Technical Report

Fibre Channel SAN Best Practices

Mathew Devanny, Daniel Chan,
Field Centers for Innovation, NetApp

March 2012 | TR-4017

TABLE OF CONTENTS

LIST OF FIGURES

1 Introduction

1.1 Storage: A High-Availability Requirement

First started in 1988 and got ANSI standard approval in 1994, Fibre Channel (FC) is now the most common connection type for storage area network (SAN). Nowadays FC SAN is already an indispensable infrastructure component in any current complex IT environment. With the proliferation of various in-house developed and packaged applications supported by various implementations from different infrastructure components, managing modern IT environment is becoming more complicated because everybody is competing with resources and expecting the best service level with minimal unscheduled downtime. Thus, a best practice guide for FC SAN design and implementation is highly desirable.

This paper highlights key considerations when designing and operating a SAN environment including redundancy, scalability, fabric configuration, zoning, supportability, monitoring, switch configurations, and security. It is assumed that the reader has a basic understanding of FC networks. For in-depth technical discussions of SAN features, refer to the reference section at the end of this paper.

2 Redundancy

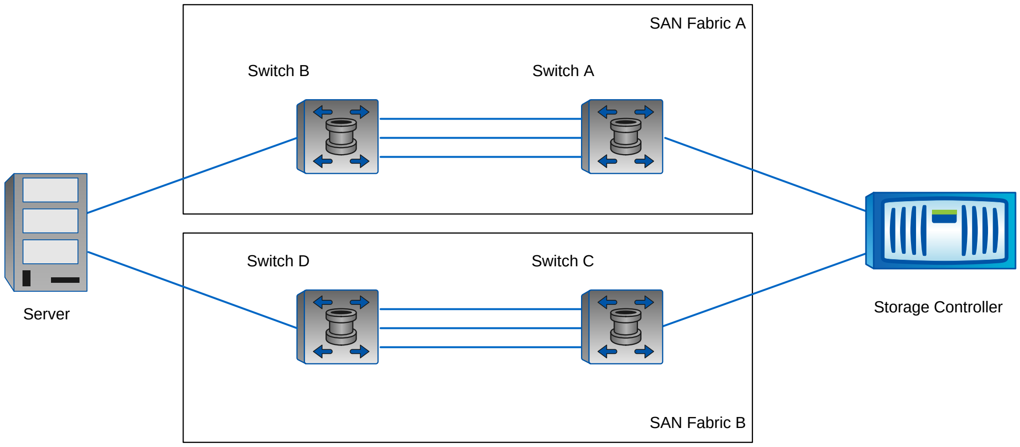
One of the most obvious methods to address high availability in a SAN is to make sure that redundancy is provided at critical points in the system.

2.1 Dual-Fabric Topology

Businesses require that their storage services be highly available because their most important applications cannot tolerate any downtime. For this reason, most (if not all) storage vendors recommend dual-fabric topologies for disk SANs.

A dual-fabric topology is effectively two separate and isolated SAN instances, which means that no single hardware or software failure can make a disk-based storage service unavailable. It also provides the ability to perform maintenance on an entire SAN fabric without affecting storage services. Figure 1 shows the total isolation provided by dual-fabric topology. For example, a failure of switch D does not affect the storage service because a path still exists through SAN fabric A.

Figure 1) Total isolation with dual-fabric topology.



Dedicated SANs that provide connectivity for tape-based backup services are not typically deployed using dual-fabric topologies because tape drives and robots often do not support multipathing. Single-fabric topologies are used instead, although many companies prefer to reuse the switches that host the disk-based storage services.

In a dual-fabric topology, each fabric should be cabled and configured close to identical. This includes the number and type of switches used, fabric configuration parameters, and where the servers and storage controllers connect into the SAN. A consistent configuration with a regular topology is easier to manage for SAN administrators and provides a more regular service to applications.

Dual-fabric topologies require twice the number of switches as a single-fabric topology, and you may be tempted to choose a single-fabric topology to save money. In this case, you must make sure that each device that is connected to the SAN uses multiple data links and that every effort is made to minimize the single points of failure. In addition, it is important that the SAN administrators negotiate with application administrators for agreed maintenance windows before the SAN is deployed. If the cost of a dual-fabric topology is prohibitive, rather than deploy a single-fabric topology it is recommended that you consider using iSCSI over an IP-SAN, or consider if file-based protocols are a practical alternative.

2.2 End-Device Connectivity

The document Fibre Channel and iSCSI Configuration Guide for the Data ONTAP 8.0 Release Family, which is found on the NetApp® Support (formerly NOW®) site, provides details about supported and recommended methods of connecting storage controllers to FC SANs.

This document includes a discussion about the preferred ports to use to maximize redundancy of hardware resources on the storage controller HBAs. NetApp recommends that servers be connected using utmost hardware redundancy by using HBAs on different PCI slots or on different PCI buses, if available.

This document also discusses various configuration maximums for each of NetApp's available storage controller models, including the maximum number of FC target ports per storage controller and the maximum fan-in ratio per target port. This might have a bearing on calculations that underpin decisions about network topologies.

Inheriting from the general rule that configuration should be identical across SAN fabrics, target ports should be split equally across both fabrics if a dual-fabric SAN is implemented.

Businesses with large nonproduction environments should commission specific target ports for this traffic. This minimizes the effect of non-production traffic on important production traffic. In addition, using port sets on the storage controllers helps to enforce traffic separation; port sets extend LUN masking configuration by fixing LUN masking configuration to a set of target ports.

Businesses that have a number of different server environments with different security classifications should commission multiple virtual SANs (vSANs) or virtual fabrics to separate traffic. Storage controller target ports must be dedicated to either classification.

Tape-based and disk-based storage services may use the same SAN switches, but they should be separated onto dedicated vSANs or virtual fabrics. Tape devices have a habit of being more temperamental than disk devices. Separating traffic prevents a misbehaving tape device from disrupting a disk-based storage service.

Finally, SAN administrators might consider using a dedicated vSAN or virtual fabric to connect storage controllers for storage virtualization (such as the NetApp V-Series) if the connection is made through SAN switches. Each vSAN or virtual fabric maintains its own zoning configuration and fabric services, making it harder for a SAN administrator to inadvertently make changes that might affect back-end storage visibility. This does not apply for configurations where the storage controllers are cabled directly to each other (that is, where switch ports are not used).

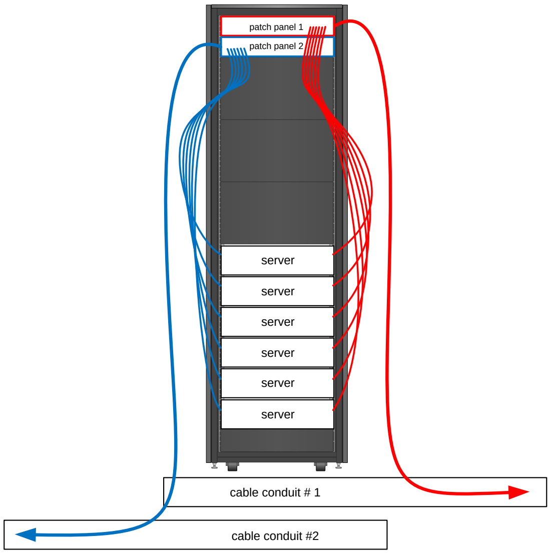
2.3 Physical Path Redundancy in Data Center

If SAN switching equipment that belongs to two separate fabrics is located in the same or adjacent racks, consider using cabling conventions that minimize the risk of a single point of failure.

