Redbooks

ibm.com/redbooks

IBM B-Type SAN Extension Platform Implementation Guide

Brian Larsen

Gary Marquard

Jon Fonseca

Mark Detrick

Tim Jeka

Storage

IBM

Redbooks

IBM Redbooks

IBM B-Type SAN Extension Platform Implementation Guide

October 2023

> **Note:** Before using this information and the product it supports, read the information in "Notices" on page xiii.

First Edition (October 2023)

This edition applies to IBM B-Type SAN Extension Platform.

This document was created or updated on October 13, 2023.

Contents

Figures

Tables

Examples

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at https://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

DS8000®	IBM Spectrum®	System z®
FICON®	IBM Z®	z Systems®
GDPS®	Redbooks®	z/OS®
HyperSwap®	Redbooks (logo) ®	z/VM®
IBM®	Resilient®	z/VSE®

The following terms are trademarks of other companies:

Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Other company, product, or service names may be trademarks or service marks of others.

Preface

This IBM® Redbooks® publication is focused on the IBM b-type SAN Extension platforms and their deployment within IBM storage solutions. IBM b-type Extension platforms provide long distance storage data transport over FCIP, FICON® over IP or IP Extension protocols.

The content within this book provides the reader an extensive look at the IBM b-type SAN Extension technology, an overview of supported IBM long distance solutions, infrastructure design and the best practices for migrating from older generation systems to the latest offerings.

The target audience of this book is network and storage administrators.

Authors

This book was produced by a team of specialists from around the world.

Brian Larsen joined Broadcom in July 1991 and has more than 35 years of professional experience in high-end processing, storage, disaster recovery, cloud, virtualization, and networking environments. Larsen is the Director of Partner Business Development and is responsible for solution and business development within all IBM divisions. In addition to his 14 years in Business Development, Larsen has held positions in sales, consulting, product management and solutions marketing.

Gary Marquard is currently a Systems Engineer for Brocade/Broadcom, delivering a broad range of storage networking solutions to Fortune 500 companies, has over 32 years of networking and storage networking experience and over 39 years overall in the IT industry. He joined Computer Network Technology (CNT) in 1991 as a Customer Support Engineer responsible for configuring, installing and troubleshooting channel extension networks for customers across the US. Through acquisitions he has worked for McDATA, Brocade, and now Broadcom as a Storage Networking Systems Engineer, responsible for understanding both technical and business objectives of the client, providing comprehensive end-to-end solution requirements definition, architecture, design, planning, and consulting. Gary began his career with Sperry/Unisys Corporation in 1984 as a mainframe Instruction Processor and Input/Output hardware specialist, responsible for troubleshooting and repairing systems both in the factory and on-site at customer data centers..

Jon Fonseca has over 20 years of professional experience in IP and opticalcommunications networking, storage, and disaster recovery solutions. As Field ApplicationEngineer, he has taken the lead in designing, implementing, and optimizing storage networksin both Open Systems and Mainframe environments around the world. Customers seek hisexpertise and consider him a trusted advisor. Johnny began his career in networking with theUnited States Marine Corps and held senior engineering positions at NTT America, Cisco,McDATA, and Brocade before joining Broadcom (via Brocade's acquisition in 2017).

Mark Detrick is a Principal R&D Engineer who has worked for over 20 years with Broadcom. Involved with many Fortune 500 large-scale IBM projects worldwide, Mark plays a consultative role in storage network design and deployment. Mark is an expert in many data center technologies, including mainframe SAN design and extension; distributed systems large-scale SAN design, FC Routing and extension; IP routing and Ethernet switching; and WAN technologies. Mark has over 30 years of experience in data center networking

environments. He understands Broadcom's ASIC technology and is a consultative resource within Brocade and externally to customers, OEMs, and resellers. Mark applies his in-depth technical knowledge of fabrics, protocols, flow control, TCP/IP, and large-scale SAN/LAN architectures. Mark is adept at next-generation data center solutions, including cloud and many types of virtualizations found in data centers.

Tim Jeka leads the technical and positioning effort of Brocade's storage networking businesswith IBM as an IBM North America Systems Engineer of Storage Networking. Tim isresponsible for driving technical awareness and positioning of Brocade's storage networkingproducts to the IBM field across the Americas. He joined Brocade in 2008, bringing over 34years of experience in Networking, Storage, and SAN technology with IBM. Before joiningBrocade, Tim was a Networking and Storage Area Networking RDS at IBM. Tim also heldTechnical Team leadership roles in IBM's Storage and Networking product divisions for over34 years with IBM

Thanks to the following people for their contributions to this project:

Vasfi Gucer, Henry Vo
IBM USA

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an IBM Redbooks residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

► Use the online **Contact us** review Redbooks form found at:

 ibm.com/redbooks

► Send your comments in an email to:

 redbooks@us.ibm.com

► Mail your comments to:

 IBM Corporation, IBM Redbooks
 Dept. HYTD Mail Station P099
 2455 South Road
 Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- ► Find us on LinkedIn:

  https://www.linkedin.com/groups/2130806

- ► Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/subscribe

- ► Stay current on recent Redbooks publications with RSS Feeds:

  https://www.redbooks.ibm.com/rss.html

1

Introduction

This IBM Redbooks publication is focused on the IBM b-type SAN Extension platforms and their deployment within IBM storage solutions. IBM b-type Extension platforms provide long distance storage data transport over FCIP, FICON over IP or IP Extension protocols.

The content within this IBM Redbooks will provide the reader an extensive look at the IBM b-type SAN Extension technology, an overview of supported IBM long distance solutions, infrastructure design and the best practices for migrating from older generation systems to the latest offerings.

This chapter has the following sections:

1

1.1 Cyber Resilient infrastructure

IBM's Cyber Resilient® infrastructure strategy is based on several elements to help IT infrastructure designers and planners to find a full range of solutions that cover the broad scope of cyber resiliency.

Cyber resiliency embraces the National Institute of Standards and Technology (NIST) security framework. This framework includes "five" dimensions of security with each dimension building on top of one another to create a holistic approach to eliminating threats. The "five" dimensions include: Identify, Detect, Protect, Recover and Response.

IBM Storage and b-type SAN and Extension platforms have strive to create solutions that integrate with higher levels of the NIST dimensions that assist with Identifying, Detecting and Responding to those threats. Through integration with IBM's software platforms like Q Radar and Safeguarded Copy early detection and identification of abnormal behavior can be recognized and actions can be taken to offset the risk. Working with Safeguarded Copy, multiple copies of data can be taken to help IT departments recover from a known good copy of data and help businesses get up and running as soon as possible.

The two NIST dimensions that the IBM Storage and b-type SAN and Extension platforms have directly focused on together is, Protect and Recover. IBM Storage is capable of replicating and recovering data between any data center sites using their flash arrays and tape subsystems. IBM flash arrays provide data replication through offerings like IBM Metro/Global Mirroring and Hyper Swap to meet very low Recovery Point Objectives (RPO) and Recovery Time Objectives (RTO). IBM tape storage can also address data protection by using their TS77xx VTL replication of block or object stores through their Grid (TCP/IP) replication solution.

There are challenges when replicating data over distance and one of those challenges is the latency (or time it takes) to move data from Site A to Site B and beyond. The second challenge is having a stable network (TCP/IP Ethernet network) that does not interrupt the flow of data. The IBM b-type SAN Extension platforms provide technology that has evolved over 30 years to offset and eliminate the impact of latency and faulty network challenges.

will review how an IT architect can create a long-distance infrastructure that supports a Cyber Resilient solution.

1.2 Data path resilience for long distance replication

The data path for a long-distance solution includes managing several protocols to transport a company's data intelligently, transparently, and securely. The data path may need to support Fiber Channel over IP (FCIP), FICON over IP or an IP extension transport or all three within the same infrastructure. IBM's b-type SAN Extension platforms have over three decades of experience in supporting all the protocols noted above, along with creating an intelligent and secure transport that ensures auto-recovery in a transparent manner.

Designing a long-distance data path is a critical element in making sure your data arrives at its destination. Having multiple paths between locations will provide redundancy and fail over in the event an outage occurs on one of the paths. This IBM Redbooks publication will review the best practices of designing and deploying long-distance transport. Figure 1-1 on page 5 will examine a more detailed look at the network paths and configuration layout.

The long-distance infrastructure design cannot be "sound" unless the foundation is "solid". The IBM b-type SAN Extension platforms provide that solid foundation. These platforms have

evolved into having on board security capabilities, hardened access, license validation and WAN encryption technologies.

All IBM b-type SAN Extension offerings are protected with:

► Hardened Fabric Operating Systems (FOS)
► Controlled access through validation systems built into the hardware
► Automated distribution of SSL certificates throughout the SAN

IBM b-type integrated security:

► Secure boot - Control processor validates integrity of FOS boot image
► Secure hardware - Establish hardware-based root of trust
► Secure software - Brocade Trusted FOS (TruFOS) Certificates ensure FOS authenticity and current entitlement.
► Secure licensing - Licensing with encryption to ensure that the licenses installed are legitimate with no tampering.

IBM b-type onboard encryption of IP WAN:

► WAN (IP) Encryption is a standard feature
► WAN Encryption is hardware based with no impact to performance.
► WAN Encryption can be enabled or disabled as needed.

In addition to the security elements mentioned previously in this section, the features included in the IBM b-type platforms provide the IT infrastructure designers flexibility when creating their data transport network.

The standard features included with these platforms include:

► eHCL (Extension Hot Code Load) - reduces downtime.
► WAN-Optimized TCP - accelerates TCP flows across the WAN.
► Brocade Extension Trunking (BET) - logically bundles circuits for bandwith scale and fail over.
► ARL (Adaptive Rate Limiting) - allows data flows to use idle WAN capacity.
► QoS for Extension Flows - prioritize flows across the WAN.

Each of these features are advanced in their ability to manage high performance storage data flows. Several of these are unique within the industry and provide the highest value possible to an IT Infrastructure Designer. Each of these advanced features will be reviewed in Chapter 3 of this book, you can also find content on these features in *IBM b-type Gen 7 Installation, Migration, and Best Practices Guide*, SG24-8497.

In "Planning for infrastructure refreshing and migration" on page 3 we will review the need to understand an aging long distance infrastructure and creating plans to refresh and migrate to the next generation platforms.

1.3 Planning for infrastructure refreshing and migration

IT infrastructure is always changing. The network will move to higher speed routers, servers and storage will be upgraded and security threats will need to be addressed. Application workloads will increase and that will require additional capacity. In addition to changes mentioned previously, the infrastructure components have a shelf life that needs to be tracked so service and support does not expire. IT infrastructure planners and designers need to take all this change into consideration and plan for refreshing and migrating to the latest technology for their long-distance storage replication solutions.

With the ever-increasing amount of data that needs to be protected and IP WAN networks moving from 1GbE to 10GbE, 25GbE, 50GbE, 100GbE, many older IBM b-type SAN Extension platforms may not support the increased workload speeds. In addition to not supporting the high-speed WAN's these platforms are reaching the end of their product life cycles. This IBM Redbooks section describes the general principles of refreshing and migrating a current environment to a new, modernized SAN Extension platform.

The IBM b-type SAN Extension offerings have evolved since the mid-1990's to it's latest Gen 7 offering that is based on a Condor 5 ASIC. The current SAN extension offerings are the IBM b-type SAN42B-R7, SAN18B-6 and SX-6 Extension blade for directors. This IBM Redbooks publication will focus on a SAN42B-R7 migration plan, but all the same principles can be applied to the other extension platforms.

1.3.1 Migration methodology and techniques

Replacing any part of your storage replication infrastructure can be highly stressful, especially if you already have a network in place that is functioning properly. Fortunately, changing from an existing IBM b-type SAN Extension platform to a newer model can help limit any migration-related anxiety. You can achieve greater performance, feature-function and security capabilities while working with a familiar platform that is built on the many generations of proven technology.

One of the most common inquiries about migrating to a new Brocade Extension platform is the question around "will I need to take an outage?" The short answer is "maybe." Users can avoid a full outage and essentially avoid any downtime if the proper methods for conversion are followed.

To remove nearly all the risk of taking a full outage, you should plan to stage and configure your new platforms to mimic or be able to accept the same configurations that you are using today. Of course, if you are going through a consolidation effort, increasing the number of IP WAN connections, or upgrading from a lower speed IP WAN to a higher speed IP WAN circuit, there will be additional steps needed to integrate those changes.

In general, the following migration methodology and techniques can be applied to any platform conversion strategy. In Chapters 5 - 7 of this IBM Redbooks, a more detailed description of best practices and configuration of the IBM b-type SAN extension platforms will be addressed.

In our migration scenario below, we will assume that you are doing a one-for-one replacement of an aging SAN Extension Platform to a new IBM b-type SAN Extension platform with no change to the Fiber Channel connections or IP WAN circuits. Using Figure 1-1 on page 5 as our example network to convert, we will walk through some guidelines for migrating the platforms.
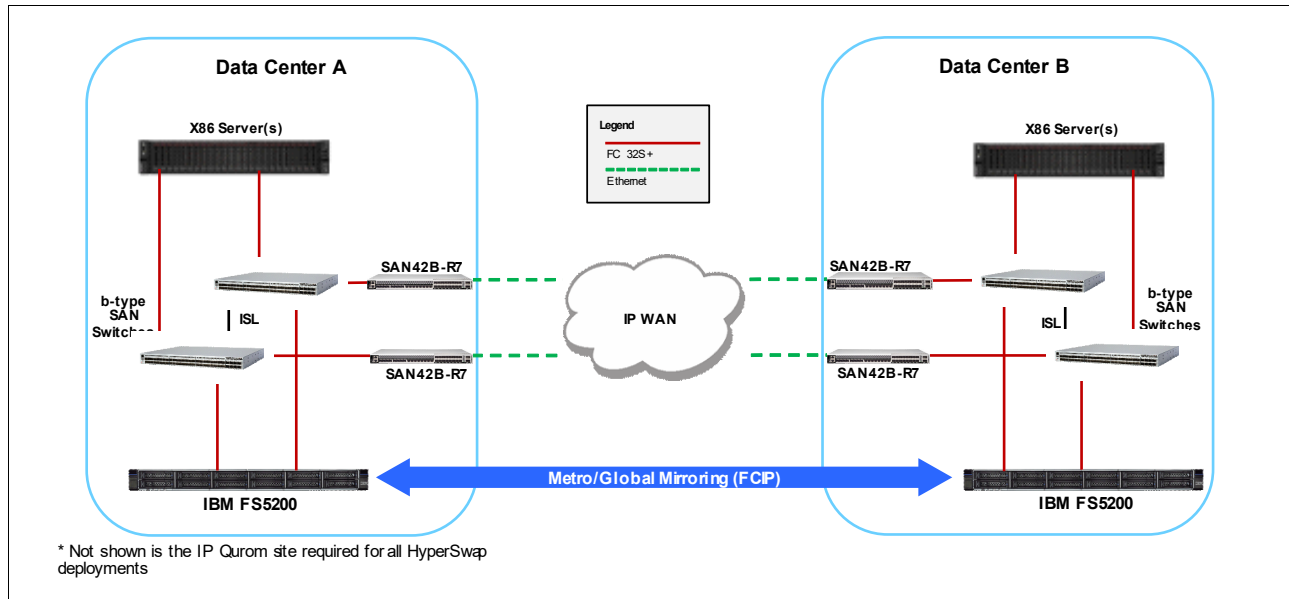
Figure 1-1   Example of a "Dual Fabric" Storage Replication Network

The "dual fabric" topology provides one of the most seamless transitions from one platform to another. By leveraging one of the fabrics (or paths), you can keep replication traffic flowing without taking a full outage.

1.3.2 General guidelines for migrating to a new SAN Extension platform

These are the general guidelines for migrating to a new SAN Extension platform:

► Preliminary work: Run the Brocade SAN Health tool to capture a snapshot of your environment and see if you have any current "network health" problems that need attention prior to the migration (see Chapter 6, "Extension refresh guidance" on page 53 on how to get started with SAN Health).
► Install and apply power to the new SAN42B-R7 in Sites A and B, leaving them disconnected from all storage or network cables.
► Gather information from the SAN42B-6s required for building and configuring the storage connections, IP interfaces, IP routes, tunnels, and circuits on the SAN42B-R7s.
► Configure the SAN42B-R7s on Site A and B with the configuration information gathered from the SAN42B-6s and disable power to all SAN42B-R7s. There are two ways to configure the SAN42B-R7s: through Brocade's SANnav management tool or through the CLI.
► Select which fabric you intend to migrate first, and if required by the storage vendor's process, suspend the path for the mirrored interface on the storage array. Review your storage vendor's administration guide for the proper process to take down one or more mirrored paths.

Note: With a dual fabric design, the other fabric/path will continue to replicate between the storage arrays but there may be some backlog in data replication since only half the bandwidth will be available.

► For the first path to migrate, disable power to the SAN42B-6 on Site A and Site B.
► Remove storage and network cables from each SAN42B-6 and reattach the cables to the SAN42B-R7s that are replacing them.

- ► Apply power to the new SAN42B-R7s in Site A and B and check the configurations and any error messages. Note: with the new path offline to the mirror, you can run the onboard traffic generator within the SAN42B-R7 to confirm that the IP WAN can support your bandwidth SLA.
- ► Once the entire path has been verified, re-enable the mirror (if required) for the path that was suspended.
- ► Monitor the mirror and network for errors and confirm that the network is running as expected.
- ► Once the initial fabric or path is working as expected, repeat these steps for the other fabric.

As you can see, by taking only one path down at a time, the administrator can reduce or eliminate the need for a complete outage.

To close out the guidelines for planning a migration from an older IBM b-type SAN Extension platform to a new version, it should be noted that a new software management platform is available. The Brocade SANnav management tool is the only management tool that supports the latest IBM b-type SAN Extension platforms. The Brocade SANnav offering has been completely re-architect with a new look and feel for the next generation of SAN infrastructure. SANnav will be reviewed in Chapter 9 of this IBM Redbooks.

In Chapter 2, "The IBM SAN42B-R7 and IBM SAN18B-6 Extension Switches" on page 7, we will provide a more detailed look at the IBM SAN42B-R7 and IBM SAN18B-6 offerings and later in this IBM Redbooks we take you through specific best practices for a long-distance network design and configuration steps in Chapter 5, "Extension best practices" on page 43 and Chapter 6, "Extension refresh guidance" on page 53.

2

The IBM SAN42B-R7 and IBM SAN18B-6 Extension Switches

This chapter provides a high level overview of the IBM SAN42B-R7 and IBM SAN18B-6 Extension Switches and contains the following sections:

- ► "IBM SAN42B-R7 and IBM SAN18B-6 product overview" on page 8
- ► "IBM and Brocade naming convention" on page 8
- ► "IBM SAN42B-R7 product description" on page 9
- ► "IBM SAN18B-6 product description" on page 11
- ► "Former b-series extension products" on page 13
- ► "Interoperability between IBM Extension Products" on page 14

2.1 IBM SAN42B-R7 and IBM SAN18B-6 product overview

The IBM SAN42B-R7 and the IBM SAN18B-6 are enterprise-class extension switches designed to provide cyber-resilient replication connectivity for enterprise storage that securely moves more data, faster over distance for continuous data protection over two protocols:

► Fibre Channel over IP (FCIP)
► IP Extension (IPEX)

These switches integrate easily into any existing IP network providing local performance over long distances with strong encryption, Brocade WAN-Optimized TCP with high-efficiency encapsulation to accelerate TCP, achieving the fastest replication speeds possible from storage devices, and ensuring in-order lossless transmission of data. They consolidate Fiber Channel, FICON, and IP storage traffic from heterogeneous devices for remote data replication (RDR), centralized backup, and data migration over long distances.

These solutions provide the following beneficial use cases:

► Data protection for both open systems and mainframe
► Multisite synchronous and asynchronous storage replication
► Accelerating IP storage across the WAN
► Operational excellence with converged bandwidth management of IP storage and FCIP across the WAN
► Enhancing the availability of FCIP and IP Extension by leveraging Extension Trunking across multiple WAN network paths
► Securing IP storage, Fibre Channel, and FICON data-in-flight across WAN infrastructures
► Centralized tape backup, recovery, and archiving for NAS, Fibre Channel, FICON, and IP-based backups
► Consolidation of replication I/O from heterogeneous arrays and multiple protocols

2.2 IBM and Brocade naming convention

Table 2-1 lists the b-type family products and their equivalent Brocade names. This publication refers to these switches using their IBM names.

Table 2-1   IBM and Brocade naming convention

IBM Name	IBM Machine Type Model	Brocade Name
IBM SAN42B-R7	8969-R42	Brocade 7850
IBM SAN18B-6	8960-R18	Brocade 7810
IBM SX6 Extension Blade	Feature Code #3892 & 3893	Brocade SX6 Blade
IBM SAN42B-R	2498-R42	Brocade 7840
IBM SAN06B-R	2498-R06	Brocade 7800

Only a feature code is given for the extension blade because the machine type and model is associated with the:

► IBM Storage Networking SAN512B-6 Model 8961-F08 (Brocade X6-8)

► IBM Storage Networking SAN256B-6 Model 8961-F04 (Brocade X6-4)

► IBM Storage Networking SAN512B-7 Model 8961-F78 (Brocade X7-8)

► IBM Storage Networking SAN256B-7 Model 8961-F74 (Brocade X7-4)

2.3 IBM SAN42B-R7 product description

The SAN42B-R7 is an enterprise-class product ideal for building a high-performance data center extension infrastructure for all replication and backup solutions. It leverages inter-datacenter WAN transport to extend open systems and mainframe storage applications over any distance, supporting Fibre Channel and FICON over IP (FCIP) and IP Extension, making it the ideal platform for all IBM storage solutions that require high performance, secure, inter-datacenter connectivity.

With twenty-four 64G FC/FICON ports, sixteen 1/10/25-GbE interfaces, and two 100-GbE interfaces, customers can achieve the bandwidth, port density, and throughput needed for maximum application performance over a wide area network (WAN). Additionally, it delivers strong security, continuous availability, unmatched scalability, and simplified operations to handle the most demanding requirements for business continuity and disaster recovery environments.

The SAN42B-R7 can be purchased as one of two models, depending on the environment and applications it will be primarily used in- Open Systems or Mainframe. Both units have all FC ports enabled, are licensed with full WAN rate capabilities with no restrictions, and have the same bundled shortwave Ethernet optics included (alternative Ethernet optics are available la carte).

The Open Systems model includes shortwave FC optics and the full software feature set with the exception of features that are specific to the mainframe.

The Mainframe model includes longwave FC optics and the full software feature set including Advanced Accelerator for FICON and FICON CUP. See Figure 2-1
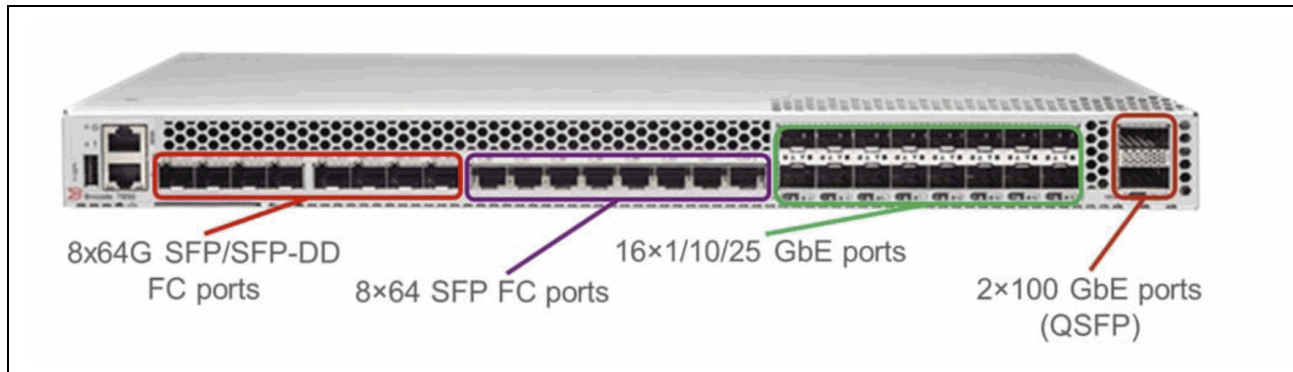
Figure 2-1   IBM SAN42B-R7 Portside

Table 2-2 provides a comparison of the hardware and software bundles associated with these two model types.

Table 2-2   SAN42B-R7 Model Bundles

Feature or SFPs (Small Form Pluggable) Included	SAN42B-R7 OS Model	SAN42B-R7 MF Model
64G SWL SFP	16	0
64G LWL SFP	0	16
10/1 GbE SR SFP	4	4
25/10 GbE SR SFP	4	4

100 GbE SR4 QSFP	2	2
Enterprise Bundle	Yes	Yes
Advanced Extension	Yes	Yes
Integrated Routing	Yes	Yes
Advanced Accelerator for FICON	No	Yes
FICON CUP	No	Yes

Note: For more than 16 FC ports, a la carte SFP-DD SWL optics are available. All Ethernet speeds covered with the bundled optics at short reach. For long reach Ethernet optics, ala carte 25/10 LR SFPs are available.

Note: There is no upgrade path from the Open Systems Model to the Mainframe Model.

2.3.1 IBM SAN42B-R7 Switch features and capabilities

The IBM SAN42B-R7 Switch offers the following features and capabilities:

- ► 1U form factor
- ► 16x physical Gen 7 FC ports (8x SFP and 8x SFP-DD)
- ► 24 total available 64G-capable ports
- ► 16x 1/10/25GbE ports
- ► 2x 100GbE QSFP ports
- ► Supports both FCIP and IP Extension
- ► Max WAN throughput: 100 Gb/s
- ► Supports compression, IPsec encryption, Brocade Extension Trunking, Adaptive Rate Limiting (ARL), Fabric Vision and WAN Test Tool
- ► Interoperable with SX6 Blade and SAN18B-6 platforms
- ► Supports Virtual Fabrics
- ► Open systems and FICON support
- ► Dual processing complexes, eHCL (Extension Hot Code Load) for WAN traffic
- ► Management through SANnav Management Portal, FOS, CLI and WebTools
- ► Security enhanced- Trusted FOS, hardware-based root of trust, secure optics, secure licensing

2.3.2 The IBM SAN42B-R7 Switch supported media types

The IBM SAN42B-R7 Switch supports the following media types:

Fibre Channel:

- ► 64G FC SFP+ LC connector (supports 16/32/64G): SWL, LWL, ELWL
- ► 32G FC SFP+ LC connector (supports 8/16/32G): SWL, LWL
- ► 2x64G FC SFP-DD SN connector (supports 16/32/64G): SWL

Ethernet:

- ► 10GbE SFP/SFP+ LC connector (supports 1/10GbE): SR
- ► 25GbE SFP/SFP+ LC connector (supports 10/25GbE): SR, LR
- ► 100GbE QSFP MPO connector (supports 100GbE): SR4

2.3.3 The IBM SAN42B-R7 Switch performance and scalability

Each IBM SAN42B-R7 switch contains two data processor (DP) complexes. DP complexes are synonymous with engines. Each DP complex contains a data processor (DP) attached to traditional switching ASICs, and consists of special purpose hardware for FCIP functions and multi-core network processors.

The switch can operate in one of two VE Modes with different VE port groupings.

6VE-Mode with 50Gbps Max tunnel bandwidth limit per group:

- ► DP0 Group 1: VE 24-26
- ► DP1 Group 1: VE 33-35

18VE-Mode with 16Gbps Max tunnel bandwidth limit per group:

- ► DP0 Group 1: VE 24-26
- ► DP0 Group 2: VE 27-29
- ► DP0 Group 3: VE 30-32
- ► DP1 Group 1: VE 33-35
- ► DP1 Group 2: VE 36-38
- ► DP1 Group 3: VE 39-41

SAN42B-R7 switch scalability and performance metrics:

- ► FCIP traffic flow to support 80 Gbps per DP FC ingress with fast-deflate and encryption
- ► Total platform FCIP throughput will be 160 Gbps with 2:1 compression ratio
- ► WAN traffic flow to support 100 Gbps total (50 Gbps per DP)
- ► IP-Extension to support 80 Gbps LAN ingress (40 Gbps per DP)
- ► Max tunnels per switch: 18
- ► Max circuits per tunnel: 10
- ► Max circuits per DP: 36
- ► Max IP-Extension TCP flows per DP: 512
- ► Max LAN Ports: 8
- ► WAN to support 250ms RTT @ 1.0% packet loss with keep-alive at 2 seconds or more
- ► WAN to support 200ms RTT @ 0.1% packet loss at all keep-alive timeout values

2.4 IBM SAN18B-6 product description

The SAN18B-6, as shown in Figure 2-2 on page 12, is a robust platform for medium-scale, multisite data center environments implementing block, file, and tape data protection solutions. The platform offers both Fibre Channel over IP (FCIP) and IP Extension technology, and is designed to handle simultaneous replication from Fibre Channel and IP storage arrays to consolidate replication workloads over WAN connections.

The SAN18B-6 offers enterprise-class capabilities to meet demanding disaster recovery requirements. With twelve 32 Gb/scapable Fibre Channel ports and six 1/10-Gigabit Ethernet (GbE) ports, this switch provides the bandwidth and throughput ideal for midrange and IP storage arrays, where native replication applications generally do not handle latency and packet loss well. Designed to be affordable, the SAN18B-6 offers flexible configurations, meeting current and future requirements.

The SAN18B-6 base model includes 4 (of 12) 32 Gbps FC ports enabled, 6 (of 6) 1/10 GbE FCIP ports, and total of 1 Gbps maximum of WAN Rate throughput. Software features included- Adaptive Rate Limiting, HW compression, FCIP FastWrite, WAN Optimized TCP, and IPSec.

The SAN18B-6 upgraded model activates the additional 8 (12 total) FC ports with and upgrades 1 Gbps WAN rate to 2.5 Gbps throughput and up to 10 Gbps of application throughput with 4:1 compression (dependent on data pattern). Upgraded software features- Advanced Extension, Enterprise Package (Fabric Vision, ISL Trunking and Extended Fabric) and Integrated Routing (IR).
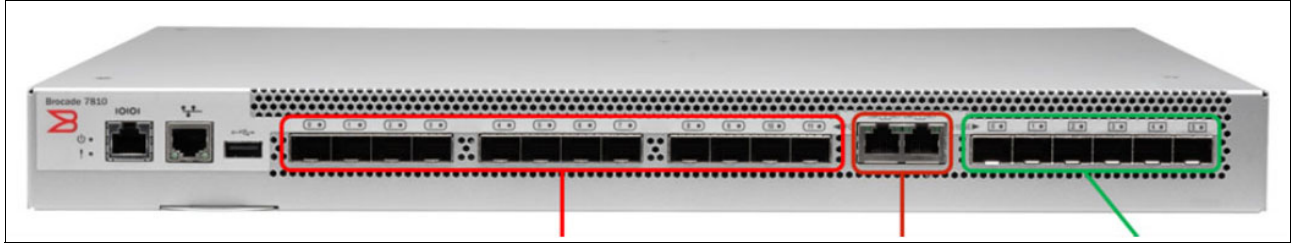
Figure 2-2   IBM SAN18B-6 portside

Table 2-3 provides a summary of the supported features and scalability limits of each model.

Table 2-3   SAN18B-6 model summary

Feature	SAN18B-6 Base Model	SAN18B-6 Upgraded Model
FC Port Limit	4	12
GE WAN Port Limit	2	6
GE LAN Port Limit	4	4
WAN Bandwidth Limit	1 Gbps	2.5 Gbps
Tunnel Limit	2	4
Circuits per Tunnel Limit	1	6
IPsec Support	Yes	Yes
Compression Support	Yes	Yes
Adaptive Rate Limiting	Yes	Yes
Fabric Vision	No	Yes
ISL Trunking	No	Yes
Extended Fabrics	No	Yes
Extension Trunking	No	Yes
FC Routing	No	Yes
Virtual Fabrics	No	No
FICON Support	No	No

Note: Qty (4) 16 Gbps FC SWL SFPs are included for the Base Model. Qty (12) 16 Gbps SWL SFPs are included for the Upgraded Model ordered from the factory. Optical GE SFPs are not included and need to be ordered a la carte.

Note: The upgraded model can be ordered from the factory or the base model can be upgraded in the field. Field Upgrade Kit includes Qty (8) 16 Gbps FC SWL SFPs

> **Note:** FICON traffic, and TS7700 Tape Grid are not supported on the SAN18B-6.

2.4.1 The IBM SAN18B-6 Switch features and capabilities

The IBM SAN18B-6 Switch offers the following features and capabilities:

- ► 1U form factor
- ► 12x physical Gen 6 FC ports (32G capable)
- ► 6x 1/10 GbE ports (GE0-GE5)
- ► 2x 1GbE copper ports (GE0-GE1)
- ► Supports both FCIP and IP Extension
- ► Max WAN throughput: 2.5 Gb/s
- ► Supports Compression, IPsec encryption, Extension, Trunking, ARL, Fabric Vision and WAN Test Tool
- ► Inter-operable with SX6 Blade and SAN42B-R7 platforms
- ► Supports open systems only
- ► Single processing complex, no eHCL (Extension Hot Code Load) for WAN traffic
- ► Management through SANnav Management Portal, FOS, CLI and WebTools

2.4.2 The IBM SAN18B-6 Switch supported media types

The IBM SAN18B-6 Switch supports the following media types:

Fibre Channel:

- ► 32G FC SFP+ LC connector (supports 8/16/32G): SWL, LWL

- ► 16G FC SFP+ LC connector (supports 4/8/16G): SWL, LWL

Ethernet:

- ► 1GbE SFP/SFP+ LC connector: SX, LX

- ► 10GbE SFP/SFP+ LC connector: SR, LR, USR

- ► 1GE SFP/SFP+ RJ-45 Copper: (CAT5 or higher)

2.5 Former b-series extension products

The IBM SX6 Extension Blade and the IBM SAN42B-R are two additional b-type extension products that are not covered in detail in this book. Both are deployed and used extensively in business continuance and disaster recovery environments and are still actively marketed at the time of this publication.

The IBM SAN06B-R was withdrawn from marketing (WFM) and will be reaching its end of service life on October 31, 2024. The recommend replacement product is the SAN18B-R.

At the time when the SAN42B-R reaches WFM, the SAN42B-R7 will be the recommended replacement product.

Table shows the former b-series extension products.

Table 2-4   Former b-series extension products

IBM Name	Generation	Available	Marketing Withdrawn	Service Discontinued
SX6 Extension Blade	Gen 6 (32 Gbps)	2016-10-11		
SAN42B-R	Gen 5 (16 Gbps)	2014-12-12		
SAN06B-R	Gen 4 (8 Gbps)	2009-11-20	2019-11-13	2024-10-31

> **Note:** The SX Extension Blade and SAN42B-R are covered in detail in the IBM Redpaper *IBM SAN42B-R Extension Switch and IBM b-type SX6 Extension Blade in Distance Replication Configurations (Disk and Tape)*, REDP-5404.

2.6 Interoperability between IBM Extension Products

The IBM SAN42B-R7, and IBM SAN18B-R use the same IBM Fabric OS that supports the entire IBM storage networking product family. This technique helps ensure seamless interoperability between Fibre Channel connected directors, switches and other extension switches. Fibre Channel specific advanced features such as Fabric Vision, ISL Trunking, Extended Fabrics and Integrated Routing are all fully compatible as well.

Interoperability between extension switches via extension tunnels requires more consideration because of the variances in FCIP data processing complexes used in the different generations of switches. At a high level Table 2-5 shows the compatibility between extension switches and blades that have not been withdrawn from marketing (WFM). Specific requirements and limitations are covered in detail in Chapter 5, "Extension best practices" on page 43 and Chapter 6, "Extension refresh guidance" on page 53..

Table 2-5   Extension technologies compatibility

Platform Connection	IBM SAN42B-R7	IBM SAN18B-6	IBM b-type SX6 Extension Blade	IBM SAN42B-R
IBM SAN42B-R7	Y	Y	Y	N
IBM SAN18B-6	Y	Y	Y	Y
IBM b-type SX6 Extension Blade	Y	Y	Y	Y
IBM SAN42B-R	N	Y	Y	Y

3

IBM SAN42B-R7 features

Extension provides optimized Fibre Channel over IP (FCIP) and IP Extension storage communications over IP networks between data centers. This section describes the features related to extension.

This chapter has the following sections:

3.1 The 3 sides of IBM b-type Extension

IBM b-type extension can be thought of as having three sides of connectivity. These three sides work together to provide FCIP (FC over Extension and FICON over Extension) and IP Extension (IP over Extension). The following sections describe each side:

► FC and FICON (within the data center)
► WAN (between the two data centers)
► LAN (within the data center)

3.1.1 The FC and FICON side

On the IBM SAN42B-R7, the FC side is a full-fledged IBM b-type switch, which runs the same Brocade Fabric OS as other Gen7 FC switches. Within IBM b-type extension platforms are Brocade FC application-specific integrated circuits (ASIC). Communication between F_Ports, E_Ports, and EX_Ports is identical to any other b-type switch of the same generation and Fabric OS version.

VE_Ports (FC and FICON side)

VE_Ports are logical E_Ports specific to extension. From the perspective of FC and FICON, a VE_Port is an E_Port and the endpoint of an ISL implemented over an IP-based tunnel. An extension connection between two VE_Ports creates an ISL between the domains and forms a fabric.

Zoning

Zoning is specific to the FC and FICON side, and end devices can be zoned across FCIP connections like any other fabric. A fabric can be set to either open or closed access as it pertains to end devices being able to communicate with or without being zoned. The best practice is closed access, which requires end devices to be zoned before communicating. Array-to-array replication ports should be zoned, which is best practice. In FICON environments, zoning is used and often matches the HCD (Hardware Configuration Definition).

Virtual Fabrics

Virtual Fabrics is an architecture to virtualize hardware boundaries. IBM SAN42B-R7 comes with Virtual Fabrics enabled. Traditionally, SAN design and management are done at the granularity of a physical platform. Virtual Fabrics allow fabric design and management to be done with port granularity.

From an extension perspective on the FC and FICON side, Virtual Fabrics applies to F_Ports, E_Ports, and VE_Ports.

3.1.2 The WAN side

FC is a data center technology not intended to traverse an IP WAN; extension was designed to address the specific intricacies of FC and FICON over an enterprise WAN. The extension WAN side connects to the IP network and has tunnels and circuits that extend across the WAN to a peer extension platform. This section focuses on the WAN side.

The WAN side consists of the following components and features of the IBM b-type extension:

► FCIP
► IP Extension
► IP WAN network

- ► VE_Ports
- ► Virtual Fabrics
- ► GE interfaces
- ► LLDP
- ► IP interfaces (IPIF)
- ► IP routes
- ► Tunnel
- ► Encryption (IPsec)
- ► Compression
- ► FastWrite
- ► OSTP (Open Systems Tape Pipelining)
- ► Advanced FICON Accelerator
- ► Circuits
- ► BET (Brocade Extension Trunking)
- ► ARL and CIR
- ► Extension Hot Code Load (eHCL)
- ► QoS
- ► Failover/failback (metrics)
- ► SLA
- ► KATOV

IBM b-type FCIP

IBM b-type FCIP is an ultrahigh-performance tunneling protocol to link FC and FICON over wide area IP networks. It is primarily used for disk remote data replication (RDR), tape backup, and migration. The extension link is an ISL transport for FC data and control frames between data centers. FCIP is supported on all IBM b-type extension platforms. IBM b-type extension tunnels are not interoperable with IBM c-type extension products. IBM b-type extension uses a proprietary tunnel protocol capable of moving FC, FICON, and IP storage across the same tunnel. See Table 3-1.

Table 3-1   FCIP Protocol

Network	WAN/MAN
Transport Stack	FC/Tunnel/TCP/IP/Ethernet
Encapsulation	Brocade HEE (High-Efficiency Encapsulation)
IP Routable	Yes

IBM b-type IP Extension

IBM b-type IP Extension is an ultrahigh-performance tunneling protocol to link IP over wide area IP networks. It is primarily used for data replication, tape backup, and migration. The extension link is an ISL transport for IP data between data centers. IP Extension is supported on all IBM b-type extension platforms. IBM b-type extension tunnels are not interoperable with IBM c-type extension products. IBM b-type extension uses a proprietary tunnel protocol capable of moving FC, FICON, and IP storage across the same tunnel. See Table 3-2.

Table 3-2   IP Extension Protocol

Network	WAN/MAN
Transport Stack	TCP/Tunnel/TCP/IP/Ethernet
Encapsulation	Brocade HEE (High-Efficiency Encapsulation)
IP Routable	Yes

IP WAN network

The IP WAN network must be capable of transporting data to the remote data center. It should offer multiple redundant, highly-available pathways. Requirements include adequate bandwidth for the data sent during peak loads.

Asynchronous and copy replication is forgiving, to a degree, if the IP network cannot accommodate the load. Short periods of insufficient bandwidth can be managed, provided extensive periods of sufficient bandwidth. In any case, insufficient bandwidth results in RPO elongation.

Synchronous replication is not forgiving. In the case of synchronous replication, the IP network must provide a level of service in which it can accommodate peak loads without congestion or packet drops; otherwise, application response times will suffer.

VE_Ports (WAN side)

VE_Ports are specific to extension, and a VE_Port may apply to two or three sides depending on the deployment. The three sides are FC/FICON, WAN, and LAN. VE_Port is a port in an FC or FICON fabric logical switch and a tunnel endpoint on the WAN side. Because extension platforms support multiple VE_Ports, multiple tunnels can be created. The number of tunnels depends on the platform's VE mode (6VE or 18VE mode).

Virtual Fabrics (WAN side)

Virtual Fabrics (VF) have different implications for the three sides. A VF logical switch separates VE_Ports, E_Ports, and F_Ports. On the WAN side, only VE_Ports are relevant. Putting a tunnel (VE_Port) into a logical switch (LS) isolates the tunnel from other tunnels (replication fabrics). Isolating a replication fabric limits aberrant IP WAN behavior to a small set of replication ports and creates a deterministic path for protocol optimization. Additionally, logical switches isolate replication FC ports from production FC ports. Such isolation is essential if a WAN connection drops and a set of FC devices have to log back into the replication fabric.

GE interfaces (WAN side)

A Gigabit Ethernet (GE) interface can be configured on the WAN or LAN sides. A WAN-side GE interface is used as a circuit portal. Circuits are mapped to GE interfaces using IP interfaces (IPIF). A WAN-side IPIF IP address is a circuit endpoint assigned to the desired GE interface and the DP of the VE_Port to be used. Each WAN-side GE interface can be used by circuits from multiple VE_Ports.

On the IBM SAN42B-R7, there are 1GbE, 10GbE, 25GbE, and 100GbE interfaces. GE interfaces are abstracted from the VE_Ports, which means that the GE interfaces are not fixed to a VE_Port. See Table 3-3.

Table 3-3   GE interfaces and speeds per platform

Platform	GE (SFP)	10GE (SFP+)	25GE (SFP28)	40GE (QSFP)	100GE (QSFP28)
IBM SAN18B-6	2: ge0-ge1 (RJ45) 6: ge2-ge7	6: ge2-ge7			
IBM SAN42B-R7	16: ge0-ge15	16: ge0-ge15	16: ge0-ge15		2: ge16-ge17
IBM SX6 Extension Blade	16: ge2-ge17	16: ge2-ge7		2: ge0-ge1	

LLDP

Link Level Discovery Protocol (LLDP) is supported on the extension GE interfaces of the IBM SAN42B-R7 extension platform. It is a vendor-neutral, open protocol referred to as IEEE 802.1AB. LLDP is a Layer 2 protocol that exchanges device identity and port numbers with the peer. LLDP applies to all extension GE interfaces on both the WAN and LAN sides.

> **Note:** LLDP works with the management port starting with FOS 9.2.0.

LLDP ensures connectivity to the intended port and device. Additionally, LLDP aids in troubleshooting to ensure the Ethernet link is functioning. If an LLDP exchange stops due to a physical layer problem (cable, optics, or configuration), the LLDP port entry times out and is removed from the list. LLDP is enabled by default, and preset global parameters are applied to all interfaces.

IP interfaces (WAN side)

An IP interface (IPIF) is the endpoint of a circuit. You must configure an IPIF at each end for each circuit you want to create. IPIFs are created before circuits are configured; therefore, plan the circuits and their associated IPIFs, GE interfaces, and VE_Ports. Depending on your intentions to use eHCL, you may also need to configure eHCL IPIFs.

Each IPIF is associated with a GE interface and DP. The IBM SAN42B-R7 has two DPs (DP0 and DP1).

Part of configuring an IP address to an IPIF is assigning the GE interface it will operate on and, in turn, the interface the circuit will use. GE interface assignment is essential because circuits physically connect to specific data center LAN switch ports. Port, optic, cable, data center switch redundancy, and resiliency are essential considerations for each circuit's physical connectivity. Tunnels span multiple physical connections by using multiple circuits assigned to different connections. Circuits can be assigned to any physical connection. One circuit cannot span more than one physical connection.

IP routes (WAN side)

From the perspective of the platform being configured, a WAN-side IP route is based on the remote IP address of a circuit. An IP route must be configured to reach the local gateway if the remote address is not on the same subnet as the local IP address. The IP route consists of the destination subnet and the local IP gateway. The local IP gateway, not to be confused with an IP Extension Gateway, is on the same subnet as the local WAN side IPIF IP address. When creating a route, specify the destination IP subnet with mask length and the gateway address. The gateway is a router interface that forwards traffic toward the destination.

Tunnels

An extension tunnel is an ISL that transports data between extension platforms. Extension tunnels allow FC and IP storage traffic to pass through a WAN optimally. A tunnel is the basis of an extension connection, composed of one or more circuits between two extension endpoints. Whether FC or IP storage, applications are unaware of the intermediate IP network. End devices do not experience WAN-associated data transmission reliability or latency issues. An extension tunnel ensures lossless transmission and in-order delivery when multiple circuits are deployed. The WAN side uses High-Efficiency Encapsulation (HEE), WAN Optimized TCP, protocol optimization (FastWrite and OSTP), encryption, compression, and QoS to transport data from the FC and LAN sides to the corresponding side at the remote end.

Encryption (IPsec)

IPsec (Internet Protocol Security) of in-flight data employs cryptographic security to ensure private, secure communications over IP networks. IPsec supports network-level data integrity, data confidentiality, data origin authentication, and anti-replay protection. It helps secure extension traffic against network-based attacks from untrusted entities.

Using IPsec is recommended for securing data transmission in a network that crosses the confines of a secure data center. IPsec is enabled at the tunnel level, not the circuit level, which means all of a tunnel's circuits are encrypted and use the same SA (security association). Different tunnels can have unique IPsec settings. IPsec uses Internet Key Exchange (IKE) to set up the SA. The key exchange can be through a pre-shared key (PSK) or a public-key infrastructure (PKI).

The following list contains the key features of AES-GCM-ESP:

► AES provides encryption with 256-bit keys.
► GCM provides data integrity with a 128-bit integrity check value.
► Peer platforms use IKEv2 key exchange for key agreement.
► IKEv2 uses UDP port 500 to communicate between platforms.
► IKEv2 traffic is protected using AES-256 encryption.
► HMAC (hash message authentication code) is a 192-bit or 128-bit GCM that checks data integrity and man-in-the-middle tampering.
► PRF (pseudo-random function) generates multiple security keys using 384-bit or 512-bit HMAC.
► MODP (modular exponential) is a 2048-bit or 384-bit Elliptic Curve Diffie-Hellman group used for IKEv2 and IPsec key generation.
► After the key lifetime of 4 hours or 2 billion frames has been reached, new keys are generated, and the old keys are discarded; rekeying is non-disruptive.
► When an SA expires, a new key is generated, limiting the amount of time or quantity of data an attacker has to decipher a key.
► Depending on the time expired or the number of frames being transferred, parts of a message might be protected by different keys generated as the SA life expires.
► ESP (Encapsulating Security Payload) is used in transport mode.
► ESP uses a hash algorithm to calculate and verify an authentication value, encrypting only the TCP header and its payload.
► Circuits from a non-secure tunnel can use the same Ethernet interfaces as circuits from a secure tunnel.
► IBM b-type IPsec is a hardware implementation (FPGA) that does not degrade or impact performance.
► IBM b-type IPsec does not preclude the use of compression or QoS.
► AES-CBC does not have an integrated integrity algorithm; therefore, HMAC-384-192 provides integrity.

Compression

IBM extension platforms provide an advanced compression architecture that supports multiple algorithms to optimize compression ratios at various throughput requirements. The available algorithms include hardware-based (FPGA) compression and software-based (HW-assisted processor) compression. The available compression modes depend on the platform and the protocol (FCIP or IP Extension).

Compression options

The following are the compression options:

► None

Compression is disabled. No data will be compressed.

► Fast-Deflate

Hardware-implemented compression. Fast-Deflate compresses data upon entering the DP and on the opposite side decompresses it upon leaving. It provides the highest throughput with the least amount of compression ratio. Because Fast-Deflate is hardware-based, it has the least propagation delay and is most appropriate for synchronous applications.

Note: The IBM SAN18B-6 does not support Fast-Deflate.

► Deflate

Hardware-assisted compression. Deflate mode can process FC or IP Extension data. Deflate's ingress speed (pre-compression) operates at a lower throughput than Fast-Deflate and faster than Aggressive-Deflate. Deflate provides more compression than Fast-Deflate and less than Aggressive-Deflate.

► Aggressive-Deflate

Hardware-assisted compression. Aggressive deflate mode can process FC or IP Extension data. The ingress speed (pre-compression) of Aggressive-Deflate has the lowest throughput of the algorithms. Aggressive-Deflate provides the most compression of the algorithms.

