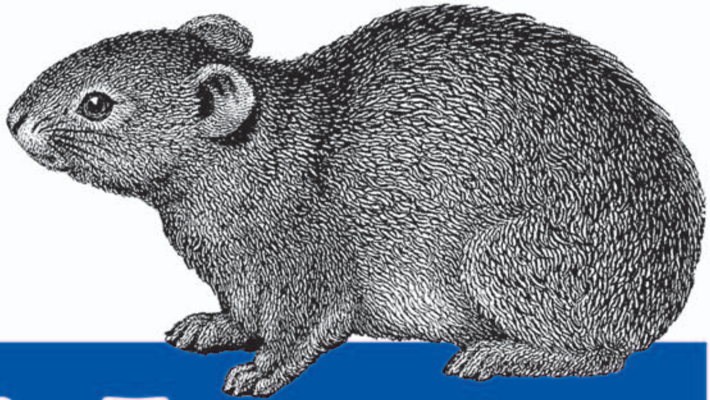
Help for Storage Administrators



Using SANs *and* NAS

O'REILLY®

W. Curtis Preston

Network Administration

O'REILLY®

Using SANs and NAS

Data is the lifeblood of modern business, and modern data centers have extremely demanding requirements for size, speed, and reliability. Multi-terabyte data stores are common; petabyte data stores are not unheard of. But given this much data, how do you ensure it's always available, that access times and throughput are reasonable, and that the data can be backed up and restored in a timely manner.

Using SANs and NAS shows you how to use the latest technologies for building large data centers. It helps you understand which complementing technologies are appropriate for your network. How do SANs and NAS differ, and how are they alike? How is NAS different from simple NFS and SMB/CIFS? What future technologies, such as iSCSI and DAFS, might affect the picture? What are the building blocks of a SAN, and how can they be assembled for effective storage solutions?

Most important, *Using SANs and NAS* pays special attention to the problems of backing up huge data stores. How do you design a cost-effective backup system that can perform a backup or recovery of your multi-terabyte storage system in a reasonable time? What virtualization technologies can be used to make backup and recovery almost instantaneous?

Using SANs and NAS is a practical book that gives storage administrators the tools they need to solve the toughest problems. It is an essential part of a storage administrator's library.

W. Curtis Preston is the author of *Unix Backup & Recovery*, and the webmaster of *storagemountain.com* (formerly *backupcentral.com*). He is also president of The Storage Group, which designs and implements storage systems for companies of all sizes.

www.oreilly.com

US $29.95 **CAN $44.95**
ISBN-10: 0-596-00153-3
ISBN-13: 978-0-596-00153-7

52995
9 780596 001537

Using SANs and NAS

Using SANs and NAS

W. Curtis Preston

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Paris • Sebastopol • Taipei • Tokyo

Using SANs and NAS
by W. Curtis Preston

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly Media, Inc. books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*safari.oreilly.com*). For more information, contact our corporate/institutional sales department: (800) 998-9938 or *corporate@oreilly.com*.

Editor: Mike Loukides

Production Editor: Mary Anne Weeks Mayo

Cover Designer: Ellie Volckhausen

Interior Designer: Melanie Wang

Printing History:

February 2002: First Edition.

This book uses RepKover™, a durable and flexible lay-flat binding.

ISBN-10: 0-596-00153-3
ISBN-13: 978-0-596-00153-7
[M] [8/07]

With great sadness, this book is dedicated to:

92 lives lost on American Airlines Flight 11
64 lives lost on American Airlines Flight 77
56 lives lost on United Airlines Flight 175
45 lives lost on United Airlines Flight 93
125 lives lost in the Pentagon
Thousands of lives missing or lost in the
World Trade Center

. . . and to the thousands of heroes who have risen from this disaster, including firefighters, paramedics, police officers, construction workers, and everyone supporting these fine people who must perform the saddest job in U.S. history.

