WUTC DOCKET: 181051
EXHIBIT: RA-3
ADMIT ☑ W/D □ REJECT □

Exh. RA-3
Docket UT-181051
Witness: Dr. Robert Akl

BEFORE THE WASHINGTON
UTILITIES AND TRANSPORTATION COMMISSION

WASHINGTON UTILITIES AND TRANSPORTATION COMMISSION, **Complainant,** **v.** **CENTURYLINK COMMUNICATIONS, LLC.,** **Respondent.**	**DOCKET UT-181051**

EXHIBIT TO TESTIMONY OF

DR. ROBERT AKL

ON BEHALF OF STAFF OF
WASHINGTON UTILITIES AND
TRANSPORTATION COMMISSION

Broadcom Storage Area Networking Design and Best Practices Design Guide

August 31, 2022

Exh. RA-3
Docket UT-181051
Page 1 of 112

BROCADE
A Broadcom Company

SAN Design and Best Practices
Design Guide

Exh. RA-3
Docket UT-181051
Page 2 of 112

Broadcom reserves the right to make changes without further notice to any products or data herein to improve reliability, function, or design. Information furnished by Broadcom is believed to be accurate and reliable. However, Broadcom does not assume any liability arising out of the application or use of this information, nor the application or use of any product or circuit described herein, neither does it convey any license under its patent rights nor the rights of others.

The product described by this document may contain open source software covered by the GNU General Public License or other open source license agreements. To find out which open source software is included in Brocade products, to view the licensing terms applicable to the open source software, and to obtain a copy of the programming source code, please download the open source disclosure documents in the Broadcom Customer Support Portal (CSP). If you do not have a CSP account or are unable to log in, please contact your support provider for this information.

Table of Contents

Chapter 1: Preface

1.1 Introduction

This document is a high-level storage area networking (SAN) design and best-practices guide based on Broadcom products and features, focusing on Fibre Channel SAN design. The storage landscape continues to modernize, and multiple choices must be made to design the right Fibre Channel architecture. Covered topics include the early planning phase, understanding possible operational challenges, and monitoring and improving an existing SAN infrastructure.

The guidelines in this document do not apply to every environment, but they will help guide you through decisions for a successful SAN design. Consult your Broadcom representative or refer to the documents in Appendix D for details about the hardware and software products.

NOTE: This is a “living” document that is continuously being expanded, so be sure to frequently check Broadcom.com for the latest updates to this and other best-practice documents.

1.2 Audience and Scope

This guide is for technical IT architects who are directly or indirectly responsible for SAN design based on Brocade® Fibre Channel SAN platforms. It describes many of the challenges that face SAN designers today in both greenfield and legacy storage environments. While not intended as a definitive design document, this guide introduces concepts and guidelines to help you avoid potential issues that can result from poor design practices. This document describes best-practice guidelines in the following areas:

- Modernizing the storage landscape
- Architecting a SAN
- SAN topologies
- Data flows
- Traffic Optimizer
- FPIN
- Predeployment infrastructure testing
- Device connections
- Scalability and performance
- Supportability
- Monitoring
- Troubleshooting
- FC Routing
- FCIP
- IP Extension
- Intelligent services
- NPIV
- Access Gateway
- Workloads

- SAN management
- Security
- Automation

NOTE: A solid understanding of SAN concepts and Brocade Fibre Channel technology is assumed. Please see Appendix D for recommended additional publications.

1.3 Approach

Although some advanced features and specialized SAN applications are discussed, these topics are covered in greater detail in separate documents. The primary objective of this guide is to provide a solid foundation to facilitate successful SAN designs—designs that effectively meet current and future requirements. This document addresses basic administration and maintenance, including capabilities to identify early warning signs for end-device (initiator or target) latency, which can cause congestion in the SAN fabric. However, you should consult product documentation and documents in Appendix D for more details. Comprehensive discussions of SAN fabric administration, storage network cabling, and Fibre Channel security best practices are covered in separate documents.

1.4 Overview

Although Brocade SAN fabrics are plug-and-play and can function properly even if left in a default state, Fibre Channel networks clearly benefit from a well-thought-out design and deployment strategy. In order to provide reliable and efficient delivery of data, your SAN topology should follow best-practice guidelines based on SAN industry standards and considerations specific to Broadcom.

This document does not consider physical environment factors such as power, cooling, and rack layout. Rather, the focus is on network connectivity edge devices to the fabric and any inter-switch links (ISLs) and software configurations.

NOTE: The scope of this document is switch-centric and does not discuss HBA, storage, end-device setup, configuration, and maintenance. Some fabric monitoring, management, diagnostics, cabling, and migrations are covered, but if you want full details, please refer to other appropriate documents.

Chapter 2: Storage Landscape

2.1 The Storage Landscape

In the IT infrastructure world, storage is a critical element. It is where the data resides. It is where secure copies of the data exist. It is the foundation of the overall performance of the IT application base. After all, no matter how many CPU cores and how much memory a server may have, every server waits for data at the same speed as any other machine.

Consequently, the scope of this environment includes early disk drives, tape for securely and cost-effectively backing up the data and software implementations that provided access, performance, and security. Furthermore, the responsibilities of the storage admin include the need to have secure copies of the data, whether through RAID configurations or site-to-site replication solutions. The mantra of storage admins is “a single copy of any data set is a single point of failure waiting for a disaster to occur.” Data loss is never an acceptable option from the application point of view.

2.2 Tipping Point—The All-Flash Data Center

Over the years, the state of the storage environments in IT has changed dramatically. A quick review of the changes takes you from the early tape systems through the evolution of the hard disk drive (HDD) into the development of RAID systems and enterprise arrays. One of the things that had traditionally been true was that storage, based on HDD building blocks, did not evolve rapidly. Changing from 5400 RPM disk drives to 7200 RPM disk drives as a performance element (more data under the head per second) took some 10 years from its inception to fully populate data centers. Other developments included the density of the magnetic signature on the drive platter and the number of platters and heads per drive. This retrospective is only helpful by denoting that the storage environment in IT did not progress as rapidly as, say, CPU development or memory performance and capacity. Those developments were in silicon, and drive development was mechanical. As a result, Moore’s Law applied to CPU and memory, but not to storage. The nature of storage has begun to change with the advent of solid-state drives (SSDs).

Initially, the progress was moderate. In a brilliant market-enabling move, the drive vendors made the SSD platform the same shape/canister as existing HDDs with the same SCSI, SAS, and SATA connectors, which meant “plug compatibility” on the back end of the array for the new technology. However, early on, the enterprise array controllers and, in the case of embedded disk drives in servers, the OS driver stack did not take advantage of the change in performance and other drive characteristics. The OS stack not taking advantage of the new technology was part of why the early “hybrid arrays,” which used a mix of traditional HDD and SSD drives on the back end, were less performant than many storage administrators had expected. Not meeting expectations impacted the adoption rate, and the IT organization did not experience as much of an issue with existing storage networks as expected. For over 40 years in IT, we have been moving bottlenecks in CPU performance, memory speed/scale, storage capacity/performance, and network speeds. Removing one bottleneck simply lets you find the next one, not unlike how widening one segment of a major highway pushes the traffic backup to the next narrow section of the highway. The result was that many customers felt that the performance gain versus the technology cost was only fit for their very highest demand applications.

With the advent of the all-flash array environment and inline features that came along, such as compression, encryption, and deduplication, the cost per terabyte became more enticing. Additionally, the newer enterprise array controllers were designed for all-flash performance characteristics and gained significant increases in both IO per second (IOPS) and latency. Added to that was the storage density, which allowed them to collapse multiple racks of HDD platforms into a partial rack of SSD with the benefits of power and cooling reduction. What was the consequence? There is a much more rapid adoption rate of all-flash arrays, which are currently more than 70% of the shipping environment.

That technology shift leads to changing demands on the design of storage area networks, which is true regardless of the technology used. If the expectation is to utilize the capacity and performance of these platforms, then serious consideration has to be given to the design.

A dedicated storage network infrastructure that provides lossless, low-latency, deterministic, scalable, and performant storage services to the applications becomes critical. Storage admins will talk about “fan-in” or “fan-out” ratios for storage platforms, which are the number of servers/applications in the network using a particular array or array port for their storage access.

Depending upon the types of applications and their performance needs, that ratio may range anywhere from low single digits to 40 to 50 servers (hundreds of virtual machines). As with any provisioning scenario, the storage admin is dealing with projections of how much capacity and performance any server/application will use. But application performance is variable; time of day, week, month, and seasonal or event-driven events cause spikes or drops in application demand.

Another consideration here is that the entire application base does not refresh simultaneously. And in fact, the more common scenario is that multiple generations of performance will exist in the environment simultaneously. The availability of service windows drives this to re-platform existing environments to new servers and storage. Some legacy applications may not have an environment that runs a current operating system or application version. One may have 10-year-old or older operating systems connected with a host bus adapter (HBA) with two or more older generations of technology behind it. How then do you balance that still critical application against the needs and performance of the newer machines?

The answer to this is a combination of topology, balanced provisioning, granular monitoring, and automated mitigation.

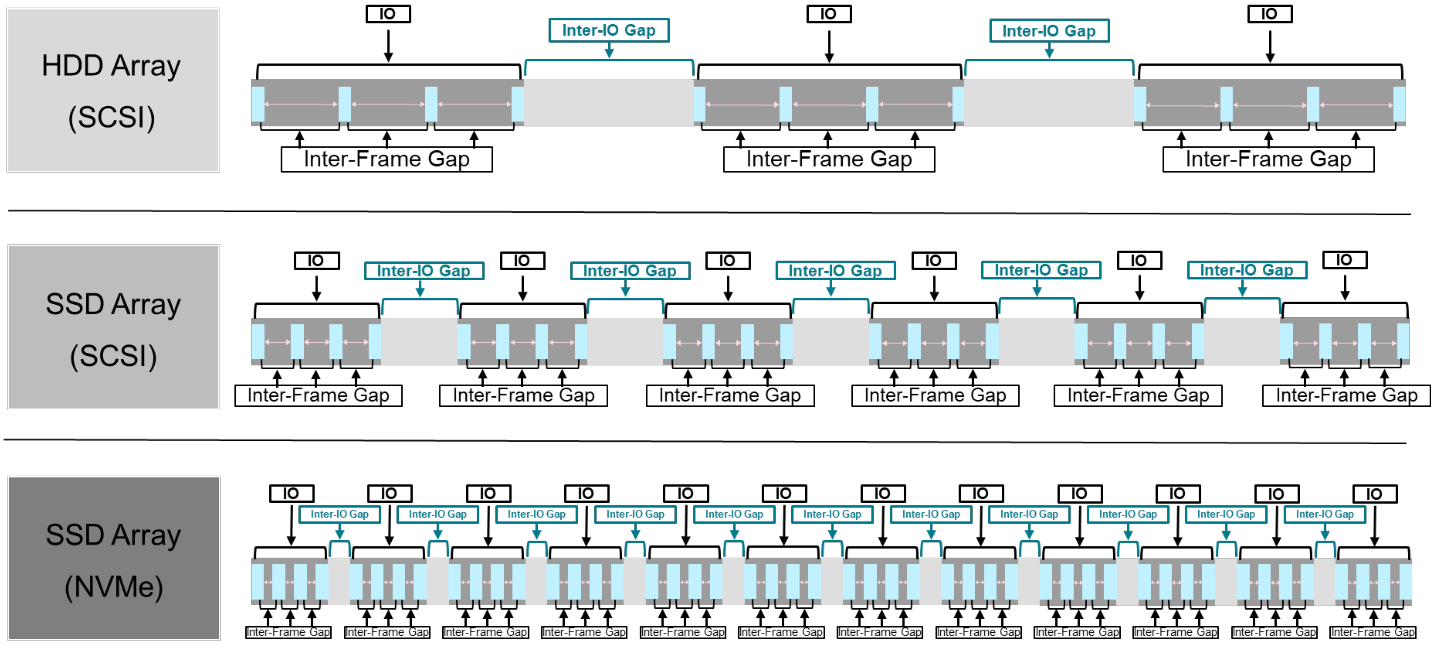
The most flexible configuration is a core-edge topology. Such an architecture allows significant scaling while keeping the number of “hops” low, which is the number of times the data transfers from one platform to another. The locality of storage connectivity for high-performance applications is a consideration that drives some storage admins to place storage ports on the same blade, switch, or port group as the application server. This design, however, does impact the flexibility of IT to move the application from one server platform to another (not a trivial consideration in a highly virtualized environment where hypervisor platforms may frequently migrate applications between server platforms).

Another alternative topology is a full-mesh design; every switch has a direct ISL to every other switch in the fabric. Full-mesh is problematic for environments with many switches since it consumes valuable ports for inter-switch links (ISLs) that otherwise would be used to connect servers and storage. Director platforms have inter-chassis links (ICLs), which allow exceptional bandwidth and scale between chassis without consuming any ports usually used for storage and server connectivity. Full-mesh ensures that no end device is more than one hop away from any other end device.

From a performance perspective, it is essential to note that the advent of the all-flash data center also means that new storage technologies, performance-based, capacity-based, or both, arrive at 18- to 24-month intervals. Technology refresh does not imply the wholesale replacement of the existing platforms, but rather that your storage network must be able to accommodate roughly two of these iterations per 4- to 5-year capital depreciation cycle. One advantage of a Fibre Channel SAN is the dual-redundant hardware and isolated nature of SAN architectures. A & B fabric architectures portend that no device defect, scheduled maintenance, accidental human event, or malicious activity will completely take storage connectivity offline. Keeping storage networking online allows for seamless technology refreshes, which applies to technology upgrades for the storage network elements and attached server and storage.

One of the additional changes required by the all-flash data center is improved monitoring, partly due to the reduction in latency, the criticality of data, and the high amount of data in-flight in modern SANs.

Figure 1: NVMe Implies Less Idle Time on the Network



As illustrated in Figure 1, the amount of idle time in the network continues to reduce. Consequently, the “event window” of a problem can be very brief, and traditional monitoring systems based on “sample rate” (inspecting perhaps one packet in 8000), while sufficient for modeling, may not provide the rapid root-cause analysis that the modern SAN requires. This situation is further exacerbated by the scale/capacity of modern all-flash storage. Specific platforms can scale above a petabyte (a petabyte [PB] is 1000 terabytes [TB]) of capacity within a mere two rack units. The consideration here is that such a platform could be hosting between 6,000 and 10,000 virtual machines or applications, and any problem or outage affecting that kind of footprint becomes intolerable. The requirement for high availability drives granular monitoring with self-optimizing and self-healing technology within the storage network. The net of the performance curve is that humans are no longer fast enough or responsive enough to address problems in the all-flash data center without forcing an outage.

2.3 NVMe

Another consideration why storage networks must be reconsidered and re-architected is Non-Volatile Memory Express (NVMe). At the device level, there are some characteristics to be aware of:

- Density – Current NVMe devices have 8 to 10 times the density of DRAM.
- Latency – Current NVMe devices have a sub-20-microsecond latency.
- Bandwidth – Current NVMe devices consume 4 PCIe Gen 3 lanes (32G).
- Streamlined software – Current NVMe software has 13 required and 25 optional commands.

Why are these characteristics important to consider? Because storage density continues to increase on a rough “Moore’s Law” schedule. Additionally, the latency of devices continues to be reduced and is already significantly lower than either HDD or traditional SSD devices. Current network environments below 64 G are potential choke points for fan-in/fan-out ratios supported by large-scale storage platforms. Another significant issue is that the language used to communicate with storage is changing for the first time in over three decades.

Taking advantage of new technology, especially its performance aspects, requires special attention to the storage network—legacy environments, whether Ethernet or Fibre Channel, will not take advantage of NVMe performance.

“Speed is the new currency of business” - Marc Benioff, CEO of Salesforce

The ability to scale to massive IOPS, drive extreme consolidation to the new storage density, and reduce the overhead on application servers with a streamlined software stack depends upon having the proper infrastructure. An NVMe over Fibre Channel fabric is a production-ready NVMe environment.

Conversion will be slow, as it will take time for all servers and applications to migrate to NVMe from SCSI. With the proper design and implementation of a Fibre Channel SAN, both NVMe and SCSI can be run concurrently on the *same* HBA, on the *same* FC switch, and to the *same* NVMe-based storage. Applications can potentially be migrated from SCSI to NVMe nondisruptively.

However, this new environment will need to be self-learning, self-optimizing, and self-healing simply because it will be too critical and performant to wait for human intervention to solve problems before they become disruptive.

Chapter 3: Architecting a SAN

The SAN planning process is similar to any type of project planning and includes the following phases:

- Phase I: Gathering requirements
- Phase II: Developing technical specifications
- Phase III: Estimating project costs
- Phase IV: Analyzing Return on Investment (ROI) or Total Cost of Ownership (TCO) (if necessary)
- Phase V: Creating a detailed SAN design and implementation plan

When selecting which criteria to meet, you should engage users, server and storage subject matter experts (SMEs), and other relevant experts to understand the role of the fabric. Since most SANs tend to operate for a long time before they are renewed, you should consider future growth as SANs are difficult to re-architect. Deploying new SANs or expanding existing ones to meet additional workloads in the fabrics requires a critical assessment of business and technology requirements. Proper focus on planning will ensure that the SAN, once deployed, meets all current and future business objectives, including availability, deployment simplicity, performance, future business growth, and cost. Tables in Appendix B are provided as a reference for documenting assets and metrics for SAN projects.

A critical aspect for successful implementation that is often overlooked is the ongoing management of the fabric. Identifying systems-level SMEs for all components that make up the SAN and adequate and up-to-date training on those components is critical for efficient design and operational management of the fabric. When designing a new SAN or expanding an existing SAN, you should consider the following parameters.

Application Virtualization

- Which applications will run under a virtual machine (VM) environment?
- How many VMs will run on a physical server?
- Under what conditions will the VMs be migrated (business and nonbusiness hours; is additional CPU or memory needed to maintain response times)?
- Is there a need for solid-state storage media to improve read response times?

Homogeneous/Heterogeneous Server and Storage Platforms

- Are blade servers or rack servers used?
- Is auto-tiering in place?
- Which Brocade Fabric OS® (FOS) versions are supported in a multivendor storage environment?
- What is the planned refresh cycle of servers and storage platforms (2 years/3 years)?

Scalability

- How many user ports are needed now?
- How many devices will connect through an access gateway?
- How many inter-switch links (ISLs)/Brocade UltraScale inter-chassis links (ICLs) are required to minimize congestion in the fabric?
- What distances for ISL/ICL connections need to be supported?
- Does the fabric scale-out at the edge or the core?

Backup and Disaster Tolerance

- Is there a centralized backup? (This will determine the number of ISLs needed to minimize congestion at peak loads.)
- What is the impact of backup on latency-sensitive applications?
- Is the disaster solution based on long-distance metro Fibre Channel (FC) ISLs or a Fibre Channel over IP (FCIP) solution?

Diagnostics and Manageability

- What is the primary management interface to the SAN (command-line interface [CLI], Brocade SANnav, or third-party tool)?
- How often will Brocade FOS and SANnav be updated?
- How is cable and optics integrity validated?

Investment Protection

- Is support needed for adding Gen 7 switches into a Gen 6 fabric?
- Is support needed for storage technologies like NVMe over Fabrics?
- What device interoperability support is required?
- Is interoperability required for other technologies such as UCS?

3.1 Operational Considerations

Even though Brocade fabrics scale in port density and performance, the design goal should ensure simplicity for the highest availability, most straightforward management, future expansion, and serviceability. Examples of this simplicity may include using a two-tier core-edge topology, avoiding using inter-fabric routing (IFR) and Virtual Fabrics (VF) where not required, and turning on port monitoring parameters for critical applications.

NOTE: Refer to the *Brocade SAN Scalability Guidelines* for currently tested and supported scalability limits. Any requirements beyond the tested scalability limits should be pretested in a nonproduction environment, orsystem resources like CPU and memory utilization should be actively monitored to minimize fabric anomalies.

3.2 Be the Pilot

Whether building a new SAN or connecting to an existing SAN, prestaging and validating a fabric or application before putting it into production ensures baseline metrics for rated throughput, latency, and expected errors based on the physical cable infrastructure, including patch panels.

3.3 Predeployment Cabling and Optics Validation

SANs built with Brocade Gen 5, Gen 6, and Gen 7 Fibre Channel switches that are equipped with 16G, or faster, optics can run Brocade ClearLink® Diagnostics. ClearLink enables predeployment testing to validate the integrity of the physical network infrastructure before operational deployment. Part of Brocade Fabric Vision® technology, Brocade ClearLink Diagnostic Port (D_Port) mode allows you to convert a Fibre Channel port into a diagnostic port for testing traffic and running electrical and optical loopback tests. The test results can be beneficial in diagnosing a variety of port and link problems. ClearLink is an offline diagnostics tool that allows users to perform an automated battery of tests to measure and validate maximum throughput speeds, latency, and distance across fiber infrastructure. ClearLink Diagnostics can also be used to verify the health and integrity of all 16G and 32G transceivers in the fabric. Diagnostics should be conducted before deployment to vet potential CRC errors caused by physical-layer issues.

A ClearLink Diagnostics port (D_Port) requires that only the individual ports attached to the tested link go offline, allowing the remainder of the ports to stay online in isolation from the link. D_Port can also be used to test links to a new fabric switch without allowing the new switch to join or even be aware of the current fabric, providing an opportunity to measure and test ISLs before they are put into production. This fabric-based physical-layer validation enables the following:

- Transceiver health check
- Transceiver uptime
- Local and long-distance measurements (5-meter granularity for 16G and 32G small form-factor pluggable [SFP] optics and 50 meters for 10G optics)
- Link latency measurements between D_Ports
- Link power (dB) loss
- Link performance

Refer to the *Brocade Fabric OS Troubleshooting and Diagnostics User Guide* for a more detailed discussion of diagnostic port usage.

Refer to “Appendix A: ClearLink Diagnostics” in the *SAN Fabric Resiliency and Administration Best Practices User Guide* for details on enhancements in each FOS release.

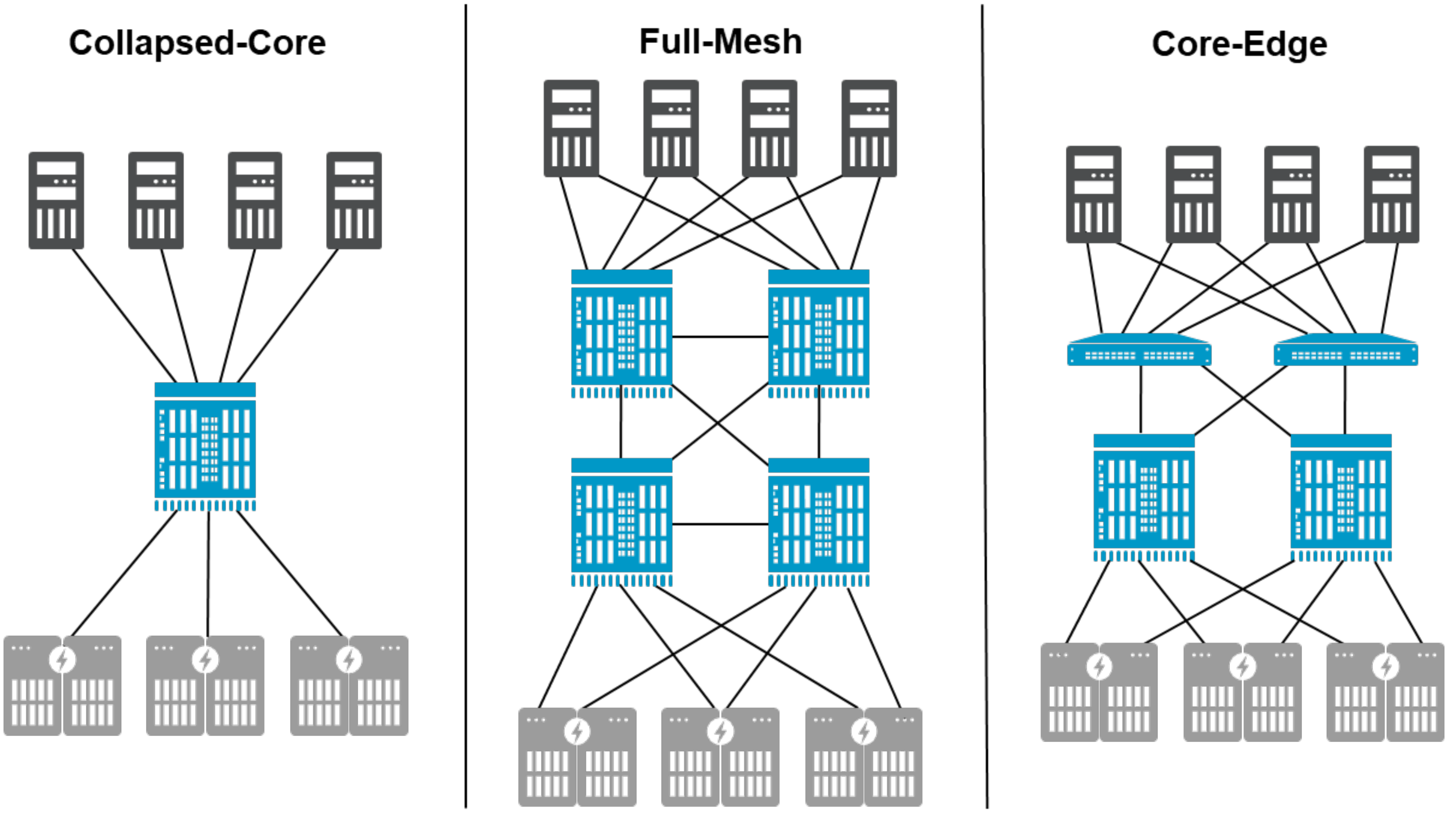
Chapter 4: SAN Design Basics

This chapter provides high-level guidelines for architecting a typical SAN. The focus is on best practices for collapsed-core, core-edge, and mesh fabrics. The discussion starts at the highest level, the data center, and works down to the port level, providing recommendations at each point along the way.

4.1 Topologies

A typical SAN architecture comprises devices on the network's edge, switches in the network's core, and the cabling connecting the devices. Topology is usually described in terms of how the switches are interconnected, such as collapsed-core, core-edge, and full-mesh. The recommended SAN topology to optimize performance, availability, management, and scalability is a tiered, core-edge topology. The core-edge approach provides good performance without unnecessary interconnections. At a high level, the tiered topology has a large number of edge switches used for device connectivity and a smaller number of core switches used for routing traffic between the edge switches, as shown in Figure 2.

Figure 2: Three Scenarios of Tiered Network Topologies



The difference between these three scenarios is device placement, where devices are attached to the network, and the associated traffic flows.

- Scenario A, collapsed-core, has localized traffic within a single platform fabric. Flows could be cut-through switched within the same ASIC on a blade. Alternatively, flows may be switched between ASICs on the same or different blades within the chassis. A collapsed-core can have small performance advantages for performance-critical latency-sensitive workloads. but it does not scale this performance beyond the ASIC's port group. Manageability is significantly reduced by having only a single platform per fabric. How many fabrics are deployed determines overall manageability.
- Scenario B, core-edge, separates storage and server connectivity, thus providing ease of management and greater scalability. A core-edge topology has only one fabric hop from server to storage, providing identical performance as full-mesh while allowing greater scalability.
- Scenario C, full-mesh, has no more than one fabric hop between server and storage, assuming the servers and storage are not connected to the same platform. Designing fabrics with UltraScale ICLs is an efficient way to save valuable FC ports. Using best-practice SAN design considerations, users can quickly build a large fabric with 3456 ports or more.

4.1.1 Collapsed-Core

The collapsed-core topology (Figure 2) places initiators (servers) and storage (targets) on the same chassis and potentially the same blade or even the same ASIC. This topology has several benefits depending on the size of the environment. collapsed-core is used when customers migrate from multiple switches to a single, dual-core architecture in which all initiators and targets can fit onto the same core switch. If future design requirements include further increasing scale with simplistic management, starting with a core-edge architecture may be beneficial in the long run.

4.1.2 Core-Edge

The core-edge topology (Figure 2) places initiators (servers) on the edge tier and targets (storage) on the core tier. For redundancy, each fabric (A & B) has two cores. Since servers and storage are on different switches, this topology provides easy management, outstanding performance, and minimal latency with data flows traversing one hop from edge to core. Storage-to-storage traffic will require two hops if the second storage platform destination is not connected on the same core. If storage-to-storage connectivity is required, the two cores within the same fabric can be connected. The disadvantage to a core-edge design is that storage and core-to-edge connections contend for expansion as the environment scales; however, director platforms are flexible, allowing ICLs for inter-switch connectivity and freeing up ports for additional devices.

4.1.3 Full-Mesh

A full-mesh topology (Figure 2) allows you to place servers and storage anywhere since communication between source and destination is no more than one hop. Using director-class switches with UltraScale ICL ports for interconnectivity is essential to this design to ensure maximum device port availability and utilization. Design this architecture with a minimum of two switches and up to nine switches in a full-mesh.

4.2 High-Performance Latency-Sensitive Workloads

Over the last few years, enterprises have come to leverage low-latency, high-throughput flash arrays for demanding, performance-sensitive workloads. Brocade's Gen 7 Fibre Channel is ideally suited to these types of workloads due to the sub-microsecond latency through the switch and the increased bandwidth offered by 32/64G throughput speeds while providing accurate I/O latency instrumentation. Performance testing has shown that 32-Gb and even 8-Gb all-flash arrays can realize dramatic benefits by connecting to a Gen 7 SAN and using Gen 7 HBAs, with gains of up to 2x over Gen 5 and Gen 6 SANs.

The Gen 6 and Gen 7 standards include forward error correction (FEC) to ensure transmission reliability and a highly deterministic data flow. FEC corrects up to 140 corrupt bits per 5280-bit frame at the receiving end of the link, avoiding the need to retransmit frames when bit errors are detected.For highly demanding workloads, a no-hop fabric connection through a one-ASIC switch like the Brocade G720 or local switching within an ASIC on a director's port blade minimizes latency to sub-microsecond speeds. Local switching performs cut-through switching of FC frames from the ingress port to the egress port when in the same port group. Some platforms contain multiple switching ASICs between data ingress and egress, like the X6/X7 directors and G730 switches. Keeping host and storage connections within an ASIC's port group minimizes latency and avoids moving data between ASICs. To find details on port groups and local switching, refer to the *Brocade Fabric OS Administration Guide* and the hardware installation guide for the appropriate product.

4.3 Redundancy and Resiliency

An essential aspect of SAN architectures is the resiliency and redundancy of the fabrics. The objective is to remove single points of failure. Resiliency is the ability of the network to continue to function by recovering from a failure. In contrast, redundancy describes the duplication of components, typically the entire fabric, to eliminate a fabric failure as a single point of failure in the overall SAN. Brocade fabrics have resiliency built into Brocade Fabric OS (FOS), which runs on all Brocade platforms. FOS can quickly "repair" and overcome most failures. For example, FSPF (Fabric Shortest Path First) quickly computes new paths when the fabric topology changes or a link goes offline. Of course, this assumes a second path exists, which is when fabric resiliency is important.

The key to high availability and enterprise-class availability is redundancy. By eliminating an entire fabric as a single point of failure, business continuance is provided through most foreseeable and unforeseeable events. At the highest level of fabric design, the complete fabric should be redundant, with two mirrored, entirely different fabrics that do not share any common SAN platforms.

Servers and storage devices should be connected to both fabrics (A & B) leveraging some form of multipath I/O (MPIO), such that data can flow across both fabrics seamlessly in either an active/active or active/passive mode. MPIO ensures that an alternate path is available if the current path fails. Ideally, redundant fabrics are identical, but at a minimum, they should be based on the same switches to ensure consistency of performance and delivery. In some cases, these fabrics are in the same location. However, two separate locations are often used to provide for disaster recovery (DR), either for each complete fabric or sections of each fabric.

Regardless of the physical geography, there are two different fabrics for complete redundancy.