- Use different fibre cables (bundles) for different SAN fabrics across the entire SAN. For example, a single 12-core fibre cable that carries traffic for both SAN fabrics is a single point of failure because an incident in the data center that affects that single cable disrupts the operation of both SAN fabrics.
- Dedicate patch panels for a particular SAN fabric.
- Use separate sides of the rack to run cables and patch leads for different SAN fabrics.
- In some large enterprise SANs, the core switches of both SAN fabrics are placed at opposite ends of the data center room, using completely separate cabling and conduits all the way to the racks that contain edges switches, initiators, and targets.

These conventions might seem extreme and should be balanced by practicality and cost; however, they have the potential to save unscheduled downtime. Figure 2 shows an example of physical path redundancy in a server rack.

Figure 2) Physical path redundancy in a server rack.



Racking layouts should be well planned. Networking racks can be the termination point for hundreds of cables; therefore, investing in wide racks (750mm) that are designed for large numbers of data cables shortens and simplifies cabling tasks.

Overcrowded racks make work time consuming and risky. It increases the likelihood that other cables are adversely affected while technicians are working in a rack. Optical fibre cables do not need to be broken or connectors dislodged for switches to experience issues; a technician handling an existing cable can temporarily affect the bit error rate (BER) simply by bending the cable.

More room in the rack makes it easier to implement good cable management practices. If technicians are not struggling for space and time, they are more enabled to address good cabling practices such as bend radii, slack considerations, and bundle weights.

2.4 Power Redundancy

Avoid the oversubscription of power. Some modular switches permit the oversubscription of power; that is to say that the total power draw of the switch under normal operating conditions is more than can be provided in the event of the failure of a single power supply. This can result in the switch powering down individual line cards until enough power is available.

Most SAN switches include at least two power supplies, which must be cabled to different power distribution units (PDUs) that are fed by different power sources. A poorly connected switch can be taken offline due the failure of a single PDU, and although a dual-fabric topology can absorb this type of failure without service interruption, it is best to avoid it if the additional power facilities are already available.

3 Scalability

Scalability is important to maintain a regular network shape that provides a consistent service to all of its clients. If it is not easy, quick, and inexpensive to extend the capacity of a SAN, then SAN administrators look for an alternative, one that might meet the immediate requirement of accommodating a new server but increases the operational complexity of the overall network.

SAN administrators must understand the maximum number of hosts, storage controllers, and tape drives that might need to be connected to the SAN before making any significant investment in switching hardware. SANs should be designed to accommodate a specific number of connections, and this number must align with the strategic IT infrastructure plan of the business.

However, it is often difficult to assess how many servers, storage controllers, and tape drives will be needed by the business in future years. Make an educated guess; select a design that is appropriate for this number, and develop a high-level contingency plan to describe how the design can be evolved to service a larger number of connections should the needs of the business change.

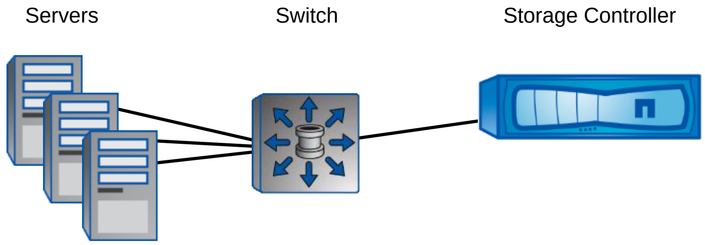
3.1 Network Shape

The biggest decision that affects scalability relates to network topology; it affects not only the overall port capacity of the SAN, but also other characteristics of the storage service. The more layers of switches, the greater the scalability; however, latency is increased. Whichever shape is used, it is important to use it consistently. Storage controllers and servers should plug into specific places in the network so that two servers do not experience different latencies through the network.

Single-Switch Topology

A single switch per fabric can be used if the port capacity ceiling is known and can be met by a fixed number of ports, and in this case a modular switch model (with line cards) is preferred. Small and medium-sized businesses often begin with a single switch as their data center is confined to a single room and their server numbers are static. Figure 3 shows a single-switch topology.

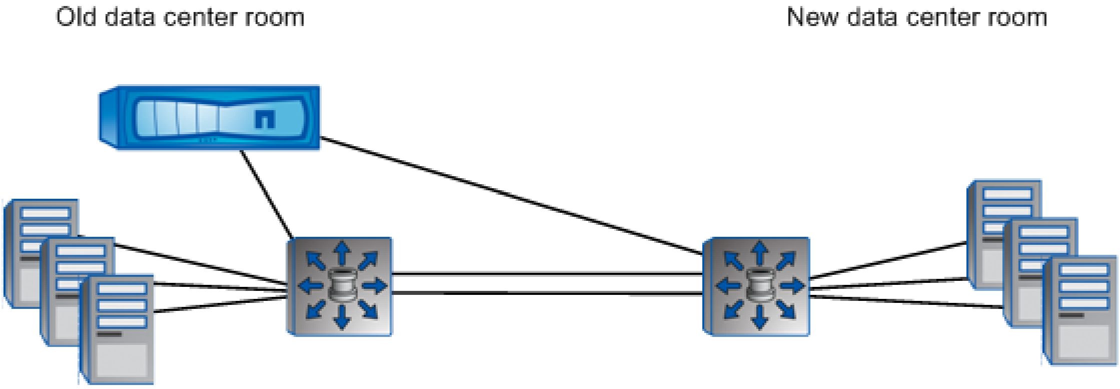
Figure 3) Single-switch topology.



A small to medium business will often rent their data center space, and in the event that the initial allotment of space is consumed, it is often not possible to have more space made available in an adjoining room; the data center service provider might only be able to make space available on another floor of the same building. Cabling all new servers back the SAN switches in the first data center room might not be feasible, even if there are empty ports available on the old SAN switches. In this case, it is common for a second switch per fabric to be purchased for the new data center room.

Figure 4 shows a common evolution of a single-switch topology.

Figure 4) Evolution of a single-switch topology.



The evolution shown in Figure 4 is convenient, but has its drawbacks:

- New storage arrays need to be connected to the SAN switches in each data center room, using twice as many ports on switches and storage controllers.
- Switch zoning needs to selectively connect initiator ports to particular target ports, something that is generally a part of larger SANs.
- There are now two modular switches per fabric, both with expensive supervisor and switching line cards.
- In an event where the second data center room is filled, another switch will be needed per fabric. In addition, an exponentially increased number of cables will be required to provide interconnects back to the existing infrastructure. This will become a full-mesh topology.