Protocol optimization

Protocol optimization features on the IBM b-type extension include the following:

► FCIP-FastWrite
► FCIP-OSTP (Open Systems Tape Pipelining)
► Advanced Acceleration for FICON
► Teradata Acceleration

FastWrite

FastWrite is an FCIP algorithm that reduces the number of round trips required to complete a SCSI write operation. FastWrite can maintain throughput levels over links that have significant latency. An RDR (Remote Data Replication) application still experiences latency; however, reduced throughput due to that latency is minimized for asynchronous applications, and response time is reduced by up to 50 percent for synchronous applications.

FastWrite is not used if the SCSI implementation uses similar techniques to achieve the same outcome. IBM replication products, like Global Mirror, do not engage FastWrite, even if FastWrite has been enabled for other traffic flows.

OSTP (Open Systems Tape Pipelining)

OSTP (Open Systems Tape Pipelining) is an FCIP algorithm that enhances open systems SCSI tape read and write I/O performance. The WAN has the most latency in the network, and OSTP provides accelerated tape read and write I/Os across the tunnel. OSTP accelerates tape read and write I/Os based on its sequential nature by reducing the number of round-trips required to complete the exchange.

Advanced Accelerator for FICON

FICON Acceleration provides specialized data management techniques and automated intelligence to accelerate FICON tape reads, writes, and IBM Global Mirror for System z® (XRC) replication while maintaining the integrity of command and acknowledgment sequences.

> **Author Comment:** Teradata Acceleration is not covered.

Circuits

Extension circuits are connections within a tunnel, and data flows between the source and destination VE_Ports. A circuit connects a local and remote IPIF (IP interface). Each tunnel contains at least one circuit; however, a tunnel can comprise multiple circuits. When there is more than one circuit, it is called a Brocade Extension Trunk (BET). Within each circuit are multiple WAN-optimized TCP sessions.

BET (Brocade Extension Trunking)

BET is an advanced IBM extension WAN side feature that benefits FCIP and IP Extension. BET enables bandwidth aggregation, and it is logically a single ISL. It performs Lossless Link Loss (LLL), guarantees in-order delivery, and lossless failover/failback for non-disruptive resiliency over the WAN. It provides redundant paths and manages load balancing.

BET is automatically enabled by creating a tunnel with multiple circuits. The tunnel uses multiple circuits to carry data between the source and destination data centers. BET does not apply to the LAN side. BET provides all the functionality of a port channel and more; it is used instead of LAG (Link Aggregation) on the WAN side.

Adaptive Rate Limiting (ARL) and Committed Information Rate (CIR)

An ARL or CIR configuration is required on every circuit to maintain, adjust, and limit the rate at which data is transmitted into the WAN. Oversubscribing the capacity of the WAN results in an egress buffer overflow and, ultimately, packet drops. Packet drops with TCP manifests into degraded performance. ARL and CIR are designed to minimize packet drops and optimize performance.

ARL utilizes WAN-Optimized TCP to determine and adjust circuit rate limits dynamically. ARL adjusts quickly to ever-changing conditions in the IP WAN network. ARL dynamically increases the rate limit up to the maximum until either it reaches the maximum or detects that no more bandwidth is available, whichever comes first. If it detects that no more bandwidth is available and ARL is not at the maximum, it periodically tests for more bandwidth. If ARL determines that more bandwidth is available, it continues to increase toward the maximum. On the other hand, if congestion events are encountered, ARL reduces the rate based on the selected backoff algorithm. ARL never attempts to exceed the maximum configured value and reserves at least the minimum configured value; everything in between is adaptive.

If the min and max are configured equally, this is called CIR.

eHCL (Extension Hot Code Load)

On the IBM SAN42B-R7 and the IBM SX6 Extension Blade, eHCL allows the non-disruptive Fabric OS firmware updates on circuits that have been configured for HA; FCIP and IP Extension traffic continue to flow during a firmware update.

Mainframe environments benefit from eHCL by supporting error-free, nonstop FICON connectivity for replication (Mirroring over FCIP) and tape (Grid over IP Extension). eHCL maintains extension connectivity across the WAN during firmware updates without disrupting active I/O, data loss, and causing out-of-order data delivery.

The IBM SAN42B-R7 and IBM SX6 Extension Blade have two data processor (DP) complexes, referred to as DP0 and DP1. A firmware update occurs on one DP at a time. eHCL does not apply to the IBM SAN18B-6 because it has only one DP. During the eHCL

process, each DP reloads sequentially while the other DP remains operational. When initiated, the update always starts on DP0. Before DP0 is updated to the new firmware, all eHCL-enabled circuits move to DP1 to maintain communication.

The failover process guarantees lossless and in-order delivery of data. eHCL uses three tunnel groups to perform non-disruptive updates. Consider the primary data center as local and the DR site as remote. The perspective in this section stays location consistent. The example assumes that the tunnel is from DP0 to DP0; however, tunnels can be created from either DP to either DP. The local backup tunnels (LBT) and remote backup tunnels (RBT) are automatically created when configuring an eHCL circuit by adding the HA IP addresses. A tunnel does not require that all circuits within it be enabled with eHCL; some circuits might be eHCL enabled while others might not.

> **NOTE:** eHCL is not supported on the IBM SAN18B-6 Extension Switch; however, the SAN18B-6 does support the configuration of remote HA addresses for peer eHCL support when connected to an IBM SAN42B-R7 or an IBM SX6 Extension Blade.

QoS (Quality of Service)

QoS refers to policies for expediting particular traffic flows over other flows. These policies are based on application delivery requirements. For example, email tolerates delays and dropped packets, but real-time voice and video do not. Some storage replication applications are more sensitive to delay than others. QoS settings provide a framework for accommodating these differences in data as it passes through the extension tunnel or IP network.

IBM extension has several Quality of Service (QoS) features. QoS on these platforms has two general categories of functionality. First, there is a two-tier classification into the protocol (FCIP and IP Extension) and priority (high, medium, and low). The second category is traffic marking with DSCP and L2CoS, which the IP network uses to prioritize flows.

There are two QoS tiers for WAN bandwidth allocation: protocol distribution and priority ratio. Each extension protocol (FCIP and IP Extension) has a QoS distribution. By default, each protocol receives 50% of the bandwidth. The distribution for a protocol can be set as low as 10% or as high as 90%, and the total must equal 100%. If a protocol path is utilized, the configured percentage of bandwidth is reserved. The other protocol cannot use the reserved bandwidth. If a protocol path is not utilized, for example, FCIP is used, and IP Extension is not, then the IP Extension bandwidth is not reserved, and FCIP can consume all the bandwidth regardless of the configured settings. The same is true if IP Extension is used and FCIP is not.

Figure 3-1 on page 24 shows the internal architecture of TCP connections that handle Priority TCP QoS (PTQ). The figure illustrates a tunnel containing a single circuit. Hybrid mode provides three QoS priorities for each FCIP and IP Extension, six QoS priorities, plus one for Class-F control traffic.
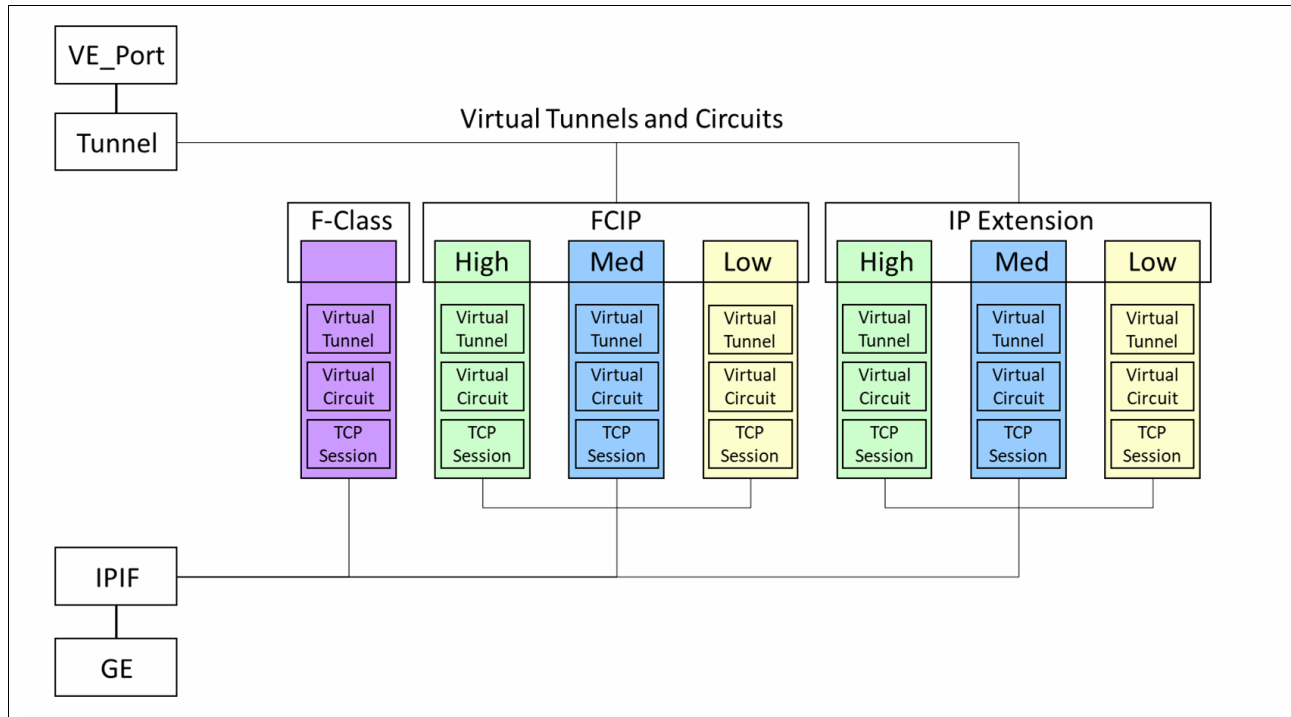
Figure 3-1   Virtual tunnels and circuits

Circuit failover

By providing a lossless link loss (LLL), BET ensures that all lost data in flight is retransmitted and reordered before being delivered to the upper-layer protocols. LLL is an essential feature to prevent interface control checks (IFCCs) on mainframes that use FICON and SCSI timeouts for open-system-based replication.

Multiple extension tunnels can be defined between pairs of extension switches or blades, but this defeats the purpose of defining a single tunnel for multiple circuits. Two tunnels between switches or blades are less redundant and fault-tolerant than multiple circuits within one tunnel.

Failover metrics

Failover circuits and groups are a feature supported on all IBM b-type SAN extension platforms. A group defines a set of metric-0 and metric-1 circuits. Each group operates autonomously. Metric-1 circuits provide failover protection for the metric-0 circuits. All metric-0 circuits in a group must fail before the metric-1 circuits become active. Typically, one metric-0 and one metric-1 circuit is configured per group, which provides one-to-one circuit protection. Circuits in a failover group must belong to the same tunnel. Failover and failback use LLL (Lossless Link Loss), which means no data is lost or delivered out of order during failovers and failbacks.

A circuit's Keepalive Timeout Value (KATOV) determines failed connectivity when a circuit goes offline. When a circuit goes offline in a Brocade Extension Trunk (BET), the data takes an alternative circuit in the trunk. In failover and failback, the data takes a metric-1 circuit. Both methods prevent interruption by seamlessly switching circuits that have gone offline.

Failover groups

The purpose of failover groups is to group circuits so that they can back up each other if one fails—a circuit failover group controls which metric-1 circuits are activated when which

metric-0 circuits fail. A specific set of metric-0 and metric-1 circuits are grouped. When all metric-0 circuits within the group have gone down, metric-1 circuits in that group become active. As each failover group is autonomous, circuits with metric 0 in other failover groups are unaffected.

For example, two circuits (metric-0 and metric-1) in one group would form a one-to-one backup. Failover groups are numbered; group ID is a value from 0 through 9, and the default is 0.

Typically, one metric-0 circuit and one metric-1 circuit are grouped. This aims to immediately replace the offline metric-0 circuit with a corresponding metric-1 circuit, avoiding a degraded tunnel. The remaining metric-0 circuits continue to operate as usual.

The following two supported configurations:

► Active-Active: This configuration is known as BET (Brocade Extension Trunking). Data is balanced across circuits based on each circuit's bandwidth weighting. All circuits are metric 0 by default.

► Active-Passive: This configuration is known as circuit failover. If all metric-0 circuits go down within a failover group, data fails over to the metric-1 circuits. If there is a failover/failback event, LLL guarantees that all data is delivered and delivered in order.

In Figure 3-2 circuit-0 is assigned a metric of 0, and circuit-1 is assigned a metric of 1. Both circuits are in the same tunnel and failover group (group 1). Circuit-1 is inactive until all metric-0 circuits within group-1 go offline. If all metric-0 circuits go offline within the failover group, traffic is transmitted and sent over available metric-1 circuits within the failover group. Failover between circuits is lossless; all lost data is retransmitted.
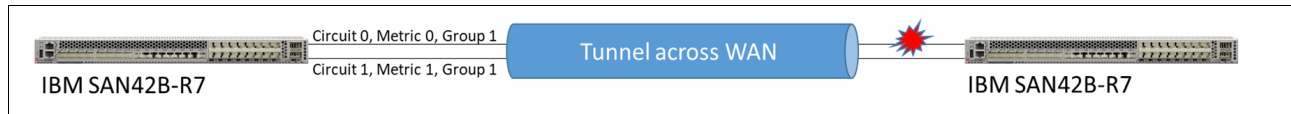
Figure 3-2   Metric-0 (Production Circuit) Failover to Metric 1 (Standby Circuit)

Circuit spillover

The spillover feature is supported on all IBM b-type extension platforms. The spillover feature enables you to configure primary circuits that are regularly used and secondary circuits that are only used during high utilization. When circuits are configured for spillover, the metric-0 circuits are utilized until the configured bandwidth is reached, after which the remaining traffic is sent over the higher metric-1 circuits.

For example, consider three 1 Gbps circuits that are configured on a tunnel as follows:

► Two circuits are metric-0 circuits
► One circuit is a metric-1 circuit
► Each circuit supports a maximum of 1 Gbps

This configuration starts at 2 Gbps and runs over the metric-0 circuits. If the traffic increases to 2.5 Gbps, the metric-0 circuits continue to support the 2 Gbps traffic, and the additional 500 Mbps runs over the metric-1 circuit.

Circuit spillover is a load-balancing method defined at the tunnel level, not per circuit. This feature allows circuits to be used only when there is congestion or during periods of high-bandwidth utilization.

SLA (Service Level Agreement)

The service-level agreement (SLA) feature provides automated testing before putting circuits into service. SLA works with Wtool (WAN Test Tool).

SLA checks circuits for packet loss and ensures the network meets that parameter before bringing the circuit into service. If other circuit verifications are required, such as throughput, congestion, and out-of-order delivery, use Wtool to run tests manually.

When configured, SLA automatically invokes Wtool and runs a circuit test. Wtool verifies the network path using the same circuit configuration, so no additional configuration is needed. After the network condition is met, the Wtool session stops, and the circuit is returned to service. If a circuit bounces, it reverts to test mode, and the SLA condition must be met before the circuit is enabled again.

KATOV (Keepalive Timeout Value)

KATOV is essential for quick error recovery and is based on application requirements. IBM b-type extension recovers data lost inflight faster than applications can detect the error and recover at that level. Each circuit has its own KATOV because it usually takes a unique path, which may vary. When the KATOV expires, the circuit is deemed offline, and the data lost in flight is retransmitted across a remaining online circuit. The data is placed in order and delivered to the upper layer protocol. No data is lost in transit or delivered out of order.

3.1.3 The LAN side (IP Extension)

IBM b-type IP Extension provides IP storage with protocol optimization, bandwidth management, resiliency, encryption, and compression benefits. It is used for IP storage applications, such as TS7700 Grid, remote hosts, databases, NAS replication, data migration, and IP backups. IP Extension can use the same tunnel (VE_Port) and circuits that FCIP uses or its own tunnel. The LAN side connects to a data center LAN so that IP storage end devices can access the IP Extension Gateway.

IP Extension has the same benefits as FCIP:

► High-speed encryption
► Bandwidth management and flow-control
► Resiliency with lossless in-order delivery
► QoS
► Compression
► Load balancing
► Lossless failover/failback

IBM IP Extension is supported on the following platforms:

► IBM SAN18B-6 Extension Switch
► IBM SAN42B-R7 Extension Switch
► IBM SX6 Extension Blade

GE interfaces (LAN side)

IP Extension cannot function without at least one LAN-side GE interface, which requires one or more GE interfaces to be in LAN mode. IP storage replication ports communicate with IP Extension Gateways (LAN side IPIF) through the data center LAN. IP Extension Gateways are accessible through the LAN side GE interfaces. GE interfaces default to the WAN side, designated as FCIP in `switchshow`. Incoming LAN traffic is not recognized on a WAN-side GE interface and is dropped.

IBM b-type extension platforms have speed-settable GE interfaces. GE speeds are not auto-negotiated; extension GE interfaces operate at the configured speed and with the optic supporting that speed. GE optics might or might not be able to operate at more than one speed. See Table 3-4.

Table 3-4   Supported LAN side GE interfaces

Extension Platform	GE Interfaces	LAN side support	Interface Speeds
IBM SAN18B-6	GE0-GE1	Up to 4 of 6 interfaces	1GbE (RJ-45)
	GE2-GE7	Up to 4 of 6 interfaces	1GbE, 10GbE
IBM SAN42B-R7	GE0-GE15	Up to 8 of 16 interfaces	1GbE, 10GbE, 25GbE
	GE16-GE17	No (WAN side only)	100GbE
IBM SX6 Extension Blade	GE0-GE1	No (WAN side only)	40GbE
	GE2-GE17	Up to 8 of 16 interfaces	1GbE, 10GbE

Any interface configuration must be removed before changing the GE interface's WAN or LAN mode. Once configured as a LAN side interface, it cannot be used as a WAN side interface for circuits.

Consider the following when you configure GE interfaces:

► Determine which GE interfaces will be used for the LAN side connectivity
► Ensure that these GE interfaces are configured in LAN mode
► Ensure that these GE interfaces are in the Default LS

Portchannels (LAG)

IBM b-type IP Extension supports portchannels on the LAN side, also called link aggregation groups (LAG). A LAG forms a single logical connection from multiple links. It depends on the platform how many links a portchannel can have and how many portchannels there can be. A LAG can be defined as static or dynamic. Dynamic LAGs use Link Aggregation Control Protocol (LACP).

Deployment types

End device IP Extension connectivity is supported via direct connections to the extension platform, layer 2 (data center LAN), and layer 3 (data center routed) networks.

Layer 2 deployment

Layer 2 communicates at the Ethernet level. End devices on the same VLAN with IP addresses on the same subnet as the IP Extension Gateway can communicate at the Ethernet level. Alternatively, IP storage can be connected directly to the IP Extension LAN ports; however, data center LAN switches scale the number of end-device Ethernet connections that may be needed. Utilizing a static route on the end device can differentiate the gateway to which it sends traffic based on the destination. Figure 3-3 on page 28 shows an IP Extension deployment using a Layer 2 method.
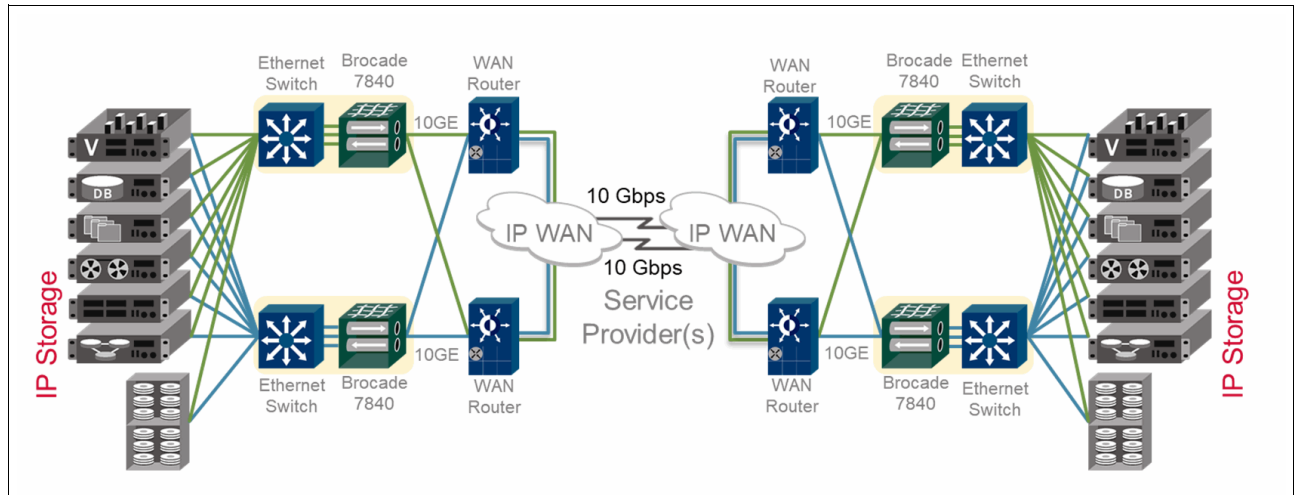
Figure 3-3   Layer 2 end device connectivity to IP Extension Gateway

With Layer 2 connectivity, configure IP storage replication with static routes and the IP Extension gateway as the next hop. The IP Extension gateway is configured on the DP (DP0 or DP1) that owns the tunnel's VE_Port. The IP Extension gateways are LAN IPIFs, which have IP addresses specific to a DP. The same IP Extension gateway cannot be configured on more than one DP. Also, the same IP Extension gateway cannot be configured on multiple DPs across platforms, which would cause an IP address conflict on the IP network. Since it is Layer 2, the end device learns the IP Extension gateway's MAC address through an ARP or Neighbor Discovery (ND) request.

When IP storage devices are connected to a Layer 2 switch, as shown in the following figure, you can use Link Aggregation 802.3ad (LAG) to connect IP Extension LAN interfaces to the switch. The IBM b-type extension platforms support static and dynamic LAG to data center switches. Dynamic LAG uses LACP. The LAG can be 802.1Q tagged if end devices live on more than one VLAN. Tagging enables multiple VLANs to communicate across the same physical LAG. The default is no tagging.

> **Note:** Only one path between the data center LAN switches and the IP Extension Gateway can exist. A LAG (portchannel) is considered a single path.

Figure 3-4 shows that the IP Extension LAN interfaces are connected to IP storage by a Layer 2 LAN switch. For a Layer 2 deployment, each VLAN requires a LAN side IPIF (IP Extension gateway). The LAN side IPIF corresponds to the DP that owns the tunnel to which the TCL matches traffic. The LAN side IPIF has an IP address on the same subnet as the end device replication ports and acts as the IP Extension gateway that forwards traffic to the remote data center.
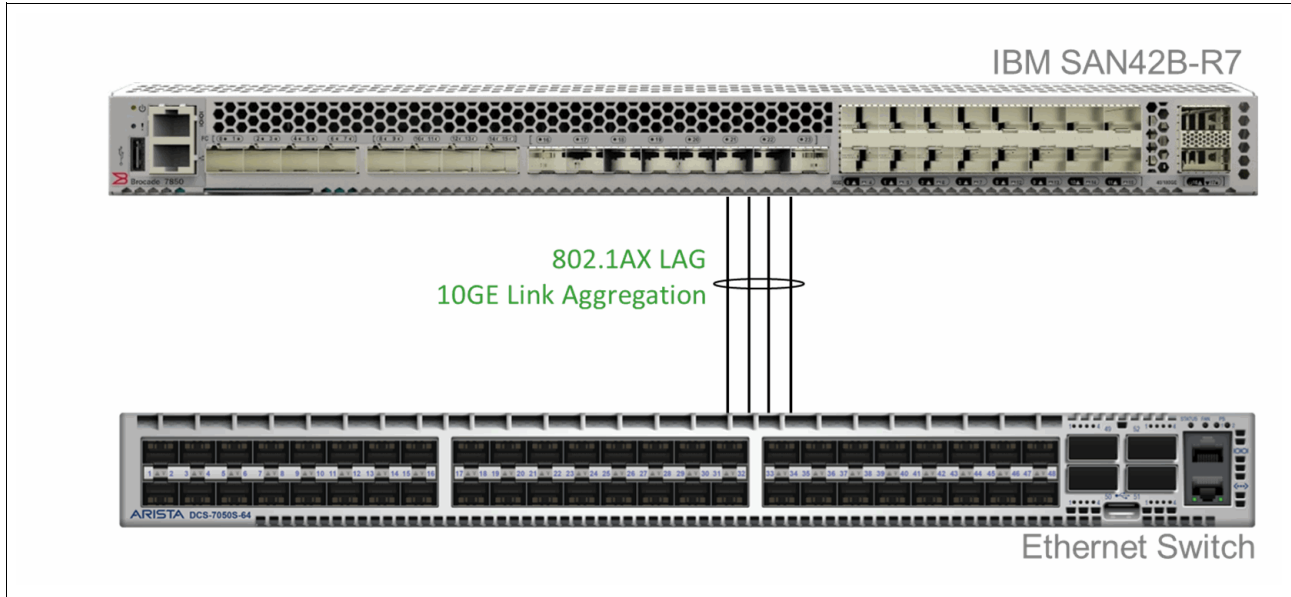
Figure 3-4   Layer 2 deployment with Ethernet Switch

Layer 3 deployment

In a Layer 3 deployment, end devices do not require special routes. Traffic is sent to the traditional default router gateway instead of sending it to the IP Extension gateway. Changes are made in the network routers, which must be configured to intercept specific traffic and forward it to the IP Extension gateway. Policy Based Routing (PBR) is a common mechanism to perform this task.

In Figure 3-5, IP Extension in a routed network configured with PBR. The network intercepts traffic matched by an access control list (ACL) and redirects the traffic to an IP Extension gateway. A Layer 3 deployment simplifies IP storage connectivity and scales to more subnets than a Layer 2 deployment.
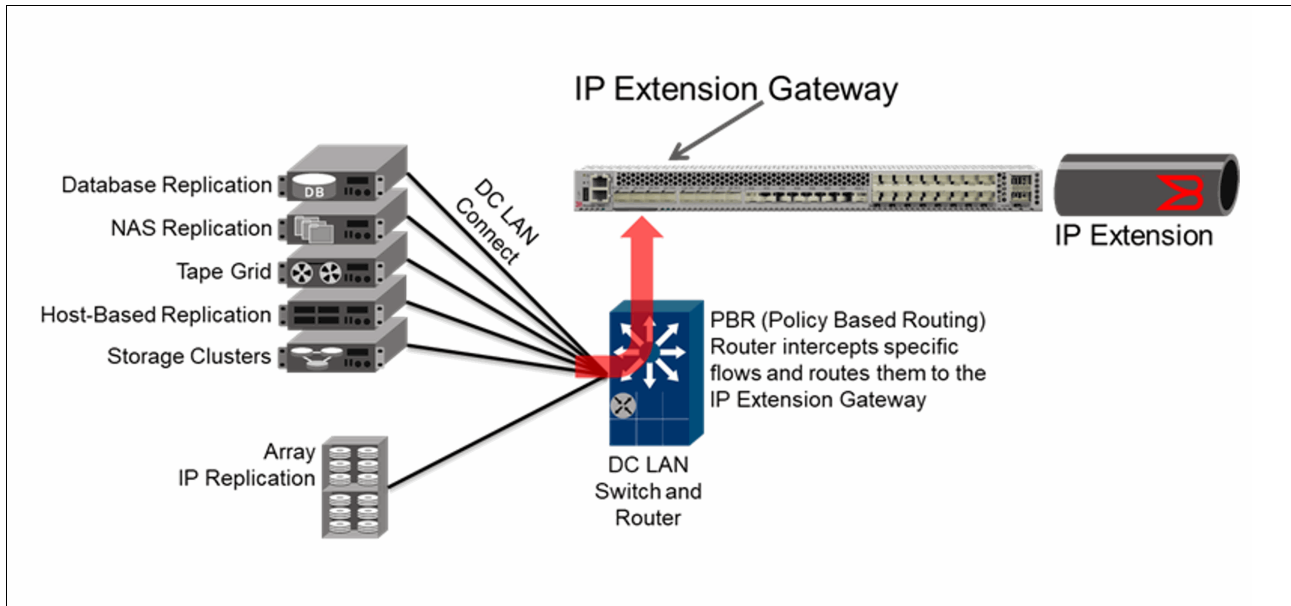
Figure 3-5   Layer 3 PBR, end device routes not needed

IP Extension considerations:

► For TCP traffic:

– The LAN side uses the Rx window for source flow control to prevent buffer overflows
– The Rx window closes when there is congestion
– The Rx window closes when the destination exhibits slow drain behavior
– Each TCP session has an autonomous Rx window; no session blocks another
– Up to 512 TCP open requests per second are allowed; additional open requests are dropped
– Limiting open requests per second mitigates Denial of Service (DoS) attacks
– Statistics are available for the number of dropped connections

Table 3-5   Maximum supported TCP sessions and RASlog warning threshold

Platform	TCP Sessions per DP	RASlog Warning Threshold
IBM SAN18B-6	256 (DP0 only)	244
IBM SAN42B-R7	512 (on each DP)	500
IBM SX6 Extension Blade	512 (on each DP)	500

IP Extension Gateway (LAN IP interface)

A LAN-side IPIF is a gateway for IP storage devices to access IP Extension, called an IP Extension gateway. End devices send IP storage traffic to IP Extension gateways to transmit data across IP Extension. The IP Extension gateways are used for layer 2 and layer 3 deployments. Each DP has its own set of IP Extension gateways. The same IP Extension gateway cannot be used on different DPs. In an L2 deployment, an IP Extension gateway is on the same subnet as the end device replication ports (same broadcast domain). IP Extension gateways support IPv4 and IPv6.

LAN-side GE interfaces can access all IP Extension gateways; therefore, all untagged (no 802.1Q) Ethernet traffic can access IP Extension gateways on either DP. If the LAN-side traffic is tagged (802.1Q), access depends on matching the incoming VLAN tag with the LAN-side IPIF VLAN ID.

In a Layer 2 deployment, the IP Extension gateway is the next-hop address for the storage end device. A layer 2 deployment requires IP routes on the end device to direct its replication traffic to the IP Extension gateway. Changes must be made to the end devices. On each end device, for each subnet that is used on an interface, an IP route must be added to the end device forwarding replication traffic to the IP Extension gateway. For each subnet on which an end device has an interface, an IP Extension gateway must be configured. These end device IP routes are required before the end device can send traffic to an IP Extension gateway.

In a layer 3 deployment, the end device's default route is used, and no changes are made to the end devices. However, a network router must be modified to intercept the IP storage flows and forward them to the IP Extension gateways.

> **Note:** LAN side end device subnets must differ on the local and remote sides.

LAN side IPIF considerations:

► An IPIF lan.dp# is an IP Extension gateway

► An IP Extension gateway cannot be used on more than one DP

► Identify the end device replication ports that will be extended through IP Extension

- ▶ Identify the subnets used on IP storage replication ports
- ▶ A separate LAN side IPIF is needed for each subnet
- ▶ An 802.1Q tagged LAG (tagged portchannel) may carry multiple VLAN IDs
- ▶ Identify the VLAN IDs associated with each subnet
- ▶ Identify the VLAN's MTU and end device replication MTU; use the smaller of the two
  - A LAN side IPIF MTU range is 1280 to 9216 bytes. The default is 1500 bytes.
  - PMTU is not supported on the LAN side IP Extension gateway interfaces
- ▶ With a Layer 2 deployment, create an IP Extension gateway for each subnet
  - Multiple IP Extension gateways on the same subnet cannot be configured on the same DP
  - An IBM SAN42B-R7 and IBM SX6 Extension Blade can have a maximum of eight IP Extension gateways per DP
  - IBM SAN18B-6 can have a maximum of four IP Extension gateways
- ▶ IPIF IP addresses and subnet masks cannot be modified; the IPIF must be deleted and recreated
- ▶ There is only one Software Virtual Interface (SVI) per DP, which means only one LAN side MAC per DP

In the following situations, separate LAN gateways must be configured:

- ▶ If the data center LAN switch is not configured to use VLAN tagging (802.1Q) on the LAG (portchannel), do not configure a VLAN ID on the LAN IPIF. Ethernet links do not operate when there is a mismatch of tagging and non-tagging. Tagged traffic can only communicate with interfaces configured for tagged traffic.
- ▶ If the LAG (portchannel) from the data center LAN switch is tagged, an IP Extension gateway must be created for each VLAN. To accommodate multiple VLANs, the LAG must be a trunk, which means the traffic is tagged. In a trunk, multiple logical ISLs pass through the physical LAG, and each VLAN is tagged with its corresponding VLAN ID. For tagged traffic to be forwarded to the correct IP Extension gateway, the LAN IPIF must be configured with the corresponding VLAN ID.

VLANs (LAN side)

When VLANs are used, they differentiate between traffic flows passing through the same physical GE interface. Data center LAN switches can forward flows out specific GE interfaces based on the VLAN ID or it can send a set of VLANs over a trunk.

LAN side VLAN IDs are configured on LAN side IPIFs. Configuring the same VLAN ID on multiple IPIFs, for example, on an IPIF for DP0 and an IPIF for DP1, is possible. IPIFs for DP0 and DP1 cannot share the same gateway IP address but can belong to the same VLAN.

A VLAN ID is needed only when the traffic from the data center network switches is tagged. The Ethernet port directly connected to a LAN side GE must be configured for VLAN tagging; otherwise, do not configure a VLAN ID on the IPIF. Usually, switch ports are members of a VLAN and not configured for VLAN tagging, so a VLAN ID on the IPIF is unnecessary. The Ethernet link cannot be established if there is a tagging mismatch between the extension platform and the data center switch.

MTU (LAN side)

Determine the supported MTU from the IP storage provider and the Network Administrator, use the smaller value. Set the MTU value on the LAN side IPIF. The default MTU is 1500

bytes. The larger the MTU, the less relative overhead and the more efficient the data transfer is.

IP routes (LAN side)

IP routes define a destination subnet and router gateway address. The destination subnet is where remote end devices reside, and the gateway is the local router that forwards the traffic. IP Extension supports layer 3 deployments. In a layer 3 deployment, a LAN-side IP route forwards traffic from the WAN side to a next-hop router on the LAN side. The local router forwards the traffic to the end devices. LAN-side IP routes are only used with layer 3 deployments. LAN-side IP routes do not forward traffic to the WAN side, which the TCL does.

A layer 2 deployment does not use LAN-side IP routes because the end devices are on the same subnet as the IP Extension gateway; therefore, no intermediate LAN-side router exists.

In a Layer 3 deployment, when traffic arrives at the local LAN side router from an IP storage end device, it is intercepted by the router and forwarded to the IP Extension Gateway. Typically, an access control list (ACL) and policy-based routing (PBR) match the incoming traffic and forward it to the IP Extension Gateway. The router must be configured to perform this function. End devices do not require any particular configuration with Layer 3 deployments.

Configuration steps for PBR are done on the router. PBR is configured with the next hop as the IP Extension Gateway. An ACL and PBR policy determines traffic sent to the IP Extension Gateway.

A LAN-side IP route to the next-hop gateway is configured on the extension platform. The IP Extension Gateway and the next-hop router are on the same subnet. The steps for configuring a LAN-side IP route are similar to configuring a WAN-side IP route. PBR forwards traffic from the end devices toward the tunnel. IP Extension LAN side IP routes forward traffic from the tunnel toward the end devices.

Figure 3-5 on page 29 shows a network topology where IBM IP Extension is the next-hop gateway for end device traffic forwarded by PBR on the local LAN side router. Traffic not diverted by PBR to IBM IP Extension continues to be routed based on the traditional routing table.

Tunnels (LAN side)

A tunnel is a WAN side object; however, IP Extension must be enabled on the tunnel if needed. Do not confuse enabling IP Extension on a tunnel with enabling hybrid mode on an extension platform. By default, tunnels do not have IP Extension enabled, even if the platform does have IP Extension enabled.

Compression (LAN side)

When IP Extension is not enabled on a tunnel, compression is set per tunnel. If IP Extension is enabled on a tunnel, compression is set per protocol (FCIP and IP Extension). A protocol-specific compression setting overrides the general compression setting.

The protocol-based compression modes can be set to default, which causes the compression setting for each protocol to be inherited from the general tunnel setting. If the tunnel's general compression is set to fast-deflate, IP Extension compression is set to none, and FCIP compression is set to fast-deflate. Only deflate and aggressive-deflate are allowed with the `--ip-compression` option.

> **Note:** IP Extension does not support Fast-deflate compression, only Deflate and Aggressive-Deflate are supported.

QoS (LAN side)

IBM b-type extension has robust QoS functionality. The sections below describe QoS distribution and marking.

Bandwidth distribution

Each protocol (FCIP and IP Extension) can configure a custom QoS protocol bandwidth distribution; the default distribution is 50/50% FCIP/IP Extension.

Each protocol has a high, medium, and low priority distribution. The default is 50%, 30%, and 20% for high, medium, and low. IP Extension has three QoS priorities; ip-high, ip-medium, and ip-low. The minimum distribution allocation per priority is 10%. The maximum distribution allocation for a priority is 90%. QoS allocations must total 100%.

FCIP and IP Extension protocol distribution percentages and priority bandwidth percentages are rate-limiting. When bandwidth is utilized in a protocol distribution or bandwidth priority, the entire percentage is reserved and not shared with other distributions or priorities. If a protocol is not configured for use or a priority is not configured for use, the bandwidth is not reserved and available. For example, if IP Extension is not being used, and low and high priorities are not configured for FCIP, medium is the default, and FCIP medium will get all the bandwidth.

Marking

IP Extension traffic can be marked with DSCP and L2 CoS (802.1P). Work with your Networking Administrators to determine if QoS marking should be used to expedite traffic in the network.

DSCP marking

Layer 3 Class of Service Differentiated Services Code Point (DSCP) refers to a specific implementation for establishing QoS policies defined by RFC 2475. DSCP uses six bits of the Type of Service (TOS) field in the IP header to establish up to 64 values to associate with data traffic priority. DSCP settings are helpful only if IP routers are configured to enforce QoS policies uniformly within the network. IP routers use the DSCP value to index per-hop behavior (PHB). Control connections and data connections can be configured with different DSCP values. Before configuring DSCP settings, determine if the IP network implements DSCP, and consult with the WAN administrator to determine the appropriate values for extension.

Layer 2 Class of Service (L2CoS) marking

VLAN traffic is routed using 802.1Q-compliant tags within an Ethernet frame. The tag includes a unique VLAN ID and Class of Service (CoS) priority bits. The CoS priority scheme Layer 2 Class of Service (L2CoS) uses three Class of Service (CoS or 802.1P) priority bits, allowing eight priorities. Consult with your WAN administrator to determine usage.

TCL (IP Extension Traffic Control List)

The traffic control list (TCL) defines how LAN side ingress traffic is mapped to tunnels. A TCL is a list of rules that can identify LAN side traffic based on IP, TCP, and Ethernet header information, such as protocols, ports, source and destination IP addresses, source and destination subnets, QoS values (DSCP or 802.1P), and 802.1Q VLAN tags. Each rule allows or denies a matched flow, acting as an input filter to an IP Extension tunnel. TCL rules are arranged by priority and provide control over LAN side traffic directed to a particular tunnel.

4

IBM solution support

As mentioned in Chapter 1, "Introduction" on page 1, the National Institute of Standards and Technology (NIST) has five dimensions of their security model and the two dimensions we are focused on in this IBM Redbooks are the Protect and Recover dimensions. IBM's Cyber Resilience offerings include a broad range of components to adhere to the NIST model but in this section, we will review the IBM Storage Solutions that help Protect and Recover your data through long-distance replication.

In this chapter will cover the following sections:

4.1 Introduction

IBM Storage Solutions that replicate data have been around for decades. However, those solutions have evolved from moving data from a local disk array to a local tape device to infrastructure that can move data anywhere in the world. These solutions provide protection of your data but also allow you to recover that data if you experience a disaster or cyber-attack.

At a high level, all the solutions we will review in this chapter can be applied to any IBM storage platform. However, depending on whether you are responsible for a mainframe environment or an open systems environment, variants of these solutions may apply.

4.2 IBM Storage Solutions supported by IBM b-type SAN Extension Platforms

- ► Metro Mirroring with Hyperswap - FlashSystems and DS8K
- ► Global Mirroring with Hyperswap - FlashSystems and DS8K
- ► Global Copy
- ► z/OS® Global Mirror, previously known as extended remote copy (XRC)
- ► SVC Stretched Cluster
- ► IBM Z® FICON Tape Extension - Virtual Tape Library (VTL)
- ► Open System Fibre Channel (FC) Tape Extension - All FC Tape
- ► TS77xx Grid Solution - Supporting Block and Object Store Replication

Figure 4-1 provides an example of an IBM Z environment where the IBM b-type SAN42B-R7 is being used in a TS77xx Grid solution. The SAN42B-R7 provides the interconnect across TCP/IP wide area networks to each of the remote TS77xx systems and helps maintain the performance and stability needed for a Cyber Resilient system.
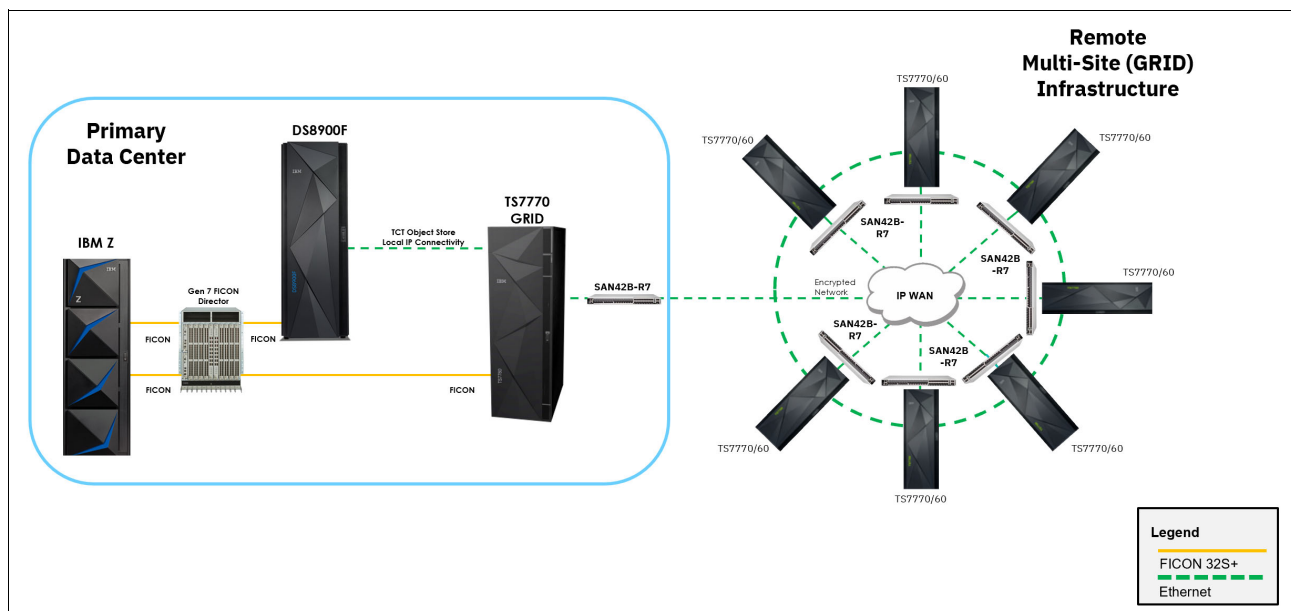
Figure 4-1   IBM Z environment using TS77xx grid for block and object store replication

The IBM b-type SAN Extension platforms support all the IBM Storage Solutions mentioned above. The SAN extension platforms can also support all of these solutions concurrently within the same physical system and network. Having this flexibility gives IT Infrastructure designers and planners the ability to protect different "tiers" of storage without needing separate equipment.

In the following sections of this chapter, a general description of the IBM Storage Solutions that leverage the IBM b-type SAN Extension will be provided. These general descriptions are not meant to be a "how to guide" for solution design. Detailed information for implementation and design of these solutions can be found in other IBM Redbooks publications specific to those IBM Storage Platforms, such as:

► *Implementation Guide for IBM Storage FlashSystem and IBM SAN Volume Controller (for IBM Storage Virtualize 8.6)*, SG24-8542

► *IBM Storage DS8900F Architecture and Implementation: Updated for Release 9.3.2*, SG24-8456

4.3 IBM Disk Mirroring solutions

IBM b-series Extension switches and blades are most commonly used for business continuity (BC) and disaster recovery (DR) by leveraging an IP routed network between datacenters. Leveraging Remote Data Replication (RDR) to transport critical data across a significant enough distance outside a catastrophic event preserves data. Data preservation permits an organization to recover operations.

IBM business continuity solutions that utilize IBM b-series Extension switches consist of the following disk replication services:

► Metro Mirror

► Global Mirror

► Global Copy

► z/OS Global Mirror, previously known as extended remote copy (XRC)

► SVC Stretched Cluster

Using combinations of remote disk replications the following multi-site solutions can also be deployed utilizing b-series extension solutions:

► 3-site Metro/Global Mirror with Incremental Resync

► z/OS Metro/Global Mirror across three sites

► IBM z/OS Metro/Global Mirror Incremental Resync (RMZ Resync)

► IBM Multiple Target Peer-to-Peer Remote Copy

► SVC Stretched Cluster with Global Mirror or Metro Mirror

All of the disk replication applications, other than z/OS Global Mirror, use FCP or Open Systems protocol end-to-end over the extension switches, so the deployment and topology for those solutions are very similar.

It is a best practice to connect directly to the storage arrays, without connecting to an intermediate FC switch whenever possible, and to always deploy redundant fabrics. The diagrams in the following sections show basic dual fabric connectivity, but not necessarily the specific zoning and pathing requirements.

z/OS Global Mirror is a FICON extension solution and therefore has unique pathing requirements and topologies compared to the other disk replication solutions.

SVC Stretched Cluster also has very specific fabric connectivity and zoning requirements that must be followed when deploying extension switches.

4.3.1 Metro Mirror

Metro Mirror is a synchronous replication solution between two storage arrays where write operations are completed on the local and remote volumes before the I/O is considered to be complete. Metro Mirror is used in environments that require no data loss in the event of a storage system failure.

Because data is synchronously transferred to the secondary storage system before considering the write to be complete, the distance between primary and secondary storage systems affects the application response time. The supported distance for Metro Mirror is 300 km (186 mi).

4.3.2 Global Mirror

Global Mirror provides a long-distance Remote Copy feature across two sites by using asynchronous technology. With Global Mirror, the data that the host writes to the storage system at the primary site is asynchronously shadowed to the storage system at the secondary site. A consistent copy of the data is automatically maintained at the remote site.

Global Mirror operations provide the benefit of supporting operations over unlimited distances between the local and remote sites, which are restricted only by the capabilities of the network and the channel extension technology. It can also provide a consistent and restorable copy of the data at the remote site, created with minimal impact to applications at the local site.

