Chapter 1

The Storage Area Network

In This Chapter

- Understanding storage area networks (SANs)
- Determining whether a SAN is right for you
- Looking at SAN layers and protocols
- Figuring out which operating systems benefit from SANs
- Discovering which applications can use or require SANs

This chapter is dedicated to helping you get a handle on what a storage area network (SAN) is, the basics of how one works, and whether one is right for your needs. You'll discover all the parts that make up a SAN, the things that make one run, and who actually makes all the different parts that you can buy. Putting a SAN together is somewhat like putting together one of those high-end stereo systems; you have many components and many different manufacturers to choose from. This chapter helps you choose the ones that suit your needs and create something that you can be proud of.

These days, becoming proficient with SANs can mean a major boost to your career. Perhaps you're bored to death in your current position and would like a change of pace. SAN administration is one of the highest-paying jobs in Information Technology (IT) today. If you add storage area networking to your résumé, you may find your phone ringing off the hook as headhunters vie to offer you a six-figure income (hey, might as well dream big).

Defining a SAN

First, the basics. In today's terms, the technical description of a SAN (Storage Area Network) is a collection of computers and storage devices, connected over a high-speed optical network and dedicated to the task of storing and protecting data.

In a nutshell, you use a SAN to store and protect data. A SAN uses the SCSI (Small Computer Storage Interconnect) and FC (Fibre Channel) protocols to move data over a network and store it directly to disk drives in block format. Today, that high-speed network usually consists of fiber-optic cables and switches that use light waves to transmit data with a connection protocol known as Fibre Channel. (A *protocol* is a set of rules used by the computer devices to define a common communication language.) More and more, regular Internet protocol (IP)–based corporate networks, and even the Internet, are being used as the network part of a SAN. IP networks that are already in place can be used by other storage connection protocols such as iSCSI (internet Small Computer Storage Interconnect) to move and store data.

Using a network to create a shared pool of storage devices is what makes a SAN different. A SAN moves data among various storage devices, allows sharing data between different servers, and provides a fast connection medium for backing up, restoring, archiving, and retrieving data. SAN devices are usually bunched closely in a single room, but they can also be connected over long distances, making a SAN very useful to large companies.

Many of today's SAN components are pretty much plug-and-play. To create a simple SAN, you just connect all the devices together with cables, and off you go. Creating larger SANs with many storage switches can become complex, though, and that's the reason for this book: to give you a handle on what you need to know about large, complex SANs.

Fiber versus Fibre

No, it isn't just a snooty way of spelling *fiber*. (Well, okay, not *only* that.) Networking geeks use the *fibre* spelling (reversing the *er* to *re*) to refer specifically to fiber-optic cables used in a SAN. The idea is to differentiate SAN cables from the optical cables used in other networks (such as TCP/IP Networks). That's because SAN devices use a different language to communicate with each other than do the devices in other networks. This is why the main protocol used in a SAN (snooty or not) is called *Fibre Channel*.

All network protocols are divided into layers, like a layer cake. All the layers in the cake are logically tied together into a *stack*. Each layer of the stack provides different functionality, and each device in the network uses the stack like a language to communicate with other devices in the network. The bottommost layer of the stack is hardware-based (as opposed to software-based), and thus is referred to as the *physical layer*.

The physical layer consists of tangible hardware stuff such as cables, switches, and connectors. This is where the fiber-optic cables are. On top of the physical layer are the software layers that make up the *protocol stack*. In a Fibre Channel SAN, those layers make up the *Fibre Channel* protocol.

Each type of network uses a different protocol to handle data. The Internet, for example, uses a protocol stack called the Transmission Control Protocol/Internet Protocol (TCP/IP). The physical layers of both Internet and SAN can transmit data as light pulses over fiber-optic cables — which (as you might expect) makes the data move nearly as fast as light. The only difference between regular fiber-optic computer networks such as the Internet and a fiber-optic SAN is the protocol and the switches used by the devices to talk to each other over the network. SANs use the *Fibre Channel* protocol and Fibre Channel switches, and the Internet uses the *TCP/IP* protocol and Ethernet switches. Fibre Channel was developed to move data really fast between computers and disk drives; TCP/IP (or "Internet Protocol") was developed to move files over long distances between computers.