"You can tear down our buildings, but you can't tear down our spirit."

—George W. Bush

Table of Contents

Preface

The alphabet soup that is the computing industry has finally made its way to the storage industry. It wasn't enough that we had NFS, SMB, CIFS, and SCSI. Now there are SANs (storage area networks) and NAS (network attached storage). As exciting as these two industries are, and as happy as I am for the solutions they bring, did they have to use terms that are palindromes? I think if those involved could have just changed the name of one of them, some of the confusion I deal with regularly would be avoided.

But, we don't have that luxury. What we've got are two industries based on the same premise: you've got a lot of data to store, and you need somewhere to put it. Both industries are trying to solve the current challenges there are with traditional storage:

Manageability
: Let's face it. Traditional, parallel SCSI systems are difficult to manage when you start talking hundreds of gigabytes or terabytes. Each disk is married to a computer physically, electrically, and logically. When a disk goes bad, replacing it live is almost never possible with parallel SCSI. And if a particular disk or set of disks is no longer needed by one system, it's difficult to reallocate them to a new system. These concerns are just a start.

Scalability
: Sixteen devices per bus? Are you kidding? With systems getting smaller and smaller, and becoming more rack-mount friendly, the backplane real estate required for all those SCSI cards just goes away. A system is needed in which you can store data that grows when needed—without a lot of hassle.

Availability
: Traditional, parallel SCSI systems aren't known for their availability. Does the term *SCSI bus reset* mean anything to you? And, since a disk can be seen only by one SCSI bus, how can you perform maintenance on the SCSI cards when they fail?

Recoverability

Ah, my favorite subject. What about the backups? How on earth are you supposed to get all this data to tape or to some sort of offsite storage device using traditional backup and recovery methodologies designed for slower networks and much less data? We need new answers here.

Many of these problems can be solved by installing a SAN or NAS. But which one should you buy? What's the difference between these two technologies? Their names look almost the same. Are they really that different? Is what your NAS salesperson saying about SANs really true? Is what your SAN salesperson saying about NAS really true? Maybe it is; maybe it isn't.

Boy, Was This Fun!

I got a lot of cooperation from a number of storage vendors while writing this book. In fact, the technical edit of this book was done by a number of people that work for storage vendors. They were very helpful, and the process was also very interesting. (It kind of reminded me of when I wrote *Unix Backup and Recovery* [O'Reilly] and was trying to get the Informix, Oracle, and Sybase guys to agree on terms.) Here are some of the comments I heard while working with both vendors and users of this technology. (These comments are in no particular order.)

- "NFS and CIFS are inefficient protocols. How can they support databases on NAS?"
- "The volume managers and filesystems of today are inefficient. The only efficient filesystem is a NAS filer."
- "Only Fibre Channel is able to run at line speed. Gigabit Ethernet is too slow."
- "The new hardware-accelerated Gigabit Ethernet NICs are way faster than Fibre Channel."
- "Fibre Channel can be routed across WANs now."
- "So can Gigabit Ethernet, and we don't need any special switches to do it."
- "You have to use NDMP to back up NAS filers."
- "NDMP doesn't have any of the usual maintenance associated with traditional client-based push agents."
- "Server-free backups are the wave of the future."
- "Server-free backups are all hype."
- "You can automatically sync this NAS filer with that NAS filer."
- "You can do the same thing with this SAN array,"
- "iSCSI will make all this SAN stuff a thing of the past."
- "iSCSI will probably take off, but Fibre Channel is here to stay."

- "I work for a SAN company. This book just sounds like NAS propaganda."
- "I work for a NAS company. This book just sounds like SAN propaganda."

Like I said, lots of fun! Frankly, I'm really glad that there is fierce competition among storage vendors. It means that you, the customer, get better products.

I worked hard to present a balanced view of these two storage options. I hope you find it useful.

What Is This Book For?

