



realtimepublishers.com<sup>tm</sup>

# *The Definitive Guide<sup>tm</sup> To*

# Building Highly Scalable Enterprise File Serving Solutions



*Chris Wolf*

---

## Introduction to Realtimepublishers

by Sean Daily, Series Editor

The book you are about to enjoy represents an entirely new modality of publishing and a major first in the industry. The founding concept behind [Realtimepublishers.com](http://Realtimepublishers.com) is the idea of providing readers with high-quality books about today's most critical technology topics—at no cost to the reader. Although this feat may sound difficult to achieve, it is made possible through the vision and generosity of a corporate sponsor who agrees to bear the book's production expenses and host the book on its Web site for the benefit of its Web site visitors.

It should be pointed out that the free nature of these publications does not in any way diminish their quality. Without reservation, I can tell you that the book that you're now reading is the equivalent of any similar printed book you might find at your local bookstore—with the notable exception that it won't cost you \$30 to \$80. The Realtimepublishers publishing model also provides other significant benefits. For example, the electronic nature of this book makes activities such as chapter updates and additions or the release of a new edition possible in a far shorter timeframe than is the case with conventional printed books. Because we publish our titles in “real-time”—that is, as chapters are written or revised by the author—you benefit from receiving the information immediately rather than having to wait months or years to receive a complete product.

Finally, I'd like to note that our books are by no means paid advertisements for the sponsor. Realtimepublishers is an independent publishing company and maintains, by written agreement with the sponsor, 100 percent editorial control over the content of our titles. It is my opinion that this system of content delivery not only is of immeasurable value to readers but also will hold a significant place in the future of publishing.

As the founder of Realtimepublishers, my *raison d'être* is to create “dream team” projects—that is, to locate and work only with the industry's leading authors and sponsors, and publish books that help readers do their everyday jobs. To that end, I encourage and welcome your feedback on this or any other book in the Realtimepublishers.com series. If you would like to submit a comment, question, or suggestion, please send an email to [feedback@realtimepublishers.com](mailto:feedback@realtimepublishers.com), leave feedback on our Web site at <http://www.realtimepublishers.com>, or call us at 800-509-0532 ext. 110.

Thanks for reading, and enjoy!

Sean Daily  
Founder & Series Editor  
Realtimepublishers.com, Inc.

Introduction to Realtimepublishers.....	i
Chapter 1: Moving Beyond Current File Serving Philosophies .....	1
State of the World.....	1
Performance Challenges .....	1
Management Challenges.....	1
Availability Challenges.....	2
Growth of Managed Data.....	2
Today's File Serving Landscape.....	2
Standalone Servers.....	3
DFS .....	5
NAS Appliances.....	7
Failover Clusters .....	8
Cluster Architecture .....	9
Shared Data Clusters.....	10
Current Storage Architectures.....	14
SCSI .....	14
SATA .....	15
FC and SANs .....	16
Switches and Hubs.....	17
Router.....	17
FCIP and iFCP .....	18
iSCSI.....	18
Clustered File Serving Gaining Momentum.....	19
High Availability .....	19
Consolidation Advantages .....	20
Drive Toward Standardization.....	20
Summary.....	21

## Copyright Statement

© 2005 Realtimedpublishers.com, Inc. All rights reserved. This site contains materials that have been created, developed, or commissioned by, and published with the permission of, Realtimedpublishers.com, Inc. (the "Materials") and this site and any such Materials are protected by international copyright and trademark laws.

THE MATERIALS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. The Materials are subject to change without notice and do not represent a commitment on the part of Realtimedpublishers.com, Inc or its web site sponsors. In no event shall Realtimedpublishers.com, Inc. or its web site sponsors be held liable for technical or editorial errors or omissions contained in the Materials, including without limitation, for any direct, indirect, incidental, special, exemplary or consequential damages whatsoever resulting from the use of any information contained in the Materials.

The Materials (including but not limited to the text, images, audio, and/or video) may not be copied, reproduced, republished, uploaded, posted, transmitted, or distributed in any way, in whole or in part, except that one copy may be downloaded for your personal, non-commercial use on a single computer. In connection with such use, you may not modify or obscure any copyright or other proprietary notice.

The Materials may contain trademarks, services marks and logos that are the property of third parties. You are not permitted to use these trademarks, services marks or logos without prior written consent of such third parties.

Realtimedpublishers.com and the Realtimedpublishers logo are registered in the US Patent & Trademark Office. All other product or service names are the property of their respective owners.

If you have any questions about these terms, or if you would like information about licensing materials from Realtimedpublishers.com, please contact us via e-mail at [info@realtimedpublishers.com](mailto:info@realtimedpublishers.com).

[**Editor's Note:** This eBook was downloaded from Content Central. To download other eBooks on this topic, please visit <http://www.realtimepublishers.com/contentcentral/>.]

## Chapter 1: Moving Beyond Current File Serving Philosophies

The challenges that face file serving have evolved over the past few years, and the methods used to meet those challenges have advanced as well. Today, many organizations view data availability as critical, allowing for very small windows of system downtime. Compounding the problems of maintaining data availability is the sheer volume of data that many organizations must manage. The industry has moved from needing gigabytes of storage a few years ago to eclipsing the terabyte or even petabyte range of managed storage.

This chapter will begin an exploration of how to build highly scalable enterprise file serving solutions by looking at the current state of the world of file serving. Along the way, you will see the many disk, server, performance, and availability choices at your disposal. After exploring the countless available options, the chapter will examine how Information Technology (IT) as a whole is modernizing its approach to file serving and data management. This chapter will provide the foundation on which to build the rest of the guide.

### State of the World

Today, file serving can be deployed in many shapes and sizes. Architecturally, there are several methods for designing and deploying file serving solutions. Many organizations don't just employ one idea or methodology but are often faced with managing a collection of disparate technologies.

#### *Performance Challenges*

Performance problems often follow the pattern of a pendulum—they flow from one extreme to another. On some levels, there are several servers not working up to capacity with physical resources under-utilized. On most networks, there are almost always other servers that are over-utilized, with users continually complaining about slow performance. On many networks, resources are present to solve the problems of high-volume file serving, but the distribution of the resources doesn't allow all file servers to cohesively meet demand.

#### *Management Challenges*

In addition to performance challenges, managing a high volume of servers is a difficult task. With each independent file server on your network, you are faced with the need to maintain system hardware, software updates, and antivirus software in addition to a host of other management tasks. To deal with the increased management requirements that are often the result of network sprawl, many organizations are looking to achieve the following:

- Consolidate for the purpose of managing and maintaining fewer servers
- Consolidate and manage storage centrally
- Scale on-demand
- Centrally manage a collection of servers as a single computing resource
- Reduce software costs such as operating system (OS) and application licensing costs

Besides the management challenges faced, file serving continues to be challenged by availability trials.

## **Availability Challenges**

In 2004, the Gartner Group determined that the average cost of downtime worldwide was \$42,000 per hour. They also found that the average network experiences 175 hours of downtime each year. Thus, based on Gartner's determinations, it should not take your organization long to recognize the importance of data availability. Even if an organization is far below the average downtime and is down for 100 hours in a year, that time would equate to potentially \$4,200,000 in lost revenue.