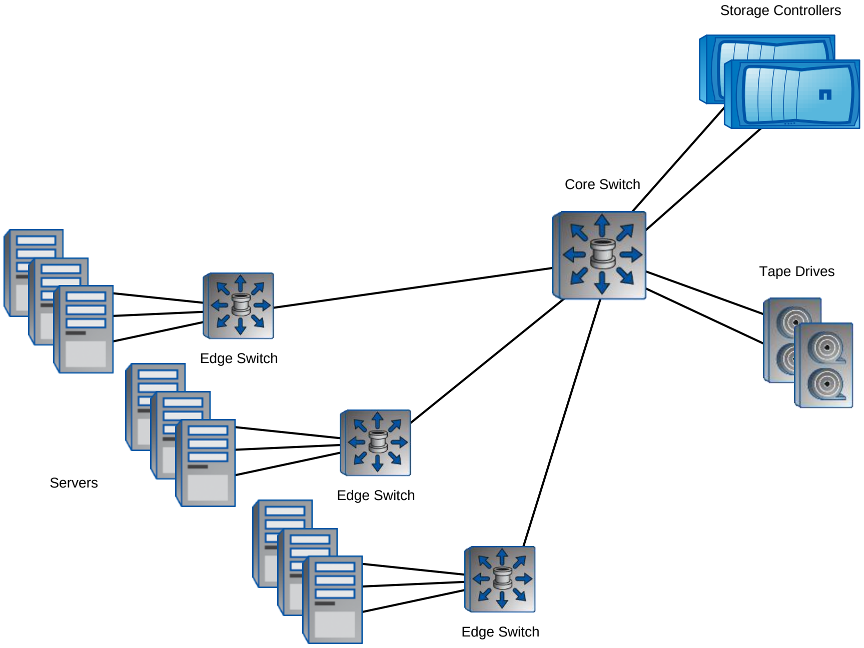
Core-Edge Topology

A core-edge topology is the simplest to manage (except for the single-switch topology) but has the added ability to scale predictably by adding edge switches. A single core switch provides connectivity to all target ports in a fabric, and another layer of switches called edge switches provides connectivity to servers. Most importantly, this consistent placement of servers and storage results in a consistent network service, though latencies will exceed those in a single-switch topology.

Edge switches might be less costly fixed-configuration switches and when deployed end-of-row can considerably reduce the expense and complexity of cabling to the servers. FCoE switches now being offered by network vendors provide the ability to connect servers with copper patch leads, making the cabling component of server provisioning even less expensive.

Figure 5 shows a core-edge topology.

Figure 5) Core-edge topology.



An extension to the core-edge topology is the edge-core-edge topology. Although scalable, a core-edge topology has finite port capacity, and in large environments it is possible that port capacity on the core switches could be exhausted. An edge-core-edge topology permits further scalability by pushing connections to storage controllers and tape devices onto dedicated edge switches, making all ports on the core switches ISLs to other edge switches.

Other Topologies

SAN design guides often discuss a number of other SAN fabric topologies such as ring topology, hub topology, and meshed topology.

A ring topology can permit unlimited scalability, although typical implementations provide a high degree of variance in characteristics of the service offered, specifically concerning latency.

The hub-and-spoke topology that is found in many network design guides is already covered by the core-edge and edge-core-edge topologies described in the preceding section. Network design guides cite the possibility of hub failure as the main drawback of hub-and-spoke, but this is overcome with dual fabrics.

Fully meshed topologies guarantee a single-hop storage service for all initiators, though it requires a larger number of interconnects to switches and storage controllers to provide service. The number of new interconnects increases with each new switch (that is, it is exponential). The preceding two-switch option described earlier in the subsection "Single-Switch Topology" is fully meshed (albeit a small mesh).

Partially meshed topologies overcome these limitations, but introduce degrees of complexity that are not present in the topologies discussed up until this point. Technically, any irregular topology shape could be considered a partial mesh, but the SAN administrators should strive to design a partial mesh, which provides storage services consistently irrespective of where the initiators and the targets are connected to the network.

A list of supported topologies can be found in the document Fibre Channel and iSCSI Configuration Guide for the Data ONTAP 8.0 Release Family.

3.2 Legacy Switches

Typically, a SAN switch is more expensive than an Ethernet switch with a similar port capacity and hardware architecture. For this reason, SAN administrators often must deal with legacy equipment more than their counterparts in the Ethernet network operations team, though as a consolation older SAN switches tend to be more useful than old Ethernet switches. Practical (thrifty) infrastructure managers tend to take an “if it ain’t broke, don’t fix it” approach to old SAN switches.

The downside is that a combination of old SAN switches and new SAN switches effectively results in two tiers of service when ideally a SAN administrator should be striving to offer a single classification of service.

In the end, it comes down to cost. The following questions should be asked:

- What does it cost to replace a batch of old 2Gbps SAN switches?
- What does it cost to recable the affected hosts to the newer 4Gbps or 8Gbps switches?
- Are these hosts being decommissioned soon?
- Can these hosts be housed in racks closer to newer SAN equipment?
- What features are missing on the old 2Gbps switches, for example, Secure Shell (SSH)?
- Are the old switches capable of running the same version of firmware as the new switches?
- If not, are the old and new versions completely compatible?

4 Fabric Configuration

SAN switches will function perfectly well out-of-the-box (that is with no additional configuration), though this is not recommended.

4.1 Hard-Coded Fabric Parameters

Switch vendors permit SAN operators to configure various fabric parameters, such as timer values and domain IDs. The SAN administrator should set some parameters, such as domain IDs, manually and leave the switch to automatically configure other parameters, such as timer values. These parameters should be consistent between the two fabrics.

Some UNIX® systems use the Fibre Channel IDs (FCIDs) assigned to them by the SAN fabric in the UNIX device filenames that represent different paths to the targets. If a switch does not use a persistent domain ID (configured manually), a switch reboot might result in a different domain ID being assigned to the switch, changing the UNIX device filenames, resulting in disk devices disappearing from UNIX hosts.

The SAN administrator should control the principal switch selection on the fabric manually through the use of switch priorities. Choose an enterprise-grade switch in a central position, such as a core switch, to operate as the principal switch.

Leave some parameters, such as fabric timer values, as default unless in the hands of a SAN expert. Only alter these values in a deliberate and well-thought-out way.

4.2 Avoid Intersite Fabrics for Large Deployments

Large SANs should avoid having fabrics that stretch across data center sites because a limited number of domain IDs can coexist in a SAN fabric. Figure 6 and Figure 7 show the differences between smaller sites and larger sites.

Figure 6) Intersite fabric design (smaller sites).

