WHITE PAPER

IMPROVING SYSTEM AVAILABILITY WITH STORAGE AREA NETWORKS

A SAN-based high-availability infrastructure can help ensure maximum business continuance

BROCADE
The intelligent platform for networking storage

By reducing or eliminating single points of failure in enterprise environments, Brocade® Storage Area Networks (SANs) can significantly improve the overall availability of business applications. A networked storage approach enables a comprehensive system design—including SAN components as well as server and storage interconnects—that provides a much higher level of availability than single-point SAN products such as directors. In fact, Brocade switch-based SANs provide an extremely reliable foundation for increasing both data and system availability in a cost-effective manner.

Executive Summary

In networked systems such as SANs—with their associated servers, fabric, and storage components, as well as software applications—downtime can occur even if parts of the system are highly available or fault tolerant. To improve business continuance under a variety of circumstances, SANs can incorporate redundant components, connections, software, and configurations to minimize or eliminate single points of failure.

While some single-point director products base their claims of high availability on redundant hardware, those products do not necessarily account for the most common causes of system downtime: human error and software failures (according to a Gartner Group study of IT system downtime). In contrast, networked Brocade SAN fabrics can account for the full range of possible failure scenarios by employing redundant components along with a resilient network that greatly increases fault tolerance. As a result, Brocade dual-fabric SANs facilitate business continuance in many more failure scenarios than single-point director products (see Table 1).

Dual-fabric SANs supplement high-availability hardware by giving applications an independently managed path to data. This approach helps ensure ongoing business continuance even during failures caused by human errors, software issues, or site failures. In addition, dual-fabric SANs enable organizations to test and upgrade new SAN firmware and software without disrupting business applications. Moreover, Brocade offers a director-class product, the SilkWorm® 12000 Core Fabric Switch, that can further increase availability, especially at the center of a core-to-edge network model.

Table 1. Brocade dual-fabric SANs help guard against the greatest number of failure scenarios.

Failure Scenario	Protection with Single-Point Director	Protection with True High-Availability Dual Fabrics
Connection failure	Yes	Yes
Maintenance	Yes	Yes
Hardware failure	Yes	Yes
Nondisruptive firmware upgrades	Yes	Yes
Human error	No	Yes
Firmware failure	No	Yes
New software testing	No	Yes
Site failure	No	Yes

The Increased Need for Higher Availability

With the emergence of the Internet and the proliferation of global e-business applications, more and more organizations are implementing computing infrastructures specifically designed for continuous data and system availability. Today, even applications such as e-mail have become critical for ongoing business operations. Faced with increased customer and internal user expectations, organizations are currently striving to achieve at least 99.999 percent (the "five nines") availability in their computing systems—a figure equivalent to less than 5.3 minutes of downtime a year.

Any additional downtime can severely impact business operations and cost valuable time, money, and resources. To ensure the highest level of system uptime, organizations are implementing reliable storage networks capable of boosting the availability of data for all the users and applications that need it. These organizations typically represent the industries that demand the highest levels of system and data availability—the utilities and telecommunications sector, brokerages and financial service institutions, and a wide variety of service providers, for example. Companies such as Morgan Stanley, Intuit, and Depository Trust & Clearing Corporation require true high availability and have implemented dual-fabric Brocade SAN solutions to ensure that their businesses continue to operate on a 24x7x365 basis. Brocade helps provide a highly available foundation for these networks with scalable Fibre Channel switches and solutions that leverage high-availability components, devices, and configurations.

High-Level Availability Objectives

System availability is less the result of individual products or devices than it is an overall philosophy of system design. Developing highly available networks involves identifying specific availability requirements and predicting what potential failures might cause outages. The first step is to clearly define availability objectives, which can vary widely from one organization to another and even within segments of the same organization. In some environments, no disruption can be tolerated while other environments might be only minimally affected by short outages. As a result, availability is a function of the frequency of outages (caused by unplanned failures or scheduled maintenance) and the time to recover from such outages.

Many organizations are addressing their availability requirements by implementing networked fabrics of Fibre Channel devices designed to provide a high-performance storage environment. These flexible SANs are based on the following principles:

- A thorough understanding of availability requirements throughout the enterprise
- A flexible design that incorporates fault tolerance through redundancy and mirroring
- Simplified fault monitoring, diagnostics, and repair capabilities to ensure fast recovery
- A minimal amount of human intervention required during failover events
- A reliable backup and recovery plan to account for a wide variety of contingencies

To make sure systems can avoid or withstand a variety of failures, SANs incorporate a wide range of capabilities, including:

- Highly available components with built-in redundancy and hot-plugging capabilities
- No single points of failure
- Intelligent routing and rerouting
- Dynamic failover protection
- Nondisruptive server and storage maintenance
- Hardware zoning for creating safe and secure environments
- Predictive fabric management

System Availability through Redundancy

One of the simplest ways to increase availability is to implement fully redundant SAN environments to help ensure that alternate devices, data paths, and configurations can support enterprise applications during failures or other problems. A critical aspect of any high-availability system, redundancy helps prevent isolated failures from causing widespread outages. In fact, redundant systems are one of the only methods for preventing failures caused by software and human error, the leading causes of system downtime. Redundant systems also enable nondisruptive maintenance and upgrades—minimizing the impact on system operations.

Implementing multiple levels of redundancy throughout a SAN environment can reduce downtime by orders of magnitude. Hardware components, servers, storage devices, network connections, and even the storage network itself can be completely redundant. A fundamental rule for improving fault tolerance is to ensure dual paths through separate components. This is especially true when physical location and disaster tolerance are a concern, since a single device cannot adequately address these issues.

Applications

One of the keys to improving availability is shifting the focus from server availability and recovery to application availability and recovery. Mission-critical applications should be supported on clustered or highly available servers and storage devices to ensure the applications' ability to access data when they need it—even in the midst of a failure. Sophisticated software applications can enable application or host failover, in which a secondary server assumes the workload if a failure occurs on the primary server. Other types of software, such as many database applications, enable workload sharing by multiple servers—adding to continuous data availability where any one of several servers can assume the tasks of a failed server.