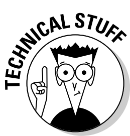
How a SAN Makes Computing Different

Using a SAN can really change how you think about computing. In the past, there was the *mainframe*, which was a gigantic computer that could run all the programs in a large business. All the computer stuff was gathered in one place called a *data center*. All the storage that the mainframe needed was directly connected to it. Everything was located and managed as a single, large entity.

The PC revolution changed a lot of things. Everything started to spread out. Data was moved off the mainframe and stored in server computers. The servers were then dispersed throughout the enterprise to bring computing power closer to the actual users. The servers became connected by a network, called a *local area network,* or LAN. This was cool because now the computing power was spread out and made more available to end users. Eventually, LANs were connected to create the Internet.

Networks enabled people who used computers in far-flung places to communicate and share information with each other. In business, problems arose when inter-networking finally took off. A great deal of data was now being stored with no effective way to manage it all. Managing all the scattered data dispersed throughout the network became a nightmare.

Because all data storage was located inside each individual server, you had no effective way to efficiently allocate storage space between all the servers. Sure, users could share files over a LAN, but you still needed a way to share access to physical disks, rather than using dedicated disks inside every server. Hence the advent of the SAN.



Since the original TCP/IP network protocols used in a LAN (Local Area Network) were built to move and share files, they had no built-in way to directly access disk drives. As a result, very high-performance applications needed direct access to *block-based* disk drives to move and store data very fast. (Data is stored as blocks on a disk drive.)

Disk drives in a SAN are stored in a dedicated storage device called a *disk array*. All the servers connect to the storage device over a high-speed network using the Fibre Channel protocol, which enables very fast access to disks over a network. Using a SAN gives businesses shared and consolidated access to data storage — available to any server connected to the SAN.

Putting a SAN in place makes individual server computers less important and more peripheral to the data stored in the SAN. After all, the data is what is important to your business. If you lose a server, you can buy a new one. If you lose your data, it's "Adiós, amigo" for your business.

Understanding the Benefits of a SAN

The typical benefits of using a SAN are a very high return on investment (ROI), a reduction in the total cost of ownership (TCO) of computing capabilities, and a pay-back period (PBP) of months rather than years. Here are some specific ways you can expect a SAN to be beneficial:

- **Removes the distance limits of SCSI-connected disks:** The maximum length of a SCSI bus is around 25 meters. Fibre Channel SANs allow you to connect your disks to your servers over much greater distances.
- **Greater performance:** Current Fibre Channel SANs allow connection to disks at hundreds of megabytes per second; the near future will see speeds in multiple gigabytes to terabytes per second.
- **Increased disk utilization:** SANs enable more than one server to access the same physical disk, which lets you allocate the free space on those disks more effectively.
- **Higher availability to storage by use of multiple access paths:** A SAN allows for multiple physical connections to disks from a single or multiple servers.
- **Deferred disk procurement:** That's business-speak for not having to buy disks as often as you used to before getting a SAN. Because you can use disk space more effectively, no space goes to waste.
- **Reduced data center rack/floor space:** Because you don't need to buy big servers with room for lots of disks, you can buy fewer, smaller servers — an arrangement that takes up less room.
- **New disaster-recovery capabilities:** This is a major benefit. SAN devices can mirror the data on the disks to another location. This thorough backup capability can make your data safe if a disaster occurs.
- **Online recovery:** By using online mirrors of your data in a SAN device, or new continuous data protection solutions, you can instantly recover your data if it becomes lost, damaged, or corrupted.