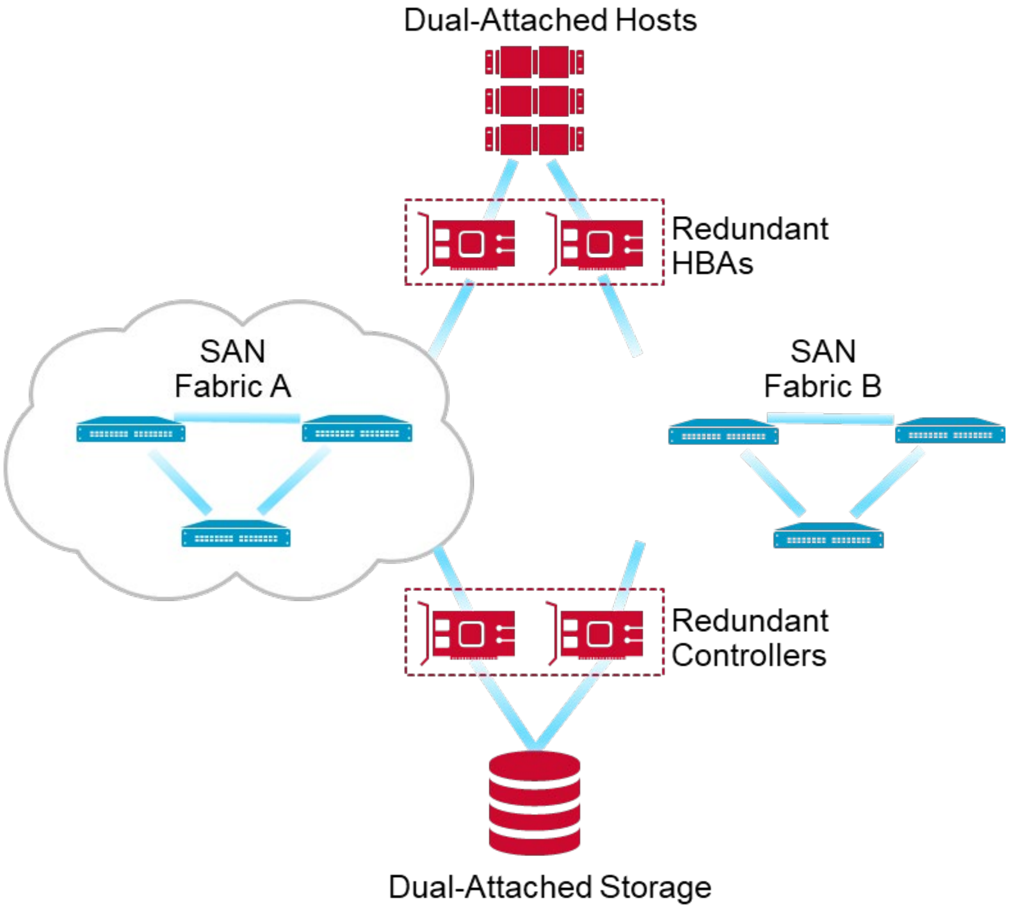
In summary, best practices for SAN design are to ensure application availability and resiliency via the following:

- Fabric redundancy to avoid a fabric being the single point of failure
- Resiliency built into each fabric to avoid a single point of failure within the fabric
- Redundant connections from each host to each fabric
- MPIO-based failover from initiator to target
- Identical architectures and platforms in each fabric
- Redundant ISL/IFL/ICL for inter-switch connectivity
- Separate storage (core tier) and server (edge tier) tiers for independent expansion
- Core switches of equal or higher performance compared to the edge switches
- Defining the principal switch to be the highest performance switch in the fabric

4.4 Switch Interconnections

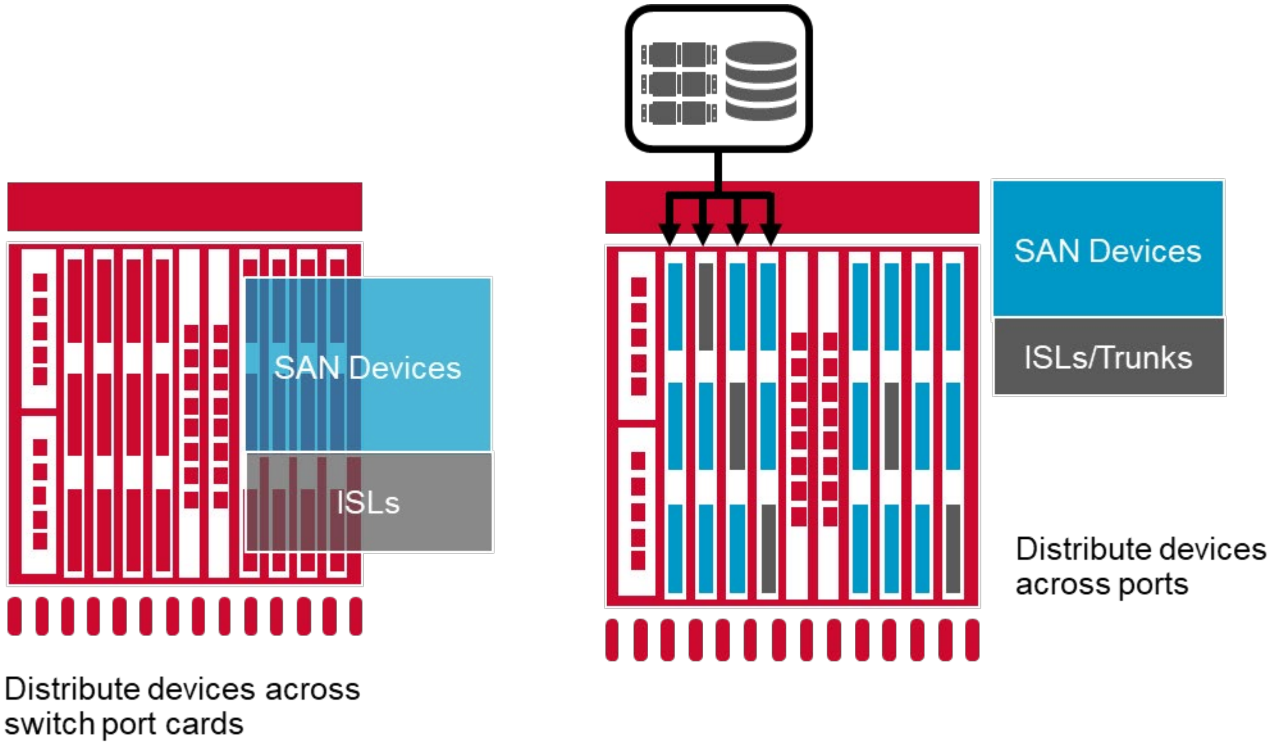
As mentioned previously, there should be at least two of every element in the SAN to provide redundancy and improve resiliency. The number of available ports and device locality (server/storage tiered design) determines the number of ISLs needed to meet performance requirements. ISL requirements for directors include a minimum of two trunks with at least two ISLs per trunk. Each source switch should be connected to at least two other switches, and so on. In Figure 3, each blue connection line represents at least two physical cables. Two physical connections provide redundancy for ports, optics, fiber patch cables, patch panels, and fiber infrastructure.

Figure 3: Connecting Devices through Redundant Fabrics



Redundant trunks on a director platform should be placed in varying port groups on different blades, as shown in Figure 4. See the appropriate hardware manual to determine port groups for the various models of port blades. For more details, refer to the *Brocade Fabric OS Administration Guide*. Whichever method is decided upon, it is crucial to be consistent across the SAN. For example, do not place ISLs on lower port numbers in one chassis, as shown in the left diagram in Figure 4, and stagger ISLs on a different chassis, as shown in the right diagram in Figure 4.

Figure 4: Examples of Distributed ISL Placement for Redundancy


SAN Devices
ISLs
Distribute devices across switch port cards
SAN Devices
ISLs/Trunks
Distribute devices across ports

NOTE: In Figure 4, ISL trunks are placed on separate ASICs or port groups. It is important to match ISL placement between devices and across fabrics to ensure simplicity in design and assist in problem investigation.

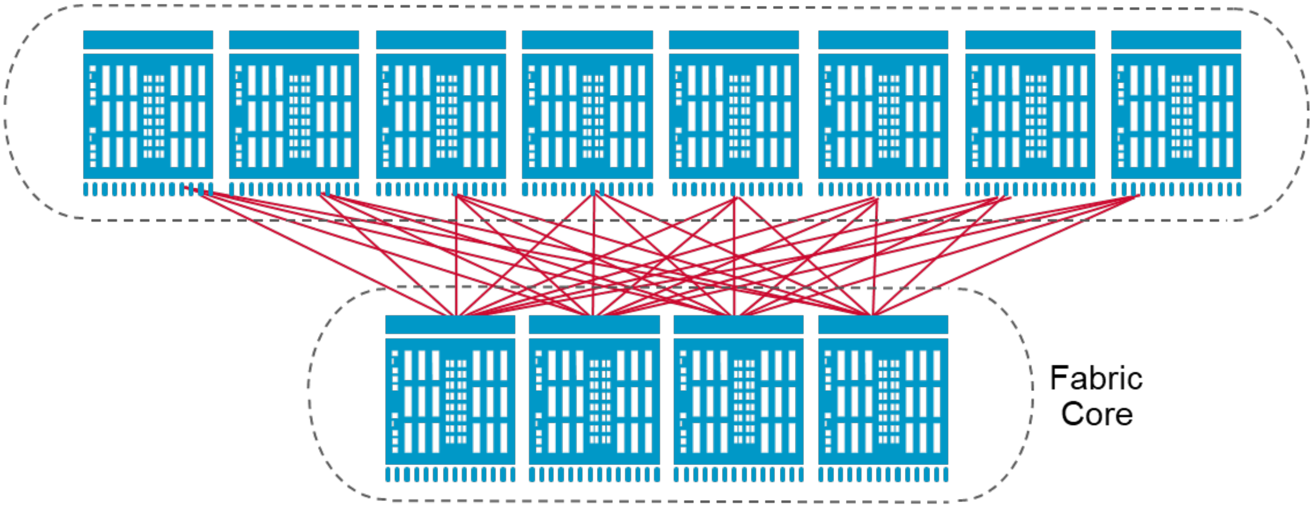
4.4.1 UltraScale ICL Connectivity for Gen 5 Brocade DCX® 8510-8/8510-4, Gen 6 Brocade X6-8/X6-4, and Gen 7 Brocade X7-8/X7-4

The Brocade DCX® 8510, X6, and X7 platforms use second-generation UltraScale ICL technology from Broadcom with optical QSFPs. The Brocade DCX 8510-8, X6-8, and X7-8 support up to 32 QSFP ports per chassis (Figure 5), and the Brocade DCX 8510-4, X6-4, and X7-4 support up to 16 QSFP ports to help preserve director ports for connections to end devices. Each QSFP port has four independent links, and each terminates on a different ASIC on the core blade.

NOTE: X6 ICL ports can connect to DCX 8510 ICL ports at 16G speeds, with the port speed on the X6 ICL configured to 16G and the 16G QSFP in the X6 ICL port.

NOTE: X7 ICL ports can connect to DCX 8510 ICL ports at 16G speeds, with the port speed on the X7 ICL configured to 16G and the 16G QSFP in the X7 ICL port.

Figure 5: Twelve-Chassis UltraScale ICL-Based Core-Edge Design

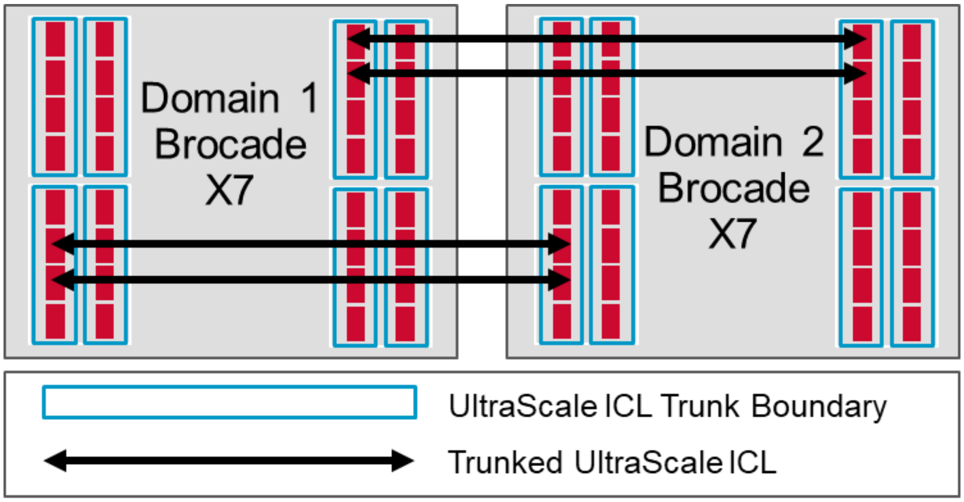


4.5 Best Practices for Brocade UltraScale ICL Connections

Each core blade in a chassis must be connected to each of the two core blades in the destination chassis to achieve full redundancy (Figure 6). In Figure 5 above, each director has 32 ICL ports, 16 on each core routing blade. There are eight edge directors and four core directors. Each edge director has two connections from each core routing blade to each core director's corresponding core routing blade. Each core director connects to the edge directors using (8 edge directors x 2 connections each x 2 core routing blades each) = 32 connections.

NOTE: A pair of ICL links are used to connect core routing blades for redundancy.

Figure 6: Minimum ICL Connections Needed between Brocade X7 Chassis

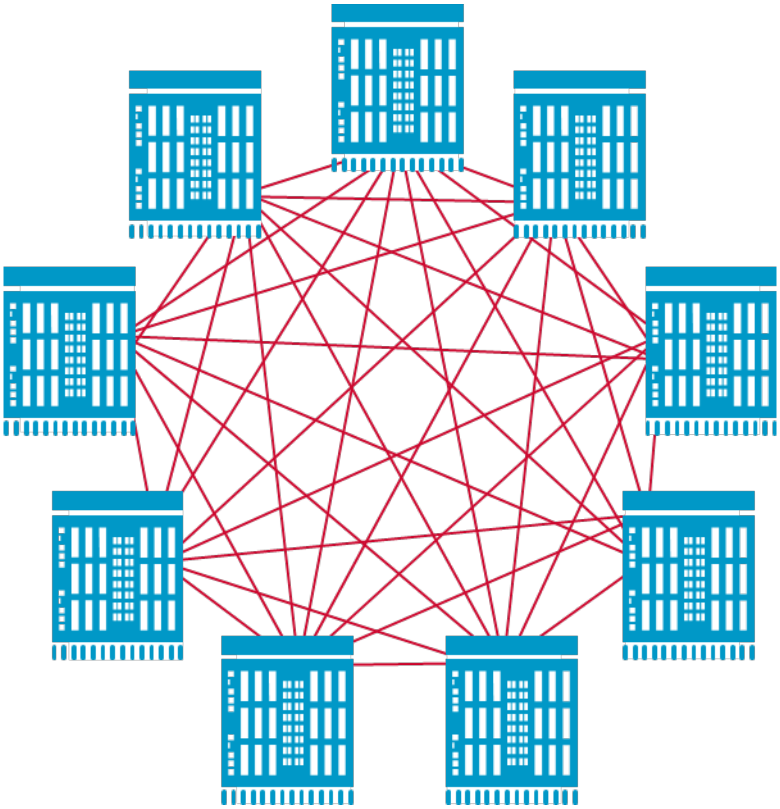


4.6 Full-Mesh Topology

A full-mesh architecture provides a single hop between source and destination. Broadcom supports a 9-director ICL mesh with up to 100-meter distances using select QSFPs and OM4 fiber. In the example shown in Figure 7, up to 4608 (9 X7 directors x 8 blades in each director x 64 ports on each blade) 64G FC device ports are supported using UltraScale ICLs with a 512 Gb/s ICL (2 ICL links x 256Gb/s for each ICL link) between each director.

Alternatively, if the full-mesh had 5 directors instead of 9, there would be 2560 end-device 64G FC ports with 1 Tb/s between each director using ICL connectivity.

Figure 7: Nine-Chassis UltraScale ICL-Based Full-Mesh Topology



NOTE: Refer to the *Scale-Out Architecture with Brocade UltraScale Inter-Chassis Links Design Guide* for details.

UltraScale ICL connections are considered a “hop of no concern” in a FICON fabric.

When using a core-edge architecture methodology, edge switches should connect to at least two core switches via trunks of at least two ISLs each. Each trunk should be attached to a different blade. Redundancy requires an identically mirrored second fabric, and end devices must be connected to both fabrics using MPIO to manage active/active or active/passive flows and failover/failback.

The following are recommendations for ISL/UltraScale ICL connectivity:

- There should be at least two core switches.
- Every edge switch should have at least two trunks to each core switch.
- Create enough small trunks. Keep trunks to two ISLs unless anticipating very high traffic volumes. Small trunks ensure that losing an entire trunk does not result in losing significant connectivity. Plan the number of trunks based on at least one trunk being offline at some point.
- Place redundant trunks on different blades.
- Trunks are in a port group, ports within an ASIC boundary.
- Allow no more than 30m in cable difference for optimal ISL trunk performance.
- Use the exact cable length for all UltraScale ICL connections.
- Use either ISL or UltraScale ICL connectivity into the same domain. Mixing the two types of connections is not supported.
- Use the same type of optics on both sides of the trunks: Short Wavelength (SWL), Long Wavelength (LWL), or Extended Long Wavelength (ELWL).

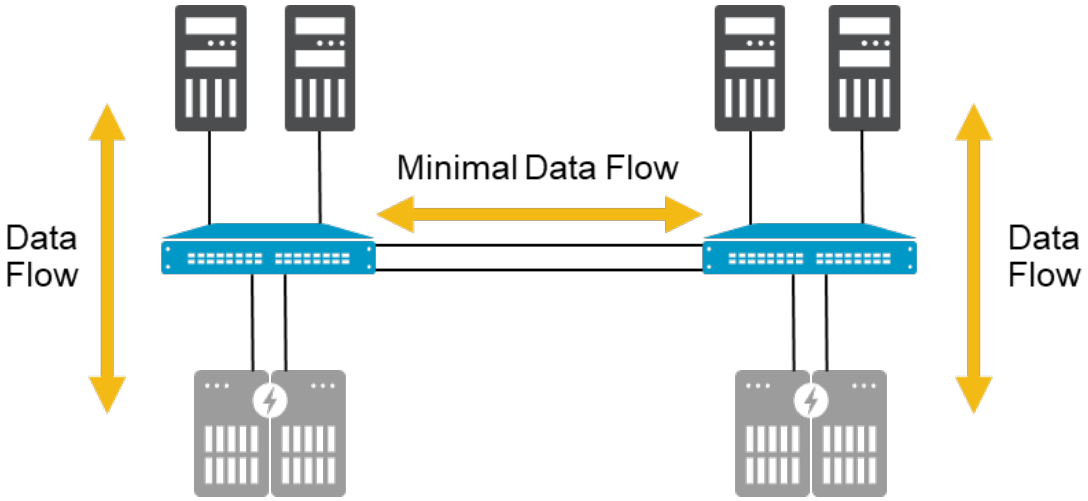
4.7 Device Placement

Device placement is a balance between traffic isolation, scalability, manageability, and serviceability. Virtualization has dramatically optimized compute platforms, driving the need for high performance and improved scalability in storage networks. Frame congestion can become a severe concern if there are interoperability issues with the end devices.

4.7.1 Traffic Locality

Designing device connectivity depends significantly on the expected data flow between devices. For simplicity, communicating hosts and targets can be attached to the same switch (Figure 8).

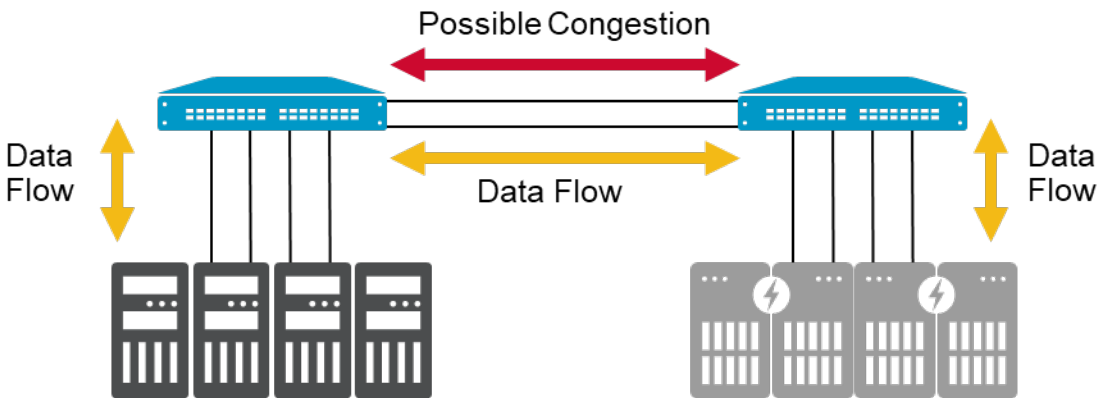
Figure 8: Hosts and Targets Attached to the Same Switch to Maximize Locality of Data Flow



However, this approach does not scale well. Given Fibre Channel's high-speed, low-latency nature, attaching these host-target pairs on different switches does not mean that performance is adversely impacted for typical workloads. With the current generation of switches, local switching is not required to gain performance or achieve low latency. Nevertheless, architects may want to switch traffic locally for mission-critical applications that depend on extremely fast response times.

In exceptional cases, multihop concerns may involve traffic congestion, specifically, inadequate inter-switch connectivity or concerns about proper resiliency (Figure 9). Often, these concerns can be mitigated by provisioning ISLs/UltraScale ICLs.

Figure 9: Hosts and Targets Attached to Different Switches for Ease of Management and Expansion



One frequently used scheme for scaling a core-edge architecture is dividing the edge switches into a storage/target tier and a host/initiator tier. This approach lends itself to easier management as well as further expansion. In addition, host and storage devices generally have different performance requirements, cost structures, and other factors that can be readily accommodated by placing initiators and targets in different tiers.

Chapter 5: Data Flow Considerations

5.1 Fan-In Ratios and Oversubscription

A critical aspect of data flow is the *fan-in ratio* or *oversubscription* in terms of source ports to target ports and devices to ISLs. Oversubscription can also be viewed from the storage array perspective, referred to as the *fan-out ratio*. The ratio is the number of device ports that share a single port, whether ISL, UltraScale ICL, or target. The ratio is always expressed from the single entity's point of view, such as 7:1 for seven hosts utilizing a single ISL or storage port.

What is the optimum number of hosts that should connect to a storage port? This question seems reasonably straightforward. However, the situation can quickly become much more complex once you consider clustered hosts, VMs, workload characteristics, and the number of LUNs (logical unit numbers) per server. Determining how many hosts to connect to a particular storage port can be narrowed down to three considerations: port queue depth, I/O per second (IOPS), and throughput. Of these three, throughput is the only network component. Thus, a simple calculation adds up the expected peak bandwidth usage for each host accessing the storage port.

In practice, though, it is improbable that all hosts perform at their maximum level simultaneously. The bandwidth of the host bus adapter (HBA) was considerably overprovisioned with traditional application-per-server deployments. However, the game changed radically with virtual servers (KVM, Xen, Hyper-V, proprietary UNIX OSs, and VMware). Conceptually, oversubscription is built into virtual servers to optimize server resources. To the extent that servers optimize resource utilization, they should reduce oversubscription proportionately. Therefore, it may be prudent to oversubscribe ports to ensure a balance between cost and performance.

Another method is to assign host ports to storage ports based on the I/O capacity requirements of the host. The intended result is a small number of high-capacity servers assigned to each storage port, which results in a large number of low-capacity VM workloads distributed across multiple storage ports.

Regardless of the method used to determine the fan-in/fan-out ratios, port monitoring should determine actual utilization and any appropriate adjustments. In addition, ongoing monitoring provides valuable heuristic data for effective expansion and efficient assignment of existing storage ports. A simple calculation method works best to determine the device-to-ISL fan-in ratio: the storage port should not be oversubscribed into the core. For example, a 32G storage port should have a 32G pipe into the core.

Chapter 6: Scalability and Performance

Broadcom products are designed with scalability in mind, knowing that most installations will continue to expand and that growth is supported with very few restrictions. However, following the same basic principles outlined in previous sections as the network grows will ensure that the levels of performance and availability will continue.

Evaluate the impact on topology, data flow, workload, performance, and perhaps most importantly, redundancy and resiliency of the entire fabric any time one of the following actions is performed:

- Adding or removing initiators:
  - Changes in workload
  - Changes in provisioning
- Adding or removing storage:
  - Changes in provisioning
  - Changes in storage media type (for example, increased deployment of flash-based storage)
- Adding or removing switches
- Adding or removing ISLs and ICLs
- Change in virtualization (workload and storage) strategies and traffic flow pattern

If these design best practices are followed when the fabric is deployed, small incremental changes should not adversely impact the availability and performance of the fabric. However, if ongoing changes occur and the fabric is not properly evaluated and updated, performance and availability can be jeopardized. Some key points to cover when looking at the current status of a production FC SAN include the following:

Reviewing redundancy and resiliency:

- Are there two or more redundant fabrics?
- Are there two or more physically independent paths between each source (initiator) and destination (target) pair?
- Does each host connect to two different edge switches?
- Are edge switches connected to at least two different core switches?
- Are inter-switch connections composed of two trunks of at least two ISLs?
- Does each storage device connect to at least two different edge switches or separate port blades?
- Are storage ports provisioned such that every host has at least two ports through which it can access LUNs?
- Are redundant power supplies attached to different power sources?
- Are zoning and security policies configured to allow for patch/device failover?

Reviewing performance requirements:

- Host-to-storage port fan-in/out ratios
- Oversubscription ratios:
  - Host to ISL
  - Edge switch to core switch
  - Storage to ISL
- Size of trunks

- Routing policy and currently assigned routes (evaluate actual utilization for potential imbalances)
- Use of FEC for all ISLs and connections to Gen 5 (if supported) and Gen 6 devices

Watching for latencies because of the following:

- Poor storage performance
- Overloaded hosts or applications
- Distance issues over constrained long-distance links resulting from changes in usage, such as adding mirroring or too many workloads
- Deteriorating optics resulting in declining signal strength and increased error rate

In Gen 6 and Gen 7 networks, storage response latency can be baselined and monitored continuously using IO Insight in conjunction with MAPS. Deal with latencies immediately; they can impact the fabric profoundly.

In summary, although Brocade SANs are designed to allow for any-to-any connectivity, and they support provision-anywhere implementations, these practices can harm the performance and availability of the SAN if left unchecked. As detailed above, the network needs to be monitored for changes and routinely evaluated for how well it meets desired redundancy and resiliency requirements.

Chapter 7: Supportability

Supportability is a critical part of deploying a SAN. Follow the guidelines in this chapter to ensure that the data needed to diagnose fabric behavior or problems has been collected.

- Configure Brocade MAPS: Leverage Brocade MAPS to implement proactive monitoring of errors and warnings such as CRC errors, loss of synchronization, and high-bandwidth utilization.
- Configure syslog forwarding: By keeping historical log messages and having all switch messages sent to one centralized syslog server, troubleshooting can be expedited and simplified. Forwarding switch error messages to one centralized syslog server and keeping historical log messages enables faster and more effective troubleshooting and provides simple monitoring functionality.
- Create a switch configuration template in SANnav to avoid configuration drift from occurring over time. You can quickly adopt existing configurations as a template for deploying new switches in the fabric, ensuring consistency across the data center.
- Follow Broadcom's management interface best practices for the data center LAN. Broadcom's best practice for a management LAN is to set up different physical broadcast domains for the management interfaces belonging to each fabric. Placing an IP router between broadcast domains and configuring each broadcast domain on a different VLAN is typical. Do not put the management interfaces from fabric A on the same VLAN as the management interfaces from fabric B. Secure each VLAN to required and authorized access only.
- Enable audit functionality: To provide audit functionality for the SAN, keep track of which administrator made which changes, usage of multiple user accounts (for example, using RADIUS), and configuration of change tracking or audit functionality along with the use of error logs/syslog forwarding.
- Configure multiple user accounts (LDAP/OpenLDAP or RADIUS): Make mandatory the use of personalized user accounts as part of the IT and SAN security policy so that user actions are tracked. Also, restrict access by assigning specific user roles to individual users.
- Establish a testbed: Set up a testbed to test new applications, firmware upgrades, driver functionality, and scripts to avoid missteps in a production environment. Validate functionality and stability with rigorous testing in a test environment before deploying into the production environment.
- Implement a serial console server: Implement remote serial access to manage switches even when there are network issues or problems during switch boot or firmware upgrades.
- Use aliases: Use aliases to give switch ports and devices meaningful names for faster troubleshooting.
- Configure `supportftp`: Configure `supportftp` for automatic file transfers. The parameters set by this command are used by `supportSave` and `traceDump`.
- Configure an NTP server: To keep a consistent and accurate date and time on all switches, configure switches to use an external time server.

Exh. RA-3
Docket UT-181051
Page 31 of 112

7.1 Firmware Upgrade Considerations

Both fixed-port and modular switches support hot code load for firmware upgrades.

- Disruptive versus nondisruptive upgrades:
  - Simultaneous upgrades on neighboring switches
  - Standard FC ports versus application and special-feature ports
- Review the *Brocade Fabric OS Release Notes* for the following:
  - Upgrade path
  - Changes to feature support
  - Changes to backward compatibility
  - Known issues and defects
- Consider a different AG firmware upgrade strategy. Brocade Access Gateways have no fundamental requirement to be at the same firmware release level as Brocade FOS. Upgrading only directors and switches minimizes the infrastructure changes required during an upgrade cycle.

Chapter 8: Monitoring

8.1 Brocade Fabric Vision Technology

Organizations face a constant struggle to both manage data growth and deliver actionable intelligence from raw data—all while meeting SLAs. As a result, even well-managed IT organizations must often make difficult choices about resource allocation, weighing the benefits of focusing more resources on monitoring, for instance, and fewer resources on planning or optimizing. With Brocade Fabric Vision technology, organizations can achieve unprecedented insight and visibility across the storage network through critical monitoring and diagnostic capabilities.

8.1.1 Monitoring and Alerting Policy Suite

Monitoring Alerting Policy Suite (MAPS) is a health and threshold monitoring tool that allows for autonomous self-monitoring of directors and switches in the fabric. It helps detect potential and active problems, automatically alerting users to those problems long before they become costly outages. MAPS is a part of the Brocade Fabric Vision feature set and is available with FOS 7.2.0 and above.

MAPS tracks a variety of SAN fabric health categories and events. Monitoring fabric-wide events, ports, bit errors, and environmental parameters enables early fault detection and isolation as well as a means to measure performance. All health monitoring categories are customizable, providing flexibility around how and what users want to monitor. Create your own monitoring groups, assign custom thresholds, and with FOS 9.0 and above, gain the same monitoring capabilities at a flow level. MAPS means users can now threshold-monitor application flows for abnormal completion times to manage SLAs. Users can also easily integrate MAPS with enterprise operations solutions.

MAPS also provides predefined monitoring policies for users to get a quick start. These policies provide thresholds derived from 20 years of best practices and customer experiences. Based on how closely users want to monitor their SAN environment, they can select from conservative, moderate, or aggressive policies. If the default policies do not meet your needs, customize the thresholds and actions, and activate your custom policy. MAPS provides notifications before problems arise, such as reporting overutilized ports approaching specified bandwidth limits, potentially leading to congestion. These insights enable SAN administrators to perform preemptive network maintenance, such as trunking or zoning, avoiding potential network failures.

MAPS also lets you define how often switches and fabric elements are measured while specifying notification thresholds. Whenever fabric elements exceed these thresholds, MAPS can automatically take action. These actions include administrative notifications using email, SNMP traps, RASLog entries, and automated actions, such as SDDQ (Slow Drain Device Quarantine) and FPIN (Fabric Performance Impact Notification).

8.1.1.1 MAPS Recommendations

Brocade MAPS is a recommended optional feature that provides threshold monitoring of multiple switch elements. For instance, MAPS monitors port groups based on the port type, allowing for different thresholds to be set within those port groups and for the port groups to be monitored simultaneously. Different port types (for example, F_Ports, E_Ports, N_Ports) tend to differ in characteristics. MAPS provides the flexibility in monitoring and alerting capabilities to address a wide variety of cases.

MAPS also allows for the monitoring and alerting of IO Insight flow metrics, providing SAN admins with notifications of performance degradation and an early alert into developing congestion issues that may impact storage response times. When support for VM Insight is enabled, potential issues can be identified end to end from the individual virtual machine to the LUN to which it is communicating.

8.1.1.2 Tips on Getting Started with MAPS

Are you new to Brocade Monitoring and Alerting Policy Suite and looking to get started with the autonomous monitoring of your SAN environments? MAPS will provide deep insights into the health of your SAN and its performance with a single click. The following are some quick tips on the initial use of MAPS. Note that a Fabric Vision license is required before taking advantage of all monitoring capabilities, which enables over 300 additional rules.

When starting with MAPS, SAN admins should begin monitoring their fabric with one of the three predefined policies (conservative, moderate, aggressive), ideally the dflt_conservative_policy. This policy will allow SAN admins to better understand what MAPS monitors, the severity of thresholds set, and the generated alerts. If the conservative policy is not meeting your monitoring needs, change to the moderate or aggressive policy. SAN admins can begin personalizing any default policies with their observed thresholds and desired actions to fit their SAN environments better.

SAN admins can implement policy customization and management through the Brocade SANnav™ Management Portal or the CLI. The following are some examples of customization.

- Clone predefined policies for customizations of individual thresholds and rules.
- Create custom monitoring groups (for example, ports, SFPs, application flows).
- Distribute policies across the SAN for uniform fabric monitoring.
- Configure MAPS actions and take advantage of automated problem mitigation.
- Create custom MAPS monitoring dashboards through SANnav Management Portal.

8.1.2 Fabric Performance Impact Monitoring

Fabric Performance Impact (FPI) monitoring leverages predefined MAPS policies to automatically detect and alert administrators to the severity of latency and identify slow drain devices that could impact network performance. This feature enables the detection of various latency severity levels, exactly pinpointing which devices are causing backpressure in the fabric and are being impacted by a bottlenecked port. MAPS and FPI then work together to automatically quarantine the slow drain devices, preventing buffer credit starvation.

8.1.3 SDDQ Explained

Lost BBCs (buffer-buffer-credits), Credit-Stall, and oversubscription (OS) often lead to fabric congestion and backpressure. Backpressure can potentially affect neighboring flows, called victim flows, resulting in widespread performance degradation. The fabric uses automated Slow Drain Device Quarantine (SDDQ) to mitigate backpressure through MAPS.

SDDQ works with MAPS and FPI monitoring to detect various congestion scenarios and then isolates problematic devices to a low-priority virtual channel (VC) with different resources. By default, in a Brocade fabric, traffic runs in medium-priority VCs. SDDQ occurs automatically and nondisruptively based on frame loss, oversubscription, and impacted performance experienced in the fabric. Note that MAPS actions can be individually enabled for any of these conditions. Once the problematic traffic flow is isolated, the backpressure in the fabric is relieved, freeing up BBCs and link bandwidth for flows in the medium-priority VCs.

Also, FPI monitoring continually checks for cleared congestion conditions on impacted devices, allowing MAPS to automatically unquarantine previously quarantined flows. The MAPS unquarantine action moves flows back into the medium-priority VCs. This process is similar to the quarantine action and is nondisruptive.

The SDDQ and unquarantine MAPS actions are supported on devices on the local switch and the remote switch (attached via ISLs) and on devices attached through Brocade Access Gateways.