I wrote this book for a few reasons. The first is that I found a good bit of confusion in the industry as to what SANs and NAS are. The second reason is that I've done a lot of really interesting backup and recovery projects with both of them and wanted to share that experience with everyone. Having said this, I believe that this book answers seven main questions:

What in the world are SANs or NAS, anyway?
: Where did they come from? How are they different (and better) than the technologies they are replacing? What are the basics of how they work? Chapter 1, *What Are SANs and NAS?*, Chapter 2, *Fibre Channel Architecture*, and Chapter 5, *NAS Architecture* answer these questions.

What is the difference between SANs and NAS?
: This is a really important question. If you can't understand the fundamental difference between these two technologies, you will never be able to answer the next question. See the end of Chapter 1.

Which is right for me?
: Obviously, I can't answer this question for you, but I can give you enough information to help make the decision. Hopefully after reading the book, you'll have a pretty good idea of how the two technologies can be implemented where you work and which is right for you. See the end of Chapter 1.

What kind of neat things can I do with a SAN that I can't do without one?
: You've heard about LAN-free and server-free backups, but you're not really sure what they are. In Chapter 4, *SAN Backup and Recovery*, they are explained in detail, including command-line information about how to do some of them yourself.

What kind of neat things can I do with NAS that I can't do without it?
: You've heard of dozens of ways to back up filers, but you're not sure what the differences are. Should you use NDMP, dump, snapshots, or some other technology? This is what Chapter 7, *NAS Backup and Recovery*, is about.

What kind of tasks will I find myself doing if I install a SAN?
: Now that you've put all of your storage onto a network—now what? What happens when the network goes down? How do you design it so that this doesn't

impact you? How do you even know that it went down? These answers are in Chapter 3, *Managing a SAN*.

What kind of tasks will I find myself doing if I install a NAS?
: How do you manage the volumes, snapshots, quotas, etc? What's it like to get users onto a NAS filer? What other tasks are you performing on a regular basis? See Chapter 6, *Managing NAS*, for the answers to these questions.

Conventions Used in This Book

The following conventions are used in this book:

Italic
: Used for program names, URLs, and for the first use of a term

`Constant width`
: Used in code examples and to show the output of commands

`Constant width italic`
: Used to indicate variables within commands

This icon designates a note, which is an important aside to the nearby text.

This icon designates a warning relating to the nearby text.

Comments and Questions

Please address comments and questions concerning this book to the publisher:

O'Reilly & Associates, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
(800) 998-9938 (in the United States or Canada)
(707) 829-0515 (international/local)
(707) 829-0104 (fax)

There is a web page for this book, which lists errata, examples, or any additional information. You can access this page at:

http://www.oreilly.com/catalog/sansnas

To comment or ask technical questions about this book, send email to:

bookquestions@oreilly.com

For more information about books, conferences, Resource Centers, and the O'Reilly Network, see the O'Reilly web site at:

http://www.oreilly.com

Acknowledgments

I got a lot of help on this book from old friends and new friends, and I wish to thank them publicly for their assistance. Without them, this book would not have come into being.

To God: Any abilities I have came from You.

To my wife, Celynn: Now you've got another book to help you sleep at night. Thank you again for all the love and support, and for taking care of our two girls while I tried to get this thing done.

To my older daughter, Nina: Thank you for continuing to ask me when I would be done with this book so that you could show it off in Mrs. Plunkett's class. Tell Alex A., Alex T., Andrew, Christina, Eduardo, Eric, Genesea, Ian, Jacqueline, K.C., Kyle G., Kyle K., Mason, Megan, Olivia, Rachel, Scott, Stephanie, and Tanner that I said hello. I love you, honey.

To my younger daughter, Marissa: Thank you for the constant breaks while writing the book, as you would come in to tell me you love me one more time. Hugs. I love you, too!

To my parents: Without your encouragement and support, I would have never had the *chutzpah* to attempt writing a book in the first place. Thanks!