- **Better staff utilization:** SANs enable fewer people to manage much more data.
- **Reduction of management costs as a percentage of storage costs:** Because you need fewer people, your management costs go down.
- **Improved overall availability:** This is another big one. SAN storage is much more reliable than internal, server-based disk storage. Things break a lot less often.
- **Reduction of servers:** You won't need as many file servers with a SAN. And because SANs are so fast, even your existing servers run faster when connected to the SAN. You get more out of your current servers and don't need to buy new ones as often.
- **Improved network performance and fewer network upgrades:** You can back up all your data over the SAN (which is dedicated to that purpose) rather than over the LAN (which has other duties). Since you use less bandwidth on the LAN, you can get more out of it.
- **Increased input/output (I/O) performance and bulk data movement:** Yup, SANs are fast. They move data much faster than do internal drives or devices attached to the LAN. In high-performance computing environments, for example, IB (Infiniband) storage-network technology can move a single data stream at multiple gigabytes per second.
- **Reduced/eliminated backup windows:** A *backup window* is the time it takes to back up all your data. When you do your backups over the SAN instead of over the LAN, you can do them at any time, day or night. If you use CDP (Continuous Data Protection) solutions over the SAN, you can pretty much eliminate backup as a separate process (it just happens all the time).
- **Protected critical data:** SAN storage devices use advanced technology to ensure that your critical data remains safe and available.
- **Nondisruptive scalability:** Sounds impressive, doesn't it? It means you can add storage to a storage network at any time without affecting the devices currently using the network.
- **Easier development and testing of applications:** By using SAN-based mirror copies of production data, you can easily use actual production data to test new applications while the original application stays online.
- **Support for server clusters:** *Server clustering* is a method of making two individual servers look like one and guard each other's back. If one of them has a heart attack, the other one takes over automatically to keep the applications running. Clusters require access to a shared disk drive; a SAN makes this possible.
- **Storage on demand:** Because SAN disks are available to any server in the storage network, free storage space can be allocated on demand to any server that needs it, any time. Storage virtualization can simplify storage provisioning across storage arrays from multiple vendors.

Finding Out Whether a SAN Is Right for You

Though SANs can offer many advantages, they aren't for everyone. If you own a small business and use just a few computers to keep it going, using a SAN is probably overkill for you. Sometimes the cost isn't justified by the benefits. The more servers you have in your organization — and the more data that you need to store — the more benefit you'll see from a using a SAN. Prices have come down a lot since the first writing of this book, but storage networking equipment isn't cheap. For example, a single high-performance host bus adapter (more about that later) can cost more than a thousand dollars; a storage switch can cost tens of thousands.

A good guideline that we use is what we call *The Rule of 16.* If you have 16 or fewer servers, using a SAN probably doesn't make sense. (Of course, you may still benefit from a less expensive NAS- or iSCSI-based solution, which we touch on later.) You can easily manage 16 or fewer servers with one person, and data-storage needs shouldn't be that high. If you use more than 16 servers, or servers that run large databases, you're a good candidate for a SAN. If you're responsible for hundreds of servers, using a SAN will probably dramatically reduce the cost of managing data.

Who should use a SAN?

You should use a SAN if you work in a large organization (more than 16 servers, or servers that run large databases) in which data management or data backup is becoming a problem. (By *server* here, we mean the hardware you buy to run your applications. When it runs your applications, it is the "server" part of a client/server implementation.) Your servers might be running out of disk space all the time, and you might have no room left in the servers to add disk drives. A business in this server pickle is a typical SAN candidate. You might have way too much data to be backed up or restored in a timely fashion. Using a SAN can fix that, too.

The following checklist details the types of server resources, both software and hardware, that should be included in a SAN:

- **Database servers:** Oracle, Sybase, SQL, DB2, Informix, AdaBase, and other databases love to make use of the extremely fast disks in a SAN.
- **File servers:** Using SAN-based storage for Windows or Unix computers acting as file servers lets you expand your file-server storage resources quickly, makes them run better, and improves overall management. Specialized devices called NAS (Network Attached Storage) servers can supply shared access to stored files over a standard TCP/IP network.