Although the cost of downtime may be obvious and is certainly backed by some pretty significant statistics from the Gartner Group, there are still countless organizations that simply deal with downtime as if it's an expected part of life in IT. In addition, many organizations believe that the cost of downtime is eliminated once systems are backed up. When a company's data is unavailable, their reputation may be damaged and customer confidence weakened as a result of the downtime. This is especially true with e-commerce. Potential customers will likely not return to an unavailable Web site and will look to other option to buy their needed solution. In many cases, if an organization's data availability is unreliable, potential customers will believe that the organization is also unreliable.

Although downtime for individual systems is inevitable, data does not have to be unavailable during that period. System patches, hardware, and software upgrades are a required factor for all networks, but the sole purpose of the network is to provide access to data. If one system must go down for maintenance, why must the data be unavailable? With clustered file serving, server maintenance or even failure will not significantly interrupt data access.

## **Growth of Managed Data**

Over the past decade, storage growth has repeatedly exceeded the projections of most network planners. Storage has continued to grow at an exponential rate, while the reliance each company has on electronic data has increased as well. The result has been a need to manage an abundance of storage while providing for fast access and high availability.

## **Today's File Serving Landscape**

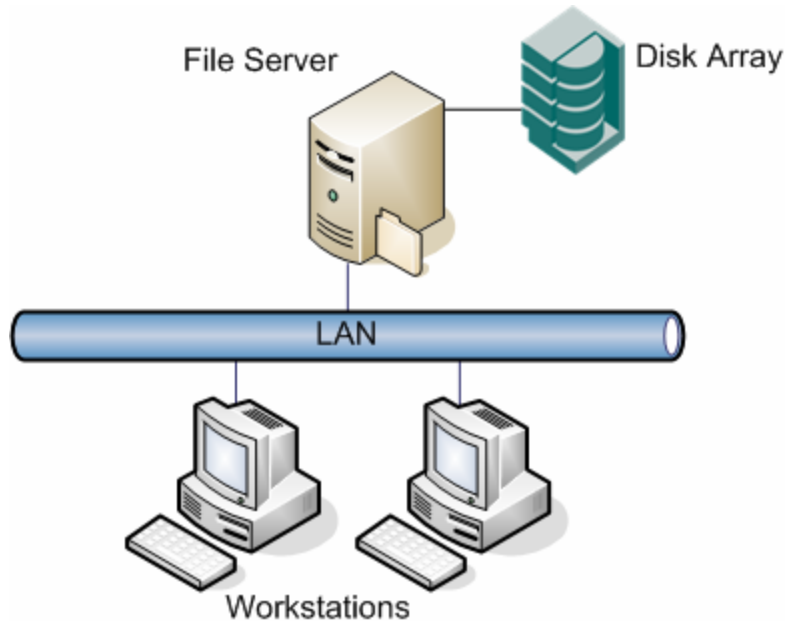
Years ago, file serving was pretty simple. Today, file serving is much more complex, and there are many approaches from which to choose. Today's approaches to file serving include:

- Standalone servers
- Distributed file systems (DFSs)
- Network Attached Storage (NAS) appliances
- Failover Clusters
- Shared data clusters

This section will look at the current role of each of these architectures as well as the advantages and disadvantages of each of these approaches.

## Standalone Servers

Standalone servers represent the root origin of file serving, and today maintain a very large presence in the file serving landscape. Figure 1.1 shows a typical standalone file server implementation.




**Figure 1.1: Standalone file server implementation.**

Notice in the figure that the storage scalability is addressed by attaching an external disk array to the server. Although the initial deployment and management of this type of architecture is usually simple at first, as the network scales, management generally becomes more difficult. The file server implementation that this figure shows is generally referred to as a *data island*. The reason is that access to the data is through a single path—the Local Area Network (LAN). Whether access is required for clients or for backup and restore operations, the data must be accessed over the LAN. For backup operations, this requirement might mean that backup and restore data is throttled by the speed of the LAN. A 100Mbps LAN, for example, would provide you with a maximum throughput of 12.5MBps (100Mbps divided by 8 bits).

Many organizations have combated the storage management shortcomings of standalone file servers by implementing either a dedicated LAN or a storage area network (SAN) for backup and recovery operations. Although this approach might solve immediate storage needs, it does little for scalability and availability. With the single file server acting as the lone access point for data access, several individual problems that occur with the server can result in the complete loss of data access. For example, any of the following failures would result in data unavailability:

- Hardware failure, such as CPU, RAM, or motherboard
- Network failure
- Power failure
- Disk failure
- Malware

Aside from any element of system hardware representing a possible single point of failure, having just one or even two access points to data can result in performance bottlenecks.

 Chapter 3 will look at ways to combat the availability and performance bottleneck issues associated with standalone file servers.

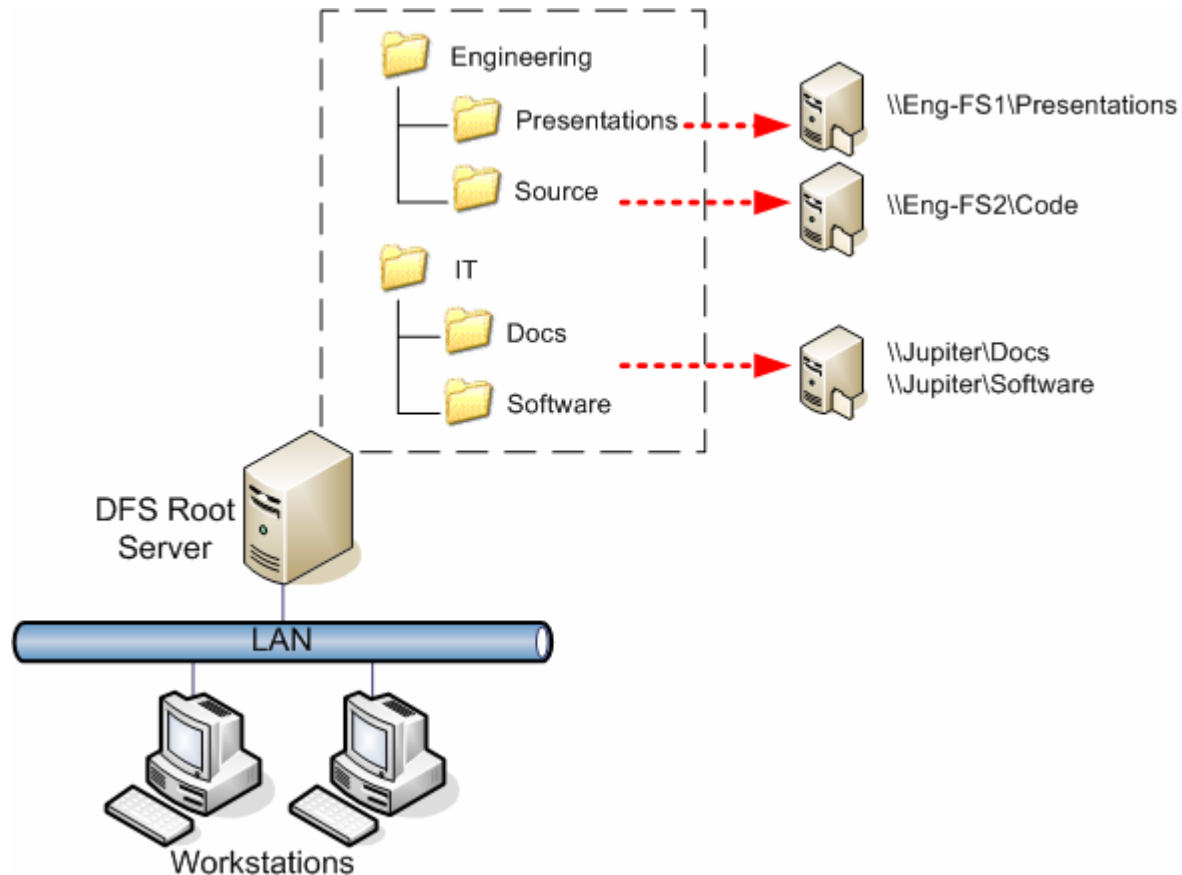
How do most organizations overcome file serving performance issues as their networks grow? Most simply add file servers. If one server is becoming overtaxed, an organization will order another server and move some of the shares on the overburdened server to the new server. This approach to growth is simple and has certainly been tested over time. However, adding servers also means that administrators have more systems to manage. This load will ultimately include additional work in hardware, software, and patch management. In addition, administrators will be faced with the task of updating login scripts to direct clients to the new servers. Thus, in addition to the cost of the new servers, there will ultimately be increased software and administrative costs associated with the addition of the new server.