In addition, many server vendors and value-added software providers offer clustering technology to keep server-based applications highly available, regardless of individual component failures. The clustering software is designed to transfer workload among active servers without disrupting data flow. As a result, clustering helps organizations guard against equipment failures, keep critical systems online, and meet increased data access expectations.

Some clustering software, such as VERITAS Cluster Server, enables application failover on an application-by-application basis. This capability enables organizations to prioritize the order of application failover. Fibre Channel SANs facilitate high-availability clustering by simplifying storage and server connectivity. Moreover, SANs can provide one of the most reliable infrastructures for server clustering, particularly when clustered servers are distributed throughout the enterprise to achieve higher levels of disaster tolerance, a practice known as "stretched clusters."

Servers and Host Bus Adapters

To ensure high availability, servers should include redundant hardware components with the dual power supplies, dual network connections, and mirrored system disks typically used in enterprise environments. Servers should also have multiple connections to alternate storage devices through Fibre Channel switches and a minimum of two independent connections to the SAN. In most cases, servers should feature dual-active or hot-standby configurations with automatic failover capabilities.

The next single point of failure to consider after the server is the path between the server and the storage. Potential points of failure on this path might include Host Bus Adapter (HBA) failures, cable issues, fabric issues, or storage connection problems. The HBA is the Fibre Channel interconnect between the server and the SAN (replacing the traditional SCSI card for storage connectivity). Using a dual-redundant HBA configuration ensures that an active path is always available. In addition to providing redundancy, this configuration might enable overall higher performance due to the additional SAN connectivity.

To achieve true fault tolerance, these multiple paths are then connected to separate dual fabrics, ensuring that no failure scenario can cause a loss of both paths. Server-based software for path failover enables the use of multiple HBAs, and typically allows a dual-active configuration that can divide workload between multiple HBAs to improve performance. The software monitors the "health" of available storage, servers, and physical paths and automatically reroutes data traffic to an alternate path if a failure occurs.

In the event of an HBA failure, host server software detects that the data path is no longer available and transfers the failed HBA's workload to an active one. The remaining HBA then assumes the workload until the failed HBA is repaired or replaced. After identifying failed paths or failed-over storage devices and resolving the problem, the software automatically initiates failback and restores the dual path without impacting applications. If desired, an administrator can manually perform the failback to verify the process.

The software that performs this failover is typically provided by system vendors, storage vendors, or value-added software developers. Software solutions—such as Hewlett-Packard SecurePath, VERITAS Dynamic Multi-Path, and EMC PowerPath—help ensure that data traffic can continue despite a path failure. These types of software products effectively remove connections, components, and devices as single points of failure in the SAN to improve availability of enterprise applications.

To help eliminate unnecessary failover, the software distinguishes between actual failures and other network events that might appear to be failures. By recognizing false failures, the software can help prevent unnecessary failover/failback effects caused by marginal or intermittent conditions. After detecting an event, the software waits to determine whether the event is an actual failure. The typical delay in the failover process can range from an instant failover (when a loss of light signal is detected) up to a minute (if the light signal is still available and the path failure is in another part of the network). These delays are usually adjustable to enable a variety of configurations and to allow more rapid recovery mechanisms, such as path rerouting in the SAN, to occur.

Storage

To improve performance and fault tolerance, many of today's storage devices feature multiple connections to the SAN. Multiple connections help guard against failures that might result from a damaged cable, failed controller, or failed SAN component, such as an optical module. The failover process for storage connections typically follows one of the following methods.

One method is transparent failover, in which a secondary standby connection comes online if the primary connection fails. Because the new connection has the same address as the original failed connection, failover is transparent to the server connection, and application performance is not affected. After the primary connection is repaired, it assumes the workload.

Another method is to use dual or multiple active connections with each connection dedicated for certain logical volumes within a given storage system. If one connection fails, the other active connections automatically assume its logical volume workload until it comes back online. During this time, the alternate connections support all logical volumes, so there might be a slight performance impact depending on workload and traffic patterns.

A third method used for storage path failover also utilizes dual or multiple active connections. In this case, however, both connections can simultaneously access the logical volumes. This design can improve performance through load balancing but typically requires host-based software. During a storage connection failure, the alternate active connection continues to access the logical volumes. After the failed connection is repaired, the other path becomes active and load balancing resumes.

All of these failover methods are designed to ensure the availability of the enterprise applications that use them. In addition, failover generally is coordinated with server software to ensure an active path to data, transparent to the application.

Mirroring

Another effective way to achieve high availability in a SAN environment is by mirroring storage subsystems. SANs enable the efficient mirroring of data on a peer-to-peer basis across the fabric. These mirroring functions contribute tremendous fault tolerance and availability characteristics to SAN-based data. Combining the mirroring functions with switch-based routing algorithms (which enable traffic to be routed around path breaks

within the SAN fabric) creates a resilient, self-healing environment to support the most demanding enterprise storage requirements. The mirrored subsystems can provide an alternate access point to data regardless of path conditions.

A common use of mirroring involves the deployment of remote sites within the enterprise. Implementing SANs through Fibre Channel switches enables the distribution of storage and servers throughout a campus, metropolitan area, and beyond. Fibre Channel overcomes many of the distance limitations of traditional SCSI connections, enabling devices to be extended over much longer distances for remote mirroring, tape backup, and disaster recovery operations.

Native Fibre Channel supports server and storage connections at distances up to 10 km, which is adequate for most large campus applications. For higher availability solutions that require disaster tolerance, even longer distances might be required to connect disaster recovery sites. Brocade switches support up to 100 km Fibre Channel distances at 100 MB/sec bandwidth. This capability enables SANs to use existing long-distance optical connections or to connect to Metropolitan Area Networks (MANs) that employ Dense Wave Division Multiplexing (DWDM) technology. In addition, tunneling technologies can be used to connect SANs to existing Wide Area Network (WAN) IP infrastructures—increasing distances between remote sites well beyond 100 km.