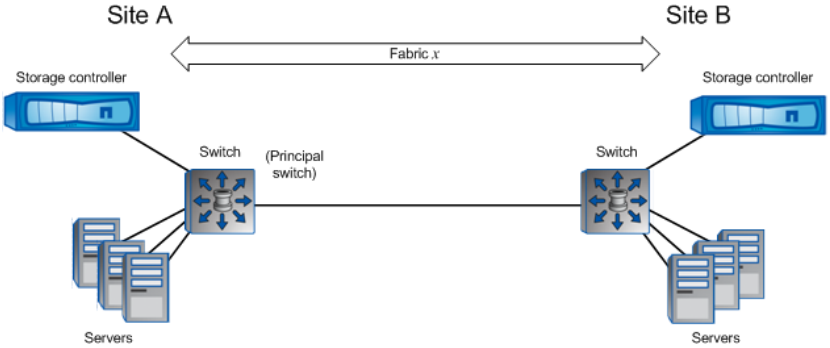
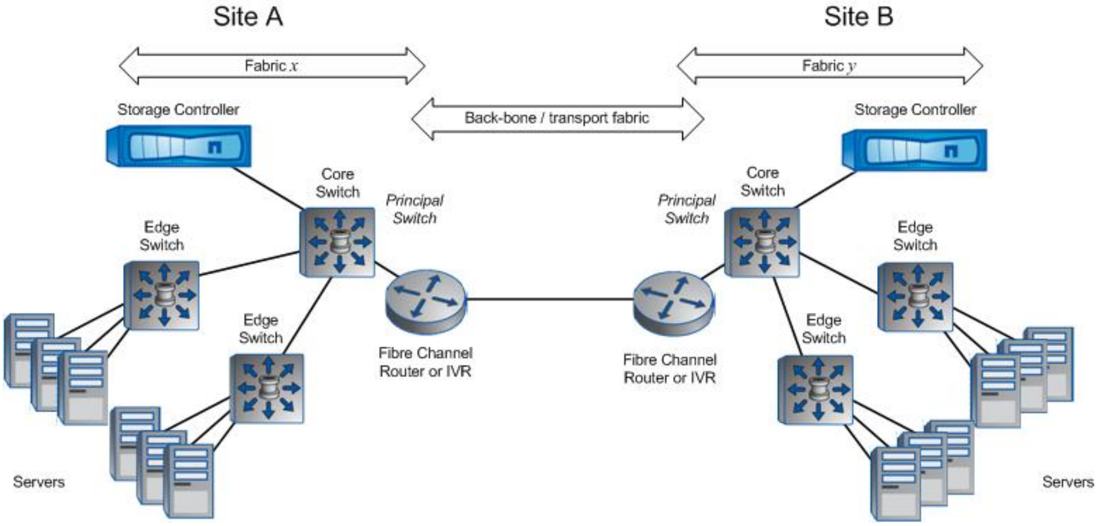


Figure 7) Intersite fabric design (larger sites).



Separating fabrics at different sites with FC routing decreases the likelihood of issues at one site (or on the intersite links) affecting the other site. It does; however, increase the complexity and might require additional licensing. Larger SANs are more able to justify this increase in complexity and cost.

Note: For either of the examples, there is a maximum supported hop count between initiator and target, which is dependent on switch vendor. For additional information, refer to Fibre Channel and iSCSI Configuration Guide for the Data ONTAP 8.0 Release Family.

4.3 Interswitch Links

When targets and initiators are located on different SAN switches, interswitch links (ISLs) become an obvious bottleneck in the network; a larger number of initiators are vying for capacity on the data links of the ISL, a term referred to as ‘‘contention’. However, the contention on ISLs does not necessarily infer that a problem exists. The SAN administrator must observe the actual utilization of the ISLs and take action if required.

If the SAN administrator observes that ISLs are being consistently saturated with traffic, following are the two options:

- Adjust multipath configuration. Multipathing configuration on the initiators can be tweaked so that the storage service uses an alternative path. This might relieve contention on ISLs by forcing the imitator to use the path through the other fabric.

- Add an additional data link to the ISL. Switches can aggregate multiple data links to form a single logical link, referred to as Cisco® port channeling or Brocade ISL trunking. A small amount of configuration is required on the switch to aggregate multiple links, and NetApp recommends configuring switches to use link aggregation even if only one physical link is associated with the logical aggregated link; this enables on-the-fly addition of physical inks to the logical link.

Saturation of ISLs should be avoided, and SAN administrators should be proactively monitoring the performance on ISLs in their SANs.

4.4 Flow Control

FC SANs implement flow control differently than Ethernet and IP networks. At the risk of oversimplification, it is moderately analogous to the traffic lights on freeway onramps; traffic is not permitted onto a data link unless the device at the other end first allows it. SAN ports are allocated buffer credits, which are replenished by the device at the other end of the data link, and each FC frame costs one buffer credit.

Flow control in SANs offers a number of tunable parameters, though they should not be altered unless the switch vendors advise doing so; there might be supportability implications if it is done incorrectly. Often, switch vendors recommend that intersite data links be configured to have a larger number of buffer credits to maximize the data link usage. Modern SAN operating systems abstract the actual buffer credit number and instead ask the operator to input the fibre path distance between the two data center sites in kilometers.

4.5 EndPoint Queue Depths

Storage vendors make specific recommendations for endpoint queue depth settings. It is important to follow these recommendations for the majority of the devices connected to the SAN fabric. Exceptions may be made for devices with particularly heavy I/O profiles, but these exceptions should be tracked.

The NetApp FC Host Utilities automatically adjust the host's queue depths to a default value recommended by NetApp. It can be modified to suit a particular workload and environment's requirements. NetApp requires installing the Host Utilities for all FC and iSCSI implementations.

4.6 Switch Hardware Architecture

Different switch vendors and switch models use different hardware architectures in their switches. Be aware of these and understand them at a high level, especially in regard to the maximum throughputs of individual ports and port groups.

Do not assume that every port on your switches can operate at a sustained line rate. Dedicate the ports that provide higher rates of sustained throughput to ISLs and storage controllers. Be aware that you might need to leave some ports unused to guarantee capacity to other ports with higher throughput requirements.

Consult the switch vendor literature and, if needed, seek training on specific switch models.

4.7 Heterogeneous Fabrics

As a general rule, NetApp does not support heterogeneous SAN fabrics. These are defined as SAN fabrics that include switches from more than one SAN vendor.

There are some exceptions, specifically in the area of embedded blade switches. In this instance, the switches might still need to operate in interoperability mode, and the SAN switch vendor's documentation should be consulted. The NetApp Interoperability Matrix Tool must also be consulted to make sure that the particular switch/blade-switch combination is supported.

5 Zoning

Zoning is a form of access control to storage services that is implemented in the storage network. It controls the initiator's access to targets.

The most important zoning recommendation is to configure zoning using consistent methods and parameters. In the day-to-day operation of a SAN, the zoning configuration is altered more than any other part of the configuration. Management complexity, especially in an operational team with several administrators, is significantly reduced with clear definitions of zoning configuration.

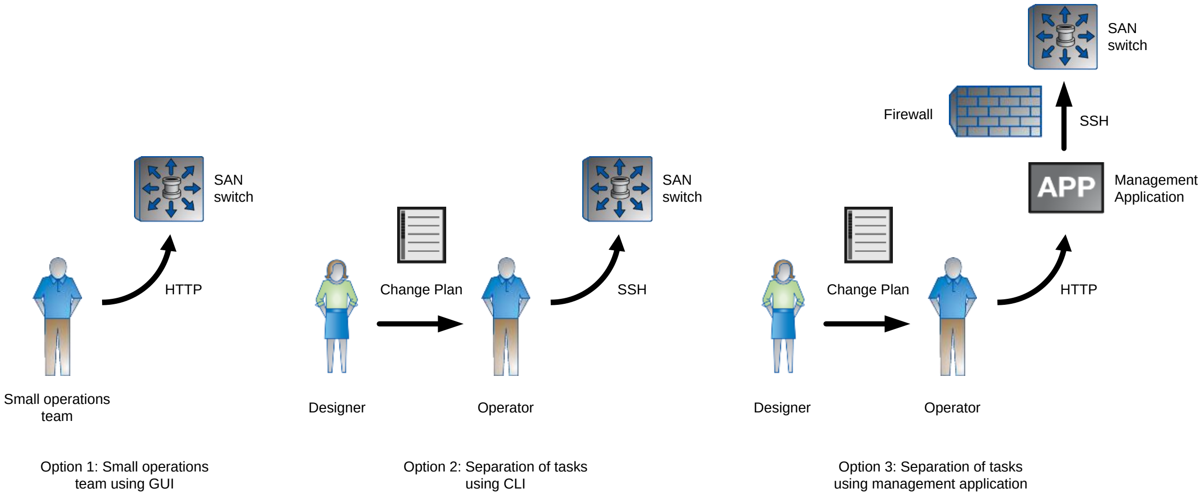
Observe the following practices:

- Single-initiator zoning is the most popular zoning method and is the one recommended by NetApp and many other SAN ecosystem vendors. This means that a single zone contains references to one initiator and one or more targets. It results in fewer overall zones than other methods and tends to match the SAN administrators' typical use cases.