Figure 4-2 shows an example of two site open systems topology for Metro Mirror or Global mirror utilizing IBM FlashSystems and the SAN42B-R7 Extension Switch.
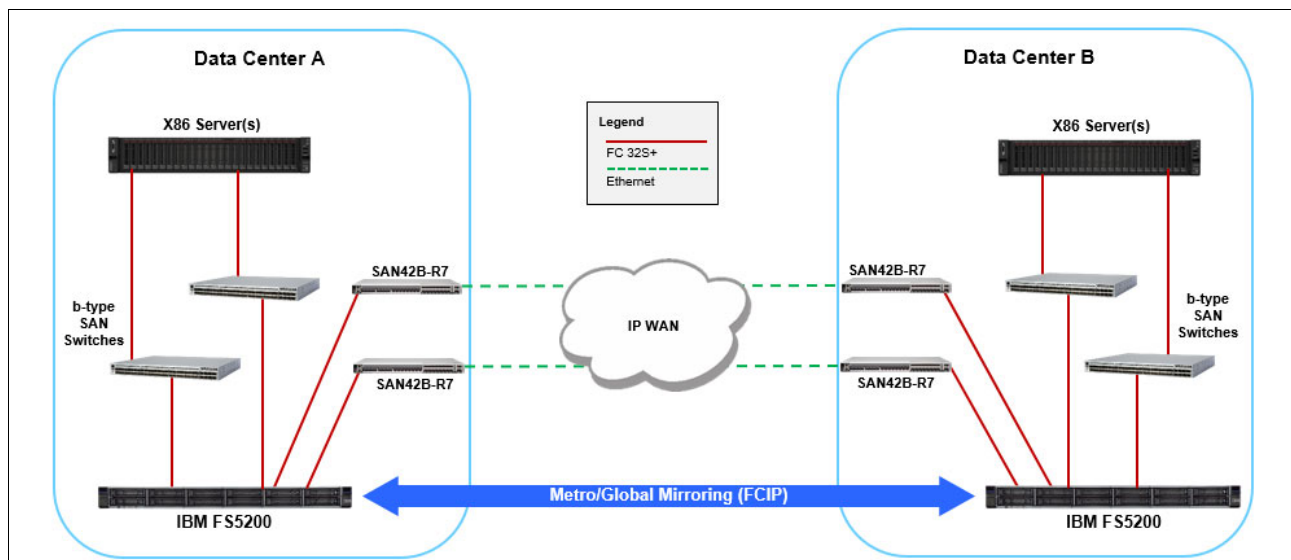
Figure 4-2   IBM FlashSystems and IBM b-type SAN Extension Mirroring

Figure 4-3 shows an example of a three site mainframe topology for Metro Mirror and Global mirror utilizing IBM DS8000® Enterprise Storage and the SAN42B-R7 Extension Switch.
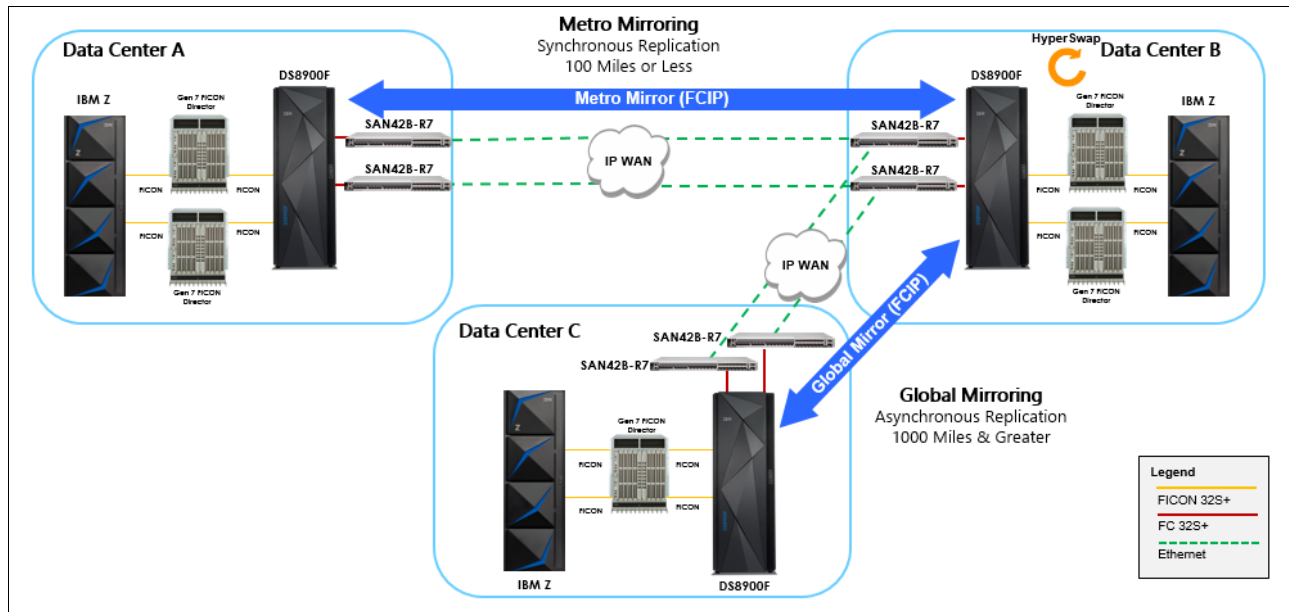
Figure 4-3   IBM Z and IBM Enterprise Storage Leverage IBM b-type SAN Extension for GDPS® configurations

4.3.3 Global Copy

Global Copy is a copy service specific to the DS8000 arrays and is an asynchronous remote copy function that is used for longer distances than are possible with Metro Mirror. Global Copy is appropriate for remote data migration, offsite backups, and transmission of inactive database logs over virtually unlimited distances.

Global Copy is also used as the data transfer mechanism for Global Mirror. With Global Copy, write operations complete on the primary storage system before the data is copied to the secondary storage system, thus preventing the primary system's performance from being affected by the time that is required to write to the secondary storage system. This method allows the sites to be separated by a large distance.

All data that is written to the primary array is transferred to the secondary array, but not necessarily in the same order that it was written to the primary. This method means that data on the secondary is not time-consistent. Making use of the data on the secondary volume requires using some technique to ensure consistency.

The SAN extension topology for Global Copy would be identical to the topology deployed for Metro Mirror or Global Mirror.

4.3.4 z/OS Global Mirror (XRC)

z/OS Global Mirror (zGM), previously known and commonly referred to as extended remote copy (XRC), provides long-distance remote mirror solution across two sites for IBM Z data using asynchronous technology.

zGM is a remote mirroring function that is available for FICON attached devices on z Systems® architectures (z/OS, z/VM®, and z/VSE®). zGM asynchronously maintains a consistent copy of the data at a remote location and can be implemented over unlimited

distances. It is a solution that consists of cooperative functions between DS8000 firmware and z/OS DFSMS software that offers accurate and rapid disaster recovery with data integrity.

Figure 4-4 shows an example topology of a z/OS Global Mirroring topology. Data replication is taking place from the Primary Data Center to the Secondary Data Center utilizing the SAN42B-R7 Advanced Acceleration for FICON feature, allowing the SDM running on the IBM z to maintain read I/O performance from the DS8000 over long distances.

> **Note:** The IBM DS8900F family will be the last platform to support z/OS Global Mirror. There will not be any new z/OS Global Mirror functions provided with IBM DS8900F.
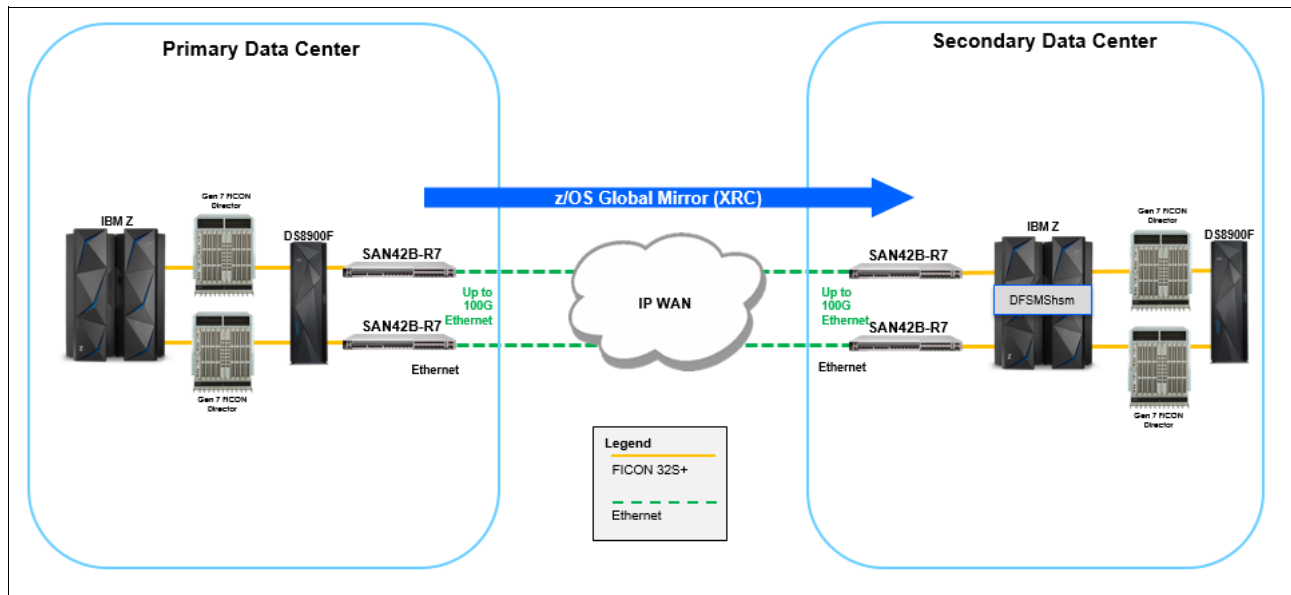
Figure 4-4   IBM Z and IBM Enterprise Storage z/OS Global Mirror (XRC) with IBM b-type FICON Extension

4.3.5 SVC Stretched Cluster

In a Stretched Cluster configuration each node from the same I/O Group is physically installed at a different site. Stretched Cluster is considered a HA solution because both sites work as instances of the production environment (no standby location is used).

When implemented, this configuration can be used to maintain access to data on the system, even if failures occur at different levels, such as the SAN, back-end storage, IBM Spectrum® Virtualize controller, or data center power. Stretched cluster configurations can support distances up to 300 km (186.4 miles).

Figure 4-5 on page 41 shows an example topology utilizing redundant fabrics, a best practice that is not unique to Stretched Clusters or HyperSwap®. However, what is unique to these IBM Spectrum Virtualize cluster configurations is a further subdivision of each of the redundant fabrics into public and private fabrics. This subdivision can be done using virtual fabrics for the public and private SANs. However, care must be taken to ensure that the public and private fabrics traverse dedicated ISLs.

> **Note:** For more information about Stretched Cluster and HyperSwap SAN design and topology best practices, reference the Redpaper *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597.
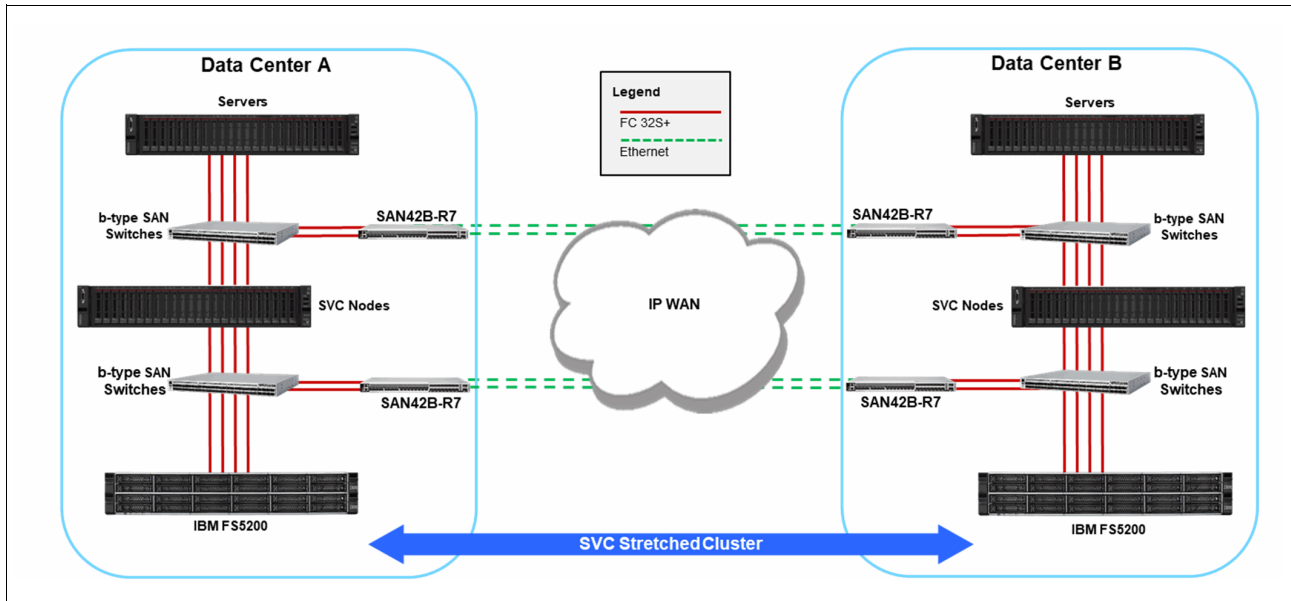
Figure 4-5   IBM SVC Stretched Cluster and IBM b-type SAN Extension

5

Extension best practices

To address disaster recovery requirements, IBM b-type extension offers a solution portfolio designed to provide organizations with flexible deployment options for replication. The IBM b-type extension platforms are robust for large-scale, multi-site data center environments implementing block, file, and tape data protection. IBM b-type extension is a purpose-built solution that securely moves data over considerable distance while minimizing the risk of disruption. These platforms deliver unprecedented performance, security, and availability to handle unrelenting data growth between in FC, FICON, and IP storage environments.

IBM b-type extension has always been the leader in innovation and was the first to develop technologies such as extension trunking (BET), FC Routing (FCR), Advanced FICON Accelerator, Adaptive Rate Limiting (ARL), FastWrite, Open Systems Tape Pipelining (OSTP), FC compression, FCIP encryption, PerPriority TCP QoS (PTQ), IP Extension, and more. These capabilities deliver unmatched value and drive down capital and operational expenses as well as losses due to unplanned downtime.

This chapter has the following sections:

5.1 The IP network

The IP WAN network must be able to transport extension TCP sessions with adequate performance and service levels. By today's standards, less than 0.1% is considered excessive packet loss for WAN networks. IBM b-type extension supports 1% packet loss if the KATOV is 2 seconds or greater and 0.1% if the KATOV is less than 2 seconds. IBM b-type extension can run circuits across unique pathways of disparate service provider networks. There is no requirement for various WAN connections to have the same bandwidth, latency, or packet loss.

On IBM b-type extension platforms, the Gigabit Ethernet interfaces are called GE interfaces. Every effort should be made to connect the GE interfaces as close to the WAN as possible to prevent excessive hops, added latency, potential congestion, and performance degradation. The supported propagation delay across an IP network is 200 ms RTT, which will usually reach anywhere on Earth. IBM b-type products do not support network switch ports that block or oversubscribe. Some Ethernet switches have host access ports that are oversubscribed or blocked, which degrades performance.

From end to end, the IP network must allow specific protocols to pass. When not using IPsec, IBM b-type extension uses TCP destination ports 3225 and 3226. IBM b-type extension selects a random ephemeral port (source port) between 49152 and 65535.

The TCP URG flag is required on IBM b-type extension and must not be dropped or modified by any intermediate network device, such as a firewall. Dropped or changed TCP URG flags by intermediate firewalls are not supported. Refer to 7.3.8, "Encryption (IPsec)" on page 80 section in this document for additional information concerning the network path when using encryption.

5.1.1 Redundancy

Multiple extension tunnels can be defined between a pair of extension domains, but doing so defeats the benefits of a single multiple-circuit extension tunnel. Defining two tunnels between a pair of switches or blades is less redundant or fault-tolerant than multiple circuits in one tunnel.

Brocade Extension Trunking (BET) provides lossless link loss (LLL), which ensures that all data lost in flight is retransmitted and reordered before being delivered to upper-layer protocols. This essential feature prevents interface control checks (IFCCs) on mainframes and SCSI timeouts for open-system-based replication.

Furthermore, the best practice is to deploy two replication fabrics (A and B) for redundancy. Each replication fabric should be identical and autonomous, with no interconnections (referred to as anan "air gap").

In practice, separate replication fabrics can be constructed from standalone IBM b-type extension platforms or logical switches on director platforms using the IBM SX6 Blade. Create one logical switch for replication per director. Do not place both of the redundant replication fabrics on the same director. Placing the FC and VE_Ports used for replication into an autonomous logical switch logically separates the replication fabric.

5.1.2 Bandwidth

The best practice for critical remote data replication (RDR) is to use a separate and dedicated WAN connection between the production data center and the backup site. Even so, a

dedicated WAN connection between data centers is often impractical. In this case, bandwidth must be, at a minimum, logically dedicated to extension. There are a few ways this can be done:

► Use QoS and prioritize extension, logically dedicating enough bandwidth to extension over other traffic.

► Use committed access rate (CAR) to identify and rate-limit certain traffic types. Use CAR on non-extension traffic to apportion and limit that traffic to a maximum bandwidth, leaving the remainder of the bandwidth to extension. Set the aggregate circuit min-comm-rates to use the remaining dedicated bandwidth, logically dedicating bandwidth to extension.

► For all traffic to coexist without congestion, massively over provisioning bandwidth is the easiest, most costly, and most common practice in extension deployments.

IBM b-type extension uses an aggressive TCP stack called WAN Optimized TCP (WO-TCP), which dominates other TCP stacks by causing them to dramatically back off. UDP-based flows may result in considerable congestion and excessive packet drops for all traffic.

The best practice is to rate-limit and flow-control traffic on the extension platforms at the source and destination and not rate-limit in the IP network. Rate-limiting extension in the IP network leads to performance problems and complex troubleshooting issues. IBM b-type extension rate limiting is advanced, accurate, and consistent; there is no need to rate limit in the IP network.

To determine the network bandwidth needed, record the number of bytes written over a month (or more). A granular recording capable of calculating rolling averages of varying lengths is helpful. It is essential to understand the number of bytes written to volumes during various interims of the day, night, weekends, end of quarter, end of fiscal year, holidays, etc. These data rates must be available for replication to maintain an adequate Recovery Point Objective (RPO).

If replicating asynchronous RDR (RDR/A) across a tunnel, calculate the average value over a finite number of minutes throughout the day and night to determine the maximum average. Remember that business transactions may increase, and replication may require significantly more bandwidth during the quarter end, the fiscal end, and certain holidays. RDR/A needs enough bandwidth to accommodate the high averages discovered over a finite period. RDR/A performs traffic shaping, which moves peaks into troughs. Averaging over too long a period may cause a backlog of writes when the troughs do not occur frequently enough to relieve the peaks. Excessive journaling is challenging to recover from, depending on available bandwidth.

The best practice for synchronous replication (RDR/S) is to dedicate the IP network without sharing bandwidth. A 10 Gbps DWDM λ is an example of dedicated bandwidth. Dedicated bandwidth virtually eliminates packet drops and reduces the need for TCP retransmits. Record the peak traffic rates if you plan to replicate RDR/S across an extension tunnel. RDR/S must be able to send writes immediately, accommodating the entire demand at any time.

Plot the recorded values into a histogram. Suppose you have calculated 5-minute rolling averages from your recorded data. Bytes written over each period will be an integer value. The x-axis is the number of bytes written in each of the 5-minute averages. The left x-axis starts at zero bytes. The right x-axis is the largest number of bytes written during an interim. Averages with the same number of bytes will occur multiple times. The y-axis is the number of times that a particular average occurred. The smallest size occurred relatively rarely. The largest size occurred relatively often. 68% of the averages fall into the middle section (see Figure 5-1 on page 46), the largest group and the group you are most concerned with. The resulting curve is a bell curve.

Based on cost, plan your WAN bandwidth to accommodate at least the first standard deviation (68%), and if possible, include the second standard deviation (95%). In most cases beyond the second deviation, occurrences are so rare that they can be disregarded in bandwidth planning.
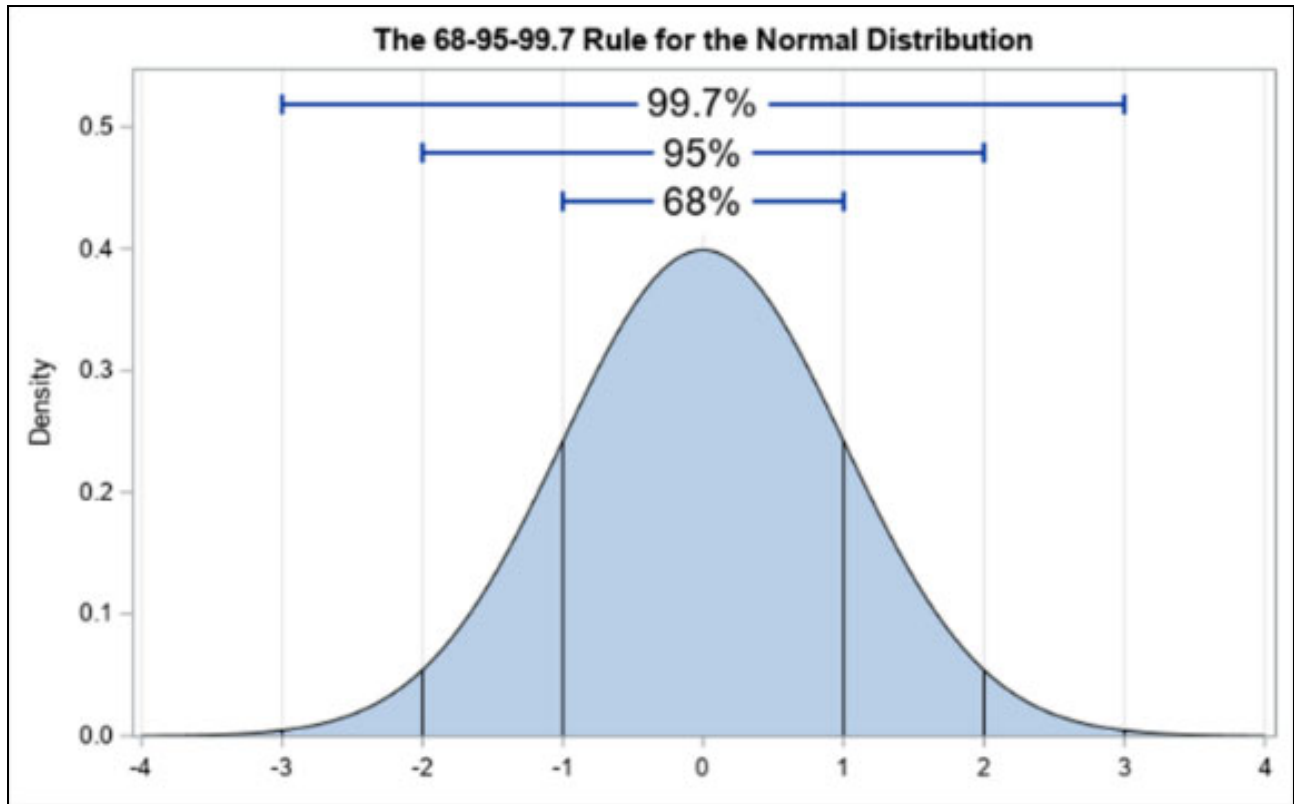
Figure 5-1   Gaussian Standard Normal Distribution

You can plan for a certain amount of compression, such as 2:1 if your data is compressible; however, data compressibility is a moving target. The best practice is to use compression as a safety margin to address unusual and unforeseen situations. Also, achieving some compression can provide headroom for potential future growth. If there is no margin, periodic demand increases may not be advantageous.

For remote tape backups, extension is used to extend a fabric to a remote VTL. Measuring tape volume in MB/h and the number of simultaneous drives in use is crucial. Bandwidth will limit the backup window even if an adequate number of drives are available.

5.1.3 IP networking

IP networking is a primary part of extension. The ability of the IP network to deliver storage traffic is directly related to storage replication completing sufficiently and reliably.

 IP networking may be limited by the following:

► In order datagram delivery. A network that does not perform best effort to deliver datagrams in order is not supported. Relatively infrequent out-of-order datagrams are not an issue.

► Less than 1% packet loss, and in some cases, less than 0.1% packet loss, depending on the KATOV. Packet loss greater than 1% is not supported.

- ► Less than 250 ms RTT, and in some cases less than 200 ms RTT depending on the KATOV. RTT greater than 250 ms is not supported.
- ► IPsec through NAT is not supported.
- ► Altering TCP flags (that is, firewalls changing TCP flags) is not supported.
- ► WAN optimization devices are not supported.
- ► Network must be able to deliver the volume of data configured on the extension boxes. Extension platforms perform flow control between the destination and source.

5.1.4 QoS

If the IP network uses QoS to parse bandwidth, work with the Network Admins to apportion the bandwidth required for storage. Determine the DSCP or L2CoS marking required for the IP network to recognize the storage traffic as part of that QoS class. Test replication across extension and the IP network to ensure that the QoS class is being recognized and is performing adequately.

5.2 The replication SAN

A storage network may have various SANs. Foremost is the production SAN, a local high-speed and high-reliability network. Secondly, there should be a replication SAN, which connects a local and remote data center. Every I/O of value is replicated to the remote site. IBM b-type extension is used to build a high-speed, high-reliability, and high-security replication SAN over an IP WAN network.

5.2.1 Connectivity

FC replication connections from storage to the extension platforms should be direct connections. If replication uses a blade in a director, extension should be separated from production via a logical switch. If the replication ports are dedicated to replication, there is no reason to connect these dedicated ports through the production network. Move the replication FC ports to the replication logical switch and move the tunnel's VE_Port to the same logical switch.

A single VE_Port should be used to connect the local and remote extension domains, and multiple circuits belonging to the VE_Port should be used. Using multiple circuits improves resiliency, increases bandwidth, and enhances redundancy.

Always connect each extension platform's power supply to independent power sources within the data center.

5.2.2 Redundancy

Replication SANs are typically deployed in redundant pairs called "A" and "B." If one of the replication fabrics goes offline, the remaining replication fabric maintains replication. Offline fabrics could be due to a planned outage or an unplanned outage.

It is common for mainframe environments to have four fabrics in their replication SAN such that, at most, 25% capacity is sacrificed during an offline replication fabric event.

Many aspects of redundancy are required beyond just the fabrics themselves. The IP network and connections require redundancy as well. Every replication fabric should have its own IP WAN connection, potentially from a unique service provider. Each replication fabric should have its own autonomous IP networking equipment; IP networking equipment should not be shared between replication fabrics or IP WAN connections.

Considering connections, ports, optics, cables, and bandwidth, each should have redundancy.

5.2.3 Failover and failback

All circuits have a metric of 0 or 1 and belong to a failover group, as shown in Figure 7 below. Metric 0 is the default and is preferred over metric 1. Metric 0 circuits are used until all metric 0 circuits within the failover group have gone offline, after which the metric 1 circuits become active. Metric 0 and metric 1 circuits belong to the same VE_Port (the same extension trunk), which means that during failover and failback, no data in flight is lost or sent out of order to the upper-layer protocol.

There are two primary cases for using metrics.

First, when a backup circuit needs to be passive until a production circuit goes offline. Passive means there is a low quantity of traffic, such as keepalives; nevertheless, there is minimal traffic.

Second, use metric 1 backup circuits when licensing doesn't permit the aggregate maximum needed to retain bandwidth after a production circuit has gone offline. For example, extension platforms calculate aggregate bandwidth based on currently active circuits, and only metric 0 or 1 can be active at any time. It is common to use a failover group for each circuit; therefore, each circuit has a backup circuit. See Figure 5-2.
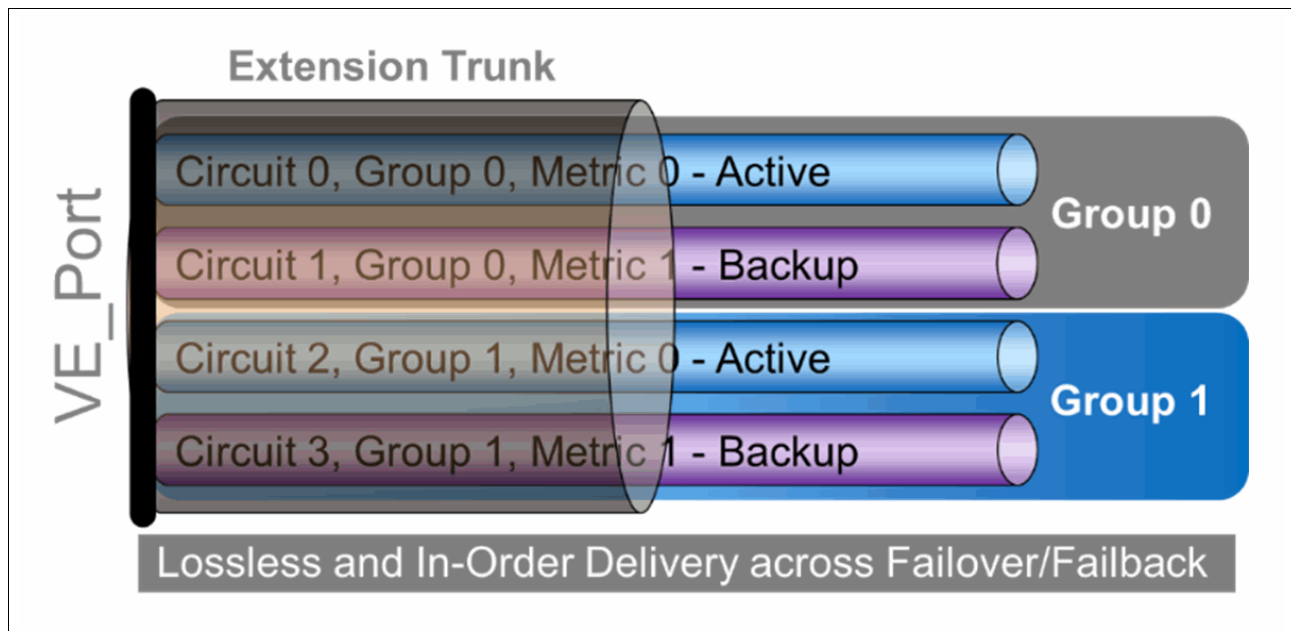
Figure 5-2   Circuit failover metrics and groups

5.2.4 KATOV

IBM b-type extension circuits use keepalives to determine circuit health and to ensure that higher-level FC or TCP/IP extended flows do not experience protocol timeouts caused by WAN or routing outages. Five keepalives are sent during the keepalive timeout value (KATOV) unless it would cause a keepalive to be greater than 1 second, in which keepalives are sent at a maximum of 1-second intervals. Each keepalive arrival resets the timer. If the timer expires, the circuit is deemed offline.

Keepalives are injected into the same transport as data. Data is never lost in transit; therefore, keepalives are never lost. Massive IP network congestion and dropped packets can cause keepalives to be delayed, which can cause the time to be extended and could cause a circuit to drop.

Keep the KATOV short enough because WO-TCP can quickly recover through brief outages or bouts of congestion. When the timer is too short, circuits can be inadvertently dropped when they should not have been. Conversely, a longer KATOV will take more time to detect failed circuits.

FICON and FCP circuits have different default keepalive values. FICON has stricter timing than FCP and must have no more than a 1-second KATOV. Specify `--ficon` when creating a tunnel with circuits that carry FICON.

If there is only one circuit in a tunnel, the proper KATOV is probably the application timeout value plus 1 second. A circuit may drop if the KATOV is set too short when the IP network has over subscription, congestion, long convergence times, or deep buffers.

If multiple circuits are in a tunnel, set the KATOV to a value less than the overall application protocol timeout, so each circuit could fail with a KATOV before the protocol timer expires. Most supported RDR application have a 6-second application protocol timeout. If there are two circuits, set a 2.5-second KATOV for each circuit. If there are three circuits, set a 1.8-second KATOV for each circuit, and so on.

5.2.5 Encryption

Would your company operate Wi-Fi with no encryption? Of course not! A best practice is to use IBM b-type extension IPsec to protect data end-to-end. When you implement IBM b-type extension, it is always prudent to enable IPsec. Data leaving its secure confines into a service provider's infrastructure is vulnerable to opportunistic attack, and for data in flight, there is no service provider security guarantee. Links must be authenticated and data encrypted to prevent eavesdropping and attacks. IBM b-type IPsec is easy and practical to deploy.

IBM b-type IPsec is a hardware implementation that operates at line rate. IPsec includes no additional licenses or costs in all base extension platforms. Encryption adds a propagation delay of approximately 5 microseconds. Pre-shared key configuration is easy.

Using IBM b-type IPsec removes the need for a firewall. A best practice is to connect extension directly to the WAN routers or switches and to avoid intermediate devices such as firewalls. IPsec IKEv2 uses UDP port 500 as its destination.

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

Firewalls are not considered best practice for the following reasons:

- ► Per DP, IBM b-type IPsec operates at line rate on the WAN side, which on the IBM SAN45B-R7 is 50 Gbps. Most firewalls cannot meet this throughput requirement.
- ► Storage traffic requires minimal propagation delay.
- ► IBM b-type IPsec encrypts data closer to the data source and destination, which is considered best practice.
- ► Firewalls and WAN optimization devices may proxy TCP sessions, resulting in remote IP packets not being identical to the originals, which is not supported. These devices are not supported.

5.2.6 Compression

Fast-Deflate, Deflate, and Aggressive-Deflate are the compression algorithms implemented on the IBM SAN45B-R7 and IBM SX6 Extension Blade. Compression operates in the tunnel scope, not per circuit. Compression must be configured identically on both ends of the tunnel; asymmetrical compression is not supported.

The IBM SAN18B-R has Deflate and Aggressive-Deflate; it does not have Fast-Deflate. The IBM SAN18B-R does not support the throughput warranting Fast-Deflate hardware and can achieve higher compression ratios using Deflate and Aggr-Deflate while meeting throughput capacity. Fast-Deflate is unavailable for IP Extension on any IBM b-type extension platform, and IP Extension compression is limited to Deflate and Aggressive-Deflate.

Compression can be configured specifically for each protocol (FCIP and IP Extension). For example, on an IBM SAN42B-R7, configure Fast-Deflate for FCIP and Deflate for IP Extension. Using these algorithms for each protocol is considered best practice because the Fast-Deflate and Deflate compression engines differ. Fast-Deflate uses an FPGA engine, and Deflate uses a hardware engine in a DP processor. 20 Gbps of Fibre Channel ingress to the Fast-Deflate engine does not consume any IP Extension Deflate or Aggressive-Deflate compression engine resources.

Compression is recommended with RDR applications, including RDR/S. Tape data is commonly compressed, and attempting to compress data again is not helpful.

5.2.7 SLA

The primary purpose of SLA (service-level agreement) is to provide automated circuit testing before placing it back into service. SLA checks the circuit for packet loss. If you must verify the circuit for network performance, such as throughput, congestion, and out-of-order delivery, use the WAN Test Tool to run additional tests manually.

Before testing can commence, configure matching SLA sessions at each end of the circuit. SLA uses the circuit configuration information to establish an SLA connection. If the circuit ends specify different transmission rates, SLA uses the lower configured rate. This allows SLA to run when circuit configurations have a minor mismatch. When the session is established, traffic starts automatically. During the test, the packet loss percentage must remain under the specified amount before the circuit can be placed into service. On an IBM SAN42B-R7, a maximum of 20 SLA sessions can be defined per DP. On an IBM SX6 Blade, up to 20 SLA sessions can be defined per DP. On an IBM SAN18B-R, a maximum of 12 sessions can be defined.

In addition to packet loss, SLA can test for timeout duration. If the timeout value is reached during an SLA session, the session is terminated, and the circuit is put into service. A timeout value of none means the test runs until the runtime and packet-loss values are met.

SLA interaction is limited while it is running. You can view statistics, and you can abort an active session. Any attempt to modify a session while it is active is blocked, which means the WAN Test Tool commands cannot be used while an SLA session is running.

Whenever a tunnel or circuit goes offline and comes back online, or when a circuit is administratively disabled and enabled, an SLA session is started. It tests the link before allowing the circuit to go back into service. Configured SLA sessions are persistent across reboots because circuit configurations are persistent across reboots, and SLA is part of circuit configuration. However, user-configured WAN Test Tool sessions are not persistent.

6

Extension refresh guidance

This section guides updating the IBM b-type extension technology from legacy platforms to the latest IBM Gen7 SAN42B-R7 platform.

In this chapter covers the following sections:

► "Introduction to migration" on page 54
► "SAN Health" on page 59

6.1 Introduction to migration

The chapter should be used to help in the migration from legacy extension platforms to modern IBM b-type Gen 7 extension platforms.

- ► FICON Logical Switch
- ► Transitioning from the IBM SAN42B-R to the IBM SAN42B-R7
- ► Migration methodology and techniques

6.1.1 Planning

As you prepare to migrate your environment from legacy technology to Gen 7, it is a good idea to consider preparing for the transition to Brocade Fabric OS 9.2.x release.

The advent of the FICON logical switch is intended to simplify setting up IBM b-type platforms for FICON configuration and management. In earlier releases of FOS, many settings required manual configuration and needed to be validated during the process, which was prone to mistakes. FICON logical switches save valuable time and prevent users from wasting time determining the cause of connectivity errors. The IBM b-type High Integrity Fabric (HIF) limits manually induced errors, and the FICON logical switch was added to further reduce configuration errors by providing inherent configuration of the necessary parameters.

In addition to simplification, the FICON logical switch provides a mechanism to abstract the FICON director definition from the physical hardware, which can now support more ports than the FICON director architecture permits. This mechanism mimics the abstraction that logical partitions (LPARs) provide on the mainframe and is a container for all FICON connectivity.

> **Note:** A FICON logical switch is required if FICON connections are to be made; otherwise, a FICON logical switch is not required.

6.1.2 FICON Logical Switch

Mainframe customers often disable logical switches (LS) when configuring or leave Virtual Fabrics (VF) enabled using only the default switch. A VF is required for FOS 9.0 and above, and a FICON LS must be created. VF is analogous to configuring logical partitions (LPARs) on an IBM Z central processor complex (CPC).

The FICON LS is required in FOS 9.2.x, so customers planning to migrate to IBM SAN42B-R7 should plan on implementing a FICON LS during deployment rather than taking an outage during a later upgrade.

- ► FICON-specific information and control are now exposed in the RESTCONF API
- ► The Allow/Prohibit Matrix (PDCM) was deprecated in FOS 9.1

The Need for a FICON Logical Switch

A FICON logical switch facilitates the following:

- ► Improve security
- ► Ensure properly configured FICON fabrics
- ► Simplify configuration

For users, a FICON LS has the following advantages:

- ► Address binding – In prior versions of FOS, determining if a port address was bound was challenging for non-expert users. A bound address means that the FC address for the port cannot change. A bound address is vital because the IBM Z channel subsystem builds FC addresses based on the link addresses defined in the Hardware Configuration Definition (HCD) tool.
- ► Although circumstances for changing the FC address assigned to a port on the switch are unlikely, should it happen, the I/O definition file (IODF) would no longer be valid, resulting in device or channel errors. This potential problem is fixed by defining a FICON LS.
- ► Ensuring FICON switch requirements do not change – It is possible to configure an LS meeting all FICON requirements and then back out specific changes after the channels come online. A FICON LS does not permit changes not conforming to FICON requirements.
- ► Since a mainframe channel checks the proper security attributes only at login time, with an LS not explicitly designated a FICON LS, the potential exists for a channel not to come online when it goes offline and comes back online. Note that login is handled by the physical channel (PCHID), not the logical channel path (CHPID). A channel login results from a CPC initial machine load (IML) or channel path maintenance for most customers.
- ► Address mode – Addressing mode restrictions depend on the switch type; these restrictions cause considerable confusion and, in some cases, limit options. A FICON LS ensures that the proper address mode is used.
- ► The address mode is essential because specific modes use the lower byte, ALPA. The lower byte must always be the same throughout the fabric because IBM Z builds FC addresses from link addresses.

6.1.3 Transitioning from the IBM SAN42B-R to the IBM SAN42B-R7

The IBM SAN42B-R (Gen 5) is two generations older than the IBM SAN42B-R7 (Gen 7) extension switch, and they cannot run the same Fabric OS version. The IBM SAN42B-R cannot deliver the same performance nor support the newer Gen 7 features. Because of this and other incompatibilities, IBM SAN42B-R and SAN42B-R7 tunnel connectivity is not supported.

When introducing the IBM SAN42B-R7 into an environment with existing IBM SAN42B-R extension switches, first deploy IBM SAN42B-R7 units parallel to the IBM SAN42B-R units, preferably one location at a time. The IBM SAN42B-R units remain connected to the remote IBM SAN42B-R units. New IBM SAN42B-R7 units connected to remote IBM SAN42B-R7 units. Interconnectivity is not necessary between generations. The conservative and deliberate process of adding IBM SAN42B-R7 units facilitates transition and performance while using the Gen 7 feature set. Eventually, connectivity will be transitioned to the refreshed network.

In Figure 6-1 the new IBM SAN42B-R7 extension platforms were added to the core and each remote data center; however, connectivity was not yet relocated to the new platforms. Because the same IP addresses are being used on the new extension platforms, no changes are needed on the IP network, provided the Ethernet speeds remain consistent. No IP Extension TCLs or storage end device routing changes are needed.
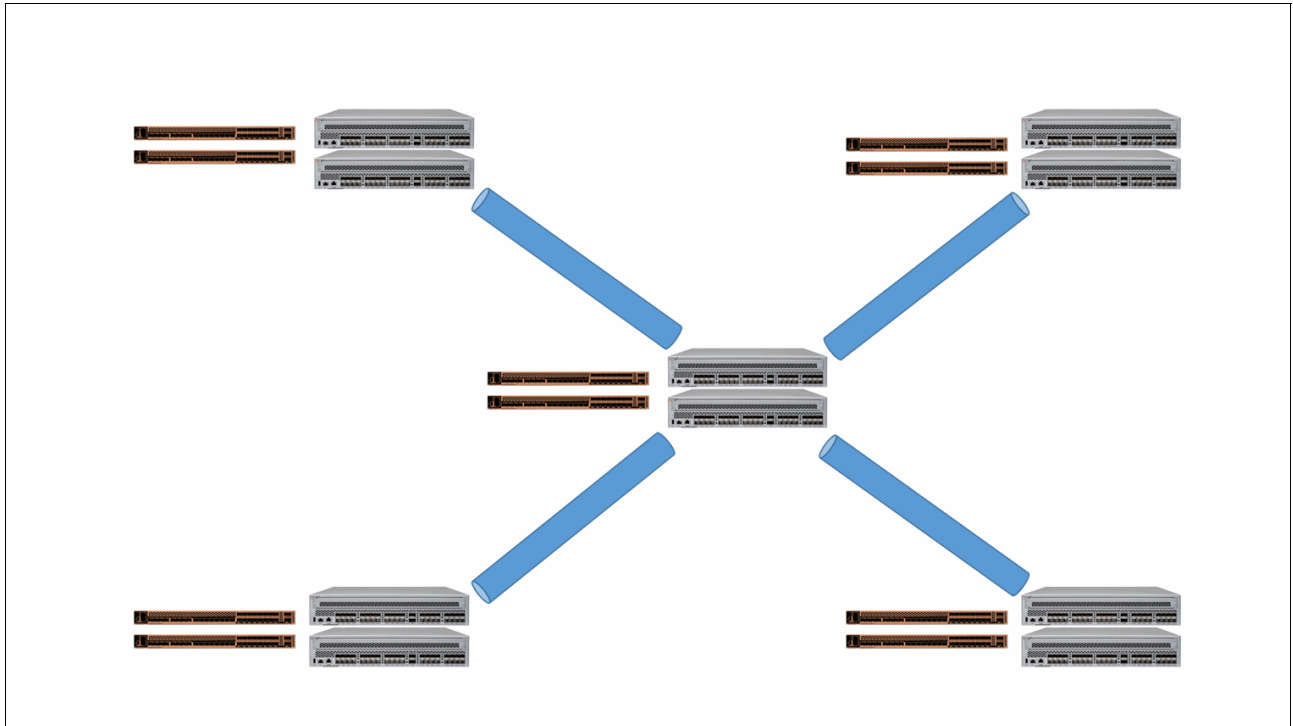
Figure 6-1   Transition to IBM SAN42B-R7 - Start, no active SAN42B-R7

In Figure 6-2, one site was migrated to the new IBM SAN42B-R7 platforms in the core and remote data centers. There are no remaining connections to the legacy extension platforms; they must be disconnected from the IP network to prevent IP address conflicts. It is best to power down the legacy extension platforms once connectivity on the new platforms has been verified.
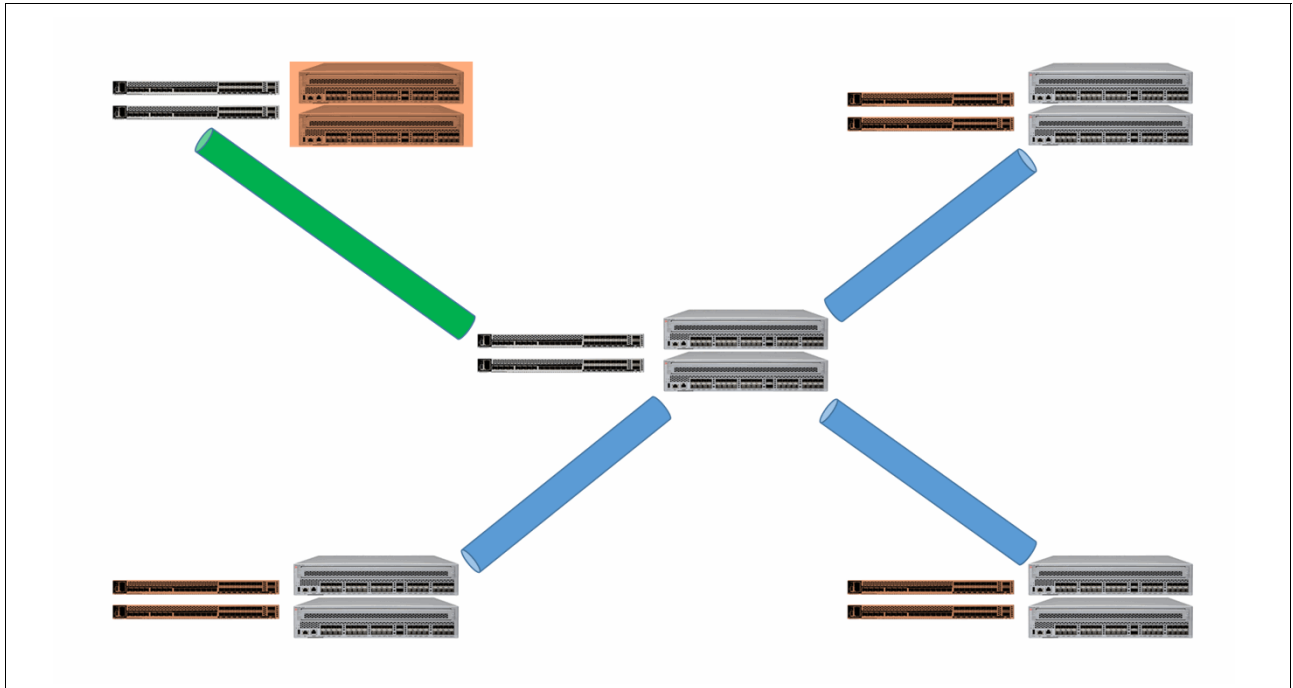
Figure 6-2   Transition to SAN42B-R7 at one location

In Figure 6-3, another site was migrated to the new IBM SAN42B-R7 platforms in the core and a different remote data center.
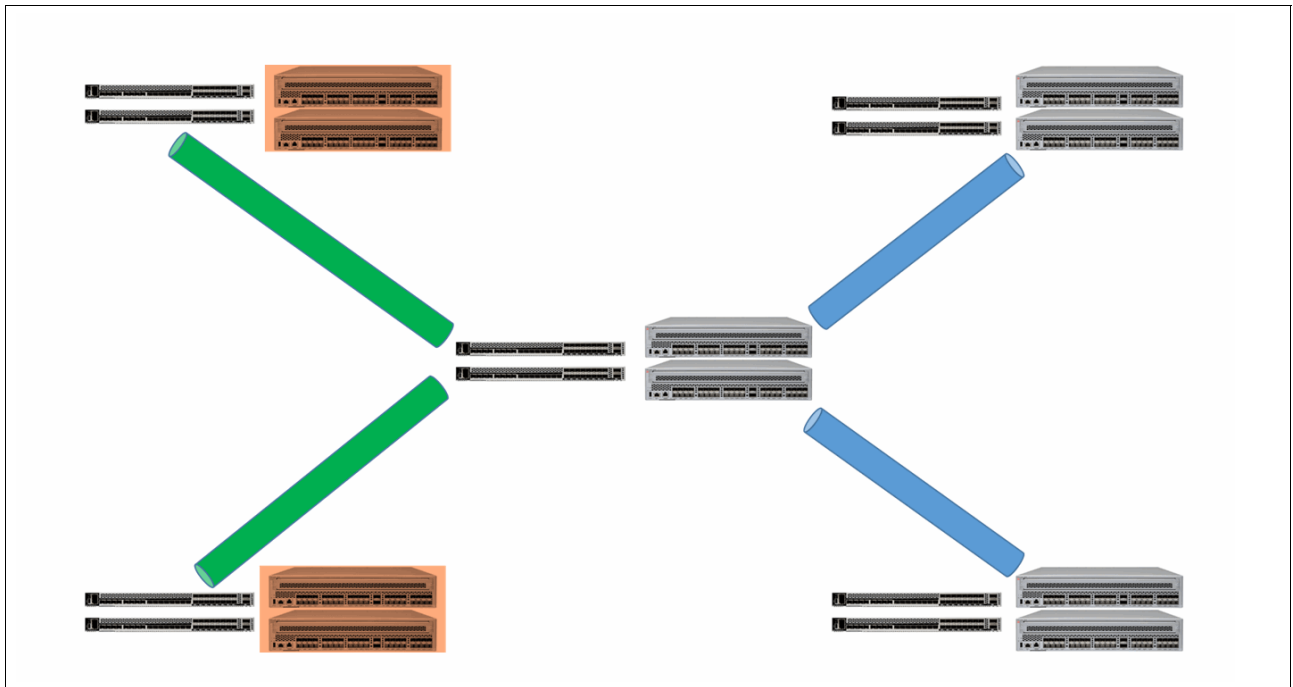
Figure 6-3   Transition to IBM SAN42B-R7, add additional site

In Figure 6-4, all the remote data centers have been migrated to the new IBM SAN42B-R7 extension platforms. All connectivity to the legacy extension platforms has been removed. Power has been disconnected from the legacy extension platforms, which are no longer operational. All storage replication traffic has been tested and verified.
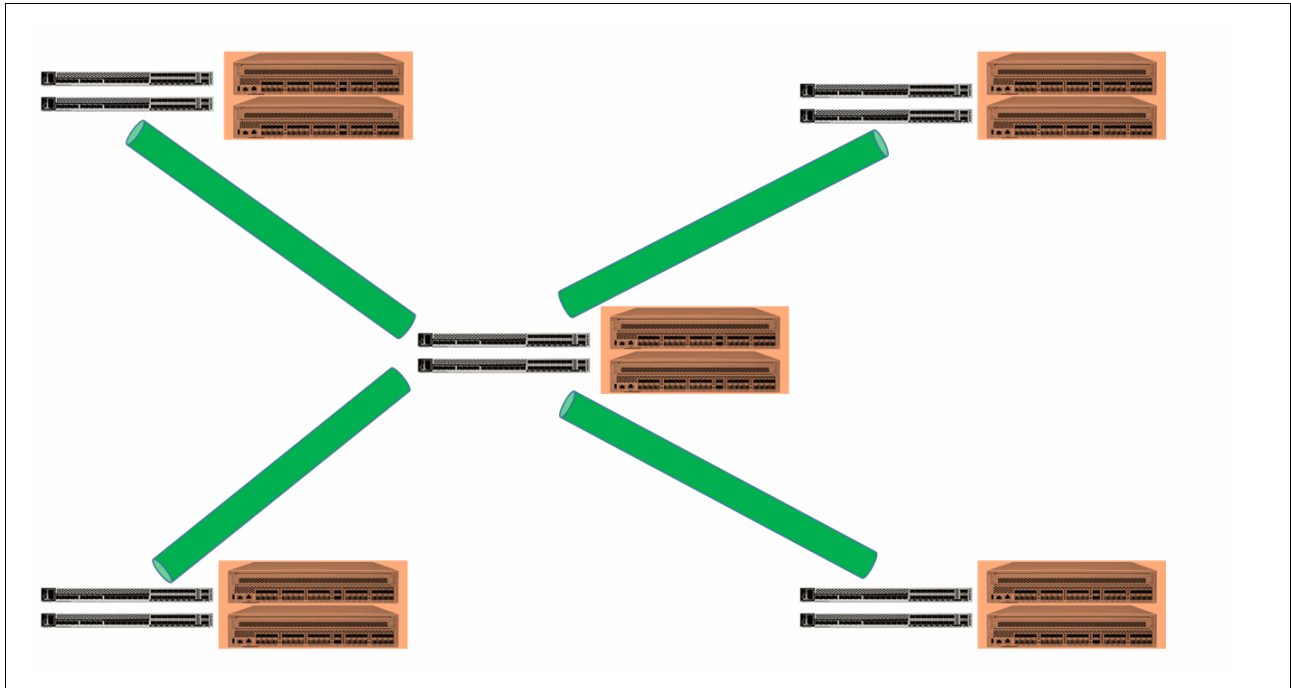
Figure 6-4   Transition to SAN42B-R7, add remaining sites

This approach provides transition flexibility. Gradually replace IBM SAN42B-R units with IBM SAN42B-R7 units – at suitable times that work for your environment. As older IBM SAN42B-R units are replaced, replication traffic gradually moves to IBM SAN42B-R7 units, and the core IBM SAN42B-R units handle less and less traffic. Once all IBM SAN42B-R are replaced, they can be de-commissioned.