Fabric Switches, Integrated Fabrics, and Core Fabric Switches

The SAN infrastructure itself is typically one of the highest availability components of storage networks. Fibre Channel switches are capable of extremely high reliability. Availability improves even further with switches that feature hot-pluggable, redundant power supplies and cooling, as well as hot-pluggable optic modules that enable single-port replacement of optics without impacting other working devices.

The Brocade SilkWorm 2000 family of switches has an availability rate of five nines (99.999 percent) as calculated by an observed Mean Time Between Failures (MTBF) of 749,000 hours for the overall system, including redundant components that further increase system uptime. The new Brocade SilkWorm 3000 series of switches provides the same high-availability features as the SilkWorm 2000 series, but adds twice the performance (at 2 Gbit/sec throughput) as well as other advanced features.

The SilkWorm 12000 Core Fabric Switch offers additional levels of redundancy with hot-swappable port cards and a redundant CPU. Dual CPUs enable organizations to upgrade and activate firmware nondisruptively and test new firmware on one of the CPUs prior to upgrading the second. An additional availability benefit of the SilkWorm 12000 is its ability to have two completely redundant 64-port switches in a single chassis. This design enables a redundant fabric with only a shared passive backplane between the two fabrics.

Although all members of the SilkWorm family of switches are designed to ensure the highest levels of hardware availability, the only way to truly ensure high availability in

all types of failure scenarios is to create a SAN design with dual fabrics. This approach eliminates single points of failure, not only from a hardware perspective, but also in the more likely case of human errors or software failures.

The ability to upgrade switches and integrated fabrics efficiently is important for testing new firmware within specific environments. Because a network of switches can provide alternate paths within a SAN—path failure can be handled transparent to applications—organizations can upgrade switches in the network without interrupting operations. Upgrades on switches with device connections are performed in conjunction with the dual-path capabilities of servers and storage, thus ensuring business continuance.

After an organization has tested the new firmware, it can be downloaded to other portions of the SAN. The ability to upgrade selected parts of the network or run different firmware versions within the SAN is a key advantage over single monolithic switch designs. For instance, a particular capability or fix for a device might need to be loaded onto only the applicable switch. Also, switch resellers often standardize on a particular version of switch firmware. As the SAN grows, it might include switches from many different resellers. As a result, organizations have the choice of continuing to use the supported firmware versions on particular switches instead of being forced to upgrade the entire system.

As scalability requirements continue to grow, organizations are seeking ways to incorporate higher end core products within their networks. These core products require higher performance than current Fibre Channel speeds: 2 Gbit/sec versus 1 Gbit/sec. In addition, core switch products should provide additional functions to ensure higher scalability, improved management, and multiprotocol support. These capabilities enable organizations to grow efficiently—enabling even larger and better performing SAN applications as well as a variety of network infrastructures.

New core fabric switch products—such as the Brocade SilkWorm 12000 Core Fabric Switch—are designed to meet both current and future core requirements for storage networking. With its speed-matching capabilities, the SilkWorm 12000 supports 2 Gbit/sec performance requirements while ensuring interoperability with existing 1 Gbit/sec SANs and Fibre Channel devices. This capability enables higher speed Fibre Channel backbones for applications that require more bandwidth than current SANs provide. Existing 1 Gbit/sec switches can be migrated to support edge connectivity as organizations deploy newer 2 Gbit/sec products in the core (see Figure 1). This approach is similar to the traditional IP-based networking model of outward expansion from the core.

By supporting future Fibre Channel technologies (such as 10 Gbit/sec Fibre Channel), the core fabric switch provides superior investment protection. Along with higher speed connectivity, a true core product should provide advanced functions to improve scalability, management, and security. For example, trunking (the ability to connect multiple ISLs together) enables high-bandwidth connections up to 8 Gbit/sec—a key component for ensuring a high-bandwidth core network.

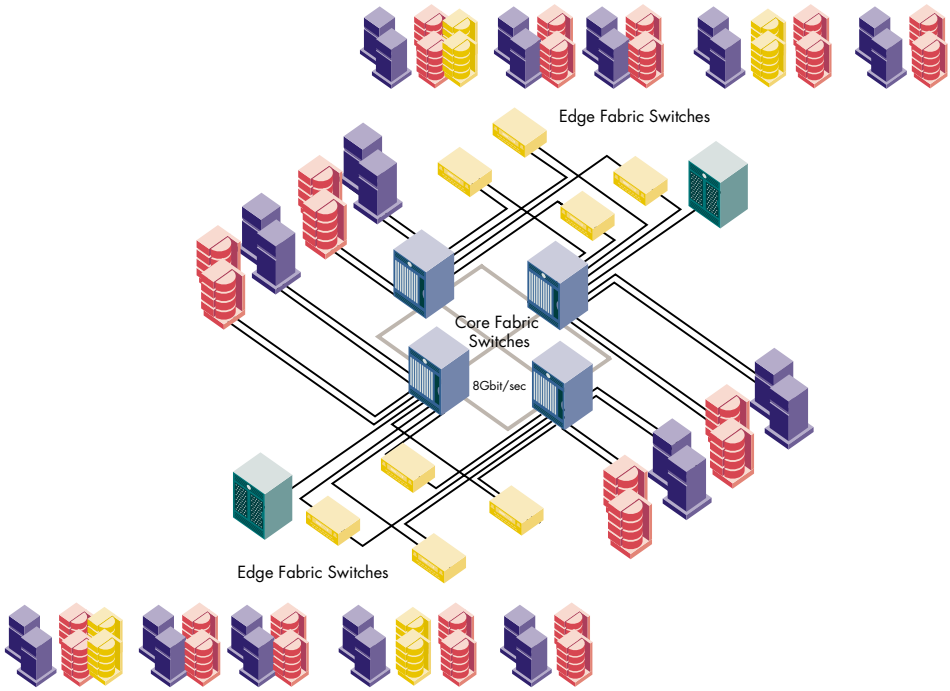


Figure 1.
A core-to-edge backbone networking model

In addition, new ASIC architectures enable Fibre Channel frame filtering capabilities that are not currently available in today's director products. Frame filtering provides a high level of advanced performance monitoring, such as monitoring end-to-end performance between a source and a destination across the fabric. Integrating this capability with storage management applications can improve monitoring while enabling capabilities such as per-usage billing.