  Zones should be configured in a way that minimizes the number of target ports that are visible to an initiator. The more target ports that are unnecessarily visible to initiator ports, the more frequently the target ports will be engaged as the initiator ports query them for LUNs. This can lead to longer path failover times and longer server reboot times.
- The main network vendors recommend that zoning should use worldwide port names (WWPNs) to identify initiators and targets, especially on the later models of SAN switches.
- This does not mean that the domain ID/port ID method of identification does not have its merits. Rather, the SAN vendors have determined that the WWPN method is better. Many SANs use the domain ID/port ID method to identify initiators and targets, and SAN administrators should consider updating their processes and procedures so that new zoning configuration uses WWPNs.
- Aliases should be used to identify WWPNs, as this means that the SAN administrator needs to type only the WWPN once, thereby reducing the likelihood of error.
- Network vendors recommend not mixing WWPN and domain ID/port ID information in a single zone.

Most SAN vendors offer two methods for making changes to zoning configuration, using either a GUI (logging on to the switch through HTTP or using a consolidated SAN management application such as Cisco Fabric Manager, recently converged into Cisco Data Center Network Manager) or the CLI of individual switches. Many SAN administrators use the GUI without considering the benefits that the CLI offers.

Figure 8) Making zoning changes: GUI versus CLI.



Smaller organizations, where change planning and implementation are performed within the same team or by the same person, might find using a GUI simpler. Making zoning changes using the point-and-click method on a GUI is conceptually much simpler for many SAN administrators.

Larger organizations, where the change planning and implementation tasks are separated, with the implementation possibly outsourced to a third-party provider, should seriously consider the CLI method. Changes using the CLI can be planned prior to implementation, then reviewed, and finally implemented by an operator copying and pasting individual commands into the terminal window. In this instance, the operator, who might be performing the change activity after hours, has a much simpler task.

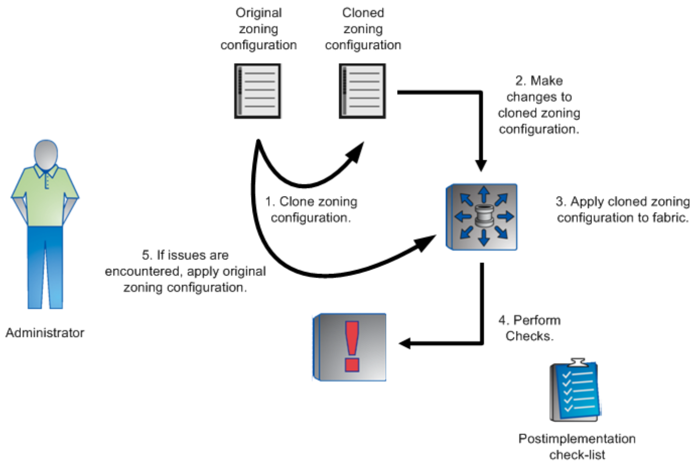
Some SAN management applications permit sending CLI commands through the GUI of the management application. This permits proxying of access to the management console of SAN switches. This is a more secure option that often suits third-party outsourcing situations without negating the benefits of using the CLI.

5.1 Zonesets or Zone Configurations

Using zonesets (Cisco) or zone configurations (Brocade) is a quick way to roll back zoning changes. Complex zoning changes on large zoning configurations can be rolled back by using a small number of commands without having to manipulate individual zones.

NetApp recommends such mechanisms, even for experienced operators. Figure 9 shows the process for making complex zoning changes with minimal commands.

Figure 9) Making zoning changes using cloned configurations.


Original zoning configuration
Cloned zoning configuration
2. Make changes to cloned zoning configuration.
1. Clone zoning configuration.
3. Apply cloned zoning configuration to fabric.
5. If issues are encountered, apply original zoning configuration.
Administrator
4. Perform Checks.
Postimplementation check-list

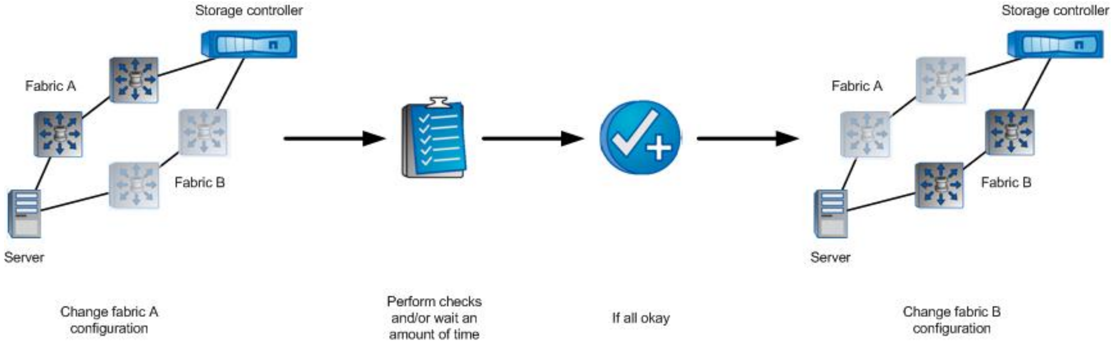
5.2 Changes to Both Fabrics

Even if a dual-fabric SAN has been deployed, a business can still suffer unscheduled downtime due to operator error. A well-operated SAN must separate implementation tasks to run on different SAN fabrics at different times.

After operators make changes to one SAN fabric, a certain amount of testing must take place to make sure that the changes have had the intended effects, and have not had have any unintended effects.

Testing for unintended effects can be quite laborious. In all but the smallest SANs, logging in to every host and checking that its disks are all still online is not feasible. Waiting for a period of time between fabric changes is useful in that it allows for unintended consequences to come to light. If an unintended consequence is revealed, the same change to the second SAN can be postponed. Figure 10 shows the steps for making changes to both fabrics.

Figure 10) Making a change to both fabrics.


Storage controller
Fabric A
Fabric B
Server
Change fabric A configuration
Perform checks and/or wait an amount of time
If all okay
Storage controller
Fabric A
Fabric B
Server
Change fabric B configuration

In determining how long to wait between implementing changes on two SAN fabrics, SAN administrators might ask themselves the following questions:

- Do we perform this type of change infrequently?
- Has this type of change had any unintended consequences in the past?
- Is it not feasible to check a statistical sample of services to make sure that they are still working?