Although adding servers to the network is an inevitable part of growth, there are other technologies that can assist in the scalability issues that surround file serving today. The next few sections will look at alternative methods that can be either substituted for or complement the addition of file servers to the LAN.



## DFS

The use of a DFS to manage file serving has been a growing trend in recent years. In short, a DFS enables the logical organization of file shares and presents them to users and applications as a single view. Thus, an organization's 200 file shares scattered across 12 servers may logically appear as if they're attached to a single server. Figure 1.2 illustrates the core concept of a DFS.




**Figure 1.2:** A simple DFS implementation.

With DFS, users can access network shares via a DFS root server. On the DFS root server, administrators can configure a logical folder hierarchy, then map each folder to a share located on another server on the network. Each physical location that is mapped in the DFS hierarchy is referred to as a *DFS link*. The link will contain the Universal Naming Convention (UNC) path to the actual location of the shared folder. When a user accesses a shared folder on the DFS server, the user will be transparently linked to another physical server on the network.


To illustrate this concept, compare DFS with traditional file serving—DFS enables administrators to present users a mapped network drive to access each share, and the administrators can simply map a single drive letter to the DFS root. Having a logical access layer in front of physical network resources offers several advantages:

- Administrators can change the physical location of shared data to support data consolidation or relocation without interrupting user access
- Replicas can be created for folders at the DFS root, allowing files to be replicated between multiple file servers
- With domain-based DFS, the DFS root can exist on multiple domain controllers, thus adding fault tolerance to the DFS root itself
- Windows DFS is closely intertwined with Active Directory (AD), enabling users to automatically be directed to shares that exist in their local site when multiple replicas of the same shared folders exist

 For more information about DFS, refer to Microsoft TechNet <http://www.microsoft.com/technet> and search using the key word “DFS.”

In being able to create replicas of DFS links, administrators can add a level of fault tolerance to the file serving infrastructure. Also, in being able to integrate with AD sites, users accessing a link that contains multiple replicas will be directed to the replica location that exists in their computers’ local site. The actual DFS root can also be replicated using domain-based DFS; thus, the DFS root will also be fault tolerant.

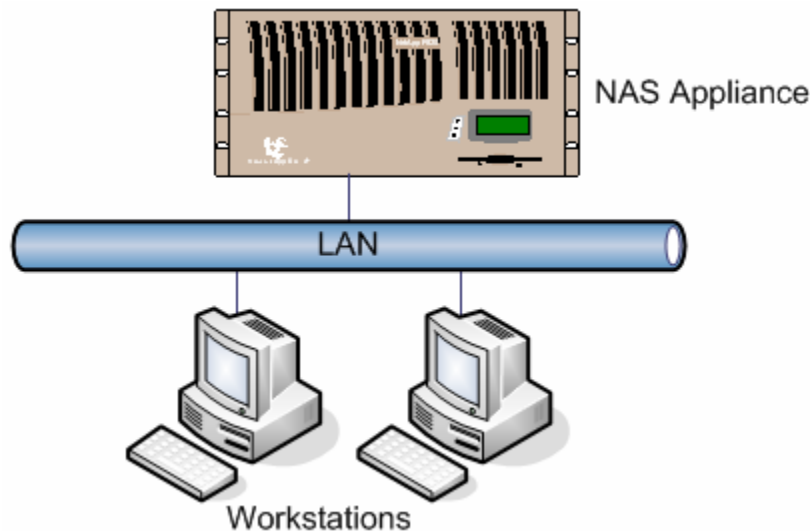
DFS solves a few of the scalability issues with file serving. With DFS in place, file servers can be added without having any impact on users and drive mappings. Availability can be increased by creating replica links for critical shares. If the replica links traverse two or more sites, an organization will also have simple disaster protection in place.

 DFS should not be considered a replacement for normal backups. Although DFS can transparently maintain multiple copies of files across two sites, it does not prevent file corruption, erroneous data entry, or accidental or intentional deletion. Thus, you should still back up your file server data to removable media and store it at an offsite facility.

Although DFS can solve some of the data access and availability concerns of standalone file servers, it does not help combat the server sprawl that administrators will have to contend with as additional servers are added to the network. Each server will still need to be maintained as a separate entity. DFS will hide the complexity of the network infrastructure to end users and applications, but administrators aren’t so fortunate. As the network grows, administrators will be faced with managing and maintaining each server on the LAN.

## NAS Appliances

NAS appliances began gaining momentum as a method to consolidate and simplify file serving in the late 1990s. NAS appliances quickly gained popularity as a result of the fact that they can be deployed quickly (often within minutes) and with support for up to terabytes of storage; several file servers are often able to be consolidated into a single NAS. Figure 1.3 shows a typical NAS deployment.



**Figure 1.3:** A simple NAS deployment.

NAS devices are labeled *appliances* because of the fact that an administrator can literally buy a NAS and plug it in. However, NAS devices have restricted software choices. By restricting the software that could be installed, if any, NAS vendors are able to guarantee the reliability of their systems. As most NAS appliances have a sole purpose of being file servers, there isn't much of a need to install applications.

Major vendors in the NAS space include Network Appliance, EMC, and Microsoft, which offers a Windows Storage Server 2003 OS. Network Appliance and EMC provide both hardware and their own proprietary NAS OS with each appliance. Microsoft does not ship NAS appliances. Instead, it provides a NAS OS to vendors such as Dell and Hewlett-Packard, who ship NAS appliances with the Windows Storage Server 2003 OS.

In being built for file serving, nearly all NAS appliances (including those from Network Appliance, EMC, and Microsoft) support the two most common network file sharing protocols: Common Internet File System (CIFS) and Network File System (NFS). Also, most NAS appliances include built-in redundant hardware as well as data management utilities.