To my wife's family: Thank you again for raising such a wonderful daughter and treating me as a member of your family.

To all those who made me who am I today (whatever that is): There are too many of you to list. This includes my friends and family growing up, my teachers and professors, and all of my coworkers over the years. Thank you for your contribution to my life.

To JF, JT, and ET: Thank you for pushing me out of the nest and letting me fly on my own.

To Todd Toles and Julie Stewart at Veritas; Grant Melvin, Greg Linn, and Eyal Traitel at Network Appliance; Roberto Basilio at Hitachi Data Systems; Charles Curtis at Compaq; Julie Stewart at Veritas; and David Eeoff at Legato: Without all your help, this book would be sadly lacking in data. Your answers to my constant questions and support while writing this book was invaluable. You are a credit to your respective companies, and I wish you and your companies continued success.

To Eyal Traitel: Thank you for writing the *Managing NAS* chapter. It's not easy writing a chapter to fit into somebody else's book, but you managed to pull it off. Thanks a bunch.

To Chely, from Eyal: "You are my one and only girlfriend and wife. Thanks to you, and to my always supporting family."

To Derek Brawdy: Thank you for helping me get the *Managing SANs* chapter started. Too bad you had to leave to go you know where and help them find you know whom. I hope you find him. Make us proud.

To the technical editors: This includes Stephen Potter, Todd Toles, Mark Perino, Melvin Grant, Greg Linn, Eyal Traitel, John Norman, Scott Aschenbach, Daniel Pigg, William Welty, Richard Hirtler, Chris O'Regan, Megan Restuccia, and Brian Kirouac. Thanks for keeping me in check. As always, the book is better because of you.

To O'Reilly & Associates: Thank you for the opportunity to bring another much-needed book to market.

To Michael Loukides: Thank you for sticking with me through this trying time. Thank you for your guidance and patience as I continually molded this book into what it is now.

To the reader: Thank you for purchasing this book. I hope it helps you to understand this confusing industry.

To everyone else: Buy the book, already! :)

In this chapter:

CHAPTER 1
What Are SANs and NAS?

Throughout the history of computing, people have wanted to share computing resources. The Burroughs Corporation had this in mind in 1961 when they developed multiprogramming and virtual memory. Shugart Associates felt that people would be interested in a way to easily use and share disk devices. That's why they defined the Shugart Associates System Interface (SASI) in 1979. This, of course, was the predecessor to SCSI—the Small Computer System Interface. In the early 1980s, a team of engineers at Sun Microsystems felt that people needed a better way to share files, so they developed NFS. Sun released it to the public in 1984, and it became the Unix community's prevalent method of sharing filesystems. Also in 1984, Sytec developed NetBIOS for IBM; NetBIOS would become the foundation for the SMB protocol that would ultimately become CIFS, the predominant method of sharing files in a Windows environment.

Neither storage area networks (SANs) nor network attached storage (NAS) are new concepts. SANs are simply the next evolution of SCSI, and NAS is the next evolution of NFS and CIFS. Perhaps a brief history lesson will illustrate this (see Figure 1-1).

From SCSI to SANs

As mentioned earlier, SCSI has its origins in SASI—the Shugart Associates System Interface, defined by Shugart Associates in 1979. In 1981, Shugart and NCR joined forces to better document SASI and to add features from another interface developed by NCR. In 1982, the ANSI task group X3T9.3 drafted a formal proposal for the Small Computer System Interface (SCSI), which was to be based on SASI. After work by many companies and many people, SCSI became a formal ANSI standard in 1986. Shortly thereafter, work began on SCSI-2, which incorporated the Common Command Set into SCSI, as well as other enhancements. It was approved in July