The more "yes" or "do not know" answers to these questions, the longer the SAN administrator should wait before performing the same change procedure on the second SAN fabric.

6 Supportability

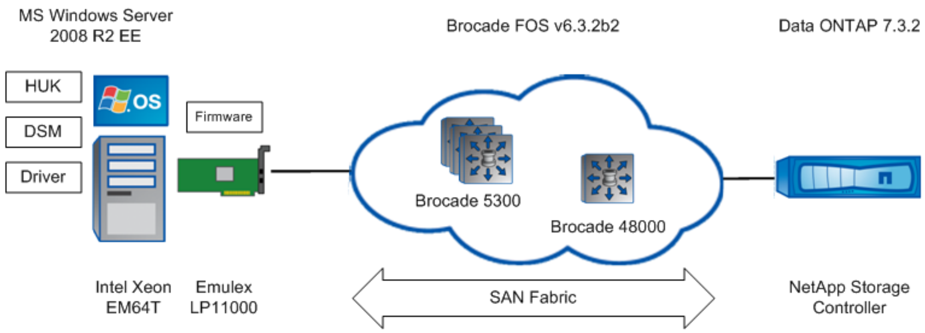
In FC SANs, supportability is of prime importance. SANs are different compared to Ethernet and IP networks, where there is an implicit notion of support, and it is this implicit supportability that is partly responsible for Ethernet's and IP's popularity over FC.

A supported configuration is the one in which the vendors of any component of your SAN, be it the switches, the HBA, or the storage controllers, will assist you if you contact them. Switch vendors, storage vendors, HBA vendors, operating system vendors, and application vendors only support very specific end-to-end configurations, and the onus is on the SAN administrator to make sure that what has been deployed in the SAN is referenced in each vendor's support matrix. If not, the applicable vendor will help if a support case is raised, although their first recommendation might be to change to a configuration that they support.

The support matrix of each applicable vendor must be consulted prior to design. NetApp's support matrix is called the Interoperability Matrix Tool.

It is imperative that supportability be addressed before deployment of a SAN because it is much more expensive to address supportability after implementation. If you have inherited responsibility for an existing SAN, NetApp strongly recommends that you assess the supportability of your SAN as a matter of priority.

Figure 11) Assessing supportability.



In Figure 11, the SAN administrator has a SAN fabric that is hosted on a mix of Brocade 48000 and Brocade 5300 switches that is running Brocade Fabric Operating System (FOS) version 6.3.2b2. The storage administrator wants to know if a new storage controller running NetApp Data ONTAP® 7.3.5 is a supported configuration. After using the NetApp Interoperability Matrix, it can be found that this configuration is indeed supported by NetApp (configuration number 20110304-094931971, status: Supported).

Next, a server administrator proposes connecting a new 64-bit Intel® server to the SAN, with an Emulex LP11000 HBA and Microsoft® Windows Server® 2008 R2 Enterprise Edition. The NetApp Interoperability Matrix confirms 21 configurations that are supported by NetApp containing these parameters. The server administrator can use any one of these supported configurations to determine which version of HBA firmware, driver, Host Utilities Kit, and so on to install.

To make sure that the other hardware vendors (the server vendor, Microsoft, and Emulex in this example) support the configuration, the server and storage administrators must consult those vendors' support matrixes.

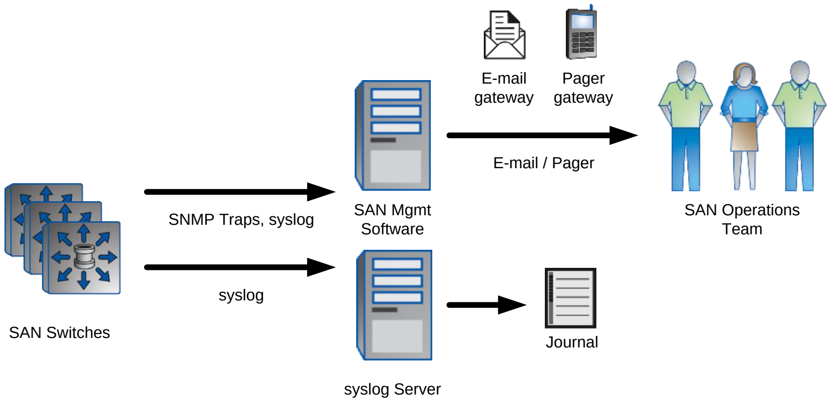
7 Monitoring Tools

A number of monitoring tools are available for SAN fabrics. This section describes the functionality that is needed to operate a highly available SAN.

7.1 SNMP Monitoring

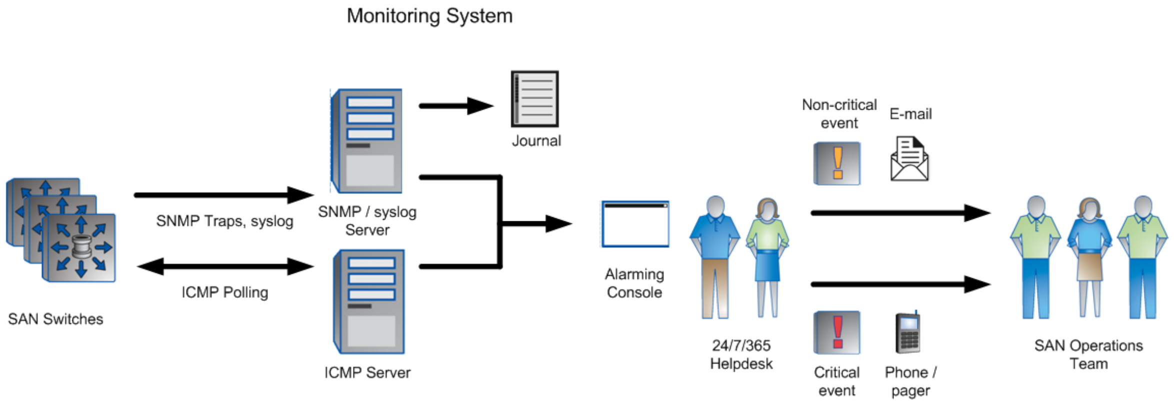
All modern SAN switches provide Simple Network Management Protocol (SNMP) daemons that send traps and respond to queries, as well as provide the ability to forward syslog messages to a remote server. These features must be implemented to provide a high-availability service; it is impossible to do so without them. It might be fine for SAN administrators to perform manual switch monitoring by logging on to each fabric's principal switch Monday through Friday, but what happens over a weekend or over a vacation period?

Figure 12) Direct SAN switch monitoring.



SAN vendor management applications, such as Cisco Fabric Manager (recently converged into Cisco Data Center Network Manager) or Brocade Data Centre Fabric Manager (recently converged into Brocade Network Advisor), can be deployed to monitor SAN switches and automatically forward notifications to SAN administrators through e-mail or pager gateways. SAN vendor management applications have the advantage of being able to understand all messages being sent by the SAN switches, but they typically cannot monitor other types of infrastructure.

Figure 13) SAN switch monitoring using a centralized monitoring system.