6.1.4 Migration methodology and techniques

The IBM SAN42B-R7 provides industry-leading performance and improved features on a platform built with proven technology. One of the most common migration inquiries is, "Will I need to take an outage, and if so, how long will that last?" The short answer is "maybe." Users can avoid an outage by following proper conversion methods.

To remove the risk of an outage, stage and configure new platforms to mimic in-use configurations. Suppose you are embarking on a consolidation effort, increasing the number of WAN connections, or upgrading to a higher-speed WAN; additional steps are needed to integrate such changes. For example, new IP addresses, reuse, or changes may be required. Additionally, new or different interface connections, IP route changes, VE_Port numbers on a new platform, bandwidth changes, rate limiting changes, and TCL changes are all things that need to be considered when migrating.

A dual-fabric topology provides a seamless transition from one generation platform to another. Using one of two paths, replication keeps flowing without an outage. Use the figures above (Figures 8 to 11) as the example network.

An outage can be eliminated by converting one path at a time. Follow the guidelines below for a one-for-one replacement of IBM SAN42B-R to an IBM SAN42B-R7 with no changes to FC connections or the IP network:

1. Preliminary work: Run Brocade SAN Health to capture a snapshot of your environment and see if you have any current "network health" problems that need attention before the migration. See Additional Resources for more information about Brocade SAN Health.
2. Gather information from the IBM SAN42B-R required to build and configure the replication connections on the IBM SAN42B-R7, such as IP interfaces (IPIF), IP routes, eHCL, IPsec, compression, tunnels (VE_Ports), circuits, QoS settings, Fastwrite, OSTP, failover metrics, failover groups, TCL, etc.
3. Install the IBM SAN42B-R7 and configure the management IP address to gain access to the platform.
4. Leave disconnected all FC and Ethernet data connections.
5. Apply power to the IBM SAN42B-R7 in Sites A and B.
6. There are two ways to configure the IBM SAN42B-R7: IBM SANnav or the CLI.
7. Configure the IBM SAN42B-R7 in sites A and B with the information gathered.
8. Select which fabric you intend to migrate first, and if required by the storage vendor's process, suspend the path for the mirrored volume. Review your storage vendor's administration guide for the proper process to take down one or more mirrored paths.

> **Note:** The other fabric will continue replicating between arrays with a dual-fabric architecture. However, there may be a replication backlog since only half the bandwidth is available.

9. Disable power to the new IBM SAN42B-R7.
10. To migrate the first path, turn off power to those IBM SAN42B-R in Site A and Site B.
11. Remove the FC and Ethernet connections from each IBM SAN42B-R, and reattach the cables to the appropriate ports on the replacement IBM SAN42B-R7.
12. Apply power to the IBM SAN42B-R7 in sites A and B.
13. Check for error conditions and messages.
14. While the new path is offline to the mirror, run the IBM SAN42B-R7 traffic generator (WAN Test Tool) to confirm that the IP network and WAN reliably support the data requirements.
15. Once the path has been verified, re-enable the mirror (if required).
16. Monitor the mirror, the IBM SAN42B-R7, and the network for errors. Confirm that replication is running as expected.
17. Once the fabric works as expected, repeat these steps for the other fabric.

6.2 SAN Health

Brocade SAN Health is an easy-to-use fabric auditing tool provided to all IBM customers and partners free of charge. SAN Health works with IBM b-type extension platforms. With a simple download and installation to any modern Windows machine with connectivity to the switches to be audited. SAN Administrators get a comprehensive report and diagram of their SAN environment and attached devices. For additional product information, go to www.broadcom.com/sanhealth.

This session will cover the flowing topics:

► SAN Health overview

► SAN Health Diagnostics Capture

► Installing SAN Health

6.2.1 SAN Health overview

An easy way to complete labor-intensive configuration checking, documentation, and inventory management tasks. With the free SAN Health Diagnostics Capture utility, you will be able to:

► Inventory devices, switches, firmware versions, and SAN fabrics

► Capture and display historical performance data

► Compare zoning and switch configurations against best practices

► Assess performance statistics and error conditions

► Produce detailed graphical reports and diagrams

SAN Health gives you a powerful tool that helps you focus on optimizing your SAN rather than manually tracking its components. Various useful features make it easier to collect data, identify potential issues, and check your results over time. As a result, you can significantly increase your productivity while enhancing your SAN operations.

6.2.2 SAN Health Diagnostics Capture

SAN Health uses a data capture application and a back-end report processing engine. After capturing switch diagnostic data, it automatically generates a Visio topology diagram and a detailed fabrics, switches, and ports report. Other helpful information includes alerts, historical performance graphs, and recommended best practices.

Because it provides a point-in-time SAN snapshot, the SAN Health Diagnostics Capture is invaluable to your change-tracking process.

You can download Brocade SAN Health Diagnostic Capture here.

With SAN Health, you can generate personalized storage network performance and inventory reports to help optimize operations. SAN Health automatically discovers, analyzes, and reports the critical characteristics of your SANs, including switch, topology, and performance details. Results are presented in standard Excel and Visio formats. Figure 6-5 through Figure 6-8 show you some example SAN Health reports.

SAN SUMMARY DETAILS FOR SAN_EXAMPLE

SWITCHES IN SAN SAN_Example

Fabric Name	Switch Name	Domain	IP Address	World Wide Name	Model	Speed	OS Ver	Ports	Unused
Storage_Edge	sw3200_32	32	192.168.163.32	10:00:00:60:69:c0:06:55	3200	2G	3.2.1a	8	1
Storage_Edge	sw4100-41	41	192.168.163.41	10:00:00:05:1e:34:56:5e	4100	4G	5.1.0d	32	24
Storage_Edge	sw3850-50	50	192.168.163.50	10:00:00:05:1e:34:12:20	3850	2G	5.0.1a	16	10
Server_Edge	sw3800_38	38	192.168.163.38	10:00:00:60:69:50:08:7e	3800	2G	3.2.0a	16	4
Server_Edge	sw48000-48	48	192.168.163.48	10:00:00:60:69:e4:25:18	48000	4G	5.1.0d	48	39
Server_Edge	sw24000-24	24	192.168.163.24	10:00:00:60:69:e2:03:b0	24000	2G	5.1.0d	32	21
Server_Edge	sw3900-39	39	192.168.163.39	10:00:00:60:69:90:0c:a3	3900	2G	5.1.0d	32	23

HEALTH AND MONITORING STATUS FOR SAN_Example

Fabric Name	Switch State Marg	Switch State OK	Power Supplies Bad	Power Supplies Marg	Power Supplies OK	Fans Bad	Fans Marg	Fans OK	Temp Sensors Low	Temp Sensors OK	Temp Sensors High	Errors Lvl1	Errors Lvl2	SNMP No	SNMP Yes	SysLog No	SysLog Yes
Storage_Edge	0	3	2	0	2	0	0	12	0	12	0	0	0	3	0	3	0
Server_Edge	3	2	1	0	7	0	0	19	0	17	0	0	0	5	0	5	0
TOTALS	4	5	4	0	10	0	0	34	0	35	0	0	0	9	0	9	0

PORT USE

Fabric Name	Port Use Disk	Port Use Tape	Port Use Host	Port Use ISL	Port Use Free	Port Use Total	Fan Out Ratios Host:Disk	Fan Out Ratios Port:ISL	Fan Out Ratios Device:ISL	Port Long Distance Modes 10km	Port Long Distance Modes 25km	Port Long Distance Modes 50km	Port Long Distance Modes 100km	Port Long Distance Modes Auto
Storage_Edge	5	0	0	16	35	56	0:5	2.5:1	0.31:1	56	0	0	0	0

ZONING METRICS

Fabric Name	Zone Database Use	Aliases Statistics Aliases	Aliases Statistics AvMem	Aliases Statistics MaxMem	Aliases Statistics Hanging	Zone Statistics Zones	Zone Statistics AvMem	Zone Statistics MaxMem	Zone Statistics Hanging	Config Statistics Configs	Config Statistics AvMem	Config Statistics MaxMem	Config Statistics Hanging
Storage_Edge	0.8% of 258k	24	1	1	19	11	4.5	15	1	1	10	10	1
Server_Edge	0.9% of 258k	30	1	1	11	11	4.9	20	1	1	11	11	1
TOTALS		54	0.7	1	30	22	3.1	20	2	2	7	11	2

Figure 6-5   Example of SAN summary details
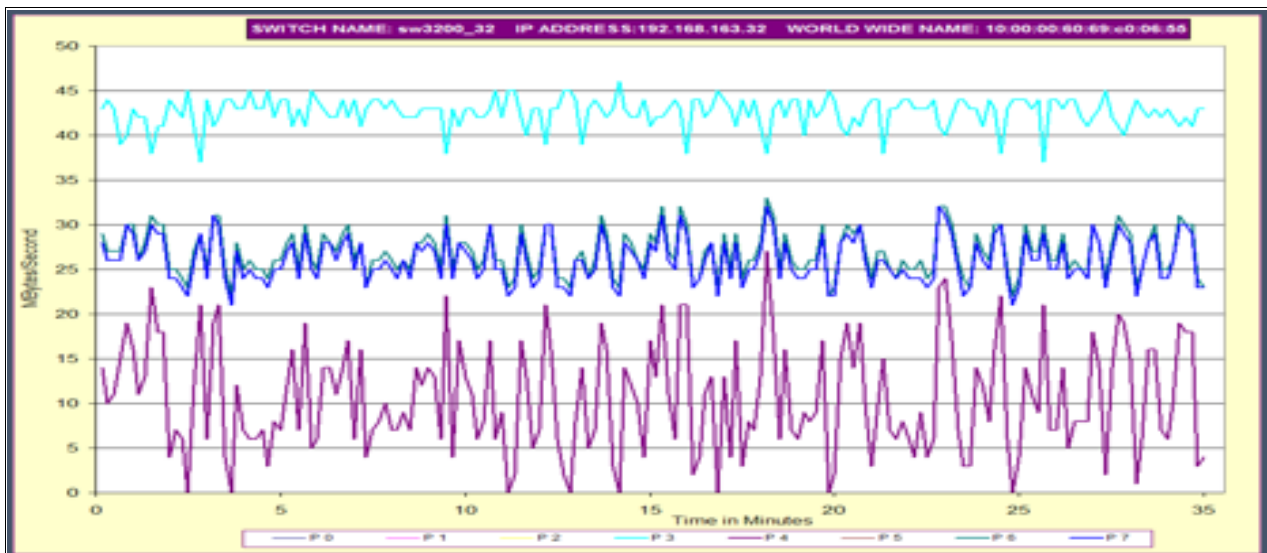
Figure 6-6   Example of historical performance graphs

Chapter 6. Extension refresh guidance   **61**
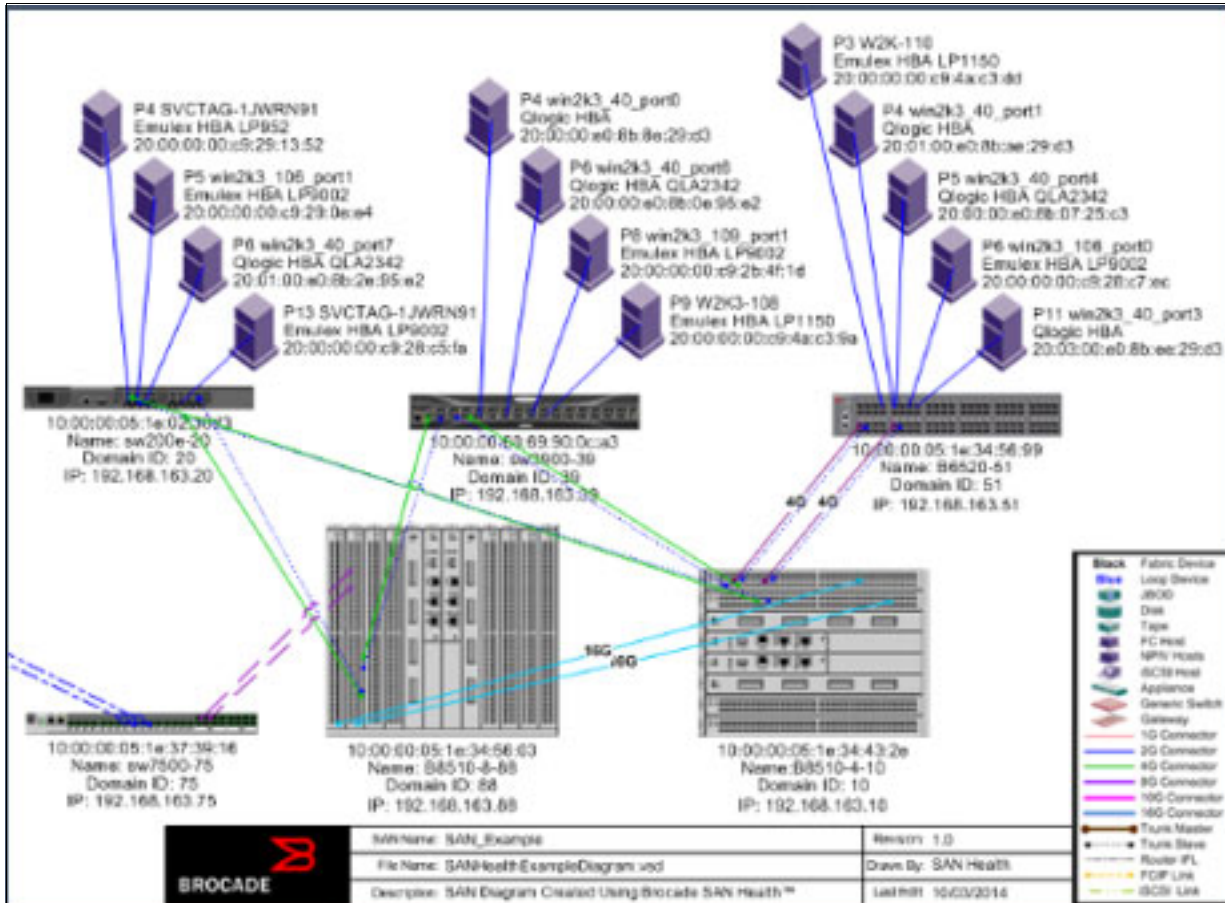
Figure 6-7   Example of architecture connectivity

Figure 6-8   Inventory, FRU, Location, Status, and Up Time List

6.2.3 Installing SAN Health

A few considerations when installing SAN Health:

► Ensure minimum desktop or laptop system requirements:
  – Intel P4 or AMD equivalent (AMD K7)
► Download SAN Health Diagnostics Capture
► Run file `InstallSANHealth.exe`
  – Install on any Windows-based PC that has TCP/IP connectivity to the SAN
  – Generates an encrypted results file (.BSH)
► Generate your reports
  – Submit results file (.BSH)
  – Submit via email or online upload
  – Receive a report generation notification email within 1 to 8 hours
► Download reports from the secure Broadcom Customer Support Portal (CSP)

If you have no Broadcom CSP account, SAN Health will create one for you.

7

IBM b-type Extension configuration

This chapter covers the basic configuration of IBM b-type Extension. The steps include:

1. Establish the replication fabric architecture.

2. Gather the necessary information to configure your replication fabric.

3. Configure each IBM b-type extension platform.

4. Validate each IBM b-type extension platform.

5. Validate the FC replication traffic across extension.

6. Redirect end-device IP storage traffic to the appropriate IP Extension gateway.

7. Validate the IP storage traffic across IP Extension.

The following topics are covered in this chapter:

7.1 Configuration assumptions and prerequisites

> **Note:** Initial FOS fundamental settings configuration are not in the scope of this document.

The assumptions listed here are specific to this deployment guide and example. You can change configuration parameters to suit your own needs and environment.

► Cabling (power, management, LAN Ethernet, and WAN Ethernet) is complete.

► IBM b-type extension requires both ends to operate with the same Fabric OS version.

► The IBM b-type extension platforms are installed, powered, and initially configured with fundamental settings such as: domain ID, DNS, NTP, and management port access.

► IBM b-type extension platforms and management network are operational.

► If IP Extension will be deployed with two data center LAN switches at a site, the two switches must form a single logical switch (i.e., vPC, VLT, MLAG, or VCC).

► You have administration access to all switches that you need to configure. The default username and password is admin and password.

► IP WAN network is provisioned, up, and operational.

► IBM SAN42B-R7 extension platforms do not have any preexisting extension configuration.

► 100GE interfaces are used for the WAN side.

► 25GE interfaces are used for the LAN side.

► The two data center LAN switches have redundant 25GE connections to the IBM SAN42B-R7 LAN side, one link from each.

► The two data center WAN switches have one IBM SAN42B-R7 100GE connection to each. There is one circuit per link.

► The LAN side uses a default MTU of 1500 bytes.

► The WAN side uses an MTU of 9216 bytes.

► You have the LAN information of your IP storage end devices.

► The IP Extension gateway is on the same VLAN as the end devices.

► No VLAN tagging is required on the LAN or WAN sides.

► No QoS implementation is required on the LAN or WAN sides.

► The WAN can accommodate up to 100 Gbps of data traffic.

► A maximum of 50 Gbps (ceiling) and a minimum of 30 Gbps (floor) of bandwidth per tunnel will suffice for the replication applications.

► The VE_Port is 24 (Tunnel ID 24). Typically, the first VE_Port is used.

► The tunnel comprises two circuits: VE24–Circuit0 and VE24–Circuit1.

► The following SAN42B-R7 Ethernet interfaces are used:

  – GE2–GE5 (IP Extension: LAN side)
  – GE0–GE1 (tunnel: WAN side).

Table 7-1 shows the WAN IP subnets and masks that are used at each circuit's endpoint.

Table 7-1   WAN IP subnets and masks used at each circuit's endpoint

Side	Subnet	Cir0	Cir1
Local WAN	172.16.1.0/24	172.16.1.3	172.16.1.4
Remote WAN	172.16.2.0/24	172.16.2.3	172.16.2.4

Table 7-2 on page 67uses the LAN IP subnets and masks that are used to assign IP Extension gateway addresses.

Table 7-2   LAN IP subnets and masks used to assign IP Extension gateway addresses

Side	LAN IP Subnet	IP Extension Gateway
Local LAN	10.0.0.0/24	10.0.0.2
Remote LAN	192.168.1.0/24	192.168.1.2

Table 7-3 shows that gateway addresses are on the same IP subnet as the WAN circuit's IP interface.

Table 7-3   WAN gateway addresses

Side	IP Extension WAN Gateway
Local WAN	172.16.1.0/24
Remote WAN	172.16.2.0/24

Table 7-4 shows the minimum and maximum values represent the adaptive rate limiting's (ARL) rate per circuit.

Table 7-4   Adaptive rate limiting's (ARL) rate per circuit

Limits	ARL Rate
Minimum	15 Gbps (15,000,000 kbps)
Maximum	25 Gbps (25,000,000 kbps)

Note: If ARL is unnecessary, set the minimum equal to the maximum to disable ARL and enable CIR (Committed Information Rate).

Metrics are not used in this example. Metrics and groups are used for failover/failback scenarios.

On your IP storage device, redirect your replication traffic from the default gateway to the new IP Extension gateway.

7.2 FC and FICON side configuration

This section covers the configuration considerations for FC and FICON specific to extension. CLI and SANnav configuration tasks are covered in detail in the *Brocade Fabric OS Extension User Guide*, *Brocade SANnav Management Portal User Guide,* and *Brocade Fabric OS Command Reference Manual* (see https://techdocs.broadcom.com/fabric-os-extension).

Internal to the IBM SAN42B-R7 are backend FC ports that connect to two Data Processors (DP). These backend ports have special processing called VE_Ports (VE) and are logical representations of E_Ports. VE_Ports are not ports on the Brocade FC ASIC. Multiple internal FC ports feed a DP. On the IBM SAN42B-R7, the maximum FC rate per DP is 100 Gbps. A VE is considered the transition point between FC and the FCIP tunnel within an extension platform.

Legacy VEX_Ports are no longer supported on IBM b-type extension platforms.

> **Note:** Array-to-array replication (RDR) is FC-based replication. Even though a volume may be written to by FICON, the array does not use FICON to replicate to the remote array. Therefore, the considerations and caveats of FICON do not apply to array-to-array replication.

> **Note:** FICON is supported only on the IBM SAN42B-R7 and the IBM SX6 Extension Blade. It is not supported on the IBM SAN18B-6.

7.2.1 Virtual fabrics

Sometimes it is necessary to create a logical switch specifically for deploying extension. Creating a logical switch (LS) for extension requires adding a VE_Port to the LS. The FID (Fabric ID) must be the same on both ends of the extension tunnel. The FID must be unique on the platform; you cannot use the same FID more than once on the same platform.

GE interfaces are not added to the replication LS. All LSs can access the GE interfaces in the default LS. Some extension portions are configured in the default LS (IPIF and IP routes), and others in the replication LS (tunnels and circuits).

Ports can be in only one logical switch; they cannot span across logical switches. If the logical switch needs to connect to other local switches, E_Ports must be added to the logical switch, or an xISL must be used. The F_Ports connecting to the end-devices communicating across extension must be in the same logical switch as the VE_Port.

Refer to the Brocade Fabric OS Users Guide for creating, configuring, deleting, and managing Virtual Fabrics.

Example 7-1 shows the CLI syntax for the `lscfg` command.

Example 7-1   lscfg command

```
SW37_7850_A:FID128:admin> lscfg
lscfg {--create | --delete | --config | --show | --change | --help}
-------------------
lscfg --create <FID> [-b | -base] [-lisldisable] [-f | -force]
lscfg --create <FID> [-n | -ficon] [-lisldisable] [-f | -force]
        FID:          Required.  Fabric ID for this switch. Must be unique.
        -base:        Optional.  Make this switch the base switch.
                                 Only a single switch may be the base
                                 switch per chassis.
        -lisldisable: Optional.  Use this option to bring up the lisl
                                 port(s) in Offline state.
                                 When not used lisl port(s) are brought
                                 Online
        -ficon:       Optional.  Creates logical switch auto-configured
```

```
                                     with FICON configurations.
        -force:        Optional.  Foregoes all user prompts.
        Note: "base" and "ficon" options are mutually exclusive.
-------------------
lscfg --delete <FID> [-f |-force]
        FID: Required.  Fabric ID for the switch to be deleted.
                        All ports must be removed from this switch
                        prior to issuing the command.
        -force: Optional.  Foregoes all user prompts.
-------------------
lscfg --config <FID> -port  port1[port2] [-q | -qsfp] [-f |-force]
lscfg --config <FID> -index  index1[index2] [-q | -qsfp] [-f |-force]
        The qsfp option is only supported with the port
        FID:    Required.  Fabric ID for the switch.
        -port: Required.   The port to be added to the switch.
                           A range may be supplied (example: -port 6-16).
        -index: Optional.  The port index or range to be added to the switch.
                           Examples:
                                 -index 28
                                 -index 7,10-12,3,5-6
        -force: Optional.  Foregoes all user prompts.
-------------------
lscfg --show [-ge] [-instance]
        -ge:           Optional.  Display GE ports.
        -instance:     Optional.  Display LS Instance.
-------------------
lscfg --change <FID> {-base | -newfid <FID>}  [-force]
        FID:    Required.  Fabric ID for the switch.
        -newfid: Optional.  The new fid for the switch.
        -base:   Optional.  Makes this switch the base switch.
                            If the current base switch is indicated,
                            this will remove the base switch
                            attribute.
        -force:  Optional.  Foregoes all user prompts.
```

Example 7-2 shows the Logical Switch listing on an IBM SAN42B-R7. The default switch (ds) is FID 128 and contains all ports. No additional logical switches have been added.

Example 7-2   Logical Switch listing on an IBM SAN42B-R7

```
SW37_7850_A:FID128:admin> lscfg --show
Created switches FIDs(Domain IDs): 128(ds)(1)
Port       0     1     2     3     4     5     6     7     8     9
-------------------------------------------------------------------
FID    128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 |
Port      10    11    12    13    14    15    16    17    18    19
-------------------------------------------------------------------
FID    128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 |
Port      20    21    22    23    24    25    26    27    28    29
-------------------------------------------------------------------
FID    128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 |
Port      30    31    32    33    34    35    36    37    38    39
-------------------------------------------------------------------
FID    128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 |
Port      40    41
-------------------
```

```
FID     128  |  128  |
```

7.3 WAN side configuration

This section covers the configuration considerations for extension's WAN side. CLI and SANnav configuration tasks are covered in detail in the *Brocade Fabric OS Extension User Guide*, *Brocade SANnav Management Portal User Guide,* and *Fabric OS Command Reference Manual* (see https://techdocs.broadcom.com/fabric-os-extension).

7.3.1 IP network considerations

IBM b-type extension uses TCP connections over the WAN. Consult the WAN carrier and IP network administrator to ensure that the network equipment on the data path supports TCP connections.

Ensure that routers and firewalls in the data path are configured to pass extension traffic through specific ports and to provide the throughput that replication through extension requires.

IP network and WAN considerations

The following are IP network and WAN considerations:

► IPsec uses Encapsulating Security Payload (ESP is protocol ID: 50) and UDP port 500 (IKEv2).

► WAN-Optimized TCP randomly selects an ephemeral (source or initiating) port between 49152 and 65535.

► WAN-Optimized TCP uses well-known destination TCP ports 3225 and 3226.

► WAN-Optimized TCP requires TCP URG flags not to be dropped or modified.

► IBM extension sets the IP DF bit (Do Not Fragment bit). Fragmentation in the IP network is not supported.

► IBM extension PMTU requires end-to-end ICMP echo and replies.

► A supported RTT of 200 ms @ 0.1% packet loss with all KATOV values.

► A maximum supported RTT of 250 ms @ 1.0% packet loss with KATOV of 2 seconds or more.

► These limits are the same on all IBM extension platforms.

To enable resiliency against a WAN failure or device outage, ensure redundant network paths are available from end to end. Ensure that the underlying WAN infrastructure supports the redundancy and performance expected in your implementation. If you configure jumbo frames on the extension platforms, ensure the network supports it.

> **Note:** IBM b-type extension sets the IP DF bit (Do Not Fragment). Fragmentation in the IP network is not supported.

> **Note:** Use the IPIF AUTO option if the network's MTU is unknown. This option will avoid unexpected circuit behavior due to unsupported MTU sizes.

7.3.2 Staging configuration method

IBM b-type extension configuration can be simplified by performing the configuration in steps. Brocade Fabric OS facilitates staging by allowing common pieces of the configuration to be entered at a time. For example, first create the tunnel with the VE_Port, IP Extension, IPsec, and Compression settings as applicable using the command:

`portcfg fciptunnel <ve> create`

Then create and add the circuits, with no settings, using the command:

`portcfg fcipcircuit <ve> create <cir_num>`

Change the command verb from create to modify for `portcfg fcipcircuit`, and add the `--min` and `--max` bandwidth values. Next, remove the `--min` and `--max` arguments and reissue the command with the `--local-ip` and `--remote-ip` addresses. Again, reissue the command with the `--local-ha-ip` and `--remote-ha-ip` addresses for eHCL. Each step is easy, less error-prone, and has a clear objective for the final configuration. Staging permits smaller, more manageable, and readable CLI input vs. long, unwieldy entries. Cut, paste, and up-arrow are helpful tools when working with staging.

7.3.3 GE interfaces (WAN side)

GE interfaces can be either WAN or LAN side. The default is WAN side. Configure GE interfaces for proper protocol and speed before continuing with other configuration. Use the command as shown in Example 7-3.

Example 7-3   portcfgge command

```
SW37_7850_A:FID128:admin> portcfgge
Usage: portCfgGe [<slot>/][<port>]
[{ --enable -autoneg | --disable -autoneg }]
[--set { -speed <speed> | -lan | -wan | -channel <channel> | -fec <fec> }]
[--show [-lmac] [--help]
Port Format:
   ge#
Args:
   --enable -autoneg      - Enable the auto-negotiate mode of the 1 GE ports
only.
   --disable -autoneg     - Disable the auto-negotiate mode of the 1 GE ports
only.
   --set -speed <speed>   - Set the port speed of the GE ports. Allowable
speeds:
                            { 1G | 10G | 25G } (port dependent)
   --set -lan             - Set the port as lan.
   --set -wan             - Set the port as wan.
   --set -channel <channel> - Set the port tunable SFP channel ID of the 10 GE
ports only.
                            Allowable channel ID Range [1] - [102]
   --set -fec <fec>       - Set the port FEC clause Supported values are
(depending on port/speed):
                            { Off | CL108 | CL91 }
   --show                 - Show the current GE port configurations.
   --show -lmac           - Show the Local MAC addresses for GE/LAN ports.
   --help                 - Show this usage statement.
```

Output from the `portcfgge` command is shown in Example 7-4 on page 72.

Example 7-4   portcfgge command output

```
SW37_7850_A:FID128:admin> portcfgge --show
 Port        Speed    Flags    Channel  FEC   Lag Name
------------------------------------------------------------------------
 ge0         10G      ----     N/A      Off   -
 ge1         10G      ----     N/A      Off   -
 ge2         10G      ----     N/A      Off   -
 ge3         10G      ----     N/A      Off   -
 ge4         10G      ----     N/A      Off   -
 ge5         10G      ----     N/A      Off   -
 ge6         10G      ----     N/A      Off   -
 ge7         10G      ----     N/A      Off   -
 ge8         10G      ----     N/A      Off   -
 ge9         10G      ----     N/A      Off   -
 ge10        10G      ----     N/A      Off   -
 ge11        10G      ----     N/A      Off   -
 ge12        10G      ----     N/A      Off   -
 ge13        10G      ----     N/A      Off   -
 ge14        10G      ----     N/A      Off   -
 ge15        10G      ----     N/A      Off   -
 ge16        100G     ----     N/A      CL91  -
 ge17        100G     ----     N/A      CL91  -
-------------------------------------------------------------------------
Flags: A:Auto-Negotiation Enabled  C:Copper Media Type
L:LAN Port G: LAG Member
```

Example: Setting the speed of ge2 to 25 Gbps:

```
SW37_7850_A:FID128:admin> portcfgge ge2 --set -speed 25G
Operation Succeeded.
```

It shows FCIP as the WAN side protocol, although IP Extension is also supported if enabled on the tunnel.

GE interface considerations
The following lists the GE interface considerations:

► An extension GE interface can be set for either WAN or LAN

► GE interfaces default to FCIP (legacy term for WAN side)

► GE interfaces are not associated with any particular set of VE_Ports until designated by an IPIF. The specified DP must own the VE_Port to be used.

► Ensure the GE interface speed (1, 10, 25 Gbps) configuration is complete before creating IPIF, IP routes, tunnels, and circuits.

7.3.4 VE_Ports (WAN side)

Tunnels are identified by their VE_Port number. The VE_Port number designates the local extension tunnel; the number can differ on the opposite side. There is no requirement for VE_Port numbers to be identical at each end of a tunnel.

Once a tunnel exists, you can create circuits and add them to the tunnel. Each tunnel can have multiple circuits, and each circuit endpoint uses an IP interface (IPIF) and associated

GE interface. Creating circuits and adding them to the same tunnel groups them into a Brocade Extension Trunk (BET). Each tunnel can use multiple GE interfaces; however, a circuit cannot span multiple interfaces.

If a VE_Port is disabled, the tunnel is brought down. Disabling and enabling a VE_Port bounces the tunnel.

Multiple VEs in different virtual fabric logical switch FIDs can share a GE interface if the interface remains in the default logical switch (FID 128). If a GE interface is moved to a non-default LS, the interface can only be used by VE_Ports in that FID. The best practice is to keep the GE interfaces in the default LS.

For BET, the DP owning the VE_Port controls all member circuits. There is no distributed processing, load sharing, in-order delivery, or LLL (lossless link loss) across DPs. FCIP failover from one DP to another is done at the FSPF level, provided another DP and the configuration and architecture permit it. The only component that can be shared is a GE interface.

IP networks can route circuits over separate paths based on the destination IP addresses, VLAN tagging, QoS marking, and other Layer 2, Layer 3, and Layer 4 attributes. GE interfaces on the IBM SAN42B-R7 provide connectivity to the network for one or more circuits.

7.3.5 Link Layer Discovery Protocol (LLDP)

LLDP is enabled by default, and preset global parameters are applied to all GE interfaces.

► IBM b-type extension supports LLDP on LAN and WAN side GE interfaces.
► LLDP is an Ethernet Layer 2 protocol and operates on the data link. It is not IP-based.
► LLDP advertises and receives network switch and port identities specific to the data link.
► LLDP uses keepalives on the GE interfaces; do not be confused with circuit Keepalive Timeout Values (KATOV)

Table 7-5 shows the LLDP timeout values and transmission intervals:

Table 7-5   LLDP timeout values

Protocol	Default Hello Interval	Hello Interval Range	Minimum Timeout	Maximum Timeout	Multiplier (mx)
LLDP	30 seconds	4 to 180 seconds	8 seconds (min hello x min mx)	1800 seconds (max hello x max mx)	2 to 10 120 seconds at default hello interval

LLDP considerations

The following are the LLDP considerations:

► The IBM SAN42B-R7 supports LLDP on LAN and WAN side GE interfaces.

► LLDP is not supported on the management port.

► LLDP is an Ethernet Layer-2 protocol and operates only on the data link. It is not IP-based or IP routable.

► LLDP advertises and receives network switch and port identities specific to the Ethernet link.

LLDP uses keepalives on the GE interfaces. These keepalives should *not* be confused with extension circuit Keepalive Timeout Values (KATOV).

Use the following command to view the LLDP relationship. SW37_7850_A is connected to SW38_7850_B on interfaces ge16 and ge17.

Typically, the default settings for LLDP are sufficient. LLDP is an open protocol and works with most networking equipment. Most networking equipment supports LLDP. Use the command below to display Ethernet connectivity to devices that support LLDP:

```
lldp --show -nbr
SW37_7850_A:FID128:admin> lldp --show -nbr
Local Dead     Remaining Remote Chassis        Tx    Rx     System
Intf  Interval Life      Intf   ID                          Name
ge16  120      104       ge16   0acd.1fd8.0000 8837  8816   SW38_7850_B
ge17  120       92       ge17   0acd.1fd8.0000 8757  8735   SW38_7850_B
```

Example 7-5 shows the CLI syntax for the **lldp** command.

Example 7-5   CLI syntax for lldp

```
SW38_7850_B:FID128:admin> lldp
Usage:
lldp --create -profile <profile_name>
lldp --delete -profile <profile-name>
lldp --config -sysname <system name>
lldp --config -sysdesc <system description>
lldp --config -mx <multiplier> [-profile <profile_name>]
lldp --config -txintvl <interval> [-profile <profile_name>]
lldp --enable
lldp --enable -tlv <tlv name> [-profile <profile_name>]
lldp --enable -port <port_num | port-range> [-profile <profile_name>]
lldp --disable
lldp --disable -tlv <tlv name> [-profile <profile_name>]
lldp --disable -port <port_num | port-range> [-profile <profile_name>]
lldp --clear -nbr [<port_num | port-range>]
lldp --clear -stats [<port_num | port-range>]
lldp --show
lldp --show -nbr [<port_num | port-range>] [-detail]
lldp --show -stats [<port_num | port-range>]
lldp --show -profile [<profile-name>]
lldp --show -port <port_num | port-range>
lldp --default
Specifying port ranges:
        port_num -- <port_num> is <[slot]/port>
        port_range -- Specifies a set of ports as a range:
       (examples: '10/6' or '10/6-9' or '10/ge6-ge9' or '10/ge6-7' or 'ge6-ge7'
         or 'ge6-7')
Actions:
--create  : Creates lldp profile. Profile name is a string of max 32 character
            Characters allowed: alphanumeric with special char underscore(_)
--delete  : Deletes the specified lldp profile
--config  : Configures global and profile-specific parameters
            -sysname : Configures system name used in LLDP Exchanges.
             sysname is up to 32 characters.
             Characters allowed: alphanumeric with special char underscore(_)
            -sysdesc : Configure system description used in LLDP.
```

```
                         sysdesc is up to 255 characters.
                         Characters allowed: alphanumeric with special char underscore(_)
                        -mx : Configures multiplier values for the lldp protocol.
                         mx range is <2-10>
                         One option: [-profile <profile_name>]
                         -profile: configures mx values on lldp profile
                        -txintvl : Configures TX interval values for the lldp protocol.
                          txintvl range is <4-180> seconds
                          One option: [-profile <profile_name>]
                          -profile: configures txintvl values on lldp profile
        --enable : Enable LLDP protocol across the switch
                    -port <port_num | port-range>
                     Enables lldp protocol on the port
                    -tlv <tlv name> [-profile <profile_name>]
                     Enables the specified TLV on the profile, if a profile is provided,
                     else enable it on the global profile
                    -port <port_num | port-range> [-profile <profile_name>]
                     Enables lldp on the port and sets profile to port if a profile
                     is specified
        --disable : Disable the LLDP protocol across the switch.
                     The suboptions are the same as --enable
        --clear   : Clears the specified options
                     -nbr [<port_num>]
                      Clear the neighbors info for all ports if no ports are specified,
                      else clear the neighbors info for the specified ports.
                     -stats [<port_num>]
                      Clear the LLDP stats info for all ports if no ports are specified,
                      else clear the LLDP stats for the specified ports.
        --show    : Show the global Configuration
                     -nbr [<port_num | port-range>] [-detail]
                      Show the neighbors info for all ports if no ports are provided,
                      else show neighbor info for the specified ports.
                      Shows details if the detail option is provided
                     -stats [<port_num | port-range>]
                      Show the LLDP stats info of all ports if no ports are provided, else
                      show the LLDP stats info for the given ports
                     -profile [<profile-name>]
                      Show the details of a LLDP profile
                    -port <port_num | port-range>
                     Show the details of a LLDP port
        --default : Sets the configuration to its default
```

7.3.6 IP interfaces (IPIF)

When configuring a circuit, the local and remote IP addresses must be specified, and optionally, the local HA IP and remote HA IP. A local IP or local HA IP cannot be assigned to a circuit until configured as an IPIF; otherwise, the error message "*Object Does Not Exist*" appears when configuring from the CLI.

The IPIF must be configured on the GE interface that will be used for the circuit. The DP specified must own the VE_Port used for the tunnel. The IP address and subnet mask is the IP address of the circuit's endpoint. The VLAN ID and MTU are optional arguments.

To configure an IPIF, use the command (example):

```
portcfg ipif ge0.dp0 create 192.168.10.10/24
portcfg ipif ge16.dp1 create 10.10.10.10/29
```

The IPIF consists of an IP address and a netmask length, optionally, an IP MTU size and VLAN ID.

To modify an existing IP address, MTU, or VLAN ID, use the command:

portcfg ipif <ge#.dp#> modify <ipaddr/subnet_length> [mtu <size>] [vlan <ID#>]

The maximum number of IPIFs supported on the IBM SAN42B-R7 is 60 per DP and 64 per GE interface.

Example 7-6 shows the **portcfg ipif** command.

Example 7-6   portcfg ipif command

```
SW37_7850_A:FID128:admin> portcfg ipif
Usage:  portCfg ipif [<slot>/]<port> { create { <ipaddr>/<pfx> | <ipaddr> netmask
<mask> } [mtu <mtu_size>] [vlan <vlan_id>] | modify { <ipaddr>/<pfx> | <ipaddr>
netmask <mask> } [mtu <mtu_size>] [vlan <vlan_id>] | delete <ipaddr> | --help }
Port Format:
   ge#.dp# (WAN side IPIF)
   lan.dp# (LAN side IPIF)
Options:
   create         - Create a new IP interface.
   modify         - Modify an existing IP interface.
   delete         - Delete an existing IP interface.
   help           - Show this usage message
Create / Modify Args:
   <ipaddr>/<pfx> [mtu <mtu>] [vlan <vlan_id>]
      or
   <ipaddr> netmask <mask> [mtu <mtu_size>] [vlan <vlan_id>]
   ipaddr         - IP Address to use for operation.
   Pfx or netmask - Prefix length or netmask (create only).
   mtu            - MTU size.
   vlan           - Specify the VLAN-ID.


Examples:

  portcfg ipif ge2.dp0 create 10.1.42.10/24 vlan 100
  portcfg ipif ge2.dp0 modify 10.1.42.10/24 vlan 101
  portcfg ipif lan.dp0 create 10.1.42.10/24 vlan 100 mtu 4000
```

IPIF considerations

► The following are the IPIF considerations:

► A tunnel can have a mixture of IPv4 and IPv6 circuits

► Each circuit must be either IPv4 or IPv6 at both ends; its endpoints cannot have a mixture of IPv4 and IPv6

► CIDR notation is supported for IPv4 and IPv6 addresses

► CIDR subnet mask of 31 is supported

► An IPv4 address with a 31-bit subnet mask has no network and broadcast addresses

- ► 31-bit subnet mask example, there is a one-bit difference in the following two addresses. They are on the same subnet.
- ► Address A: 192.0.2.0/31
- ► Address B: 192.0.2.1/31
- ► Specifying an IP MTU is optional; if not specified, the default is 1500 bytes
- ► The minimum supported MTU is 1280 bytes
- ► The maximum supported MTU is 9216 bytes
- ► The MTU can be set manually, or it can be set to AUTO
- ► Auto uses Path MTU discovery (PMTU) to determine the MTU

Using 31-bit subnet mask IP addresses, as defined in RFC 3021, reduces the number of IP addresses consumed by networking devices for point-to-point connectivity. Before RFC 3021, the longest subnet mask for point-to-point links was 30 bits, which meant that the all-zeros (the subnet itself) and all-ones (the broadcast) IP addresses were wasted, which was 50% of the IP addresses.

IPv6 addressing

The WAN side implementation of IPv6 uses unicast addresses for circuit IPIF. The link-local unicast address is automatically configured on the interface; however, using the link-local address space for a circuit endpoint is not permitted. Site-local unicast addresses are not allowed for circuit endpoints.

> **Note:** IPv6 addresses can exist with IPv4 addresses on the same interface, but each circuit must be configured as IPv6-to-IPv6 or IPv4-to-IPv4. Mixed IPv6-to-IPv4 circuits are not supported.

IPv6 considerations

The following are IPv6 considerations:

- ► IPv6 IPsec is supported.
- ► IPv6 PMTU (Path MTU) discovery is supported.
- ► IPv6 interfaces have unique unicast addresses.
- ► Users must statically configure IPv6 addresses and routes.
- ► IPv6 interfaces cannot be configured with anycast addresses.
- ► IPv6 interfaces cannot be configured with multicast addresses.
- ► The IPv6 8-bit Traffic Class field is defined by the configured Differentiated Services field for IPv6 (RFC 2474).
- ► The DSCP (Differentiated Services Code Point) marking configuration is done per circuit using the portcfg fcipcircuit DSCP arguments.
- ► IPv6 optional Extension Headers are not supported.
- ► IPv6 router advertisements and stateless address auto-configuration (RFC 2462) are not supported.
- ► IPv6 Neighbor Discovery (ND) and RFC 4443 ICMPv6 message types are supported.
- ► IBM b-type Extension platforms use portions of the Neighbor Discovery protocol (RFC 4861).
- ► Hop limits, such as Time to Live (TTL), are learned from Neighbor Advertisements.

► Neighbor's link-local addresses are learned from Neighbor Advertisements.

► Each GE interface's IPv6 link-local address is automatically set at startup and advertised to neighbors.

► The user does not configure an interface's link-local address.

IPv6 with embedded IPv4 addresses

When using IPv6 within an IPv4 network, only IPv4-compatible IPv6 addresses are supported. Only the low-order 32 bits of the address can be used as an IPv4 address (the high-order 96 bits must be all zeros). This allows IPv6 addresses to be used on an IPv4 routing infrastructure that supports IPv6 tunneling over the network. Both endpoints of the circuit must be configured with IPv4-compatible IPv6 addresses. IPv4-to-IPv6 connections are not supported. IPv4-mapped IPv6 addresses are not supported because they are intended for nodes that support IPv4 only when mapped to an IPv6 node.

IPIF VLAN tagging (802.1Q)

When a VLAN ID (802.1Q) is configured on an IPIF, all IPIF traffic is tagged with that VLAN ID (802.1Q). If no VLAN ID is configured, circuit traffic is not tagged. The peer GE interface in the data center must be configured the same; it is either tagging (a trunk port) or not tagging (an access port). Access ports are the most common method of connecting GE interfaces to data center switches.

IPIF MTU and PMTU

IBM b-type extension supports jumbo frames, which is the IP datagram MTU size. The smallest supported MTU is 1280 bytes, and the largest is 9216.

IBM b-type extension supports PMTU (path maximum transmission unit). PMTU is sending a set of various known-size ICMP (Internet Control Message Protocol) datagrams across the IP network to determine the maximum size that was successfully received. The ICMP datagrams are marked Do Not Fragment. Based on the largest ICMP Echo Reply received, the PMTU discovery process sets the IP MTU for the circuit's IP interface (ipif).

PMTU does not discover an MTU greater than 9100 bytes. If the MTU is larger than 9100 bytes, the best practice is to configure the MTU manually. Suppose the MTU of the IP network is known. In that case, the best practice is to set it manually, avoiding values that may be less optimal and eliminating the additional time required to bring up a circuit.

If a circuit bounces, the PMTU discovery process is initiated when reestablishing the circuit. Each circuit initiates the PMTU discovery process before coming online. This is required because circuits may have gone offline due to an IP network failure and rerouted to a new path with different network characteristics. The PMTU discovery process requires additional time when bringing a circuit online.

If the PMTU discovery cannot communicate with the peer switch, the circuit will use the smallest supported MTU (1280 bytes). PMTU requires that ICMP Echo Requests and Replies are permitted end-to-end across the WAN IP network. As a rudimentary check, ping the remote IBM extension WAN side IPIF.

7.3.7 IP routes

When configuring WAN side IP routes on the IBM SAN42B-R7, up to 128 routes per GE interface and a maximum of 120 routes per DP can be created.

The following example in Figure 7-1 shows an IP route configuration for a layer 3 routed network. A layer 3 routed network means the local and remote subnets differ. Notice that the

IPIF IP address and router gateway are in the same subnet. IPIFs communicate to the WAN through a router gateway.

The GE interface and DP are specified when creating an IP route. Circuits are configured on a VE_Port owned by the DP and with an IP address configured on an IPIF, which puts the circuit on a specific GE interface. DPs own a set of VE_Ports. The DP specified for the IPIF, the IP route, and the VE_Port must align. For example, specifying the command `portcfg ipif` and `portcfg iproute` with `ge2.dp0` puts the IPIF and iproute on ge2 and DP0.

IP routes do not span across GE interfaces or DPs. The same iproute must be added to all GE interface/DP combinations when a subnet is used for multiple circuits, including eHCL.

For example, on the IBM SAN42B-R7, VE24 is owned by DP0; therefore, creating an IPIF must designate the desired GE interface and DP0.