8.1.4 Flow Vision

Flow Vision was designed as a diagnostics tool and is supported on all Brocade SAN platforms running Fabric OS 7.2 and later. Flow Vision provides SAN administrators with traffic flow visibility in the fabric and can copy traffic flows for later analysis. Flow Vision also allows for test-flow generation at line-rate speeds to prevalidate SAN hardware performance and connectivity. Use the flow generation capability before operational deployment where possible to confirm optimal health and the ability to support spikes in throughput.

For mission-critical applications, consider running Flow Vision constantly to keep a historical record of application performance profiles and intermittent irregularities. For application owners who may frequently call, run Flow Vision regularly when time permits to verify good fabric health and to preempt lurking issues.

8.1.5 IO Insight

IO Insight, also known as Flow Monitor, is supported by Broadcom's Gen 6 and Gen 7 Fibre Channel switching products, which provide deeper flow-level IO statistics. These statistics include storage device latency and IOPS metrics such as first IO response time, IO completion time, and the number of pending IOs for a specific initiator and target or target and LUN. IO Insight provides IO workload monitoring and early detection of storage performance degradation.

IO Insight metrics should be monitored for storage devices that support critical applications. These IO Insight metrics should be added to MAPS policies and dashboards to report storage response time and performance degradation. This reporting is of tremendous value for performance-sensitive workloads, enabling administrators to meet critical SLAs. IO Insight provides application and storage administrators feedback on performance over time on device reliability and performance optimization. For example, pending IOs will measure the current queue depth of an HBA and can be used to fine-tune the server queue depth configuration.

Beginning with Fabric OS 9.0, IO Insight autonomously learns all flows traversing a switch with no user configuration required, as it is enabled by default. Once switches are discovered via SANnav Management Portal, telemetry data is automatically propagated to the management platform, which can then be utilized for flow-level and application-level investigation.

Refer to the *Brocade Fabric OS Flow Vision User Guide* for configuration and usage details on Flow Vision and IO Insight.

8.1.6 VM Insight

The VM Insight feature essentially provides the same IO and performance-level metrics that IO Insight provides but for individual virtual machines. This feature can logically distinguish individual VM flows down to the LUN, even if other VMs share the same LUN. VM Insight allows for unprecedented visibility for monitoring the health and performance of applications running on virtual machines.VM Insight integrates with MAPS, allowing users to threshold-monitor and alert on VM-level flow-performance metric deviations, similar to IO Insight/Flow Monitor.

This feature is available on Gen 6 platforms running Fabric OS 8.1 and later and on all Gen 7 platforms.

8.2 SANnav Management Portal Monitoring Overview

SANnav Management Portal is Broadcom's GUI-based management platform, tightly integrating with Fabric OS. From feature configuration to analysis of gathered telemetry data and events, SANnav provides actionable insight to SAN administrators.

Brocade SANnav uses Fabric OS features to detail device health, congestion, and flow telemetry, investigate concerns, troubleshoot issues, and customize dashboards. See Section 16.1.2, SANnav Management Portal, for more detail around the management platform monitoring capabilities.

8.3 Troubleshooting

8.3.1 ClearLink Diagnostics (D_Port)

For SANs built with Brocade Gen 5, Gen 6, or Gen 7 Fibre Channel switches equipped with 16-Gb or higher optics, Brocade ClearLink Diagnostics enables predeployment testing to validate the integrity of the physical network infrastructure before operational deployment. Part of Brocade Fabric Vision technology, ClearLink is an offline diagnostics tool that allows users to perform an automated suite of tests to measure and validate maximum throughput speeds, latency, and distance across fabric links. ClearLink Diagnostics can also be used to verify the health and integrity of all 16G and 32G SFP+ transceivers and 32G QSFP transceivers in the fabric on a one-by-one basis. Diagnostics should be conducted before production or when an excessive number of CRC errors occur, potentially due to a physical layer issue.

A ClearLink Diagnostics port (D_Port) requires that only the individual ports attached to the tested link go offline, allowing the remainder of the ports to stay online in isolation from the link. ClearLink can also be used to test links to a new fabric switch without allowing the new switch to join or even be aware of the current fabric, providing an opportunity to measure and test ISLs before they are put into production. This fabric-based, physical-layer validation enables the following:

- Transceiver health check
- Transceiver uptime
- Local and long-distance measurements (5-meter granularity for 16G and 32G Small Form-factor Pluggable [SFP] or 32G Quad Small Form-factor Pluggable [QSFP] optics and 50-meter granularity for 10G SFP optics)
- Link latency measurements between D_Ports
- Link power (dB) loss
- Link performance

8.3.2 Recommendation: D_Port On-Demand

When an on-demand D_Port-capable switch or chassis comes online, the switch checks if the other end of the connection supports dynamic D_Port mode. If dynamic D_Port is supported on the opposite end, the switch changes the remote port to D_Port mode and starts a diagnostic test automatically. After successfully completing the test, the D_Port changes to normal port mode.

For Brocade ClearLink Diagnostics guidelines and restrictions, refer to the *Brocade Fabric OS Troubleshooting and Diagnostics User Guide* for a more detailed discussion of diagnostic port usage.

8.3.3 Forward Error Correction (FEC)

Forward error correction (FEC) is a data transmission error correcting method, which includes redundant error-sensing data. Error-correcting code ensures virtually error-free transmission. FEC supports the following data transmissions:

- When 10/16G FEC is enabled, it can correct one burst of up to 11-bit errors in every 2112-bit transmission, whether the error is in a frame or a primitive.
- When 32G FEC is enabled, it can correct up to 7 symbols in every 5280-bit transmission. A symbol consists of 10 bits, so there are 528 symbols in every 5280-bit transmission.
- When 64G FEC is enabled, it can correct up to 15 symbols in every 5440-bit transmission. A symbol consists of 10 bits, so there are 544 symbols in every 5440-bit transmission.

Because FEC is optional at 10G and 16G speeds, the Transmitter Training Signal (TTS) was extended to include a means to negotiate FEC capabilities. FEC is negotiated and activated when both sides of the link have FEC enabled. The FEC active indicator in Fabric OS indicates whether FEC was successfully negotiated. FEC uses unused bits within the signaling protocol to generate an error-correcting code (ECC) and correct bits as needed.

Refer to the *Brocade Fabric OS Administration Guide* for FEC configuration options and limitations.

8.3.4 Buffer Credit Loss Detection and Recovery

Enable both credit loss detection and recovery on your Brocade platforms since it is disabled by default. Buffer Credit Loss Detection and Recovery enables Brocade hardware to detect and automatically recover any lost credits that occur on backend ports with no user intervention.

BBC recovery allows links to recover after one or more buffer credits are lost. The recovery feature must be enabled. BBC recovery is supported on E_Ports, EX_Ports, and F_Ports. If a credit loss is detected, a recovery attempt initiates. BBC recovery is accomplished via a link reset in which performance is maintained and frame and BBC counters are reset.

Buffer credit recovery is enabled automatically across any long-distance connection for which the E_Port, EX_Port, or F_Port buffer credit recovery mechanism is supported.

8.3.5 RASLog Messages

RASLog messages report significant system events and information (failure, error, and critical conditions) and are used to show the status of high-level user-initiated actions. RASLog messages are forwarded to the console, configured syslog servers, and configured Simple Network Management Protocol (SNMP) traps or informs. SANnav Management Portal can be used as a RASLog receiver.

The following are the severity levels for messages and their descriptions:

- 1 = CRITICAL

  Critical-level messages indicate that the software has detected severe problems that will cause a partial or complete failure of a subsystem if not corrected immediately; for example, a power supply failure or a temperature rise must receive immediate attention.

- 2 = ERROR

  Error-level messages represent an error condition that does not impact overall system functionality significantly. For example, error-level messages might indicate timeouts on specific operations, failures of certain operations after retries, invalid parameters, or failure to perform a requested operation.

- 3 = WARNING

  Warning-level messages highlight a current operating condition that should be checked, or it may lead to a failure in the future. For example, a power supply failure in a redundant system relays a warning that the system is no longer operating in redundant mode unless the failed power supply is replaced or fixed.

- 4 = INFO

  Info-level messages report the current nonerror status of the system components, for example, detecting the online and offline status of a fabric port.

8.3.6 Audit Log Messages

Event auditing is designed to support post-event audits and problem determination based on high-frequency events of specific types such as security violations, zoning configuration changes, firmware downloads, and certain types of fabric events. Audit messages flagged as only AUDIT are not saved in the switch error logs. The switch can be configured to stream audit messages to the switch console and forward the messages to specified syslog servers. The audit log messages are not forwarded to an SNMP management station. There is no limit to the number of audit events.

For any given event, audit messages capture the following information:

- User Name: The name of the user who triggered the action.
- User Role: The access level of the user, such as root or admin.
- Event Name: The name of the event that occurred.
- Event Information: Information about the event.

8.4 Monitoring the Switches

You should seriously consider implementing some form of monitoring of each switch. Issues often start relatively benign and gradually degrade into more severe problems. Monitoring the logs for warning, critical, and error severity messages will go a long way in avoiding many problems.

- Plan a centralized collection of RASLogs and perhaps Audit logs via syslog. You can optionally filter these messages relatively easily through some simple scripting programs, or you can perform advanced correlation using an event management engine.
- Brocade platforms are capable of generating SNMP traps for most error conditions. Consider implementing some sort of alerting mechanism via SNMP or email notifications.

8.5 Latencies

Latency has many causes:

- Slow devices such as hosts and storage arrays
- Oversubscribed devices
- Long-distance links
- Servers that are not responding rapidly to previous I/O requests
- Degraded cables and failing SFPs due to I/O retries

Very little can be done in the fabric to accommodate end-device latencies, which are typically addressed through other means. Array latencies can be dealt with by array or LUN reconfiguration or data migration and tech-refresh. Long-distance problems might require more metro bandwidth or adjustment of the switches' distance settings. Applications might require tuning to improve performance. Failing fiber infrastructure and SFPs must be identified and replaced. At best, Brocade fabrics can help identify problem sources. Broadcom has worked diligently to enhance Fabric OS RAS features congruent with ever-changing customer requirements. Some of these features are described briefly in the following sections.

Exh. RA-3
Docket UT-181051
Page 38 of 112

8.6 Misbehaving Devices

All fabrics, regardless of equipment vendor, are vulnerable to the effects of misbehaving devices, that is, a server or storage device that, for whatever reason, stops functioning correctly. The effects of such behavior can be severe, causing other applications to fail intermittently, fail over, or stop altogether. The fabric can do nothing to anticipate such behavior. Broadcom has implemented several new features designed to rapidly detect misbehaving devices and isolate them from other devices in the fabric.

Isolating a single server has much less impact on applications than disabling a storage array port. Typically, a storage port services many applications, and the loss of that storage can severely impact all the applications connected to it. One of the advantages of a core-edge design is that it is straightforward to isolate servers from their storage and ensure that any action applied to a host port for a given behavior can be very different than the action applied to a storage port for the same behavior.

Detailed guidance on monitoring for misbehaving devices and configuring fabrics to respond to developing issues can be found in the *SAN Fabric Resiliency and Administration Best Practices User Guide*.

8.7 Design Guidelines

- **Transaction-based systems**: Ensure that ISL/UltraScale ICLs traversed by transaction-based systems accessing storage do not contain an excessive number of flows. The fan-in from initiators should not exceed a ratio of 10 to 1. Also, ensure that there is as little interference from other applications as possible so that latencies and congestion from other sources (called perpetrator flows) do not affect the overall performance of the applications (called victim flows).
- **I/O-intensive applications**: Bandwidth is the most common constraint for I/O-intensive applications. Modern fabrics typically provide more bandwidth than is needed except for the most powerful hosts. Ensure that these high-performing systems do not interfere with other applications, particularly if utilization spikes at specific times or if batch runs are scheduled. Add more paths (ISLs or trunks) through the fabric when in doubt.
- **Clusters**: Clusters often have behavioral side effects that must be considered, particularly true during storage provisioning. It is possible, for example, for a cluster to inundate the fabric and storage arrays with LUN status queries and other short-frame requests. This behavior can cause frame congestion in the fabric and stress the arrays' control processors. In the cluster, spread LUNs across as many arrays as possible.
- **Congestion**: Traffic congestion in some cases can be remedied by adding more ISLs or BT (Brocade Trunk) links. Of course, this assumes the congestion is between switches. In many cases, congestion occurs at a host or storage port; in which case, congestion must be addressed with the nodes experiencing the congestion. Brocade Gen 7 FPIN technology can be leveraged to mitigate end device congestion issues. Additionally, end devices should be configured with modern HBAs and drivers to optimize traffic across lossless Fibre Channel networks that use FPIN.
- **Misbehaving devices**: Little can be done in the fabric to mitigate the effects of a badly behaving device other than to remove it from the fabric. Brocade Fabric OS Port Fencing technology is designed to isolate rogue devices. Port Fencing works with MAPS to disable a port when a specific threshold is reached. In combination with FPI monitoring, Port Fencing can detect and isolate high-latency devices. High-latency devices frequently impact many other devices in the fabric.
- **Initiator and targets**: If possible, isolate host (initiator) and storage (target) ports onto separate switches for greater management and control over misbehaving and high-latency devices. The effect on an application is often less severe if a host port is disabled versus a storage port. A storage port services the flows from many servers.

Chapter 9: FC Routing

9.1 Overview and Purpose

A large SAN may have thousands of end devices, which could inundate or exceed fabric scalability, fabric services, convergence timeliness, and user manageability. FCR (FC Routing) constrains fabric services within the edge fabric(s) and the backbone fabric. Fabric services do not span entire fabrics with FCR, and without merging, end devices communicate across multiple edge fabrics. Fabric services are self-contained within each edge fabric or backbone, for example, the name server.

Limiting fabric services to within each edge fabric is done for various reasons:

- The overall SAN can scale to a much larger relative size than the maximum scalability of the fabric services within each edge fabric.
- Within an edge fabric, FCR reduces switch domains and managed zones.
- Edge-fabric disturbances and reconfigurations affect only the local fabric services, thereby providing fault isolation.
- FCR increases security because end devices cannot communicate outside of an edge fabric unless explicitly zoned.

9.2 Edge Fabrics

Edge fabrics are traditional fabrics, except they are connected to backbone EX_Ports. Edge fabrics contain end devices and may be connected to other edge fabrics via the backbone fabric. Backbones have EX_Ports facing edge fabric E_Ports, and edge fabrics have E_Ports facing the EX_Ports. There are no EX_Ports in an edge fabric.

Generally, edge fabrics follow the core-edge or collapsed-core architecture, the same as traditional fabrics. Unique to FCR is edge-fabric interconnectivity to a backbone.

9.3 Inter-Fabric Links

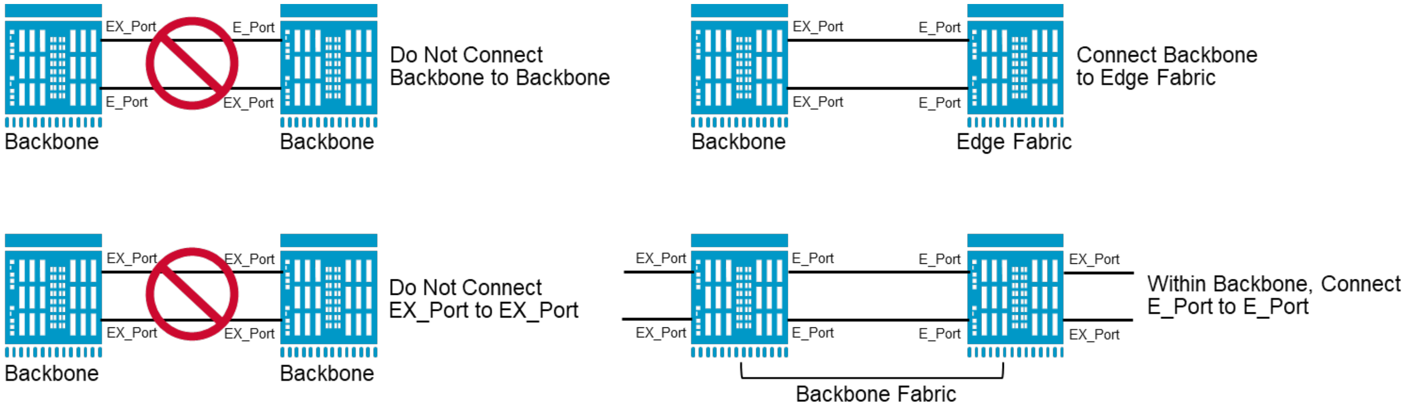
An inter-fabric link (IFL) connects an EX_Port to an E_Port. It is a type of inter-switch link (ISL) that spans from an edge fabric to a backbone.

Provision enough IFLs between each edge fabric and the backbone to accommodate the projected peak traffic load, plus planning for IFL outages or due to a bad optic or fiber.

9.4 Backbone Fabrics

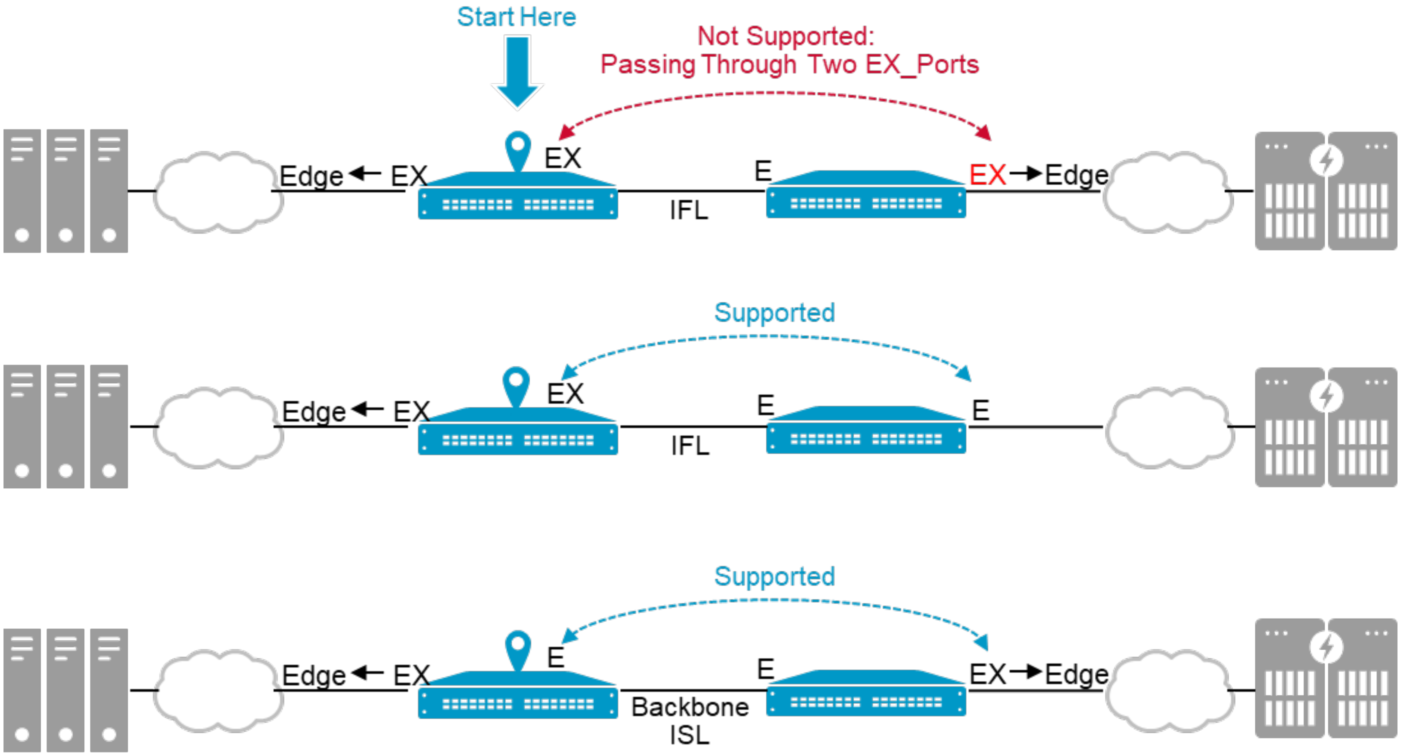
Backbone fabrics contain EX_Ports and are the demarcation for fabric services. A fabric may contain one backbone. A backbone can be a dedicated backbone with no connected end devices, or it may have connected end devices—both architectures are supported (Figure 10). Do not connect a backbone to a backbone, which is not supported; see Figure 10. Additionally, a backbone fabric may or may not contain Extension (FCIP) links; see Section 9.8, FCR and Extension.

Figure 10: Supported Backbone Architectures



EX_Ports are fabric-service demarcation points. Fabric services do not pass beyond an EX_Port. Two EX_Ports cannot be connected. EX_Ports connect only to E_Ports. Topology supportability is determined by starting within an FC router and moving toward the end device; traffic cannot pass more than one EX_Port along the path to the end device. If more than one EX_Port is passed, the architecture is unsupported. (See Figure 11.)

Figure 11: Supported FCR Architectures



There are many factors to consider when designing backbone fabrics. Backbone fabrics vary based on size, requirements for redundancy, and distance between edge fabrics. Generally, SAN architecture recommendations apply equally to backbone fabrics. There should be redundant fabrics, and each fabric should have redundant paths to every edge fabric. Consider the following factors when identifying the best switch platforms and backbone topology, including interconnections. The number of edge fabrics can impact the backbone topology and how they attach. Brocade FCR can be enabled on standard FC ports; a license may be required in some cases.

Composition of edge fabrics:

- **Scale and interoperability**: Ensure that director and switch platforms can support the scale and interoperability needed.
- **Legacy SAN platforms**: Anywhere in the SAN, earlier directors/switches or firmware may impact supported features, manageability, and interoperability.
- **Advanced SAN applications and features**: Some advanced SAN applications and features may not be compatible with FCR or a particular platform type.

Projected inter-fabric traffic patterns:

- **Quantity (bandwidth utilization)**: Provision enough ISLs within the backbone to accommodate projected peak traffic loads that traverse the backbone.
- **Bursty versus peak traffic**: Bursty traffic is a sudden spike that rapidly dissipates. It is not the same as peak traffic, which may not be bursty. Infrequent, bursty traffic can be forgiving. Traffic bursts may cause temporary response time increases due to congestion. Buffer credits may be withheld until a burst subsides. Such congestion is less likely with traffic patterns of a continuous nature.
- **Small versus large frame size**: Fibre Channel is a high-speed, low-latency protocol. It relies on buffer-to-buffer credit (BBC) flow control. This mechanism is a fundamental part of FC and provides lossless data communications. A sequence of small frames uses the same number of BBCs as a series of large frames. On the other hand, large frames use more bandwidth. In other words, a large amount of small-frame traffic can fully utilize available buffers while consuming only a minimal amount of bandwidth. Therefore, consider not only bandwidth but also the typical frame size. For instance, FC compression creates primarily smaller FC frames. If the bulk of frames is expected to be smaller, additional buffers should be allocated to the paths handling those I/O patterns. Pay extra attention to this type of congestion because congested backbones adversely impact the performance of all connected edge fabrics. When in doubt, overprovision IFLs.
- **Distance (location of fabrics)**: Long-distance IFLs require adequate bandwidth and BBCs to prevent data transmission congestion and droop, respectively. Consider all potential traffic flows that may traverse the long-distance links. Long-distance links have more latency, simple physics time = distance/rate. Therefore, overprovisioned long-distance links may prevent oversubscription such that unexpected bursts do not adversely impact flows.
- **Virtual Fabrics (VF)**: All EX_Ports must reside in the base switch. The base switch does not support ISL R_RDY mode. If a logical switch has XISL enabled, you cannot connect an EX_Port to that logical switch. The base switch is similar to a backbone switch, and a base fabric is like a backbone fabric. All switches in a backbone fabric must have the same backbone fabric ID, which must be unique relative to any edge fabric.

Potential growth:

- **Number of fabrics**: If the number of fabrics is likely to increase, then deploy backbone fabrics to readily accommodate additional edge fabrics and additional traffic loads.
- **Size of fabrics**: If the size of edge fabrics is likely to grow, and the inter-fabric traffic is expected to grow accordingly, provision additional IFLs and ISLs such that the capacity of available paths stays well ahead of current usage. That way, incremental growth on the edge can be accommodated without immediately upgrading the backbone.
- **Amount of traffic between fabrics**: If the inter-fabric traffic is expected to grow even without growth in the individual edge fabrics, then provision additional IFLs and ISLs such that the capacity of available paths stays ahead of current usage. That way, incremental increases in data flow across the backbone can be accommodated without immediately upgrading the backbone. Make sure that you allow for plenty of room for backbone expansion.

NOTE: Refer to the *Brocade SAN Scalability Guidelines* for FCR scalability limits.

Consider using FCR under the following conditions:

- There are requirements for added scalability.
- There are benefits to compartmentalizing manageability.
- Enhanced security is required.
- There is a limited number of initiator-target pairs shared between edge fabrics.
- There is a limited number of LUNs shared between edge fabrics.
- Archiving devices, such as tape libraries, must be shared.

The implementation and configuration of inter-fabric links (IFLs in the case of FCR) should be based on the expected data volume between the backbone and edge fabrics and the desired level of redundancy. Some architectural examples of FCR topologies follow.

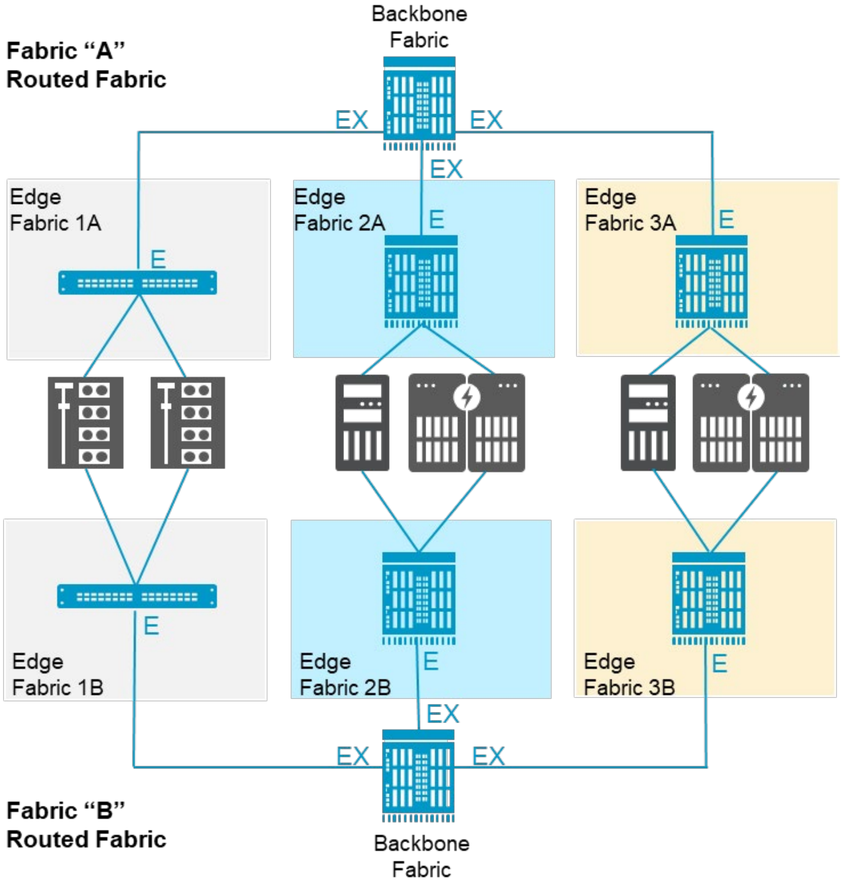
Except in the case of tape, which often has only a single pathway, there should always be A and B fabrics, each with IFL redundancy. A routed fabric environment consists of one or more edge fabrics interconnected by one or more backbone fabrics. Multiple backbone fabrics are parallel and belong to only the A or B fabric, not both. A backbone fabric can be a single switch, multiple switches, or a core-edge topology. These topologies are valid for the edge fabrics as well.

In Figure 12, the architecture consists of three edge fabrics and a backbone fabric. A and B fabrics are shown. The “A” backbone connects to each edge fabric via EX_Ports. EX_Ports in the backbone connect to E_Ports in the edge fabric to form IFLs. Each backbone must have a unique backbone fabric ID (BBFID), and all switches within that backbone must have that same BBFID. The default is 128, and when a single backbone is deployed, as in Figure 12, no BBFID needs to be configured because the default will suffice. An alias can be assigned to BBFIDs.

Each edge fabric must have a unique edge-fabric ID (EFID), and all EX_Port connections to that edge fabric must use that EFID. Each EX_Port is configured with the corresponding EFID belonging to the edge fabric that it connects. E_Ports are not configured with any additional parameters when connecting to EX_Ports.

A collapsed-core backbone is a relatively straightforward FCR architecture.

Figure 12: Routed A and B Fabric Collapsed-Core Architecture



In Figure 13, a separate backbone fabric is not deployed. Instead, the middle fabric is assigned as the backbone, and end devices connect directly to the backbone. There are three fabrics, each with its own self-contained fabric services.

Not having a separate backbone fabric limits the topology from being an interconnected full-mesh. There are only two connections coming out from the center edge fabric, and there is no connection between the left and right edge fabrics. Such a design violates the previously mentioned supported FCR architectures by creating a situation in which more than one EX_Port may be traversed from inside an FC router to the ultimate destination device.

This design may be used when the cost of an additional fabric for the backbone is prohibitive.

Figure 13: Common-Backbone, Dual Collapsed-Core Architecture

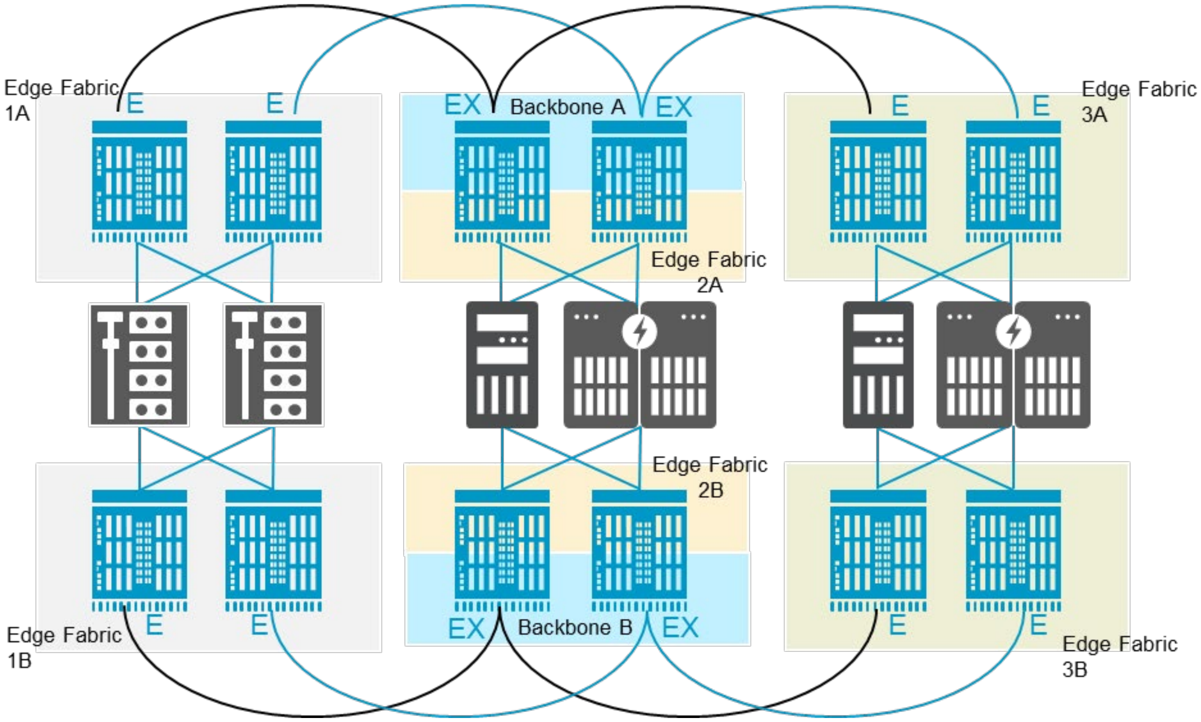
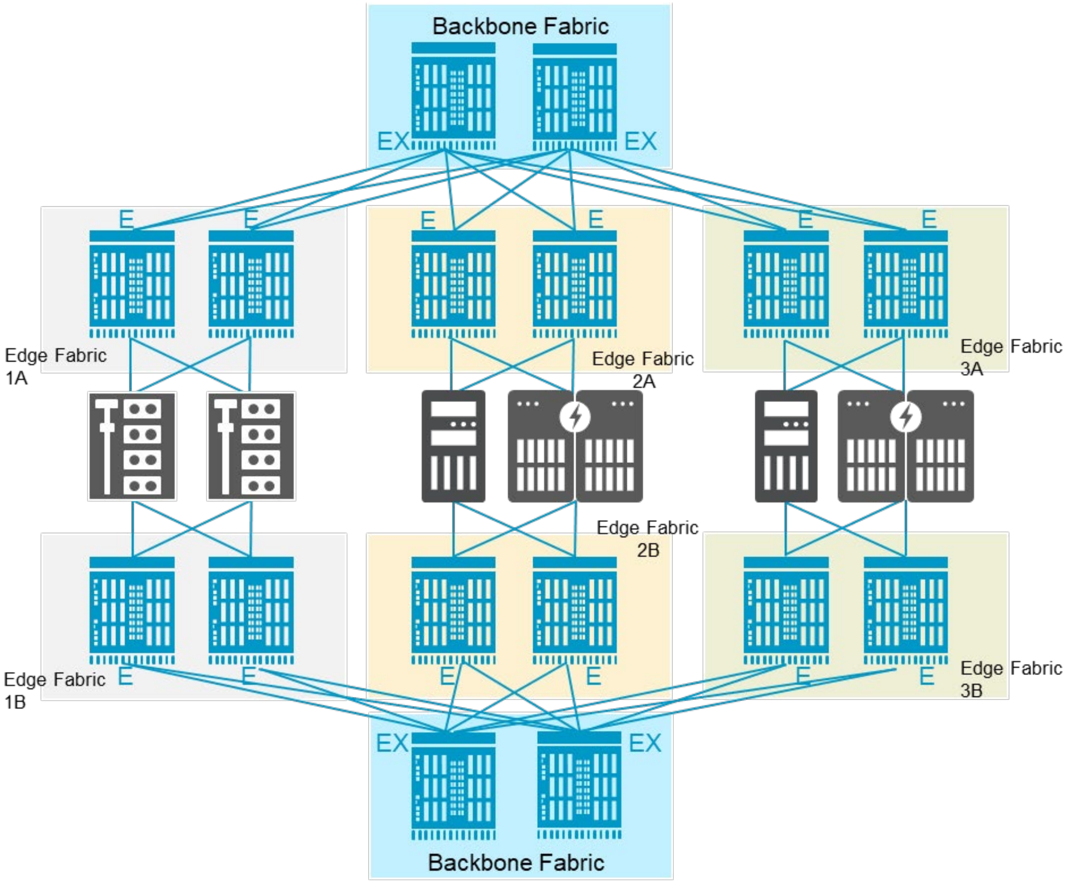


Figure 14 shows a routed SAN with A and B fabrics, each having a dual-core backbone and a unique BBFID. The EX_Ports are exclusively in the backbone, and fabric services do not pass beyond the EX_Ports. There are three edge fabrics, each with its own EFID. There are multiple IFLs to each edge fabric. The dual-core backbone architecture is highly redundant, resilient, and scalable for critical enterprise applications demanding zero downtime. Considering the dual-core backbone's scalability, it is relatively easy to manage operationally.