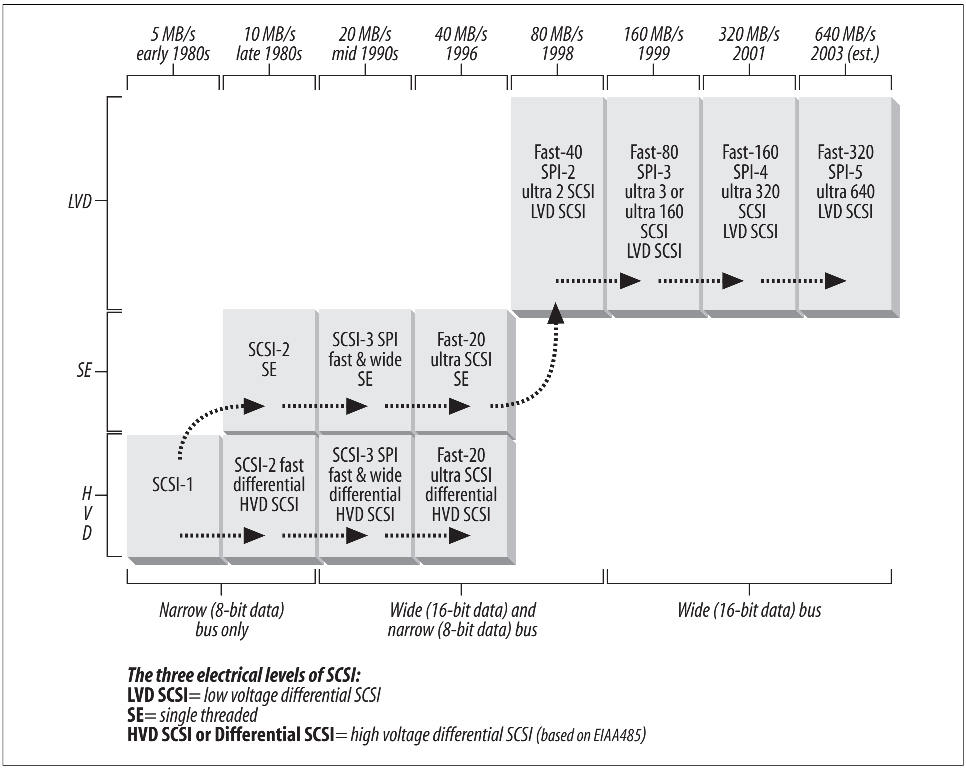


Figure 1-1. Generations of SCSI

1990.* Although SCSI-2 became the de facto interface between storage devices and small to midrange computing devices, not everyone felt that traditional SCSI was a good idea. This was due to the physical and electrical characteristics of copper-based parallel SCSI cables. (SCSI systems based on such cables are now referred to as parallel SCSI, because the SCSI signals are carried across dozens of pairs of conductors in parallel.) Although SCSI has come a long way since 1990, the following limitations still apply to parallel SCSI:

- Parallel SCSI is limited to 16 devices on a bus.
- It's possible, but not usually practical, to connect two computing devices to the same storage device with parallel SCSI.
- Due to cross-talk between the individual conductors in a multiconductor parallel SCSI cable, as well as electrical interference from external sources, parallel SCSI has cable-length limitations. Although this limitation has been somewhat

* This brief history of SCSI is courtesy of John Lohmeyer, the chairman of the X10 committee.

overcome by SCSI-to-fiber-to-SCSI conversion boxes, these boxes aren't supported by many software and hardware vendors.

- It's also important to note that each device added to a SCSI chain shortens its total possible length.

Some felt that in order to solve these problems, we needed to change the physical layer. The most obvious answer at the time was fiber optics. Unlike parallel SCSI, fiber cables can go hundreds of meters without significantly changing their transmission characteristics, solving all the problems related to the electrical characteristics of parallel SCSI. It even solved the problem of the number of connections, since each device in the loop had its own transmitting laser. Therefore, additional devices actually increase the total bus length, rather than shorten it.