In the example, the following is used:

▶ Side A:
  – Circuit Subnet 10.0.0.0/24 (255.255.255.0)
  – Circuit IPIF 10.0.0.6
  – Circuit eHCL IPIF 10.0.0.7
  – Local Gateway 10.0.0.1

▶ Side B:
  – Circuit Subnet 192.168.0.0/24 (255.255.255.0)
  – Circuit IPIF 192.168.0.6
  – Circuit eHCL IPIF 192.168.0.7
  – Local Gateway 192.168.0.1

Example `IPIF` command:

```
Side A: portcfg ipif ge2.dp0 create 10.0.0.6/24
Side B: portcfg ipif ge2.dp0 create 192.168.0.6/24
```

Example `eHCL IPIF` command:

```
Side A: portcfg ipif ge2.dp1 create 10.0.0.7/24
Side B: portcfg ipif ge2.dp1 create 192.168.0.7/24
```

The IProute must be assigned to the same IPIF (ge#.dp#), for example, the command:

```
Side A: portcfg iproute ge2.dp0 create 192.168.0.0/24 10.0.0.1
Side B: portcfg iproute ge2.dp0 create 10.0.0.0/24 192.168.0.1
```

eHCL is the exception; the IPIF and IP route are configured on the opposite DP, for example, the commands:

Example `eHCL IPIF` command:

```
Side A: portcfg ipif ge2.dp1 create 10.0.0.7/24
Side B: portcfg ipif ge2.dp1 create 192.168.0.7/24
```

Example `eHCL iproute` command:

```
Side A: portcfg iproute ge2.dp1 create 192.168.0.0/24 10.0.0.1
Side B: portcfg iproute ge2.dp1 create 10.0.0.0/24 192.168.0.1
```
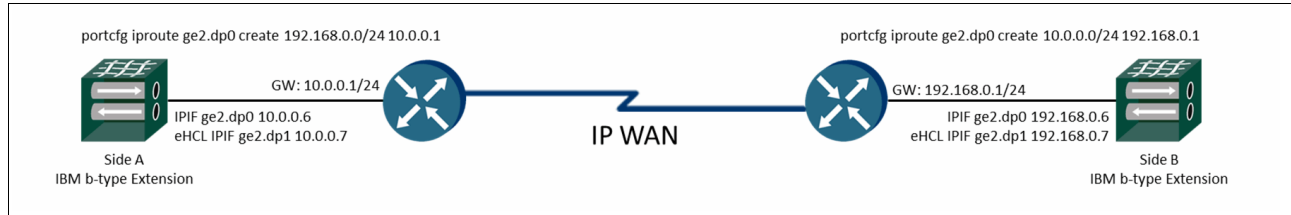
Figure 7-1   Configuring IP routes

Example 7-7 shows the **portcfg iproute** command syntax

Example 7-7   Portcfg iproute command syntax

Portcfg iproute command syntax:

```
SW37_7850_A:FID128:admin> portcfg iproute
Usage:  portCfg iproute [<slot>/]<port> { create { <ipaddr>/<pfx> | <ipaddr>
netmask <mask> } <gateway> | modify { <ipaddr>/<pfx> | <ipaddr> netmask <mask> }
<gateway> | delete { <ipaddr>/<pfx> | <ipaddr> netmask <mask> | --help }
Port Format:
    ge#.dp# (WAN side IPIF)
    lan.dp# (LAN side IPIF)
Options:
    create         - Create a new IP route.
    modify         - Modify an existing IP route.
    delete         - Delete an existing IP route.
    help           - Show this usage message
Create / Modify Args:
    <ipaddr>/<pfx> <gateway>
       or
    <ipaddr> netmask <mask> <gateway>
    ipaddr         - IP Address to use for operation.
    pfx/netmask    - Prefix length or netmask.
    gateway        - Gateway IP address to use (create/modify only).
```

Examples:

Side A:

```
portcfg iproute ge2.dp0 create 192.168.0.0/24 10.0.0.1
portcfg iproute ge2.dp1 create 192.168.0.0/24 10.0.0.1
portcfg iproute ge3.dp0 delete 192.168.0.0/24
```

Side B:

```
portcfg iproute ge2.dp0 create 10.0.0.0/24 192.168.0.1
portcfg iproute ge2.dp1 create 10.0.0.0/24 192.168.0.1
```

7.3.8 Encryption (IPsec)

Before IPsec can be enabled, an IPsec policy must be defined. Multiple IPsec policies can be defined; however, only one policy can be applied to a tunnel. All circuits in the tunnel use the same IPsec policy. When you define an IPsec policy, you can select between two profile types. The PSK (Pre-Shared Key) profile lets you specify a key when configuring IPsec. The PKI (public-key infrastructure) profile uses certificates.

When using a PSK (Pre-Shared Key), both ends of the secure tunnel must be configured with the exact key string. If both ends are not configured with the same key, the IKEv2 session will not form, preventing the tunnel from coming up. The IPsec policy name can differ at each end, but the key must be identical. The pre-shared key is a string of 16 to 64 alpha-numeric characters.

IBM b-type IPsec supports a cryptographic suite that is CNSA and Suite B compliant, which requires generating and importing compliant X.509 end-entity certificates and issuing CA's certificates used to sign them. This requires each end of a secure connection to have access or a means to look up the CA certificate for verification purposes. PKI enables the configuration of the Elliptic Curve Diffie-Hellman (ECDH) for key agreement and the Elliptic Curve Digital Signature Algorithm (ECDSA) for peer authentication.

PKI considerations

The following are PKI considerations:

- ► CNSA and Suite B compliant.
- ► X.509 certificates are supported.
- ► X.509 Certificate Revocation Lists (CRL) are not supported.
- ► ECDSA certificates are supported on the IBM SAN42B-R7 platforms.
- ► Non-ECDSA certificates are not supported.
- ► PKI support is restricted to key-size P384 and hash-type SHA384.

> **Note:** The PKI profile uses certificates. The PKI profile complies with the Commercial National Security Algorithm (CNSA) and Suite B. For information on configuring and managing certificates using the `secCertMgmt` command, refer to "SSL Configuration Overview" in the *Brocade Fabric OS Administration Guide* at https://techdocs.broadcom.com/fabric-os-extension

> **Note:** Both ends of the extension tunnel must run the same Brocade Fabric OS version when running IPsec.

An IPsec policy can be modified while assigned to a tunnel or WAN Tool session. Sometimes, the local and remote sides get out of sync preventing the tunnel from coming up, and an authentication error is indicated. See IPsec IKEv2 Authentication Failures for information on how to restart IKEv2 authentication.

IPsec considerations

The following applies to IPsec:

- ► IPsec encryption is at the tunnel level, not per circuit.
- ► IPsec supports IPv6 and IPv4-based tunnels
- ► There are IPsec-specific statistics
- ► Network Address Translation (NAT) is not supported
- ► Authentication Header (AH) is not supported
- ► The `--connection-type` tunnel option must be the default (type is not specified)
- ► A single predefined mode of operation is used for protecting TCP traffic over a tunnel:
  - AES-GCM-ESP: AES (Advanced Encryption Standard described in RFC 4106)
  - GCM (Galois Counter Mode)

– ESP (Encapsulating Security Payload)

IPsec uses the following sequence of events:

1. IPsec and IKEv2 (Internet Key Exchange version 2) policies are created on both tunnel ends.

2. IKEv2 exchanges policy information on each end of the connection. The connection will not come online if the policy information does not match. Some exchanged security parameters include the encryption and authentication algorithms, Diffie-Hellman key exchange, and the SAs.

3. Data is transferred between IPsec peers based on the parameters and keys stored in the SA database.

4. When authentication and IKEv2 negotiation are complete, the IPsec SA is ready for data transmission.

5. SA lifetimes terminate through deletion or expiration. SA lifetime equates to approximately:
   – 2 billion frames of data traffic passed through the SA
   – 4-hour expiration time

6. When the SA is about to expire, an IKEv2 rekey occurs to create a new SA on both ends, after which data starts using the new SA. IKEv2 and SA rekeying is non-disruptive.

Table 7-6 shows the algorithm selection for a Suite B-compliant configuration supported in Brocade Fabric OS.

Table 7-6   IPsec CNSA and Suite B Attributes

Attribute	CNSA and Suite B in Fabric OS 9.2.0
Authentication	ECDSA-P384
Diffie-Hellman	ECDH-P384
Integrity	HMAC-384-192
Encryption	AES-256-CBC (IKE), AES-256-GCM (data)
Pseudo-Random Function	PRF-HMAC-384

Example 7-8 shows the IPsec CLI syntax.

Example 7-8   IPsec CLI syntax:

```
SW37_7850_A:FID128:admin> portcfg ipsec-policy
Usage:  portCfg ipsec-policy <name> { create [<args>] | modify [<args>] | delete
| restart | --help }
Name Format:
    <string> - The IPSec Policy name(Min 1 character, Max 31 characters).
             Cannot contain the following special characters:
             ;$!#`/\><&'"=,?.*^{}()
Option:   create  - Create the specified IPsec Policy
          modify  - Modify the specified IPsec Policy
          delete  - Delete the specified IPsec Policy
          restart - Restart all inactive IKE sessions for this policy
          help    - Show this usage message
Optional Arguments:
  -p,--profile { preshared | pki } -
                                - Set the IPsec-Profile.
  -k,--preshared-key <16-64> (FIPS Mode: <32-64>) -
```

```
                                    -  String value for preshared key (for
                                       authentication method "SHARED_KEY").
     -K,--keypair <keypair name>   -  Name of the keypair. Max 31 Chars (for
                                       authentication method "ECDSA_P384").
     -h,--help                     -  Show the IPsec-Policy configuration usage
                                       statement.
```

Example of creating an IPsec policy used in a tunnel:

`portcfg ipsec-policy myPolicy create --preshared-key <your_secret_PSK>`

Example of using an IPsec policy in tunnel 24:

`portcfg fciptunnel 24 modify --ipsec myPolicy`

CLI syntax for **`seccertmgmt`** command:

`SW38_7850_B:FID128:admin> seccertmgmt`

Generate Extension CSR:

```
seccertmgmt generate -csr extn -keypair_tag <keypair_name>
[-type ecdsa] [-keysize P384] [-hash sha384] [-years <x>] [-f]
```

Generate Extension Self-signed Cert:

```
seccertmgmt generate -cert extn -keypair_tag <keypair_name>
[-type ecdsa] [-keysize P384] [-hash sha384] [-years <x>] [-f]
```

Import CA and Extension Signed Cert:

```
seccertmgmt import -cert extn -certname <certificate name>
[-keypair_tag <keypair_name>] [-protocol {scp|ftp}]
[-ipaddr <IP address>][-remotedir <remote directory>]
[-cacert <preimported_local_ca_cert>] [-login <login name>]
[-password <password>]
seccertmgmt import -ca {-client|-server} extn -certname <certificate name>
[-protocol {scp|ftp}] [-ipaddr <IP address>]
[-remotedir <remote directory>] [-login <login name>]
[-password <password>]
SW37_7850_A:FID128:admin> seccertmgmt show -cert extn
List of extn files:
Certificate Files:
--------------------------------------------------------------------------------
Protocol   CA                 SW             CSR      PVT Key Passphrase Keypair Tag
--------------------------------------------------------------------------------
List of local CERT files
EXTN    NA              IBMEXT.pem   Exist   Exist   NA        IBMEXT
List of remote CERT files
```

7.3.9 Circuits

Extension circuits are individual connections within a tunnel, and data flows between the source and destination VE_Ports. Each tunnel can contain a single circuit or multiple circuits. A circuit connects a pair of IPIF (IP interfaces). Each IPIF must have a unique IP address. Each circuit has a unique pair of source and destination IP addresses. A circuit's local and remote IP addresses can be from the same subnet, which is referred to as Layer 2 (L2), or

from different subnets, referred to as Layer 3 (L3). L2 and L3 architectures are supported on the WAN side with FCIP and IP Extension.

After creating a tunnel, create its circuits. Start with circuit-0 and add each circuit incrementing the circuit number by 1. Creating a circuit for each unique path through the IP WAN infrastructure is best. Circuits require identical settings at each end.

> **Note:** The WAN side subnets have no restrictions. They can be layer 2 or 3, as described above. The WAN side subnets are flexible to the environment of the organization. The subnets do not have to be different (L2) but can be (L3). On the LAN side, the local and remote subnets must be different.

When the remote IPIF is not on the same subnet as the local IPIF, an IP route to the remote subnet through the local gateway must be configured for the interface DP pair.

Local and remote, tunnel, and circuit settings must match to obtain a UP state. For example, if you configure IPsec on a tunnel, each end of the tunnel must be configured to use the same IPsec parameters, such as the PSK (pre-shared key). Other circuit parameters must be consistent, such as MTU, min/max bandwidth, QoS distribution and priority percentages, and keepalive timeout (KATOV) values.

All circuits are point-to-point. To form a circuit, there is a source and destination IP address; on the opposite side, the same addresses are merely swapped.

Circuit considerations

The following requirements and limitations apply to extension circuits:

► Each circuit has a unique pair of source and destination IP addresses.

► An IP interface (IPIF) defines an IP address associated with a circuit endpoint and GE interface.

► The maximum rate of a circuit cannot exceed the GE interface speed.

► On a GE interface, defining multiple IPIFs allows multiple circuits on that interface.

► A VE_Port with multiple circuits is a Brocade Extension Trunk (BET).

► If a circuit's source and destination IP addresses are not on the same subnet, an IP route must be configured on both ends.

► Metric-0 circuits are active under normal operating conditions within the failover group.

► Metric-1 circuits are active when all metric-0 circuits within the failover group have gone offline.

► A failover group control in which metric-1 circuits become active when metric-0 circuits in the group go offline.

► When a spillover is configured, failover groups are not used.

► When a spillover is configured, metric-1 circuits are used when the capacity of the metric-0 circuit is exceeded.

► 4:1 Rule: Within a tunnel, consider the ARL minimum rates; the difference between the slowest min rate and the fastest min rate can be no greater than 4x.

  – For example, the minimum rates of Cir0 = 1 Gbps and Cir1 = 4 Gbps is supported; however, Cir0 = 1 Gbps and Cir1 = 5 Gbps is not.

  – If circuits are configured with rates greater than 4x apart, the bandwidth may not be fully utilized.

► Minimum supported bandwidth

– IBM SAN18B-6 is 20 Mbps

– IBM SAN42B-R7 is 50 Mbps

– IBM SX6 is 20 Mbps

► Circuit settings required to match on both ends cause circuits not to come up if there is a mismatch.

► Circuit settings that should be identical on both ends, such as minimum and maximum bandwidth values and KATOV, will indicate an error condition.

Example 7-9 shows the `portcfg fcipcircuit` command.

Example 7-9   Portcfg fcipcircuit command

```
SW37_7850_A:FID128:admin> portcfg fcipcircuit
Usage:   portCfg fcipcircuit [<slot>/]<port> { create [<args>] | modify [<args>] |
delete | --help } <cid>
Option:    create - Create the specified tunnel/circuit
           modify - Modify the specified tunnel/circuit
           delete - Delete the specified tunnel/circuit
           help   - Show this usage message
Optional Arguments:
  -a,--admin-status { enable | disable } -
                                  - Set the admin-status of the circuit.
  -S,--local-ip { <ipaddr> | none } -
                                  - Set local IP address.
  -D,--remote-ip { <ipaddr> | none } -
                                  - Set remote IP address.
     --local-ha-ip { <ipaddr> | none } -
                                  - Set local HA IP address. This allows for HCL
                                    operations on local switch. [Not available on
                                    7810]
     --remote-ha-ip { <ipaddr> | none } -
                                  - Set remote HA IP address. This allows for HCL
                                    operations on remote switch.
  -x,--metric { 0 | 1 }         - Set the metric. 0=Primary 1=Failover.
  -g,--failover-group <0-9>     - Set the failover group ID.
  -b,--min-comm-rate { <kbps> | <mbps>M | <gbps>G } -
                                  - Set the minimum committed rate in
                                    Kbps|Mbps|Gbps.
  -B,--max-comm-rate { <kbps> | <mbps>M | <gbps>G } -
                                  - Set the maximum committed rate in
                                    Kbps|Mbps|Gbps.
     --arl-algorithm <mode>     - Set the ARL algorithm. Allowable modes are {
                                    auto | reset | step-down | timed-step-down }.
  -C,--connection-type <type>   - Set the connection type. Allowable types are {
                                    default | listener | initiator }.
  -k,--keepalive-timeout <ms>   - Set keepalive timeout in ms.
     --dscp-f-class <dscp>      - Set DSCP marking for Control traffic.
     --dscp-high <dscp>         - Set DSCP marking for FC-High priority traffic.
     --dscp-medium <dscp>       - Set DSCP marking for FC-Medium priority
                                    traffic.
     --dscp-low <dscp>          - Set DSCP marking for FC-Low priority traffic.
     --dscp-ip-high <dscp>      - Set DSCP marking for IP-High priority traffic.
     --dscp-ip-medium <dscp>    - Set DSCP marking for IP-Medium priority
```

```
                                                traffic.
          --dscp-ip-low <dscp>        -  Set DSCP marking for IP-Low priority traffic.
          --l2cos-f-class <l2cos>     -  Set L2CoS value for Control priority traffic.
          --l2cos-high <l2cos>        -  Set L2CoS value for FC-High priority traffic.
          --l2cos-medium <l2cos>      -  Set L2CoS value for FC-Medium priority
                                         traffic.
          --l2cos-low <l2cos>         -  Set L2CoS value for FC-Low priority traffic.
          --l2cos-ip-high <l2cos>     -  Set L2CoS value for IP-High priority traffic.
          --l2cos-ip-medium <l2cos> -  Set L2CoS value for IP-Medium priority
                                         traffic.
          --l2cos-ip-low <l2cos>      -  Set L2CoS value for IP-Low priority traffic.
          --sla { <sla-name> | none } -
                                      -  Set the SLA name for this circuit.
          --help                      -  Print this usage statement.
```

Example:

```
portcfg fcipcircuit 24 create 0 --min-comm-rate 1G
```

7.3.10 Compression

Set the desired compression algorithm per protocol if the tunnel has IP Extension enabled.
The IBM SAN42B-R7 allows the configuration compression at a protocol level per FCIP or IP
Extension. The per-protocol compression configuration overrides the general compression
setting for a tunnel. See Table 7-7.

Table 7-7   Extension protocol compression choice

Compression Algorithm	FCIP	IP Extension
Fast Deflate	Yes (SAN42B-R7 and SX6 blade)	No
Deflate	Yes	Yes
Aggressive Deflate	Yes	Yes

Compression is set at the tunnel level, meaning compression is not set per circuit. Setting
compression is disruptive. Compression can be set for the entire tunnel (regardless of
protocol) or per protocol (FCIP and IP Extension). To set compression per protocol, the tunnel
must have IP Extension enabled. The IBM SX6 must be configured in hybrid mode to enable
IP Extension on tunnels.

Compression considerations
The following considerations apply to ompression:

► The overall algorithm throughput depends on the algorithm and platform.

► The compression ratio depends on the compressibility of the data and the algorithm.

► The available compression algorithms depend on the protocol (FCIP and IP Extension).

► Enabling IP Extension on a tunnel enables compression algorithm selection at the protocol
  level.

► Protocol-level compression overrides the general compression setting.

► The available compression algorithms depend on the platform

– IBM SAN18B-6 does not have fast-deflate

Use the following guidelines in the tables Table 7-8 through Table 7-10 on page 87 to assign a compression algorithm for a tunnel:

Table 7-8   Assigning compression on the IBM SAN18B-6

Total WAN Side Bandwidth	Compression Algorithm	Protocol
1.5 Gbps and higher	Deflate	FCIP and IP Extension
1.5 Gbps and lower	Aggressive Deflate	FCIP and IP Extension

Table 7-9   Assigning compression on the IBM SAN42B-R7

Total WAN Side Bandwidth	Compression Algorithm	Protocol
10 Gbps and higher	Fast Deflate	FCIP only
5 Gbps to 10 Gbps	Deflate	FCIP and IP Extension
5 Gbps and lower	Aggressive Deflate	FCIP and IP Extension

Table 7-10   Assigning compression on the IBM SX6 Extension Blade

Total WAN Side Bandwidth	Compression Algorithm	Protocol
4 Gbps and higher	Fast Deflate	FCIP only
2 Gbps to 4 Gbps	Deflate	FCIP and IP Extension
2 Gbps and lower	Aggressive Deflate	FCIP and IP Extension

Note: Throughput for all compression modes depends on the compression ratio achievable for the data. IBM and Broadcom make no promises, guarantees, or assumptions about the compression ratio that user data may achieve.

Note: Each end of a tunnel must have the same compression setting.

Note: Fast-Deflate compression mode is not supported for IP Extension on any platform. Deflate and Aggressive Deflate function for both FC and IP Extension protocols.

Note: Compression is rate-limiting if the selected algorithm has insufficient throughput to accommodate the flows. Fast-Deflate is processed on a different engine than Deflate and Aggr-Deflate in each DP; Deflate and Aggr-Deflate are processed on the same engine.

See Example 7-10 for setting the compression mode when creating or modifying a tunnel.

Example 7-10   Setting the compression mode when creating or modifying a tunnel

```
SW37_7850_A:FID128:admin> portcfg fciptunnel
Usage:  portCfg fciptunnel [<slot>/]<port> { create [<args>] | modify [<args>] |
delete | --help }
```

```
Option:    create - Create the specified tunnel/circuit
           modify - Modify the specified tunnel/circuit
           delete - Delete the specified tunnel/circuit
           help   - Show this usage message
Optional Arguments:
  -c,--compression <mode>       - Set the compression mode. Allowable modes are
                                  { none | deflate | aggr-deflate | fast-deflate
                                  (Not available on 7810) }.
     --fc-compression <mode>    - Set the compression mode. Allowable modes are
                                  { none | deflate | aggr-deflate | fast-deflate
                                  | default }.
     --ip-compression <mode>    - Set the compression mode. Allowable modes are
                                  { none | deflate | aggr-deflate | default }.
```

7.3.11 ARL and CIR

ARL is configured on a per-circuit basis, not per tunnel, because each circuit can take a unique path and have a different bandwidth. The ARL minimum and maximum bandwidths are configured when a circuit is created or modified. Circuit bandwidths might be limited based on the licensing or capacity of the platform. Circuits within a tunnel can have the same or different amounts of bandwidth.

Table 7-11 describes the maximum bandwidth that a single circuit can have on each platform:

Table 7-11   Maximum circuit bandwidth

Extension Platform	Max Circuit Bandwidth (base unit)	Max Circuit Bandwidth (upgraded/full unit)
IBM SAN18B-6	1 Gbps	2.5 Gbps (GE interface permitting)
IBM SAN42B-R7	25 Gbps (GE interface permitting)	
IBM SX6 Blade	10 Gbps (GE interface permitting)	

ARL considerations

The following considerations apply to ARL:

► The aggregate of circuit minimum-rate bandwidth settings should not exceed the available WAN bandwidth. For example, given a dedicated 10 Gbps WAN, the aggregate of the ARL minimum rates should be no more than 10 Gbps. There is no limitation for FC ingress rates because flow control (BBC) limits incoming FC data.

► The aggregate of the minimum configured bandwidth values on a GE interface cannot exceed the interface's speed.

► Configure minimum rates from all tunnels on a DP so that the aggregate does not exceed the DP's capacity.

► 5:1 Min/Max Committed Rate Rule: The ratio between a circuit's minimum and maximum rates cannot exceed five times. For example, if the minimum is set to 2 Gbps, the maximum for that circuit cannot exceed 10 Gbps. This limit is enforced.

► 4:1 Lowest to Highest Circuit Rule: The ratio between any two circuits in the same tunnel, the highest bandwidth, should not exceed four times the lowest bandwidth. For example, if one circuit is configured with a minimum of 2 Gbps, another circuit cannot be configured

with a maximum exceeding 8 Gbps. This rule is not enforced in software but is recommended for proper operation.

► For any circuit, the minimum and maximum bandwidth values must match on the local and remote ends.

ARL is enabled only when a circuit's minimum < maximum bandwidth value. If you configure the minimum equal to the maximum, for example, 1G bps, ARL is not used. Such a configuration is referred to as CIR (Committed Information Rate).

ARL FSPF link cost calculations

Fabric Shortest Path First (FSPF) is a link-state path selection protocol that directs the traffic along the shortest path between the source and the destination. FSPF calculations are based on the link cost; every link has a cost. The link cost is calculated by summing the maximum rates of all active circuits in the tunnel. The calculation is not per circuit but per tunnel because each tunnel is effectively an ISL.

The following formula is used:

► If the bandwidth is or exceeds 2 Gbps, the link cost is 500.

► If the bandwidth is less than 2 Gbps but greater than or equal to 1 Gbps, the link cost is 1,000,000 divided by the bandwidth in Mbps.

► If the bandwidth is less than 1 Gbps, the link cost is 2000 minus the bandwidth in Mbps.

If there are multiple parallel tunnels between the same domains, set identical static link costs for the VE ports and configure lossless DLS, which helps to minimize FC frame loss during bandwidth updates caused by circuit bounces.

Note: Multiple tunnels between the same domains are supported but not recommended.

Committed Information Rate (CIR)

ARL can be disabled, and a single committed information rate (CIR) can be used. By setting min=max, ARL is disabled, and the value used for min and max is the CIR. Most commonly, CIR is used when line rate is needed. For example, there is a 100 Gbps WAN, no contention, and two 25 Gbps circuits are being used; ARL is unnecessary. In this case, each circuit would use min = max = 25 Gbps.

Note: ARL min and max values are not optional. Circuits will not come up until the min and max values have been configured. The values must be the same at both ends.

ARL Backoff Algorithm

Although ARL provides shared WAN bandwidth, the amount of replicated storage data for extension connections continues to grow, which needs larger and faster links. On all supported extension platforms, the enhanced response time of ARL provides faster rate-limiting adaptation, which permits optimized throughput of the extension traffic and competing flows.

The back-off mechanism implemented by ARL is optimized to increase the overall throughput. ARL dynamically preserves bandwidth and evaluates network conditions to see whether additional back-offs are required. ARL maintains the round-trip-time (RTT) state information to better predict network conditions and to allow intelligent and granular decisions about proper adaptive rate limiting. ARL examines the prior stateful information when it encounters a network error. Rate-limit decisions are made using the ARL algorithm. When configured for

automatic, ARL dynamically determines which algorithm to use based on changing network conditions.

The ARL algorithm for backing off traffic can be configured on all extension platforms. The default is automatic, and the ARL logic determines the best approach.

The ARL algorithm choices are as follows:

► Automatic (default)

► Static reset

► Modified multiplicative decrease (MMD) or step-down

► Time-based decrease or timed step-down

Note: Only change the ARL backoff algorithm if instructed by your support organization.

See Example 7-11 for ARL and CIR Comm-Rate CLI configuration.

Example 7-11   ARL and CIR Comm-Rate CLI configuration

```
SW37_7850_A:FID128:admin> portcfg fcipcircuit
Usage:   portCfg fcipcircuit [<slot>/]<port> { create [<args>] | modify [<args>] |
delete | --help } <cid>
Option:    create - Create the specified tunnel/circuit
           modify - Modify the specified tunnel/circuit
           delete - Delete the specified tunnel/circuit
           help   - Show this usage message
Optional Arguments:
-b,--min-comm-rate { <kbps> | <mbps>M | <gbps>G } -
                                  - Set the minimum committed rate in
                                    Kbps|Mbps|Gbps.
  -B,--max-comm-rate { <kbps> | <mbps>M | <gbps>G } -
                                  - Set the maximum committed rate in
                                    Kbps|Mbps|Gbps.
    --arl-algorithm <mode>    - Set the ARL algorithm. Allowable modes are {
                                    auto | reset | step-down | timed-step-down }.
Example:
  portcfg fcipcircuit 24 create 0 --min-comm-rate 100000
```

7.3.12 Extension Hot Code Load (eHCL)

To enable high availability (HA) for circuits, in addition to the customarily configured IPIFs and IP routes, you must also configure the HA IPIFs and IP routes on the opposing DP.

We assume configuring a tunnel between a local and remote DP0. The main tunnel (MT) uses IPIF IP addresses on the local DP0 going to the remote DP0. The LBT uses IPIF IP addresses on the local DP1 to the remote DP0. The RBT uses IPIF IP addresses on the local DP0 to the remote DP1. You only configure the production IP addresses, the HA IP addresses, and the production tunnel and circuits; do not create the backup circuits (LBT and RBT), as they are created automatically. This collection of circuits and tunnels forms the HA tunnel group.

When configuring a production tunnel and circuits, you designate IP addresses specific to eHCL through the HA arguments in the **portcfg fcipcircuit** command.

```
portcfg fcipcircuit <ve#> modify <cir#> --local-ha-ip <IP>
```

```
portcfg fcipcircuit <ve#> modify <cir#> --remote-ha-ip <IP>
```

eHCL requires two IP addresses per circuit, with one HA IP address at each end. Each circuit has four IP addresses: the production local and remote IP addresses and the eHCL local and remote HA IP addresses used during firmware updates. Each HA IP address has its own IPIF. Local IP addresses are typically part of the same subnet, and remote IP addresses are part of the same subnet, with different subnets at each site. All IP addresses must be able to communicate across the IP WAN infrastructure. If the production IPIF is on DP0, then the eHCL IPIF is on DP1; conversely, if the production IPIF is on DP1, then the eHCL IPIF is on DP0. On different GE interfaces and DPs, an IP route to the remote site must be configured on an IPIF for each subnet a circuit will go to.

The IBM SAN18B-6 has only one DP and does not support eHCL by itself; however, when connected to an IBM SAN42B-R7 or an IBM SX6 Extension Blade, it does support eHCL on those platforms. Therefore, the local HA option on the IBM SAN18B-6 is not configurable.

```
portcfg fcipcircuit <ve#> modify <cir#> --local-ha-ip <IP>
```

However, the HA option on the remote side is required on the IBM SAN18B-6 to facilitate eHCL operation on the IBM SAN42B-R7 and IBM SX6 Extension Blade.

```
portcfg fcipcircuit <ve#> modify <cir#> --remote-ha-ip <IP>
```

When the IBM SAN42B-R7 and the IBM SX6 Extension Blade are configured for eHCL with an IBM SAN18B-6 at the other end, only the `--local-ha-ip` option should be used. If you use the `--remote-ha-ip` option on the IBM SAN42B-R7 or the IBM SX6 Extension Blade, the tunnel will not reach an ONLINE state and remain DEGRADED. Only the MT and RBT groups appear on the IBM SAN18B-6, and only the MT and LBT groups are found on the IBM SAN42B-R7 or IBM SX6.

The below command functions for the IBM SAN18B-6, even though it does not support eHCL. This command on the IBM SAN18B-6 provides DP connectivity information for the MT and RBT groups.

```
portshow fciptunnel --hcl-status
```

eHCL performs coordinated upgrades, meaning the firmware update can be initiated simultaneously on both tunnel ends. Fabric OS coordinates the update process to maintain online, lossless, in-order data delivery.

eHCL considerations:

The following considerations should be made when implementing eHCL:

► eHCL was designed for FICON environments such as IBM XRC, FICON tape, and TS7700 Grid.

► In-order delivery is guaranteed.

► No data is lost inflight during the update process.

► eHCL does not cause a FICON interface control check (IFCC).

► eHCL was designed for FC replication, such as IBM Global Mirror.

► eHCL implementation requires proper planning.

► eHCL has a circuit scope and must be configured per circuit; only circuits that specify a local and remote HA IP address are protected. If eHCL does not protect a circuit, it goes offline during the firmware update process.

► eHCL supports all features, such as Virtual Fabrics (VF) and FC Routing (FCR).

► eHCL supports asynchronous and synchronous FC/FICON environments and IP Extension.

► eHCL issues the RASlog warnings and error messages (WARN or ERROR).

► The production circuit (regular operating circuit) is called the MT.

► The local backup circuit to the remote switch is the LBT.

► The tunnel from the remote switch back to the local switch is the RBT, which is used when the remote side performs its firmware update.

► The IPIFs for an MT and LBT must be on opposing DPs. For example, if the MT IPIF is on DP1, then the LBT IPIF is on DP0.

► MT tunnel and circuit parameters are cloned on the LBT and RBT, including the circuit properties such as QoS markings, ARL, FastWrite, OSTP, and FICON Acceleration.

► The total switch capacity temporarily decreases by 50% as only one DP complex remains operational during eHCL.

► Redundant replication fabrics typically have A and B pathways. The other fabric provides bandwidth during a firmware update as well.

► The management of ARL during eHCL is handled on a VE_Port basis. ARL is temporarily disabled when a VE_Port begins the failover process, and no other VE_Ports on a DP are affected. When failover/failback is complete, ARL is enabled for that VE_Port.

► No configuration changes, including changing tunnel or circuit parameters, are allowed during the eHCL process. New device connections requiring zone checking can experience a timeout during the CP reboot phase of the firmware download. A CP performs zone checks and must be active to process new SID/DID connection requests, such as Port Logins (PLOGI).

► If parallel tunnels (VE_Ports) are configured between local and remote sites, and if each tunnel has multiple circuits, you must manually set a static link cost on the VE_Port; otherwise, FC traffic might be disrupted during the eHCL activity.

► When Teradata Emulation is enabled on an extension tunnel, eHCL is not supported. You must perform a disruptive firmware update.

Bandwidth planning might include dedicating one of the two DP complexes to HA or limiting each DP to a maximum of 50% of the licensed capacity to reserve adequate bandwidth for high-availability operations. Most firmware updates support eHCL; however, not every Brocade Fabric OS release guarantees a firmware update capable of eHCL. Refer to the Brocade Fabric OS Release Notes for potential caveats. The firmware at each end of the tunnel must be compatible; start with the same version of Brocade Fabric OS on each end. Regular production operation of extension with disparate Brocade Fabric OS versions is not tested and can introduce instability, aberrant behavior, prevent successful tunnel formation, or exhibit impaired feature functionality.

eHCL does not require a different communication path. The production IP WAN network used for extension tunnel communications is used for the two eHCL backup tunnels (LBT and RBT) that exist alongside.

Do not configure specific pairs of VE_Ports in different logical switches with different Brocade Fabric OS routing policies. For example, on an IBM SX6 Extension Blade, one logical switch has Exchange-Based Routing (EBR), and the other has Port-Based Routing (PBR), a typical configuration in mainframe environments. The logical switches use VE_Ports 24 (FID 1) and 33 (FID 2). During eHCL, these VEs share backend virtual channels (VC) when on the HA path. This configuration can cause unexpected results. Table 7-12 on page 93 shows the VE_Port pairs to avoid. On the IBM SX6 Extension Blade, the pairing restriction is per blade.

Table 7-12   VE_Port pairs and differing LS traffic policies

IBM SAN42B-R7 Extension and IBM SX6 Blade Conflicting VE_Port Pairs with Differing LS Traffic Policies			
IBM SAN42B-R7		IBM SX6 Blade	
DP0	DP1	DP0	DP1
24	33	16	26
25	34	17	27
26	35	18	28
27	36	19	29
28	37	20	30
29	38	21	31
30	39	22	32
31	40	23	33
32	41	24	34
		25	35

7.3.13 Circuit QoS

Based on its FCIP or IP Extension priority (high, medium, and low), traffic can be marked with DSCP or 802.1P (L2CoS). The marked value is user configurable; there is no default. Each protocol (FCIP and IP Extension) is marked independently.

QoS marking is configured per circuit, not per tunnel. Each circuit takes its own IP network route, which could have its own QoS paradigm. The default is no marking. Marking is not enforcement. It is up to the IP network to perform QoS enforcement based on the markings. Therefore, working with the network administrators to establish enforcement is essential.

The following two options are available for QoS marking:

► DSCP (IP marking, layer 3)

► L2CoS (802.1P Ethernet marking in 802.1Q VLAN tag header, layer 2)

Protocol bandwidth distribution

In the first tier of extension QoS, each protocol (FCIP and IP Extension) receives a proportion of the overall configured bandwidth, referred to as protocol distribution. Distribution allows apportioning bandwidth between FCIP and IP Extension.

FCIP and IP Extension protocol distribution are rate-limiting to bandwidth. If a protocol is utilized, its total bandwidth percentage is reserved and unavailable to the other protocol. Distribution occurs at the tunnel level, not per circuit. For example, you can specify a QoS distribution ratio of FCIP 60% and IP Extension 40%.

QoS protocol bandwidth distribution considerations:

► The default distribution is 50/50%.

► The distribution (FCIP:IP Extension) must total 100%.

► 10% is the minimum value.

- ► 90% is the maximum value.

- ► Distribution ratios must be the same on both ends of a tunnel.

- ► Protocol distribution between FCIP and IP Extension is rate-limiting. If a protocol (FCIP or IP Extension) is used, the total bandwidth percentage is reserved and unavailable to the other protocol. If a protocol is not being used, the protocol distribution is ignored, and 100% is used by the one protocol being used.

See the CLI syntax for setting the protocol (FCIP and IP Extension) distribution in Example 7-12. The tunnel must have IP Extension enabled; otherwise, only FCIP is enabled.

Example 7-12   CLI syntax for setting the protocol (FCIP and IP Extension) distribution

```
SW37_7850_A:FID128:admin> portcfg fciptunnel
Usage:  portcfg fciptunnel [<slot>/]<port> { create [<args>] | modify [<args>] |
delete | --help }
Option:   create - Create the specified tunnel/circuit
          modify - Modify the specified tunnel/circuit
          delete - Delete the specified tunnel/circuit
          help   - Show this usage message
Optional Arguments:
  -p,--distribution [<mode>,]<percentage ratio,...> -
                                 - Set protocol distribution ratio for the tunnel.
                                   mode:protocol ratio:<fc>,<ip>
  -Q,--fc-qos-ratio <high>,<med>,<low> -
                                 - Set the bandwidth ratio for FC priorities.
                                   [distribution:protocol only].
  -I,--ip-qos-ratio <high>,<med>,<low> -
                                 - Set the bandwidth ratio for IP priorities.
                                   [distribution:protocol only].
  -q,--qos-bw-ratio { <ratio> | default } -
                                 - Set the QoS bandwidth percentages for FC
                                   and/or IP or restore the defaults with
                                   'default' option. Ratio syntax: FCIP-Only
                                   Tunnels: <fcHigh>,<fcMed>,<fcLow> Hybrid
                                   Tunnels:
```

<fcHigh>,<fcMed>,<fcLow>,<ipHigh>,<ipMed>,<ipLow>

Example of setting the FCIP and IP Extension distribution to 60% and 40%:

```
portcfg fciptunnel 24 modify --distribution 60,40
!!!! WARNING !!!!
Modify operation can disrupt the traffic on the fciptunnel specified for a brief
period of time. This operation will bring the existing tunnel down (if tunnel is
up) before applying new configuration.
Continue with Modification (Y,y,N,n): [ n]     y
Operation Succeeded.
```

Example: IP Extension not enabled.

```
portcfg fciptunnel 24 modify --distribution 50,50
Tunnel modify failed: Tunnel configuration parameter out of range.
Tunnel 24 is configured for FCIP only. Cannot change the distribution ratio.
```

Enable IP Extension on a tunnel.

```
SW37_7850_A:FID128:admin> portcfg fciptunnel 24 modify --ipext enable
```

Error when distribution does not total 100%.

```
SW37_7850_A:FID128:admin> portcfg fciptunnel 24 modify --distribution 60,60
Tunnel modify failed: Tunnel configuration parameter out of range.
Tunnel 24 User-group Bandwidth Ratio is invalid. All User-groups must add up to
100%.
```

Priority bandwidth distribution

In the second tier of extension QoS, each protocol (FCIP and IP Extension) supports its own priority (high, medium, and low) distribution. Traffic classified into high, medium, and low priorities is allotted a percentage of the bandwidth. Priority distribution has a tunnel scope. It is configured at the tunnel level, not per circuit.

A set of internal virtual tunnels is created for each protocol and priority. Created within each circuit are seven virtual tunnels, one F-Class, three for IP Extension (H/M/L), and three for FCIP (H/M/L). Medium priorbty is the default for both protocols. IBM extension uses Priority-TCP-QoS (PTQ), which opens a set of WAN-Optimized TCP sessions for each protocol's priority. Flows are automatically mapped to the TCP session of its corresponding priority. No TCP session carries more than one priority.

The priority is based on the virtual circuit (VC) that carries the data into the DP. For FC, zoning with QoS prefixes assigns the flows to a specific set of VCs; in turn, VCs are mapped to extension priority TCP sessions. For example, if data enters through a high VC, it is placed into high TCP sessions; if it enters through a low VC, it is placed into low TCP sessions. The Traffic Control List (TCL) matches flows and assigns them to the correlating priority TCP sessions for IP Extension.

High, medium, and low are each assigned a percentage of bandwidth; the default is 50%, 30%, and 20%. The ratios must add up to 100%. If traffic flows on a priority, its bandwidth becomes reserved and is unavailable to other priorities. Suppose only the medium priority (default) is used, for example. In that case, no flows have been designated as high or low, and all bandwidth is made available to medium, even if the medium queue remains configured at 30% (default).

An egress queue is serviced once every millisecond, and the amount of traffic sent on that queue is based on the priorities being used and the configured ratios.

► The default distribution is 50/30/20% (H/M/L).

► Medium is the default priority.

► The distribution (H/M/L) must total 100%.

► 10% is the minimum value.

► 90% is the maximum value.

► Distribution ratios must be the same on both ends of a tunnel.

► FCIP QoS assignment (H/M/L) is configured using zoning.

► IP Extension QoS assignment (H/M/L) is configured using TCL.

► Priority distribution between high, medium, and low is rate-limiting. If a priority is used, its total bandwidth percentage is reserved and unavailable to other priorities. If a priority is not being used, its percentage of bandwidth is available for the other priorities to use.

► If your storage device supports Fibre Channel CS_CTL prioritization, the CS_CTL values in the FC header to prioritize the QoS traffic.

DSCP marking

Differentiated Services Code Point (DSCP) operates across IP (layer 3). Its implementation for establishing QoS policies is defined in RFC 2475. DSCP uses six bits of the Type of Service (TOS) field in the IP header to establish up to 64 different classes of service.

Enterprise network administrators assign service classes and configure the network to handle traffic accordingly. DSCP settings are helpful only if routers and switches are configured to enforce such policies uniformly across the network. Extension control, high, med, and low-priority flows can be configured with different DSCP values.

The IP network must implement Per Hop Behavior (PHB) with the appropriate DSCP enforcement before enabling DSCP marking.

> **Note:** Whenever IPsec is enabled on a tunnel, all priorities inside a circuit must use the same QoS markings.

Example 7-13 shows the CLI syntax for setting DSCP marking on a circuit.

Example 7-13   CLI syntax for setting DSCP marking on a circuit

```
SW37_7850_A:FID128:admin> portcfg fcipcircuit
Usage:  portCfg fcipcircuit [<slot>/]<port> { create [<args>] | modify [<args>] |
delete | --help } <cid>
Option:   create - Create the specified tunnel/circuit
          modify - Modify the specified tunnel/circuit
          delete - Delete the specified tunnel/circuit
          help  - Show this usage message
Optional Arguments:
    --dscp-f-class <dscp>     - Set DSCP marking for Control traffic.
    --dscp-high <dscp>        - Set DSCP marking for FC-High priority traffic.
    --dscp-medium <dscp>      - Set DSCP marking for FC-Medium priority traffic.
    --dscp-low <dscp>         - Set DSCP marking for FC-Low priority traffic.
    --dscp-ip-high <dscp>     - Set DSCP marking for IP-High priority traffic.
    --dscp-ip-medium <dscp>   - Set DSCP marking for IP-Medium priority traffic.
    --dscp-ip-low <dscp>      - Set DSCP marking for IP-Low priority traffic.
```

Example of setting the DSCP marking for FCIP High to 46 on tunnel 24, circuit 0:

```
portcfg fcipcircuit 24 modify 0 --dscp-high 46
```

Example of setting the DSCP marking for IP Extension low to 3 on tunnel 24, circuit 0:

```
Portcfg fcipcircuit 24 modify 0 --dscp-ip-low 2
```

Example of disabling DSCP marking for FC on tunnel 24, circuit 0:

```
portcfg fcipcircuit 24 modify 0 --dscp-high 0
```

L2CoS marking (802.1P)

VLAN traffic uses 802.1Q tags within an Ethernet header. The tag includes a VLAN ID and the Class of Service (CoS) priority bits. To set L2CoS, the IPIF must be configured with a VLAN ID. The 802.1P L2CoS priority scheme uses three priority bits, allowing eight priorities. Consult with your WAN administrator to determine if setting L2CoS is useful and which L2CoS value can be used.

Enterprise network administrators assign classes of service and configure the network to handle traffic accordingly. L2CoS settings are helpful only if routers and switches are

configured to enforce such policies uniformly across the network. Extension control, high, med, and low-priority flows can be configured with different L2CoS values.

The IP network must implement Per Hop Behavior (PHB) with the appropriate L2CoS enforcement before enabling L2CoS marking.

> **Note:** Whenever IPsec is enabled on a tunnel, all priorities inside a circuit must use the same QoS markings.

Example 7-14 shows the CLI syntax for setting L2CoS marking on a circuit.

Example 7-14   CLI syntax for setting L2CoS marking on a circuit

```
SW37_7850_A:FID128:admin> portcfg fcipcircuit
Usage:   portCfg fcipcircuit [<slot>/]<port> { create [<args>] | modify [<args>] |
delete | --help } <cid>
Option:    create - Create the specified tunnel/circuit
           modify - Modify the specified tunnel/circuit
           delete - Delete the specified tunnel/circuit
           help   - Show this usage message
Optional Arguments:
    --l2cos-f-class <l2cos>   - Set L2CoS value for Control priority traffic.
    --l2cos-high <l2cos>      - Set L2CoS value for FC-High priority traffic.
    --l2cos-medium <l2cos>    - Set L2CoS value for FC-Medium priority traffic.
    --l2cos-low <l2cos>       - Set L2CoS value for FC-Low priority traffic.
    --l2cos-ip-high <l2cos>   - Set L2CoS value for IP-High priority traffic.
    --l2cos-ip-medium <l2cos> - Set L2CoS value for IP-Medium priority traffic.
    --l2cos-ip-low <l2cos>    - Set L2CoS value for IP-Low priority traffic.
```

Example of setting the L2CoS marking for FCIP to 3 on tunnel 24, circuit 0:

portcfg fcipcircuit 24 modify 0 --l2cos-high 3

Example of setting the L2CoS marking for IP Extension to 2 on tunnel 24, circuit 0:

portcfg fcipcircuit 24 modify 0 --l2cos-ip-low 2

Example of disabling L2CoS marking for IP Extension on tunnel 24, circuit 0:

portcfg fcipcircuit 24 modify 0 --l2cos-ip-low 0

7.3.14 FastWrite

FastWrite has a tunnel scope; it is not configured per circuit. The best practice is to have the same version of Brocade Fabric OS at both ends of the tunnel. Either only FastWrite or FastWrite and OSTP are enabled at both ends.

If multiple tunnels with protocol optimization are required between the exact two domains, a Virtual Fabric Logical Fabric (LF) must be implemented. FastWrite requires a deterministic FC path between the initiator and the target. When multiple protocol-optimized tunnels exist on the same platform, two logical fabrics can be implemented with one tunnel to provide a deterministic path for each tunnel. Non-controlled, parallel, equal-cost multipath tunnels are unsupported between two domains when protocol optimization is enabled on one or more tunnels. Protocol optimization requires the source ID and destination ID (SID/DID) pairs to remain on a specific tunnel during the exchange of I/Os.

> **Note:** IBM b-type extension recommends identical Fabric OS versions at each end. When planning Fabric OS upgrades or downgrades, it is recommended that both ends of an extension tunnel have the same Fabric OS version

Example 7-15 shows the CLI syntax for setting FastWrite.

Example 7-15   CLI syntax for setting FastWrite

```
SW37_7850_A:FID128:admin> portcfg fciptunnel
Usage:  portCfg fciptunnel [<slot>/]<port> { create [<args>] | modify [<args>] |
delete | --help }
Option:   create - Create the specified tunnel/circuit
          modify - Modify the specified tunnel/circuit
          delete - Delete the specified tunnel/circuit help
- Show this usage message
Optional Arguments:
  -f,--fastwrite              - Enable / Disable the fastwrite option.
```

Example of enabling Fastwrite:

```
SW37_7850_A:FID128:admin> portcfg fciptunnel 24 modify --fastwrite enable
!!!! WARNING !!!!
The fastwrite feature is incompatible with multiple equal-cost paths.
Please ensure that there are no multiple equal-cost paths in your fabric before
continuing.
Continue with operation (Y,y,N,n): [ n] y
!!!! WARNING !!!!
Delayed modify operation will disrupt traffic on the fcip tunnel specified. This
operation will bring the existing tunnel down (if tunnel is up) for about 10
seconds before applying the new configuration.
Continue with delayed modification (Y,y,N,n): [ n]      y
Operation Succeeded.
```

Example of disabling Fastwrite:

```
SW37_7850_A:FID128:admin> portcfg fciptunnel 24 modify --fastwrite disable
!!!! WARNING !!!!
Delayed modify operation will disrupt traffic on the fcip tunnel specified. This
operation will bring the existing tunnel down (if tunnel is up) for about 10
seconds before applying the new configuration.
Continue with delayed modification (Y,y,N,n): [ n]      y
Operation Succeeded.
```

7.3.15 OSTP

OSTP has a tunnel scope; it is not configured per circuit. FastWrite and OSTP have to be enabled at both ends. To use OSTP, you must enable FastWrite as well.

> **Note:** Best practice is to have the same version of Fabric OS at both ends of the tunnel.

If multiple tunnels with protocol optimization are required within the same platform, A Virtual Fabric Logical Fabric (LF) must be implemented. OSTP requires a deterministic FC path

between the initiator and the target. When multiple protocol-optimized tunnels exist on the same platform, two logical fabrics can be implemented with one tunnel to provide a deterministic path for each tunnel. Non-controlled, parallel, equal-cost multipath tunnels are unsupported between two domains when protocol optimization is enabled on one or more tunnels. Protocol optimization requires the source ID and destination ID (SID/DID) pairs to remain on a specific tunnel during the exchange of I/Os.

Control blocks created during FCP traffic flows

For FCP/SCSI traffic flows, tunnel processing creates control block structures based on the SID/DID pairs used between devices. Upon enabling FastWrite or OSTP (read or write), additional structures and control blocks are created for each logical unit number (LUN) based on SID/DID pairs. If multiple SID/DID pairs can access the same LUN, FCP processing in an emulated tunnel creates multiple control blocks for each LUN.