An added benefit of frame filtering is improved security. Organizations can isolate frames down to individual protocols and restrict access by device and by the type of data being sent. In addition to increasing security, this capability improves control over the SAN fabric—especially when it is used for protocols other than SCSI.

Even with current director technology and the emergence of core fabric switches with their associated high availability, organizations should implement dual-SAN configurations to facilitate business continuance. Dual-SAN configurations help protect against the one major factor that a single device cannot: user error.

For instance, if an administrator accidentally zones out a server from its storage with a single SAN, both connections to the single director or switch would be affected. In contrast, a dual-SAN approach ensures that a path is always available, regardless of what problems occur in one SAN. After changes have been made and verified on one of the SANs, the changes can be applied to the other. This method helps ensure business continuance, as does a geographically distributed dual-SAN configuration that guards against localized physical disasters.

Another single point of failure is software, which can cause connected devices or even the entire single SAN or director to fail, leaving servers without a path to their storage. Again, with the dual-SAN approach, servers always have a connection to their storage to ensure business continuance even if a software failure causes a disruption in the connection.

Networking in the Fabric

Although dual SANs provide an excellent way to enhance availability, organizations should also utilize redundancy within the SANs to ensure the highest level of availability. Just as organizations have embraced client/server networking to overcome the limitations of the mainframe-centric IT infrastructure, many are taking a similar approach to SANs. A networked SAN is a flexible architecture that can be easily implemented and quickly adapted to changing requirements—extending the availability characteristics of hardware and software components directly into the SAN fabric.

Redundancy in the SAN fabric is built by networking switches, integrated fabrics, and core fabric switches—each with its own redundant components and separate control processors—to provide a robust mission-critical SAN solution. With dual-connected servers, switches, and storage, a networked fabric helps ensure high availability and business continuance. Because of its distributed nature, the network provides a more resilient infrastructure than any single-point product can. With an infrastructure of switches, organizations can also grow their network to meet high port-count needs. As new products provide actual core networking capabilities—such as 2 Gbit/sec support and trunking—organizations can build a higher speed core to meet future SAN performance requirements.

Networking a fabric of switches not only increases availability but also enables design flexibility and "pay-as-you-grow" scalability. After all, no matter how many ports it has, no single product can always meet every storage need. Because of this, networks are the only reliable way to achieve steady SAN growth.

With the advent of higher speed core products, the value will be in higher bandwidth capabilities rather than just increased port count. As a result, the rules will change for scaling the core fabric. For example, today's SAN infrastructures require port aggregation to solve problems such as backup and storage consolidation. For port aggregation large directors are overkill, since the network is used to provide "fan-out" connectivity to storage resources, and the bandwidth is limited not by the network but by the number of storage ports available. A network of smaller switches enables the SAN to support the appropriate level of bandwidth for these storage ports. In addition, a network can easily scale by increasing the number of switches for more server connectivity—thereby providing a lower overall total cost of ownership.

With future SAN technologies—such as virtualization, distributed databases, and global file systems—bandwidth requirements will undoubtedly increase and require a higher bandwidth core backbone to scale efficiently. To support these higher bandwidth requirements, organizations can implement a 2 Gbit/sec backbone that would still be compatible with legacy 1 Gbit/sec SAN infrastructures.

To help lower the cost of ownership and ensure uptime through efficient management, Brocade Fabric Manager software enables the management of groups of switches as well as multiple fabrics. This global approach simplifies the management of multiple elements with a single management console designed to increase user productivity. By enabling features such as automatic firmware loading and activation, topology mapping, roll-up status, and global policy management, Fabric Manager helps organizations focus on the SAN as a whole.

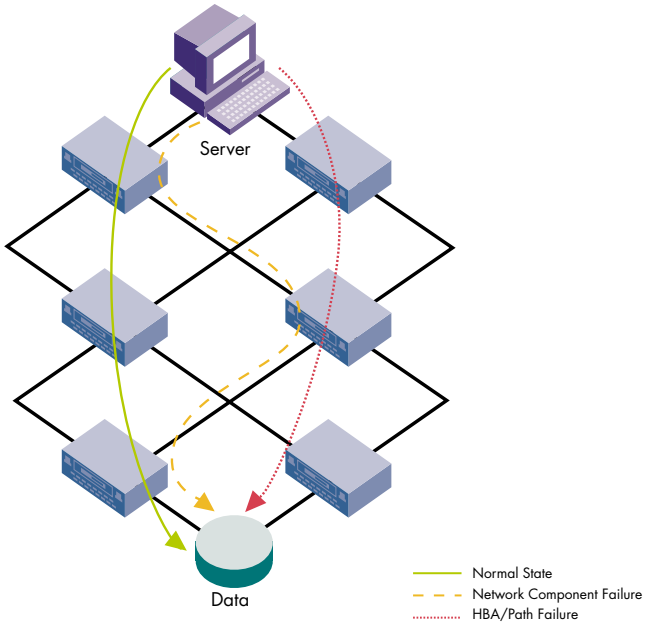
Beyond management, all Brocade fabric products automatically discover each other and any connected devices. Based on this discovery, these products create routing tables used by the entire fabric. No programming of the fabric is necessary, because device information is spread throughout the fabric as new devices log in. The network can scale in size at the edge as well as at the core switch level to provide higher bandwidth and redundant connectivity. In fact, SAN fabrics can feature multiple levels of availability—including meshed tree topologies of switches, single fabrics with dual connectivity, and dual fabrics with dual connectivity for environments that require the highest levels of availability.