The problem was, how do you take a protocol that was designed to be carried on many conductors in parallel and have it transmitted over only one conductor? The first thing that needed to be done was to separate the SCSI specification into multiple levels—a lesson learned from network protocol development. Each level could behave any way it wanted, as long as it performed the task assigned to it and spoke to the levels above it and below it according the appropriate command set. This was the beginning of the SCSI-3 specification. This separation of the various levels is why the SCSI-2 specification is contained in one document, and the SCSI-3 specification spans more than 20 documents. (Each block in Figure 1-2 represents a separate document of the SCSI-3 specification.)

Once this was done, it became possible to separate the physical layer of SCSI from the higher levels of SCSI. Once the physical layer was given this freedom, limitations caused by the physical layer could be addressed. You can see this separation in the SCSI-3 Architecture Roadmap in Figure 1-2, which was graciously provided by the T10 committee, the group responsible for defining the SCSI-3 architecture. It shows five alternatives to SPI (the SCSI Parallel Interface), including serial SCSI, Fibre Channel, SSA, SCSI over ST, and SCSI over VI. A relatively recent addition to this list that has been gaining acceptance is the iSCSI protocol. iSCSI uses IP as the transport layer to carry serial SCSI traffic.

As of this writing, iSCSI is gaining ground and market share but is still very new. Once it's in full swing, I'll prepare a second edition to this book that includes iSCSI coverage. More information about iSCSI is available in Appendix A.

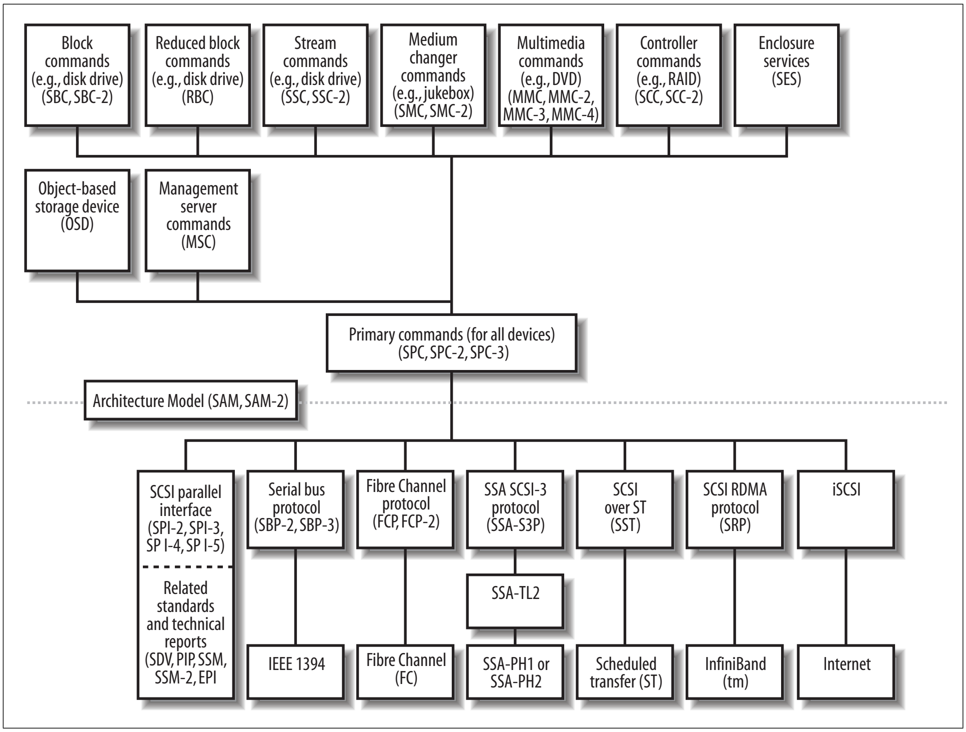


Figure 1-2. SCSI-3 Architecture Roadmap

The most popular alternative to SPI is the *Fibre Channel Protocol*. Fibre Channel, in contrast to SPI, is a serial protocol that uses only one path to transmit signals and another to receive signals. Fibre Channel offers a number of advantages over parallel SCSI:

Distance
: The distance between devices is no longer important. You can place individual devices up to 10 kilometers apart using single-mode fiber and theoretically go unlimited distances with bridging technologies such as ATM. (It's impractical at this time to use this for online storage due to the latencies involved in the long distance between the host and storage. However, most people using ATM bridging use it to make remote copies, not for live data access.)