Figure 14: Dual-Core Backbone Routed Fabric



9.5 Redundancy

FCR SAN redundancy is achieved by:

- Using best practices within the edge fabrics (core-edge or collapsed-core architectures).
- Using best practices within the backbone fabric(s) (core-edge or collapsed-core architectures).
- Deploying dual backbone fabrics for each fabric (A and B). The need for redundancy versus cost and operations must be considered. Ask yourself what the purpose of the routed SAN is? What happens if routing between edge fabrics goes offline, yet the edge fabrics themselves remain online?
- Parallel IFLs between the backbone and edge fabrics, including ports, optics, and cable redundancy.

9.6 Avoiding Congestion

A routed fabric must be evaluated for bandwidth and potential utilization between all endpoints as with any traditional flat fabric, which means calculating the traffic flowing in and out of every edge fabric, and providing enough bandwidth into and across the backbone to accommodate traffic. The same ISL guidelines and best practices connect edge fabrics via IFLs for improved utilization and resiliency. A higher performance edge fabric versus an underperforming backbone may result in an oversubscribed backbone, which during peak loads can lead to congestion, higher latency, and longer storage response times. If the edge fabric has 64G FC ISLs, the backbone fabric must have 64G FC ISLs as well. Before upgrading an edge fabric, upgrade the backbone to avoid congestion and oversubscription issues.

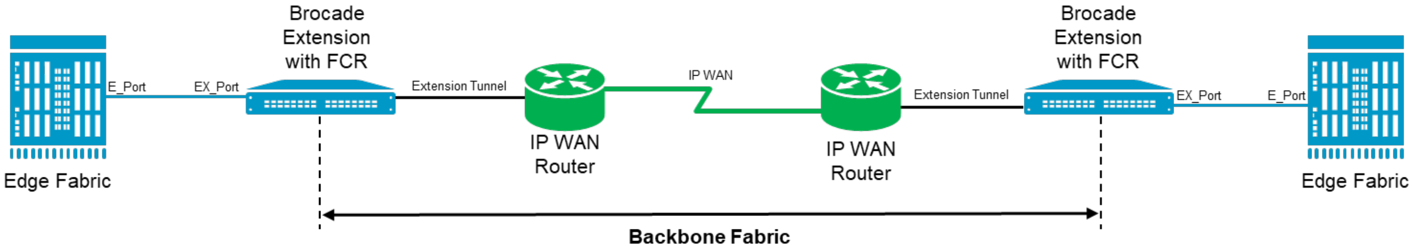
9.7 Available Paths

An optimal approach is to have multiple Brocade trunking paths between edge fabrics to spread traffic across available resources. Never attach both A and B fabrics to the same backbone device. Connecting A and B edge fabrics to the same backbone device destroys the air gap between A and B and is *not* considered a redundant architecture and best practice. From the perspective of FC, you should adhere to the concept of an “air gap” from host to storage. A common device connected to both A and B fabrics can cause a SAN-wide outage. If an air gap is implemented, faults on one fabric cannot affect the other fabric. These faults can manifest from defects in hosts, drivers, the fabric operating system, the fabric hardware, the storage hardware, the storage software, and human error. It is not relevant that FCR keeps fabric services separate because faults within one large routed fabric can transcend FCR, causing the entire SAN to fail.

9.8 FCR and Extension

FCR can be used within a single data center or across campus data centers and between edge fabrics connected by FCIP over a metropolitan area network (MAN) or wide-area network (WAN), as shown in Figure 15. A Brocade Extension tunnel is an ISL (VE_Port to VE_Port). A VE_Port is an E_Port that is an endpoint of an Extension tunnel. Each Extension platform becomes part of the backbone fabric. EX_Ports on the Extension platforms connect to the edge fabrics via one or more IFLs.

Figure 15: Fibre Channel Routed Fabric over Extension



More information about Extension can be found in Chapter 12.

9.9 FCR Design Guidelines and Constraints

The following are some of the key metrics and best practices for routed SAN topologies:

- Keep A and B fabrics separated from host HBA to storage ports from an FC perspective, referred to as an *air gap.* Air gaps do not include FCIP because, in an IP network, Ethernet switches and IP routers do not merge as FC fabrics do. Extension VE_Ports should never connect fabric A to fabric B, which is the same as cross-connecting a traditional ISL, resulting in the connection of fabric A to fabric B.
- Localize traffic within an edge fabric to the greatest extent possible.
- Have a predefined schema for assigning domains within the SAN. For example, edges, cores, switches, EFIDs, translate domains, and BBFIDs should be within specific ranges to avoid domain overlap.
- Consider upgrading backbone fabrics before upgrading edge fabrics to avoid oversubscription and congestion.
- Have no more than one long-distance ISL or Extension between source and destination during normal operations. An additional hop may be used for high availability during an outage. For example, in a triangle architecture in which the primary link has gone down, the remaining two legs of the triangle can be used as a backup path. However, latency and response times will likely be longer.
- Long-distance links are within the backbone and not between an edge fabric and the backbone. Edge fabrics will be isolated from disruption because fabric services are not extended beyond the EX_Port. Most often, long-distance links are the primary cause of instability.
- Logical SAN (LSAN) zones are only for end devices that communicate from edge fabric to edge fabric across a backbone. In other words, do not make zones within edge-fabric LSAN zones.
- Redundant backbones for each fabric improve resiliency, meaning full redundancy; therefore, fabric A would be fully resilient if one of its backbones were to fail, as would fabric B. Both fabric A backbones would have to fail before relying solely on fabric B to maintain operations.

Different redundant backbone fabrics that share connections to the same edge fabric must have unique BBFIDs. Refer to the case where there are redundant fabric A backbones and redundant fabric B backbones. There are no cross-connections between A and B fabrics, nor are there cross-connections between the parallel backbones within fabric A or B.

Chapter 10: Virtual Fabrics (VF)

Virtual Fabrics (VF) is an architecture to virtualize hardware boundaries within a SAN platform. Traditionally, SAN design and management are done at the granularity of a physical switch. Virtual Fabrics allows SAN design and management to be done at the granularity of a port.

Virtual Fabrics is a suite of related features to customize logical fabrics based on requirements. Virtual Fabrics consists of the following specific features:

- Logical switches
- Logical fabrics
- Device sharing

Hardware-level fabric isolation is accomplished through the concept of a logical switch, which provides the ability to partition physical switch ports into one or more “logical” switches. Logical switches are then connected to form logical fabrics. As the number of available ports on a switch continues to grow, partitioning switches gives storage administrators the ability to take advantage of high-port-count switches by dividing physical switches into different logical switches. Without VF, an FC switch is limited to 512 ports. A storage administrator can then connect logical switches through various ISLs to create one or more logical fabrics.

There are three ways to connect logical switches: a traditional ISL, an IFL (EX_Port used by FCR), or an extended ISL (xISL). An ISL can be used only for regular L2 traffic between the connected logical switches, carrying only data traffic within the logical fabric of which the ISL is a member. One advantage of Virtual Fabrics is that logical switches can share a common physical connection, and each logical switch does not require a dedicated ISL. For multiple logical switches, in multiple logical fabrics, to share an ISL, Virtual Fabrics supports an xISL connection, which is a physical connection between two base switches. Base switches are a type of logical switch specifically intended for intra-fabric and inter-fabric communication. Base switches are connected together via xISLs and form a base fabric.

Once a base fabric is formed, VF determines all physically associated logical switches and logical fabrics via the base fabric and the best route between each. For each local logical switch, a logical ISL (LISL) is created for every destination logical switch reachable via the base fabric. Thus, an xISL is a physical link between base switches, carrying the virtual connections. In addition to xISLs, a base fabric also supports EX_Ports for communication between logical fabrics. An FCR (FC Routing) link between an EX_Port and an E_Port is called an IFL (inter-fabric link). Base switches interoperate with FC routing via an EX_Port in the base fabric or EX_Ports in a separate backbone fabric.

10.1 Use Case: FICON and Open Systems (Intermix)

Virtual Fabrics enables customers to share FICON and FCP (SCSI and NVMe) traffic on the same physical platform. As chassis densities increase, this is a viable option for improved hardware utilization while maintaining director-class availability. The primary reasons for moving to an Intermix environment are the following:

- Array-to-array RDR of FICON volumes (Most array replication uses FCP for FICON volumes.)
- ESCON-FICON migration
- Sharing of infrastructure in a nonproduction environment
- Reduced TCO
- Growth of zLinux on the mainframe

From a SAN design perspective, consider the following guidelines when considering FICON Intermix:

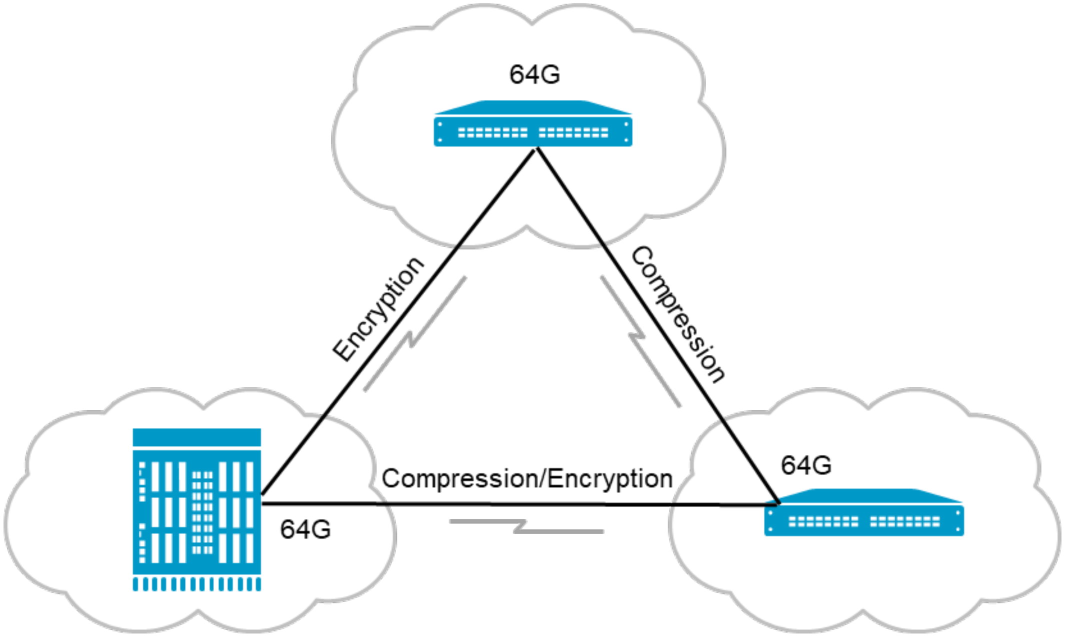
- Connect devices across port blades (connectivity from the same device should be spread over multiple blades).
- A one-hop-count architecture applies; however, there are “Hops of No Concern” in some cases. Refer to the *Brocade FICON/FCP Intermix Best Practices Guide* for details.

Chapter 11: Fibre Channel Intelligent Services

11.1 In-Flight Encryption and Compression

Brocade Gen 6 and Gen 7 Fibre Channel platforms support both in-flight compression and encryption at a port level for both local and long-distance ISL links (see Figure 16). In-flight data compression is a valuable tool for saving money when bandwidth caps or bandwidth usage charges encumber transferring data between fabrics. Similarly, in-flight encryption enables a further layer of security with no key management overhead when transferring data between local and long-distance data centers besides the initial setup.

Figure 16: Latency for Encryption and Compression



As the frame is processed, enabling in-flight ISL data compression or encryption increases ASIC latency. At each stage (including encryption, compression, and local switching), the approximate latency is 6.2 microseconds (see Figure 16).

11.1.1 Virtual Fabric Considerations: Encryption and Compression

E_Ports in logical switches, base switches, or default switches can support encryption and compression. Both encryption and compression are supported on xISL ports but not on LISL ports. If encryption or compression is enabled and ports are moved from one LS to another LS, it must be disabled before moving to another LS.

11.1.2 Guidelines: In-Flight Encryption and Compression

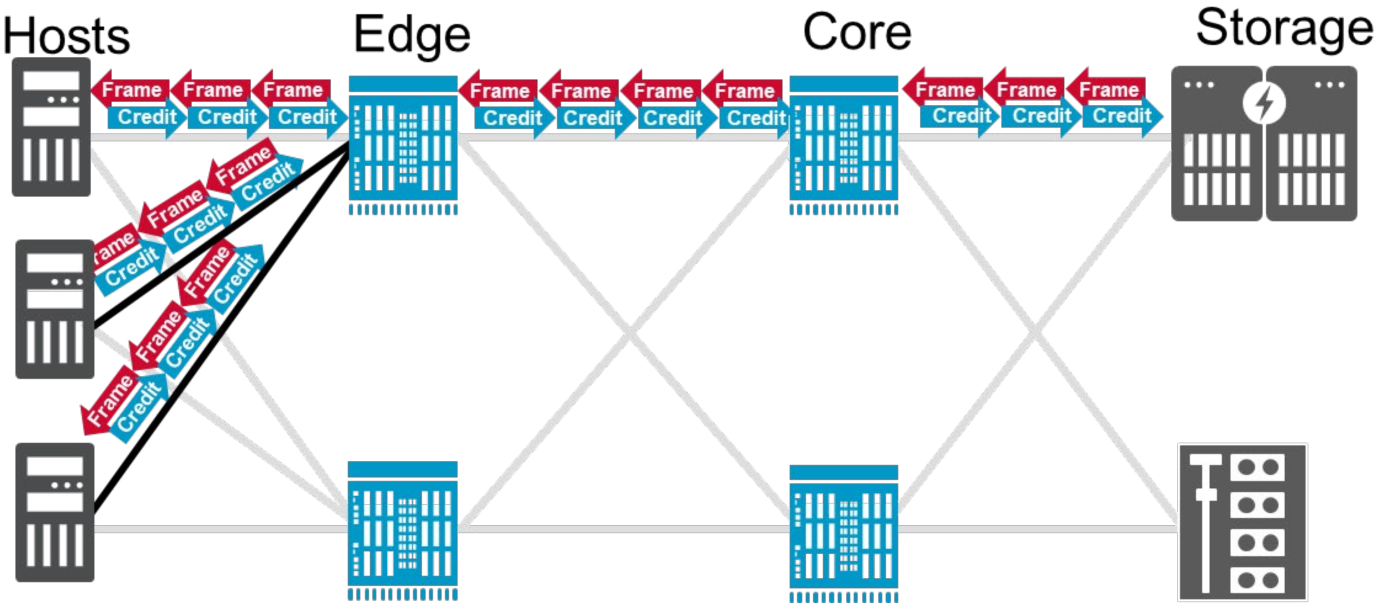
Refer to the *Brocade Fabric OS Administration Guide* for the latest information.

11.2 Fabric Notifications

Fibre Channel networks can be elusive to troubleshoot because flows are difficult to visualize, and the affected devices are not likely to correspond with the problem cause. Fibre Channel uses a credit-based flow-control mechanism (see Figure 17), with inherent congestion characteristics due to head of line blocking. Broadcom introduces a hardware, software, and management solution for achieving congestion reduction and elimination called Fabric Notifications.

Collecting transport characteristic data from various sources, evaluating it, and disseminating it to interested devices allows for faster and sometimes automatic problem resolution. End devices can employ primary response and recovery mechanisms. Fabric information is helpful for end devices, and end devices have helpful information for the fabric and peer end devices. Fabric Notifications play a crucial role in collecting and disseminating information among interested and related devices.

Figure 17: Freely Moving Lossless Credit-Based Flow-Control FC Network



Fabric Notifications addresses four issues: congestion (oversubscription and credit stall), link integrity, and SCSI command delivery failure.

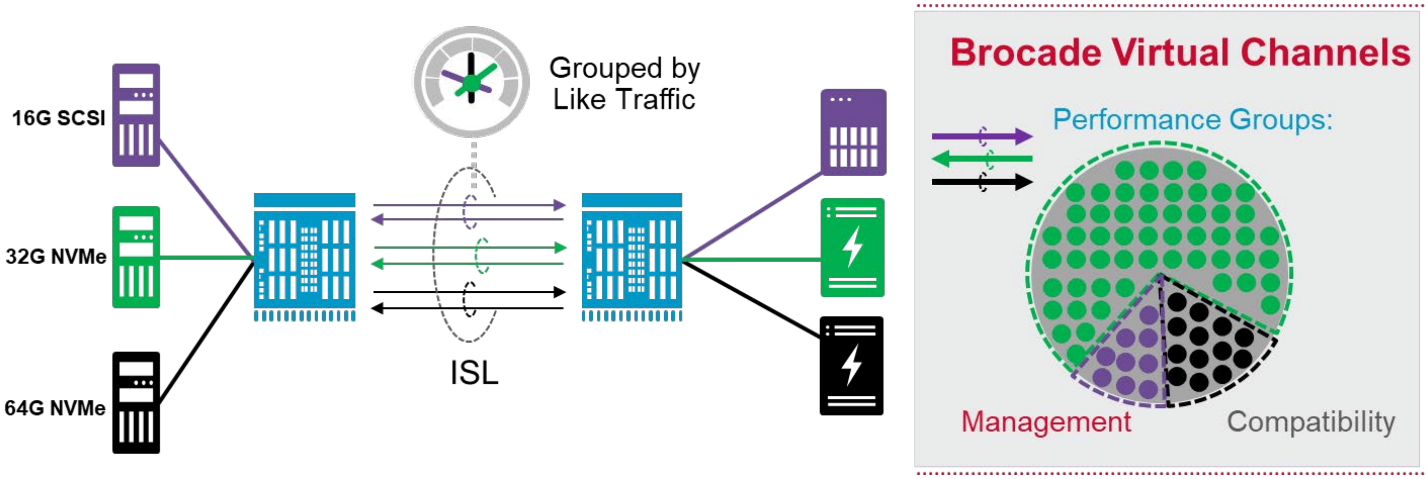
11.3 Traffic Optimizer

For years, Brocade virtual channels (VCs) worked perfectly well at 1, 2, 4, 8, and 16-gigabit rates by leveraging multiple logically independent paths. To a degree, virtual channels mitigated interference between slower flows impeding faster flows. Optionally, critical faster flows could be manually assigned to a high QoS VC, and slower flows could be manually assigned to a low QoS VC to prevent such interference.

Technology evolves, and Broadcom has optimized VC efficiency by enhancing effectiveness to targeted flow characteristics. Demands on an enterprise SAN have never been more significant due to other complimentary technology evolutions such as NVMe, AFA, and host virtualization. Flows compete for resources, and head of line blocking is not an option and must be efficaciously dealt with, which is where Traffic Optimizer helps.

Brocade Traffic Optimizer technology takes VCs to the next level (see Figure 18). Traffic Optimizer organizes and manages traffic flows using performance groups (PGs), and fabric resources are allocated based on performance groups. Flows are assigned to a VC based on the destination port speed and protocol (SCSI or NVMe). Brocade fabrics know the destination port speed and protocol for every flow.

Figure 18: Host to Storage via Traffic Optimizer VCs



Brocade Gen 7 hardware added more VCs. E_Ports and EX_Ports use 16 new Traffic Optimization VCs, which are assigned as follows: Four VCs per speed (<16, 32, 64) for a total of 12 VCs, plus 4 VCs dedicated to NVMe. Benefiting from an NVMe storage investment requires NVMe dedicated resources in the SAN.

QoS VCs have not changed; there are 2 for low, 4 for medium (default), and 5 for high. QoS VCs are used when traffic has been designated to a high or low priority. Slow Drain Device Quarantine (SDDQ) uses the QoS low VCs.

Chapter 12: Extension

12.1 Extension Common Principles

12.1.1 Ethernet RJ-45, SFP, SFP+

Depending on the Extension platform, a variety of Ethernet interfaces are available. The Brocade 7840 Extension Switch and SX6 Extension Blade have two 40GE interfaces and sixteen GE/10GE interfaces. The Brocade 7810 Extension Switch has two RJ-45 interfaces and six GE/10GE interfaces; not all can be used simultaneously. Indeed, the interfaces selected must be capable of carrying the bandwidth required by the storage application and configured for the circuit or circuits passing through. For example, you cannot put a 2Gb/s circuit on a GE interface.

Additionally, 7840 and SX6 GE/10GE interfaces are in groups, and the speed must be consistent within a group. The following is a list of ports that block each other when their speed is not set the same. The best practice is to use 2, 3, 4, 5, 10, 11, 12, and 13 first and then move on to doubling up within a port group.

Please check the *Brocade Fabric OS Extension User Guide* for additional information.

Ethernet interfaces in the same group:

- 2 & 6
- 3 & 7
- 4 & 8
- 5 & 9
- 10 & 14
- 11 & 15
- 12 & 16
- 13 & 17

Remember that the Ethernet interfaces do not perform speed negotiation between 1GbE and 10GbE. The speed is user-configurable only. An SFP (1GbE) and an SFP+ (10GbE) are different optics, and neither can change to the speed of the other. If you need 1Gb/s connections, you must order 1GbE optics; likewise, if you need 10Gb/s connections, you must order 10GbE optics.

The Brocade 7810 has GbE and 10GbE interfaces. Two interfaces (GE0 and GE1) are RJ-45 copper ports (1Gb/s only). Copper is enabled by default. These copper ports are mutually exclusive with optical ports GE0 and GE1. Either copper is enabled, or the optics are enabled, not both. There is no disadvantage to using one set of interfaces over another.

12.1.2 Brocade Extension Trunking (BET)

Brocade Extension Trunking (BET) is an exclusive Broadcom feature that offers the following benefits:

- In-order delivery
- Remediation of data lost in flight
- Bandwidth aggregation
- Granular load balancing
- Lossless failover/failback

A tunnel that contains more than one circuit is a Brocade Extension Trunk. A VE_Port defines a tunnel endpoint. BET forms a single ISL between two VE_Ports. BET circuits terminate at the VE_Port on each side; therefore, only a single tunnel is load-balancing across the circuits. A circuit is a connection that is defined by a source and destination IP address and other configuration parameters such as QoS, min/max rate limits, and KATOV.

NOTE: Compression and IPsec are at the tunnel level, not the individual circuit level.

A circuit is assigned to an Ethernet interface via the IP address (ipif). Often, each circuit is assigned to its own dedicated Ethernet interface; however, this depends on the interface's speed, the cumulative max rates of the circuits, and the port redundancy required. Ethernet interfaces physically connect to a DC LAN switch or WAN router. Connect one interface to switch "A" and the other interface to switch "B." Circuits load-balance with a high degree of granularity. BET will prevent transmission data loss when a network device fails or goes offline or when optic instability, cable damage, a human error, or a service-provider disruption occurs.

Circuits may have varying characteristics; for instance, circuits may experience different latency, take different paths (like DC LAN switches/WAN routers), and belong to different service providers. Circuits that belong to a VE_Port can have bandwidth differences up to 4x the lowest bandwidth circuit. For example, if the cir0 min rate is 1Gb/s, the min rate on cir1 can be no more than 4Gb/s.

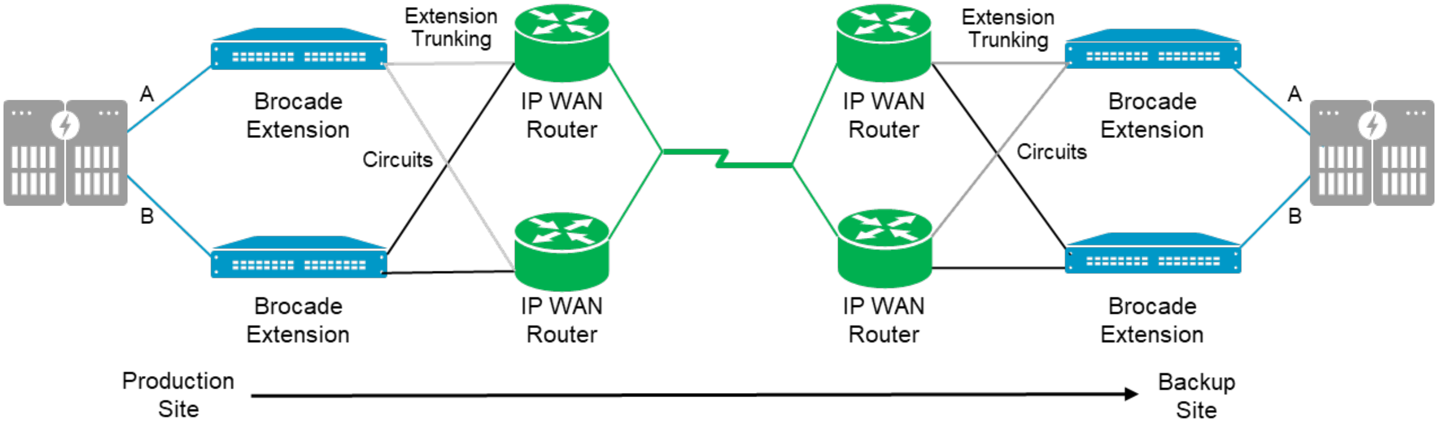
Using BET is considered best practice versus using multiple VE_Ports. Multiple VE_Ports between the same two domains will use one of the following methods to route traffic across the VE_Ports:

- EBR (Exchange-Based Routing: default)
- DBR (Device-Based Routing)
- PBR (Port-Based Routing)

The architecture shown in Figure 19 is a dedicated high-availability replication SAN with no connections to the production SAN. Array replication ports are nearly always dedicated; therefore, these ports should not connect to the production fabric.

About BET, there is an extension trunk for the "A" path (grey - top to top Extension boxes) and an extension trunk for the "B" path (black - bottom to bottom Extension boxes). The IP network merely connects these point-to-point circuits to each associated endpoint.

Figure 19: Four-Box High-Availability Replication SAN



All data centers have redundant routers/switches. Connecting each Brocade Extension platform to each router/switch for redundancy and resiliency is best practice. Without BET, this design would require two VE_Ports per Extension box.

Available VE_Ports are not the issue; performance, resiliency, and redundancy are paramount. It is best to use one VE_Port with BET to form a trunk that consists of two circuits between the domains.

Typically, tape can take only a single SAN path, but there are exceptions. Brocade Extension Trunking (BET) is logically a single path. Take advantage of BET to implement redundant network paths with each circuit. Tape must transparently fail over/fail back paths without data loss while maintaining in-order delivery; otherwise, tape jobs fail.

A single tunnel between domains is required for disk and tape protocol optimization (FastWrite for disk and Open Systems Tape Pipelining for tape). Use multiple circuits for redundancy, resiliency, and failover protection. Virtual Fabrics Logical Switches (VF LS) must be configured, ensuring that all sequences from every exchange traverse the same VE_Port in both directions if multiple tunnels are required. There can be only one VE_Port per LS.

12.1.3 IPsec

As a best practice, use Brocade IPsec to protect data end-to-end. When Brocade Extension is implemented, it is always prudent to enable IPsec. All data that leaves the secure confines of a data center into a vulnerable infrastructure guarantees no security, and no service provider guarantees data security in flight. Links must be authenticated, and data must be encrypted to prevent attacks and eavesdropping. Brocade IPsec is easy and practical to deploy. Would your company operate WiFi with no encryption? Of course not!

Brocade IPsec is a hardware implementation that can operate at line rate. IPsec is included in base Extension platforms, and there are no additional licenses or costs for encryption. Encryption adds a minor propagation delay, approximately 5 µs. Pre-shared key (PSK) and CA and self-signed certificates are configurable.

Brocade IPsec is Suite B and CNSA compliant and implements the latest encryption technologies, such as AES 256, SHA-512 HMAC, IKEv2, and Diffie-Hellman. Rekeying occurs in the background after approximately 2 billion frames or every 4 hours, and the process is nondisruptive.

Firewalls are *not* considered best practice for the following reasons:

- Brocade IPsec can operate up to 20Gb/s per data processor (DP), which is WAN-side full line rate, and 40Gb/s if implementing both DPs. Most firewalls cannot meet this throughput requirement.
- Broadcom requires minimal propagation delay because it is storage traffic.
- Brocade IPsec encrypts data closer to the source and destination of the data, which is considered best practice.
- Firewalls, WAN optimizers, or any network device using TCP proxy sessions resulting in emergent remote side TCP segments not identical to the original entrant TCP segments are not supported.

12.1.4 Adaptive Rate Limiting

Adaptive Rate Limiting (ARL) is an integral part of most Extension designs. When more than one circuit feeds into the same WAN link or when the WAN is shared with other traffic, ARL is an essential component. ARL is used to manage data sent into the IP network based on the minimum and maximum rate settings and the available WAN bandwidth between that range. There may be a single WAN connection or multiple WAN connections—the cumulative bandwidth of the circuits assigned to a particular WAN link matters. Multiple WAN connections are evaluated independently.

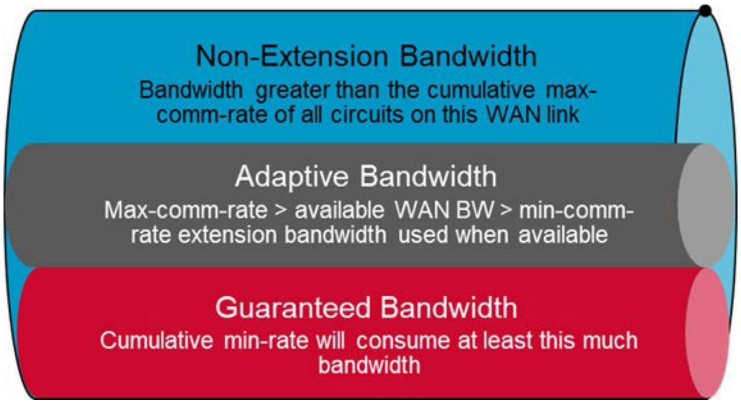
See Figure 20 and assume a 1Gb/s WAN. The ARL max-comm-rate is set to either the GE interface line rate or the maximum available WAN bandwidth, whichever is lowest. In this case, the max-comm-rate is set to 1Gb/s.

Each circuit is configured with a floor (min-comm-rate) and ceiling (max-comm-rate) bandwidth value in kbps (kilobits per second). Assume adequate bandwidth from the storage platform; the minimum circuit bandwidth will never be less than the min-comm-rate and will never be more than the max-comm-rate. The available bandwidth to the circuit will adjust automatically between the min and max based on IP network conditions. A congestion event causes the rate limit to readjust toward the minimum. An absence of congestion events causes it to rise toward the maximum. If the current rate is not at the maximum, ARL will periodically attempt to adjust upward; however, if another congestion event is detected, the rate will remain stable.

The ARL min-comm-rate is set to the max-comm-rate divided by the number of circuits feeding the WAN. In this example, 1Gb/s ÷ 4 = 250Mb/s. The min-comm-rate is set to 250Mb/s. When all the circuits are up, each will run at 250Mb/s. In an extreme case in which three circuits have gone offline, the remaining circuit will run at 1Gb/s. At 1Gb/s, all the WAN bandwidth continues to be consumed, and the replication application remains satisfied.

When more than one circuit feeds a WAN link, the two circuits equalize and utilize the available bandwidth, as shown in Figure 20. When an interface or the entire platform goes offline, ARL will re-adjust to utilize bandwidth that is no longer being used by the offline circuits, which maintains the utilization of WAN bandwidth during periods of maintenance or failures.

Figure 20: ARL Min and Max Comm Rates


Non-Extension Bandwidth
Bandwidth greater than the cumulative max-comm-rate of all circuits on this WAN link
Adaptive Bandwidth
Max-comm-rate > available WAN BW > min-comm-rate extension bandwidth used when available
Guaranteed Bandwidth
Cumulative min-rate will consume at least this much bandwidth

In a shared WAN, consider bandwidth as separated into three distinct areas:

- Guaranteed Bandwidth (0 → min)
- Adaptive Bandwidth (min → max)
- Non-Extension Bandwidth (max → WAN bandwidth)

Referring to the graphic above, ARL manages the bandwidth in each category. Red is the minimum bandwidth utilized by Extension, the amount of bandwidth reserved exclusively for Extension. It is important to note that the minimum bandwidth used is the aggregate of *all* circuit minimums on the WAN. Blue is reserved exclusively for other traffic sharing the WAN. The bandwidth where the blue section starts is the aggregate of *all* maximum circuit values. Gray is the area between the top of the red section and the bottom of the blue section. Extension circuits may use this bandwidth when it is available, meaning other applications sharing the WAN are not currently using it. There are many ways in which ARL can be used, and this was just one example.

