by a segment of end users and signed off by management, the system moves into production mode. This is where the application first encounters the storage infrastructure.

Moving into production mode is a tricky business. In large systems, other activities come into play such as user education, migrating users from older systems into new systems, and ensuring that the application's business data stays in sync from one system to the next.

However, as the new business applications progress to this point, implementation will produce some interesting challenges, especially in the absence of storage infrastructure information and planning. Now it's a matter of how the application works with an

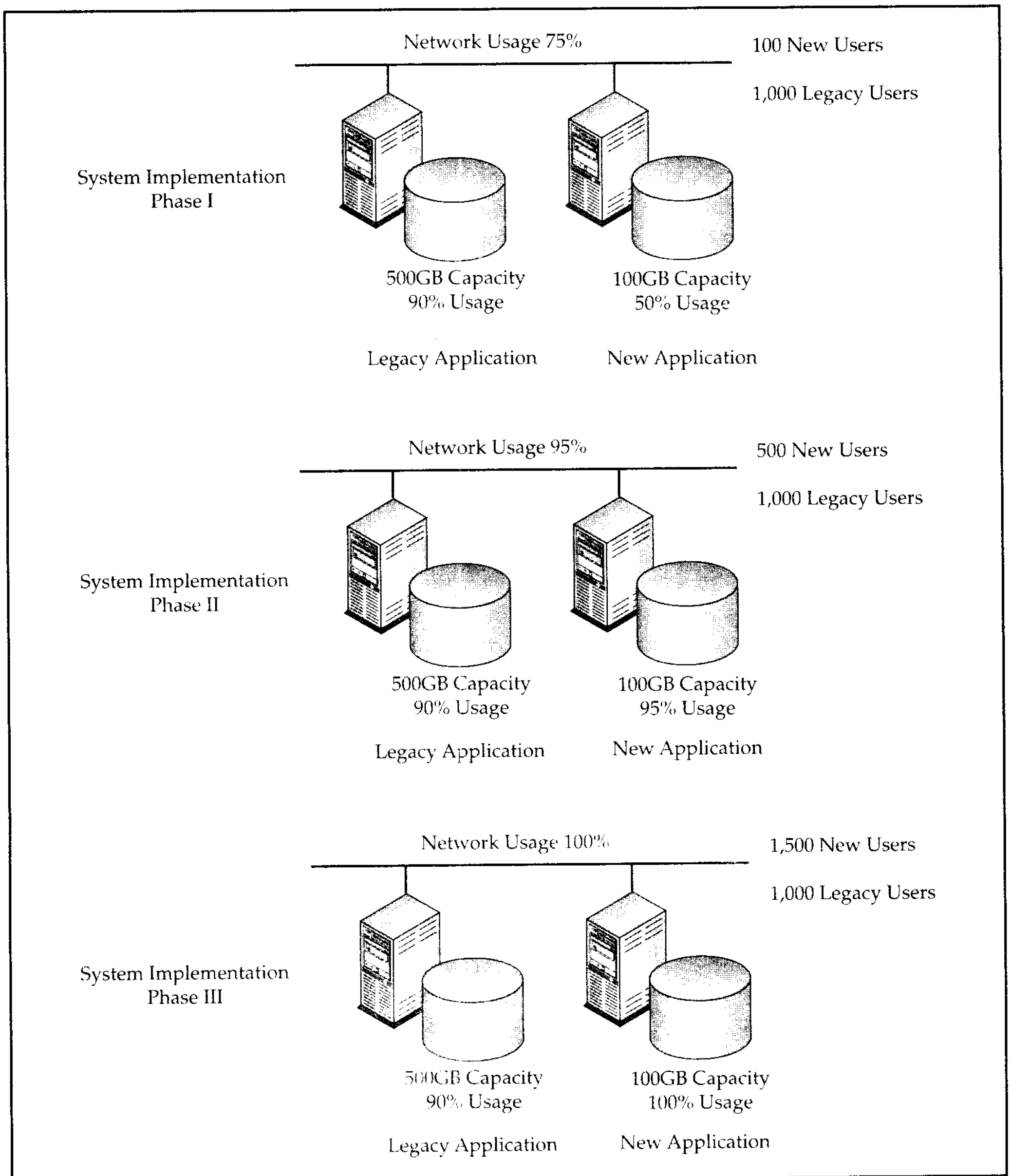

Figure 2-5. *Application implementation challenges*

unknown server and storage infrastructure. If the storage infrastructure has not been taken into account, the window of application non-linear performance (see Chapter 1) becomes very thin and quickly results in an unresponsive application as soon as users and data are applied in production mode. (See Figure 2-6.)