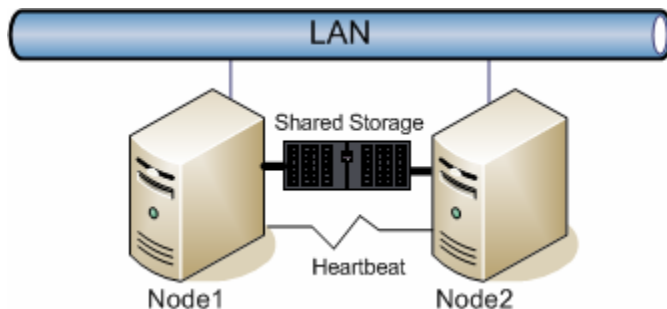
The popularity of NAS has been attributed primarily to its ability to be quickly deployed as well as the relative simplicity of administration of the NAS appliance. Nearly all NAS appliances come with a simple to use Web-based administration tool.

As with other file serving approaches, NAS has a few drawbacks. Most NAS appliances come with proprietary hardware and a proprietary OS. This shortcoming limits the flexibility of the device in the long run. For example, an older and slower NAS appliance cannot later be used as a database server. Also, the nature of proprietary solutions requires the purchaser to return to the same NAS vendor to purchase hardware upgrades. Another challenge that has recently plagued NAS is sprawl. For many network administrators that bought into the NAS philosophy of file serving, adding capacity means adding another NAS. In time, many organizations have accumulated several NAS appliances that are all independently managed.

## Failover Clusters

Another approach to file serving involves the use of *clusters*. The simple definition of a *cluster* is two or more physical computers collectively hosting one or more applications. A major advantage to clusters is in the ability for an application to be able to move from one node to another in the cluster. The process of an application moving to another node is known as *failover*. A shared storage device between all nodes in the cluster is needed so that an application will see a consistent view of its data regardless of the physical node that is hosting it. With these capabilities, when many think of the term *cluster*, they quickly realize the benefits of availability provided by clustering.

The two primary architectures available for file serving clusters are failover clusters and shared data clusters. The difference between these architectures lies in how the cluster's shared storage is accessed. With failover clustering, one node in the cluster exclusively owns a portion of the shared store resource. If an application in the cluster needs to fail over to another node, the failover node will need to mount the storage before bringing the application online. Figure 1.4 illustrates a failover cluster.



**Figure 1.4:** Failover cluster with SCSI-attached shared storage.

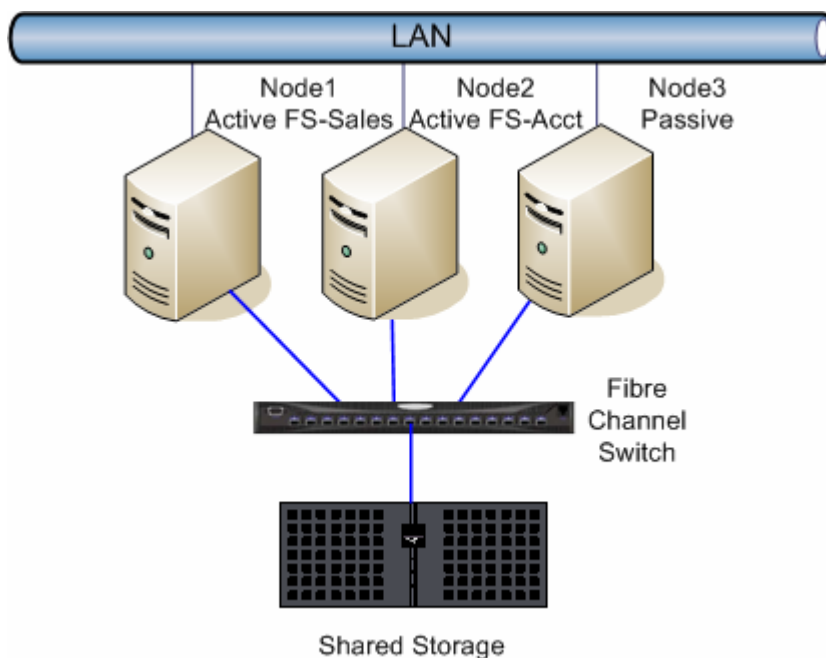
Notice that a heartbeat connection is also shown in the illustration. The heartbeat represents a dedicated network over which the cluster nodes can monitor each other. In this way, a node can determine whether another node is offline. If no dedicated heartbeat network is present, the cluster nodes will monitor each other over the LAN.

In a simple failover cluster, one node hosts an application, such as a file server inside a virtual server. The virtual server acts as an addressable host on the network and has a unique host name and IP address. The second node, the passive node, monitors the first node for failure. If the first node becomes non-responsive, the second node will assume control of the virtual server. Many popular OS vendors offer failover clustering support with their OSs. For example, Microsoft Windows Server 2003 (WS2K3) Enterprise Edition and Red Hat Enterprise Advanced Server 4.0 with the add-on Cluster Suite both support as many as 8-node failover clusters. The open source High-Availability Linux Project offers support for failover clusters of 8 nodes or more.

There are plenty of available failover clustering solutions on the market today. However, vendors are also starting to embrace shared data clusters, which offer the same level of fault tolerance as failover clusters, but several additional benefits as well.

## Cluster Architecture

Clusters are typically described as either N-to-1 or N-Plus-1. In an N-to-1 architecture, one node in the cluster is designated as the passive node, leaving it available to handle failover if an active node in the cluster fails. Figure 1.5 shows a 3-node N-to-1 cluster.



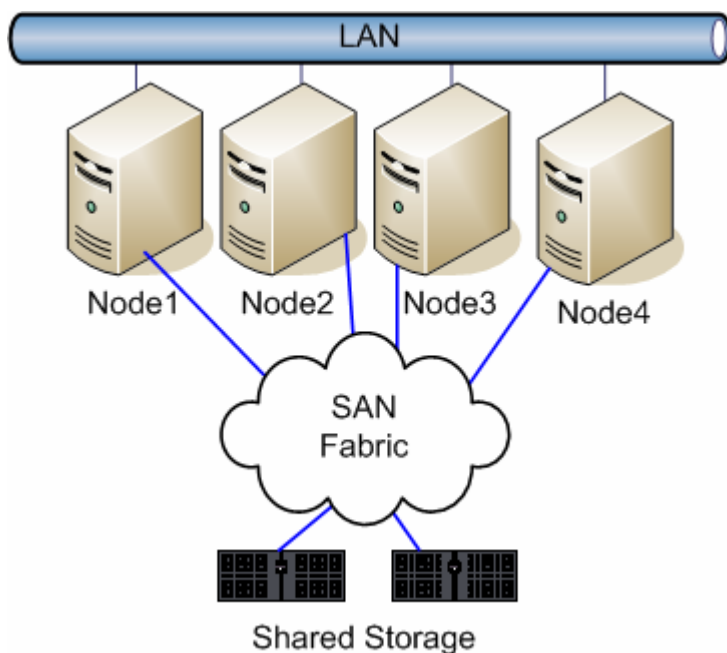
**Figure 1.5: A 3-node N-to-1 cluster.**

Notice that Node1 is active for the virtual server FS-Sales and Node2 is active for the virtual server FS-Acct. If either active node fails, Node3 will assume its role. In this architecture, Node3 is always designated as the passive node, meaning that when the primary active node returns online following a failure, the service will fail back to the primary node. Although this approach offers simplicity, having automatic fail back means that the failed service will be offline twice—once during the initial failover and again during the fail back.