12.2 WAN Side

Brocade Extension conceptually has three sides: FC/FICON, WAN, and LAN. This section addresses the WAN side, which faces the WAN and forms tunnels and circuits that traverse the WAN. There are only VE_Ports and non-LAN Ethernet interfaces on the WAN side.

12.2.1 LLDP

A best practice is to use Link Layer Discovery Protocol (LLDP) as an indicator of Ethernet connectivity. LLDP is used by both the Extension platform administrator and the network administrator. LLDP information is logged on both ends. Network administrators see what is connected to their ports. Storage administrators see what is connected to their Ethernet ports. For troubleshooting and add/move/changes, LLDP is a convenient tool.

LLDP is enabled by default and operates on all Ethernet interfaces except the mgmt interface. Most DC LAN Ethernet switches support LLDP, and by default, it is enabled.

There are specific TLV (type, length, value) that should be enabled and disabled when connecting Extension to a DC LAN switch:

- Enable
  - Chassis ID
  - System Name
  - System Capabilities
  - System Description
  - Port ID
  - Port Description
  - Management Address
- Disable
  - dcbx
  - fcoe-app
  - fcoe-lls
  - dot1
  - dot3

12.3 FCIP

12.3.1 Compression

Use compression with Remote Data Replication (RDR) applications, including RDR/S. Commonly, tape traffic is compressed, and compressing data again has no benefit.

NOTE: Broadcom makes no guarantees, warranties, or claims as to the actual compression ratio that will be achieved with customer-specific data.

There are three modes of compression besides disabled: Fast-Deflate, Deflate, and Aggressive-Deflate.

12.3.1.1 Fast-Deflate

Fast-Deflate typically gets about a 2:1 compression ratio. Fast-Deflate is a hardware-implemented (FPGA) compression algorithm suitable for synchronous applications and adds a mere 10 µs of propagation delay and accommodates significant bandwidth rates. In hybrid mode on the Brocade 7840 and SX6 Blade, a 2:1 fast-deflate compression ratio can accommodate 20Gb/s per DP of FC traffic. In FCIP mode, the Brocade 7840 and SX6 Blade can accommodate 40Gb/s per DP of FC traffic.

12.3.1.2 Deflate

Deflate typically gets about a 3:1 compression ratio. Deflate accommodates up to 16Gb/s ingress from the FC side per DP. Deflate has been designed to work efficiently with circuits up to 5Gb/s per DP. Deflate is a software-processed algorithm with a hardware assist. Software-based algorithms take longer to process and may not suit synchronous applications.

12.3.1.3 Aggressive-Deflate

Aggressive-Deflate takes the trade-off between the compression ratio and the compression rate further. Aggressive-Deflate typically gets about a 4:1 compression ratio. The maximum rate per DP is 10Gb/s ingress from the FC side. Per DP, Aggressive-Deflate has been designed to work efficiently with circuits up to 2.5Gb/s. Aggressive-Deflate is a software-based algorithm and is not suitable for synchronous applications.

12.3.2 FCIP Architectures

Extension is most commonly used for business continuance via disaster recovery. Leveraging Remote Data Replication (RDR) and remote tape applications to transport critical data across a significant enough distance outside a catastrophic event preserves data. Data preservation permits an organization to recover operations.

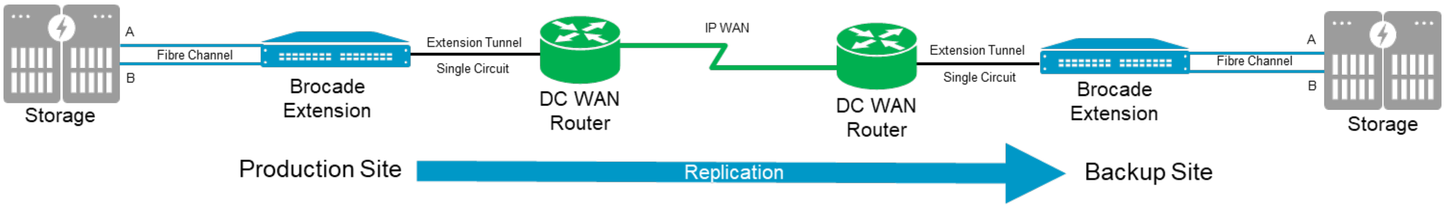
RDR is typically disk-array–to–array communications. The local storage array at the production site sends data to the array at the backup site. RDR can be done via native FC if the backup site is within a reasonable distance and there is WDM or dark fiber between the sites. However, a cost-sensitive and ubiquitous IP infrastructure is commonly available, not native FC connectivity.

High-speed Brocade Extension adds about 50 to 75 µs of propagation delay per pass through the platform (four passes = 0.2 to 0.3 ms added RTT), which is appropriate for asynchronous RDR and may be acceptable for synchronous RDR as well.

Best-practice deployment is to connect array N_Ports directly to Extension F_Ports, and not connect through a production fabric. Storage replication ports are dedicated to RDR and have no host traffic. Nevertheless, valid reasons remain to connect via a production fabric, such as tape applications, and the Extension platform cannot accommodate all the array ports.

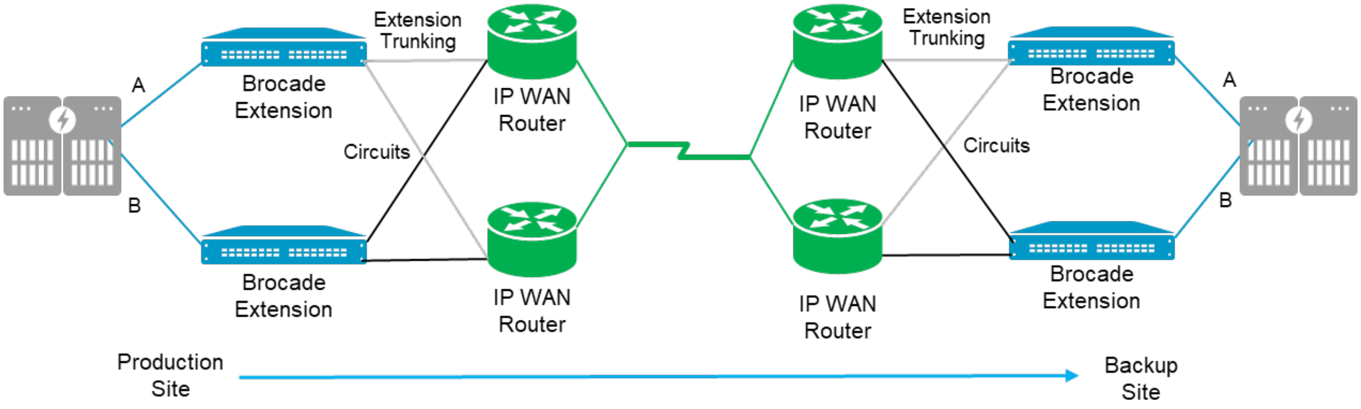
A single Extension platform can be directly connected to “A” and “B” storage controllers, referred to as a two-box solution, one at each DC, as shown in Figure 21.

Figure 21: Basic Nonredundant Extension Architecture



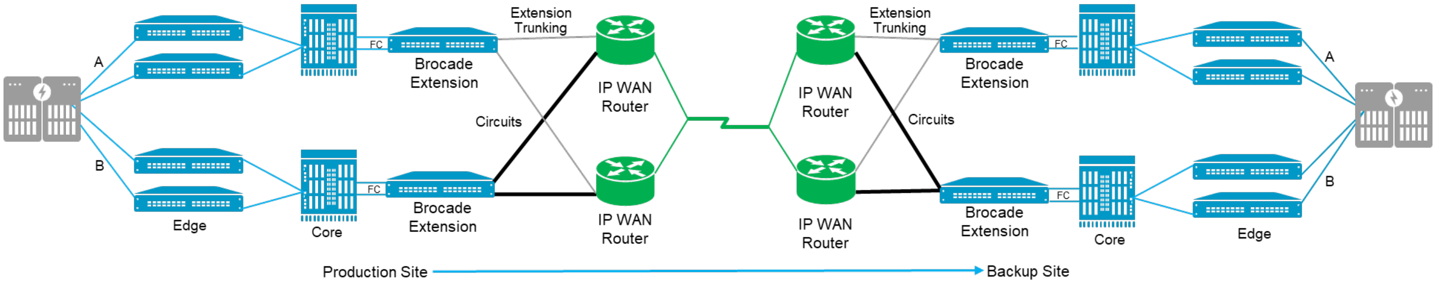
Alternatively, when one Extension platform is dedicated to the “A” fabric or controller, and a physically different Extension platform is dedicated to the “B” fabric or controller, this is referred to as a four-box solution, as shown in Figure 22. A single WAN link for both paths may be used, or different service providers may be used depending on requirements, cost, and recovery tolerance.

Figure 22: FCIP Architecture with Dedicated Extension for Each Controller



A production fabric can be extended using Extension as shown in Figure 23, but do not do so unless there is a compelling reason. The most common reason is distributed systems tape to gain connectivity to many devices then pipelining the tape traffic back to a DR DC.

Figure 23: Extension of a Routed SAN

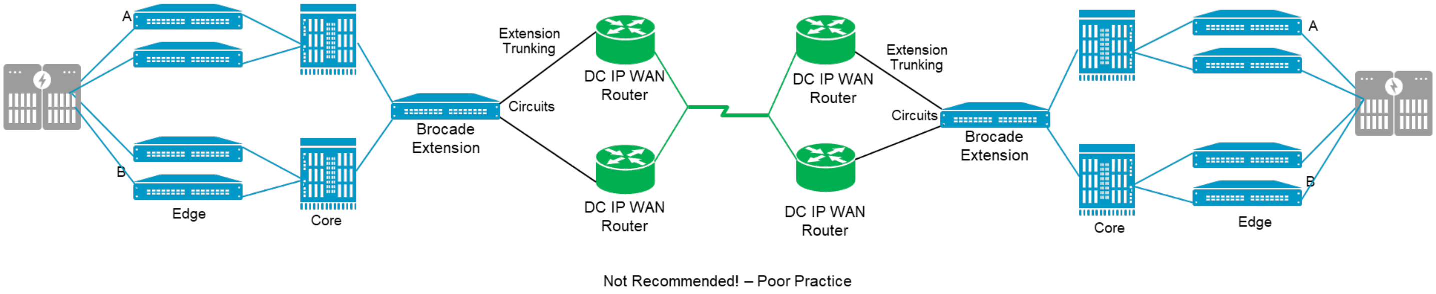


In environments that require Extension attached to a production SAN, it is not best practice to interconnect the same Extension platform to both “A” and “B” fabrics. A best practice is to have two separate and redundant fabrics in a production environment, especially if the organization could suffer financial losses during an outage. Even momentary SAN outages can “blue screen” or hang servers, forcing a reboot, which takes a lot of time in most situations.

For maximum availability, it is best practice to divide a SAN into “A” and “B” fabrics, which implies that there is an “air gap” between the two autonomous fabrics from the server HBA to the storage ports. There are no physical links between the two fabrics. Servers, storage, and VMs are equipped with drivers that monitor pathways and send data accordingly. The driver fails traffic over to a remaining path when a path is detected down.

Extension via FCR to a production fabric cannot connect to both the "A" and "B" fabrics, as shown in Figure 24. Do not use this type of architecture. Without FCR, the fabrics would merge into one big fabric, which destroys any notion of redundant autonomous fabrics. If FCR is used, the fabrics do not merge; however, a common Linux kernel device still runs attached to both fabrics. If maximum availability is the goal, this architecture is unacceptable and considered a poor practice due to its high risk. Additionally, an architecture with a common device to A and B fabrics is susceptible to human error, which could bring down the entire SAN.

Figure 24: Two-Box Solution Connected to Both Production Fabrics: Poor Practice



When connecting Extension to production fabrics, each production fabric should be designed using best-practice traditional core-edge concepts. Since storage connects directly to the core in a core-edge design, Extension switches connect to the core, or an Extension blade is placed in a core director. Standalone Extension platforms should be connected to a fabric with at least two inter-switch links (ISLs) for redundancy.

In a four-box solution, it is inappropriate to make ISL cross-connections between the two Extension platforms and the "A" and "B" fabrics because of the same reasons discussed above, a common Linux kernel and human error.

Cross-connecting circuits from a tunnel to various Ethernet DC LAN switches or IP network devices is encouraged. Circuits that traverse the IP network are point-to-point and can take alternate resilient and redundant paths without merging the A and B fabrics.

12.4 IP Extension

IP Extension accelerates, secures, and manages supported IP storage applications by leveraging Extension technology. Much of the technology is shared across the FCIP and IP Extension protocols, and they share the same tunnel. Enemies to TCP/IP-based replication are latency and packet loss. IP Extension overcomes performance degradation caused by these inherent characteristics; plus, it provides encryption.

This section covers unique points to IP Extension design, best practices, and architectures.

12.4.1 LAN Side

As mentioned previously, Brocade Extension can be represented by having three sides. One of those sides is the LAN side. The LAN side is specific to IP Extension, and it is used to connect IP storage Ethernet ports via the connected LAN.

IP Extension supports connectivity of multiple DC LANs via VLAN tagging (802.1Q) on the Ethernet links between the Extension platform and the DC LAN switches. An IP Extension gateway (`ipif lan.dp#`) must be configured and specified for each VLAN.

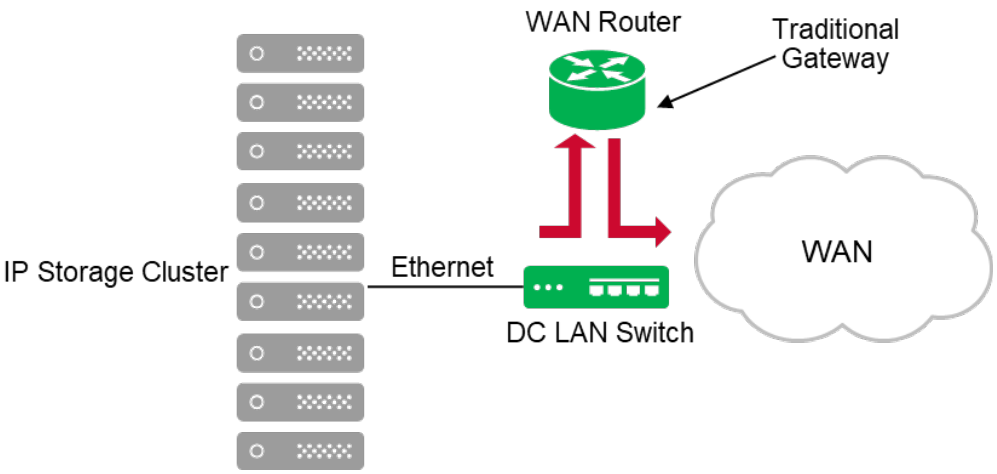
NOTE: IP subnets used for LAN-side end devices on each end of the IP Extension tunnel cannot be the same subnet.

12.4.1.1 IP Extension Gateway

IP Extension is the gateway for traffic meant to cross the Extension tunnel. If IP storage traffic is not forwarded to the IP Extension gateway, it will not utilize IP Extension.

In Figure 25, traffic comes from the IP storage cluster into the DC LAN switch. The DC LAN switches forward traffic to the traditional router gateway. The router may be an inherent part of the DC LAN switch or not. The router sends the traffic toward the destination.

Figure 25: Traditional DC Gateway

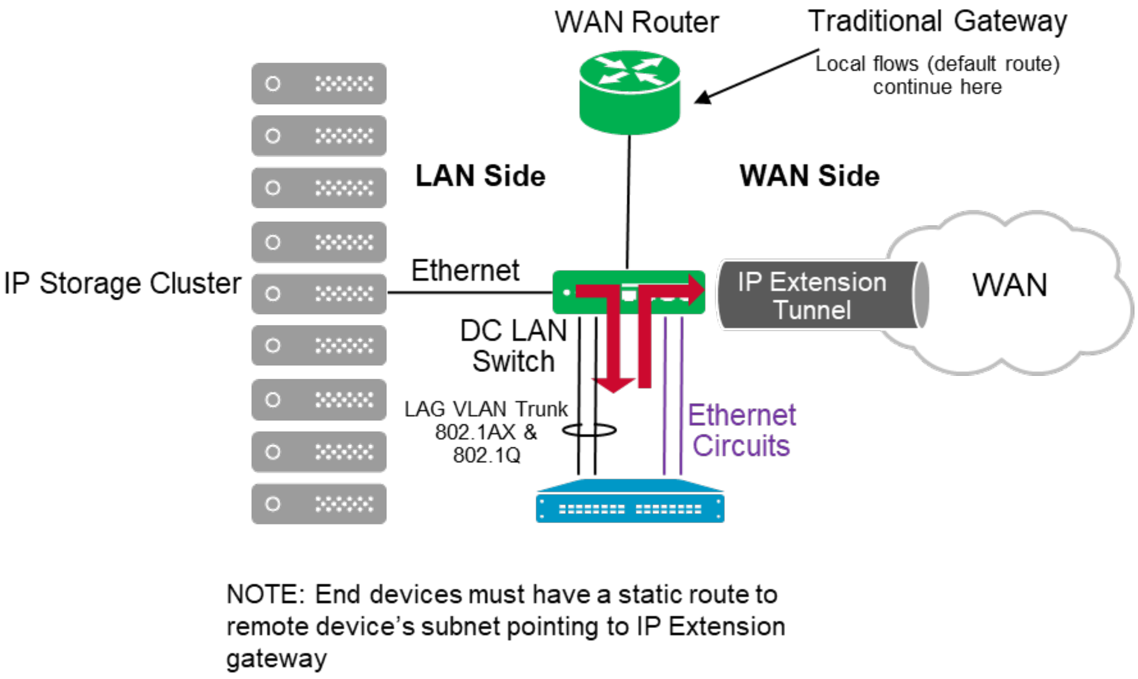


With IP Extension, it is required that the end device have a static route or some route that is more specific (more specific means the mask is longer than the default mask) than the default route that forwards remote subnet traffic to the IP Extension gateway. The remote end device is on the remote subnet. The default route is used only when a more specific route does not exist. The default route stays pointed to the traditional router gateway.

It is essential to keep in mind that putting IP Extension in the path and removing it involves merely activating or deactivating the static routes on the end devices. Traffic goes to in-path IP Extension when static routes direct it. Without static routes, traffic goes to the traditional router, which is out of the path.

In Figure 26, traffic from the remote IP storage cluster is forwarded to the IP Extension gateway, where the TCL evaluates it. A TCL is evaluated once when the TCP connection is initially opened. If the traffic matches a rule allowing it to enter the tunnel, it is sent into the specified tunnel, referred to as a target in TCL. The IP Extension traffic is now on the WAN side.

Figure 26: IP Extension Gateway


WAN Router
Traditional Gateway
Local flows (default route) continue here
LAN Side
WAN Side
IP Storage Cluster
Ethernet
IP Extension Tunnel
WAN
DC LAN Switch
LAG VLAN Trunk 802.1AX & 802.1Q
Ethernet Circuits
NOTE: End devices must have a static route to remote device's subnet pointing to IP Extension gateway

NOTE: End devices must have a static route, for example: for remote subnet w.x.y.z/m, use IP extension gateway a.b.c.d.

12.4.1.2 GE Interfaces

GE interfaces are either WAN facing or LAN facing. GE interfaces cannot do both and must be configured for one. The default is WAN. LAN-side connectivity can be made from GE/10GE interfaces on all platforms. The Brocade 7840 and SX6 Blade must be in app-mode "hybrid" (supports both FCIP and IP Extension) before a GE interface can be configured to be LAN facing with a maximum of 8 of the 16 interfaces. The Brocade 7810 has no app-mode setting; it has only hybrid mode; four of the six active interfaces can be configured as LAN facing.

The Brocade 7810 has two copper (RJ-45) ports, which can only do 1Gb/s. There is no advantage or disadvantage to using the copper ports (GE0 and GE1) over the reciprocal optical ports (GE0 and GE1). Speed is the only limitation on the copper ports.

NOTE: On any Extension platform, in-band management is not supported on data interfaces.

12.4.2 Compression

Compression operates in the tunnel scope. Compression cannot be configured per circuit. Compression must be configured identically on both ends of the tunnel; asymmetrical compression is not supported. IP Extension compression is limited to Deflate and Aggressive-Deflate. Fast-Deflate is not available for IP Extension on the Brocade 7840 and SX6 Blade. The Brocade 7810 does not have Fast-Deflate because the platform does not support the capacity that Fast-Deflate accommodates.

Compression can be configured specifically for each protocol (FCIP and IP Extension). For example, on a Brocade 7840, configure Fast-Deflate for FCIP and configure Deflate for IP Extension, which is considered best practice in a hybrid environment because the Fast-Deflate and Deflate compression engines are separate. Fast-Deflate is hardware-based, and Deflate uses hardware assist processing. 20Gb/s of FC ingress to Fast-Deflate does not consume the IP Extension capacity available for Deflate or Aggressive-Deflate.

Table 1: WAN-Side Egress Rate: Which Compression Algorithm to Use for IP Extension?

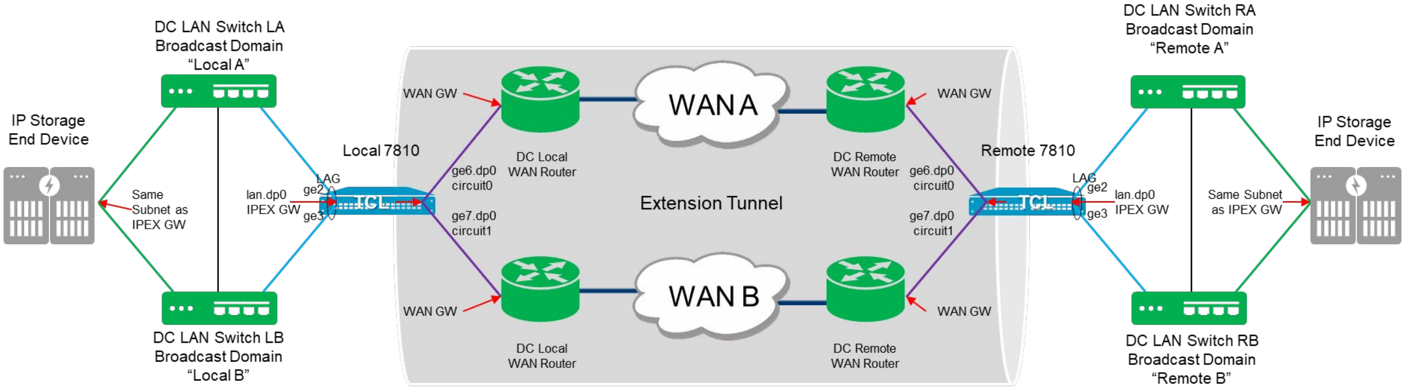
Hybrid Mode	Brocade 7810	Brocade 7840	Brocade SX6 Blade
Disabled (1:1)	2.5Gb/s (DP max egress)	20Gb/s (DP max egress)	20Gb/s (DP max egress)
Fast-Deflate (2:1)	N/A (not on the platform)	N/A (no IP Extension)	N/A (no IP Extension)
Deflate (3:1)	1.5Gb/s to 2.5Gb/s	10Gb/s to 15Gb/s	10Gb/s to 15Gb/s
Aggressive-Deflate (4:1)	20Mb/s to 1.5Gb/s	20Mb/s to 10Gb/s	20Mb/s to 10Gb/s

NOTE: Compression ratios are approximate. Broadcom makes no warranties, guarantees, or claims about the actual compression ratio achieved with customer-specific data.

12.4.3 IP Extension Architectures

Data flow through an IP Extension architecture starts at the originating storage device. A static route on that device indicates that if the destination is subnet “x”, use the IP Extension gateway. The IP Extension gateway is on the same L2 network; the end device forwards traffic to the gateway. Traffic that matches a traffic control list (TCL) is put in a tunnel and forwarded via the WAN to the remote side. The traffic is removed from the tunnel and forwarded to the destination end device on the remote side. See Figure 27.

Figure 27: Data Flow through IP Extension Architecture

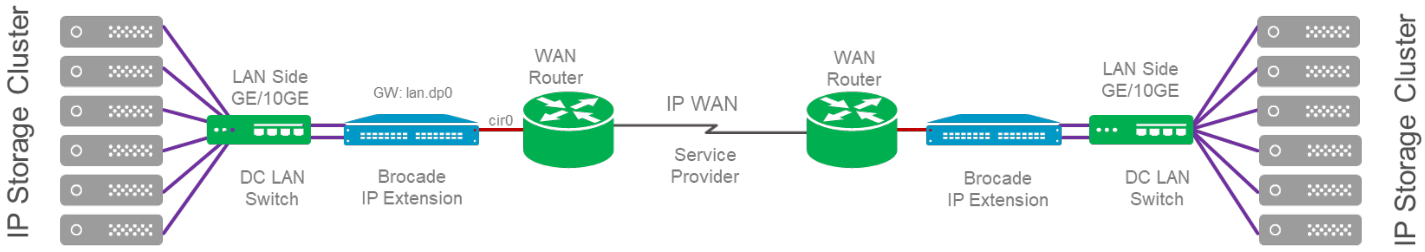


Various IP Extension architectures can be built—these range from simplistic and cost-effective to more complex, higher availability, and higher capacity architectures.

12.4.3.1 Two-Box Solutions

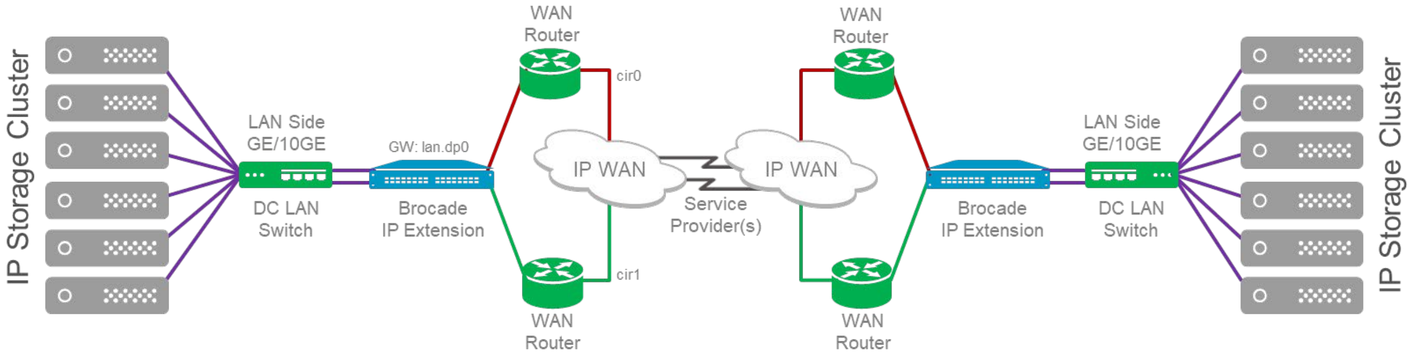
A two-box solution, not using Brocade Extension Trunking (BET), is typically used with the Base Brocade 7810. The Base Brocade 7810 is the most cost-effective level of Extension platform. It does not support BET; however, BET can be enabled with an upgrade license. Enabling BET is required to create more than one circuit per tunnel; therefore, the architecture in Figure 28 shows only a single circuit.

Figure 28: Brocade 7810 Base Unit Architecture



The two-box solution using BET is a more common architecture than without BET. Connecting WAN-side interfaces to A and B switches/routers increases availability; see Figure 29. The tunnel remains up and undisturbed when a switch/router/optic/cable/WAN goes offline. All traffic will be delivered and delivered in order. In this architecture, the DC LAN switches and Brocade Extension platform remain single points of failure.

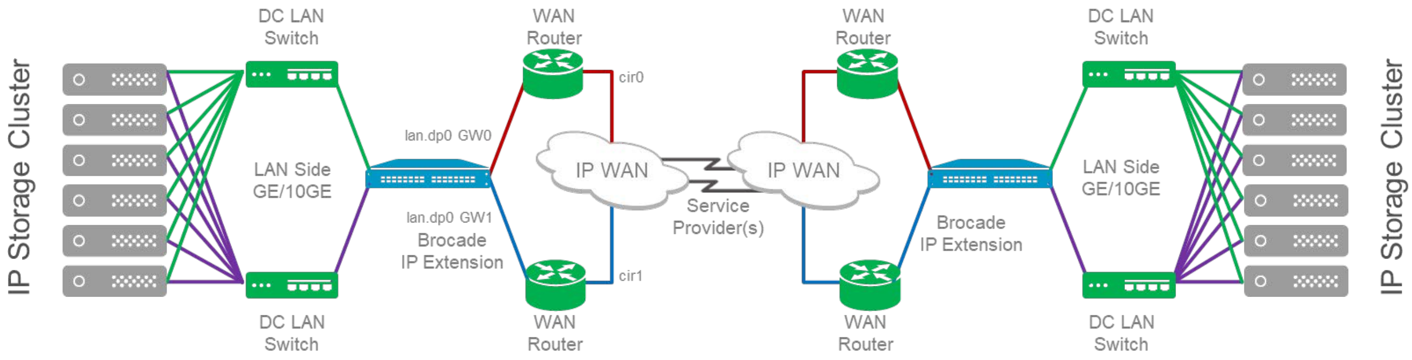
Figure 29: Two-Box Solution Using BET: Single DC LAN Switch



The two-box solution using BET with dual DC LAN switches eliminates the DC LAN switches as single points of failure. This architecture requires that the IP storage devices be capable of multiple gateways specific to their Ethernet interfaces. As shown in Figure 30, each end device has dual NICs, each connected to a different DC LAN switch. The DC LAN switch, in turn, connects to the Extension platform. The same IP Extension gateway cannot accommodate the two separate DC LAN switches. The green path requires its own VLAN and IP Extension gateway, as does the purple path. The Brocade Extension platform is a single point of failure in this architecture.

Exh. RA-3
Docket UT-181051
Page 64 of 112

Figure 30: Two-Box Solution Using BET: Dual DC LAN Switches



12.4.3.2 Four-Box Solutions

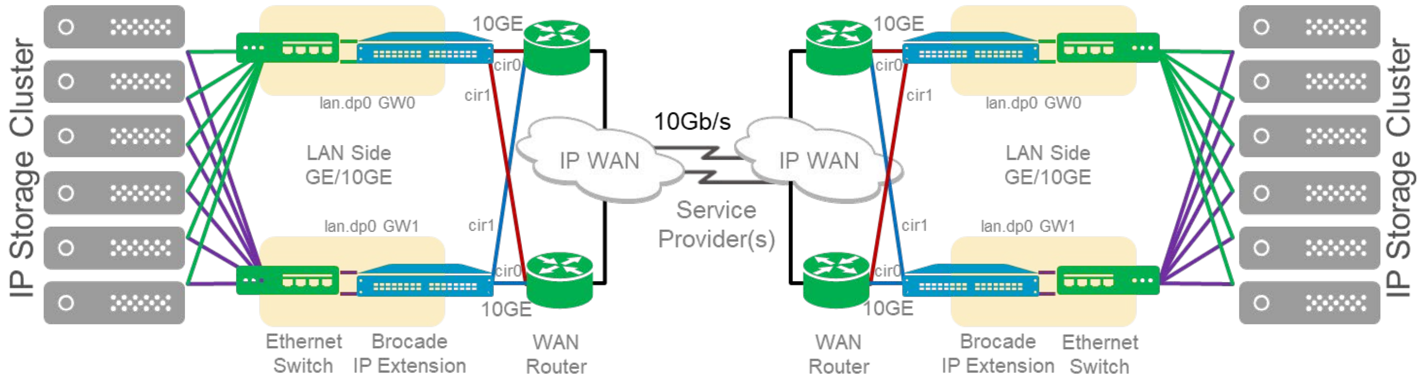
Avoiding a single point of failure may be paramount in some environments. A four-box Brocade Extension solution is possible when the end device supports the ability to accommodate more than one gateway. Deploying more than one local IP Extension platform requires more than one gateway. Brocade IP Extension on the LAN side does not offer Virtual Router Redundancy Protocol (VRRP) or Hot Standby Router Protocol (HSRP), which provide a single virtual gateway. Therefore, the end device must be capable of the following:

- A different subnet, static route, or gateway per Ethernet interface or set of Ethernet interfaces
- The same subnet with the ability to configure a unique static route per Ethernet interface or set of Ethernet interfaces

The DC LAN switches in this architecture are not configured to form a single logical Ethernet switch; see Figure 31. The two DC LAN switches are separate and autonomous. The NAS cluster can configure multiple unique gateways, not including the default gateway. The IP Extension gateway is not the default gateway; IP Extension gateways are just for the IP Extension traffic. When sending replication traffic to the remote NAS cluster, the green connections forward traffic to lan.dp0 GW0 (top), and the purple connections forward traffic to lan.dp0 GW1 (bottom).