A third process in the application development cycle, albeit unplanned, is a post-implementation analysis and understanding of the real storage and data access requirements of the application system. This generally results in quickly provisioning additional resources into the storage configurations to handle the production processing. Due to the immediacy of this action, the resources are provisioned at a macro level without accounting for growth or scalability.

Obviously, the third process should be integrated into the application's development process, probably somewhere between the external and internal design activities. However, this requires two major pieces of information. First is the availability of storage solutions to offer to the application designers. Secondly, a more detailed understanding of input and output processes is needed to enhance the application's design through I/O processing and storage requirements beforehand. (More information on this can be found in Chapters 6, 7, and 8.)

It is important to note that enterprise storage infrastructures deal with multiple applications and should be seen in context of an enterprise view rather than addressed piecemeal for each application system. This becomes problematic, however, during an implementation frenzy when the new applications are being deployed online. This points out another level of difficulty: developing an effective storage strategy that works with the traditional client/server storage model. The ability to look at the requirements from a cumulative perspective, noting the data size, characteristics, and access requirements on an enterprise scale, requires storage configurations that support multiple applications in a scalable manner. The application design-development-implementation process will not change, it will only get faster and less resource-aware. Accounting for it through an enterprise storage infrastructure strategy helps, but this must be integrated with storage solutions that mediate the challenge. (This forms a whole topic outside the scope of this book. Additional information can be found later in Chapter 17.)

The Benefits of Storage Networks on Business Applications

As stated previously, storage networking technology is being implemented to deal with extreme congestion and scalability issues within enterprise data centers. (See the Metro Data Areas sidebar earlier in this chapter.) IT management can expect the following benefits with the design and implementation of storage networking configurations and solutions.