N-Plus-1 clustering offers a different approach. With N-Plus-1, a standby (passive) node can assume control of a primary node's service when the primary active node fails. However, when the active node returns to service, it then assumes the role of the passive node. Thus, in time, the active node for each service managed by the cluster may be completely different than at the time the cluster was originally set up. However, automatic failback is not an issue with this approach, thus providing for better overall availability.

## Shared Data Clusters

Shared data clustering can also provide the benefit of high performance as well as load balancing. Shared data clusters differ from failover clusters in how they work with shared storage. In a shared data cluster, each node in the cluster simultaneously mounts the shared storage resources. The approach provides far superior performance over failover clusters because mount delays are not encountered when an application tries to failover to another physical node in the cluster. With shared data clusters, multiple nodes in the cluster can access the shared data concurrently; with failover clusters, only one node can access a shared storage resource at a time. Figure 1.6 shows a shared data cluster. Notice that one of the key differences with the shared data cluster is that a SAN is used to interconnect the shared storage resources.




**Figure 1.6:** Shared data cluster with SAN-attached storage.

 The elements of the SAN cloud are discussed later in this chapter.

Shared data clusters have steadily grown in popularity as a result of their ability to satisfy many of the problems facing today's file serving environments. In particular, shared data clusters can offer the following benefits:

- Provide more effective utilization of hardware resources
- Provide for simple scalability to accommodate growth
- Provide for high availability

Depending on whom you ask, industry analysts have found that average server CPU consumption runs from 8 percent to 30 percent. Most organizations have several servers that exhibit similar performance statistics. For example, consider an organization that has two servers that average 10 percent CPU utilization. Consolidating the servers to a single system will not only allow hardware to be more effectively utilized but also reduce the total number of managed systems on the network.

 Several organizations have turned to virtual machines as a means to further consolidate server resources. Companies such as VMware and Microsoft provide excellent virtualization tools in this arena. Although virtualization might make sense in many circumstances, a virtual machine is still a managed system and will need to endure patch and security updates as with any other system on the network. Virtual machines provide an excellent benefit in consolidation, especially when consolidating legacy OSs running needed proprietary database applications, but they are not always the best fit for file serving. Consolidating to virtual servers running on top of clusters not only allows you to maximize your hardware investment but also reduces the number of managed systems on your network.

Like traditional failover clustering, shared data or cluster file system architectures involve the use of virtual servers that are not bound to a single physical server. Virtual servers that exist in the cluster can move to another host if their original host becomes unavailable.

Where cluster file systems differ is in their fundamentally unique approach to clustering. In traditional clustering, each virtual server has its own data that is not shared with any other virtual server. In shared data cluster computing, multiple virtual servers can export the same data.

To summarize the key components of shared data cluster, consider the following common characteristics:

- **Modular**—Several dense servers are grouped to support mission-critical file serving and application needs.
- **Adaptive**—Physical resources in the cluster can be dynamically allocated to meet performance requirements.
- **High availability**—Virtual servers are enabled to fail over to available physical resources if a failure occurs.
- **Shared data**—Servers in the cluster concurrently access shared data via a SAN. Concurrent access provides for near instantaneous failover.
- **Platform independence**—Hardware of each node in the cluster does not need to be identical or even from the same vendor.
- **Management layer**—Intelligence exists that oversees and ensures cohesion of physical and logical elements in the cluster.

**Modular**

In being modular, the cluster should support the logical grouping of physical resources to support the demand and quantity of virtual servers that are needed. In being able to group both physical server resources as well as storage resources, management is relatively simple. On the outside, shared data clusters can look intimidating. For this architecture to succeed, it's important for the management of resources to be simple. Modularization provides this simplicity.


**Adaptive**

Shared data clusters have the ability to take advantage of both high-performance clustering and failover clustering. To meet the needs of applications, additional servers can be redeployed to virtual server groups to accommodate demand. Additional virtual servers and applications can usually be added with minimal to no investment.

**High Availability**

To support high availability, virtual servers in the cluster can failover to other nodes. If data access via one physical server in the cluster is interrupted, another physical server can take control of a virtual server in the cluster. Also, shared data clusters provides for a unique data sharing architecture that allows failovers to typically complete within seconds.

With file servers running as virtual servers hosted by a shared data cluster, data access does not need to be unavailable for several hours due to scheduled or unscheduled downtime. Instead, if a node in the cluster needs to go offline (or is taken offline by system failure), the application hosted by the node can simply be moved to another node in the cluster. With failover generally taking seconds to complete, user access would be minimally disrupted.

 Not all clustering products support application failover during upgrades. Some products will require all servers be taken down simultaneously during an upgrade. Administrators should consult their cluster product vendor prior to performing any cluster maintenance to verify that clustered applications will remain available during any system upgrades.

**Shared Data**

With shared data, many traditional failover cluster architectures employ a shared-nothing architecture. With shared-nothing clustering, one or more servers share storage, but in reality only one server can use a shared physical disk at a time. The argument for this approach has long been that concurrent I/O operations from multiple sources could corrupt the shared hard disk, so it is best that the disk only be mounted on one physical server at a time. Ultimately, this means that traditional architectures in which software running on the servers in the cluster simply will not run properly if multiple physical servers are concurrently accessing the same disk space. However, this architecture will result in slow failovers in the event of a failure due to one node needing to release the storage resource and then the failover node needing to mount the storage resource.

With shared data clusters, each node in the cluster mounts the shared storage on the SAN. Thus, during a failover, no delay is incurred for mounting storage resources. To insure data integrity, the cluster's management layer uses a distributed lock manager (DLM). The DLM allows multiple servers to read and write to the same files simultaneously. The DLM also provides for cache coherence across the cluster. True cache coherence is what allows multiple servers to work on the same application data at the same time. This feature is what allows shared data clustering to offer both high performance and high availability.



**Platform Independence**

In being platform independent, cluster computing allows the use of preferred hardware for the assembly of the cluster's inner infrastructure. Platform independence makes it much easier for organizations to get started with cluster computing, and as servers in the cluster age, those servers can potentially be used for other purposes within the organization.

**Management Layer**

The role of the management layer within cluster computing is to not only modularize physical resources such as servers and storage but also provide failover and dynamic allocation of additional resources to meet performance demands.

As shared data clusters are a new and different approach to clustering, there are currently few choices available that can provide the complex management functionality of cluster computing-driven server infrastructure. The lone vendor that can fully deliver shared data clusters today is PolyServe; however, there are other storage vendors that offer consolidation and availability solutions such as Network Appliance and EMC (but each of these solutions is hardware centric).