Speed
: Although the bus speed of a single gigabit Fibre Channel connection is now slower than the fastest parallel SCSI implementations, you can trunk multiple Fibre Channel connections together for more bandwidth. (Also, Fibre Channel is much faster than the majority of installed SCSI today, which is usually 20 or 40 MB/s. Additionally, 2-Gb Fibre Channel is now available.)

Millions of devices connected to one computer

You can connect one million times more devices to a serial SCSI card (i.e., a Fibre Channel host bus adapter [HBA]) than you can to a parallel SCSI card. Parallel SCSI can accept 16 devices, and a Fibre Channel fabric can accept up to 16 million.

Millions of computers connected to one device

You can easily connect a single storage device to 16 million computers. This allows computers to share resources, such as disk or tape. The only problem is teaching them how to share!

> Most current implementations place limitations on the number of devices that can be connected to a single fabric. It's unclear at this time how close to 16 million devices we will ever get.

I should probably mention that Fibre Channel doesn't necessarily mean fiber optic cable. Fibre Channel can also run on special copper cabling. I cover this in more detail in Chapter 2.

What Is a SAN?

At this point, my definition of a SAN is as follows:

> A SAN is two or more devices communicating via a serial SCSI protocol, such as Fibre Channel or iSCSI.

This definition means that a SAN *isn't* a lot of things. By this definition, a LAN that carries nothing but storage traffic isn't a SAN. What differentiates a SAN from a LAN (or from NAS) is the protocol that is used. If a LAN carries storage traffic using the iSCSI protocol, then I'd consider it a SAN. But simply sending traditional, LAN-based backups across a dedicated LAN doesn't make that LAN a SAN. Although some people refer to such a network as a storage area network, I do not, and I find doing so very confusing. Such a network is nothing other than a LAN dedicated to a special purpose. I usually refer to this sort of LAN as a "storage LAN" or a "backup network." A storage LAN is a useful tool that removes storage traffic from the production LAN. A SAN is a network that uses a serial SCSI protocol (e.g., Fibre Channel or iSCSI) to transfer data.

A SAN isn't network attached storage (NAS). As mentioned previously, SANs use the SCSI protocol, and NAS uses the NFS and SMB/CIFS protocols. (There will be a more detailed comparison between SANs and NAS at the conclusion of this chapter.) The Direct Access File System, or DAFS, pledges to bring SANs and NAS closer together by supporting file sharing via an NFS-like protocol that will also support Fibre Channel as a transport. (DAFS is covered in Appendix A.)

It's common for a NAS filer to be comprised of a filer head with SAN-attached storage behind it.

In summary, a SAN is two or more devices communicating via a serial SCSI protocol (e.g., Fibre Channel or iSCSI), and they offer a number of advantages over traditional, parallel SCSI:

- Fibre Channel (and iSCSI) can be trunked, where several connections are seen as one, allowing them to communicate much faster than parallel SCSI. Even a single Fibre Channel connection now runs at 2 Gb/s in each direction, for a total aggregate of 4 Gb/s.
- You can put up to 16 million devices in a single Fibre Channel SAN (in theory).
- You can easily access any device connected to a SAN from any computer also connected to the SAN.

Now that we have covered the evolution of SASI into SCSI, and eventually into SCSI-3 over Fibre Channel and iSCSI, we'll discuss the area where SANs have seen the most use—and the most success. SANs have significantly changed the way backup and recovery can be done. I will show that storage evolved right along with SCSI, and backup methods that used to work don't work anymore. This should provide some perspective about why SANs have become so popular.