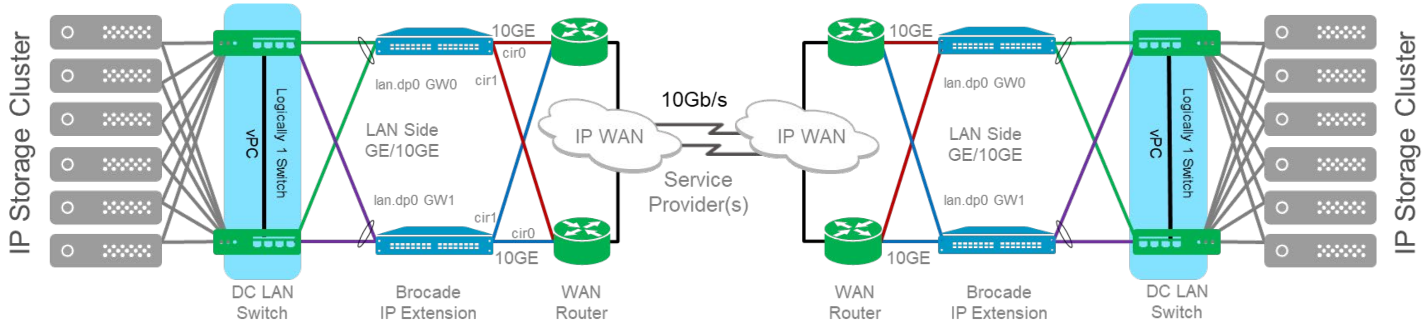
There are two Brocade extension trunks (red and blue) on the WAN side with two circuits each. Each circuit connects to a different WAN switch/router. While most data centers implement VRRP or HSRP, connecting to different WAN switches remains necessary if a switch goes down, an optic fails, or a cable is damaged. The gateway used on the WAN side is the virtual gateway created by VRRP or HSRP and is accessible from both WAN side switches. The Brocade Extension platforms are not single points of failure in this architecture.

Figure 31: Dual Connected High-Availability Architecture



The DC LAN switches in Figure 32 are logically a single Ethernet switch. A virtual port channel (vPC) connection allows the switches to join, appearing as one switch to the network. Brocade Extension can form a port channel with a single switch or across two switches. A port channel across both switches requires the DC LAN switches to be configured as a single logical switch. The purpose of this architecture is to gain redundant DC LAN switches. When one of the switches goes offline, the other switch continues to forward traffic to the IP Extension gateway. The Brocade Extension platforms are not single points of failure in this architecture.

Figure 32: Four-Box Solution with Single Logical DC LAN Switch



Chapter 13: SAN Design for Critical Workloads

With all-flash arrays (AFAs) being today's standard in enterprise data centers and transitioning to FC-NVMe AFAs well underway, critical business applications are increasingly dependent on consistent low-latency, high-throughput storage performance for demanding performance-sensitive workloads. When designing a SAN, it is vital to consider the placement of critical workloads relative to storage placement, the fan-in ratio to storage ports and ISLs/trunks.

Protecting critical workloads is crucial. Brocade SAN technology provides measures such as Traffic Optimizer (TO), Fabric Performance Impact Notifications (FPIN), Monitoring Alerting Policy Suite (MAPS), and Slow Drain Device Quarantine (SDDQ) to avoid workload interference that may experience congestion behavior. Ideally, the most demanding and critical workloads have dedicated storage ports, possibly even a dedicated array, and the shortest possible path through the SAN. The purpose is to avoid any interference from other workloads that may result in congestion or backpressure, which could adversely impact the performance of critical workloads.

13.1 Placement of Servers with Business-Critical Workloads

With core-edge SAN designs, connecting critical workload servers directly to the core alongside the storage ports is often advantageous. This practice works well when the number of business-critical workloads is easily defined and limited to a subset of servers, and there is an adequate number of core ports available.

Suppose the number of business-critical servers exceeds the number of available ports on the core; in that case, it will be necessary to connect the business-critical servers to the edge switches. The most common model is to use dedicated edge switches for business-critical servers to remove competing flows and decrease the fan-in ratio of servers to ISLs.

An alternative is to evenly distribute business-critical servers across the edge switches, assuming that workloads even out with other less demanding workloads. Although a logical approach, the practice has demonstrated that using this model is operationally complex to guarantee optimal performance for business-critical workloads.

13.2 Business-Critical VMs

In today's data centers, it is not uncommon for some or all business-critical workloads to be running on VMs. Combining this with the digital business environment, the value, criticality, and performance requirements for a given application often change throughout the application's life cycle. Inevitable change means that it can be difficult or impossible to predict the future requirements for an application. Planning placement from the beginning is not simple—luckily, hypervisors can move VMs between hosts without disruption and migrate storage when necessary. To apply the same principle to bare metal server placement for business-critical applications, deploy dedicated hypervisor clusters connected to the core or high-performance edge switches. Use these hypervisor clusters only for business-critical and performance-sensitive VM workloads.

Visibility into each VM's workload on the same datastore, backed by LUN or NSID, can be achieved with VM Insight. VM Insight enables storage administrators to monitor VM-level application performance and set baseline workload behavior. This information can be used to quickly determine whether the storage fabric is the source of performance anomalies. Storage administrators can plan placement and provision based on application requirements and fine-tune the infrastructure to meet service-level objectives based on VM Insight.

Chapter 14: Access Gateway and NPIV

This chapter goes over SAN design considerations when using Access Gateway (AG) and N_Port ID Virtualization (NPIV). Considerations are primarily related to increasing density and scale in the SAN. In addition, we describe AG default port mapping and best practices on how to design port mapping for specific SAN designs while ensuring balance and failover.

Refer to the *Brocade Fabric OS Access Gateway User Guide* for detailed information on deploying and configuring Access Gateway.

Standards-based NPIV can connect multiple F_Ports to the fabric on a single N_Port. These F_Ports can be virtual initiator ports from a single physical HBA port connecting to the fabric, such as a hypervisor with virtual HBAs or a storage device with multiple virtual target ports on the same physical HBA port.

An NPIV use case is a switch configured in NPIV mode connected to a fabric as an Access Gateway (AG). AG does not participate in the fabric as a switch; instead, it extends the number of ports connected to the fabric without increasing the number of domains. Additionally, a switch in NPIV mode is a way to connect a different vendor's switch to the fabric. Brocade fabrics support connecting switches in NPIV mode from other vendors who follow the NPIV standards.

Common NPIV use cases include the following:

- Using Brocade Access Gateway to increase port count without increasing the number of domains
- AG deployed in a POD architecture
- Connecting blade-server chassis embedded switches
- Connecting other vendor switches (including UCS-FIs)
- Storage arrays with a virtual storage controller architecture presenting separate virtual target ports *behind* the same physical target port
- Hypervisors with virtual HBAs provisioned to VMs for Raw Device Mapping (RDM) storage allocation

When connecting a switch in AG mode, the F_Ports connect to the fabric as N_Ports rather than E_Ports, as illustrated in the following two figures. Figure 33 shows a switch in native mode with all devices connecting to F_Ports and switch-to-switch connections as E_Ports (ISLs).

Figure 33: Switch Functioning in Native Mode

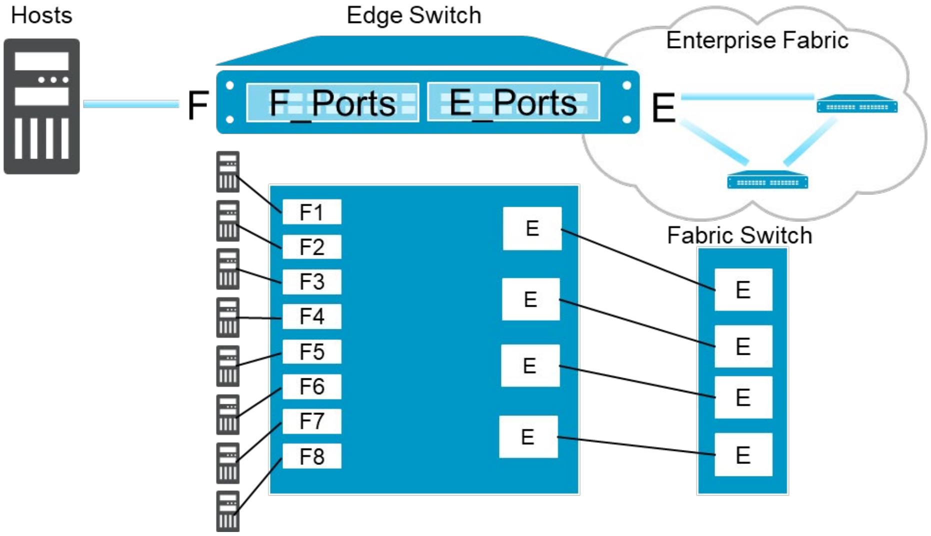
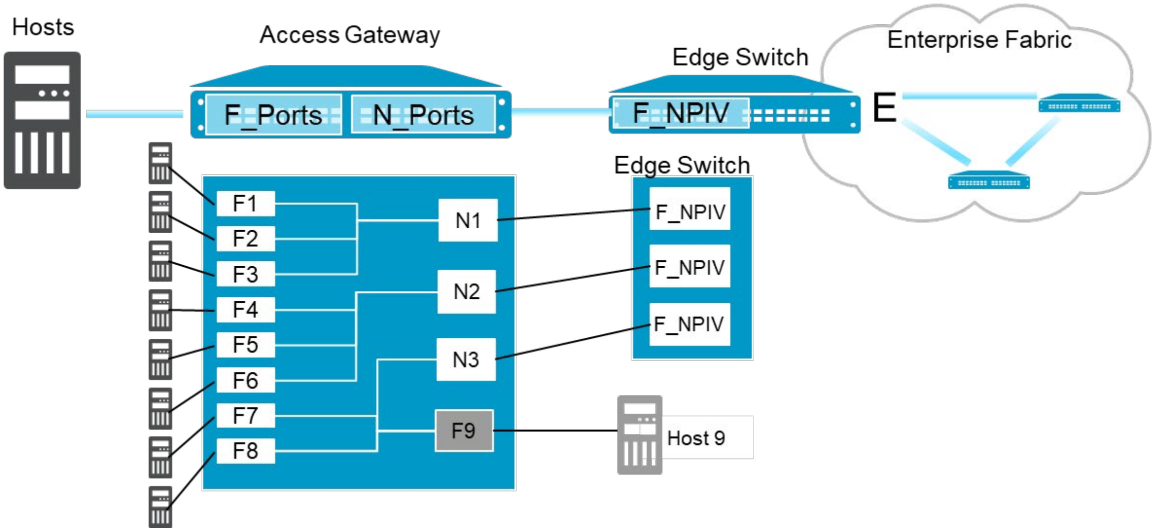


Figure 34 shows a switch in Access Gateway mode with all devices connecting to F_Ports, then mapping to N_Ports, providing the AG to switch connectivity.

Figure 34: Switch Functioning in Access Gateway Mode



Switches in AG mode are logically transparent to the host and the fabric. Therefore, you can increase the number of hosts accessing the fabric without increasing the number of switch domains. AG mode simplifies configuration and management in a fabric by reducing the number of domain IDs and ports; a fabric-specific configuration is *inherited* from the fabric and unavailable on the AG.

The main reason for using AG mode is to achieve scalability with many small switches. In an environment with many blade servers, the embedded switches can quickly encroach on the total domain count limit. Placing these switches in AG mode means that they will not consume a fabric domain ID.

With current Fabric OS levels, AG functionality is enriched; although, there are still some scenarios in which full switch functionality is more advanced. Therefore, deciding to use AG involves an evaluation of the environment to determine if AG is a design option. In a fabric with many legacy devices, identifying and isolating misbehaving devices is easier in a complete switch environment.

For configurations with hosts and targets connected to the same AG, traffic must pass through the AG to a fabric switch. If AG is not used, local traffic can be handled within the embedded switch and does not need to traverse the AG to the fabric and then back. The theoretical domain limit in a single fabric is 239, but most fabrics are typically limited to a much smaller number (56 is the maximum number of domains supported in Brocade fabrics). The domain count limit typically comes into play when a large number of small-port-count switches are deployed. Large-bladed server deployments, for example, can push the domain count over recommended limits when embedded switches are implemented. FC switches in blade server enclosures typically represent fewer than 32 ports.

NPIV was initially developed to provide access to Fibre Channel devices from IBM mainframes and improve the efficiency of mainframe I/O for virtualized environments.

Exh. RA-3
Docket UT-181051
Page 69 of 112

14.1 Benefits of the Brocade Access Gateway

- **Scalability**: You can add many Access Gateways to a fabric without increasing the domain count. A major scalability constraint is avoided when small-port-count switches or embedded switches are part of an infrastructure. Registered state change notifications (RSCNs) are also significantly reduced; only those related to the initiators on the downstream Access Gateway ports are passed on through to the fabric. Since it is essentially a device, the Access Gateway can connect to more than one fabric from its upstream ports. Brocade Access Gateways can be cascaded to reduce the number of fabric connections required to support the attached hosts' given workload or traffic level.
- **Error isolation and management**: Most initiator errors are not propagated through the fabric. Disconnecting an upstream port, for example, does not cause a fabric rebuild. Most management activities on the Brocade Access Gateway are also isolated from the fabric. One possible scenario is server administrators managing the Access Gateways and storage administrators simply providing LUNs and zoning support for the servers using NPIV.
- **Increased resiliency**: The Brocade Access Gateway supports F_Port trunking, which increases the resiliency of connections into the fabric. Losing a trunk member simply reduces the bandwidth of the upstream trunk. Although a few frames may be lost, no host connections are affected.
- **Other**: Hosts or HBAs can be configured to automatically fail over to another upstream link should the one they are using fail. Brocade Access Gateway also implements advanced features such as adaptive networking services, trunking, hot code load, Brocade MAPS, Brocade ClearLink, credit recovery, and forward error correction.

14.2 Constraints

The advantages of the Brocade Access Gateway are compelling, but there are constraints:

- Although the benefits are much more evident for servers, Brocade Access Gateway supports storage devices, but the traffic must flow through to the fabric, which has its limitations.
- The maximum number of NPIV connections per upstream port is 254.
- The number of Brocade Access Gateways per switch is limited only by what the fabric switches cansupport.

The primary factors to consider:

- The total number of devices that attach to the fabric through the Access Gateways
- The number of devices per Access Gateway N_Port
- The total number of devices attached to the switch and fabric

Refer to the *Brocade SAN Scalability Guidelines* for details.

The number of fabrics an AG can be connected to is limited by the number of N_Ports on theAG. Most deployments require two AG connections to each fabric. Note that connecting different upstream ports to different fabrics does not reduce the requirement for redundancy. All attached servers should have dual paths to storage through different fabrics and separate AG.

14.3 Design Guidelines

Use the Brocade Access Gateway when you deploy blade servers, have many low-port-count switches, or need to connect servers in different fabrics from a bladed enclosure. AG can be valuable to separate the blade enclosure management so the server administrators manage the enclosure and the storage administrators manage the fabric. Management separation is provided through the NPIV connection, which allows the AG to be managed separately by, for example, integrated blade-server enclosure management tools, without any risk to the fabric.

14.4 Monitoring

Brocade Access Gateway has been enhanced to include features found in the standard version of Brocade FOS, such as Port Fencing, device security policies, FPI monitoring, and SDDQ. However, monitoring and troubleshooting NPIV flows are less feature-rich than traditional flows.

14.5 Maintenance

Usually, there is no requirement to maintain AG firmware levels synchronized with the fabric firmware levels. Broadcom supports other vendors' NPIV-enabled devices where firmware synchronization is not possible. Maintaining firmware levels can be significant in large fabrics with many Access Gateways. The version of Brocade FOS running on fabric switches can be upgraded at one time and AGs at another time, dramatically reducing the amount of change required to the infrastructure during a single maintenance window.

See the *Brocade Fabric OS Release Notes* to determine if a synchronized Brocade FOS upgrade of Brocade Access Gateway devices is required.

14.6 Access Gateway Mapping

When a switch operates in AG mode, you must specify the AG device's routes to direct traffic from the devices on its F_Ports to the fabric ports connected to its N_Ports. The routes must be preprovisioned. The process of provisioning routes in AG mode is called mapping. By comparison, a switch operating in Native mode determines the best routing path between its F_Ports.

You can create two types of maps: port maps and device maps. Port maps are required. Device maps are optional and assign device WWNs to N_Ports and N_Port groups. Port mapping and device mapping operate as follows.

14.6.1 Port Mapping

Port mapping ensures that all traffic from a specific F_Port goes through the same N_Port. An F_Port is mapped to an N_Port or an N_Port group. To map an F_Port to an N_Port group, map the F_Port to an N_Port that belongs to that port group. All F_Ports mapped to an N_Port group are part of that N_Port group.

14.6.2 Device Mapping

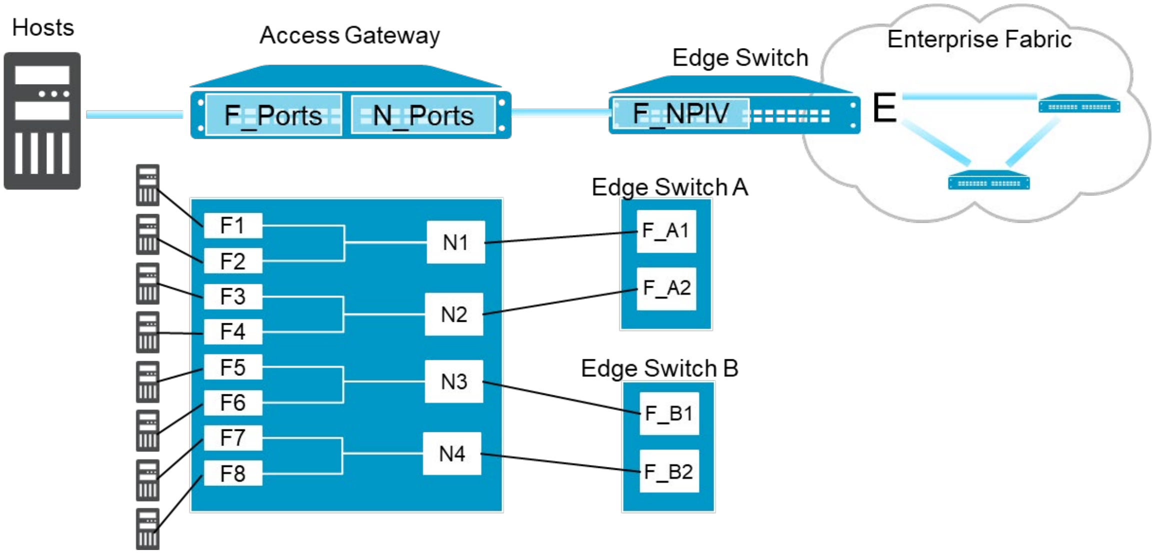
Device mapping is optional. Port maps must exist before you can create device maps. Device mapping allows a virtual port to access its destination device regardless of the F_Port where the device resides. Device mapping also allows multiple virtual ports on a single physical machine to access multiple destinations residing in different fabrics.

The preferred method is to map a device WWN to an N_Port group. When a device WWN is mapped to an N_Port, and a failover N_Port is specified, the device can reach the fabric through the primary or secondary N_Ports only. However, when a device WWN is mapped to a port group, it can log in to the fabric until the last N_Port in the particular port group remains online.

You can map a device to multiple groups. Alternatively, you can map a device to a specific N_Port.

F_Ports must be mapped to N_Ports before the F_Ports can come online. Figure 35 shows an example in which eight F_Ports are mapped evenly to four N_Ports on a switch in AG mode. The N_Ports connect to the same fabric through different edge switches.

Figure 35: Port Mapping Example



The following table shows the port mapping illustrated in the figure. F_Ports F1 and F2 map to N_Port N1; F_Ports F3 and F4 map to N_Port N2, and so forth.

Table 2: Description of Port Mapping

Access Gateway		Fabric	
F_Port	N_Port	Edge Switch	F_Port
F1, F2	N1	Switch A	F_A1
F3, F4	N2	Switch A	F_A2
F5, F6	N3	Switch B	F_B1
F7, F8	N4	Switch B	F_B2

14.6.3 Default Port Mapping

When you enable AG mode on a switch, a default mapping is used for the F_Ports and N_Ports.

The following table describes the default port mapping for a G720. Refer to the *Brocade Fabric OS Access Gateway User Guide* for default mappings on all supported hardware platforms.

Table 3: Access Gateway Default Port Mapping for a G720

Brocade Platform	Total Ports	F_Ports	N_Ports	Default Port Mapping
G720	64	0-39, 48-63	40-47	0-6 mapped to 40 7-13 mapped to 41 14-20 mapped to 42 21-27 mapped to 43 28-34 mapped to 44 35-39, 48-49 mapped to 45 50-56 mapped to 46 57-63 mapped to 47

NOTE: By default, failover and failback policies are enabled on all N_Ports.

The default mapping can be changed to meet specific requirements for your environment. For more information, refer to the *Brocade Fabric OS Access Gateway User Guide*.

Exh. RA-3
Docket UT-181051
Page 73 of 112

Chapter 15: Security

Many SAN security components are related to SAN design, and deciding to use them depends on requirements rather than network functionality or performance. When you design your SAN security policy, you do not need to implement and enable every available security feature. Some security features add performance overhead, others affect administrator productivity, and others have associated implementation costs. There is a balance between features and the value of protecting assets and the chance of exploiting a vulnerability.

One clear exception is the zoning feature used to control device communication. Proper zoning is key to fabric functionality, performance, and stability, especially in more extensive networks. Other security-related features are primarily mechanisms for limiting access and preventing attacks on the network, often mandated by regulatory requirements, while not required for SAN operation.

This chapter covers best practices for secure communication within the SAN and secure access and protection of the SAN infrastructure.

15.1 Zoning: Controlling Device Communication

Brocade zoning plays a crucial role in managing device communication. Managing device communication is essential for effective, efficient, and secure storage network use. A SAN's primary responsibility is for the flow of data between devices, and zoning specifies which device is permitted to communicate with another. When zoning is enforced, devices not in the same zone cannot communicate.

In addition, zoning protects from disruption. Fabric changes result in notifications (RSCNs) sent to the fabric and end devices. Zoning bounds the scope of RSCNs. Delivery is limited to devices within the zone and only when a change occurs. This limit reduces the processing overhead on the switch by reducing the number of RSCNs delivered, and it limits the impact in rare cases where a faulty HBA creates "noise." Thus, only devices in the zones impacted by the change are disrupted. Based on this, the best practice is to create "single-initiator single-target zoning," zones with one initiator and one target, so that changes to initiators or targets do not impact other initiators or targets. Disruptions are minimized, as illustrated in Figure 36. In addition, the default zone setting should be set to "No Access," which is what happens when zoning is disabled. Devices are isolated when zoning is disabled.

Zones can be defined by either the device's connected switch port or the device's World Wide Name (WWN). Although it takes more effort to use WWNs in zoning, it provides excellent flexibility. For example, if a device is moved anywhere in the fabric, it will maintain a valid zone membership.

15.1.1 Peer Zoning

As the number of hosts increases, configuring and maintaining single-initiator zones becomes challenging. In addition, the storage requirement of having a unique zone for each initiator and target could grow to exceed the maximum database size. "One-to-many zoning" defines a zone with one target and many initiator members. Peer zoning has the advantage of being easier to manage and avoids exceeding the maximum zone database size. Zoning initiators together in this manner results in the less effective use of hardware resources and greater RSCN traffic. Prior to the availability of peer zoning, it was not uncommon to zone multiple initiators of the same kind, typically the same OS, together with a single or even multiple targets to achieve operational efficiency when provisioning. However, by today's standards, it is not best practice. SANs that use peer zoning provide operational efficiency and effective single-initiator zoning while reducing the zone database size.

Peer zoning allows a *principal* device to communicate with *nonprincipal* devices within the zone as if it were a single-initiator, single-target zone. Nonprincipal devices within a peer zone can communicate only with the principal device. They cannot communicate with each other—nor can principal devices communicate with other principal devices. Peer zones are operationally simple and reduce database size.

A peer zone can have one or multiple principals. In general, storage ports are assigned as principals. Multiple principal members in a peer zone are used when all the nonprincipals (initiators) in the zone are to share the same target (storage) ports.

The peer zone members can be defined as WWNs or aliases specifying WWNs or “domain, port.” When defining peer zoning, you cannot mix WWNs and “domain, port” or associated aliases.

15.1.2 Target-Driven Zoning

Target-driven zoning is a variant of regular peer zoning. The user specifies the configuration in a regular peer zone. The principal device itself defines a target-driven peer zone. This device is usually a storage array but does not have to be.

Target-driven zoning manages the zoning using a third-party management interface to manage the device and the switch interactions. Target-driven zoning must first be enabled on the F_Port connected to the principal device’s N_Port.

Refer to the principal device vendor’s manual to determine the commands and options to construct a target-driven peer zone.

Refer to the applicable *Brocade Fabric OS Administration Guide* for additional details and considerations.

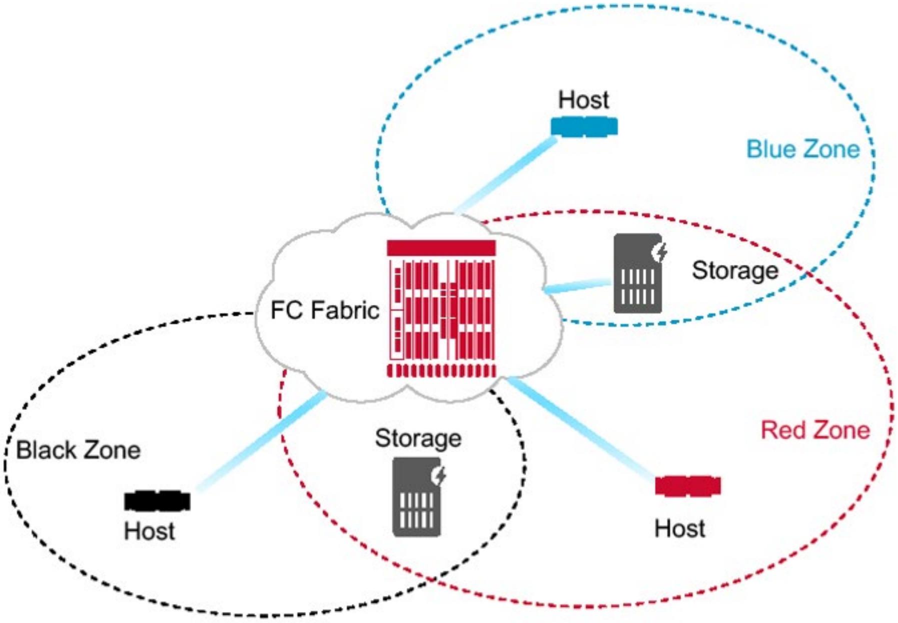
15.1.3 Zone Management: Duplicate WWNs

In a virtual environment like VMware or HP’s Virtual Connect, it is possible to encounter duplicate WWNs in the fabric, most often as a transient condition. Duplicate WWNs impact switch responses to fabric service requests like “get port WWN,” which results in unpredictable behavior. Additionally, it represents a security risk by enabling spoofing of the intended target. The fabric’s handling of duplicate WWNs is not meant to be an intrusion detection tool but rather a recovery mechanism. Before Brocade FOS 7.0, when a duplicate entry was detected, a warning message was sent to the RAS log, but no effort was made to prevent the device login of the second entry.

Available since Brocade FOS 7.0, the handling of duplicate WWNs is as follows:

- Same switch: The choice of which device stays in the fabric is configurable. The default is to retain the current device.
- Local and remote switches: Remove both entries.
- Zoning recommendations include the following:
  - Always enable zoning.
  - Use peer zoning or single-initiator zoning.
  - Define zones using device World Wide Port Names (WWPNs).
  - Set default zoning to No Access.
- Follow vendor guidelines for preventing the generation of duplicate WWNs in a virtual environment.

Figure 36: Example of Single-Initiator to Single-Target Zones



15.2 Securing the SAN Infrastructure

One of the operational advantages of a Brocade SAN is quickly adding a new switch into the fabric. A SAN administrator need only connect a new switch to an available port on an existing switch via an ISL and then power up the new switch. A unique domain ID is automatically assigned, and the configuration files are downloaded to the new switch. However, from a security perspective, this time-saving administrative ease-of-use capability also means that anyone with a switch and physical access could potentially connect to an existing fabric and gain fabric control. If an attacker with admin or root access on a rogue switch were to use this technique, the attacker would now have admin and root privileges for the entire fabric.

There are several layers of defense available to secure and protect the SAN. The following list (in order of ease of deployment) describes best-practice configurations (layers) to secure the SAN. At a minimum, the first four layers should be deployed in any environment, and the remaining layers should be deployed depending on the security requirements in your organization.

- Persistently disable unused ports.
- Prevent switch ports from becoming E_Ports.
- Configure monitoring, alerting, and logging.
- Use a strict fabric-wide consistency policy where possible.
- Use the SCC policy to restrict switch connections to the fabric.
- Use an FCS policy to further restrict security configuration changes.
- Use DCC policies to restrict device access by WWN and physical switch ports.
- For more sensitive environments, use DH-CHAP to authenticate devices that join a fabric.

The first and simplest line of defense is to persistently disable all unused ports, preventing someone without management privileges from connecting to the fabric. It is vital to use the persistent disable option to ensure that disabled ports remain disabled after a reboot or power cycle. Otherwise, an attacker could unplug the switch to re-enable unused ports.

The second line of defense is to prevent ports from becoming E_Ports. If an unused port remains enabled, a new switch will not join the fabric since the port cannot become an E_Port.

The third line of defense is to configure monitoring, alerting, and logging to ensure visibility into any unexpected switch events and attacks.

The fourth line of defense is to use a fabric-wide consistency policy to ensure that all switches in all fabrics consistently avoid a “weak-link” that could be exploited in an attack.

Subsequent layers use an access control list (ACL) defense, described in more detail in the following section.

It is unnecessary to implement every one of these lines of defense to prevent unauthorized access to a fabric. The number of security layers an organization decides to implement depends on their requirements, data sensitivity, environment sensitivity, and tolerable risk. In reality, very few organizations implement all of these levels. It is up to each organization to establish the acceptable risk and decide which features should become part of operations.

15.3 Access Control Lists (ACL)

Access control lists (ACLs) provide network security via policy sets. Brocade FOS provides several ACL policies, including a Switch Connection Control (SCC) policy, a Fabric Configuration Server (FCS) policy, a Device Connection Control (DCC) policy, an IP Filter policy, and others. The following subsections briefly describe each policy and provide basic guidelines.

A more in-depth discussion of ACLs can be found in the *Brocade Fabric OS Administration Guide*.

15.3.1 SCC Policy (Switch Connection Control)

The SCC policy restricts fabric elements from joining a fabric, particularly Brocade FOS platforms. Only switches specified in the policy are allowed to join the fabric. All other attempts to join will fail authentication, resulting in the E_Ports being segmented due to a security violation.

Use the SCC policy in environments where there is a need for strict control of fabric members. Since the SCC policy can prevent switches from participating in the fabric, it is essential to review and adequately maintain the SCC ACL regularly.

15.3.2 FCS Policy (Fabric Configuration Server)

Use the FCS policy to restrict the source of fabric-wide settings to one FC switch. The policy contains the WWN of one or more switches. The first online WWN in the list becomes the primary FCS. If the FCS policy is active, only the primary FCS can make and propagate fabric-wide parameters. These parameters include zoning, security (ACL) policy databases, and other settings.

Use the FCS policy in environments where there is a need to strictly control fabric settings. As with other ACL policies, it is essential to regularly review and adequately maintain the FCS policy.

15.3.3 DCC Policy (Device Connection Control)

The DCC policy restricts devices via WWN from attaching to an FC port. The policy specifies the FC port and one or more WWNs allowed to connect to that port. The DCC policy set is comprised of the DCC policies defined for each FC port. Not every FC port must have a DCC policy, and only those ports in the active policy set enforce access control. A port in the active DCC policy set allows only those WWNs to connect and log in to the fabric. All other WWNs fail authentication when attempting to connect, which results in the corresponding F_Port being disabled due to the security violation.

Use the DCC policy in environments where there is a need for strict control of fabric members. Since the DCC policy can prevent devices from participating in a fabric, it is essential to review and adequately maintain the DCC policy regularly.

15.3.4 Policy Database Distribution

Security policy database distribution provides a mechanism for controlling the distribution of each policy on a per-switch basis. Switches can use individually configured policies to either accept or reject a policy distribution from another switch in the fabric. In addition, a fabric-wide distribution policy can be defined for the SCC and DCC policies with support for strict, tolerant, and absent modes. These modes can be used to enforce whether the SCC and DCC policy must be consistent throughout the fabric.

- Strict mode: All updated and new policies of the type specified (SCC, DCC, or both) must be distributed to all switches in the fabric, and all switches must accept the policy distribution.
- Tolerant mode: All updated and new policies of the type specified (SCC, DCC, or both) are distributed to all switches in the fabric, but the policy does not need to be accepted.
- Absent mode: Updated and new policies of the type specified (SCC, DCC, or both) are not automatically distributed to the other switches in the fabric; policies can still be manually distributed.

Together, the policy distribution and fabric-wide consistency settings provide a range of control on the security policies from no control, little control, to strict control.

Refer to the *Brocade Fiber Channel Security Best Practices* for a detailed discussion of SAN security concepts and issues.

15.3.5 Authentication Protocols

Brocade FOS supports both Fibre Channel Authentication Protocols (FCAPs) and Diffie-Hellman Challenge Handshake Authentication Protocols (DH-CHAPs) on E_Ports and F_Ports. Authentication protocols provide additional security during link initialization by assuring that only the desired device/device type connects to a given port.

15.4 Secure SAN Management

User account and privilege management are cardinal to secure SAN management, with strong policies for accounts and passwords in combination with separation of duties and assigned privileges on a need-to basis only.

The following list outlines best practices for secure SAN management:

- Allow only secure protocols to connect the switches (SSH, HTTPS, and SNMPv3).
- Use unique user accounts with proper roles and privileges (RBAC).
- Change default passwords on *all* default accounts, and, ideally, do not use default accounts.
- Create and enforce password policies (strength, history, expiration, and lockout).
- Use a centralized account and password management methods such as RADIUS, TACACS+, or LDAP.