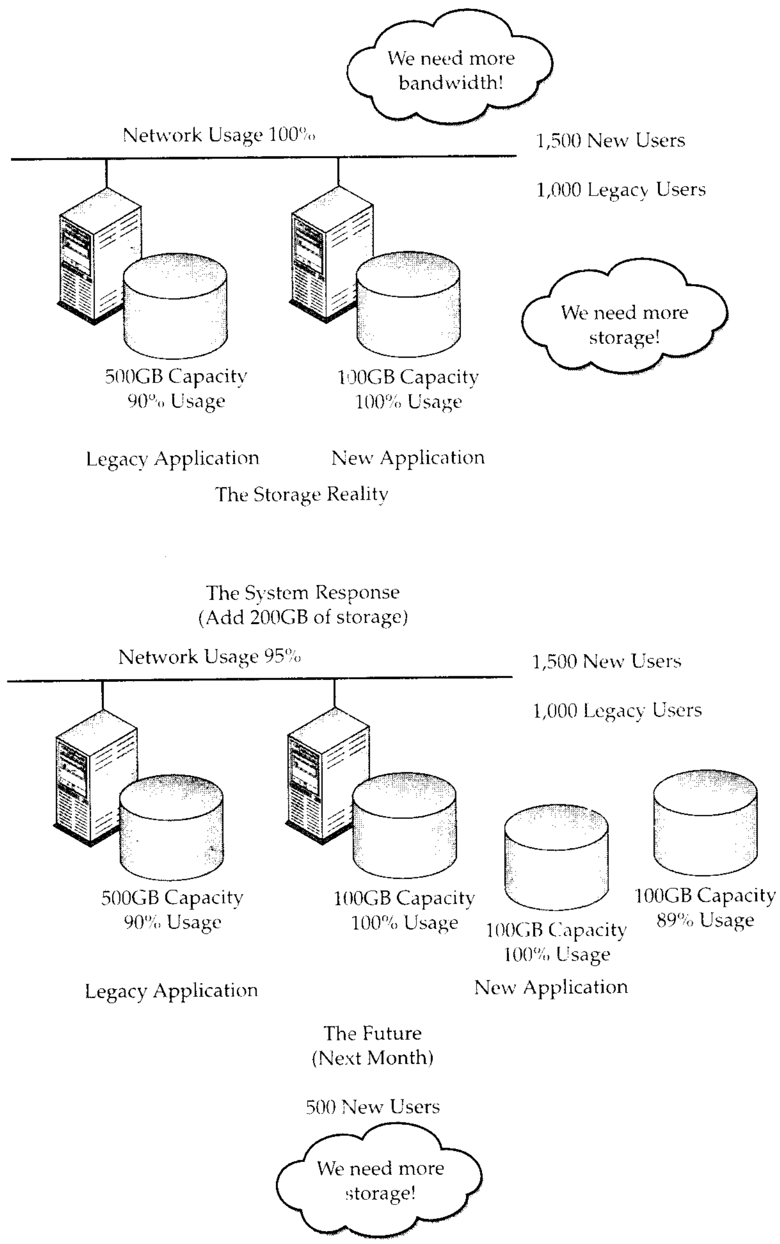

Figure 2-6. *The realities of application implementation*

Scalability of Access to Business Application Data

Figure 2-7 demonstrates the shift from centralized data access to networked data access for both SAN and NAS configurations, respectively. The platform benefits are clear, more servers can connect and utilize storage devices concurrently, thereby eliminating congested access to single servers, as well as the complexities and overhead involved in maintaining multiple copies of data for each server. From the business application perspective, this allows the application logic to be distributed throughout the servers, maintains centralized access to the data associated with the application, and eliminates a single point of failure for the application.

Note *One associated attribute of this configuration could be a storage protection strategy of data mirroring or RAID functionality to reduce the single point of failure for data availability.*