Backup and Recovery: Before SANs

A long time ago in a data center far away, there were servers that were small enough to fit on a tape. This type of data center led to a backup system design like the one in Figure 1-3. Many or most systems came with their own tape drive, and that tape drive was big enough to back up that system—possibly big enough to back up other systems. All that was needed to perform a fully automated backup was to write a few shell scripts and swap out a few tapes in the morning.

For several reasons, bandwidth was not a problem in those days. The first reason was there just wasn't that much data to back up. Even if the environment consisted of a single 10-Mb hub that was chock full of collisions, there just wasn't that much data to send across the wire. The second reason that bandwidth wasn't a problem was that many of the systems could afford to have their own tape drives, so there wasn't a need to send any data across the LAN.

Gradually, many companies or individuals began to outgrow these systems. Either they got tired of swapping that many tapes, or they had systems that wouldn't fit on a tape any more. The industry needed to come up with something better.

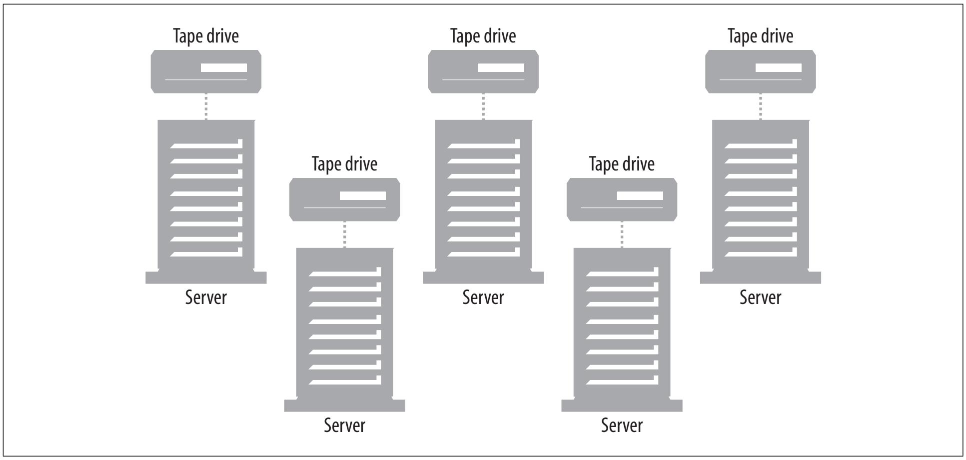


Figure 1-3. Backups in the good old days

Things Got Better; Then They Got Worse

A few early innovators came up with the concept of a centralized backup server. Combining this with a tape stacker made life manageable again. Now all you had to do was spend $5,000 to $10,000 on backup software and $5,000 to $10,000 on hardware, and your problems were solved. Every one of your systems would be backed up across the network to the central backup server, and all you needed to do was install the appropriate piece of software on each backup "client." These software packages even ported their client software to many different platforms, which meant that all the systems shown in Figure 1-4 could be backed up to the backup server, regardless of what operating system they were running.

Then a different problem appeared. People began to assume that all you had to do was buy a piece of client software, and all your backup problems would be taken care of. As the systems grew larger and the number of systems on a given network increased, it became more and more difficult to back up all the systems across the network in one night. Of course, upgrading from shared networks to switched networks and private VLANs helped a lot, as did Fast Ethernet (100 Mb), followed by Etherchannel and similar technologies (400 Mb), and Gigabit Ethernet. But some had systems that were too large to back up across the network, especially when they started installing very large database servers that contained 100 GB to 1 TB of records and files.

A few backup software companies tried to solve this problem by introducing the *media server*. In Figure 1-5, the central backup server still controlled all the backups, and still backed up many clients via the 100-MB or 1000-Mb network. However, backup software that supported media servers could attach a tape library to each of