15.4.1 Role-Based Access Controls

One way to limit access to a fabric is through user roles. Brocade FOS has predefined user roles, each authorizing a subset of CLI commands. Predefined roles are known as Role-Based Access Controls (RBAC) and are associated with user login credentials. RBAC alignes users with their function and authority. To enforce separation of duties, users are granted specific privileges based on an organization's security model. A role could be read-only, allowing users to only view information but not modify or delete it. At the opposite end of the spectrum, there is a role granting full admin privileges. Other roles fall in between. Roles can be customized for specific functions, such as an operator or a security administrator.

15.5 Securing Management Interfaces

Management interfaces are a vulnerable point in any IT infrastructure; therefore, protecting them should always be a high priority and reasonably straightforward. The following list outlines the measures to protect the management interfaces:

- Use a separate VLAN (or private VLANs) for the management network.
- Use secure protocols to access management interfaces (SSH, HTTPS, and SNMPv3).
- Disable the equivalent unsecure protocols (Telnet, HTTP, and SNMPv1).
- Limit the entry points for management with an IP Filter policy, and use an FCS policy if necessary.

A straightforward technique for protecting management interfaces is to use a separate and dedicated VLAN and subnet to isolate the management network from user and production networks. An isolated network limits access to SAN administrators. Since insiders can be a significant threat, it is always good practice to use only secure protocols to encrypt the communications between management workstations and the devices being managed. Encrypted communications can be accomplished using SSH, HTTPS, and SNMPv3. Disable equivalent unsecured protocols: Telnet, HTTP, and SNMPv1.

15.5.1 IP Filter

The IP Filter policy is a set of rules applied to the IP management interfaces as a packet-filtering firewall. According to the IP Filter rules, the firewall permits or denies (drops) the traffic going through the IP management interfaces.

Fabric OS supports multiple IP Filter policies defined at the same time. A name and associated type identify each IP Filter policy. Two IP Filter policy types, IPv4 and IPv6, exist to provide separate packet filtering for IPv4 and IPv6. You cannot specify an IPv6 address in the IPv4 filter, nor can you specify an IPv4 address in the IPv6 filter. There can be up to six different IP Filter policies defined for both types. Only one IP Filter policy for each IP type is active at a time.

The IP Filter policy restricts access to the Ethernet management ports. Only the IP addresses listed in the IP Filter policy are permitted to connect to the specified TCP/UDP port (if specified). IP Filter rule construction can vary.

The IP Filter policy should be used to allow secure protocols only. For additional security in environments where there is a need for strict control of fabric management access, the source addresses or subnet from which SAN administration is done should be specified as the only IP addresses with access permitted. As with other ACL policies, regularly reviewing and adequately maintaining the IP Filter policy is essential.

Chapter 16: Brocade Management Platform and Data Gathering

16.1 SANnav™ Management Suite

16.1.1 SANnav Global View

To address the management needs of large-scale SAN environments or environments distributed globally, SANnav supports a hierarchical management model, where a higher-level "global" application view provides comprehensive visibility, summarization, and seamless navigation across multiple instances of SANnav Management Portal. SANnav Global View allows a user to drill down into any instance and perform detailed monitoring, investigation, and troubleshooting based on events rolled up into the global view.

16.1.2 SANnav Management Portal

SANnav allows you to efficiently manage your SAN infrastructure through various easy-to-use functions. SANnav implements a highly scalable client-server architecture for SAN management. With a modern browser-based UI, SANnav eliminates the need for a Java-based thick client. The SANnav user interface is designed based on real-world use cases and user workflows, providing a highly intuitive user experience. SANnav uses a micro-services architecture based on Docker container technology that allows it to scale to meet the management needs of small and large SAN environments and those that may change over time. This scalable architecture allows SANnav to support new functionality without causing performance degradation to the application.

SANnav Management Portal allows the management of one or more SAN fabrics in the same or different geographical locations, and it supports a maximum of 15,000 physical SAN ports. For environments larger than15,000 ports, you can deploy multiple SANnav Management Portal instances.

SANnav Management Portal does not replace Brocade Web Tools or the Fabric OS command-line interface.

16.1.3 SANnav Deployment: Requirements and Scalability

SANnav is available for Linux-based servers (VM or bare metal) or for virtual appliances (OVA). The following table specifies the server and virtual-machine requirements for deploying SANnav Management Portal.

Table 4: Requirements for SANnav Installation on a Bare Metal Server or VM

Product/Edition	Max. Switch Ports/Instances under Management	Operating System	Host Type	CPU	Memory	Hard Disk
Brocade SANnav Management Portal Base Edition (Manages switches only, no directors)	600	RHEL 7.7, 8.0, & 8.1 CentOS 7.7, 8.0, & 8.1	Bare Metal, ESXi VM	16 Cores	48 GB	600 GB
Brocade SANnav Management Portal Enterprise Edition (Required to manage directors)	Up to 3,000	RHEL 7.7, 8.0, & 8.1 CentOS 7.7, 8.0, & 8.1	Bare Metal, ESXi VM	16 Cores	48 GB	600 GB
	Between 3,000 and 15,000	RHEL 7.7, 8.0, & 8.1 CentOS 7.7, 8.0, & 8.1	Bare Metal, ESXi VM	24 Cores	96 GB	1.2 TB
Brocade SANnav Global View	Up to 20 SANnav Management Portal instances	RHEL 7.7, 8.0, & 8.1 CentOS 7.7, 8.0, & 8.1	Bare Metal, ESXi VM	16 Cores	32 GB	450 GB

Exh. RA-3
Docket UT-181051
Page 80 of 112

The following table shows the virtual machine specifications required for the OVA deployment of SANnav Management Portal.

Table 5: Requirements for SANnav Virtual Appliance (OVA) Installation

Requirement	Base or Enterprise License with up to 3,000 Ports, Included with the SANnav OVA Package	Enterprise License with up to 15,000 Ports
Operating System	CentOS 8.0	CentOS 8.0
Server Package	VMware ESXi / vCenter 6.7 or higher	VMware ESXi / vCenter 6.7 or higher
CPU	16 Cores	24 Cores
CPU Sockets	2	2
Memory (RAM)	48 GB	96 GB

16.2 Getting Started with Brocade SANnav Management Portal

16.2.1 Fabric Discovery

Discover your fabrics quickly through SANnav Management Portal to begin managing and monitoring all entities that belong to such fabrics. When devices running Fabric OS 9.0 and later are discovered, streaming of telemetry data to the SANnav server is automatically initiated for your monitoring and troubleshooting needs with no additional configuration required.

16.2.2 User Management

Access to SANnav Management Portal is controlled by authentication and authorization of users. Authentication is the process of validating user names and passwords. Authorization is the process of validating the roles and areas of responsibility (AORs) for each user.

Roles and AORs

In SANnav, SAN administrators should set appropriate roles and fabric permissions for the intended users. Specific roles help to minimize potential issues and permit changes to be made by only users with sufficient rights. SANnav offers the required flexibility to fully customize user roles and areas of responsibility to fit a wide range of customer needs.

Third-Party Authentication and Authorization Integration (LDAP, RADIUS, TACACS+)

Customers can also configure SANnav to authenticate user names and passwords to use an external server. Optionally, external servers can also be used for user authorization involving roles and AORs. The SANnav Management Suite supports external server authentication and authorization via LDAP, RADIUS, and TACACS+.

16.2.3 Device Enclosure Validation

SANnav introduced support for the automatic creation of host and storage enclosures of devices, given that such devices support FDMI. Moreover, for devices in the SAN that do not support FDMI, SANnav provides a way to create and edit desired host and storage enclosures manually.

Device enclosures provide users with an intuitive way to keep track of device ports and the physical device they live in—enclosures aid SANnav in assessing device health and forming accurate topology views.

Exh. RA-3
Docket UT-181051
Page 81 of 112

16.2.4 Monitoring and Alerting Policy Suite

With SANnav, SAN administrators can enable, configure, and distribute monitoring policies across each Brocade platform in the SAN, making it easy for users to obtain consistent monitoring across their environment or monitor specific switches of interest more or less aggressively.

If you are already familiar with MAPS and have existing policies that have been tailored over time, SANnav supports the import, customization, and distribution of such policies.

16.2.5 Dashboard Monitoring

SANnav Management Portal offers a set of predefined monitoring dashboards aimed to satisfy the monitoring needs of most customers:

- **Health Summary Dashboard**

  The Health Summary Dashboard monitors all SAN entities: hosts, storage, switches, and fabrics, based on established Brocade best practices, MAPS violations, and hardware statuses. SANnav provides insights and recommended actions to assess issues further.

- **Network Port Traffic Conditions Dashboard**

  The Network Port Traffic Conditions Dashboard monitors ports in the fabric for congestion while visually tracking severity over time. This dashboard allows users to take a deeper look into the causes of congestion through a port and flow-level investigation.

- **Extension Dashboard**

  The Extension Dashboard monitors extension-related statistics, providing insights into tunnels and circuits.

In addition, users can create custom monitoring dashboards from the 80+ status and performance widgets available; for example, a MAPS Database Dashboard that monitors for all violations occurring in your SAN while providing easy access to troubleshooting actions.

16.2.6 Reporting

SANnav reporting, like the dashboards, is user-customizable and offers an extended set of widgets to choose from. Reports can be scheduled, run on-demand, viewed in SANnav, and downloaded in various formats (HTML, PDF, CSV).

Reports can be used to keep track of SAN device inventory, zoning configurations, top talkers, congested devices, events of interest, and more. SANnav also allows custom filters to be applied to reporting templates, providing users with the flexibility of reporting on only what is of interest.

16.2.7 FOS Version Management

Make use of the SANnav built-in Fabric OS repository, where users can upload and manage FOS versions along with release notes. The FOS repository enables users to simultaneously perform single or multiple FOS upgrades across multiple switches, leading to substantial time savings.

16.2.8 Event Management

Using Event Management features within SANnav allows trap configurations, SNMP traps registrations, syslog events, and other information from switches. Users can also view, search, and filter event logs. SNMP traps and syslog messages can also be configured and forwarded to external destinations.

The following are some of the actions that you can configure for events when triggered:

- Generate email alerts.
- Trigger SupportSave.
- Enable maintenance mode to suppress events in the event log.

You can also perform robust event analysis by filtering events using custom network scopes, date ranges, and stackable filters. Users can perform searches within the event log and generate event reports on demand.

In addition, Event Management provides several event-managing widgets that you can add to dashboards, such as the Top N Events, Events Summary, and Health Violations widgets.

16.2.9 Northbound Telemetry Streaming Support

Northbound streaming provides support to securely stream performance and flow metrics from the switch to an external Kafka cluster (the northbound server). Streaming gives you access to a large set of data from one or more managed fabrics that can be used to build customized reports or applications.

Raw SNMP metrics from Performance Monitor data and raw flow metrics received from switch streaming will be streamed to the northbound server.

Configuration of the streaming interface is controlled using the REST API.

Refer to the *Brocade SANnav Management Portal REST API and Northbound Streaming Reference Manual* for northbound streaming and configuration information.

16.2.10 SANnav Backup and Restore

SANnav allows you to back up the SANnav server data and restore it as required, such as in scenarios where the data is deleted or corrupted. Also, you can use the backup when you want to bring up a new SANnav server. Creating a backup helps to protect the server's data and configuration in the event of a disaster, such as a server failure. You have options to exclude large data sets, which can save restoration time in some situations. You can perform a scheduled daily or weekly backup or an on-demand backup.

16.2.11 Backup Recommendations

The recommendations for backup follow:

- Perform a full backup weekly or monthly as daily backups may slow down the server.
- Ensure that your backup location has enough disk space before backing up your data.
- Ensure that your backup location is different from where SANnav is installed.
- For scheduled backups, occasionally check if the backup data size has abnormal patterns (files being too large or too small).
- Periodically verify that enough disk space exists to complete scheduled backups successfully. SANnav does not purge older backups.

16.3 Brocade SAN Health

Brocade SAN Health is a free tool that securely captures, analyzes, and reports comprehensive information about Brocade fabrics running Brocade Fabric OS and Cisco MDS running SANOS/NXOS. SAN Health can perform the following tasks:

- Taking inventory of devices, switches, firmware versions, and SAN fabrics.
- Capturing and displaying historical performance data.
- Comparing zoning and switch configurations against best practices.
- Assessing performance statistics and error conditions.
- Alerting the administrator of relevant technical support bulletins (TSBs).
- Producing detailed reports (in Microsoft Excel) and diagrams (in Microsoft Visio).

Download Brocade SAN Health and find details and instructions on using it at www.broadcom.com/sanhealth.

Chapter 17: Brocade Support Link

Brocade Support Link (BSL) is available on all switches running Fabric OS 8.2.1c or later. This chapter goes over the main features and considerations for enabling and deploying Brocade Support Link.

BSL depends on enabling Active Support Connectivity (ASC) on all switches in the SAN. When ASC is enabled and configured, every switch periodically collects configuration settings, log messages, and port metrics and forwards that data to the Support Link Server (SLS) at Broadcom. The data is comparable to the data collected using SAN Health and a subset of SupportSave. The collected data is exclusively event logs, configuration data, diagnostics data, and metadata. It does not include any user data transported by the SAN.

The transfer of collected data can be performed directly from the switches to the SLS or by using the Brocade Active Support Connectivity Gateway (ASC-G). Using an ASC-G is considered best practice.

Deploying with ASC-G also provides additional functions and offerings such as Data Collection Assistant and Automated Case Creation, which are unavailable when switches send data directly to the SLS.

17.1 BSL Features

Brocade Support Link provides a tiered model of offerings based on support contracts with Broadcom or the OEM. For more detail, contact your sales representative.

- Configuration, Performance, and Inventory (CPI) reports

  The CPI report provides a comprehensive package of all configuration, inventory, and performance data for the past 24 hours from the date of CPI report generation. This report is valuable for inventory-related tasks such as documentation, planning, and SAN operations.

- Best Practice Assessment (BPA) reports

  The BPA report provides a best practice assessment of the SAN across hundreds of metrics in areas of security, monitoring, health, utilization, performance, scalability, and comparison of configuration consistency across fabrics. The assessment includes scorecards for easy tracking, a findings list with recommended priorities, and explanations for corrections. The exact syntax for executing the recommended action is provided for each recommendation, enabling SAN operators to follow best practices based on two decades of Broadcom experience. In addition, BPA identifies any known defects applicable to the SAN and calls out proposed actions.

- Fabric Analytics (FA)

  Fabric Analytics continually processes all telemetry data from the SAN in an analytics engine to identify the root-cause for current or potential performance issues related to congestion and physical-layer issues. The Proactive Health Summary report provides a comprehensive view of performance-impacting issues within the past 24 hours. A comparison is made to a baseline of the past week. It provides trending to proactively identify potential issues before they impact performance. An in-depth analysis is provided for each identified issue, identifying culprits and adversely impacted “victims” in each fabric.

- Extension Reports

- Extension reports are provided as an add-on to Fabric Analytics and deliver comprehensive analysis and identification of deviation in performance on Extension tunnels and circuits.

- Automated Case Creation (ACC)

  Automated Case Creation integrates SAN monitoring with support. When critical events are identified in the SAN, a support case is automatically created. The Technical Assistance Center (TAC) can immediately start troubleshooting and resolving cases without human initiation or interaction. ACC requires the deployment of an ASC Gateway.

- Data Collection Assistant (DCA)

  Data Collection Assistant is enabled with ASC-G as a centralized point for SupportSave triggering or end-user scheduling. In addition, DCA can be combined with ACC; when a critical event occurs, ACC creates a support case. Secondly, DCA triggers SupportSaves from the switches necessary in determining the root cause.

17.2 Deployment Options

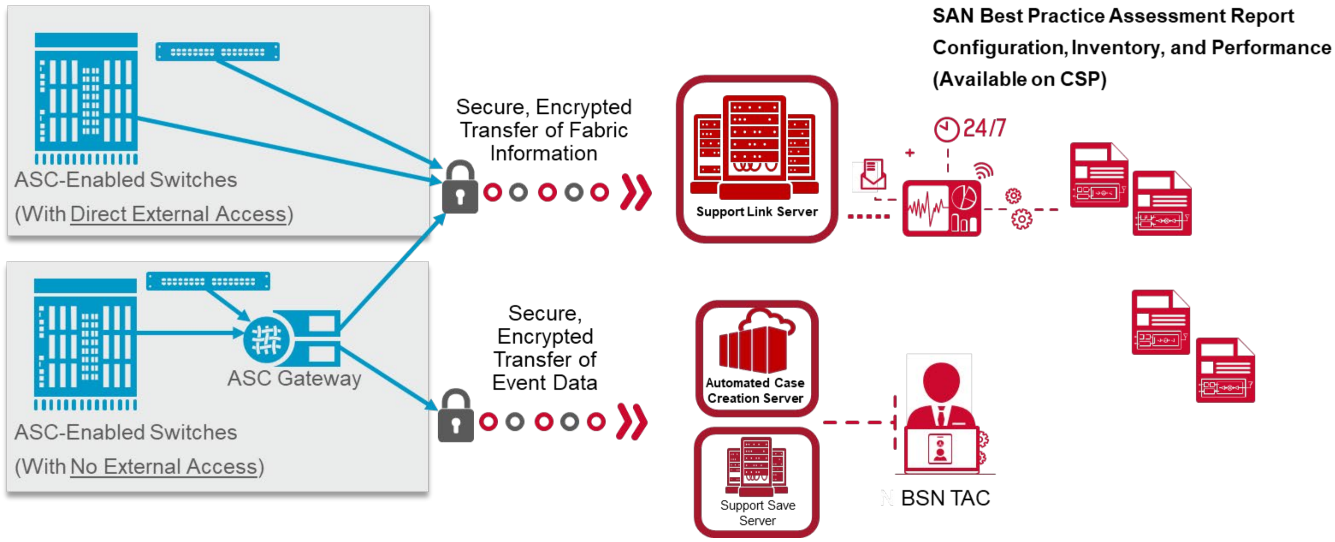
When deploying BSL, the first decision is to enable the switches to use the ASC-G to provide transport to the Support Link Server or enable the switches to send directly to the SLS, as illustrated in Figure 37.

The general recommendation is to deploy with ASC-G unless it is impossible to provide the infrastructure services necessary for the ASC-G(s) or if the SAN environment is minimal, one to two switches.

Only the ASC-Gs need access through the firewall. A best practice is to deploy two ASC-Gs for redundancy. In addition, features such as DCA and ACC are only available in deployments with the ASC-G.

Illustrated in Figure 37 are the two different ASC connectivity methods for transporting BSL data to Broadcom.

Figure 37: Active Support Connectivity Architecture


ASC-Enabled Switches
(With Direct External Access)
ASC Gateway
ASC-Enabled Switches
(With No External Access)
Secure, Encrypted Transfer of Fabric Information
Secure, Encrypted Transfer of Event Data
Support Link Server
SAN Best Practice Assessment Report
Configuration, Inventory, and Performance
(Available on CSP)
24/7
Automated Case Creation Server
Support Save Server
BSN TAC

17.3 BSL Deployment with ASC-G

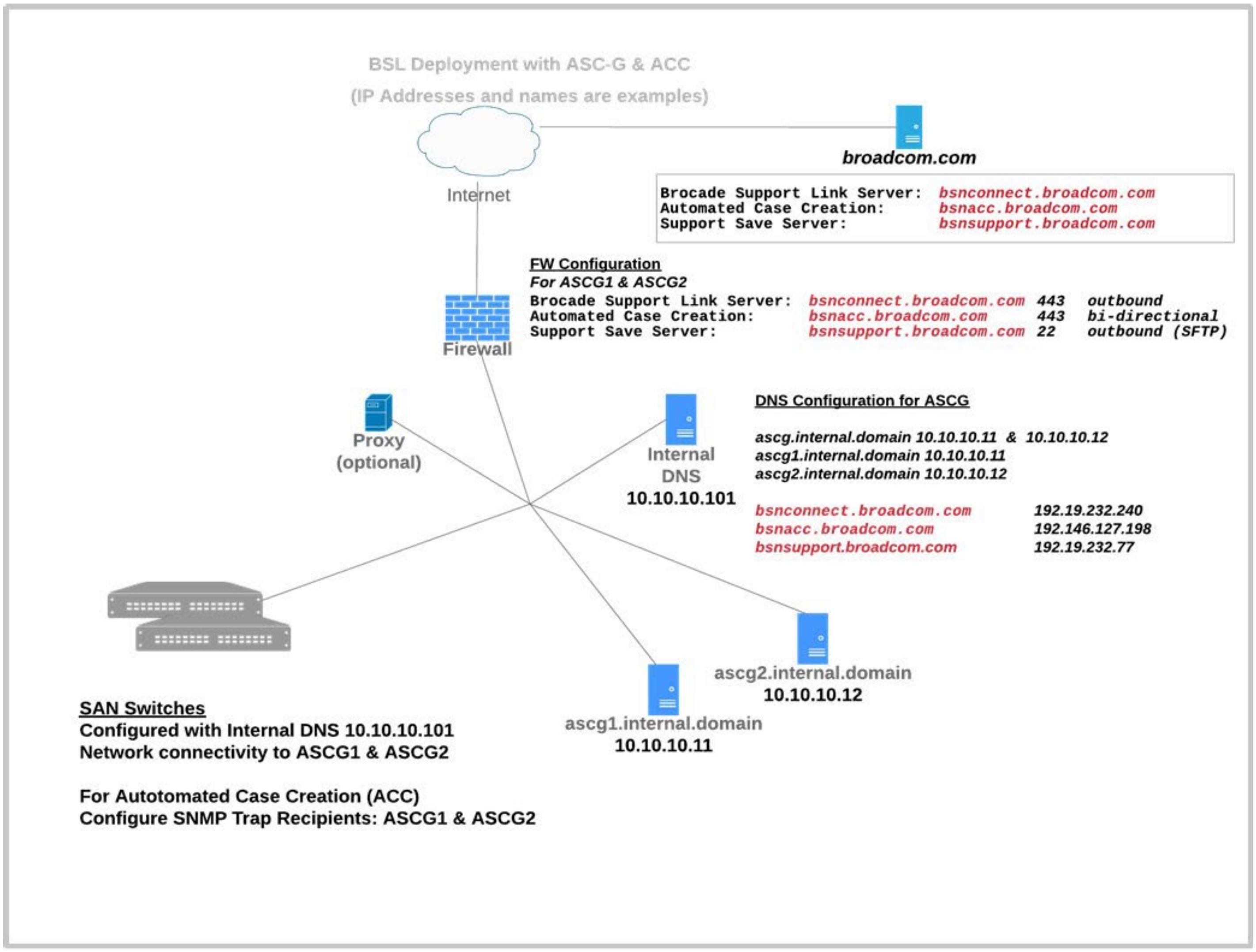
Deployment of BSL with the ASC Gateway is illustrated in Figure 38, with each of the infrastructure requirements described in more detail. In general, deploy two or more ASC-Gs per site for redundancy purposes. ASC-Gs are sized to handle 200 switches per instance.

In Figure 38, `ascg1.internal.domain` and `ascg2.internal.domain` are the two instances of the ASC Gateway deployed. The common name for the two instances, `ascg.internal.com`, is used as a common DNS record for configuration on the SAN switches. All IP addresses on the internal network are examples.

The internal DNS servers are configured on the switches to resolve the ASC Gateway address. When the ASC Gateway instances use the same DNS server, the DNS server must also resolve Brocade Support Link Server (bsnconnect.broadcom.com), and when ACC and DCA are implemented, bsnacc.broadcom.com and bsnsupport.broadcom.com must be resolvable.

The firewall must be configured to allow HTTPS outbound from the ASC-G instances to the SLS and ACC servers. The firewall must allow SFTP outbound from the ASC-G instances to the SupportSave server to upload SupportSaves with DCA. If a proxy is used for Internet access, the proxy configuration must be performed on the ASC-G instances.

Figure 38: BSL Deployment with the ASC Gateway



```
Brocade Support Link Server:  bsnconnect.broadcom.com
Automated Case Creation:      bsnacc.broadcom.com
Support Save Server:          bsnsupport.broadcom.com
```

FW Configuration
For ASCG1 & ASCG2

```
Brocade Support Link Server:  bsnconnect.broadcom.com  443   outbound
Automated Case Creation:      bsnacc.broadcom.com      443   bi-directional
Support Save Server:          bsnsupport.broadcom.com  22    outbound (SFTP)
```

DNS Configuration for ASCG

ascg.internal.domain 10.10.10.11 & 10.10.10.12
ascg1.internal.domain 10.10.10.11
ascg2.internal.domain 10.10.10.12

bsnconnect.broadcom.com	*192.19.232.240*
bsnacc.broadcom.com	*192.146.127.198*
bsnsupport.broadcom.com	*192.19.232.77*

SAN Switches
Configured with Internal DNS 10.10.10.101
Network connectivity to ASCG1 & ASCG2

For Autotomated Case Creation (ACC)
Configure SNMP Trap Recipients: ASCG1 & ASCG2

17.4 DNS Registration of ASC-G Instances

The ASC Gateways must be registered on the internal DNS server for the ASC-G switches to perform name resolution. The common name in this example is ascg.internal.domain, and it is used as a joint DNS record for the ASC Gateways and individual records for each instance to use DNS round-robin as a load-balancing mechanism. Use of a load balancer is also supported, in which case the joint DNS record must point to the load balancer VIP for the ASC-Gs.

Example DNS records for ASC Gateways: ascg.internal.domain10.10.10.11 & 10.10.10.12

ascg1.internal.domain 10.10.10.11

ascg1.internal.domain 10.10.10.12

In addition, the ASC-G instances must be able to resolve for bsnconnect.broadcom.com; for Automated Case Creation, bsnacc.broadcom.com; and for Data Collection Assistant, bsnsupport.broadcom.com.

17.5 Certificate Authority to Sign SSL Certificates

A signed SSL certificate is installed on the ASC-Gs during configuration for the switches to trust the ASC-Gs for HTTPS transfer. The certificate can be signed by a public CA or an enterprise CA; self-signed certificates from Fabric OS 8.2.2 are also supported. The SSL certificate and private key use a common name for both ASC-G instances; in this example, the FQDN `ascg.internal.domain` represents both ASC-G instances.

NOTE: When using an enterprise CA or self-signed certificate, the certificate must be installed on the switches to validate the signed certificate on the ASC Gateways.

When configuring for ACC, a certificate must be created on each switch before being trusted by the ASC-G instances. This procedure is performed during the onboarding process for ACC.

Refer to the *Brocade Active Support Connectivity Gateway User Guide*.

17.6 Hypervisor/VM/Server on Which to Install ASC Gateways

The ASC Gateway is available as both an OVA and an installable (package). Depending on the organizational policies for infrastructure components in your environment, either deploy the OVA version in a VMware vSphere environment or install the installable version on a server/VM provided by your infrastructure team. RHEL, Oracle Linux, and CentOS are supported.

Refer to the *Brocade Active Support Connectivity Gateway User Guide* for specific information for either option.

17.7 Enabling ASC on the Switches

ASC is enabled and configured on the switches using the `supportlink` command.

Whether the switches send directly to the Brocade Support Link Server or use the ASC Gateway, it is mandatory to specify a user name. The user name is an email address. Subsequently, users from the same email domain with access to the Broadcom Customer Support Portal (portal.broadcom.com) can be authorized to request and receive BPA reports. Use a common user name, for example, operators@domain.com, although it can simply be an admin's email address, for example, admin@domain.com, for all switches from the same organization.

In addition, use the User Group Tag to group switches that you want to appear in the same report. Tags can be based on location or another identifier for one or more fabrics.

NOTE: You can always request a BPA report, which includes *all* switches configured with the same user name.

Exh. RA-3
Docket UT-181051
Page 88 of 112

Enabling ASC on the switches includes configuring the frequency and time of day for ASC data collection. In general, the recommended settings are to send data daily between 1 AM and 4 AM or another time considered off hours.

Refer to the *Brocade Active Support Connectivity Gateway User Guide* for additional configuration details.

The switch will upload ASC data daily to the Brocade SLS with the ASC configuration complete. In addition, you can manually trigger ad hoc data collection and upload on demand.

Chapter 18: Automation

18.1 Overview and Purpose

Most, if not all, IT administrators have first-hand experience in managing the growing complexity of enterprise infrastructure, including SANs. According to a report from the Enterprise Strategy Group, “The cost and complexity of protecting and storing data are increasing, and IT leaders are responding with attempts to optimize better and automate storage—but they need better tools.”

Broadcom is uniquely positioned to spot and understand the impact of automation, helping organizations get more from their SAN infrastructure.

Broadcom offers a combination of SAN automation with RESTful APIs and a SAN management platform to help organizations drive greater efficiency from their SANs. Automation and efficiency are accomplished through a variety of means:

- Brocade SAN automation, provided with a multilayer architecture.
- RESTful API support on switches and management tools.
- Broadcom’s Ansible management framework, designed to eliminate repetitive tasks, simplify management, and orchestrate across the SAN infrastructure.

18.2 Motivation to Automate

The following are five reasons why organizations should embrace SAN automation:

- Reducing human error and streamlining operational processes have never been more crucial. As organizations move to digitize and adapt to new workloads, data availability, processing time, and agility in provisioning on-demand applications become the business’s life-blood. These new digitized workloads demand a more efficient and expedient infrastructure management approach, leaving no room for human error. As a result, storage administrators need to be freed from repetitive manual tasks such as configuration management, reporting, documenting inventory, and troubleshooting. Instead, IT organizations need SAN automation to help them automate and orchestrate repetitive tasks, significantly improve efficiency, and decrease the risk of operational mistakes.
- Demand for more accurate and more frequent infrastructure reports is on the rise. It is not just IT managers who crave information on storage performance, utilization, and forecasting; business stakeholders are also asking for and expecting this data on demand. No one wants to wait for a slot when storage administrators can allocate time to produce a report. This information should be available as frequently as business demands dictate—all at the click of a button. Automation provides this kind of responsiveness that traditional manual storage management processes cannot deliver. Still, it can also be customized so that all stakeholders get more accurate data aligned to their responsibility.
- SAN configuration management must be streamlined. With more enterprise applications demanding access to more data and virtual machines, deploying and configuring servers, storage, and the network has become more time-consuming and complex than ever. By streamlining SAN operations through automation, application provisioning workflows are simplified across hypervisor, network, and storage, delivering agility and responsiveness to meet dynamic business demands.
- IT service delivery is not always as responsive as the business demands. As organizations become increasingly reliant on world-class IT services for proactive, agile business decision-making, they must identify and eliminate bottlenecks to IT service delivery. These enhanced IT services must be delivered without hiring more storage administrators or boosting SAN-related CapEx spending. SAN automation is the only viable option to drive increased agility and closely align IT services with fast-changing business needs.

- Consistent configuration validation is a must. Manual configuration changes occur with greater frequency as enterprises diversify their IT architectures in general and their storage architectures specifically. SAN automation ensures the validation of consistent configuration parameters across the different SAN fabrics to facilitate troubleshooting of frequent alerts without reliance on manual intervention.

Broadcom automation solutions leverage RESTful APIs to facilitate solutions architecture, share best practices, and get to production faster.

18.3 Overview of the REST API

The FOS REST API is a programmable web service interface for Brocade Fabric OS that can manage Brocade SAN switches across a fabric. This API uses standard HTTP methods to perform Create, Read, Update, and Delete (CRUD) operations on the fabric configuration data. It provides an interface for provisioning, status, and validation operations using the YANG data model described in the YANG 1.1 RFC, but not the data store managed with NETCONF. An Apache webserver embedded in Fabric OS is used to serve the API.

The RESTful API approach lets you think of a network device as a webserver. Automation can send and receive transactions to or from a network device by using standard web-based tools just as it would send transactions to and from a website. Transactions of this nature mean that they happen over a secure socket using HTTP rules to handle the exchange. The data appears in XML or JSON depending on the RESTful API services implemented on the networking device.

To interact with a SAN (or other) device, you need to consult its RESTful API reference to learn, among other things, what “uniform resource identifiers” (URIs) you need to use. (Simply put, URIs are identifiers that can be used as part of a web address.) According to the documentation, the URI for accessing a list of zones in the active configuration is as follows:

```
GET <base_URI>/running/zoning/defined-configuration/
```

The model used to represent state and configuration information is expressed in a modeling language called Yang. Yang describes the structure of the different elements inside the model and describes whether each element is read-only or read-write. It describes the type of data that the element can hold, such as string or integer. It shows the relationship among various elements, the other nested elements they contain, their peer elements, and the parent elements that contain them. Here is a segment of the description of a zone in Yang:

```
list zone {

key "zone-name"; description

"List of the members in the zone. The members can be identified only as a WWN, domain,
index, or zone alias.";

leaf zone-name {

type zoning-name-type; description

"The zone name.";

}

leaf zone-type {

type zone-type-type; description

"The zone type. Not that target zone types cannot be created or modified (only deleted).";

}

container member-entry { description

"The zone member."; leaf-list entry-name {

type zone-member-type; min-elements 1; description

"List of the members in the zone. The members can be identified only as a WWN, domain,
index, or zone alias.";
}


leaf-list principal-entry-name {

when "../..zone-type=1 or ../../zone-type=2"; type zone-member-type;

min-elements 1; description

"List of the principal members in the peer zone. The members can be identified only as a
WWN, domain, index, or zone alias.";

}

}

}
```

Ordinarily, more information goes into a Yang module, such as revisioning and governance information; this listing omits them for brevity. Thus, the Yang description is complete, but it is also wordy. Although this precision is necessary when interacting with the model programmatically, it is sometimes helpful to get a global view of the abstraction provided by the model to see how the data is structured.