- **Backup servers:** Connecting all your servers (including backup servers) to a SAN enables you to back up your data through a SAN rather than through a LAN — and SAN-based backup is dramatically faster.
- **Voice/video servers:** Voice and video servers tend to push large amounts of data very quickly. That's what SANs are built to do.
- **Mail servers:** Using SAN-based storage for mail servers enables quick restoration of data in case of corruption or viruses. It also lets you back up your mail servers faster, and you can use clusters as mail servers.
- **High-performance application servers:** A SAN's capabilities benefit applications for managing documents, scientific computations, customer relationships, billing, data warehouses, and other high-performance business functions.

Who should not use a SAN?

You don't really need to use a SAN if your organization is small (16 servers or fewer) or where data management, application performance, or backup is not currently a problem for you.

For that matter, the technology you have may *not* be a good fit with a SAN. Here's a checklist of the types of servers that should *not* be included in a SAN. Such servers are usually better off staying on their internal disk drives; they don't benefit from SAN-based storage (which is also more expensive):

- **Web servers:** Computers set up as Web servers don't usually have large storage needs; they're usually connected to larger servers that run the databases from which Web pages are automatically built. Although Web servers are good candidates for NAS, database servers can make better use of SAN disks.
- **Infrastructure servers:** Server applications that handle the chores of network infrastructure — such as Domain Name Servers (DNS), Windows Internet Naming Servers (WINS), and Domain Controllers (DC, PDC) — are better left on the server computers' internal disks. They don't need a lot of disk space, and their performance requirements are minimal.
- **All desktop PCs:** Personal computers are not good SAN candidates because they usually connect to corporate servers for any applications that require high performance. Those corporate servers, however, *could* use a SAN.
- **Servers needing less than 10GB of storage:** Face it: Internal storage is cheaper than SAN storage. If your server has no performance problems and will never need more than 10GB of storage space, leave it alone.

- **Servers that don't need fast access to data:** If performance is good already and you don't mind maintaining the server separately, don't bother hooking it up to a SAN.
- **Servers that have to share files:** Such servers are better off connected to a Network Attached Storage (NAS) server. NAS servers store and transfer data as files, and not blocks of data, so they don't need the high-speed Fibre Channel protocol used in a SAN. NAS devices are best for file-based uses such as user home directories and shared documents.

Dissecting a SAN (The Four Ps)

We divide this section into four parts, which we call *the four Ps* — namely the parts, protocols, players, and platforms you can choose from when creating a SAN. We don't go into all the gory details because it would take up too much space here and most likely be better for bedtime reading (you're getting sleeeepy). We just give you a general overview of the following:

- **The parts:** All the hardware you use to create a SAN; the switches, cables, disk arrays, and so forth
- **The protocols:** The languages that the parts use to talk to each other
- **The players:** The folks who build the parts
- **The platforms:** The computer applications that benefit from using SAN

The Parts of a SAN

It's most convenient to imagine the parts of a SAN in three layers. The top layer is the *host layer,* which includes the server computers and everything that goes into them. The middle layer is the *fabric layer,* which includes all the cabling and switches that connect everything. The bottom layer is the *storage layer,* where all the storage devices are located.

The host layer

The major components in this layer are the servers themselves, the host bus adapters (HBAs, which include a part called the Gigabit Interface Converter, or GBIC), and all the software running on the server that enables the host bus adapter to communicate with the fabric layer.

The host bus adapter (HBA)

The server connects to the SAN through a *host bus adapter* (HBA) — an I/O adapter card that fits inside your server and connects it to the fabric layer.

The Gigabit Interface Connector (GBIC)

The Gigabit Interface Converter (GBIC) is where the cable plugs into the HBA card. Every HBA has a GBIC that snaps into an opening in the card or is soldered to the card. The openings in the GBIC extend out the back of the server so you can plug in the cable. The GBIC houses the laser and electronics that convert the data inside your server into light pulses that travel over the cables. GBICs are used not only in the HBA, but in every device in the SAN. Anywhere an optical cable has to be plugged in, you find a GBIC.