The following control blocks are created:

► Initiator-Target Nexus (ITN) structure: Each FCP-identified SID/DID flow is recorded in an ITN data structure.

► Initiator-Target-LUN (ITL) control block: Each specific LUN on a SID/DID flow has an ITL control block created for the flow.

► Turbo-Write Block (TWB) structure: FCP emulation processing creates a TWB structure for each outstanding FC exchange.

Example of enabling OSTP without Fastwrite being enabled:

```
SW37_7850_A:FID128:admin> portcfg fciptunnel 24 modify --tape-pipelining enable
!!!! WARNING !!!!
Delayed modify operation will disrupt traffic on the fcip tunnel specified. This
operation will bring the existing tunnel down (if tunnel is up) for about 10
seconds before applying the new configuration.
Continue with delayed modification (Y,y,N,n): [ n]      y
Tunnel modify failed: Tape-Pipelining feature requires Fastwrite to be configured.
```

Example of enabling OSTP with Fastwrite enabled:

```
SW37_7850_A:FID128:admin> portcfg fciptunnel 24 modify --tape-pipelining enable
!!!! WARNING !!!!
Delayed modify operation will disrupt traffic on the fcip tunnel specified. This
operation will bring the existing tunnel down (if tunnel is up) for about 10
seconds before applying the new configuration.
Continue with delayed modification (Y,y,N,n): [ n]      y
Operation Succeeded.
```

7.3.16 Advanced FICON Accelerator

The IBM SAN42B-R7 Advanced Accelerator for FICON supports Exchange Based Routing (EBR). This support comprises two significant changes:

► VT and egress VC assignments occur at the start of each FICON exchange.

► Advanced FICON Accelerator processing creates a new internal object called an exchange. The exchange object is created when a new sequence (OXID) is received from the channel or the control unit and exists during the active sequences. The exchange control block is similar to the FCP/SCSI TWB structure but for FICON flows.

This change should improve FICON over FCIP performance as FICON flows will not all be mapped to the same VT or VC; therefore, head-of-line blocking is avoided in the WAN and FC egress paths.

Control blocks created during FICON traffic flows

For FICON traffic flows, tunnel processing creates a control block structure based on the SID/DID pairs called a FICON device path block (FDPB). When the FICON emulation is enabled, additional control blocks are created for each SID/DID pair, logical partition (LPAR) number (FICON channel block structure), and LCU number (FICON control unit block structure), and each FICON device address on those LCUs. FICON Exchange control blocks are also created if the switch is configured to operate in EBR mode.

The total number of FICON device control blocks (FDCBs) that are created over a FICON emulating tunnel is represented by the following equation:

```
FDCBs = Host Ports × Device Ports × LPARs × LCUs × FICON Devices per LCU
```

This number increases rapidly in extended direct-attached storage device (DASD) configurations, such as those used for IBM z/OS Global Mirror, also known as Extended Remote Copy (XRC).

FDCB example

In the following example, the tunnel extends two channel path IDs (CHPIDs) from a System Data Mover (SDM) site to a production site. There are also two SDM-extended LPARs, the IBM DS8000 production controllers have 32 LCUs per chassis, and each LCU has 256 device addresses. Based on the previous equation, the number of extended FICON device control block images created is as follows:

```
2 Host Ports × 2 Device Ports × 2 LPARs × 32 LCUs × 256 Devices per LCU = 56,536
FDCBs
```

Advanced FICON Accelerator considerations

Use the output from the command:

portshow xtun slot/ve_port -fcp -port –stats

Output from the command below helps determine how a tunnel configuration is affecting tunnel control block memory:

portshow xtun slot/ve_port -dram2

As a guideline, up to 80% of the tunnel DP control block memory pool (dram2) should be allocated for SID/DID pair-related control blocks (ITNs, ITLs, FDPBs, FCHBs, FCUBs, and FDCBs). When more than 80% of the pool is allocated, consider redesigning the tunnel configuration to ensure a continuous operation. Consider the number of SID/DID pairs in the tunnel configuration when redesigning it and determine whether new switches, chassis, or blades are necessary to reduce the impact on DRAM2.

DP complexes such as the IBM SAN42B-R7 extension platform and the IBM SX6 extension blade generate the XTUN-1008 RASlog message when the following percentages of DRAM memory remain:

► 50%

► 25%

► 12.5%

- ► 6.25%

- ► 0.05%

RASlog messages include information regarding the amount of allocated memory, free memory, and the total size of the pool. Refer to the RASlog messages to determine if you must reduce the size of the extended configuration or plan for additional switch resources.

The IBM b-type extension DPs are limited to the number of FICON device control blocks (FDCB) and extended LUNs (ITL) listed in Table 7-13.

Table 7-13   FDCB and ITL per DP

Platform	FDCB	ITL
IBM SAN42B-R7	512,000	200,000
IBM SAN18B-6	0 (FICON not supported)	30,000
IBM SX6 Extension Blade	512,000	200,000

7.3.17 Circuit failover

A metric paradigm is used on circuits for failover and failback. Metric 0 is used for regular production, and metric 1 is used for standby. When configuring a circuit, metric 0 is the default; all circuits are active if no metrics were configured, which is typical. When all circuits are active and a circuit goes down, data fails over to the remaining circuits using LLL. The available bandwidth is the aggregate of the remaining active circuits. Within a failover group, metric-1 circuits are not included in the aggregate bandwidth if metric-0 circuits are online.

Failover metrics

Failover metrics and groups are features supported on IBM b-type extension platforms. A group defines a set of metric-0 and metric-1 circuits. Within that group, failover protection is provided. All metric-0 circuits in a group must fail before metric-1 circuits become active. When the metric-0 circuits within a group fail, the metric-1 circuits are immediately activated. Each failover group is independent.

Typically, one metric-0 and one metric-1 circuit are configured per group, which provides one-to-one protection. Circuits in a group must be connected to the same VE_Port to have LLL, meaning no data is lost during failovers and failbacks.

A circuit's Keepalive Timeout Value (KATOV) is the amount of time before detecting an offline circuit.

Failover and failback of circuits prevent interruptions by seamlessly switching offline circuits to a designated backup. To manage failover, each circuit is assigned a metric value, either 0 (default = production) or 1 (failover), which manages the failover from one circuit to another within a group.

BET uses LLL; no data in-flight is lost during a failover or failback event. LLL retransmits data over another metric-0 circuit if a circuit goes offline. In Figure 7-2, circuit-0 and 1 are both metric-0 circuits. Circuit-0 has gone offline, and transmission fails to circuit-1, which has the same metric. Traffic at the time of failure is retransmitted over circuit-1. LLL ensures all data is delivered and delivered in the same order as it was sent.

Figure 7-2   BET failover of metric-0 circuits

Failover groups

Circuit failover groups control which metric-1 circuits are activated when a set of metric-0 circuits have gone offline. A set of metric-0 and metric-1 circuits are configured into groups. When all metric-0 circuits within the group have gone offline, the metric-1 circuits become active. As each failover group is autonomous, circuits in other failover groups are unaffected. The purpose of a failover group is to sort circuits so metric-1 circuits can back up the metric-0 circuits if they fail. For example, two circuits (metric 0 and metric 1) in one group would form a one-to-one backup. Failover groups are numbered; group ID is a value from 0 through 9, and the default is 0.

Typically, one metric-0 and one metric-1 circuit are grouped. This aims to replace the offline metric-0 circuit with an online metric-1 circuit as quickly as possible. As a result, the problem of a degraded tunnel is avoided, and other metric-0 circuits in different groups remain operational.

The following two types of configuration are supported:

► Active-Active: Data is weight balanced across multiple circuits. All circuits are metric 0.

► Active-Passive: If all metric-0 circuits go offline within a failover group, data fails over to the metric-1 circuits. During failover/failback, LLL guarantees that all data is delivered and delivered in order.

In Figure 7-3 , circuit-0 is assigned a metric of 0, and circuit-1 is assigned a metric of 1. Both circuits are in the same tunnel (VE24). Circuit-1 does not become active until all the metric-0 circuits within its failover group have gone offline. If all metric-0 circuits go offline within the failover group, the traffic is retransmitted and sent over available metric-1 circuits within the failover group. Failover between like metric circuits or between different metric circuits is lossless.

Figure 7-3   Failover to metric -1 circuit

There might be differences in behavior between the metric-1 circuits and the production circuits (metric-0) if the configuration of the metric-1 circuits differs from the configuration of the metric-0 circuits. Also, if the metric-1 circuits' WAN path has different characteristics, the behavior might differ from production.

With circuit failover groups, you can control which metric-1 circuits are activated if metric-0 circuits fail. To create circuit failover groups, define a set of metric-0 and metric-1 circuits within a joint failover group. When all metric-0 circuits within the group fail, metric-1 circuits are activated. A failover group is autonomous.

Typically, there is only one metric-0 circuit in a group, so when the metric-0 circuit fails, a specific metric-1 circuit takes over, known as one-to-one failover grouping. A one-to-one configuration prevents a tunnel from operating in a degraded mode because a subset of the overall metric-0 circuits failed.

Circuit failover considerations and limitations

Circuit failover groups operate under the following conditions:

► Each failover group is independent and operates autonomously.

► All metric-0 circuits in a failover group must fail before any metric-1 circuits become active.

► All metric-1 circuits in a failover group are used if there are no metric-0 circuits.

► Circuits can participate in only one failover group.

► Both ends of a tunnel must have the same failover groups defined.

► Tunnel and circuit states indicate a misconfiguration error if failover group configurations are invalid.

► Modifying the metric is a disruptive operation.

► Modifying the failover group ID is a disruptive operation.

► Failover groups do not define load balancing over circuits.

► If no failover group is defined, the default operation is: all metric-0 circuits must fail before metric-1 circuits become active.

► For a failover group to be valid, at least one metric-0 and one metric-1 circuit must be added; otherwise, a warning is displayed.

► The number of failover groups is limited by the number of circuits created per tunnel on the extension platform.

► Consider available WAN bandwidth requirements when configuring failover groups.

See Table 7-14.

Table 7-14   Number of failover groups per platform

IBM b-type Extension Platform	Number of Failover Groups on Platform
IBM SAN18B-R	Up to 3 valid groups per 6-circuit tunnel
IBM SAN42B-R	Up to 5 valid groups per 10-circuit tunnel
IBM SAN42B-R7	Up to 5 valid groups per 10-circuit tunnel.

Bandwidth calculation during circuit failover

When all metric-0 circuits within the failover group fail, the bandwidth of metric-1 circuits is not calculated as the available bandwidth in a tunnel. Consider the following circuit configurations for circuits 0 through 3:

► Circuits 0 and 1 are metric-0

  – Circuit 0 is created with a maximum rate of 1 Gbps

  – Circuit 1 is created with a maximum rate of 500 Mbps

  – The combined bandwidth of circuits 0 and 1 is 1.5 Gbps

► Circuits 2 and 3 are metric-1

  – Circuit 2 is created with a maximum rate of 1 Gbps

  – Circuit 3 is created with a maximum rate of 1 Gbps

  – The combined bandwidth of circuits 2 and 3 is 2 Gbps

  – The metric 1 bandwidth is held in reserve

► The following occurs during a circuit failure:

– If either circuit 0 or 1 goes offline, traffic continues to flow over the remaining metric-0 circuit. The available bandwidth is still considered 1.5 Gbps.

– If circuits 0 and 1 go offline, failover to circuits 2 and 3; the available bandwidth is updated to 2 Gbps.

– If a metric-0 circuit comes online, the metric-1 circuits return to standby status, and the available bandwidth is updated as each circuit comes online. For example, if circuit-0 is recovered, the available bandwidth is updated to 1 Gbps. When circuit-1 recovers, the available bandwidth is updated to 1.5 Gbps.

Circuit failover examples

During a circuit failover, the following events occur:

► If circuit-0 fails, circuit-2 becomes active and data is load-balanced over circuit-1 and circuit-2.

► If circuit-1 fails, circuit-3 becomes active and data is load-balanced over circuit-0 and circuit-3.

► If both circuit-0 and circuit-1 fail, circuit-2 and circuit-3 become active and data is load-balanced over circuit-2 and circuit-3.

For circuit failover in a tunnel, two failover groups, each with two circuits, are shown in Table 7-15.

Table 7-15   Circuits with two failover groups

Failover Group ID	Tunnel Bandwidth	FSPF Link Cost if Circuit Goes Offline	In use for tunnel data?
	500 Mbps	1500	If active, yes
	500 Mbps	1500	Only when circuit 0 fails
	1000 Mbps	1000	If active, yes
	1000 Mbps	1000	Only when circuit 1 fails

The following table shows a tunnel with five circuits. Two circuits have a failover group defined (Group 1), and three circuits are left in the default failover group (Group 0). In this example, all data is initially balanced across circuits-0, circuit-1, and circuit-2 because they all have metric 0. The following events occur when various circuits go offline:

► Circuit-0 goes offline

– Circuit-3 becomes active in group-1

– Data is balanced over circuit-1, circuit-2, and circuit-3

– Circuit-0 fails over to circuit-3 in group-1

– Circuits-1, 2, and 3 have equal cost (each has a bandwidth of 500 Mbps).

► Additionally, circuit-2 goes offline:

– Data is balanced over circuit-1 and circuit-3

– No other circuits are active

– Circuit-1 and circuit-3 are the only online circuits because circuit-4 becomes active only when both circuit-1 and circuit-2 are offline.

– If circuit-1 and circuit-2 go offline, circuit-4 becomes active and data is balanced over circuit-0 and circuit-4. Circuit-0 in group-1 is online, and circuit-1 and circuit-2 in group-0 failover to circuit-4.

▶ Circuit-0, circuit-1, and circuit-2 go offline:

– Circuit-3 and circuit-4 become active

– Data is balanced over circuit-3 and circuit-4

– Circuit-1 and circuit-2 fails over to circuit-4 in group-0

– Circuit-0 fails over to circuit-3 in group-1

The Table 7-16 describes a tunnel with five circuits. Three circuits are in failover group-0 (default group), and two are in failover group-1.

Table 7-16   Circuit failover groups

Failover Group ID	Circuits in Tunnel	Tunnel Bandwidth	FSPF Link Cost if Circuit Goes Offline	In Use for Tunnel Data?
	Circuit-1 Metric 0	500 Mbps	1500	If active, yes
	Circuit-2 Metric 0	500 Mbps	1500	If active, yes
	Circuit-4 Metric 1	1000 Mbps	1000	Only when circuit-1 and circuit-2 fail
	Circuit-0 Metric 0	500 Mbps	1500	If active, yes
	Circuit-3 Metric 1	500 Mbps	1500	Only when circuit-0 fails

Active-active circuits

An active-active configuration has two circuits, and both are configured with the same metric, metric 0. One circuit uses ge0, and the other circuit uses ge1. Both circuits are configured for 10 Gbps and are sending data. The load is balanced across the two circuits. The effective tunnel bandwidth is 20 Gbps.

1. Create IPIFs for circuits (no VLAN or MTU specified).

```
portcfg ipif ge0.dp0 create 192.168.10.10/24
portcfg ipif ge1.dp0 create 192.168.10.11/24
```

2. Create IP routes for the IPIF.

```
portcfg iproute ge0.dp0 create 10.0.0.0/24 192.168.10.1
portcfg iproute ge1.dp0 create 10.0.0.0/24 192.168.10.1
```

3. Create Tunnel 24.

```
portcfg fciptunnel 24 create --ipsec MyIPsecPolicy --compression deflate
```

4. Create two circuits to tunnel 24.

```
portcfg fcipcircuit 24 create 0
portcfg fcipcircuit 24 create 1
```

5. Add circuit IP addresses .

```
portcfg fcipcircuit 24 modify 0 --local-ip 192.168.10.10
--remote-ip 192.168.10.20
portcfg fcipcircuit 24 modify 1 --local-ip 192.168.10.11
--remote-ip 192.168.10.21
```

6. Add 10 Gbps bandwidth to each circuit.

```
portcfg fcipcircuit 24 modify 0 --min 10G --max 10G
portcfg fcipcircuit 24 modify 1 --min 10G --max 10G
```

7. Display tunnel, circuits, bandwidth, failover metrics, and group settings.

```
portshow fciptunnel -c
```

> **Note:** If the source and destination addresses are on different subnets, you must configure IP routes to reach the gateway to the destination. See "IP routes" on page 78.

Active-passive circuits

The following example shows an active-passive configuration with two circuits. Circuit-0 has metric-0 (default, production), and circuit-1 has metric-1 (failover). Both circuits are grouped into failover group-1. One circuit uses ge0, and the other circuit uses ge1. In this example, circuit-1 is the failover circuit because it has metric-1. When circuit-0 goes down, the traffic fails over to circuit-1. The effective bandwidth of the tunnel in this example is 1 Gbps.

Perform the following steps to configure an active-passive circuit:

1. Create IPIFs.

```
portcfg ipif ge0.dp0 create 192.168.10.10/24
portcfg ipif ge1.dp0 create 192.168.10.11/24
```

2. Create IP Routes for IPIFs.

```
portcfg iproute ge0.dp0 create 10.0.0.0/24 192.168.10.1
portcfg iproute ge1.dp0 create 10.0.0.0/24 192.168.10.1
```

3. Create Tunnel 24.

```
portcfg fciptunnel 24 create --ipsec MyIPsecPolicy --compression deflate
```

4. Create two circuits on tunnel 24.

```
portcfg fcipcircuit 24 create 0
portcfg fcipcircuit 24 create 1
```

5. Add circuit IP addresses.

```
portcfg fcipcircuit 24 modify 0 --local-ip 192.168.10.10
--remote-ip 192.168.10.20
portcfg fcipcircuit 24 modify 1 --local-ip 192.168.10.11
--remote-ip 192.168.10.21
```

6. Add 10 Gbps bandwidth to each circuit.

```
portcfg fcipcircuit 24 modify 0 --min 10G --max 10G
portcfg fcipcircuit 24 modify 1 --min 10G --max 10G
```

7. Add metrics and groups to each circuit

```
portcfg fcipcircuit 24 modify 0 --metric 0 --failover-group 1
portcfg fcipcircuit 24 modify 1 --metric 1 --failover-group 1
```

8. Display tunnel, circuits, bandwidth, failover metrics, and group settings.

```
portshow fciptunnel -c
```

> **Note:** If the source and destination addresses are on different subnets, you must configure IP routes to the destination addresses.

Failover considerations

Circuit failover groups operate under the following conditions:

► Failover group IDs can range from 0-9.

► The default failover group is 0.

► Up to 5 failover groups can be defined per tunnel.

► Circuits can participate in only one failover group.

► Both ends of a circuit must have the same failover group ID.

► Each failover group is independent and operates autonomously.

► All metric-0 circuits in a failover group must fail before the metric-1 circuits are utilized.

► All metric-1 circuits in a failover group are used if there are no online metric-0 circuits.

► Tunnel and circuit states indicate a misconfiguration error if circuit failover group configurations are invalid.

► Modifying the metric is a disruptive operation.

► Modifying the failover group ID is a disruptive operation.

► Circuit failover groups do not define load balancing over circuits.

A valid failover group requires at least one metric-0 and one metric-1 circuit; otherwise, a warning is displayed. The number of failover groups per tunnel is limited by the number of circuits that can be created on the extension platform as follows:

► The IBM SAN18B-6 can configure up to three valid groups on a 6-circuit tunnel.

► The IBM SAN42B-R7 can configure up to five valid groups on a 10-circuit tunnel.

► The IBM SX6 Extension Blade can configure up to five valid groups on a 10-circuit tunnel.

Consider circuit bandwidth calculations when configuring failover circuit groups. A tunnel is an ISL with an FSPF cost associated with it. The FSPF cost can change as circuits go on and offline. For a single route to the destination domain, a changing FSPF cost is inconsequential. Suppose there are multiple routes to the destination domain. A changing FSPF cost can cause disruption. In this case, static FSPF costs can be associated with various tunnels to control traffic routes.

7.3.18 Circuit spillover

Primary circuits (metric-0) are regular production circuits. Secondary circuits (metric-1) are spillover circuits. Spillover circuits are used during high-utilization or congestion periods.

Primary and secondary circuits are controlled using the circuit's metric field. For example, when a tunnel is configured for spillover, traffic uses the metric-0 circuits until their maximum bandwidth is reached. When the bandwidth in the metric-0 circuits is exceeded, the metric-0 and metric-1 circuits are used. Conversely, when the utilization drops below the maximum on the metric-0 circuits, the metric-1 circuits are no longer used.

Tunnel failover remains intact. If a tunnel has spillover configured and the metric-0 circuits within a failover group go offline, the metric-1 circuits within the group are activated. Whether configured for failover or spillover, metric-1 circuits behave as backup circuits. Unlike spillover, failover requires all metric-0 circuits to go offline before the metric-1 circuits are used within a failover group.

Circuit failover groups can be configured along with spillover and behave the same without spillover configured; however, only the metric value determines whether a circuit is considered primary or spillover. Failover groups do not affect which circuits are used for spillover. For example, if a tunnel has four circuits, group-0 has a metric-0 and a metric-1, and group-1 has a metric-0 and a metric-1. When the metric 0 circuits become saturated, the metric 1 circuits start to be used. Also, if the metric-0 circuit in group-0 goes offline, the

metric-1 circuit becomes active, and the same would be true in group-1 if its metric-0 went offline.

Due to how spillover is implemented, the observed behavior might be different than expected. For example, consider traffic flows with high, medium, and low QoS, with corresponding bandwidth percentages of 50/30/20. The active circuits carry QoS traffic according to the percentage allocated to each priority. If the QoS low-priority traffic reaches saturation before the bandwidth allocation limit, it spills over from a metric-0 circuit to a metric-1 circuit.

For example, consider QoS traffic flows designated as high, medium, and low. The high QoS traffic flow can be assigned to metric-1 circuits, while the medium and low QoS traffic flow can be assigned to metric-0 circuits. In this instance, the spillover circuits (metric 1) are used even though the metric 0 circuits are not saturated. When the metric-0 circuits are saturated, additional traffic will spill over to the metric-1 circuits.

Circuit spillover considerations

In the event of a circuit spillover, the following factors should be considered:

► Spillover has a tunnel scope. It is not configured per circuit.

► When a tunnel is configured for spillover, all metric-1 circuits are treated as metric-0 circuits when calculating aggregate bandwidth restrictions.

► Failover groups are not used when configuring the tunnel for spillover, and any defined failover groups are ignored.

► Spillover behavior is similar to failover behavior in that if metric-0 circuits are not available, metric-1 circuits are used.

Example 7-16 shows the CLI syntax to change the failover or spillover setting.

Example 7-16   CLI syntax to change the failover or spillover setting

```
SW37_7850_A:FID128:admin> portcfg fciptunnel
Usage:   portCfg fciptunnel [<slot>/]<port> { create [<args>] | modify [<args>] |
delete | --help }
Option:    create - Create the specified tunnel/circuit
           modify - Modify the specified tunnel/circuit
           delete - Delete the specified tunnel/circuit
           help   - Show this usage message
Optional Arguments:
  -L,--load-leveling { default | failover | spillover} -
                                 - Set load leveling algorithm.
```

Example of enabling spillover:

```
portcfg fciptunnel 24 modify --load-leveling spillover
Operation Succeeded.
```

Example of disabling spillover:

```
portcfg fciptunnel 24 modify --load-leveling default
Operation Succeeded.
```

7.3.19 SLA (Service-Level Agreement)

An SLA session must be configured at each end of the tested circuit. If the circuit configurations specify different transmission rates, SLA uses the lower rate. This allows SLA to start even when circuit configurations mismatch. If the circuit configurations specify

different packet loss values, SLA uses the lower value. The lower the packet loss value, the better the circuit quality, ensuring the circuit adheres to better quality. When an SLA session is triggered, traffic starts automatically. During the test, packet loss must remain under the specified percentage before the circuit returns to service.

SLA considerations

The following are the SLA considerations:

► The default SLA runtime is 5 minutes.

► The default SLA timeout is 0 (expiration time not set).

► If setting the SLA timeout, it must be greater than the runtime value.

► After a circuit goes offline, there are two ways a circuit can be placed back into service:

   – Packet loss remains under the specified loss percentage for the SLA runtime.

   – SLA timeout was set and expired; the packet loss did not remain under the specified loss percentage during runtime.

► If a timeout value is not set, SLA testing continues indefinitely until packet loss remains under the specified loss percentage for the runtime.

► An active SLA session can be aborted but cannot be modified.

► An active SLA session is limited to viewing statistics.

► During eHCL, SLA is disabled, and no SLA sessions are created until after all eHCL operations have been completed. After all eHCL operations are complete, SLA is re-enabled.

► IBM SAN42B-R7: Up to 20 SLA sessions can be defined per DP

► IBM SX6 Extension Blade: Up to 20 SLA sessions can be defined per DP

► IBM SAN18B-6: Up to 12 SLA sessions can be defined

Attempts to modify an active SLA session are blocked. Wtool commands cannot be used while an SLA session is running. Active SLA sessions can be aborted, and SLA statistics can be viewed.

Configured SLA sessions are persistent across reboots because they are part of the circuit configurations, unlike user-configured Wtool sessions, which are not persistent.

Example 7-17 shows the CLI syntax for creating an SLA profile.

Example 7-17   CLI syntax for creating an SLA profile

```
SW37_7850_A:FID128:admin> portcfg sla
Usage: portcfg sla <name> { create --loss <percentage> [--runtime <minutes>]
[--timeout {<minutes> | none}] | modify [--loss <percentage>] [--runtime
<minutes>] [--timeout {<minutes> | none}] | delete | --help }
Name Format:
<string> - The SLA name(Min 1 character, Max 31 characters).
          Cannot contain the following: ;$!#`/\><&'"=,?.*^{}()
Option:  create - Create the specified SLA
          modify - Modify the specified SLA
          delete - Delete the specified SLA
          help  - Display the help message.
Create/Modify Command Arguments:
    --loss <percentage>  - Set the packet loss percentage (required for create).
                           Range: 0.05 - 5.0
    --runtime <minutes>  - Set the time to run under the
```

```
                              threshold to consider the SLA passed
                            Default:5 minutes
                            Range: 1 - 1440
    --timeout { <minutes> | none } - Set the timeout period to fail the
                            SLA. Default: no timeout
                            Range: 0 - 2880
                            NOTE: 0 = No Timeout
    --help              - show the sla usage statement
```

Example of creating an SLA with loss profile only:

portcfg sla networkA create --loss 0.5
Operation Succeeded.

Example of creating an SLA with loss (required), runtime (default 5 minutes), and timeout (default is no timeout) profiles:

portcfg sla networkB create --loss 1 --runtime 10 --timeout 30
Operation Succeeded.

Example of applying the SLA profile to a tunnel's circuit:

SW37_7850_A:FID128:admin> portcfg fcipcircuit 24 modify 0 --sla networkA
Operation Succeeded.

7.3.20 KATOV (Keepalive Timeout Value)

KATOV is essential for error recovery, and its setting should be based on application requirements. KATOV is configured per circuit. Check with your IP storage application provider to determine the I/O timeout.

When the KATOV sum from all circuits in a tunnel is slightly less than the I/O timeout, this is the appropriate KATOV. For example, a mirroring application has a 6-second I/O timeout. There are three circuits belonging to the tunnel. Set the KATOV to 2 seconds on each circuit, which allows the maximum number of retries across all circuits before an I/O times out by the application.

A 2-second to 3-second KATOV is often optimal for IP Extension and should be considered.

> **Note:** A 250 ms RTT with a maximum of 1% packet loss is supported at a 2 sec or more KATOV.

> **Note:** A 200 ms RTT with a maximum of 0.1% packet loss is supported at a 2 sec or less KATOV.

KATOV considerations

The following are the KATOV considerations:

► KATOV has a circuit scope. It is configured per circuit.

► The default for a non-FICON tunnel, circuit KATOV is 6 seconds.

► KATOV must be the same on both ends of a circuit.

 – If the local and remote KATOV do not match, the tunnel uses the shorter of the configured values.

► FICON best practices:

– A VE_Port used for FICON should be in a FICON logical switch.

– A tunnel that is used for FICON should be created with the FICON argument.

– FICON requires a KATOV of 1 second or less on each circuit.

– Tunnels not initially created with the FICON argument must be modified to become FICON tunnels by correcting the KATOV to 1 second or less. This will not happen automatically.

► The aggregate KATOV across all circuits must be less than the application's I/O timeout. If the I/O timeout is less than the aggregate KATOV, the I/O will time out before the extension platform can recover. IBM extension retries I/Os across each remaining online circuit.

► Changing the KATOV can be disruptive.

Example 7-18 shows the CLI syntax for setting the KATOV to 2 seconds on tunnel 24, circuit 0.

Example 7-18   CLI syntax for setting the KATOV to 2 seconds on tunnel 24, circuit 0:

```
SW37_7850_A:FID128:admin> portcfg fcipcircuit 24 modify 0 --keepaline 2000
Usage:   portCfg fcipcircuit [<slot>/]<port> modify [<args>] <cid>
Optional Arguments:
  -k,--keepalive-timeout <ms>  -  Set keepalive timeout in ms.
```

Example of setting the KATOV:

```
SW37_7850_A:FID128:admin> portcfg fcipcircuit 24 modify 0 --keepalive-timeout 2000
!!!! WARNING !!!!
Modify operation can disrupt the traffic on the fciptunnel specified for a brief
period of time. This operation will bring the existing tunnel down (if tunnel is
up) before applying new configuration.
Continue with Modification (Y,y,N,n): [ n]     y
Operation Succeeded.
```

7.4 LAN side configuration

IBM b-type IP Extension provides acceleration and encryption for IP storage replication. IP Extension uses VE_Ports as logical endpoints for a tunnel, even if no FC or FICON traffic is being transported. Connectivity between two extension platforms creates a fabric. This is of no concern if the platforms are dedicated to IP Extension. Forming a long-distance fabric can be a concern if the platforms are utilized for FC or FICON switching (not FCIP). A separate Virtual Fabric LS for the FC or FICON ports is recommended in this case.

IP Extension uses IP Extension gateways as a communication port with the IP storage end-devices (L2 deployment) or the IP network (L3 deployment). The IP Extension Gateway becomes the gateway for data flows headed to the remote data center.

End-device TCP sessions are locally terminated at the IP Extension Gateway, termed TCP proxy, and reformed on the remote side. This process has the advantage of providing local acknowledgments resulting in acceleration. WAN-optimized TCP (WO-TCP) is used to transport between data centers, not to the storage end devices.

IP Extension requires an understanding of the following points:

► A LAN side IPIF is an IP Extension gateway.

- ► Each DP requires its own unique IP Extension gateway.
- ► IP Extension forms a tunnel between the data centers allowing end devices to communicate securely.
- ► The local and remote IP Extension gateways must be on different subnets.
  - – The same LAN side subnet cannot extend from one data center to the other.
  - – The same subnet can be used on the WAN side between data centers.
- ► On the WAN side, connectivity between data centers can use:
  - – Same subnet (L2, for example dense wavelength-division multiplexing (DWDM) network).
  - – Different subnets (L3, for example routed network).
- ► IP Extension optimizes TCP-based traffic only; other traffic types are supported but not optimized.
- ► Layer 2 deployment.
  - – IBM IP Extension is the next-hop gateway for IP storage flows between data centers.
  - – A LAN side IPIF is created in the same broadcast domain (same subnet) as the attached end devices.
  - – The end device is configured with a specific route for the destination IP address or subnet that forwards replication traffic to the IP Extension gateway.
- ► Layer 3 deployment.
  - – IP Extension platforms support LAN side IP routes to forward data to a destination by the data center's IP network.
  - – LAN side IP routes do not forward traffic to the WAN, TCL is used for the WAN.
  - – End devices send traffic to their traditional network gateway. The gateway router intercepts specific data flows and redirects them to the IP Extension gateway. Policy Based Routing (PBR) is frequently used on an IP router to intercept and forward traffic to a specific gateway.
- ► LAN and WAN side GE interfaces support Link Level Discovery Protocol (LLDP).
- ► LAN side GE interfaces support dynamic and static portchannels (LAGs). A LAG is a group of physical Ethernet links that logically form a single link.
- ► IP Extension uses a traffic control list (TCL) to direct traffic to the correct tunnel. Even if there is a single tunnel. If ingress traffic does not match an "allow" TCL rule, the default behavior is to drop the traffic.
- ► Double VLAN tagging, often called QinQ or nested VLAN tagging, is not supported.

> **Note:** Creating a TCL is required, even if a single tunnel (VE_Port) is implemented. IP Extension does not function without a configured TCL at each end.

IP Extension requires IBM SX6 Extension Blade to be in hybrid mode. Enabling hybrid mode is disruptive because a blade reboot is required to load the hybrid mode image.

7.4.1 IP Extension considerations

The following are IP Extension considerations:

- ► The WAN side configuration should be completed before configuring IP Extension.
  - – IP Extension must be enabled on the tunnel

- ► IP Extension requires a TCL to be configured; the default is to deny traffic.
- ► The number of LAN side TCP connections is limited, and each platform must be considered:
  – Exceeding the number of supported TCP sessions prevents establishing additional end-device TCP sessions.
  – IBM SAN18B-6: 256 TCP and 32 UDP connections (1 DP).
  – IBM SAN42B-R7: 512 TCP and 64 UDP connections per DP (2 DPs).
  – IBM SX6 Extension Blade: 512 TCP and 64 UDP connections per DP (2 DPs).
- ► Balance the load on each DP and the number of TCP sessions.
  – Estimate bandwidth and TCP usage when planning end devices per DP.
  – Each DP has its own IP Extension Gateway.
  – Determine which IP Extension Gateway will be used by each end device.
  – Configure the end device or router accordingly.
- ► Perform the following tasks to configure the IP Extension configuration:
  – Create LAN-side GE interfaces.
  – Create LAN side LAG (portchannel) on GE interfaces.
  – Create LAN side IPIF (IP Extension Gateway). On the LAN side IPIF, add VLAN ID and set MTU if applicable.
  – Create LAN side IP routes, if applicable (layer 3 deployment only).
  – Verify the LAN side IP connectivity.
  – Create TCL.
- ► LAN side GE interfaces do not perform Ethernet (Layer 2) switching. Based on the TCL, traffic is matched and sent over the tunnel <u>or</u> dropped. As a result, loops cannot form, and STP is not required.

For TCP traffic, the following considerations apply:

- ► The LAN side uses the RX window to control source data and prevent buffer overflow. However, the RX windows close when the network experiences congestion or the destination end device exhibits slow drain behavior.
- ► Up to 512 TCP open requests per second are allowed; additional open requests are dropped. Limiting open requests per second mitigates Denial of Service (DoS) attacks.
- ► Statistics are available for the number of dropped connections.

See Table 7-17.

Table 7-17   Maximum supported TCP sessions and RASlog warning threshold

Platform	TCP sessions per DP	RASlog Warning Threshold
IBM SAN18B-R	256 (1 DP)	244
IBM SAN42B-R7	512 (2 DPs)	500
IBM SX6 Extension Blade	512 (2 DPs)	500

7.4.2 Tunnels (LAN side)

A tunnel requires IP Extension to be enabled; it is not enabled by default when creating a tunnel. This tunnel attribute should not be confused with enabling hybrid mode. A tunnel enabled for IP Extension has a protocol bandwidth distribution setting and IP Extension-specific QoS priority percentages. Tunnels can carry FCIP and IP Extension traffic, which is recommended because the tunnel manages the flow control for both protocols and prevents congestion.

Example 7-19 shows the CLI syntax for enabling IP Extension on tunnel 24.

Example 7-19   CLI syntax for enabling IP Extension on tunnel 24

```
SW37_7850_A:FID128:admin> portcfg fciptunnel 24 modify --ipext enable
!!!! WARNING !!!!
Modify operation can disrupt the traffic on the fciptunnel specified for a brief
period of time. This operation will bring the existing tunnel down (if tunnel is
up) before applying new configuration.
Continue with Modification (Y,y,N,n): [ n]      y
Operation Succeeded.
```

7.4.3 GE interfaces (LAN side)

IP Extension cannot function without at least one LAN side GE interface, which requires one or more GE interfaces to be in LAN mode. IP storage replication ports communicate with IP Extension Gateways (LAN side IPIF) through the data center LAN. The IP Extension Gateways are accessible through the LAN side GE interfaces. GE interfaces default to the WAN side, designated as FCIP in a `switchshow`. Incoming LAN traffic is not recognized on a WAN side GE interface and is dropped.

All Brocade Extension platforms have speed-settable GE interfaces. GE speeds are not auto-negotiated; extension GE interfaces operate at the configured speed and with the optic supporting that speed. GE optics might or might not be able to operate at more than one speed, for example, 1/10GE and 10/25GE optics. See Table 7-18.

Table 7-18   Supported LAN Side GE interfaces

Extension Platform	Platform's GE Interfaces	LAN side supported?	Interface Speeds
IBM SAN18B-R	GE0-GE1	Yes	1GbE
	GE2-GE7	Yes	1GbE, 10GbE
IBM SAN42B-R7	GE0-GE15	Yes	1GbE, 10GbE, 25GbE
	GE16-GE17	No	100GbE
IBM SX6 Extension Blade	GE0-GE1	No	40GbE
	GE2-GE17	Yes	1GbE, 10GbE

A GE interface's configuration must be deleted prior to changing its WAN or LAN mode. Once configured as a LAN side interface, it cannot be used as a WAN side interface for circuits.

► The IBM SAN42B-R7 has sixteen 1/10/25GE interfaces in which up to eight can be configured as LAN side.

► The IBM SX6 Extension Blade has sixteen 1/10GE interfaces in which up to eight can be configured as LAN side.

► The IBM SAN18B-R has six GE interfaces available (6 x 1/10GE optical ports or 4x 1/10GE optical ports plus 2x 1GE RJ-45 copper ports), in which up to four interfaces can be configured as LAN side. The number of available interfaces depends on whether the upgrade license has been installed.

Consider the following when you configure GE interfaces:

► Determine which GE interfaces will be used for the LAN side connectivity.

► Ensure that these GE interfaces are configured in LAN mode.

► Ensure that these GE interfaces are in the Default LS.

Configuring LAN side GE interfaces for IP Extension:

1. Connect to the switch and log on as admin.

2. Configure the necessary GE interfaces for LAN mode using the command:

portcfgge <ge#> --set -lan

3. Use the command, as shown in Example 7-20, for more information.

Example 7-20   portcfgge command

```
portcfgge
switch:admin> portcfgge --help
Usage: portcfgge [<slot>/][<port>] [<args>]
Port Format:
  ge#
Args:
  --enable -autoneg      - Enable the auto-negotiate mode of the 1 GE ports.
  --disable -autoneg     - Disable the auto-negotiate mode of the 1 GE ports.
  --set -speed <speed>   - Set the port speed.
                           Allowable speeds are [1G|10G|25G]
  --set -lan             - Set the port as lan.
  --set -wan             - Set the port as wan.
  --set -channel <channel> - Set the port tunable SFP channel ID of the 10 GE
ports only.
                           Allowable channel ID Range [1] - [102]
  --show                 - Show the current GE port configurations.
  --show -lmac           - Show the Local MAC addresses for GE/LAN ports.
```

4. Use **portcfgge <ge#> --set -lan** to configure a GE interface to be on the LAN side. The following example puts port GE7 in LAN mode.

    switch:admin> portcfgge ge7 --set -lan
    Operation Succeeded.

5. Use the **portcfgge --show** command to verify that the GE interface is in LAN mode, show the speed, and see if GE interfaces are part of a LAG. See Example 7-21.

Example 7-21   portcfgge --show command

```
switch:admin> portcfgge --show
Port Speed Flags Channel FEC Lag Name
--------------------------------------------------------------------------
ge0  1G    AC--  N/A     Off  -
ge1  1G    AC--  N/A     Off  -
```

```
ge2  10G   ----  N/A    Off  -
ge3  10G   ----  N/A    Off  -
ge4  1G    ----  N/A    Off  -
ge5  10G   ----  N/A    Off  -
ge6  10G   ----  N/A    Off  -
ge7  10G   -L    N/A    Off  -
----------------------------------------------------------------------------
Flags: A:Auto-Negotiation Enabled C:Copper Media Type
L:LAN Port G: LAG Member
```

The **switchshow** command can verify the GE interface speed and mode:

```
switch:admin> switchshow
switchName:    switch
switchType:    178.0
switchState:   Online
switchMode:    Native
switchRole:    Principal
switchDomain: 20
switchId:      fffc14
switchWwn:     10:00:88:94:71:60:99:7e
zoning:        ON (EXT-Testing)
switchBeacon: OFF
FC Router:    OFF
Fabric Name:  REPLICATION
HIF Mode:     OFF
Index Port Address Media Speed State      Proto
==================================================
0       0   140000  id    N32   No_Light   FC
1       1   140100  id    N32   No_Light   FC
2       2   140200  id    N32   No_Light   FC
3       3   140300  id    N32   No_Light   FC
4       4   140400  --    N32   No_Module  FC
5       5   140500  --    N32   No_Module  FC
6       6   140600  --    N32   No_Module  FC
7       7   140700  --    N32   No_Module  FC
8       8   140800  --    N32   No_Module  FC
9       9   140900  --    N32   No_Module  FC
10      10  140a00  --    N32   No_Module  FC
11      11  140b00  --    N32   No_Module  FC
12      12  140c00  --    --    Offline    VE
13      13  140d00  --    --    Offline    VE
14      14  140e00  --    --    Offline    VE
15      15  140f00  --    --    Offline    VE
        ge0         cu    1G    Offline    FCIP Copper Disabled (Unsupported blade
mode)
        ge1         cu    1G    Offline    FCIP Copper Disabled (Unsupported blade
mode)
        ge2         id    10G   No_Light   FCIP
        ge3         id    10G   No_Light   FCIP
        ge4         id    10G   No_Module  FCIP
        ge5         id    10G   No_Module  FCIP
        ge6         id    10G   No_Light   FCIP
        ge7         id    10G   No_Light   LAN
```

7.4.4 Portchannels (LAG)

Configuring a portchannel (LAG) for IP Extension is optional; however, you can only connect a single Ethernet link from the extension platform to the data center LAN without a LAG. A single link does not provide redundancy; a LAG provides redundant ports, cables, and optics. The recommended practice is to configure a LAG when connecting to a data center LAN.

LAN side IP Extension links cannot connect to more than one data center LAN switch unless the switches are logically one switch. For example, VCC, vPC, and MLAG are technologies that logically form a single switch. When connecting to multiple data center LAN switches that are logically a single switch, the LAN side links must be in a portchannel.

A LAG can be configured with or without 802.1Q VLAN tagging. When a link or LAG has VLAN tagging enabled, it is called a trunk and can carry multiple VLANs. A separate LAG is not required for each VLAN. A single LAG can accommodate multiple VLANs if it has VLAN tagging enabled; otherwise, only the data center switch's access VLAN can communicate across the LAG.

All links in a LAG must operate at the same speed. The IBM SAN42B-R7 and IBM SX6 Extension Blade support LAGs with up to four links. The maximum number of LAGs is two, each with two links.

The IBM SAN18B-6 has two 1 Gbps RJ-45 copper ports that support LAN side LAGs. A LAG can have a mix of copper and optical 1 Gbps interfaces.

A LAG is given a name when it is configured. GE interfaces are added to a LAG. The interface speeds (1, 10, 25 Gbps) and auto-negotiation (1GE only) settings must match all interfaces added to the LAG. Individual LAG interfaces can be enabled or disabled.

See Table 7-19.

Table 7-19   Maximum portchannels per platform

Extension Platform	Max IP Extension Gateways per DP	Max LAN interfaces per platform	Max LAGs per platform	Max links per LAG
IBM SAN18B-6	4	4	2	2
IBM SAN42B-R7	8	8	4	4
IBM SX6 Extension Blade	8	8	4	4

Note: An IP Extension LAN side LAG cannot span multiple IBM SX6 Extension Blades.

Note: A GE must be configured as LAN before adding it to a LAG. LAG does not enhance the ability of links to come online. Each link must be capable of coming online independently

Note: There are no WAN side LAGs. The WAN side uses BET to achieve enhanced functionality compared to LAG.

LAN side LAG considerations

The following considerations apply to LAN side LAG:

► LAGs are either static (manual configuration) or dynamic (uses LACP).

► IP Extension LAN side interfaces support jumbo frames. A jumbo frame can be up to 9216 bytes.

► IP Extension LAGs support VLAN tagging (802.1Q). Each VLAN has its own IP Extension Gateway, up to 8 gateways per DP.

► IP Extension LAGs do not support stacked tagging (QinQ).

► NIC teaming is not supported.

Perform the following steps to configure a LAG (portchannel):

1. Connect to the switch and log on as admin.

2. Use the command `lacp --help` to view the available options.

```
switch:admin> lacp --help
lacp cli usage:
lacp --help
lacp --show
lacp --config -sysprio <priority>
lacp --default
Sets all the configurations to default
```

3. For dynamic LAGs, set the global LACP priority from 0 through 65535; the default is 32768. Use the command:

`lacp --config -sysprio <priority>`

To view the setting, use the command:

`lacp --show`

Set the LACP system priority to 100:

`switch:admin> lacp --config --sysprio 100`
`switch:admin> lacp --show`
LACP system priority: 100
LACP System ID: 0x8000,00-05-33-74-85-42

4. Use the **`portchannel --help`** command to view the available options. See Example 7-22.

Example 7-22   portchannel --help

```
switch:admin> portchannel --help
portchannel cli usage:
portchannel --help
portchannel --create <po_name> -type <static|dynamic> [-key <po-number>]
portchannel --delete <po_name>
portchannel --enable <po_name>
portchannel --disable <po_name>
portchannel --add <po_name> -port <port-num|port-range> [-timeout <l/s> Dynamic
Port-Channel]
portchannel --remove <po_name> -port <port-num|port-range>
portchannel --show -detail | -static | -dynamic | -all | <po_name> | -stats
[<po_name>]
portchannel --config <po_name> -autoneg <on|off> [Not applicable on Empty
port-channel]
portchannel --config <po_name> -type <static|dynamic>
```

```
portchannel --config <po_name> -rename <new-po-name>
portchannel --config -port <port-num|port-range> -timeout <s|l> [Dynamic
port-channel]
portchannel --reset -stats [<po_name>]
Specifying port ranges:
  port_num -- <port_num> is <[slot]/port>
  port_range -- Specifies a set of ports as a range:
  (examples: '10/6' or '10/6-9' or '10/ge6-ge9' or '10/ge6-9')
Actions:
  --show : Show the configured portchannel details
           -static  : Show all configured static portchannels.
           -dynamic : Show all configured dynamic portchannels.
           -all     : Show all configured portchannels.
           -detail  : Show portchannels in detail.
           -stats   : Show all statistics on FCIP portchannels.
  --create : Creates portchannel.
             Portchannel name is a string of max 31 character
             Characters allowed: alphanumeric with special char _-
             -type : Specify the type of portchannel.
             -key  : Specify the Key of portchannel. key range is <1-1000>
  --delete : Delete the specified portchannel
  --config : Configures portchannel and member port-specific parameters
             -type     : Change type from dynamic to static and vice versa.
             -rename   : Change name of the portchannel.
             -autoneg : Configure Auto-neg on or off for portchannel of speed 1G.
             -port <port_num | port-range> : The port number or range of port
             -timeout : Configures member port timeout option.
              Possible options are S and L
  --enable  : Enable the specified portchannel
  --disable : Disable the specified portchannel
  --add     : Add member port to portchannel
             -port <port_num | port-range>
             The port number or range to be added
  --remove : Remove member port from portchannel
             -port <port_num | port-range>
             The port number or range to be removed
  --reset  : reset portchannel parameters
             -stats : reset sttistics for FCIP portchannels
```

5. Create either a static or dynamic LAG (portchannel).

switch:admin> portchannel --create MyDynLag -type dynamic -key 555
switch:admin> portchannel --create MyStaticLag -type static -key 100
switch:admin> portchannel --show
```
Name         Type     Oper-State  Port-Count Member Ports
---------------------------------------------------------
MyDynLag     Dynamic  Offline     0
MyStaticLag  Static   Offline     0
```

> **Note:** GE interfaces are not portchannel members after configuration. Add ports individually or by specifying a range

6. Add GE ports to the portchannel.

switch:admin> portchannel --add MyStaticLag -port ge16-17

```
WARNING: While making configuration changes the modified LAN GE ports will be
disabled. Please manually enable the modified LAN GE ports after completing all
the configuration changes.
switch:admin> portchannel --show -all
Name         Type     Oper-State Port-Count Member Ports
-----------------------------------------------------------------
MyDynLag     Dynamic  Online      1          ge6
MyStaticLag  Static   Offline     3          ge15 ,ge16 ,ge17
```

In static LAGs, LAN-side GE interfaces are disabled when added to the LAG and must be re-enabled after completing the LAG configuration. When adding an interface to a disabled dynamic LAG, the interface is disabled. When you remove an interface from a disabled dynamic LAG, the interface is enabled. If a disabled dynamic LAG is deleted using the `portchannel --delete` command, all member interfaces are enabled.

Set auto-negotiation for all 1 Gbps interfaces in the LAG using the **portchannel --config** command:

```
switch:admin> portchannel --config slag101 -autoneg on
```

> **Note:** GE interfaces are disabled after changing the auto-negotiation configuration. Fabric OS automatically re-enables the ports.