An open-source tool called pyang can parse the Yang model and produce a tree representing the model's elements. The listing includes information about each element, whether it is read-only or read-write, a list, optional, or nested. Here is the representation of the zoning model in tree form:

```
module: brocade-zone

+--rw brocade-zone

+--rw defined-configuration

|    +--rw cfg* [cfg-name]

|    |     +--rw cfg-name    zoning-name-type

|    |     +--rw member-zone

|    |     +--rw zone-name*  zoning-name-type

|    +--rw zone* [zone-name]

|    |     +--rw zone-name   zoning-name-type

|    |     +--rw zone-type?  zone-type-type

|    |     +--rw member-entry

|    |     +--rw entry-name* zone-member-type

|    |     +--rw principal-entry-name*   zone-member-type

|    +--rw alias* [alias-name]

|    +--rw alias-name  zoning-name-type

|    +--rw member-entry

|    +--rw alias-entry-name* union

+--rw effective-configuration

+--rw cfg-name?  zoning-name-type

+--rw checksum?  string

+--rw cfg-action?uint8

+--rw default-zone-access?   uint8

+--ro db-max?    uint32
```

Exh. RA-3
Docket UT-181051
Page 92 of 112

```
+--ro db-avail?  uint32


+--ro db-committed?    uint32
+--ro db-transaction?  uint32
+--ro transaction-token?     uint32
+--ro db-chassis-wide-committed?   uint32
+--ro enabled-zone* [zone-name]
+--ro zone-name  zoning-name-type
+--ro zone-type? zone-type-type
+--ro member-entry
+--ro entry-name*union
+--ro principal-entry-name*  union
```

18.4 Simple Automation Example

In this example, the <base_URI> is http://<our device IP address>/rest. Begin by creating a login session with a switch in the fabric by executing the following command (which you type as a single line):

```
curl -X POST -v -u admin:password http://10.18.254.37/rest/login
```

- `curl` is the command's name.
- `-X POST` specifies the POST HTTP method (instead of GET).
- `-v` specifies verbose output to access the authorization string in the header of the response used in the next step.
- `-u admin:password` specifies the credentials to use.

The last parameter is the uniform resource identifier (URI) for `curl` to use to log in. (The URI value is described in the RESTful API reference.)

This command establishes the session used for the following commands. The following is a trace of its execution:

```
*    Trying 10.18.254.37...
*    Connected to 10.18.254.37 (10.18.254.37) port 80
(#0)
*    Server auth using Basic with user 'admin'
   >  POST /rest/login HTTP/1.1
>    Host: 10.18.254.37
   >  Authorization: Basic YWRtaW46cGFzc3dvcmQ=
   >  User-Agent: curl/7.47.0
   >  Accept: */* >
<    HTTP/1.1 200 OK
<    Date: Wed, 31 Jan 2018 16:01:24 GMT
<    Server: Apache
<    Authorization: Custom_Basic YWRtaW46eHh4OjNkYTllZmM3NzMxYjk4OGU2ODg1YzZkMGRjNWJlM
zMyNjBhZDYxZThkOWQ2MWMxNzNiMGVlMjU3YmM2OTcyYjA=
< Cache-Control: no-cache
< X-Frame-Options: DENY
< Content-Secure-Policy: default-src 'self'
< X-Content-Type-Options: nosniff
< X-XSS-Protection: 1; mode=block
< Connection: close
< Transfer-Encoding: chunked
< Content-Type: application/yang-data+xml
```

Next, you perform a GET of the URI to return the current configuration using the Custom Basic value returned from the login for authentication:

```
curl -v -H "Authorization: Custom_Basic
YWRtaW46eHh4OjNkYTllZmM3NzMxYjk4OGU2ODg1YzZkMGRjNWJlM
zMyNjBhZDYxZThkOWQ2MWMxNzNiMGVlMjU3YmM2OTcyYjA="
http://10.18.254.37/rest/running/zoning/defined-configuration
```

By default, `curl` uses the GET method, so you do not need to specify it. `-H "Authorization: Custom_Basic YWR...jA="` is the authentication and session identifying string returned in the previous command. `-H` places the string into the GET request header as seen in the following trace:

```
*    Trying 10.18.254.37...
*    Connected to 10.18.254.37 (10.18.254.37) port 80
(#0)
   >  GET /rest/running/zoning/defined-configuration
   HTTP/1.1
   >  Host: 10.18.254.37
   >  User-Agent: curl/7.47.0
   >  Accept: */*
   >  Authorization: Custom_Basic YWRtaW46eHh4OjNkYTllZmM3NzMxYjk4OGU2ODg1YzZkMGRjNWJlM
   zMyNjBhZDYxZThkOWQ2MWMxNzNiMGVlMjU3YmM2OTcyYjA=
>
<    HTTP/1.1 200 OK
<    Date: Wed, 31 Jan 2018 16:09:39 GMT
<    Server: Apache
<    Cache-Control: no-cache
<    X-Frame-Options: DENY
<    Content-Secure-Policy: default-src 'self'
<    X-Content-Type-Options: nosniff
<    X-XSS-Protection: 1; mode=block
<    Connection: close
<    Transfer-Encoding: chunked
<    Content-Type: application/yang-data+xml <
<?xml version="1.0"?>
```

```
<Response>

<defined-configuration>

<cfg>

<cfg-name>CFG_FABRIC_A</cfg-name> <member-zone> <zone-name>CLUSTER1</zone-name>

<zone-name>Z_AIXHOST_FCS2_VMAX01_SN1234_9F0 </zone-name>

...

<alias> <alias-name>esx66_5d3d00</alias-name> <member-entry> <alias-entry-
name>10:00:8c:7c:ff:5d:3d:00 </alias-entry-name>

</member-entry>

</alias>

</defined-configuration>

</Response>

* Closing connection 0
```

The results appear as an XML data segment structured according to the Yang model's description, so it is crucial to have access to that model along with the RESTful API manual. The models can be found on GitHub as a repository among Broadcom's repositories at http://github.com/brocade/yang. The RESTful API manual can be retrieved from the Broadcom website. Having retrieved the zoning information from the fabric, you should close the session using the CLI command (the results are omitted to save space):

```
curl -v -H "Authorization: Custom_Basic
YWRtaW46eHh4OjNkYTllZmM3NzMxYjk4OGU2ODg1YzZkMGRjNWJlM
zMyNjBhZDYxZThkOWQ2MWMxNzNiMGVlMjU3YmM2OTcyYjA=" http://10.18.254.37/rest/logout
```

In FOS version 8.2.2 or later, the REST API session-less operation allows you to provide authentication credentials directly for each GET request. Essentially, the FOS REST API session-less operation completes the login, GET operation, and logout as one complete request. You can use only basic authentication formats for REST API session-less operation, including plain text or Base64.

The following example shows a GET request using plain-text authentication:

```
curl -u admin:password http://10.155.2.190/rest/running/brocade-media/media-rdp>"
```

18.5 Ansible as an Alternative

The previous section shows an example of an approach that uses a procedural methodology. The workflow starts at the beginning, executes a series of steps, and then terminates. Most traditional programs work this way.

Ansible takes a declarative approach. Rather than provide sequential steps, Ansible describes each host in an inventory. The description appears in a document called a *playbook*. For example, Ansible describes a host state where the application is already installed rather than providing steps to install a particular application. When you run the playbook, Ansible takes no action if the application is already installed. If the application is not installed, Ansible calls installation routines to bring the host into the desired state without requiring the administrator to write any specific steps.

In the realm of storage networks, the use of a declarative language means that you can describe switches and fabrics where, for example, a zone is already configured with the proper hosts and storage arrays. When you run the Ansible playbook, those zones are defined as needed, and the hosts and storage arrays are added to them if necessary.

With some other declarative automation utilities, installing an agent on each host that the utility manages is necessary. This agent retrieves the commands from a command center and runs them on the localhost. Ansible is unique in that it does not require agents. To make switch state changes, Ansible establishes a secure shell session to a proxy and sends it a small Python script. The script performs the necessary operations using the switch API and removes itself from the host.

You need two different skill sets to implement an Ansible solution successfully. First, you must understand the most common playbook operations. These operations are coded and installed for use by the playbooks. As vendors announce support for Ansible, they also provide script libraries for the most common tasks. Suppose there is a required task, but it is not available in the official Ansible distribution. In that case, the open-source community may provide code for that task in publicly available repositories.

Second, you must understand your business needs to provide ongoing playbook development. The person who maintains the playbooks does not need to be a programmer and does not need to know how remote system operations occur. That person needs to know only the desired outcomes and should be able to construct playbooks in YAML, the markup language used by Ansible. The following is an example of an Ansible playbook:

```
---

- hosts: edgeSwitches vars_files:

-../fos_passwords.yml gather_facts: False


tasks:


- name: run fos commands brocade_fos_command:

switch_login: “{{switch_admin_account}}” switch_password: “{{switch_password}}”
switch_address: “{{switch_ip_address}}” command_set:

- command: alicreate "SampleAlias1", "10:23:45:67:76:54:32:10"
```

The three dashes at the beginning are part of the YAML specification. The `hosts` section identifies automation target switches. You can keep sensitive information in a separate file, as demonstrated by the `fos_passwords.yml` line. The name of a variable in double braces such as `{{switch_password}}` indicates variable substitution. The variable file specified in `vars_files` tells where to find external variables.

18.6 SANnav REST API

SANnav Management Portal and SANnav Global View provide end-to-end visibility into enterprise SANs. These tools detect, analyze, and take action based on SAN behavior and performance, helping administrators get to the root of problems faster and remediate them fully. Storage administrators can troubleshoot across the storage fabric in as little as 30 seconds. These capabilities are unprecedented with any other shared storage infrastructure architecture.

The SANnav REST API provides functionality that complements the Fabric OS REST APIs. The SANnav REST API includes support for the following SANnav features:

- List fabrics managed by the SANnav server.
- List the seed and principal switch data for each fabric.
- List switches (members) in the fabric.
- Display FCR topology information for routing topologies such as edge-to-edge, backbone-to-edge, and edge-to-backbone.
- Configure event forwarding.
- Acknowledge or unacknowledge events.
- Display a list of filtered events.
- Search inventory.

18.7 Conclusion

SAN automation is a critical element in IT modernization and digital transformation. It helps organizations handle storage-related processes more efficiently without hiring more administrators or adding to the storage infrastructure CapEx budget. SAN automation is a high-leverage approach to turning network storage into a strategic asset.

Broadcom’s commitment to SAN automation, combined with its long-standing leadership in storage fabrics and technical innovation, makes it an ideal candidate for your IT infrastructure automation strategy.

Appendix A: Optical Cables

Table 6: Supported Distances Based on Cable Type and Data Rates

Speed Name	OM1 Link Distance 62.5-µm Core and 200 MHz*km	OM2 Link Distance 50-µm Core and 500 MHz*km	OM3 Link Distance 50-µm Core and 2000 MHz*km	OM4 Link Distance 50-µm Core and 4700 MHz*km	OS1 Link Distance 9-µm Core and ~Infinite MHz*km
1GFC	300	500	860	*	10,000
2GFC	150	300	500	*	10,000
4GFC	50	150	380	400	10,000
8GFC	21	50	150	190	10,000
10GFC	33	82	300	*	10,000
16GFC	15	35	100	125	10,000
32GFC	—	20	70	100	10,000

Table 7: LWL Optics Support (SFP+)

Transceiver Data Rate (Gb/s)	Distance (km)
4	4, 10, & 30
8	10, 25
10	10
16	10
32	10

Appendix B: Fabric Details

This appendix provides example checklists and tables that you can use to identify dominant factors, including facilities that will have an impact on the SAN design.

Table 8: Current Fabrics

SAN/Fabric	No. of Switches	Type of Switches	Total Ports	Domains	No. of Servers	No. of Storage Devices	Location	Notes
Fabric 1								
Fabric 2								
Fabric 3								
Fabric 4								
Fabric 5								

Table 9: Individual Fabric Details

SAN/Fabric	Domain Number	Serial Number	Model	Speed	WWN	IP Addresses	Brocade FOS/M-EOS Version	Notes
Switch 1								
Switch 2								
Switch 3								
Switch 4								
Switch 5								

Table 10: Device Details

Servers & Storage	Vendor	Model	WWN	Alias	Zone	OS Version	Application	Fabric/ Switches	Notes
Server 1									
Server 2									
Server 3									
Storage 1									
Storage 2									
Storage 3									

Exh. RA-3
Docket UT-181051
Page 99 of 112

Table 11: Metrics and Impact on SAN Design and Performance

Metric	Source	Impact
Servers in the SAN	Estimate/Brocade SAN Health	Normal operations
Host Level Mirroring	Estimate	Distance, ISL congestion, traffic levels
Clusters (MSFT, HACMP, NetApp) Average number of nodes Workload level	Estimate Estimate: High/Med/Low	In-band heartbeat, frame congestion, host fan-in, traffic isolation
Virtualization: VIO Server No. of servers Consolidation ratio	Estimate Estimate Estimate	Frame congestion, edge traffic increase per port, server fan-in on target ports, device latencies
Virtualization: VMware No. of VMware servers Consolidated ratio? Shared VMFS? DRS? RDM? I/O intensive?	Estimate Estimate Yes/No Yes (%)/No Yes (%)/No High/Med/Low Yes/No	Frame congestion, device latencies, and SCSI2 reservations

Exh. RA-3
Docket UT-181051
Page 100 of 112

Table 12: Consolidated SAN Snapshot

SAN Requirements Data (Complete for Each SAN)	
Fabric Information	
Target number of user ports per fabric	
Target number of total ports per fabric	
Target number of switches per fabric (number of switches/switch type, total switches)	
Number of fabrics	
Number of sites in environment	
Topology (core-edge, ring, mesh, other)	
Maximum hop count	
Expected growth rate (port count)	
Fabric licenses	
SAN Device Information	
Number/types of hosts and OS platforms	
Number/types of storage devices	
Number/types of tapes	
Number/types of HBAs	
Other devices (VTL/deduplication appliance)	
Total number of SAN devices per fabric	
Customer requirement for failover/redundancy, reliability of SAN (multipathing software utilized)	
Application Details	
SAN Application (Storage Consolidation, Backup and Restore, Business Continuance)	
Fabric management application(s)	
Performance	
Maximum latency (ms)	
Targeted ISL oversubscription ratio (3:1, 7:1, 15:1, other)	

Exh. RA-3
Docket UT-181051
Page 101 of 112

Table 13: Application-Specific Details

<table>
<tr><th colspan="3">Backup/Restore Infrastructure</th></tr>
<tr><td colspan="3"><b>Servers</b></td></tr>
<tr><td>System</td><td>OS Version, Patch Level</td><td>HBA Driver Version</td></tr>
<tr><td>Server 1/HBA</td><td></td><td></td></tr>
<tr><td>Server 2/HBA</td><td></td><td></td></tr>
<tr><td>Server 3/HBA</td><td></td><td></td></tr>
<tr><td colspan="3"><b>Backup Software</b></td></tr>
<tr><td>Vendor</td><td>Version</td><td>Patch</td></tr>
<tr><td colspan="3"><b>FC Switch</b></td></tr>
<tr><td>Vendor</td><td>Model</td><td>Firmware</td></tr>
<tr><td>Brocade</td><td></td><td></td></tr>
<tr><td colspan="3"><b>Storage</b></td></tr>
<tr><td>Vendor</td><td>Model</td><td>Firmware</td></tr>
<tr><td>Array 1</td><td></td><td></td></tr>
<tr><td>Array 2</td><td></td><td></td></tr>
<tr><td colspan="3"><b>Tape Library</b></td></tr>
<tr><td>Vendor</td><td>Model</td><td>Firmware</td></tr>
<tr><td>Library</td><td></td><td></td></tr>
</table>

NOTE: Keep a similar table for each application.

Table 14: Quantitative Analysis: Radar Maps

SAN/Storage Admin Concerns	Rank (1 is Low, 10 is High)	Notes
ISL utilization	8	Is traffic balanced across ISLs during peaks?
Switch outage	1	Have there been switch outages? If so, what was the cause?
Zoning policy	6	Is the zoning policy defined?
Number of switches in the fabric	10	Is the current number of switches a concern for manageability?
Scalability	6	Can the existing design scale to support additional switches, servers, and storage?
Redundancy	10	Is the existing SAN redundant for supporting a phased migration or firmware update?
Server: high availability	10	Does the cluster software fail over reliably?
Storage: high availability	10	Do the LUNs fail over reliably?
Available disk pool	6	Is there a sufficient disk pool to support additional apps?
Management tools for SAN	4	Are the right tools used for SAN management?
Application response	7	Have there been any instances of slow application response but no outage?

Appendix C: Terminology

Term	Brief Description
ACC	Automated Case Creation is part of the Brocade Support Link (BSL) suite.
ACL	Access control list is a list of permissions associated with a system resource (object).
AFA	All FLASH Array.
AG	Access Gateway is an FC switching product from Broadcom.
air gap	Air gap is the separation that creates true redundancy between the A & B fabrics in a SAN. Other than the hosts, storage, and management connections, there are no other connections between the A & B fabrics.
ARL	Adaptive Rate Limiting is a two-tier rate limiter. A minimum value and maximum value are configured. Always push at least the min; never push more than the max.
ASC-G	Active Support Connectivity Gateway is used with Brocade Support Link.
ASIC	Application-specific integrated circuit made by Broadcom for Brocade FC switching products.
base switch	Base switch of an enabled virtual fabric mode switch, providing a common communication infrastructure that aggregates traffic from logical switches in the physical switch in a common base fabric.
BBC	Buffer to Buffer Credit is part of the FC flow control mechanism.
BBFID	Backbone Fabric ID is an ID number identifying a specific backbone fabric used in FCR.
BET	Brocade Extension Trunking is a trunking feature specific to an Extension tunnel when it is comprised of more than one circuit.
BPA	Best Practice Assessment, part of Brocade Support Link.
BSL	Brocade Support Link.
BT	Brocade Trunk allows a group of ISLs to merge into a single logical link, enabling traffic to be distributed dynamically at the frame level. BT egress queueing reduces I/O response time.
BW	Bandwidth, as in link bandwidth or WAN bandwidth.
CA	Certificate Authority is an authorized entity that signs identity certificates.
certificate	A certificate provides authentication of the identity claimed. Within the National Security System (NSS) public key infrastructure (PKI), identity certificates may be used for authentication or for both authentication and digital signatures.
ClearLink Diagnostics	Diagnostics tool that allows users to automate a battery of tests to verify the integrity of optical cables and transceivers in the fabric.
CPI	Configuration, Performance, and Inventory (CPI) reports provide a comprehensive package of all configuration, inventory, and performance data.
CPU	A central processing unit (CPU), also called a central processor, main processor, or just processor, is the electronic circuitry that executes instructions comprising a computer program.

Exh. RA-3
Docket UT-181051
Page 104 of 112

CRC	A cyclic redundancy check (CRC) is an error-detecting code commonly used in digital networks and storage devices to detect accidental changes to digital data.
CRUD	Create, Read, Update, Delete.
DBR	Device-Based Routing routes flows based on SID/DID.
DCA	Data Collection Assistant is enabled with ASC-G as a centralized point for SupportSave triggering or end-user scheduling.
DCC	Device Connection Control is a FOS-based security feature.
DH-CHAP	Diffie-Hellman Challenge Handshake Authentication Protocol is an FC-SP protocol that provides authentication between switches and connected devices.
DID	Domain ID is the domain assigned to the switch or logical switch. It must be unique within the fabric.
DNS	Domain Name Service is a network service used to resolve human readable fully qualified domain names into an IP address or some other information.
D_Port	A Fabric Port configured in ClearLink Diagnostics testing mode for cable and optics integrity testing.
DP	Data processor is used for Extension to put data into circuits and tunnels, perform encryption and compression, and perform all other Extension functions.
DR	Disaster recovery is an organization’s plan to mitigate and recover from catastrophes to their infrastructure and staff.
DS	A default switch is a logical switch in a virtual fabric mode switch and is automatically created when Virtual Fabrics is enabled.
EBR	Exchange-Based Routing routes flows based on OxID/SID/DID
EFID	Edge Fabric ID is the ID number assigned to an edge fabric.
ELWL	Extended Long Wavelength is an SFP type that has very long distance capabilities.
E_Port	A standard Fibre Channel mechanism that enables switches to network with each other.
Extension	Extension is a technology used to transport FC (FCIP) or IP (IPEX) over an IP network in an optimized manner.
EX_Port	A type of E_Port that is a Fibre Channel router port and the demarcation point of fabric services. EX_Ports connect to edge fabrics.
FC	Fibre Channel.
FCAP	FC Authentication Protocol.
FCIP	Fibre Channel over IP enables Fibre Channel traffic to flow over an IP WAN.
FCP	FC Protocol.
FCR	FC Routing enables multiple fabrics to share devices without merging the fabrics.
FCS	Fabric Configuration Server is a Brocade FOS feature.
FDMI	Fabric Device Management Interface enables discovery of devices such as FC host bus adapters (HBAs).
FEC	Forward Error Correction is a technique used for controlling errors in data transmission over ultra-high-speed connections, unreliable transmission media, and noise-susceptible environments.
FICON	Fiber Connect is the IBM proprietary name for the ANSI FC-SB-3 Single-Byte Command Code Sets-3 Mapping Protocol for the FC protocol.
FID	Fabric ID is a unique identification number within the fabric.

FMS	FICON Management Server is a Brocade FOS feature.
FOS	Fabric Operating System, which runs on Brocade switches and directors.
FPI	Fabric Performance Impact is a Brocade FOS feature involving thresholds in MAPS.
FPIN	Fabric Performance Impact Notification.
F_Port	Fabric Port is a port in the fabric to which an N_Port is attached.
GE	Gigabit Ethernet.
HBA	Host bus adapter.
HDD	Hard disk drive.
HSRP	Hot Standby Router Protocol.
HTTP/HTTPS and SFTP	Hypertext Transfer Protocol and Hypertext Transfer Protocol Secure (HTTPS) are a combination of HTTP with the Secure Socket Layer (SSL)/Transport Layer Security (TLS) protocol. Secure FTP uses the same SSL/TLS protocol to secure FTP.
ICL	Inter-chassis links are connections between directors using the ICL ports.
IFL	Inter-fabric links are connections between a backbone fabric and edge fabricsing EX_Ports.
IOPS	I/Os per second, the rate of I/Os.
ISL	Inter-switch links are used for connecting switches and directors using E_Ports.
LF	Logical fabric, the fabric comprised of multiple logical switches in a virtual fabric.
IP	Internet Protocol (IPv4 or IPv6).
IP Extension (IPEX)	The encryption and optimization of IP storage transport using Extension.
IPEX GW	IP Extension Gateway is an interface on the LAN side of Extension to which IP storage sends its traffic.
KATOV	Keepalive Timeout Value used by Extension circuits to determine how long before going offline.
LAN	Local area network, typically within the data center.
LDAP	Lightweight Directory Access Protocol is an open, vendor-neutral, industry-standard application protocol for accessing and maintaining distributed directory information services over an IP network.
LLDP	Link Layer Discovery Protocol (LLDP) is a vendor-neutral link layer protocol used by network devices for advertising their identity, capabilities, and neighbors on a local area network based on IEEE 802 technology, principally wired Ethernet.
LS	Logical switch, part of a virtual fabric enabled switch, managed the same as physical switches.
LSAN	Logical SAN refers to the type of zoning used for FCR between devices in different edge fabrics.
LUN	A logical unit number is a number used to identify a logical unit, which is a device addressed by the SCSI protocol or SAN protocols that encapsulate SCSI, such as Fibre Channel or iSCSI.
LWL	Long Wavelength is an SFP capable of long distances, usually over single-mode fiber.
MAN	A metropolitan area network (MAN) is a computer network that interconnects users with computer resources in a geographic region of the size of a metropolitan area.
MAPS	Monitoring Alerting Policy Suite is a FOS feature that constantly monitors itself for potential faults and automatically alerts the user to detected problems.
MMF	Multi-mode optical fiber is a type of optical fiber mostly used for communication over short distances, such as within a building or on a campus. Multi-mode links can be used for data rates up to 100Gb/s.
MPIO	Multipath I/O is a fault-tolerance and performance-enhancement technique that defines more than one physical path between a computer system and its storage.

NAS	Network attached storage.
NIC	Network interface card.
NPIV	N_Port ID Virtualization is an FC feature whereby multiple FC Node Port (N_Port) IDs can share a single physical N_Port.
N_Port	Node Port is a port on the end device (host or storage).
NSID	NVMe Namespace ID.
NVMe	NVM Express (NVMe) or Non-Volatile Memory Host Controller Interface Specification (NVMHCIS) is an open, logical-device interface specification for accessing a computer's non-volatile storage media usually attached via a PCI Express (PCIe) bus.
OVA/OVF	An OVF package consists of several files placed in one directory. The entire directory can be distributed as an Open Virtual Appliance (OVA) package, which is a tar archive file with the OVF directory inside.
Oversubscription	A condition in which more devices might need to access a resource than that resource can fully support.
PBR	Port-Based Routing is routing based on the port numbers.
PG	A performance group is a set of virtual channels dedicated to traffic based on certain criteria.
port channel	A port channel allows grouping of several physical Ethernet links to create one logical Ethernet link for the purpose of providing fault-tolerance and more bandwidth to switches.
port group	A set of ports that have in common a particular function or feature, such as BT.
QoS	Quality of Service is a traffic queueing mechanism that prioritizes data based on the initiator and target zoning designation (high, medium, or low). QoS also applies to Extension traffic.
QSFP	Quad small form-factor pluggable is a four-lane SFP and can accommodate four times the speed.
RADIUS	Remote Authentication Dial-In User Service (RADIUS) is a networking protocol that provides centralized authentication, authorization, and accounting (AAA) management for users who connect and use a network service.
RAID	Redundant Array of Inexpensive Disks or Redundant Array of Independent Disks is a data storage virtualization technology that combines multiple physical disk drive components into one or more logical units for the purposes of data redundancy and performance improvement.
RASLog	Reliability Availability Serviceability Log is the logging facility on Brocade FOS-based platforms.
RBAC	Role-based access control (RBAC) or role-based security is an approach to restricting access to authorized users and authorizing specific functions based on their defined role.
RDR	Remote Data Replication is asynchronous replication that occurs between one data center and another over distance.
redundancy	Duplication of components, including an entire fabric, to avoid a single point of failure in the network (fabrics A and B are identical).
resiliency	The ability of a fabric to recover from failure; could be in a degraded state but functional (for example, an ISL failure in a trunk group).
REST and RESTful	Representational state transfer (REST) is a widely accepted set of guidelines for creating stateless, reliable web-based APIs. RESTful web APIs are typically based on HTTPS methods to access resources via URL-encoded parameters and the use of JSON or XML to transmit data.
RSCN	Registered state change notification (RSCN) is an FC fabric's notification sent to specific nodes in case of fabric changes.
RTT	Round trip time.

SAN	Storage area network, typically comprised of two or more fabrics.
SCC	Switch Connection Control is a Brocade FOS security feature.
SCSI	Small Computer System Interface is a set of standards for physically connecting and transferring data between computers, storage, and peripheral devices.
SDDQ	Slow Drain Device Quarantine is a Brocade FOS feature used to remove slow devices from the fast lanes and place them into their own slower traffic lane. SDDQ mitigates or eliminates head of line blocking.
SFP and SFP+	Small form-factor pluggable is a compact, hot-pluggable, network interface module used for both telecommunication and data communications applications.
SLA	A service-level agreement is an organizational agreement to maintain a specific level of infrastructure operational availability.
SMF	Single-mode fiber is an optical fiber designed to carry only a single mode of light. SMF fiber is associated with LWL and can carry light farther distances than MMF.
SNMP	Simple Network Management Protocol is an Internet standard protocol for collecting and organizing information about managed devices on IP networks and for modifying that information to change device behavior.
SSD	Solid state disk.
SSH	Secure Shell is a client-to-host protocol used to form a secure encrypted connection to the CLI.
SSL	Transport Layer Security (TLS) is the successor of the now-deprecated Secure Sockets Layer (SSL). SSL is a cryptographic protocol designed to provide communications security over a computer network.
SWL	Short Wavelength is associated with MMF (multi-mode fiber) and tends to carry light for short distances.
TCL	Traffic control list is part of the configuration of IP Extension. Matching traffic is assigned to a particular tunnel.
TCO	Total cost of ownership is the cost when accounting for all aspects of the business including acquisition, learning, staffing, operations, downtime, installation, maintenance, future replacement, product lifecycle, and so on.
TCP	Transmission Control Protocol (TCP) is one of the main protocols of the Internet protocol suite.
TIZ	A traffic isolation zone controls inter-switch traffic by creating a deterministic path for traffic flowing between a specific set of source ports. NOTE: The TIZ feature has been deprecated.
TLS	The Transport Layer Security protocol aims primarily to provide cryptography, including privacy (confidentiality), integrity, and authenticity through the use of certificates, between two or more communicating computer applications.
TLV	Type, length, value is an encoding scheme used for optional informational elements in certain protocols.
TO	Traffic Optimizer is a Brocade FOS feature used to sort traffic by characteristic into the most optimal traffic lanes to expedite delivery through the fabric.
UDP	User Datagram Protocol (UDP) is one of the core members of the Internet protocol suite. Applications send messages referred to as datagrams to other hosts on an IP network.
UI	A user interface can be a browser or some other interactive form with the application.
URI	Uniform Resource Identifier (URI) is a unique sequence of characters that identifies a logical or physical resource used by web technologies.
VC	Virtual channels create multiple logical data paths (queues) across a single physical link.
VE_Port	Virtual Expansion Port is specific to Extension. It is a type of E_Port that faces the WAN and connects to the IP network. VE_Ports connect to VE_Ports in remote data centers.

VF	Virtual Fabrics is a suite of related features that enable customers to create logical switches and logical fabrics.
VLAN	Virtual LAN is a logically separated LAN within the physical LAN devices. VLANs are assigned and identified by number.
VM	Virtual machine is a virtual server instance created on a physical server that may have multiple VMs running on it. Each VM can be constructed with its own set of characteristics.
VRRP	Virtual Router Redundancy Protocol (VRRP) is a networking protocol that provides for a Virtual IP (VIP) and Virtual MAC (VMAC) for a set of router gateways on a LAN. The VIP and VMAC can provide failover/failback from one GW to another.
WAN	Wide area network is a network that typically reaches beyond the metropolitan area and is IP based.
WWN	World Wide Name is a unique identifier used in storage technologies. There are WWNs for ports (pWWN or WWPN) and nodes (nWWN or WWNN). There may be other WWNs on a device as well.
UltraScale ICL	UltraScale inter-chassis link, used for connecting director chassis (Gen 6 and Gen 7) without using front-end device ports.
xISL	xISL is an ISL that uses tagging to connect two virtual fabric base switches. All logical switches in each chassis that have enabled the use of xISL can communicate across the physical xISL links forming logical fabrics.

Appendix D: References

D.1 Software and Hardware Product Documentation

FOS documentation is located in two different locations within Broadcom for Brocade products and software. The publicly facing content is located on Broadcom.com. The nonpublic documentation is located on the Broadcom Customer Support Portal (CSP). Please refer to the documentation for your particular Fabric OS release.

Broadcom.com (https://www.broadcom.com/products/fibre-channel-networking)

Contains all user guides, reference manuals, white papers, eBooks, product briefs, administrative guides, compatibility guides, and case studies.

- *Brocade Fabric OS Administration Guide*
- *Brocade Fabric OS Command Reference Manual*
- *Brocade Fabric OS MAPS User Guide*
- *Brocade Fabric OS Flow Vision User Guide*
- *Brocade Fabric OS Access Gateway User Guide*
- *Brocade Fabric OS Extension User Guide*
- *Brocade X6-4 Director Hardware Installation Guide*
- *Brocade X6-8 Director Hardware Installation Guide*
- *Brocade X7-4 Director Hardware Installation Guide*
- *Brocade X7-8 Director Hardware Installation Guide*
- *SAN Fabric Resiliency and Administration Best Practices User Guide*
- *Brocade SANnav Flow Management User Guide*
- *Brocade SANnav Management Portal Installation and Migration Guide*
- *Brocade SANnav Management Portal REST API and Northbound Streaming Reference Manual*
- *Brocade SANnav Management Portal User Guide*
- *Brocade SANnav Global View User Guide*

Broadcom Customer Support Portal (CSP) (https://portal.broadcom.com/group/support/docsafe/downloads)

Contains supported Fabric OS (FOS) software, supported SANnav software, Target Path selection guides, and release notes.

This content *requires a valid registered account*. It is available only to approved Brocade Direct Support (BDS and BSS) customers and authorized Brocade OEMs with valid entitlement on their products.

- *Brocade Fabric OS Release Notes*
- *Brocade Fabric OS Upgrade Guide*
- *Brocade Fabric OS Troubleshooting and Diagnostics Guide*

D.2 Compatibility, Scalability, and Target Path

Broadcom.com

- *Brocade Fabric OS 8.x Open Systems Compatibility Matrix*

Broadcom Customer Support Portal (CSP)

- *Brocade SAN Scalability Guidelines: Brocade Fabric OS v9.X*
- *Brocade Fabric OS Target Path Selection Guide*

D.3 Brocade SAN Health

- www.broadcom.com/sanhealth

D.4 Brocade Bookshelf

- *NVMe over Fibre Channel for Dummies*
- *Networking Next-Gen Storage for Dummies*
- *Brocade Mainframe Connectivity Solutions*
- *SAN Automation for Dummies*

D.5 Other

- *The SNIA Dictionary*
- *SAN System Design and Deployment Guide*

Revision History

53-1004781-03; May 2022

- Added updates for FOS 9.1.
- Reviewed, updated, and editied the design guide in its entirety.

53-1004781-02; September 1, 2020

- Added updates for FOS 9.0 and Gen 7.
- Added the “Automation” chapter.

53-1004781-01; November 23, 2016

- Initial release.