Fiber-optic cables

Fiber-optic cables are unique in that they are really part of all three layers in a SAN (such as the GBICs where the cables are plugged in). These cables, which connect everything in a SAN, use glass fibers to transmit light waves from one device to another. You can use one of three optical cable types, depending on the distance between connections and the wavelength of light used to transmit data. (See Chapter 2 for more information.)

The fabric layer

The *fabric layer,* or the middle layer of a SAN, is the actual network part of a SAN. The *network* — where all the cables are connected — is also where you find hubs, switches, gateways, and routers, which tie all the cables together into a logical *and* physical network. Its components include

- **Hubs:** A *hub* is a simple electronic device that physically connects the cables into a logical loop of cable. This is why hub-based SANs are called SAN *loops.* The hub has connection points — *ports* — where the cables get plugged in. These ports use GBICs to connect the cables to the hub. In a hub, the light coming in from a cable can pass through the hub to a device connected to another port. The light travels around the loop to each port in the hub. Because hub ports are connected in a loop, only one device can communicate through a hub at one time.
- **Switches:** A *switch* is a smart electronic device that physically connects cables. Switches are the heart of a SAN network. This is where a lot of the intelligence resides. The switches reliably route your data from the host layer to the storage layer.

  TIP

  Think of a switch as working like a telephone switchboard operator. Every incoming call gets connected to its destination over the wires in the switchboard, and the operator knows which wire to plug in where to make this happen.

- **Gateway**: A gateway (also referred to as a *bridge*) is a smart electronic device that physically or logically enables devices to communicate over one protocol to talk to devices that use a different protocol. For example, an iSCSI gateway can connect hosts that use the iSCSI protocol to storage devices that use the Fibre Channel protocol in a Fibre Channel SAN fabric.
- **Router**: A *router* is another smart device that physically or logically routes data between two individual networks.

The storage layer

The *storage layer* is where all your data resides on the SAN. This is the layer that contains all the disk drives, tape drives, and other storage devices, like optical storage drives. The storage layer's devices include some intelligence, such as Redundant Array of Inexpensive (or Independent) Disks (RAID) and snapshot or other data-replication technologies to help protect data. The capabilities of the storage devices can affect what you can do with a SAN.

Storage arrays

A disk is a disk — two disks are (okay) a couple of disks, and an *array* of disks is just a bunch of disks (also called a JBOD) all located in the same place. But a *storage* array adds extra intelligence to the controllers within the array — which allows you to do cool stuff like RAID, so it's no longer just a bunch of stupid disks. The intelligence built into the storage *controllers* in the storage array is what enables this additional functionality.

A *storage array* is a big box that has a bunch of disk drives in it, running smart code called *firmware* that makes it more intelligent. Of course, you could go to a computer store and buy a bunch of hard drives, but how would you connect them to your server? Today's storage arrays use fast, dedicated microprocessors to run complex software that makes them more useful than they'd be if you just connected a bunch of disks to your servers. (More on storage arrays in Chapter 2.)

The storage arrays connect to the fabric layer with cables that run from the devices in the fabric layer to the GBICs in the ports on the array. Many types of storage arrays are available, but they come in two basic flavors: modular and monolithic. Both these types use built-in computer memory to help speed up or *cache* access to slow disk drives; each uses the memory cache differently. Memory is expensive, so the more expensive monolithic arrays usually have more cache memory than modular arrays. Here's a closer look.

Modular arrays

Modular arrays have fewer port connections than do monolithic arrays; they usually store less data, and connect to fewer servers. They're designed so you can start small, with only a few disk drives, adding more drives to the array as your storage needs grow. Modular arrays come with shelves that hold the disk drives. Each shelf can hold between 10 to 16 drives, depending on the model and manufacturer. Modular arrays usually fit into industry-standard 19" racks, so you can have all your servers and SAN disks in the same rack.