The following RASlog example shows that Fabric OS has reenabled the ports:

```
2023/02/16-22:37:18 (GMT), [ESM-2801], 94, FID 128 | PORT 0/GE7, INFO, Awing-2,
Port ge7 speed set to 1G, auto negotiation Enabled, FEC clause Off [API].
2023/02/16-22:37:19 (GMT), [PORT-1008], 95, FID 128, INFO, Awing-2, GigE Port
(0/GE7) has been enabled.
```

7. Enable or disable a LAG with the **portchannel --enable | --disable** command.

> **Note:** By default, the admin state of a portchannel is enabled.

```
switch:admin> portchannel --enable MyStaticLag
switch:admin> portchannel --show -all
Name         Type     Oper-State Port-Count Member Ports
-----------------------------------------------------------------
MyDynLag     Dynamic  Online      1          ge6
MyStaticLag  Static   Online      3          ge15 ,ge16 ,ge17
```

8. View detailed information about a LAG using the following command. See Example 7-23.:

Example 7-23   portchannel --show -detail

```
switch:admin> portchannel --show -detail
Name  :test
Type  :Dynamic Key :1
Speed :1G
Admin-state: Enable Oper-state : Online
Admin Key: 0001 - Oper Key 0001
LACP System ID: 0x8000,c4-f5-7c-01-31-4a PART System ID: 0x0001,00-24-38-9b-03-00
Portchannel Member count = 2
Port Oper state Sync Timeout Auto-Negotiation
--------------------------------------------------------------------------
*ge5 Online 1 Long Disabled
ge6 Offline 0 Long Disabled
```

```
Name :static
Type :Static Key :2
Speed :1G
Admin-state: Enable Oper-state : Offline
Portchannel Member count = 2
Port Oper state Auto-Negotiation
-------------------------------------------------
ge0 offline Enabled
ge1 Offline Enabled
```

9. Reset the statistics with the **portchannel --reset** command. See Example 7-24.

Example 7-24   portchannel --reset

```
switch:admin> portchannel --reset
switch:admin> portchannel --reset -stats
switch:admin> portchannel --reset -stats [<po_name>]
switch:admin> portchannel --show -stats
Name: static_lag1 LAG Name
Speed: 10G LAG Speed
Admin-State: Enable LAG Admin State
Oper-State: Online LAG Operational Status
InPkts: 0 LAG Received Frames
InOctets: 0 LAG Received Octets
InUcastPkts: 0 LAG Received Unicast Frames
InMcastPkts: 0 LAG Received Multicast Frames
InBcastPkts: 0 LAG Received Broadcast Frames
InVlanPkts: 0 LAG Received VLAN Frames
InPausePkts: 0 LAG Received Pause Frames
InDiscards: 0 LAG Received Frames Discarded
InErrors: 0 LAG Received Frames with Errors
OutPkts: 0 LAG Transmitted Frames
OutOctets: 0 LAG Transmitted Octets
OutUcastPkts: 0 LAG Transmitted Unicast Frames
OutMcastPkts: 0 LAG Transmitted Multicast Frames
OutBcastPkts: 0 LAG Transmitted Broadcast Frames
OutVlanPkts: 0 LAG Transmitted VLAN Frames
OutPausePkts: 0 LAG Transmitted Pause Frames
OutDiscards: 0 LAG Transmitted Frames Discarded
OutErrors: 0 LAG Transmitted Frames with Errors
CRCErrors: 0 LAG CRC Errors
CarrierErrors: 0 LAG lost carrier sense
JabberErrors: 0 LAG Jabbers
LAG Uptime: 3 Days 8 Hours 43 Mins 27 Seconds
```

7.4.5 Layer 2 geployment

Layer 2 communicates at the Ethernet level. End devices on the same VLAN with IP addresses on the same subnet as the IP Extension Gateway communicate at the Ethernet level. Alternatively, IP storage end devices can directly connect to the IP Extension LAN ports; however, a data center LAN switch scales the number of available Ethernet ports for end devices.

Adding an end device static route, the end device can differentiate the gateway to which it sends traffic based on the destination IP address or subnet. The assumption is that the end

device can add static routes. The following Figure 7-4 shows an IP Extension deployment using direct connections, a Layer 2 method. End device direct connectivity typically requires its replication ports to be dedicated to that traffic. The IBM IP Extension platforms use the TCL to determine which tunnel (location) to send the traffic to. See Figure 7-4.
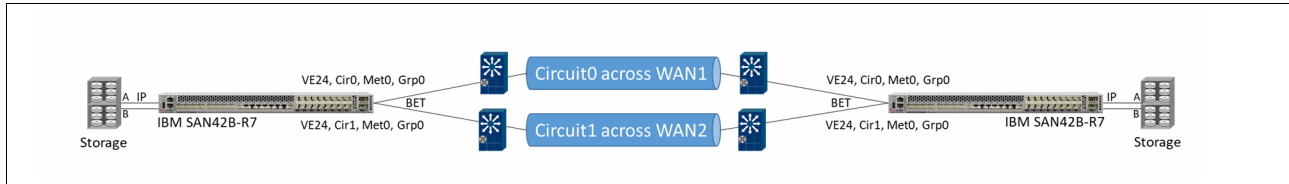
Figure 7-4   IP Storage direct connectivity to IP Extension

In the case of L2 deployment, configure the IP storage with the IP Extension Gateway as the next-hop for replication traffic. The IP Extension Gateway is configured on a DP (DP0 or DP1) owning the VE_Port of the tunnel. The IP Extension Gateways are a LAN side IPIF specific to a DP. The same IP Extension Gateway cannot be configured on more than one DP on an extension platform. An IP Extension Gateway configured on multiple platforms will cause an IP address conflict on the data center network. Based on the next-hop configured on the storage end device, the end device learns the IP Extension Gateway's MAC address through an ARP or Neighbor Discovery (ND) request. See Figure 7-5,
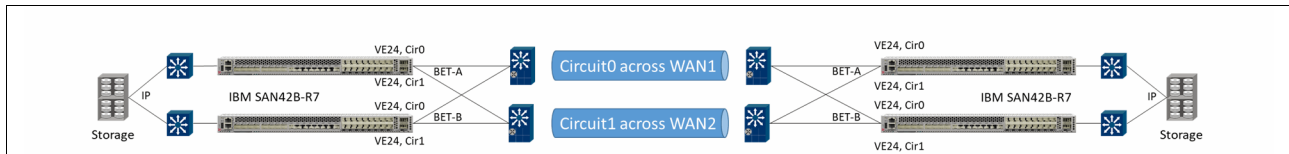
Figure 7-5   : IP Storage LAN connectivity to IP Extension

When an IP storage end device is connected to a Layer 2 LAN switch, as shown in the above figure, you can use Link Aggregation 802.3ad (LAG) to connect IP Extension LAN interfaces to the switch. The IBM extension platforms support static and dynamic LAG to data center switches. Dynamic LAG uses LACP. The LAG can be 802.1Q tagged if end devices live on multiple VLANs. Tagging enables multiple VLANs to communicate across the same physical LAG. The default is no tagging.

> **Note:** Only one path between the Layer 2 LAN and the IP Extension LAN side interfaces can exist. A LAG is considered a single path.

> **Note:** NIC teaming is not supported.

As shown in Figure 7-5 above, the IP storage is connected to the IP Extension LAN side interfaces by a Layer 2 LAN switch. At least one IP Extension Gateway (LAN side IPIF) must be configured to receive the replication traffic on the LAN network. For a Layer 2 architecture, each connected VLAN requires a LAN-side IPIF used as an IP Extension gateway. The LAN side IPIF should have an IP address on the same subnet as the end device and the same VLAN ID as the connected Ethernet network.

7.4.6 Layer 3 deployment

In the case of Layer 3 deployment, storage end devices do not require special routes and usually send traffic to the router gateway instead of sending replication traffic to an IP

Extension gateway. Some storage end devices cannot send replication traffic to an alternate gateway (no ability to add static routes). Changes are made on the router in the IP network. The router is configured to intercept specific flows and forward them to an IP Extension gateway. Policy Based Routing (PBR) is the most common method of performing this task in an IP network.

Figure 7-6 below shows IP Extension used in a routed network configured with PBR. The image is the same as in Figure 5 above; the difference is where the routing occurs, on the end device (static routes) or the router (PBR). The IP network intercepts traffic matched by an access control list (ACL) and redirects the traffic to an IP Extension gateway. A Layer 3 deployment simplifies IP storage over IP Extension and scales the number of subnets available compared to a Layer 2 deployment.
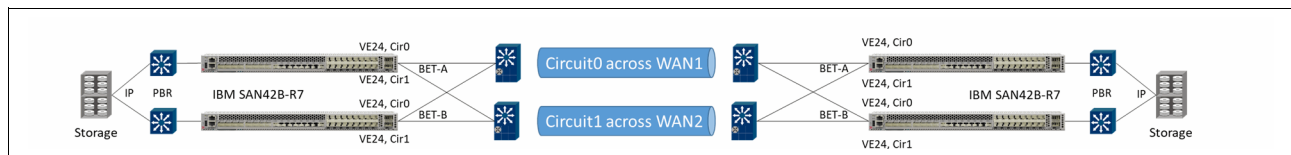
Figure 7-6   IP Extension Layer 3 deployment architecture

In a Layer 3 deployment, the IP storage end devices can be located on different subnets from the IP Extension gateway. The IP Extension gateway only requires a point-to-point network with the IP router, relying on the IP router to communicate with the end devices.

The best practice is to use a specific subnet on the end devices for IP Extension replication, which significantly eases PBR configuration. Nevertheless, IP storage end device replication ports may still exist on various subnets. A LAN-side IP route is added to IP Extension for each LAN-side destination subnet. IP Extension relies on the next-hop local IP router to deliver traffic to the end devices.

7.4.7 IP Extension Gateway

A LAN-side IPIF is a gateway for IP storage devices replicating data across IP Extension. Each DP has its own set of IP Extension gateways, one for each LAN side subnet IP storage end devices are using. On the LAN side, IPv4 and IPv6 are supported with IP Extension.

An IP Extension gateway address cannot be used on more than one DP or duplicated anywhere else in the network.

LAN-side GE interfaces can access all IP Extension gateways. All untagged (no 802.1Q) Ethernet traffic can access any IP Extension gateway on either DP that does not have an IPIF VLAN ID configured. If the LAN side traffic is tagged (uses 802.1Q), then access depends on matching the incoming VLAN tag with the same IPIF VLAN ID of the IP Extension gateway.

The IPIF for a lan.dp# interface forms an IP Extension gateway that IP storage end devices use to access IP Extension. IP storage traffic is sent to the IP Extension gateways. The IP Extension gateways are used for layer 2 and layer 3 deployments.

In a Layer 2 implementation, the IP Extension gateway is the next-hop address for the storage end device's replication traffic. A layer 2 deployment requires IP routes on the end device to direct the replication traffic to the IP Extension gateway. Changes must be made to the end devices. For each end device, an IP route must be added to forward replication traffic to the IP Extension gateway for each remote subnet replication is using. For each subnet, an IP Extension gateway must be configured. These end device IP routes are required before the end device can forward traffic to the IP Extension gateway.

In a layer 3 implementation, the end device's default route to the data center's gateway is its next-hop address; therefore, no changes should be made to end devices. However, it is necessary to modify the data center's router so that IP storage replication flows destined for remote sites are intercepted and forwarded to the IP Extension gateway.

End device replication ports can be on the same subnet or different subnets within the same site; however, networks separated by IP Extension cannot be on the same subnet; they must be on different subnets.

The following considerations apply to a LAN-side IPIF (IP Extension gateway):

► An LAN-side IPIF (lan.dp#) is referred to as an IP Extension gateway.

► The same IP Extension gateway IP address cannot be used on another DP.

► Identify the end device replication ports that will be extended through IP Extension.

► Identify the subnets that the end-devices replication ports are using:

   a. With a Layer 2 deployment, create one IP Extension gateway (IPIF lan.dp#) from an IP address on each subnet.

   b. Identify which VLANs are used.

► LAN side IPIF MTU range is 1280 to 9216 bytes.

► PMTU is not supported on LAN side IP Extension gateway interfaces (IPIF lan.dp#).

► IPIF IP addresses and subnet masks cannot be modified; the IPIF has to be deleted and recreated.

► Multiple same subnet IP Extension gateways cannot be created on the same DP.

► Each unique VLAN ID on an 802.1Q tagged LAG needs its own LAN side IPIF (lan.dp#).

An IBM SAN42B-R7 and IBM SX6 Extension Blade can have a maximum of eight IP Extension gateways per DP, and the IBM SAN18B-R can have a maximum of four IP Extension Gateways.

All LAN side IPIF form one Software Virtual Interface (SVI) per DP, which means there is only one LAN side MAC per DP, which is why only a single link to the data center LAN is supported.

In the following situations, distinct LAN gateways must be configured:

► If the data center LAN switch's LAG is not configured to use VLAN tagging (802.1Q), do not associate a VLAN ID with the LAN side IPIF (lan.dp#). Ethernet links will not come online when one side is <u>not</u> tagging and the other is tagging (mismatch). Tagged traffic can only talk to interfaces configured for the tagged traffic.

► To accommodate multiple VLANs, the LAG must be a trunk, which means the traffic is tagged. In a trunk, multiple logical ISLs pass through the same physical LAG, and the traffic of each VLAN is tagged. For tagged traffic to be forwarded to the correct IP Extension gateway, the LAN side IPIF must be configured with the corresponding VLAN ID. An IP Extension gateway is created for each expected VLAN if the LAG is tagged.

Perform the following steps to configure an IP Extension gateway:

1. Connect to the switch and log on as admin.

2. Use the command to create a LAN-side IP Extension gateway:

   `portcfg ipif <lan.dp#> create <gatewayIPaddress>/<maskLength>`

3. The VLAN and MTU arguments are optional. The VLAN default is none, which means there is no tagging. The MTU default is 1500 bytes. The example shows how to enter the

IP address and subnet using CIDR IP address notation (/subnet mask length) or the netmask argument.

The IP Extension gateway is local to the platform and registered with DP0, which is essential when associating the gateway with a tunnel (VE_Port) - both must be on the same DP.

4. Using the CIDR notation (/24):

```
switch:admin> portcfg ipif lan.dp0 create 10.0.0.4/24
```

5. Using the netmask notation:

```
switch:admin> portcfg ipif lan.dp0 create 10.0.0.4 netmask 255.255.255.0
```

6. Delete an IPIF:

```
switch:admin> portcfg ipif lan.dp0 delete 10.0.0.4/24
```

While an IP Extension gateway IPIF is in use, it can be deleted. Clean-up enforcement checks are not done on a LAN side IPIF.

> **Note:** If a LAN-side IPIF is accidentally deleted, all IP Extension flows that use that IP Extension gateway will be disrupted.

7. Use the command below to display IPIF interfaces.

```
portshow ipif
```

The following example shows two LAN side IP Extension gateways that are configured on slot 7 DP0 and two WAN side circuit IPIFs configured on slot 7 DP0:

▶ 10.0.0.1 on VLAN 100 with MTU 1500

▶ 10.0.1.1 on VLAN 200 with MTU 9216

▶ 192.168.60.20 on GE4

▶ 192.168.10.107 on GE17

```
switch:admin> portshow ipif
Port        IP Address / Pfx    MTU   VLAN  Flags
-------------------------------------------------------
7/ge4.dp0   192.168.60.20 / 24  1500  0     U R M
7/ge17.dp0  192.168.10.107 / 24 1500  0     U R M
7/lan.dp0   10.0.0.1 / 24       1500  100   U R M
7/lan.dp0   10.0.1.1 / 24       9216  200   U R M
-------------------------------------------------------
Flags: U=Up B=Broadcast D=Debug L=Loopback P=Point2Point R=Running I=InUse
N=NoArp PR=Promisc M=Multicast S=StaticArp LU=LinkUp X=Crossport
```

7.4.8 IP routes (LAN side)

IP Extension supports Layer 2 and Layer 3 deployments. LAN-side IP routes are only used with layer 3 deployments. Layer 3 deployments use IP Extension to deliver traffic to a destination router for delivery to the end device's subnet. A LAN-side IP route in a layer 3 deployment is used to forward arriving traffic to the next-hop router in the destination data center. With layer 2 deployments, the end devices are on the same subnet as the IP Extension gateway, and there is no need for an intermediate router. LAN-side IP routes do not forward traffic to the WAN side. IP routes define a destination subnet where an end device resides and the gateway address that can forward the data to that subnet.

In a Layer 3 deployment, when traffic arrives at the local router from an IP storage end device, it is intercepted by the router and forwarded to the IP Extension gateway. Typically, an access control list (ACL) and policy-based routing (PBR) match the incoming traffic and forward it to the IP Extension gateway. The router must be configured to perform this function. Storage end devices do not require a particular configuration with Layer 3.

Configuration steps for PBR are done on the router. An ACL and PBR policy determines precisely what traffic is sent to the IP Extension gateway. The router is configured with the IP Extension gateways to forward the traffic.

A LAN-side IP route to the next-hop gateway is configured on the extension platform. The IP Extension Gateway and the next-hop gateway are on the same subnet. The steps for configuring a LAN-side IP route are similar to configuring a WAN-side IP route. PBR forwarded traffic from an end device toward the tunnel does not use the IP Extension LAN side IP route. The IP Extension LAN side IP route forwards traffic from the tunnel toward the network's IP router. The network's IP router forwards the traffic to the end device.

Figure 7-7 shows an example network topology where IBM b-type IP Extension is the next-hop gateway for traffic forwarded by PBR from the DC router. Traffic not diverted by PBR to IBM IP Extension is routed based on the traditional routing table.
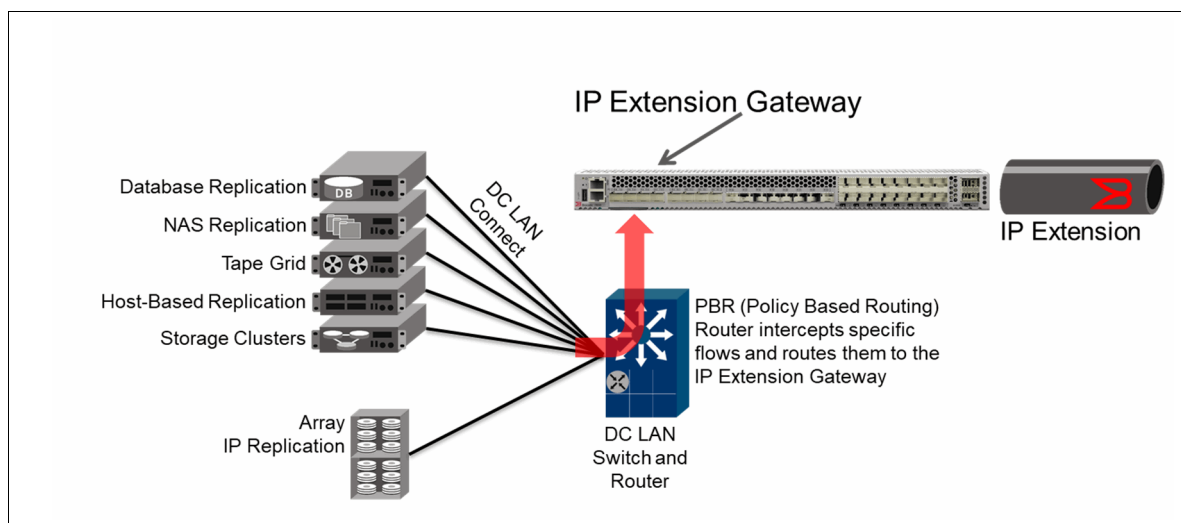
Figure 7-7   Network PBR, Interception, and Diversion to IP Extension Gateway

PBR is implemented in a network router. It is possible to have a Layer 3 deployment on one end and a Layer 2 deployment on the opposite end. The router and IP storage configurations are generic in the below configuration steps. For detailed steps, refer to your network equipment documentation.

Configuring LAN side IP routes for Layer 3 IP Extension deployment:

1. Connect to the switch and log on as admin.

2. Use the `portcfg ipif lan.dp#` command to configure a LAN side IPIF on each end. The portcfg ipif command creates a LAN side IPIF (IP Extension Gateway) on DP0. The following example shows the netmask keyword; however, the CIDR nomenclature without the keyword will also work (/29):

```
admin:switch_siteA> portcfg ipif lan.dp0 create 172.16.1.4 netmask 255.255.255.248
admin:switch_siteB> portcfg ipif lan.dp0 create 172.16.2.4 netmask 255.255.255.248
```

3. Use the `portcfg iproute lan.dp#` command to specify the next-hop gateway address. Perform this action on each end.

```
admin:switch_siteA> portcfg iproute lan.dp0 create 10.1.0.0/24 172.16.1.1
admin:switch_siteB> portcfg iproute lan.dp0 create 10.2.0.0/24 172.16.2.1
```

These commands configure a LAN-side IP route with the destination subnet of the local storage end device.

The IP Extension gateway and the data center's gateway connect via a LAG, and both gateways use a point-to-point subnet to communicate.

4. On the routers in data centers A and B, configure the portchannel (LAG) and router interface that is connected to the IBM extension platforms. Usually, there are two routers, and each has a gateway interface. Virtual IP addresses (VIPs) are created between them using protocols such as VRRP and HSRP. Using VRRP or HSRP protocols allows one router to be active and the other for failover. For example, router-1 has 172.16.1.2, router-2 has 172.16.1.3, and the shared VIP is 172.16.1.1; therefore, for this subnet, the IP Extension GW is 172.16.1.4 as shown below.

DC Router Site A:

IP: 10.1.0.1 mask 255.255.255.0

DC Router Site B:

IP: 10.2.0.1 mask 255.255.255.0

> **Note:** For detailed instructions, refer to the documentation provided with your router equipment.

The following steps provide generic information:

5. Configure PBR on network routers in data center A and B to redirect IP traffic destined for remote end devices. Designate the IP Extension gateway as the next-hop gateway.

DC Router Site A:

► Destination subnet: 10.2.0.0/24
► Source subnet: 10.1.0.0/24
► Next-Hop Gateway: 172.16.1.4

DC Router Site B:

► Destination subnet: 10.1.0.0/24
► Source subnet: 10.2.0.0/24
► Next-Hop Gateway: 172.16.2.4

6. Configure the replication ports with IP addresses from the network subnet used for IP Extension (in this example, 10.1.0.0/24) on the IP storage end devices in data centers A and B. Configure non-replication ports with IP addresses typically used in the data center for non-replication traffic (in this example, 192.168.1.0/24).

7. All IP addresses are examples.

Site A: IP storage replication port:

► IP: 10.1.0.10 mask 255.255.255.0

Site A: IP storage non-replication port:

► IP: 192.168.1.10 mask 255.255.255.0

Site B: IP storage replication port:

► IP: 10.2.0.10 mask 255.255.255.0

Site B: IP storage non-replication port:

► IP: 192.168.2.10 mask 255.255.255.0

8. Typically, this step is not required since the default route of the IP storage end device already sends traffic to the gateway of the data center's router.

Site A: IP Storage Default Route

► Subnet: 10.1.0.0

► Mask: 255.255.255.0

► Gateway 10.1.0.1

Site B: IP Storage Default Route

► Subnet: 10.2.0.0

► Mask: 255.255.255.0

► Gateway 10.2.0.1

7.4.9 TCL (Traffic Control List)

Each TCL rule is identified by a name that is assigned when created. TCL rules are modified or deleted by name. A TCL name contains 31 or fewer alphanumeric characters. Each name must be unique within the Extension platform. Rules are local to each platform and are not communicated across platforms.

When traffic matches an allow rule, the rule specifies which tunnel and QoS priority to assign that traffic. Medium is the default QoS priority. The default QoS marking is none. Unless a specific DP is configured, a rule denying traffic in a chassis system with multiple IBM SX6 Extension Blades applies to all DPs across all blades.

TCL rules have an order of precedence, each with a unique priority integer less than 65535. Smaller numbers are a higher priority (evaluated first), and larger numbers are a lower priority (evaluated last). 65535 is the lowest priority (highest number); therefore, it is the last rule enforced. A TCL priority number can be modified to reposition the rule within the list. Leave gaps between numbers for flexibility; for example, count by 100s.

Note: The TCL priority number must be unique across all active TCL rules within an IP Extension platform. For example, if a director chassis has multiple IBM SX6 Extension Blades, the priority numbers must be unique across all blades. If a TCL is defined as priority 10, that same priority cannot be used for another TCL rule on a different blade in the same chassis. A check is performed when the rule is enabled to ensure the priority value is unique.

Note: The IBM SAN42B-R7 and IBM SX6 Extension Blade provide 1024 defined TCL rules with up to 128 active rules. The IBM SAN18B-6 provides 256 defined TCL rules with up to 32 active rules.

> **Note:** At least one TCL rule must be configured. One default rule exists to deny all traffic; therefore, all traffic is denied if no other rule is created. The default rule cannot be removed or modified. One or more TCL rules must be configured to match traffic to the tunnel to allow traffic through an IP Extension tunnel.

A TCL rule has one of two actions, allow or deny. The allow rule specifies the target tunnel into which matching traffic should be directed. Upon the first match, further TCL processing terminates.

TCL considerations

The following considerations apply to TCL:

► Traffic matching an "allow" rule is sent to the rule's specified target.

► Target is the VE_Port number of the tunnel.

► Target must be an IP Extension-enabled tunnel.

► The IP Extension gateway (ipif lan.dp#) must specify the target's DP.

► Allow rules are automatically activated on the target VE_Port's DP.

► Traffic matching a "deny" rule is dropped.

► A target is not specified in a deny rule.

► Deny rules are activated on all DPs in the chassis.

► Directly connected end device ports; all traffic is intended for the remote data center.

► IP routes on end devices are managed such that only the desired traffic is allowed to use the IP Extension LAN gateway. All other traffic uses the default gateway.

No target can be specified in a deny rule because matching traffic will be dropped. Deny rules are pushed to all DPs in the chassis, accounting for traffic that might arrive at a DP. Traffic not destined for a particular IP Extension Gateway (LAN side IPIF unicast address) cannot be forwarded to a specific DP. For example, Broadcast, Unknown unicast, and Multicast (BUM traffic) traffic have no specific destination; therefore, a specific IP Extension Gateway is undeterminable. LAN side interfaces cannot determine which IPIF to forward this traffic to, so all DPs receive the deny rules, and all DPs must be capable of handling such traffic.

If the deny rule has clear matching criteria for processing specific flows, the deny rule can be assigned to a specific DP. The rule is pushed to the specified DP, denying any matching traffic on that DP only.

After traffic is matched to an allow rule, it is sent to the tunnel and further TCL processing stops. The target parameter specifies the tunnel to which the matched traffic will be routed.

Optionally, you can specify a traffic priority (high, medium, low). For example, when configuring the TCL rule, specify `--target 24` or `--target 24-high`. The traffic is sent over the IP Extension medium-priority if no priority is specified. IP Extension traffic priorities are egress scheduled separately from FCIP. Each protocol (FCIP and IP Extension) has QoS priorities in the egress scheduler.

The TCL is evaluated once when an end device's TCP stream performs the initial handshake. If TCL rules change, including priority numbers, enabled/disabled, QoS, terminated/non-terminated, etc., those changes do not affect existing streams. There is no effect on an existing stream because it has already completed its TCP handshake, and the TCL is not referenced again. If you do not see a change after changing a rule, it might be

because the traffic streams already exist. Newly established streams that match changed rules will demonstrate the change.

Before TCL changes can take effect, the VE_Port (tunnel) must be disabled and re-enabled. Disabling a VE_Port disrupts all traffic passing through the tunnel (FCIP and IP Extension). All IP Extension TCP sessions will be reset.

All traffic that appears at the IP Extension LAN interface is intended to pass through the tunnel. The simplest solution in these situations is to configure an allow rule that passes the source and destination subnets of the replication ports.

The best practice approach is configuring the fewest number of allow rules to pass the required traffic without creating an allow-all rule. Unmatched traffic encounters the default deny-all rule. You can isolate a subset of the allowed traffic by first using a more specific allow rule before a more general allow rule. For example, you can allow specific source and destination subnets with a specific protocol number followed by a rule with just the source and destination subnets.

When IP routes on the end device or PBR on the routers are configured correctly, no undesired traffic should be sent to the TCL. Undesired traffic is not matched and is dropped. Nonetheless, you can configure specific deny rules to ensure that certain devices or addresses cannot communicate across the IP Extension tunnel. When configuring and troubleshooting a complex rule set, there is an increased possibility of error.

> **Note:** When using Virtual Fabric Logical Switches (VF LS), IP Extension LAN-side GE interfaces must be in the Default LS. IP Extension LAN-side GE interfaces cannot be assigned to different logical switches. The lan.dp0 and lan.dp1 IPIFs behind the LAN-side GE interfaces are not part of VF LS contexts and cannot be assigned to an LS. When an IP datagram arrives at an IP Extension gateway, it is processed by the TCL.

LAN side GE interfaces must remain in the Default LS. Even though a VE_Port can reside within an LS other than the Default LS, there is no requirement for LAN-side GE interfaces to reside in that LS too. The TCL directs traffic to a specified VE_Port regardless of the LS in which it resides.

The target tunnel specified in a TCL allow rule must be native to the DP processing the TCL. IP storage traffic arrives on a deterministic DP based on the IP Extension gateway, a LAN IPIF (lan.dp#) on either lan.dp0 or lan.dp1.

If the TCL on the DP processing the incoming traffic matches traffic to a tunnel, the tunnel must be owned by that DP; otherwise, the TCL finds no valid match, and the traffic is dropped.

8

IBM b-type Extension architectures

This chapter discusses IBM b-type Extension architectures and has the following sections:

- ► "Fabric architecture" on page 132
- ► "FCIP architectures" on page 132
- ► "Extension with FCR" on page 135
- ► "IP Extension architectures" on page 136
- ► "Use case" on page 136
- ► "The Tunnel (The LAN side)" on page 136
- ► "IP Extension gateway" on page 137
- ► "GE interfaces (LAN side)" on page 138

8.1 Fabric architecture

Figure 8-1 shows a 1x1 architecture with one IBM SAN42B-R7 located at each site. There are two sites, and they are referred to as local and remote.

► Local refers to the location where data originates.

► Remote refers to the location where data is replicated to.

► Each extension platform has three distinct "sides": LAN, WAN, and FC, each side has unique configuration requirements.

► Each WAN link carries one circuit and represents a distinct path ("A" or "B").

► BET (Brocade Extension Trunking) refers to using multiple circuits (two in this case) that comprise a single tunnel and are represented by a VE_Port. The circuit members aggregate into the tunnel's overall bandwidth, take diverse paths for greater availability, and enable failover/failback, and the tunnel never loses or delivers out-of-order data. This means that the tunnel passes through both WAN links. Two or more circuits per tunnel are best practices and strongly recommended.
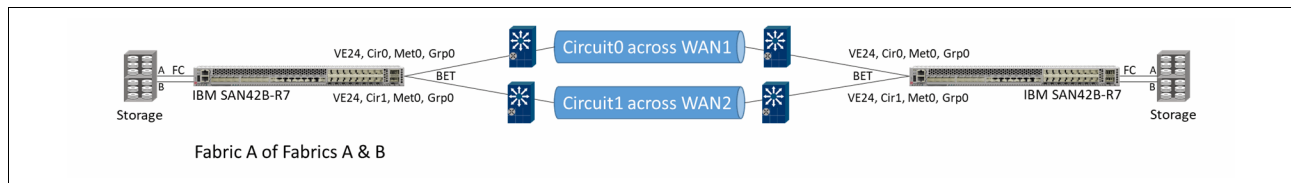
Figure 8-1   FCIP Replication Fabric Architecture (Fabric A of A & B)

> **Note:** You may have chosen a different extension environment than illustrated in this example. For example, you may have used two extension platforms at each location instead of one.

> **Note:** Only a single connection is allowed: a single physical connection or one logical connection made from multiple connections. Link Aggregation Group (LAG) uses multiple links for redundancy while maintaining a single logical connection. The reference architecture uses a LAG on the LAN side.

8.2 FCIP architectures

Data preservation permits an organization to recover, and extension is commonly used for business continuity via disaster recovery (DR). Preserve data by leveraging RDR and remote tape applications to transport critical data beyond a potentially catastrophic event.

RDR is typically 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 Fibre Channel if the backup site is within a metro distance, enough buffer-to-buffer credits exist, and there is fiber service between sites. However, cost-sensitive, ubiquitous IP infrastructure is often available, not dark fiber or native FC.

IBM b-type extension is ultra-high speed and ultra-low propagation delay per extension platform (four passes per round trip (0.3 ms). It is appropriate for both asynchronous RDR and synchronous RDR applications. A best-practice deployment connects array N_Ports

directly to extension F_Ports and does not connect through the production fabric. Frequently, replication has its own SAN because replication ports are dedicated, have no host traffic, and extension firmware updates can be done at unique intervals. Nevertheless, there remain valid reasons to connect via the production fabric, such as tape applications, and when there are more replication ports than the extension platform can accommodate.

8.2.1 Two-box FCIP solution

A single extension platform in each data center is referred to as a two-box solution, as shown in Figure 8-2 below. There is not extension platform redundancy in this architecture. A single extension platform can be directly connected to multiple storage replication ports.
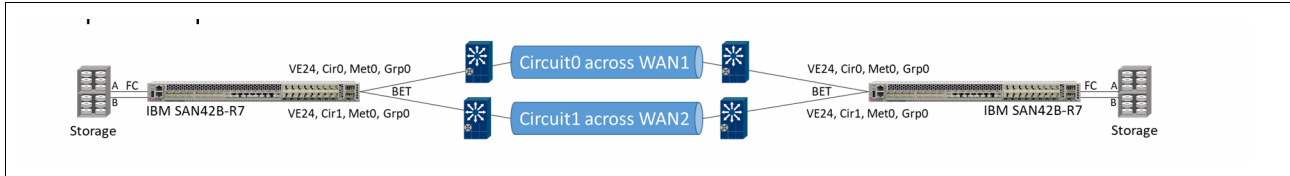
Figure 8-2   Non-redundant, basic extension architecture

8.2.2 Four-box FCIP solution

When the extension platform is dedicated to the A fabric and a physically different extension platform is dedicated to the B fabric, this is referred to as a four-box solution, as shown below in Figure 8-3 below. The four-box solution is most common and considered a best-practice implementation.

Each replication fabric has its own BET consisting of two circuits. Two WAN connections are being used; each has two circuits, one from each replication fabric. There are two WAN routers in each location. If a WAN connection or router goes offline, both fabrics continue to operate, albeit with diminished bandwidth. A single WAN link for both paths may be used, or different service providers may be used depending on the user's tolerance to disruption and cost.
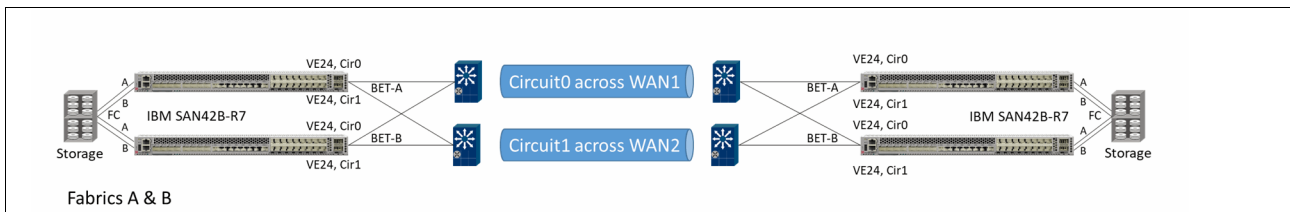
Figure 8-3   FCIP dedicated Extension Fabric for Each Controller

8.2.3 Four-box FCIP solution connected to production fabric

A production fabric can be extended, as shown in Figure 8-4 below, although this is only recommended when there are compelling reasons. A common reason is distributed tape or host-based replication in which many devices generate traffic to a DR site. In a tape scenario, OSTP may be used in a virtual fabric logical switch that isolates the tape connections and VE_Port.

Figure 8-4   Dual fabric FCIP architecture extending production fabric

In environments that require extending a production SAN, do not interconnect the same extension platform to both the 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 a SAN outage. Even momentary SAN outages can cause servers to stop responding, forcing a reboot and consistency check, which in most situations takes significant time.

Building a redundant SAN with A and B fabrics is the best practice for maximum availability, implying an air gap between the two autonomous fabrics from the server to the storage. There are no physical links between the two fabrics. Servers, storage, and VMs have drivers that monitor pathways and send data accordingly. When a path is detected as down, the driver fails over the traffic to a remaining path.

When using extension with FCR to connect production edge fabrics, do not connect both the A and B fabrics, as shown in Figure . Without FCR, both fabrics would merge into one big fabric destroying any notion of redundant autonomous fabrics. If FCR is used, the fabrics do not merge; however, both fabrics are attached to a common Linux platform. If maximum availability is a goal, this architecture is unacceptable, risky, and considered poor practice. A SAN with a common platform to A and B fabrics is susceptible to human, hardware, software, and environmental errors; each can bring down the entire SAN. See Figure 8-5.
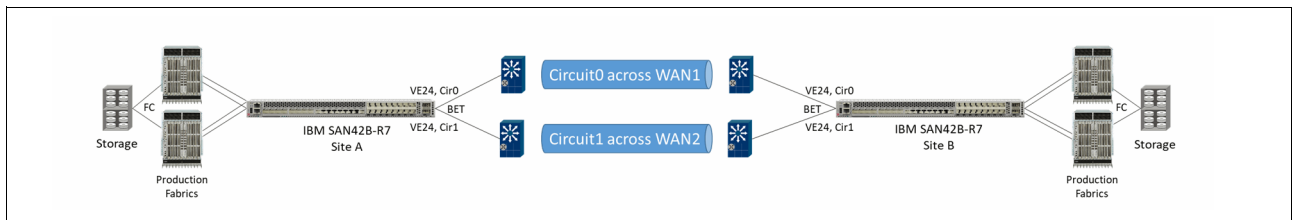
Figure 8-5   Poor practice: Two-box solution connecting both production fabrics - Not recommended

Using best-practice and traditional core-edge concepts when connecting extension to a production fabric increases reliability and manageability and facilitates troubleshooting. Since storage connects directly to the core in a core-edge design, standalone extension switches connect to the core, or an extension blade is placed in a core director. Connect extension platforms to a production fabric with at least two inter-switch links (ISL) for redundancy.

In a four-box solution, connecting an ISL between the "A" and "B" extension platforms is inappropriate for the same reasons discussed above, a common Linux instance, hardware failures, and human errors.

Cross-connecting circuits from a tunnel to various Ethernet data center 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.

8.3 Extension with FCR

An IBM b-type extension tunnel traverses a WAN, and a WAN has very different characteristics than a Fibre Channel ISL within a data center; however, an FCIP link is considered a Fibre Channel ISL. An extension tunnel traversing the WAN is essentially an ISL over a WAN. WAN flapping can disrupt fabrics.

Use FCR if a production SAN must be isolated from potential WAN instability. Disruption stems from fabric services that attempt to converge with each WAN flap. Convergence requires CPU processing, and excessive convergence may lead to excessive utilization. If the CPU can no longer process tasks promptly, instability ensues. Limiting fabric services to a local edge fabric and not permitting services to span the WAN prevents large-scale and taxing convergence. The best practice is to construct completely separate production and replication SANs. FCR is unnecessary when only replication ports are connected to a dedicated fabric or logical switch.

Refer to Figure below, isolated fabrics connected to FCR EX_Ports are called edge fabrics. EX_Ports are the demarcation points used to contain fabric services. Fabric services do not pass beyond an EX_Port, forming an edge. FCR constrains fabric services to within edge fabrics and edge fabrics are connected via a backbone fabric. Extension can make up a backbone fabric.

The following are basic architectures with and without FCR:

► The most straightforward architecture is no FCR. An independent replication SAN is the most common implementation for distributed systems and mainframes and is the easiest to manage. Directly connecting storage replication ports to extension is highly reliable and easy to implement, manage, and troubleshoot. See Figure 8-5 above.

► An edge-backbone-edge architecture with FCR in which edge fabrics bookend a transit backbone fabric; see Figure 8-6 below. The backbone fabric has EX_Ports that are outward-facing to the edge fabrics. Extension is located within the backbone fabric using VE_Ports to communicate across the WAN.
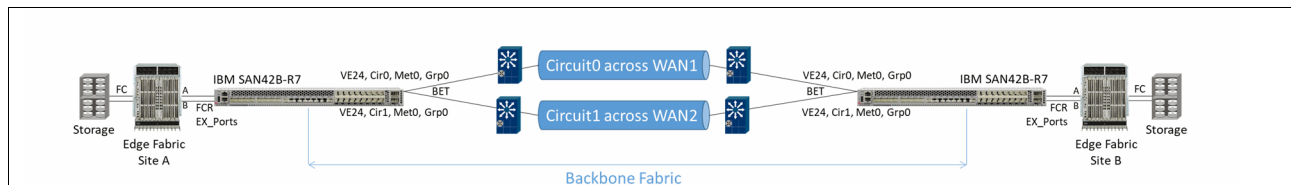
Figure 8-6   Edge-backbone-Edge extension with FCR

> **Note:** When a mainframe host writes FICON to a volume on a direct-access storage device (DASD and the DASD performs RDR to a remote DASD, the array-to-array replication does not use FICON; the array-to-array replication is FCP based.

> **Note:** Mainframe FICON environments do not support FCR.

8.4 IP Extension architectures

IBM b-type IP Extension technology accelerates, secures, and manages bandwidth for supported IP storage applications like IBM TS7700 Grid. FCIP and IP Extension share the same tunnel; the transport technology is the same. Latency and packet loss are enemies to TCP/IP replication and cause poor performance. IP Extension overcomes performance degradation caused by inherent characteristics of most WAN connections. Plus, IP Extension provides encryption without degrading host-device computation performance.

IBM b-type extension can be represented by having three sides, and one of those sides is the LAN side. The LAN side is specific to IP Extension and used to connect IP storage, typically via the data center LAN. IP Extension supports the connectivity of multiple data center LANs over a single Ethernet trunk using VLAN tagging (802.1Q). Using a tag, the Ethernet trunk between the extension platform and the LAN switch identifies each VLAN. An IP Extension gateway (`ipif lan.dp#`) must be configured for each VLAN.

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

8.5 Use case

IBM b-type IP Extension includes replication acceleration, data encryption, bandwidth management, data migration, network visibility, troubleshooting tools, high availability, congestion avoidance, and tape grids.

Figure 8-7 shows a high-availability and secure architecture used in a VM-NAS environment replicating traffic between a primary site and a cloud site. The cloud site provides compute elasticity and DR. The data is quickly replicated to maintain a coherent RPO.
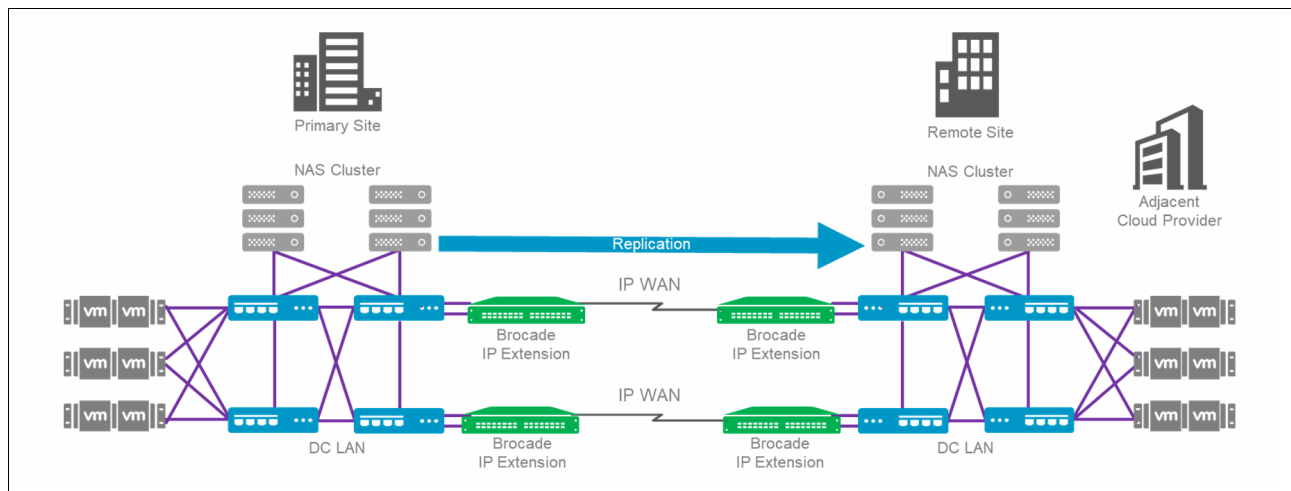
Figure 8-7   IP Storage over IP Extension

8.6 The Tunnel (The LAN side)

VE_Ports for IP Extension represent the tunnel ID. Unlike FCIP, IP Extension data does not flow through VE_Ports. Though, disabling a tunnel's VE_Port will disables it, even if only doing IP Extension.

Between the same domains, the best practice is to use one VE_Port for FCIP and IP Extension to manage the bandwidth of both protocols. Managing congestion and the bandwidth of both protocols cannot be done when different VE_Ports are used.

8.7 IP Extension gateway

IP Extension is a gateway for IP storage traffic crossing an extension tunnel. If IP storage traffic is not forwarded to the IP Extension gateway, it cannot benefit from IP Extension.

I Figure 8-8 below, traffic from the IP storage cluster comes into the data center LAN switch. The data center LAN switch forwards traffic to the traditional router gateway, which may be an inherent part of the LAN switch. The router sends the traffic toward the destination.

> **Note:** The LAN side subnet used for the IP Extension gateway must differ on the local and remote ends.
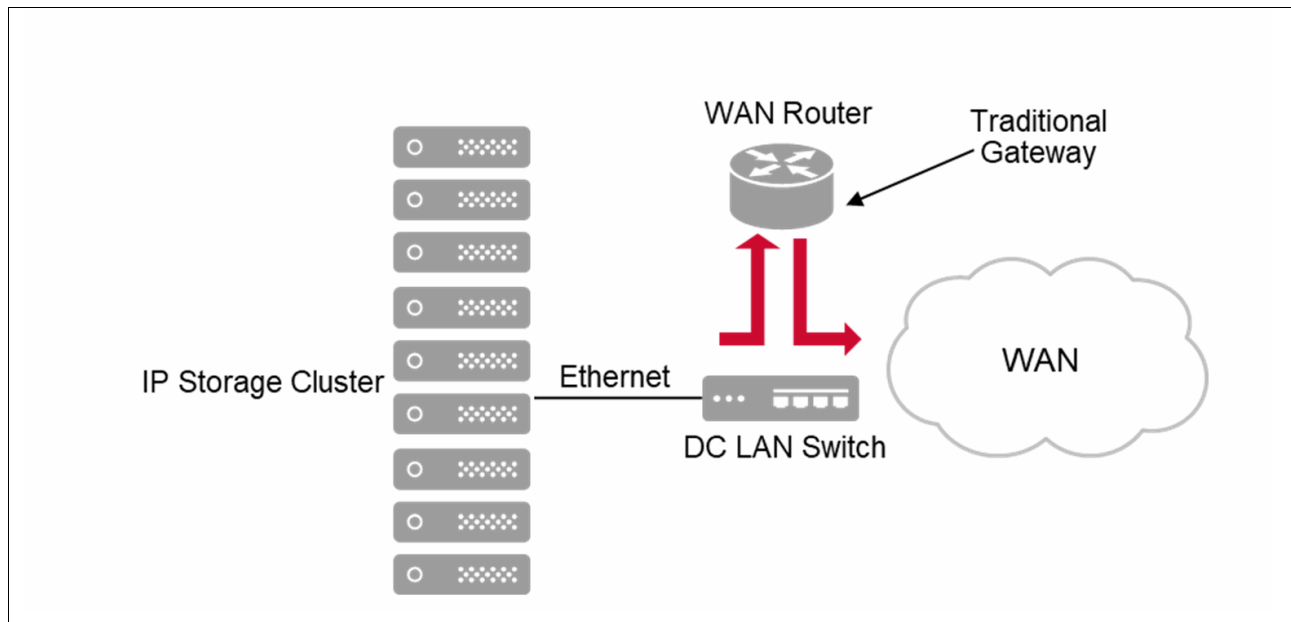
Figure 8-8   Traditional data center gateway

With IP Extension, it is required that the end device have a static route or some route that is more specific than the default route. The remote end device is on the remote subnet. The more specific static route forwards traffic destined for the remote subnet to the IP Extension gateway. The default route points to the traditional router gateway and is used for traffic not headed to the remote subnet.

Remember that putting IP Extension in the path and removing it involves activating or deactivating the end devices' static route that redirects the traffic to the IP Extension gateway. Traffic goes to the IP Extension gateway with static routes; IP Extension is in the path. The traffic goes to the traditional router without static routes; IP Extension is out of the path. Additionally, there is a method of intercepting traffic flows at the data center's router and diverting the traffic from there; in this case, no routes are needed on the end devices. The router intercept technology is called policy based routing (PBR) and is supported by IBM b-type extension.