Core-to-Edge Topology for Fabrics

One of the easiest ways to increase availability in a SAN is to network switches in a core-to-edge topology, which provides high performance and scalability. This topology connects devices to edge switches, which then connect to central interconnecting switches that in turn connect to other parts of the SAN or other devices (see Figure 2). A core layer of switches enables scaling of both the central bandwidth of the SAN and additional attachment of edge switches. As 2 Gbit/sec products are implemented, the central 1 Gbit/sec switches can be migrated to the edge to enable an even higher speed core fabric.

Starting from scratch, organizations can build a full 2 Gbit/sec solution to provide the highest port counts with the highest performance. For example, eight SilkWorm 3800s in a core-to-edge architecture can produce a 96-port SAN running at 2 Gbit/sec end to end. This number can be scaled down to 64 ports or scaled up to hundreds of ports by increasing the core or edge switches as needed.

Figure 2.
Failover in a single fabric with dual connectivity



A Single Fabric with Dual Connectivity

Dual attachment of servers and storage devices to a single fabric can contribute to a highly available system through the inherent reliability of the network itself. Although it is unable to protect against all failure scenarios (as described in dual-fabric designs), networking can improve overall system performance. The network redundancy concepts described in this section also improve the overall availability of each of the dual SANs covered in the following section.

Each high-availability switch within a networked fabric combines with the other switches to provide distributed functions such as error detection, login requests, state change notification, and management. The sharing of fabric functions enables workload sharing while helping to ensure that no single failure can disrupt the entire system. Because the switches run their own independent firmware and do not share memory like director-class products do, there is a much lower chance of a single switch impacting the entire network.

For instance, if a director-class switch experiences a firmware problem, it is likely to occur when the first CPU goes down and the secondary CPU comes online—a situation that might bring down the director itself. If there were a problem within the firmware, it would likely occur on both CPUs, possibly introducing a single point of failure within the SAN. As a result, organizations using a director or core switch need to ensure that the product can support multiple versions of firmware in case they need to quickly roll back the system to a previous version.

In contrast, the SilkWorm 12000 enables organizations to load new firmware, non-disruptively activate it, and test it while keeping the secondary CPU on the original firmware. This approach enables rapid rollback if a problem arises. In a network of switches, each switch runs its own code image—meaning that the possibility of a firmware failure is typically limited to only a single switch (which is not a single point of failure if there are
dual connections).

The high availability of the fabric stems from the near-statistical impossibility of experiencing a second failure during the repair of the first failure. The probability of two simultaneous hardware failures in the same fabric is extremely low, well into eleven nines of availability.

In addition, a Fibre Channel switched fabric can handle changes to the network—such as a link, optical, or switch failure—with minimal impact on all the connected devices. For example, if a central switch providing interconnectivity to other switches fails, the neighboring switches immediately detect the failure and inform other switches and devices within the fabric. Fabric Shortest Path First (FSPF), the standard routing protocol for Fibre Channel, detects the failed route, determines the next shortest route for data traffic, and updates the routing table. This change typically occurs within two seconds or less (the error timeout for frames within the fabric). By utilizing redundant links and alternate paths within the network, FSPF provides an effective way to cope with failed links as well as switch and SAN device outages (see Figure 2).

After a new route is established, data flow continues without requiring path failover from the storage or servers. From an application perspective, there is no downtime—just a brief pause for rerouting. The overall management of the failure is handled as a component of the system, and the management software is notified by the other switches in the fabric. This is a distributed monitoring process free of any single point of failure, as opposed to a dedicated management workstation.

Another example of a failed switch in a single networked fabric might involve an edge switch directly attached to storage or servers. If the failure is a redundant component, such as a power supply or fan, it does not affect the fabric or devices. If the failure is a port optic (GBIC or SFP) failure, only the single port is affected and the storage or servers' dual connection failover process occurs. Because the GBIC can be hot-swapped, replacement does not impact the switch or fabric.

If a motherboard fails on an edge switch in a networked fabric, the attached storage and server devices utilize their dual-path failover process. This process can be almost instantaneous, depending on several factors—such as the type of failure, the path failover software used (and its settings), and the type of storage used. Just like the motherboard on an edge switch, a port card failure on a director-class product initiates the path failover capabilities of the storage or server. This occurs because the card is not one of the redundant components of the system. From this perspective, both a network and a director would have to utilize the alternate path until the failed path is repaired.

Some storage and failover software applications immediately fail over if they detect a loss of light in the port optics, since it would be an obvious failure. Other failures in which the light is not lost might take longer to detect. This delay is typically used to filter out a thrashing event—such as repeated failover/failback attempts—due to intermittent or marginal conditions. After the failover occurs, an alternate path is used. This path might already be active, since most dual-connected devices support dual-active connections for load balancing and redundancy. As a result, the storage or server might detect only an increase in traffic rather than a change of state.

In dual-path configurations, only one side (server or storage) needs to initiate a path failover. For example, if a failure affects the storage, only the storage (not the server side) needs to alter paths. The fabric reroutes data from the server to the new storage connection.

To minimize repair time, all fabric switches should support auto-discovery. This capability enables a replacement switch to be added to the existing network and powered up, where it can then automatically discover the fabric and learn fabric-wide information such as zoning and name services. In addition, automatic recovery features can minimize recovery time and limit potential errors. These features provide easy scalability of the SAN when new switches are being added for increased connectivity.

A Dual Fabric with Dual Connectivity

Although a single SAN with dual connections can improve the availability of a system, a single fabric cannot protect against all failure scenarios. Using a dual fabric is the only way to protect against human error, site failures, and some software failures to provide even higher levels of availability. Because dual fabrics are approximately the same cost as a single fabric with dual connections (same port count, just separate SANs), it is highly recommended that the dual-SAN design be used to address the most failure scenarios and ensure business continuance. The new Fabric Manager software simplifies the management of multiple switches and multiple fabrics to allow a top-down approach and automate many of the management tasks. As a result, the management of a dual-fabric SAN is not necessarily more complicated than management of a single-fabric SAN.