**Built to Scale**

Another aspect of shared data clustering that has led to its popularity has been its simple growth model. As load increases, nodes can simply be added to the cluster. Although many failover cluster architectures experience trouble scaling, shared data clusters that can run on both Windows and Linux OSs support scaling to 16 nodes or beyond. This type of flexibility eliminates much of the guesswork of growth and capacity planning. With shared data clusters supporting a high number of maximum nodes, administrators can add nodes as needed rather than purchase based on capacity that may be planned 18 months out.

**The Cost Factor**

Shared data clusters offer several advantages, but those advantages come with a price. Shared data clusters typically share a common storage source. The shared storage is usually interconnected to the cluster nodes via a fibre channel SAN. Although shared storage contributes to some of the benefits mentioned earlier (and several more discussed in Chapter 6), it comes at a higher cost than traditional direct attached storage (DAS). However, although cost can lead to initial sticker shock, the surprise often quickly passes when the cost of downtime is compared with the cost of the shared storage infrastructure with your data hosted on industry standard Intel-based architecture and the need to scale performance. To understand the savings, look past the cost of DAS on a single server. With shared storage, after the initial infrastructure investment, there is little difference in the cost of actual storage. When compared with the cost of Intel servers over proprietary UNIX or NAS appliances, the cost savings of shared data clustering is often estimated at 8 to 10 times the cost of the proprietary equipment. Thus, the shared data approach provides not only better utilization of storage resources, better availability, and better performance, but also substantial cost savings.

In terms of complexity, storage architectures are less intimidating once the technologies available have been explored. The following sections highlight these technologies.

## Current Storage Architectures

Today, there are several ways to deploy storage on a LAN. Among the most popular choices are:

- SCSI
- Serial ATA (SATA)
- Fibre Channel (FC)
- Internet SCSI (iSCSI)

This section will take a brief look at each of these technologies as they relate to building a better file serving infrastructure.

### SCSI

SCSI has long been the core storage architecture for high-performance file serving. Although this disk architecture has lost significant ground to FC, most organizations still employ several SCSI storage devices on their networks. The first generation of SCSI offered throughput of as fast as 5MBps; today Ultra320 SCSI can push data at a rate of as fast as 320MBps. With SCSI device support, the size of the SCSI bus will ultimately determine the number of devices that can be connected to the bus. For example, narrow SCSI has an 8-bit bus, which allows it to support as many as 8 devices, including the SCSI host bus adapter (HBA).

Wide SCSI has a 16-bit bus, which allows for support for as many as 16 devices. By using logical unit numbers (LUNs), SCSI buses can support more than this limitation. SCSI IDs are used to identify each device on the bus. By default, each SCSI HBA uses an ID of 7. For narrow SCSI, IDs of 0 to 7 are valid; whereas 0 to 15 are valid IDs for wide SCSI. Table 1.1 shows the different SCSI bus types available today.

Bus Type	Bus Width (Bits)	Bandwidth (MBps)	Maximum Cable Length (m)		
			SE	LVD	HDV
SCSI-1	8	5	6	---	25
SCSI-2	8	5	3	---	25
Wide SCSI	16	10	3	---	25
Fast SCSI	8	10	3	---	25
Fast Wide SCSI	16	20	3	---	25
Ultra SCSI	8	20	1.5	---	25
Ultra SCSI-2	16	40	3	---	25
Ultra2 SCSI	16	80	---	12	25
Ultra160 SCSI	16	160	---	12	---
Ultra320 SCSI	16	320	---	12	---

*Table 1.1: SCSI bus type comparison.*



Note that the table lists cable lengths only for a SCSI bus standard supported for a particular SCSI bus type. LVD cable lengths are not listed until Ultra2 SCSI, which was the first SCSI standard to support the LVD bus type.



For more information about SCSI, visit Gary Field's SCSI Info Central at <http://www.scsifaq.org>.

SCSI runs into major scalability problems with shared storage architectures. In nearly all failover cluster implementations, shared storage connected via SCSI supports a maximum of 2 nodes. This scalability limitation has led many organizations to move away from failover clustering. Although failover clusters can run on SANs, the products of many vendors still behave as if they're SCSI attached, thus diminishing their attractiveness.

For greater scalability, many organizations are moving toward shared data clusters that interconnect shared storage to cluster nodes via an FC SAN. Although FC provides the data transport in the SAN, FC disk arrays attached to the SAN may contain internal FC, SCSI, or SATA disks. With the ability to offer scalability and support for all major disk storage architectures, it's easy to see why FC has become the leading storage interconnect in the industry.

## **SATA**

SATA drives have become increasingly popular due to their lower cost (compared with SCSI) and comparable speeds. The first SATA standard provided for 150MBps data transfer rates. In response to this standard, SCSI vendors quickly met the challenge, and, in turn, SATA began to offer 300MBps with its SATA II standard. At 300MBps, SATA II is still slightly slower than Ultra320 SCSI, but is now a viable cost-effective option in high-performance file serving.

Also, many storage vendors have jumped on the SATA bandwagon, with several vendors such as Hitachi and Sun Microsystems offering SATA disk arrays. The rise of SATA has been pushed by several storage vendors that have built SATA storage devices that can be interconnected to FC SANs.



For more information about SATA, refer to the SATA International Organization homepage at <http://www.sata-io.org>.

## FC and SANs


Today, FC is the predominant architecture for interconnecting shared storage devices. The high adoption rate of FC has been fueled by its several advantages over SCSI:

- Speed—4Gbps FC mediums offer data transfer rates as fast as 512MBps
- FC SANs support as many as 16 million devices
- FC supports cable lengths as long as 10KM

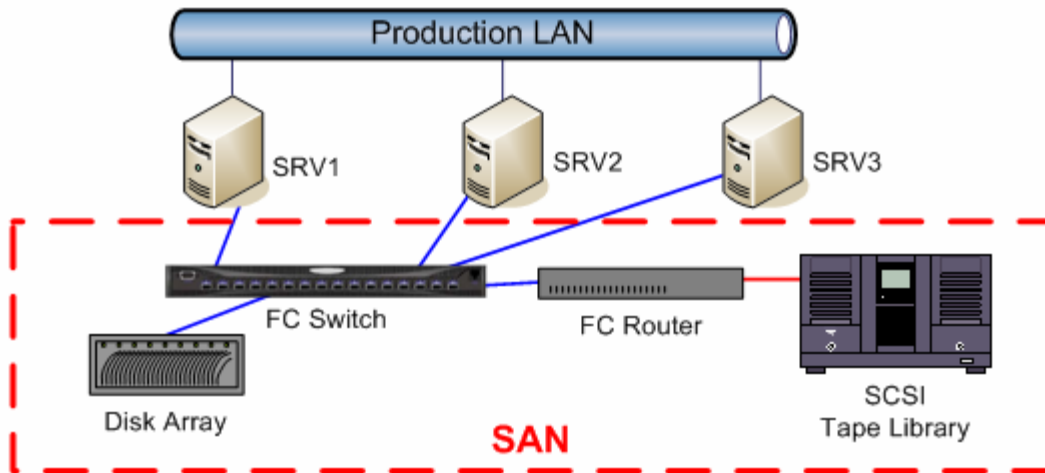
One of FC's greatest benefits is that this architecture allows for interconnecting storage devices via a dedicated SAN. SANs provide the following benefits:

- Storage resources can be pooled and shared by all servers
- Backup performance will likely increase dramatically
- Scalability issues can be more easily managed
- Shared data clusters can scale as high as 16 nodes or beyond, depending on the clustering application

Each server connected to the SAN can potentially access any storage resource on the SAN. SANs enable the maximized use of storage resources by creating better opportunities to allocate unused resources to other servers. This setup has significantly aided data backups. Now, a server no longer has to send its data over the LAN to access a tape library for backup, for example. Instead, the server can directly access the library via the SAN. Backup vendors such as Symantec (formerly VERITAS) and CommVault have architectures that support sharing of backup targets in a SAN. Now servers are no longer faced with network bottlenecks while backing up their data. The term *LAN-free* is often used to describe this backup approach. Other backup methods such as server-free and server-less are also available by using enterprise-class backup products and interconnecting storage resources via a SAN.