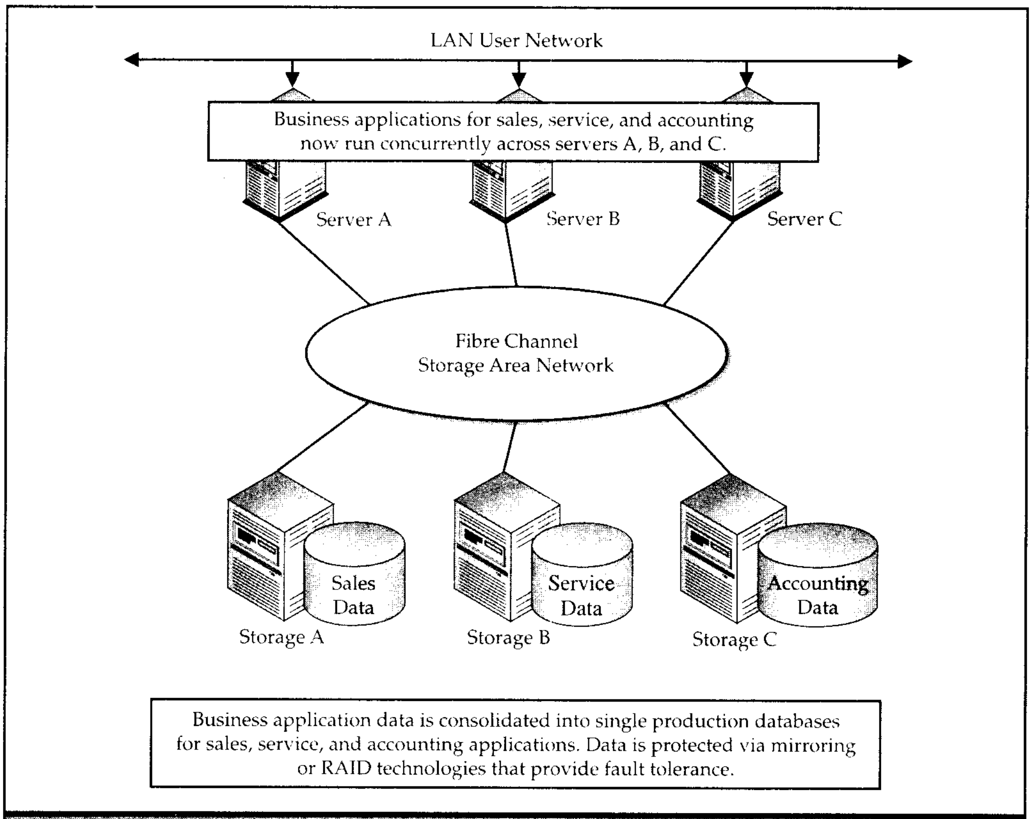

Figure 2-7. *Moving to networked data access with a SAN*

Consolidation of Business Application Data

Since maintaining multiple, synchronized copies of data is no longer necessary, storage devices previously connected directly to servers can now be used more efficiently by serving multiple servers with multiple sets of data. Eliminating the copying/synchronization process increases business application availability and reliability by decreasing data maintenance down times (copying/synchronization) and enhancing data integrity by eliminating state/timing differences from multiple copies of application data.

Faster Access to Data Through Improved Bandwidth and Centralized Configurations

The increased bandwidth associated with moving to storage network configurations has clear advantages in the amount of data transferred. Data now moves up to two times faster using Fibre Channel protocols than directly connected storage devices using SCSI protocol. Within SAN environments, this requires a new network using Fibre Channel (FC) protocol. NAS configurations offer even higher levels of Ethernet speed when configured on dedicated subnetworks that do not conflict with client network traffic. Both technologies allow for configurations that can sustain an increased number of business application transactions while setting a foundation for maintaining consistent and reliable response times.

Note *NAS boxes are used as thin servers with large storage capacities that remain dedicated to processing I/O requests through a distributed file system identified to subscribing application servers. As a result, NAS configurations become dedicated I/O extensions to multiple servers. Usage will continue to increase capability to run popular relational databases in scope as NAS becomes a reality.*

Increased Business Application Availability Through Efficiencies in Storage Management

As noted previously, there are multiple maintenance/support applications necessary to maintain platform environments. Figure 2-8 demonstrates the capability of devices within a SAN configuration to communicate with each other, thereby allowing many of the sever-based, datacentric maintenance/support applications to be optimized. Support applications can now be offloaded from the initiating server, which means data is copied from storage network device to storage network device (for example, disk-to-disk, disk-to-optical, disk-to-tape, and so on). Most business applications suffer during this maintenance process because data is unavailable during the time it is being copied. Although necessary for maintaining storage backup procedures and policies, this type of operation (in other words, copying data directly from device to device) can