Compared to single fabrics, dual-redundant fabrics simplify the concept of high availability and fault tolerance, because all components are redundant and parallel—from the server connections to the fabric itself. Any edge switch failure within one of the fabrics causes the redundant fabric to assume the additional data flow. The remote possibility of subsequent failures during any repair sequence means that the system can easily achieve much better than five nines of availability. Because the port counts are generally the same for a single large fabric and two smaller dual fabrics, the differentiating factor in availability levels is typically based on management capabilities and overall system design.

A primary difference between a single fabric and a dual fabric is the reaction during a failover. Because storage and server connections are typically split between dual fabrics, connections for both the source and destination paths might need to fail over to the alternate fabric (rather than just one side within a single fabric). In either case, failover is transparent to applications. Again, because most devices support dual-active paths, failover might involve only an increase of data flow to the alternate path rather than a state change from inactive to active. Any failures to a central switch within each fabric are routed around—enabling continuous operation without any failover on the storage or servers (see Figure 3).

Many organizations use dual fabrics in a geographically separate SAN-based disaster-ready environment to ensure that operations can continue if a disaster disrupts operations at one

of the locations. To ensure true high availability, the switchover from the primary site to the backup site must occur within a company's defined parameters of allowable downtime. Links up to 100 km are possible with Brocade Extended Fabrics firmware—a capability that supports long-distance disaster-tolerant networks by using DWDM or dark fiber lines in WAN or MAN configurations.

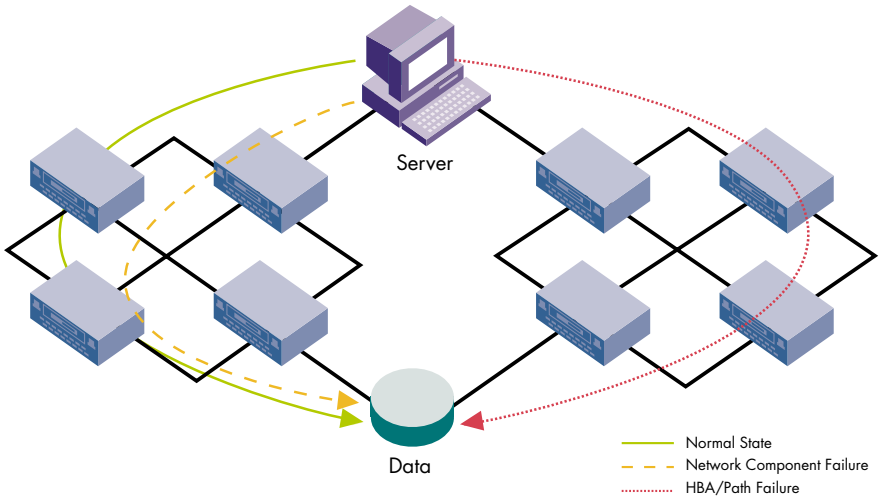


Figure 3.
Failover in a dual fabric with dual connectivity

In regard to the human error factor, the possibility of changing a zone inappropriately, moving a cable to the wrong port, loading an incorrect driver or firmware, shutting off power to a part of the network, and many other scenarios would disrupt a single fabric or director solution. In contrast, dual fabrics, especially if they are separated into different locations in the data center or even different data centers, ensure that a path will continue to operate until the problem is resolved. From the application perspective, the system continues to operate as normal.

Reliability Calculations for Single and Dual Fabrics

The reliability and high availability of either a dual fabric or a single fabric with dual connections is based on the same principles used to measure other network or RAID storage systems. Although the probability of a single failure is very low, the probability of another failure within the repair period of the first is nearly statistically impossible. Reliability is further increased by the fact that in a RAID system or a SAN fabric, downtime must be caused by a specific component failure—not just any component failure. Because the likelihood of another failure is so remote, the following calculation is often used as a conservative model for determining overall system reliability.

For example, in a 6-switch fabric with dual connections between devices, the availability during a single switch or port failure is 99.9995 percent. The probability of any secondary failure within the remaining 5-switch fabric is even lower in this conservative calculation. Refer to Figure 6 in Appendix A to see this calculation using the Markov probability modeling analysis.

For dual-redundant fabrics, only a failure in the other fabric needs to be considered since the active path would be in the active fabric. For example, in a 3-switch dual-redundant configuration—which provides approximately the same port count as a

single 6-switch configuration—the model would be as shown in Appendix A (Figure 7). The probability of a single switch failure within the dual 3-switch fabrics is similar to a single fabric. Availability would be 99.9995 percent. The probability of a second failure in the other 3-switch fabric during this repair cycle is only fractionally higher than in the single fabric. This difference is due to a fewer number of switches that can potentially fail.

These examples help demonstrate why organizations rely on highly available SANs to ensure business continuance. The fact that a single fabric has nearly the same availability rate as a dual-redundant fabric means that configuration decisions can be based solely on design philosophy, management infrastructure, and physical location requirements rather than availability concerns.

Zoning

The easiest way to increase system availability is to prevent failures from ever occurring—typically by monitoring fabric activity and performing corrective actions prior to an actual failure. By leveraging advanced SAN features such as zoning and predictive management, organizations can deploy a much more reliable and resilient SAN environment.

To help prevent localized failures from impacting the entire fabric, specific parts of SANs can be isolated through the use of zoning—in which defined zones limit access between devices within the SAN fabric. Organizations can specify different availability criteria at the connection, node, and network level to address the potential impact of certain types of outages. For instance, several minor outages in one environment might be much less destructive than a single large outage in another environment—even if the total amount of downtime is the same. The use of zoning helps limit the types of interactions between devices that might cause failures, and thus prevents outages.

Especially as organizations build larger SANs with heterogeneous operating systems and storage systems, zoning is an effective way to prevent failures. Although zoning can be implemented through either software or hardware techniques, hardware zoning provides the most secure method.

Software zoning in a Fibre Channel SAN fabric is enabled by the use of the Simple Name Server (SNS), a list of devices allowed in the zone. Devices within the zone need only check the list to learn what other devices are allowed. Only the absence of knowing about the other devices prevents a zoned device from communicating with them. Because software zoning does not physically block unauthorized data from being sent, it works under the assumption that all devices are well behaved and that they have no malicious intentions. An appropriate analogy is an unlisted phone number that is not in the phone directory. Anyone could still cause the unlisted phone number to ring, either accidentally or by getting the phone number through some other means.