Modular arrays are perfect for smaller companies looking to install a SAN on a limited budget. They're also good for large companies with many remote offices, because they are much cheaper and smaller than big monolithic arrays, so they can be placed into smaller offices. Modular arrays almost always use two controllers with separate cache memory in each controller, and then mirror the cache between the controllers to prevent data loss. Most modern modular arrays have between 16 and 32GB of cache memory.

Monolithic arrays

Monolithic arrays are those big, refrigerator-size collections of disk drives you see sitting next to mainframes in a data center. These disk arrays are loaded with advanced technology that almost always prevents them from going down. Monolithic arrays can accommodate hundreds of disk drives, can store data for a lot more servers than a modular array can, and usually connect to mainframes. Monolithic arrays have many controllers, and those controllers can share direct access to a *global* memory cache (up to hundreds of gigabytes) of fast memory. This method of sharing access to a large global or *monolithic cache* is why these arrays are also called monolithic.

Modular versus monolithic in large-scale enterprise use

At larger scales of operation, modular arrays are often used as *midrange arrays* and monolithic arrays are often used as *enterprise arrays.* The main difference here, however, is functional: Although some enterprise-class arrays *can* be modular in design, they can also connect to and store mainframe data (which a modular array usually can't do). Typically enterprise-class monolithic arrays are much more expensive, and have better built-in redundancy features that make them extremely reliable.

Whether modular or monolithic, each array type has its advantages and disadvantages. Modular arrays are generally less expensive but can handle large-scale workloads if you add enough disk shelves or controller shelves to do the job. When you add controller shelves, you get more horsepower. When you add more disk shelves, you get more storage.

Modular arrays are designed from the ground up to be extremely fast when connected to just a few servers. If you need to add servers, you just buy more controllers. Many companies like that kind of flexibility. Monolithic arrays, on the other hand, can be connected to mainframe computers. They also usually have many more physical ports on them to connect to the SAN, allowing many more servers to use the array. Many companies use monolithic arrays to help consolidate more storage into less space without losing performance when servers are added. Monolithic arrays are almost always more expensive than modular arrays, but you get what you pay for.

The SAN Protocols

As mentioned earlier, a protocol is a type of computer language used by a computer system to communicate with other devices. By *language,* we don't mean a programming language. It's more like a set of agreed-upon methods — a way for computers to communicate so they can cooperate in moving data over the network.

Each type of computer device uses a different protocol to communicate with other devices. After two devices find a common language, they establish a communication session by greeting each other with a friendly exchange of code called a *handshake.* In effect, they have a conversation to find things out about each other and to negotiate the best or fastest way to communicate.

There are two major protocols (languages) used in Fibre Channel SANs: the Fibre Channel protocol (used by the hardware to communicate) and the Small Computer System Interface (SCSI) protocol (used by software applications to talk to hard drives). Here's a closer look:

- **Fibre Channel protocol:** This is the language used by the HBAs, hubs, switches, and storage controllers to talk to each other. The Fibre Channel protocol is a *low-level language;* it's the means of communication between actual hardware components, and not between the applications that run on the hardware.

  Actually, two protocols make up the Fibre Channel protocol: Fibre Channel Arbitrated Loop (FC-AL), which works with hubs; and Fibre Channel Switched (or FC-SW), which works with switches. (Chapter 2 has more on the Fibre Channel protocols.)

  Fibre Channel is the building block of the SAN highway. It's like the road of the highway, where other protocols can run on top of it, just as different cars and trucks run on top of an actual highway. In other words, if Fibre Channel is the road, then SCSI is the truck that moves the data cargo down the road.
- **SCSI protocol:** This is the language used by SAN-attached server applications on the server computers to talk to the disk drives. This protocol lies on top of the Fibre Channel protocol.

This book is focused on Fibre Channel-based storage networks, so we only briefly touch on other protocols such as iSCSI (the SCSI protocol used over an IP network rather than a Fibre Channel network) and the Infiniband-based protocols (such as iSER and SRP) that can also be used to create a high-speed storage network. Infiniband itself can be a whole other book; it's used increasingly in GRID computing — connecting many low-cost servers over a high-speed network to act like one *very* fast computer. Storage is always the slowest part of any computer, so using a high-speed SAN with a GRID is essential. NASA and the CIA use GRID computing networks to gather and analyze massive amounts of data.