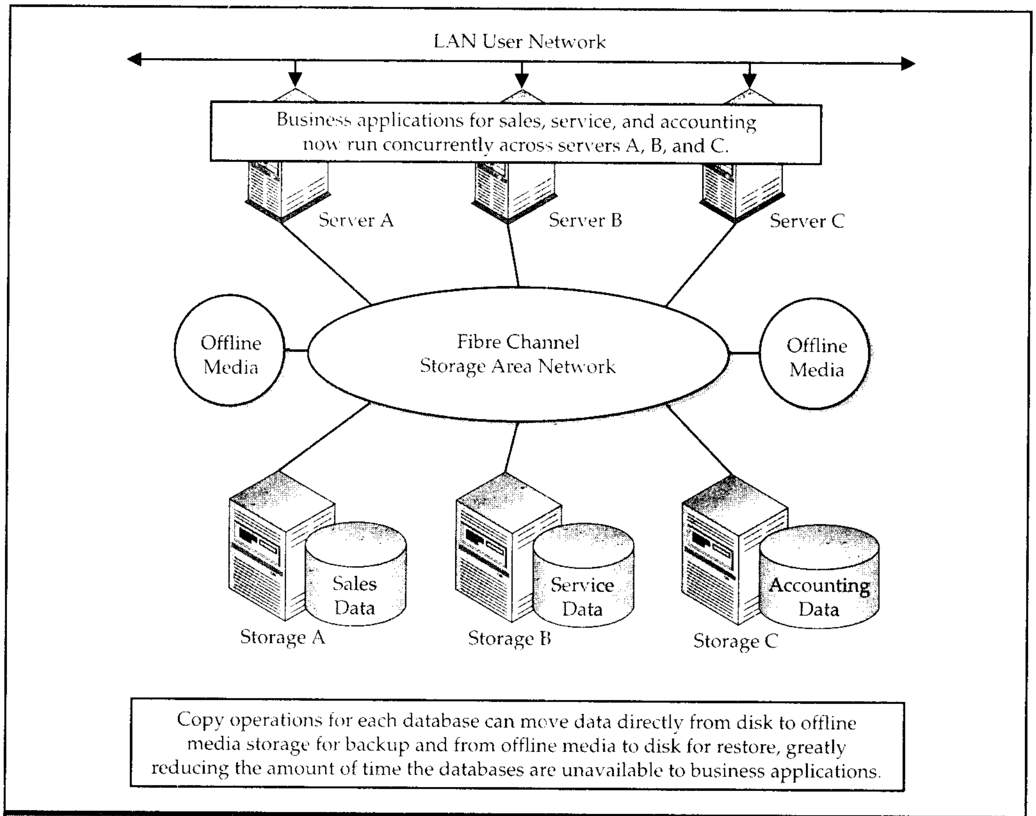

Figure 2-8. *Device-to-device communications within a SAN*

greatly improve the availability of business applications by reducing the time data is unavailable. Unfortunately, NAS suffers from its server-based architecture (albeit a thin server) and attachment to an Ethernet network. Backup and recovery operations are generally handled by a dedicated sever within the subnetwork. Presently, NAS configurations do not allow for optimization strategies such as device-to-device communication, as done in SAN configurations.

The Effects of Storage Networks on IT Organizations

A quote for your consideration: "Storage networking? Why yes, storage has always been networked. I know every morning when I come in and log in to the network, I access the sales tracking files positioned on the server. I also know that when the server goes down, I can't get to those files shared by the sales territory managers. Networked storage being new? It's always been in the computer systems I've used. Yep, couldn't live without it."

This manager of computer applications is obviously commenting on the new concept of networking storage. In his mind, storage has always been networked and is certainly not a new concept to him. In fact, as you've probably deduced, he's developed his professional activities to be dependent on storage shared on a network, across departments, and accessible to both headquarters and remote personnel. What he's talking about is the concept of the client/server storage model that has existed for many years and is now taken for granted in most global 2000 companies within the world.

And now, another quote from another IT professional: "Storage networking? No, we haven't gone in that direction yet. We're still waiting for the technology to become more mature and stable. However, we have plans to implement a SAN within the next year's planning cycle. We're still relying on our direct attached storage systems to get us through these next budget cycles. Now, to be fair, I'm not sure what the web guys are doing. They may have some other configuration they're working with."

What this IT manager of technical services is commenting on is the company's direction in implementing storage networking in the form of a SAN. Same company, same network, as described by his peer in the applications area, although in his mind, storage networking is a completely different animal than the traditional client/server storage model of storage networking.

If we next talk to the Internet/intranet infrastructure manager, we may yet find another story.

"Storage networking? Well, not really. We have Network Attached Storage that supports our web servers. That's really different from storage networks, isn't it? I'm not sure. In any event, our NAS storage boxes serve us well in allowing multiple storage locations that the web servers can go to for web pages, storing customer transaction data, customer files, and profiles. Of course, much of the transactional data gets downloaded to the application servers for processing accounts and updating the real financial records. So, I guess, yes we have networked storage across our enterprise. I think the server guys are looking at SANs though, so you may want to talk with them."

The point is that storage networking can and will be implemented across a broad range of services within the data center. Although it is increasingly available through two distinct architectural models, SAN and NAS, it is not readily recognized or understood within the data center. In order to utilize this new technology effectively, you must understand what's driving the storage infrastructures (which we have done in Chapters 1 and 2), as well as the details of each storage networking model, which will allow you to design and implement effective solutions.

Most managers responsible for application design and development think of storage as a necessary systems infrastructure—one that should always be there—so we must proceed with the knowledge that some things can be changed, while others can only be influenced. Part I of this book was designed not only as an introduction to storage networking but as an initial discussion regarding storage networks within the data center. If your job involves the technical services of your data center, you may want your peers in application development to read Part I (or perhaps purchase their own copy). It may spark an ongoing dialogue that finds each of you moving toward activities that drive and build a storage networking infrastructure unique to your environment.