Hardware zoning in a Fibre Channel SAN fabric provides an additional level of protection (see Figure 4). After a hardware zone is established, a switch creates a table of devices that can communicate with other devices in the fabric. Only traffic from devices in this zoned list is passed to the destination. All unauthorized devices are blocked and

dropped by the actual switch ASIC hardware. The analogy for hardware zoning is similar to caller ID blocking in the phone system, where only an approved list of phone numbers can cause the telephone to ring. Any numbers not on the list would get a busy signal and not be connected.

Brocade has provided secure hardware zoning since the introduction of the SilkWorm 2000 family. With the release of the Brocade 2 Gbit/sec family of switches such as the SilkWorm 3800 and SilkWorm 12000, zoning enables all types of hardware-based zoning for higher security. Whether specified at the port or device level, the zone is now hardware-based.

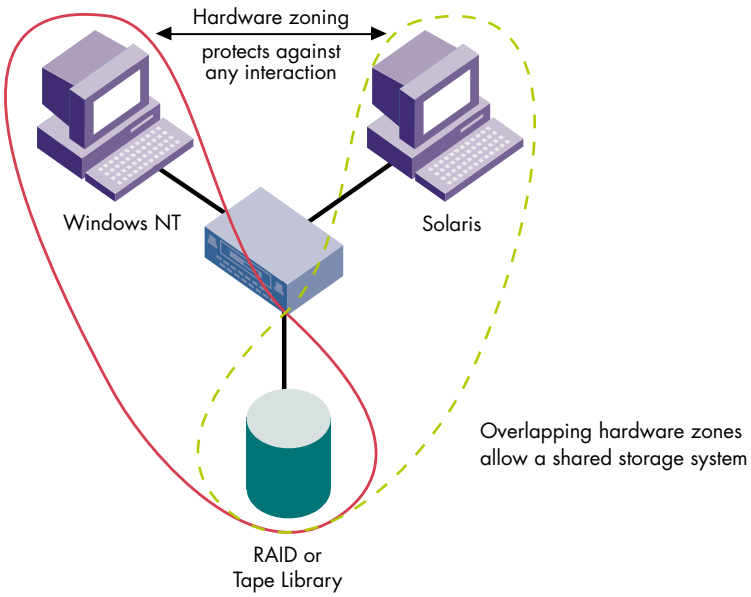


Figure 4.
Hardware zoning to ensure reliability in a heterogeneous SAN

Overlapping hardware zones allows devices to share resources such as a tape library or RAID connection, but not be in the same zone. As a result, this approach increases security within the SAN. However, not all products provide this detailed level of zoning. To enable the most secure hardware-enforced zoning, organizations should consider overlapping zones a critical requirement.

By blocking the flow of unauthorized data and control information, hardware zoning can restrict interaction between devices within the fabric. For example, a server in a heterogeneous environment might attempt to log in to all devices within the fabric, whether those devices represent other servers or storage. If the server resides in a zone with all the other servers, it could cause unnecessary and potentially erroneous data flow when attempting to communicate with them. If login attempts occur repeatedly, they could disrupt operations or even cause a failure in other server connections. As a result, zoning servers—especially heterogeneous severs—provides an additional level of assurance against potential failures.

Without zoning, failing devices that are no longer following the defined rules might attempt to interact with other devices in the fabric. (This type of event would be similar to an Ethernet device causing collisions on only a segment of a network.) With zoning, these failing devices cannot affect devices outside of their zone.

Other critical requirements for zoning are the ability to overlap hardware zones and to implement fabric-wide hardware zoning. The ability to overlap hardware zones is essential for secure storage resource sharing. The ability to share storage resources through logical volume allocation requires that multiple servers share a physical path into the storage. If hardware zones cannot overlap, all devices that utilize this physical path must reside in the same zone—creating potential interaction and disruption between devices. Overlapping hardware zones enable each server (or groups of servers) to reside in a zone with the storage connection while another hardware zone can also see the storage connection. The two distinct zones and all their devices cannot communicate with each other—only with the shared storage system.

To be truly effective, hardware zoning must be available across the enterprise fabric. Hard zones should be able to transcend physical switch products and allow devices to be securely zoned, regardless of their location in the fabric. Fabric-wide zoning, which enables information to be distributed to all switches in the fabric, can be accessed and managed from any switch in the fabric. In addition, the flexibility to locate devices at the most appropriate place within the fabric can help ensure reliability by enabling remote physical placement of devices and connections.

Fabric Management

As the network grows and interconnectivity expands across the enterprise, SAN manageability plays a greater role in ensuring high availability. Comprehensive management tools can simplify administrative tasks while helping to identify potential failures. One of the most effective ways to ensure the availability of a SAN is to use predictive threshold monitoring systems that enable organizations to predict and monitor basic environmental levels, as well as fabric-level events and errors over time.

Advanced fabric-wide switch monitoring and diagnostics enable both reactive and proactive servicing—helping to ensure rapid identification of failed components. Brocade Fabric Watch, a robust fabric-monitoring software tool, enables each switch in the SAN to constantly monitor the network and its attached nodes for potential faults—automatically detecting potential problems before they create costly outages. This innovative tool enables organizations to:

- Receive event notifications when switch and fabric elements exceed thresholds
- Quickly identify and isolate faults
- Utilize a single comprehensive fabric monitoring solution rather than multiple, vendor-specific software solutions

For example, Fabric Watch enables organizations to set thresholds to alert them to performance issues within the fabric (see Figure 5). The threshold could be set for a particular device or a connection between switches, for example. This technique enables organizations to potentially add a server or a new connection prior to impacting application performance or potentially taxing the server and causing a failure.

Threshold monitoring could also include error statistics or link status information to alert users to a possible cable or optics problem prior to that problem impacting operations. As a result, predictive failure capabilities can help resolve a situation before a failure occurs.