 Chapter 6 will provide examples of all SAN-based backup configurations, including LAN-free, server-free, and server-less as well as several examples of how organizations are consolidating storage resources by connecting their servers to SANs.

Disk arrays as well as backup devices can be shared on a SAN. In the past, many in IT addressed storage by guessing how much storage a server would need when it was initially requisitioned, and if the server needed more disk resources, more would be ordered at a later date. For servers for which the estimate was too high, disk resources would go unused. The ability to collectively pool physical disks in a SAN enables the allocation of disk space to servers as needed. The bottom line with SANs is that their implementation is a natural part of the progression toward consolidation. Figure 1.7 shows a basic SAN.



**Figure 1.7:** A SAN that consists of a switch, router, disk array, and tape library.

Notice that three servers are sharing a disk array and tape library. The switch and router are used to interconnect the storage devices on the SAN. FC SAN hardware devices share the same names of devices that you have already come to know and love with LANs. The primary devices that drive a SAN include:

- Switches and hubs
- Routers (also known as bridges)

## Switches and Hubs

Switches and hubs are used to interconnect devices on the SAN. Their role on the SAN is similar to a switch or hub on a LAN. Hubs are older FC devices that support a topology known as FC-Arbitrated Loop (FC-AL), which is the SAN equivalent to a token ring network. Switches dominate today's SAN landscape and work similarly to Ethernet switches. SANs connected via a switch are said to be a part of a Switched Fabric topology. This setup is similar to the Ethernet switches. With a switch, dedicated point-to-point connections are made between devices on the SAN, allowing the devices to use the full bandwidth of the SAN. With FC-AL hubs, bandwidth is shared and only one device can send data at a time. Among the popular switch vendors today are Brocade, McData, and Cisco Systems. Another very popular device on the SAN is the router.

## Router

Routers are devices that are used to connect an FC SAN to a SCSI device. The job of the device is to route between a SCSI bus and an FC bus. The router is a very important consideration when planning to implement a SAN, as it allows an organization to connect existing SCSI storage devices (disk arrays and libraries) to the SAN. This connection prevents the loss of the initial SCSI storage investment. The two most popular router vendors today are ADIC and Crossroads.

For more information about FC and SANs, refer to the excellent online resources: Storage Networking Industry Association at <http://www.snia.org>, Fibre Channel Industry Association at <http://www.fibrechannel.org>, and Legato System's SAN Academy at <http://www.sanacademy.com>.

## FCIP and iFCP

The cheapest transmission medium is the Internet, which requires IP. With this in mind, wouldn't it be useful to be able to bridge SANs in two sites together through the Internet? In order for this to happen, you will need a device capable of doing the FC-to-FCIP translation. Some FC switches have integrated FCIP ports that allow you to do so. However, FCIP doesn't provide any means to directly interface with an FC device; instead, it's a method of bridging two FC SANs over an IP network.

Internet FC Protocol (iFCP) is much more robust than FCIP. Like FCIP, iFCP can also be used to bridge FC switches over an IP network. However, this protocol also provides the ability to network native IP storage devices and FC devices together on the same IP-based storage network. With the rise of gigabit Ethernet networks, consider iFCP as a way to provide full integration between an FC and IP network. Another rising protocol that provides the same level of hardware integration over gigabit Ethernet is iSCSI.

## iSCSI

iSCSI works very similarly to iFCP, except that instead of encapsulating Fibre Channel Protocol (FCP) data in IP packets, SCSI data is encapsulated. In being designed to run over Ethernet, iSCSI enables the leveraging of existing Ethernet devices on a storage network. For example, consider an organization that purchases new gigabit Ethernet switches for an iSCSI SAN. As technology improves and the organization decides to upgrade to faster gigabit switches, the older switches can be used to connect hosts on the LAN. FC switches don't offer this level of flexibility.

iSCSI architecture involves a host configured as an iSCSI target. The iSCSI target can be a server with locally connected storage or a storage device that natively supports iSCSI. Clients that access the storage over the network using the iSCSI protocol are known as *initiators*. Initiators need to have iSCSI client software installed in order to access the iSCSI target. Figure 1.8 shows a typical iSCSI environment, showing two initiator hosts and one iSCSI target.

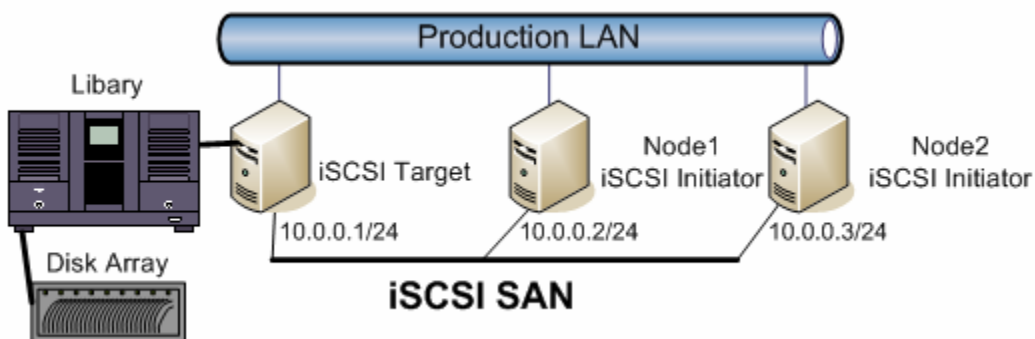


Figure 1.8: A small iSCSI SAN.

As iSCSI is a newer and maturing protocol, there are not as many storage devices that support iSCSI as those that support FC. As more devices become available, expect competition to cause the price of both iSCSI and FC SANs to drop even further.

## Clustered File Serving Gaining Momentum

To get past the reliance of data on an individual system, clustered file serving has emerged as the primary means for maintaining data availability. In short, clustering allows a *virtual server* to run on top of any physical server participating in the cluster. Virtual servers have the same characteristics as physical file servers—a name, IP address, and the ability to provide access to data. However, they differ in the fact that they are not dependent on a single piece of hardware to remain online. Instead, if a virtual server host's hardware fails, the virtual server can simply move to another host. The result is that the virtual server is only offline for a few seconds while moving to another physical host, compared with several minutes or hours of unavailability in the event of a server failure.