Larger or mission-critical SANs might be required to integrate into specialized monitoring applications such as HP OpenView or IBM's Tivoli suite. These larger monitoring applications have the advantage of being able to monitor all types of infrastructure with a single application and are more likely to be monitored 24x7x365 by human operators.

Event correlation can be performed with a consolidated monitoring application. An air conditioner failure in the data center is easier to diagnose with a consolidated monitoring application; if nearly every piece of infrastructure sends an SNMP trap reporting high temperature, human operators are able to use their judgment and call the facilities technicians rather than automatically alerting a list of infrastructure administrators.

It is important to include Internet Control Message Protocol (ICMP) polling of switches. A troubled switch might not have enough time to send an SNMP trap before it goes offline, essential in larger and mission-critical SANs.

7.2 Syslog

Syslog can be used for monitoring, but it is also an important diagnostic tool. Analyzing syslog messages is important for picking up errors and issues in a SAN fabric before they become incidents and for assessing the root causes of issues after they have become incidents. It is important that syslog servers journal their raw input so that it can be used for this diagnostic purpose; some SAN vendor management applications do not permit administrators to access this information in its raw form and provide information only under error conditions.

After going to the trouble of centralizing syslog messages, they should be used. Administrators should periodically sweep the syslog repository and assess what is happening in the SAN. Every message means something; it is the operator’s job to determine whether or not the message is important. If the messages are slightly cryptic, check the switch vendor’s support sites to understand it. Perform a Google search or open a support case with the vendor if required.

Record the message and understand whether it requires human intervention. Share this information with the other operators on your team.

7.3 CLI Audit Logs and Role-Based Access Control

In SANs that are managed by a large operational staff or that are managed by a third party, CLI audit logs should be configured to be sent by the switch to a centralized location. Role-based access control (RBAC) must be implemented on the SAN, and only a small number of trusted individuals should have knowledge of the root passwords. In addition, access to the centralized location must be limited and monitored so that records cannot be altered or destroyed.

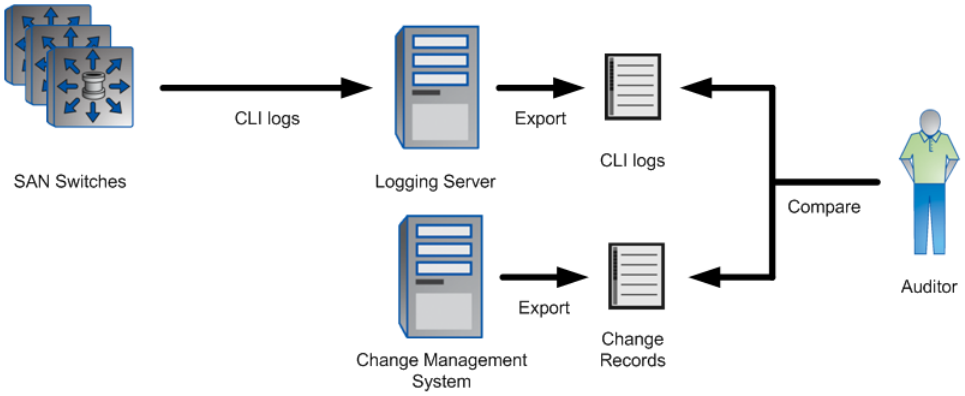
This might seem slightly extreme for many environments unless specifically required for legal compliance. However, consider that any operator who works on a team with three or more staff members has plausible deniability of their actions unless RBAC is implemented. Effective root cause analysis is a prerequisite for any improvement attempt, and unless unskilled operators can be identified, root cause analysis is extremely difficult if operator error is suspected.

Here is one example that explains how RBAC and CLI auditing has improved operator behavior. An operational team in a business was plagued by a severity one incident rate of approximately six per month. Operational staff would work late into the night, unpaid, to repair issues and consequently had little time to allocate to project work during the day. The project managers were under extreme work pressure, and the operational staff were exhausted, which led to a nonconducive working environment. To deal with the situation, the team leader implemented RBAC and configured CLI audit logs to be sent to a central location. Overnight, the severity one incident rate plummeted effectively to zero when the operational staff became aware that their actions would be visible to their peers.

Of course, it is important to stress to operational staff that implementing these measures is not an attempt to blame particular individuals when things go wrong; rather, it is so that the team might perform root cause analysis effectively to make sure the problems do not recur.

If implemented, it is important to check CLI audit logs periodically, especially if some or all operational tasks are outsourced to a third party. CLI audit logs are typically used in one of two ways. Either they are untouched until required for root cause analysis or they are periodically audited, which is more common in organizations that have legal compliance requirements.

Figure 14) CLI audit log review.



Every organization should have a policy that determines when and how changes to infrastructure configuration must be made. For example, an organization might have a policy stating that all configuration changes on the SAN must be accompanied by a record in the change management system. An internal auditor would periodically trawl the CLI audit logs and check that each line item is referenced by a corresponding change management record or that a statistical sample of the line items is referenced by a corresponding record.

If the operational team has a large staff, management should periodically commission this type of activity. It need not be exhaustive; an outline is enough to explain whether or not the change management policies of the organization are being followed.

7.4 Performance Monitoring

Gathering performance data for SAN switches is essential for optimal operations of SAN fabrics. SAN administrators need to be able to view utilization of network capacity at the various critical points in their network to address potential performance bottlenecks before they occur.

Detailed performance data is also important to answer queries from other administrators. IT performance issues tend to roll downhill, often ending up in the hands of network administrators before server and application administrators have genuinely attempted their own investigations. SAN administrators often field performance questions from other administrators, and it is usually up to the SAN administrators to remove the burden of proof from themselves (that is, to demonstrate that the source of the problems is not in the SAN).

The NetApp OnCommand™ Insight suite, using the performance license, gathers detailed performance data for switch ports. It also captures detailed performance data for virtual machines, as well as performance data for storage arrays down to the volume level and LUN level. It automatically associates host names with volume and LUN performance data, providing an exact end-to-end view of the performance of storage services. Importantly, it is an agentless product that can be deployed quickly and easily to existing environments.

8 Switch Configurations

8.1 Back up Configurations Periodically

Switch configuration should be backed up both periodically and before every configuration change to the switch. Backing up the configuration periodically permits the switch to be rebuilt if a hardware failure occurs.

Backing up the configuration before change activities makes sure that the operator can roll back the change activity, if required. SAN administrators typically provide two roll-back procedures, a soft roll-back procedure that reverses the changes that were made as part of the change or a hard roll-back procedure where the entire backed-up configuration is applied to the switch and the switch is rebooted.

Configuration does not only mean software configuration. It also means the cabling records and the versions of software binaries running on the switch. Perform backups of these as well, making sure that all of the relevant software binary files are backed up. For example, Cisco MDS switches require both a kickstart file and the main software image file.

8.2 Standardize Configuration

In many SANs, each switch has a slightly different configuration, though consider the following questions:

- Why should the configuration of the core switch in fabric A be different than the configuration of the core switch in fabric B?
- Why should the configuration of two edge switches in fabric B look completely different? Should one edge switch use SNMP agent version 2c and the other use version 3? Should the uplink on one edge switch use port 1 and the uplink on the other use port 24?

As a general rule, switch configurations should be identical for switches that have the same function, excepting comment fields, description fields, or fabric parameters such as domain IDs.