To enable more efficient management of a single or dual SAN, Fabric Manager incorporates Fabric Watch and other management features to roll up the events into a single view. This approach allows organizations to treat the fabrics as a global entity rather than manage each switch separately. Although managing a single device might appear to be easier, having software that allows individual switches to interact as one enables the same management simplicity while improving scalability and availability.

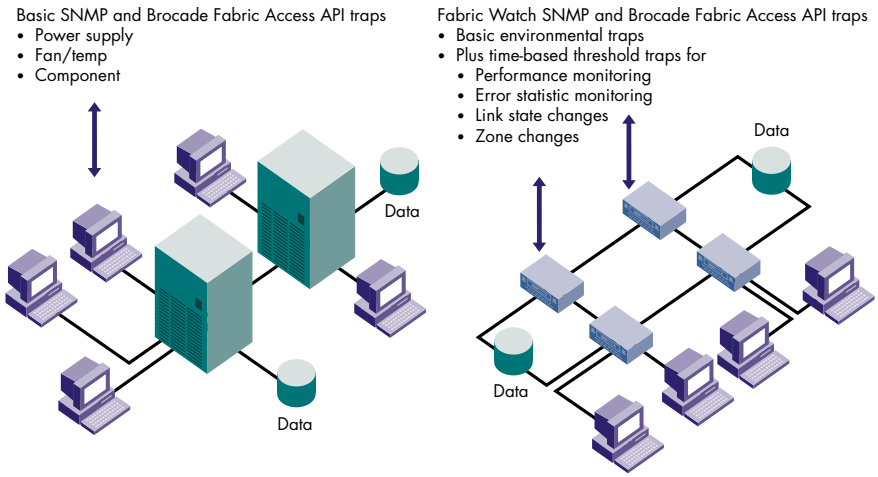


Figure 5.
Visibility into the SAN from Fabric Watch

The Key to High Availability

Achieving higher availability through redundancy and fault tolerance begins with a thorough understanding of specific system uptime requirements. True high availability stems from the use of highly available components in a networked infrastructure designed to tolerate a variety of failures.

Key points to consider in fault-tolerant design are how much of the system would be affected by a certain type of failure, and how long it would take to repair the problem. Proper system design—as well as the use of management software and zoning techniques—can help ensure that an isolated failure at one location does not lead to multiple failures in other areas of the network.

A critical factor in eliminating the possibility of a complete system outage is avoiding all potential single points of failure—through redundancy of components, devices, connections, and paths. In addition, physically separating devices can help build fault tolerance by protecting against localized physical disasters. Because director-class products are more difficult to distribute than a network of fabric switches and core switches, they are more vulnerable to localized disasters, and represent a potential single point of failure that might impact the fault tolerance of the entire system.

The ability to tolerate failures and ensure continuous system operation is the key factor in providing high availability for the overall system. Multiple connectivity paths, clustering techniques, and dual fabrics all contribute to a fault-tolerant solution. In addition, the use of redundant paths, components, and software can lower the MTBF rate to a point where large networked systems can achieve true high availability.

Today, Brocade switch-based SAN fabrics are providing a resilient solution to help leading financial, service provider, utility, and telecommunications organizations ensure high data availability and the maximum system uptime possible. As a result, organizations that deploy networked SAN fabrics are strategically positioned to achieve maximum business continuance while minimizing the risk of costly outages.

Along with switches, Brocade provides the software and features required to deploy highly available enterprise SAN infrastructures. Brocade SAN environments utilize redundant components, multiple connections, mirroring and clustering techniques, and multiple connections—as well as single- and dual-fabric configurations. These elements all contribute to an extremely resilient SAN infrastructure that helps maximize the overall availability of enterprise applications. As a result, organizations receive an excellent return on their technology investment and the assurance that their business is well protected.

For more information about Brocade high-availability SAN solutions, visit **www.brocade.com**.

Appendix A.

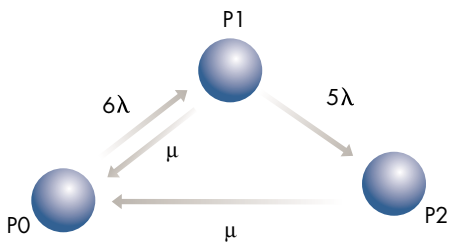


Figure 6.
Availability model for a 6-switch redundant fabric

$6\lambda P0 = \mu(P1 + P2)$
$P1 (5\lambda+\mu) = P0\ 6\lambda$
$P2\ \mu = P1\ 5\lambda$

Given:
P0 + P1 + P2 = 1
P0 = normal conditions
P1 = single failure, system up
P2 = multiple failures, system down
λ = failure rate = 1.39×10^{-6} failures/hr
μ = repair time = 0.5 hours with spare

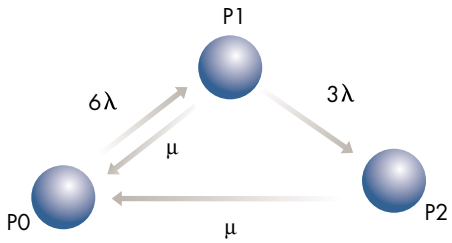


Figure 7.
Availability model for a 3-switch dual-redundant fabric

$6\lambda P0 = \mu(P1 + P2)$
$P1 (3\lambda +\mu) = P0\ 6\lambda$
$P2\ \mu = P1\ 3\lambda$

Given:
P0 + P1 + P2 = 1
P0 = normal conditions
P1 = single failure, system up
P2 = multiple failures, system down
λ = failure rate = 1.39×10^{-6} failures/hr
μ = repair time = 0.5 hours with spare

Corporate Headquarters
San Jose, CA USA
T: (408) 487-8000
info@brocade.com

European Headquarters
Geneva, Switzerland
T: +41 22 799 56 40
europe-info@brocade.com

Asia Pacific Headquarters
Tokyo, Japan
T: +81-3-5402-5300
apac-info@brocade.com

Latin America Headquarters
Miami, FL USA
T: 305-716-4165
latinam-sales@brocade.com