### High Availability

Keeping data available means keeping everything in the data path available. This goal is most often secured through redundancy. Storage itself can achieve redundancy through Redundant Array of Inexpensive Disks (RAID). Redundant switches can be added to the data path on the network, preventing against a switch failure. Redundant switches can be added to a SAN. Finally, physical servers themselves can be made redundant through clustering. Figure 1.9 illustrates an example of a highly available file serving architecture.

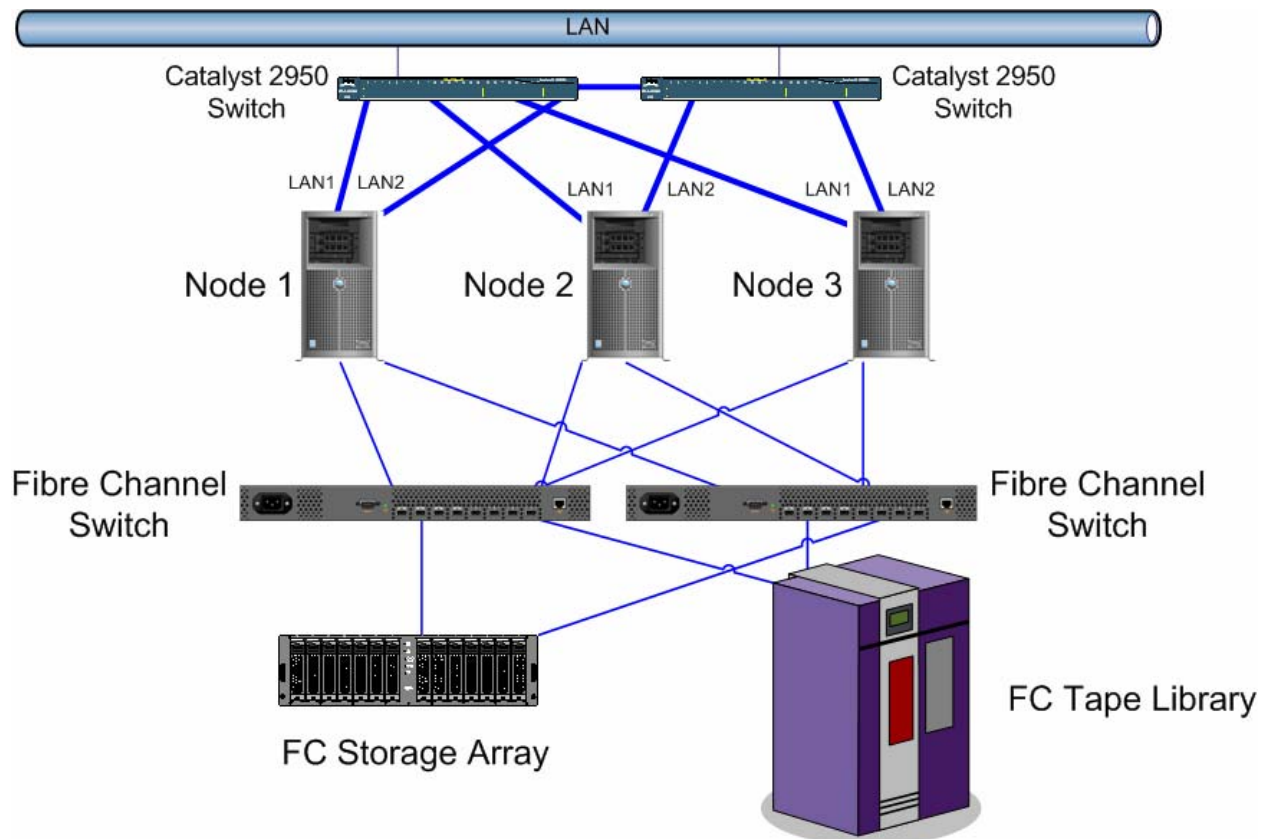



Figure 1.9: An example of a high-availability clustering architecture.

 Considerable time is spent exploring adding redundancy to the complete data path in Chapter 3.

## **Consolidation Advantages**

Newer cars have a lot more parts inside. Although the additional parts may equate to more features, such as power windows, these additions also mean that there are more parts that can break. On a network that employs 200 servers, each part on each server represents a potential failure. Reducing the number of servers on the network ultimately reduces the number of potential failures.

PolyServe recently studied the benefits of consolidating file servers to a clustered file system running on standard hardware and found the following:

- Procurement costs are reduced by as much as 70 percent
- Physical and logical file server use and storage consumption are reduced by as much as 80 percent
- Operational costs are reduced by at least 50 percent
- File server downtime is reduced by almost 100 percent

Thus, consolidating to clustered file system (CFS)-based file serving easily equates to quantifiable savings. An administrator who wants to lower the number of system management headaches needs a way to quantify proposals for new technologies in order to get them approved. If data unavailability is reduced from 175 hours per year to 1 hour per year, for example, an organization may see a production savings of more than 4 million dollars, according to the Gartner survey cited earlier.

## **Drive Toward Standardization**

Movement toward standardized hardware on Intel-based platforms has steadily gained ground over the past decade. Moving away from proprietary hardware solutions gives organizations true independence with their hardware investments. As hardware ages, it can be used in other roles, such as in application serving of a less critical database application. When mission-critical servers are upgraded, the original server systems can be used for other roles within the organization.

Having standard non-proprietary hardware also offers complete flexibility with OS and application choices. A Windows box could easily become a Linux box or vice-versa as the need arises. As needs on the network change, systems can be moved to where they're most needed. With proprietary solutions, this level of flexibility is typically not possible.

The push toward standard platforms has gone past the major OS vendors and extended to application and service vendors. Running servers on standardized hardware ultimately means far more applications are available to select from.

The bottom line with the movement toward standardization is that administrators and end users benefit the most. Organizations have better and less expensive products and much more to choose from when making purchasing decisions. The competition that has been steadily expanding in the non-proprietary market will only continue to benefit the industry with innovation fueling further competition.



## Summary

With increased need for performance and availability of files, shared data clusters have steadily emerged as the architecture of choice to meet many organizations' file serving needs. Shared data clusters offer superior scalability and a significantly lower cost than point-level proprietary solutions such as the offerings of many NAS vendors. With this type of momentum, it appears that shared data clusters will continue to experience rapid growth in the years to come. Deploying a shared data cluster architecture as part of a consolidated and highly available server infrastructure can provide a resilient and flexible architecture that can scale as an organization grows.

The next chapter digs deeper into the problems plaguing modern architectures and looks further into how these problems are being solved. The rest of the guide will explore specific examples of how to optimize the data path for performance and availability and provide examples of increasing performance, availability, and scalability of both Windows and Linux file serving solutions.

## Content Central

[Content Central](#) is your complete source for IT learning. Whether you need the most current information for managing your Windows enterprise, implementing security measures on your network, learning about new development tools for Windows and Linux, or deploying new enterprise software solutions, [Content Central](#) offers the latest instruction on the topics that are most important to the IT professional. Browse our extensive collection of eBooks and video guides and start building your own personal IT library today!

## Download Additional eBooks!

If you found this eBook to be informative, then please visit Content Central and download other eBooks on this topic. If you are not already a registered user of Content Central, please take a moment to register in order to gain free access to other great IT eBooks and video guides. Please visit: <http://www.realtimepublishers.com/contentcentral/>.