Even though most storage array manufacturers now use Fibre Channel disks in their storage arrays, the disks themselves still use the legacy SCSI protocol to communicate with applications over the Fibre Channel network. All the SCSI messages are *encapsulated* (packaged) into the Fibre Channel protocol. It's kind of like writing a letter to your dear Aunt Sally. (Aunt Sally is your disk drive here.) You write a letter (your data) and address the envelope (a SCSI block) to Aunt Sally (your disk). You want it to get there fast, though, so you put the letter into a FedEx package (you encapsulate the SCSI block in the Fibre Channel frame) and send it off. The Fibre Channel switch in the SAN opens the FedEx package (Fibre Channel frame), looks at the original address on the envelope (SCSI block), and sends it along its merry way at light speed to Aunt Sally (your disk).

How SAN devices communicate

Using English as a metaphor, think of a typical protocol conversation like this:

HBA in the server: "Hey! How are you? I'm in this server, and I'm trying to find a disk drive to store this data. Who are you?"

Switch: "Hi. I'm a Fibre Channel switch. I see that you can speak Fibre Channel. Let's talk using the new version 2 dialect, okay?"

HBA in the server: "Okay. Look, do you know of any good disk drives I can use to store the data?"

Switch: "Sure, according to your address, I've been authorized to give you access to a drive on my Port 3. Would you like to speak with her? Remember, SCSI drives speak a different language. Do you speak SCSI?"

HBA in the server: "Nope, but the server's application does! Thanks. I'll have him send you all the data using the SCSI protocol. Can you forward this to the disk?"

Switch: "Done deal. Hey, SCSI drive on Port 3, here's a message for you!"

At this point, the session is established; the switch now passes SCSI messages through to the disk drive. The drive acknowledges the messages and does what the server tells it to do.

All Fibre Channel devices work this way. The language for communication with storage devices is SCSI. Fibre Channel is just the FedEx way of getting it there faster, like a postal deliverer running at light speed. SANs work by giving the SCSI protocol a free ride on top of the Fibre Channel protocol to make communication happen much faster.

The SAN Players

The *players* are the companies that are the driving force in the SAN industry. Hundreds of companies are selling SAN equipment these days, each selling products that fit into a particular niche. You can break the players down into the different types of products that they sell. Some companies can sell everything you need, including servers. Server companies sometimes buy other companies' products and resell them as their own. (Most of us can't be good at *everything* these days.) You can get a listing of companies that develop products for SAN from the Storage Network Industry Alliance (SNIA; `www.snia.org`), a consortium of companies all working together creating standards for storage area networks.

The SAN Platforms

The *platforms* are the types of servers that can benefit from using a SAN and are appropriate for SANs. As indicated earlier in this chapter, not all servers should be hooked up to a SAN.

TIP

The operating system running on your server requires a driver. A *driver* is a small bit of software (detailed in Chapter 7) that enables the HBA in the server to talk to the other elements in the SAN. Some operating-system platforms support HBA drivers; some don't. You might need the latest version of your operating system to use a SAN if earlier versions don't support the needed drivers. For example, older versions of Windows NT, such as Windows NT 3.51, don't support SAN drivers. The same is true for older Apple-based networks. If you're running later versions of Unix or newer Windows server environments, you should be fine.