For example, you should have two basic configurations for a SAN with a core-edge topology: one for a core switch and one for an edge switch. The configuration of every switch in the SAN should be identical to either one of these two basic configurations (except for switch name, interface descriptions, and domain IDs).

Standardizing your configurations permits you to save time recalling the difference between each of your different SAN switches and more time focusing on more strategic issues. With more time to address strategic issues, you will save money and time operating your SAN.

NetApp suggests you make your SAN switch configurations look as similar as practical to the configuration of your Ethernet switches. Obviously, Ethernet and FC are two very different protocols, but at a fundamental level, they are the same; they both switch packets between different ports using a particular forwarding logic. Naming conventions and other nonfunctional configuration, such as SNMP agent versions, syslog configuration, and Authentication, Authorization, and Accounting (AAA) configuration, can be made similar. The more similar the configurations, the more likely it is that the network administrators and SAN administrators are able to share workload.

8.3 Bare-Metal Restore Guides

NetApp strongly recommends that you create an operational procedure to formalize the instructions explaining how to build or rebuild a switch.

This decreases the amount of time that it takes to rebuild a switch in the event of a hardware failure. Although a single-switch failure should not affect the availability of the storage services when using a dual-fabric topology, until the failed switch is rebuilt, the storage services are at risk of interruption if another failure were to occur in the SAN.

If you document the procedure to rebuild the switch, it can be performed by any available member of the operational team. This leads to a shorter time that the SAN is exposed to the risk of another failure.

Unless hot spare switches are kept at the data center site, this procedure must reference the vendor’s return merchandise authorization (RMA) process. Therefore, it is important that your customer information is kept up to date in the vendor’s systems and that all of the information required for the RMA is readily available. When writing this procedure, proactively contact your switch vendor to confirm what information is required.

8.4 Configuration Auditing

Without access to every server, SAN administrators typically assume that their configuration, especially their zoning configuration, is correct unless demonstrated otherwise. Therefore, they have learned to double-check and triple-check their configuration before it is implemented. However, issues with SAN configuration are often only discovered in the event of a failure.

Small businesses do not encounter this problem. If the overall number of hosts in the data center is small, the SAN administrator, who more often than not is also the server administrator, can check every host after any zoning changes are made. However, this is infeasible for a large business.

The NetApp OnCommand Insight suite, using the Assure license, parses zoning configuration gathered from SAN switches and LUN masking configuration from storage controllers to proactively identify potential issues. It can apply user-defined policies to SAN configuration, checking (for example) that nonproduction servers can see at least two paths to storage controllers and that production servers can see at least four paths, and proactively alerting if any of these policies are breached.

8.5 Network Diagrams

SAN administrators should consider maintaining network diagrams. As previously mentioned, simpler network topologies are preferred, and if this advice is followed, it is feasible that SAN administrators might not need to refer to diagrams to help them conceptualize their networks. However, there are situations where diagrams can be especially useful:

- Large network topologies, specifically partially meshed topologies, are difficult to conceptualize as not all ISLs have the same characteristics. The sheer number of switches in large multiroomed data centers can also make it difficult to track devices.
- Where a large amount of operational staff turnover is experienced. This might be the case if SAN administration is outsourced to a third-party provider. Diagrams (even for a simple topology) often help to bring new staff up to speed with the network configuration.

9 Security

Security is an area that polarizes many SAN administrators and confuses the others. Much time and money can be spent implementing security mechanisms that often have no immediately obvious benefit. This section describes a number of general best practices that apply to nearly all environments. Some environments, however, might require additional security measures. Deploy these measures for a specific reason; otherwise, it will fail to provide the value worth the investment.

You should deploy security measures in your SAN to address specific security requirements. These security requirements are written to address specific security risks. Working internally or through external consultants, ascertain which security risks apply to your organization and how to address those risks. At the very least, you can commission a series of risk assessment workshops using the existing infrastructure administrators of your organization. It is important that you present the raw output from these workshops to management. Even if your managers are not technically inclined, they are the people who are accountable for both risk and expenditure. Management should ultimately decide which risks are acceptable and which are not.

The security requirements of the applications that are hosted on your infrastructure might dictate security requirements for your infrastructure. In this instance, you might be asked to design and deploy your SAN in accordance with well-known security standards.

FC traffic sometimes traverses IP networks, as in the case of Fibre Channel Internet Protocol (FCIP), generally to provide intersite connectivity. If the IP network includes segments that are not under the administrative control of the organization, the FC switches at either end of the FCIP tunnel should at least require each other to authentication themselves.

If licensing permits, use dedicated routing table instances and virtual LAN (VLAN) tags for the FCIP traffic on the IP network. This keeps the forwarding configuration extremely simple and minimizes the likelihood that traffic is forwarded along unintended paths.

As with all FC network design, minimizing latency is of paramount importance. Use the minimum number of IP switches in your path between the SAN instances and make sure that these devices are data center grade, multilayer switches or routers.

Control console port access to your SAN switches. Make sure that authentication applies to the console port and, if hosting your equipment in external data centers, make sure that your racking infrastructure is lockable.

As previously discussed, RBAC and CLI auditing should be used, but there are other recommendations regarding AAA configuration. Use a centralized user database for user authentication and try to avoid locally administered accounts. Administering user accounts consumes precious time that should be used to administer infrastructure, not user account information. Centralizing user authentication using, for example, Active Directory®, means that managing the user accounts becomes the responsibility of the HR team and not the SAN operations team.

Most companies configure two basic roles on their SAN switches, a read-write role with complete administrative privileges and a basic read-only role.

Additional roles are generally required only by larger SANs. In this instance, clearly define the specific use cases for each group of users and be sure that the role configuration in the SAN matches these use cases. Overzealous authorization configuration makes things difficult for your SAN administrators; one command might be permitted, but another seemingly similar command might not be permitted. Infrastructure administrators either are in your organization’s “circle of trust” or are not. If you permit them with administrative privileges and they abuse it, there is always CLI auditing to keep them honest.

Consider centralizing any authorization configuration on a centralized AAA server. This simplifies your switch configurations.

As with any network device, nonessential services and processes in the SAN operating systems should be disabled. Telnet should be disabled in favor of SSH. HTTP daemons should be disabled in favor of HTTPS daemons.

References

- Fibre Channel and iSCSI Configuration Guide for the Data ONTAP 8.0 Release Family
  https://library.netapp.com/ecm/ecm_get_file/ECMM1280845.
- Interoperability Matrix Tool
  http://now.netapp.com/matrix/mtx/login.do
- Principles of SAN Design, Second Edition; Judd, Josh; 2007; Infinity Publishing
- Storage Area Network Fundamentals; Gupta, Meeta; 2002; Cisco Press

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

NetApp provides no representations or warranties regarding the accuracy, reliability, or serviceability of any information or recommendations provided in this publication, or with respect to any results that may be obtained by the use of the information or observance of any recommendations provided herein. The information in this document is distributed AS IS, and the use of this information or the implementation of any recommendations or techniques herein is a customer’s responsibility and depends on the customer’s ability to evaluate and integrate them into the customer’s operational environment. This document and the information contained herein may be used solely in connection with the NetApp products discussed in this document.

Go further, faster®

NetApp®

www.netapp.com