Three types of server platforms are good to use in a SAN: big, fast Intel- or AMD-based servers; big, fast servers that can run the different flavors of Unix; and mainframes. No surprise that these are the more expensive and powerful systems that need to store a lot of data. Most server platforms have drivers that allow them to be hooked up to a SAN environment. Whether doing so makes sense depends on the type of application running on it and the amount of disk storage that the server needs. Here are the *minimum* operating-system versions you can include in a SAN:

- Microsoft Windows NT 4.0
- Microsoft Windows 2000 or later
- Sun Solaris versions 2.6 or above
- HP-UX version 10.2 and higher
- IBM AIX version 4.2 and higher
- HP Tru64 Unix version 4.0F and higher
- HP Open VMS version 7.2 and higher
- Novell Netware version 4.11 and higher
- SGI IRIX version 6.5 and higher
- Sequent DYNIX version 4.5 and above
- All the various flavors of Linux (such as Red Hat, SuSE, and their cousins)
- IBM OS/390 Mainframe MVS, or Z/OS

Always check with your SAN vendor to find out whether the disk array you're purchasing supports the operating systems you're using.

Applications that benefit from a SAN

Most applications running on a server would benefit from faster access to the disk drives that the application is trying to use. Using a SAN instead of disks inside the server not only makes disk access faster (SAN disk access is at light speed) but also makes managing those disks much easier. If you're building a SAN, this list is a guideline for choosing servers to hook up:

- **Any server-class computer running a high-performance application:** By *server class,* we mean anything with at least lots of memory (2 to 4 GB or more) and a fast Intel, AMD, or Reduced Instruction Set Computer (RISC)-class processor. (***Note:*** This *isn't* the kind of chip you'd find in your normal desktop-type PC, Web server, or infrastructure server such as a DNS or domain server.)
- **Any server computer with expanding disk-storage needs:** Using a SAN makes it easy to allocate more storage to a server without having to bring the server down.
- **Any database-type application server:** Databases require very fast disk access. A SAN can provide this kind of fast disk access.

- **Any backup server:** Backup servers have tape drives or disk drives connected to them to back up your data so you can restore it if your disks crash. Using SAN-connected tapes or disks to back up your data relieves the strain of backing up your disks across your computer network — and also makes backup happen much faster. Backup servers benefit greatly from SANs; backup moves a *lot* of data.
- **Any virtual-server solution:** Server virtualization makes it easy to make one physical server look and act like many servers. Server virtualization software or hardware benefits from the performance a SAN offers, and by the ability to share the external storage among the virtual machines. (More on virtualization for both servers and storage in Chapter 15.)

Applications that require a SAN

Only a handful of applications actually require the use of a SAN. These are usually newer applications, designed specifically for SAN capabilities:

- **Cluster applications:** *Cluster applications* are created by tying a group of servers together via a fast network and then allowing those servers to access the same disks' storage where the application is installed. This allows for very scalable and highly available applications; if one of the servers fails, another server in the cluster can pick up where the first one left off. GRID computing (mentioned earlier in the chapter) is another example of applications that need a SAN to work right. Common cluster applications are IBM HACMP (`www.ibm.com`), Solaris Cluster 3.0 or above (`www.sun.com`), Compaq/HP TruCluster (`www.hp.com`), Oracle Failsafe Cluster (`www.oracle.com`), Oracle Real Application Clusters (`www.oracle.com`), Microsoft Cluster Server (MSCS) (`www.microsoft.com`), HP MC/Serviceguard Clusters (`www.hp.com`), and Novell Netware Cluster Services (`www.novell.com`).
- **SAN backup applications:** SAN-based backup software is optimized for using SAN hardware. The backup software includes intelligence that takes advantage of what SAN offers. When using SAN backup software, you can back up your data directly over the SAN to a tape drive, which makes backup run much faster. Common backup software that has this capability includes Veritas NetBackup (`www.veritas.com`), Tivoli Storage Manager (`www.ibm.com`), Veritas Backup Exec Enterprise Edition (`www.veritas.com`), CA ARCserve with the SLO option (`www.ca.com`), Legato NetWorker (`www.legato.com`), and CommVault (`www.comvault.com`).
- **Server-virtualization solutions:** *Server virtualization* hardware and software such as VMware, Virtual Iron, Microsoft Hyper-V, Egenera, and others need the disk-sharing capability of a SAN (especially fail-over for applications between physical servers), and they also gain from the performance benefits that a SAN has to offer.