Figure 8-9.below shows that replication traffic to remote IP storage is directed to the IP Extension gateway. The TCL evaluates the incoming TCP 3-way handshake. If the traffic matches an "allow" rule that session's traffic enters the specified tunnel, it is sent to the tunnel (target). The IP Extension traffic is now on the WAN side.
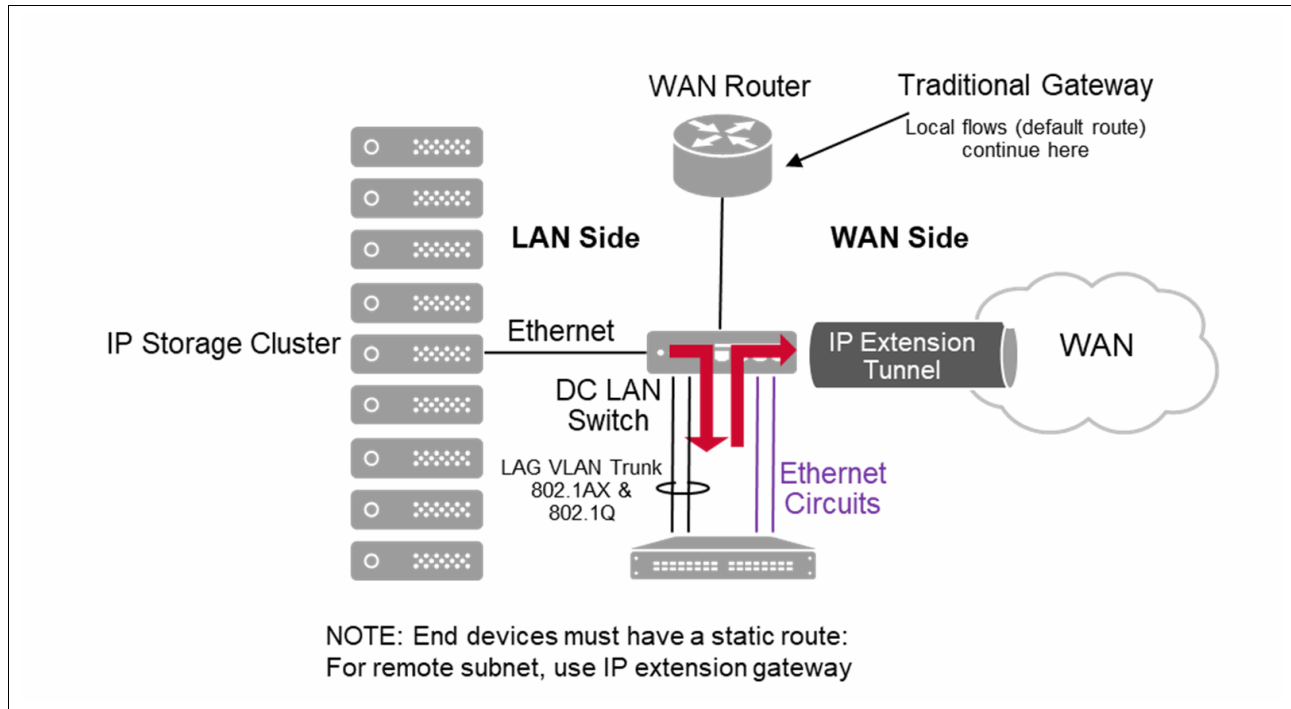
Figure 8-9   IP Extension gateway

8.8 GE interfaces (LAN side)

GE interfaces are either WAN (tunnel) facing or LAN (IP Extension) facing. An interface cannot do both and must be configured for one or the other. The default is WAN facing. LAN-side connectivity can be made from 1GbE, 10GbE, and 25GbE interfaces.

Specific to the IBM SX6 Extension blade, it must be in the hybrid (FCIP and IP Extension) application mode before a GE interface can be configured as LAN. The IBM SAN18B-R and IBM SAN45B-R7 have no app-mode setting; they only have hybrid mode. The maximum number of LAN side interfaces on the IBM SAN45B-R7 and IBM SX6 is 8 out of the 16 GE interfaces. On the IBM SAN18B-R, the maximum number of LAN side interfaces is 4 out of the 6 GE interfaces.

The IBM SAN18B-R has two copper (RJ-45) ports that operate only at 1 Gbps. There is no advantage or disadvantage to using the copper ports (GE0 and GE1) over the SFP ports (GE2 and GE3). Speed is the only copper port limitation.

9

IBM SANnav 2.3

SANnav is the next-generation SAN management application suite for Brocade SAN environments. SANnav allows you to efficiently manage your SAN infrastructure through various easy-to-use functions.

This chapter describes how to deploy IBM SANnav Management Suite 2.3 in your SAN infrastructure. It also includes detailed steps for installing the IBM SANnav Management Portal and SANnav Global View 2.3.

This chapter also describes all the system and server requirements before you start the SANnav Management suite installation. The system and server requirements are included for the deployment of SANnav Management Portal and SANnav Global View.

> **Note:** Refer to the Brocade SANnav Management Portal User Guide, 2.3.x for detailed information:
> https://techdocs.broadcom.com/us/en/fibre-channel-networking/sannav/management-portal/2-3-x.html.
>
> You can also refer to the IBM Redbooks publication *IBM SANnav Management Portal v2.2.X Implementation Guide*, SG24-8534.

This chapter includes the following topics:

9.1 Management Suite 2.3 release overview

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 workflows, providing a highly intuitive user experience.

SANnav uses a micro-services-based architecture that is based on Docker container technology. This architecture allows SANnav to scale to meet the management needs of both small and large SAN environments and those environments that may change over time. This scalable architecture also allows SANnav to support new functionality in the future without causing degradation to the performance of the application.

To address the management needs of large-scale SAN environments or those environments that are distributed by function or location, SANnav supports a hierarchical management model. In this model, a higher-level "global" application, SANnav Global View, provides comprehensive visibility, summarization, and seamless navigation across multiple instances of the SANnav Management Portal application.

There are two distinct SANnav product offerings:

► SANnav Management Portal
► SANnav Global View

9.1.1 Management Portal v2.3 release overview

IBM SANnav Management Portal v2.3.0 is a major software release introduced to support Fabric OS (FOS) v9.2.x and to provide major enhancements or new features. This chapter highlights the new features, support, capabilities, and changes in the SANnav Management Portal v2.3.0 release.

Note that this document applies only to the IBM SANnav Management Portal product. There is a separate Release Notes document for the Brocade SANnav Global View v2.3.0 release.

Within this document, SANnav Management Portal might also be referred to simply as "SANnav" or "SANnav MP".

9.1.2 Upgrade to SANnav v2.3 important considerations

Broadcom recommends customers stay on the latest SANnav target path version for the highest level of stability. Customers should consider upgrading to SANnav v2.3.0 in the following cases:

► Upgrade of SAN platforms to FOS v9.2.x, which requires SANnav v2.3.0 or later
► Benefit from new SANnav v2.3.0 features or enhancements.

9.2 IBM SANnav Management Portal

IBM SANnav Management Portal allows the management of one or more fabrics in the same or different geographical locations. It supports managing a maximum of 15,000 physical SAN ports.

For environments larger than 15,000 ports, you can deploy multiple SANnav Management Portal instances, which SANnav Global View can manage.

> **Note:** Up to two instances of the SANnav Management Portal may be used to monitor the same fabric.

Use the SANnav Management Portal to monitor and manage fabrics, switches, ports, and other SAN elements. Dashboards provide summary health status and performance information, from which you can drill down to get detailed views. You can sort and search your inventory using filters and tags to find exactly the information you want. A highly flexible reporting infrastructure allows you to generate custom graphical or tabular reports.

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

Refer to the *Brocade SANnav Management Portal User Guide* for detailed information:

https://techdocs.broadcom.com/management-portal

9.2.1 IBM 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 users to drill down into any instance and perform detailed monitoring, investigation, and troubleshooting based on events rolled into global view.

You log in to the SANnav Global View application and add portals to SANnav Global View, which then uses information in the portals to build a global view of the dashboard and inventory.

> **Note:** SANnav Global View is a separate product from SANnav Management Portal, and it requires different installation and licensing.

> **Note:** SANnav Global View is compatible only with SANnav Management Portal instances that are running the same version as Global View.

Refer to the *Brocade SANnav Global View User Guide* for detailed information:

https://techdocs.broadcom.com/us/en/fibre-channel-networking/sannav/global-view/2-3-x.html

9.2.2 What is new in SANnav Management Portal v2.3.0?

SANnav v2.3.0 provides new features and enhancements that simplify and automate common and frequent operations. The following new features or feature enhancements are provided in various functional areas of SANnav:

► Server platform deployment, installation, upgrade, and migration (including Disaster Recovery).

► Security and infrastructure: Provide security features and enhancements in all areas (SANnav server and managing Switch and FOS security).

► SANnav licensing.

► FOS Certificates Management.

► FOS Firmware platform specific download (PSD) management.

► Call Home.

► Discovery.

► Inventory: Simplify device ports to enclosure mapping using host and storage mapping policies.

► Zoning: Simplify daily zoning tasks with new or enhanced workflows such as Zone Database snapshots and zone policies.

► Configuration policy management: Accelerate the deployment of new switches, hosts, and targets with enhanced features.

► Flow management: Quickly identify issues with device ports with the new IO Health & Latency widget.

► Events and violations.

► UI/UX and usability changes: enhanced overall UI/UX usability features in Inventory, topology, dashboards and reporting

9.2.3 New hardware platforms supported in SANnav Management   Portal v2.3.0

The following are the new hardware platforms supported in SANnav Management Portal v2.3.0:

► Brocade/IBM SAN42B-R7

► Brocade FC 64-64 Blade

9.2.4 SANnav Management Portal v2.3.0 supported SAN switches

Starting with SANnav v2.3.0, support for various SAN hardware platforms and FOS versions will be reduced.

This affects support for SANnav customers as follows:

► An end user reports an issue on an unsupported hardware platform and/or unsupported FOS release.

► To receive support, the end user must reproduce the issue with a supported hardware platform and/or FOS version. Broadcom recommendation to end users is to upgrade to supported hardware platforms and/or FOS versions configurations before deploying SANnav MP v2.3.0.

9.2.5 New hardware platforms and FOS support matrix supported in SANnav 2.3

The officially supported matrix of supported hardware platforms and FOS versions is listed in

Note: that no Gen4 platform is officially supported with SANnav v2.3.0. SANnav will continue to recognize and discover/manage these no longer supported platforms, however, support may be limited in some cases.

Note: Switches running unsupported FOS versions (such as FOS v7.4.x or any FOS v8.x non target path releases) may be managed by SANnav, however, issues specific to those FOS firmware versions will not be addressed by Broadcom.

Switch Type	Hardware Model	FOS Version(s) Supported (*)
Gen 7 Switches	▪ Brocade G720 ▪ Brocade G730 ▪ Brocade X7-4 ▪ Brocade X7-8 ▪ Brocade 7850	▪ FOS v9.1.0b ▪ FOS v9.1.1b ▪ FOS v9.2.0
Gen 6 Switches	▪ Brocade G610 ▪ Brocade G620 ▪ Brocade G620 (switchType 183) ▪ Brocade G630 ▪ Brocade G630 (switchType 184) ▪ Brocade 7810 Extension Switch ▪ Brocade X6-4 ▪ Brocade X6-8 ▪ Brocade MXG610s Blade Server SAN I/O Module ▪ Brocade G648	▪ FOS v9.0.1e1 ▪ FOS v9.1.0b ▪ FOS v9.1.1b ▪ FOS v9.2.0
Gen 5 Switches	▪ Brocade 7840 Extension Switch ▪ Brocade DCX 8510-4 ▪ Brocade DCX 8510-8 ▪ Brocade 6505 ▪ Brocade 6510 ▪ Brocade 6520 ▪ Brocade M6505 Blade Server SAN I/O module ▪ Brocade 6542 Blade Server SAN I/O module ▪ Brocade 6543 Blade Server SAN I/O module ▪ Brocade 6547 Blade Server SAN I/O module ▪ Brocade 6548 Blade Server SAN I/O module ▪ Brocade 6558 Blade Server SAN I/O module	▪ FOS v8.2.3d

(*) Not all FOS versions listed in this column are supported on all hardware model platforms. Refer to the FOS and SANnav User Guide for details of which FOS version is supported by which platform.

(**) For new Gen7 hardware models (7850) <u>only</u> FOS v9.2.0 is supported.

Figure 9-1   The officially supported matrix of supported hardware platforms and FOS versions

9.2.6 Brocade SANnav Management Portal deployment

SANnav Management Portal v2.3.0 can be deployed either on a single bare-metal host, virtual machine (VM) or as an Open Virtual Appliance (OVA). The following two tables provide details of server requirements in each case.

Figure 9-2 on page 145 shows the bare-metal requirements.

Max Switch Ports Under Management (Base or Enterprise)	Operating System	Host Type	Minimum vCPU	Memory	Hard Disk
Small 600 Ports (Base) or, 3000 (Enterprise)	RHEL 8.4, 8.6 (*)	Bare metal/VMware ESX7.0 VM Bare metal/HyperV Windows Server 2022 VM	16 vCPUs	48 GB	600 GB
Large 15000 (Enterprise)	RHEL 8.4, 8.6 (*)	Bare metal/VMware ESXi 7.0 VM Bare metal/HyperV Windows Server 2019 VM	24 vCPUs	96 GB	1.2 TB

Figure 9-2   Server requirements - for physical environments

- ► (*) Other RHEL releases are not explicitly qualified or supported.
- ► Specifically, RHEL 8.2, 8.3, 8.5, 8.7 are not officially supported but installation and running SANnav on these versions is allowed upon user acceptance with conditional support.
- ► RHEL 8.0, 8.1 are not supported; the installation script exits if RHEL 8.0 or 8.1 is running on the SANnav host.
- ► RHEL 9.0 is not supported; the installation script exits if RHEL 9.0 is running on the SANnav host.
- ► The recommended CPU speed is 2000 MHz. Running SANnav with lower CPU speed may result in lower performance.
- ► The recommended number of physical CPU sockets is 2.

Figure 9-3 shows the hypervisor requirements.

Max Switch Ports Under Management (Base or Enterprise)	Supported Hypervisor	Host Type	Minimum vCPU	Memory	Hard Disk
Small 600 Ports (Base) or, 3000 (Enterprise)	VMware ESXi 7.0	VMware ESXi VM	16 vCPUs	48 GB	600 GB
Large 15000 (Enterprise)	VMware ESXi 7.0	VMware ESXi VM	24 vCPUs	96 GB	1.2 TB

Figure 9-3   Server requirements - for virtualized environments

- ► SANnav MP v2.2.1 OVA packages Rocky Linux 8.6 in the .ova file
- ► The OVA deployment allows user to select a small or large deployment configuration.
- ► The recommended CPU speed is 2000 MHz. Running SANnav with lower CPU speed may result in lower performance.
- ► The recommended number of physical CPU sockets is 2.

9.2.7 Browser requirements

The latest versions of the following web browsers are supported for a SANnav Management Portal v2.3.0 client:

- ► Chrome (Windows, Linux, MacOS)

> ► Firefox (Windows, Linux)
>
> ► Edge (Windows)

Note: Refer to *Web Tools User Guide and Release Notes* for supported list of browsers for Web Tools launch for all FOS versions (FOS v8.x and below – Java required, and FOS v9.x and above – no Java required). See https://docs.broadcom.com/doc/FOS-821-Web-Tools-UG.

9.2.8 SANnav 2.3 software upgrade

Refer to the "Upgrade and Migration Overview" section of the *Brocade SANnav Management Portal Installation and Migration Guide* for complete details (https://techdocs.broadcom.com/us/en/fibre-channel-networking/sannav/management-portal-installation-and-migration/2-2-x.html). See Figure 9-4.

Current Version	New Version	Supported?	Comments
SANnav 2.1.x and earlier	SANnav v2.3.0	No	Support only Major releases up to "N-1" upgrades. N=3 for SANnav v2.3.x
SANnav v2.2.0x	SANnav v2.3.0	No	SANnav v2.2.0 was BR GA 12/15/2021. Therefore, it is not a valid upgrade path to SANnav v2.3 (over 1 year) as per Brocade support policy.
SANnav 2.2.1x	SANnav v2.3.0	Yes	To upgrade from SANnav v2.2.1 to SANnav v2.3 in OVA deployment case *only*, it is required to perform an inline upgrade to SANnav v2.2.2 first, and then proceed with upgrade to SANnav 2.3.
SANnav 2.2.2x	SANnav v2.3.0	Yes	

Figure 9-4   Supported upgrade and migration paths to SANnav v2.3.0

> ► SANnav v2.2.2.x to SANnav v2.3.0 OVA upgrade and migration requires full extraction of the OVA and upgrade/migration due to disruptive OS change (CentOS 7.9, Rocky 8.6)
>
> ► Refer to SANnav Installation and Upgrade Guide for details before attempting SANnav MP upgrade in all deployments.

9.2.9 SANnav 2.3 licensing

Brocade SANnav Management Portal can be licensed in either a Base or Enterprise version. SANnav Management Portal Base enables management of up to 600 ports residing on fixed port switches or embedded blade switches, but it cannot be used to manage ports from any directors (4-slot or 8-slot).

SANnav Management Portal Enterprise enables management of up to 15,000 ports from any embedded switch, fixed port switch, or director class products. See Figure 9-6 on page 148.

Product Offerings	Description
SANnav Management Portal Base	Manages up to 600 ports from fixed-port or embedded switches but does not manage directors.
SANnav Management Portal Enterprise	Manages up to 15,000 switch ports from any type of switch including directors (either 4-slot or 8-slot).

Figure 9-5   Product offerings

SANnav Management Portal uses a subscription-based licensing model, which allows the product to function for the duration purchased. The SANnav Management Portal license must be renewed and installed in a timely manner to keep the product functioning without disruption.

Removal of trial period with SANnav v2.3.0

SANnav Management Portal v2.3.0 no longer provides a trial period built into the product, which allows the product to be used for a specific duration from the day of installation, without requiring a license.

New customers wanting to try out the SANnav software for a short duration may contact Brocade Support.

Customers wanting to trial the SANnav product may do so with previous versions of SANnav as follows:

► SANnav versions up to and including SANnav v2.2.0 have a 90-Day Trial period embedded.

► SANnav v2.2.1.x and v2.2.2.x have a 30-Day Trial period embedded.

Removal of 30-day grace period (available after license expiration)

The SANnav Management Portal license 30-day grace period (available after license expiration) is now removed in SANnav v2.3.0

With SANnav v2.3.0, when the license expires, the functionality will be restricted following the expiration date. A user will no longer be allowed to login to the server from the UI.

The SANnav server will continue to run and monitor the environment, but the UI will not be available (except for the ability to install a new license)

New license file expiration date

Beginning with SANnav v2.3.0, the SANnav license file (`license.xml` file) must be applied to the SANnav server within 30 days of creation of the SANnav license file.

This 30-day expiration is completely independent of the SANnav subscription expiration date. Refer to the SANnav Management Portal v2.3.0 User Guide, section Licensing for details on how to regenerate the SANnav license file (.xml file) and how to apply it to the SANnav server should this happen.

Export renewal request

With SANnav v2.3.0, a user may Export (download) any valid SANnav License information into a local client file. This helps customers and OEMs with ordering a SANnav license renewal for the current license.

► User may export the current Active, Active (Released) or Expired SANnav license details to renew the current license.

► The Export Renewal Request menu will show the following:

– Current License Expiration Date.

– Renewal License Start Date (one day after the current expiration)

– Renewal License End Date: By default this is set to one year after the Renewal Start Date, but the user can change it to any arbitrary date in the future (duration must be between 60 Days and seven years).

► SANnav calculates the number of days between the start and end renewal dates in days (renewal end – renewal start, expressed in days).

► The Export Renewal Request will download and generate a file (on the client specified browser default "Download Folder") containing all the relevant information for the customer to request the renewal quote.

► SANnav will generate a new SRV (SANnav Renewal Verification) Code as part of the Export Renewal Request to be used when placing an order for a license renewal.

► Example SRV Code - SRVS999D0777FMX12345 Refer to the SANnav Management Portal v2.3.0 User Guide, section Licensing for details on how to export the License Renewal Request file from the SANnav UI.

9.2.10 Downloading the SANnav Management Portal Software

> **Note**: Refer to the *IBM SANnav Management Portal v2.2.X Implementation Guide,* SG24-8534 for detailed information:

The SANnav software can be downloaded from IBM FixCentral. The detailed steps are outlined below.

1. Navigate to IBM FixCentral: https://www.ibm.com/support/fixcentral/

2. Search for "SANnav" and select the version you want to download. See Figure 9-6.

Figure 9-6   Find product

3. Select the installed version (in case of update) or **All f**or a new installation. See Figure 9-7 on page 149.
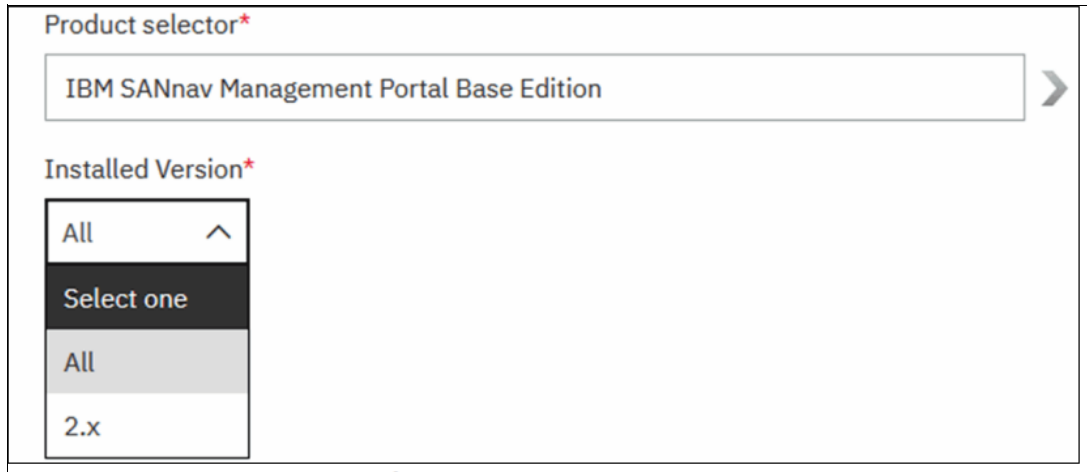
Figure 9-7   Select the installed version (in case of update) or All for a new installation

4. Click **Continue**.

5. Under Select fixes, choose the **SAN Storage Networking b type** fix pack. See Figure 9-8.
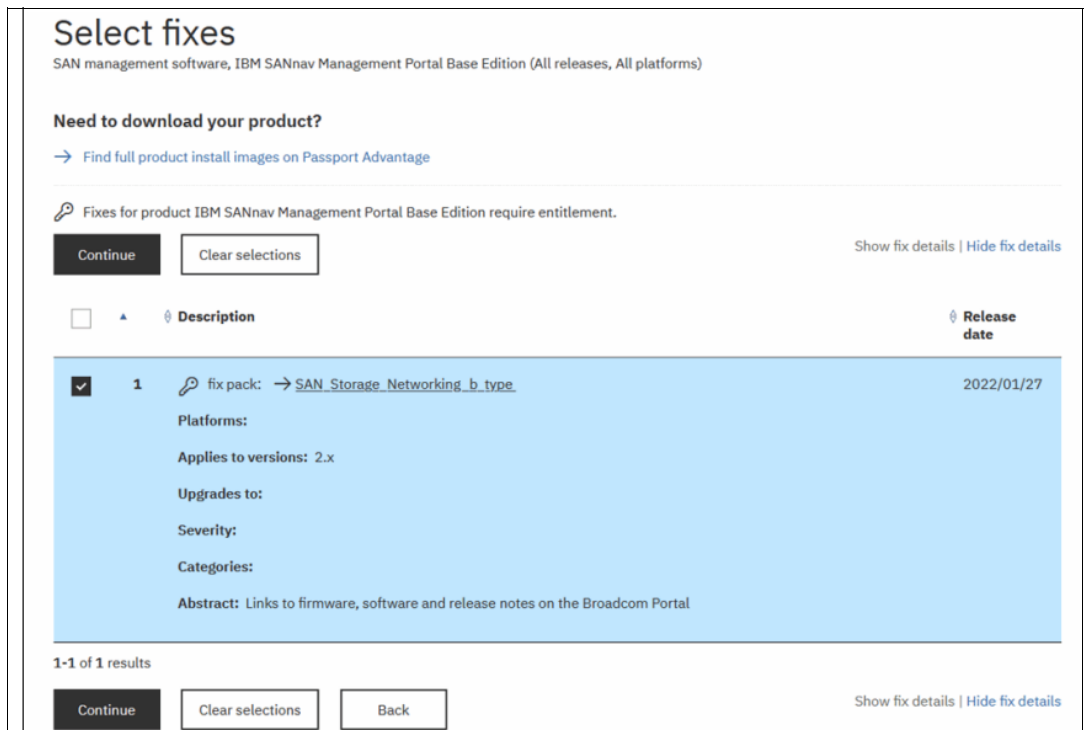
Figure 9-8   Select fixes

6. Click **Continue.**

7. To pass the entitlement check, you need to enter the machine type and the serial number.

Please provide the serial number of the machines for which Machine Code update(s) are designated and will be installed (each a "Target Machine").

The Type Number is a 4-digit number (usually followed by a 3-character Model identifier) printed on the exterior of your IBM system. It may be the first part of an ID labeled "Model" or "System Model" ID.

The Serial Number is a 7 digit ID labeled "S/N" on the exterior of your IBM system. Dash ("-") characters may be omitted.

The Country selection is based on the location of your IBM system.

See ⬚ more information for details about this page, and the actions available below.

Country

Germany ⌄

Machine type	Machine Serial Number
1.	9241	78▮▮▮▮▮

+ Add another

Upload machine type and serial number data

Browse

Save to file

If you are a third party representative of an IBM customer who has been duly authorized by the IBM customer to download Machine Code update(s), then by downloading Machine Code update(s), you agree to comply with all obligations of the IBM customer with respect to any Machine Code or Machine Code updates. Any copying, reproduction, distribution or installation of Machine Code updates, other than as expressly authorized by IBM, is prohibited.

Continue | Clear settings | Back

Figure 9-9   Entitlement check on FixCentral

8. Click **Continue**. You will be presented with a customized link to the Broadcom software portal. See Figure 9-10 on page 151.

Download files using HTTPS

SAN management software, IBM SANnav Management Portal Base Edition (All releases, All platforms)

Order number: 439635855

Show normalized list | Hide normalized list

---

fix pack: SAN_Storage_Networking_b_type

SAN_Storage_Networking_b_type

The following files implement this fix.

None

The remaining files for this fix can be downloaded from a third party location.

☐ LinkToBroadcomSWPortal

Back

Figure 9-10   Link to the Broadcom software portal

9. Click on the link to go to the Broadcom software portal.

10. Click **Continue** on the screen informing you that you are leaving the IBM web site

11. On the Broadcom page enter your email address and the captcha. See Figure 9-11

Welcome to Assist Portal

In order to enter Broadcom Assist Portal you must enter your company email address.

Email:*   john.doe@company.com

Retype Email:*   john.doe@company.com

Captcha:*   FMMF   FMMF ✔

NOTE: Once you click submit please ensure you check for a verification email by "bsn.web@Broadcom.com" to complete this process. If you do not see this email in your inbox please check your Junk folder. For information on Broadcom's privacy practices and commitment to protecting your privacy, please view our Privacy Policy.

SUBMIT

Figure 9-11   Enter your email address and the captcha

11. Click **Submit.**

12. You will be sent a verification code to your email address. See Figure 9-12.

Thank you for visiting Broadcom Assist for IBM on Broadcom assist portal to access Software Downloads and Documentation.
Please use the verification code - **yOB** ▮▮▮▮▮▮ **SDV** to verify your login.

Thank you,
Broadcom Assist Team
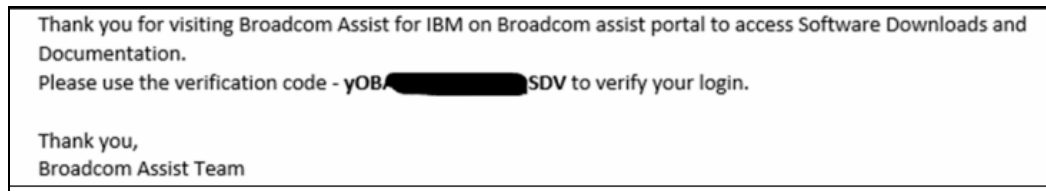
Figure 9-12   Verification code in your email

13. Enter the verification code and the captcha on the Broadcom Assist Portal. See Figure 9-13.

Figure 9-13   Entering Broadcom verification code and captcha

14. Select the files to download. See Figure 9-14.

Figure 9-14   Select files to download

15. Read and accept the end user license agreement (EULA) and click **I Accept**.

Figure 9-15   End user license agreement

16. In the next step you can download your file(s) and save them for the installation.

9.2.11 SANnav Management Portal v2.3.0 scalability features

Figure 9-16 on page 154 shows the SANnav Management Portal v2.3.0 scalability features.

Feature	Scalability Limit – SANnav Management Portal Base	Scalability Limit – SANnav Management Portal Enterprise
Maximum number of **SAN ports managed**	600	15,000
Maximum number of **end device ports managed**	2000	40,000
Maximum number of **end device ports per fabric**	10,000	
Maximum number of **Hosts** managed through vCenter discovery (*across all vCenter instances*)	300	
Maximum number of **events** stored	2 million	
Maximum number of **MAPS violations** stored	2 million	
Port statistics stored	5-minute samples are stored for up to 30 days. 1-hour data is stored for 30 days. 1-day aggregated data is stored for 30 days. 2-second samples are collected for up to 3 days for a maximum of 100 user-selected Gen 6 or Gen 7 ports. These ports can be on the same switch or across multiple Gen 6 or Gen 7 switches. Data is retained for 14 days.	
Extension Tunnel Statistics stored	5-minute samples are stored for up to 30 days. 1-hour data is stored for 30 days. 1-day aggregated data is stored for 30 days. 5-second samples are collected for up to 3 days for a maximum of 100 circuits (only supported for the SX6 Blade and 7810 switch). These circuits can be on the same switch or across multiple switches. Once data collection is complete, the data is retained for 14 days.	
Maximum number of **Flows** Supported	- Enterprise Edition (15K ports) "large" platform <u>only</u>. No support on "small" platforms 100k Flows tested and validated. Up to 150K Flows maximum allowed (not blocked). Blocked > 150K flows.	

Figure 9-16   Scalability feature

9.2.12 Important notes

Consider the following when deploying SANnav in your environment:

► The network latency between SANnav clients to the SANnav Management Portal server and between SANnav Management Portal server to the switches must not exceed 100ms. If the latency is higher than 100ms, then communication time-outs may occur and cause undesirable behavior.

► When configuring the VM for SANnav installation, make sure the MTU size of the network interface is set to 1500, otherwise SANnav will not receive Port Performance data for switches running Fabric OS less than v8.2.1b.

► Cockpit web console for Linux cannot co-exist with SANnav Management Portal. • SE Linux is not supported (Enforcing and Permissive).

► SANnav is expected to be installed and run on a dedicated host. If any other application is installed on the host, it is mandatory to un-install it before starting the SANnav installation.

► SANnav application performance may be affected during operations like SANnav backup and support data collection. It is recommended to schedule SANnav backup during application idle time.

- ► Applying TruFOS certificate from SANnav fails on FOS v9.1.1. Users need to upgrade to FOS v9.1.1a or use CLI to install TruFOS certificate on switches running FOS v9.1.1.
- ► Disaster Recovery (DR) is supported for SANnav Management Portal on VM or OVA deployments only. SANnav MP DR is not supported on bare metal deployments. DR is not supported on Global View (all deployments)

9.2.13 Launching SANnav Management Portal

Launch SANnav Management Portal to monitor and manage your fabrics. To do that, open your browser and enter the IP address or fully qualified domain name (FQDN) of the SANnav Management Portal server. See Figure 9-17.

You can use HTTP or HTTPS, as shown in the following examples:

- ► `http://192.0.2.0`
- ► `https://192.0.2.0`
- ► `http://sannavserver.company.com`
- ► `https://sannavserver.company.com`
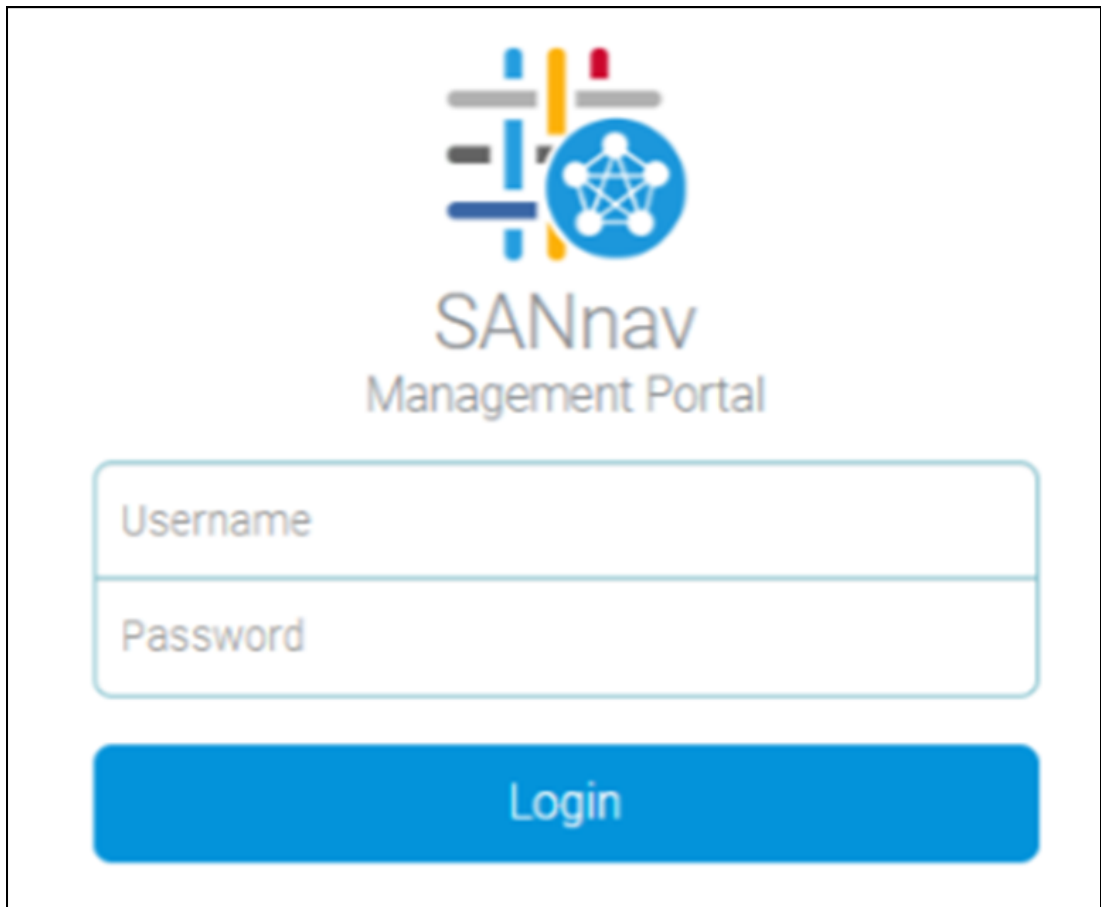
Figure 9-17   SANnav Management Portal

Refer to the *Brocade SANnav Management Portal User Guide* for detailed information:

`https://techdocs.broadcom.com/management-portal`

9.2.14 Overview of the user interface

SANnav Management Portal allows you to manage one or more SAN fabrics in multiple locations. Once familiar with the basic components of SANnav, you can quickly start monitoring and managing your fabrics. Figure 9-18 shows the basic layout of the SANnav user interface.
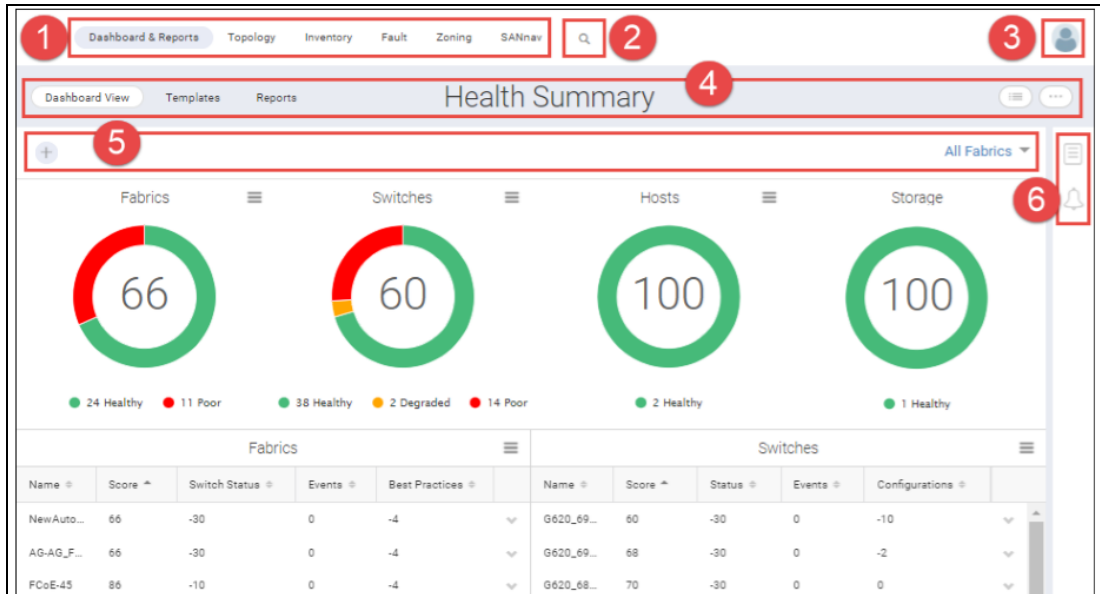
Figure 9-18   Basic layout of the SANnav user interface

Author Comment: We need to describe what these panes are.

Detail pages for a switch example: SAN42B-R7

Clicking a fabric name, switch name, or port name in a table opens a detail page for that object. The detail page displays additional information about the object and may contain other actions that you can perform. The detail page is different depending on the context. For example, the detail page for a fabric on the Inventory page (Figure 9-19) is different from the detail page for the same fabric on the Discovery page (Figure 9-20 on page 157),
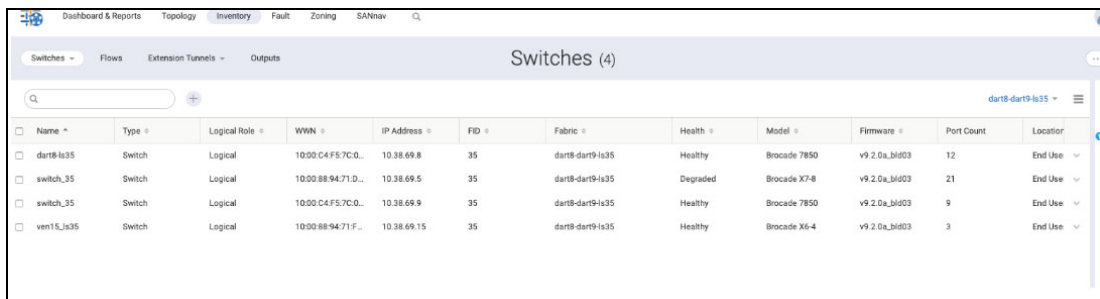
Figure 9-19   Detail page for Inventory page
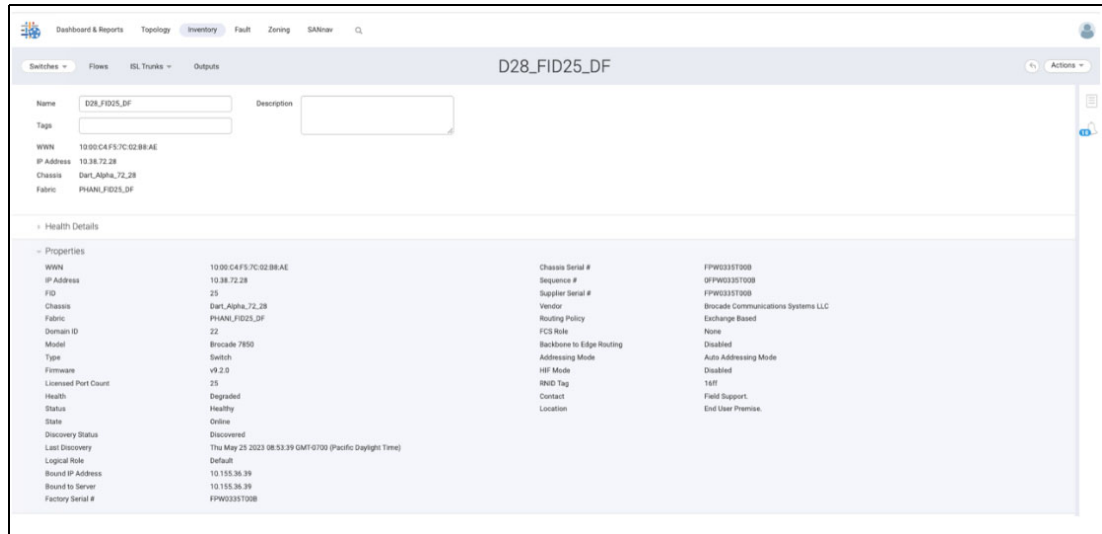
Figure 9-20   Detail page on the Discovery page

Tables example: SAN42B-R7

When you select table entries, selecting the checkbox in the top-left column does not select all entries, but selects only the entries that have been loaded into the user interface. To select all entries, you must scroll to the bottom of the table until all entries are loaded into the user interface, and then select the checkbox in the top-left column. This behavior applies to all tables. See Figure 9-21.

Figure 9-21   Filtering violations: SAN42B-R7 example

You can view violations in the Violations tab. The Violations tab displays the violations for the switches for which a rule threshold has been crossed.

> **Note:** You must apply violation filters to generate violation reports.

The violations can be filtered based on the following columns:

- All

- Measures
- Port Type
- Rule Name
- Severity
- Source Address
- Source Name

You can filter violations that are based on the following violation categories:

- All
- Backend Port Health
- Extension GE Port Health
- Extension Health
- FRU Health
- Fabric Health
- Fabric Performance Impact
- Flow Collection Aggregation
- IO Health
- IO Latency
- Port Health
- Security Violations
- Switch Resource
- Switch Status Policy
- Traffic Performance
- Virtual Machine Violations

When you select a violation category other than the All option (for example, Fabric Health), the violation column drop-down menu displays the All option. With the All option, you can filter all violations by category. The Value field does not appear when you select the All option in the violation column. The rule name is applicable for both single and multiple violation category selections. It is auto-populated after you type three letters. You can also filter the violations based on the port type, measures, threshold values, and severity.

> **Note:** The value for the measures is the measure name as listed in the Measure Name column.

When you are viewing violations, the filter bar on the Violations window provides the following options for filtering the displayed violations:

► Click the **+** button to add existing filters and create violation filters.

► Click the **All Fabrics** drop-down to select a fabric.

► Click **Last 30 Minutes** to select the date range for displayed violations. In addition to filtering the violations, you can also click the hamburger icon ( ) to hide or display specific columns of data.

To create a custom violation filter, perform the following steps:

1. Click **Fault** in the navigation bar, and then select the **Violations** tab. The Violations window displays. The Violations window lists the MAPS violations based on the selected fabric.

2. Click the **+** button on the left side of the filter bar. See Figure 9-22 on page 159.

	Rule Name	Rule Condition	Category	Actions	Measure	Measure Value	Entity Name	Port Type	Source Name	Source Address	FID	Last Occurred
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	98.62	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 13:11:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	100.00	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 13:10:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	98.63	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 13:09:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	100.00	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 13:08:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	98.65	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 13:07:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	98.61	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 13:06:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	100.00	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 13:05:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	98.63	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 13:04:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	100.00	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 13:03:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	98.61	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 13:02:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	100.00	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 13:01:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	98.63	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 13:00:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	100.00	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 12:59:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	98.62	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 12:58:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	98.61	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 12:56:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	100.00	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 12:55:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	98.60	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 12:54:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	100.00	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 12:53:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	98.61	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 12:52:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	98.62	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 12:51:4...
⚠	defALL_TARGET_PO...	ALL_TARGET_PORTS...	Fabric Performance I...	RAS Log Event	RX Performance (RX)	100.00	port22	F-Port	switch_35	10.38.69.9	35	May 02, 2023 12:50:4...

Figure 9-22   Added filter window

The Add Filter window displays. You can add existing violation filters, or you can create new violation filters.

Topology visualization example: SAN42B-R7

Using SANnav Management Portal, you can easily view and navigate a visual representation of the elements in your SAN topology based on a selected context. This visual representation enables you to focus on the information in the topology view instead of a complex network of devices and connections.

The Topology page displays graphical representations of the fabrics. For example, after you discover a fabric, you might want to view the topology to see a pictorial representation of the connected switches and devices.

You can display a topology for the following contexts:

► Fabric context: Displays all switches in the fabric and in other directly connected fabrics.

► Switch context: Displays all fabrics, switches, and devices that are directly connected to the selected switch.

► Switch port context: Displays all entities that are connected to the selected switch port.

► Host or storage context: Displays the connectivity to edge switches, fabrics, and other devices that are zoned with the selected device.

► Host port or storage port context: Displays edge switches, fabrics, and other device ports that are zoned with the selected device port.

► Zone context: Displays all zone members, including involved fabrics.

Note: If an icon on the topology page is grayed out, it means that the associated object is unavailable. The topology shows information that is related to discovered fabrics only.

For this reason, for FC Routing, you should discover all fabrics (backbone and edge fabrics) in the same instance of SANnav Management Portal.

Topology views are a snapshot in time, and they are not automatically updated. You can update the topology view by clicking the refresh icon in the upper-right corner of the page. Also, if you navigate away from the Topology page, when you return to the page, the view is updated with the latest data.

The Browse Topology page (Figure 9-23) displays a pictorial view of the fabric. This view is the fabric context, so the topology displays all switches in the fabric.

> **Note:** The fabric name that is shown in the context navigation panel at the top of the window.
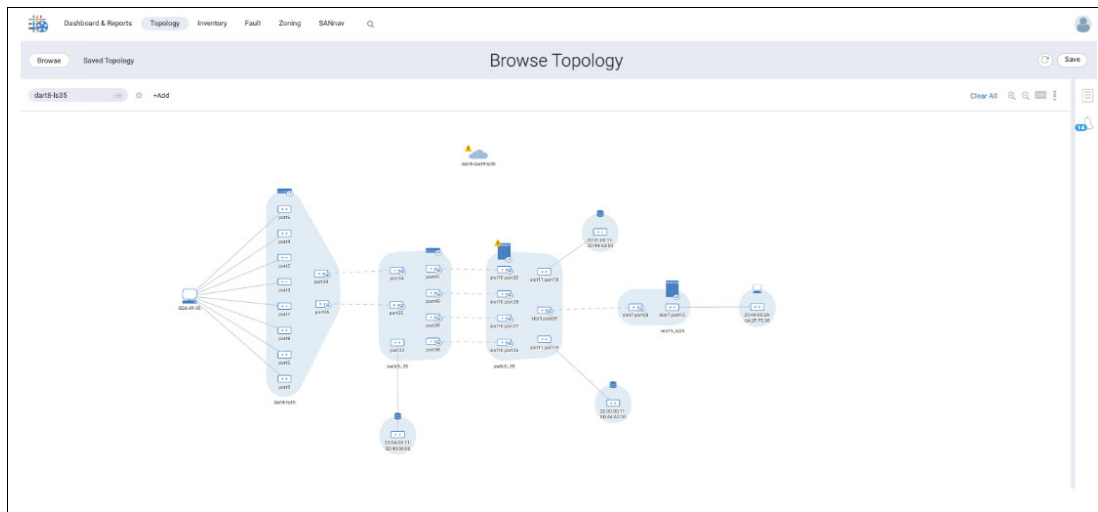
Figure 9-23   Browse Topology page

9.2.15 Extension tunnels and circuits

Using SANnav Management Portal, you can set up both Fibre Channel over IP (FCIP) tunnels and IP Extension tunnels.

Brocade extension products support both FC/FICON-based data flows and IP-based storage data flows. Brocade extension solutions maximize replication and backup throughput over distance using data compression, disk and tape protocol acceleration, and WAN-optimized TCP. Brocade extension supports applications such as remote data replication (RDR), centralized backup, and data migration.

Brocade extension uses the existing IP wide area network (WAN) infrastructure to connect Fibre Channel and IP fabrics between distant endpoints, which are impractical or costly using native Fibre Channel or IP connections. The basis of the connection is the extension tunnel, which is built on a physical connection between two extension switches or blades. Extension tunnels allow Fibre Channel and IP traffic to pass through the IP WAN. The extension tunnel connections ensure lossless transmission and in-order delivery of FC and IP frames. The Fibre Channel fabric and all targets and initiators, whether FC or IP, are unaware of the presence of the IP WAN.

The extension tunnel provides load balancing across separate network paths, optimization for extended links, rate limiting to ensure optimal performance, and lossless link loss (LLL) recovery.

Two major classifications exist based on the type of I/O:

- ► FCIP
- ► IP Extension

FCIP tunnels

FCIP tunnels are typically established across WANs. Management of the tunnels often involves monitoring the FCIP traffic across the WAN and monitoring the associated FC traffic at each end of the tunnel. SANnav Management Portal provides ample tools for monitoring both types of traffic.

IP Extension tunnels

The IBM SAN42B-R7 Extension Switch, the IBM SAN42B-R Extension Switch, the IBM SAN18B-6 Extension Switch, and the IBM SX6 Extension Blade support IP Extension.

Extended IP traffic receives the same benefits as traditional FCIP traffic. IP Extension provides Layer 3 extension for IP storage replication.

The deployment models include the following configurations:

- ► Direct connection (IP storage with the IBM SAN42B-R7 Extension Switch, the IBM SAN42B-R Extension Switch, or the IBM SX6 Extension Blade
- ► LAN ports connected to a Layer 2 switch SANnav Management Portal supports configuration of hybrid tunnels (FCIP + IP Extension).

SANnav Management Portal supports configuration of hybrid tunnels (FCIP + IP Extension).

Viewing Extension tunnels

The following steps describe how to view the list of extension tunnels and explain the possible values for the Status column in the tunnel list.

1. Click the **SANnav** tab, and then select **SAN Configuration** → **Extension Tunnels Management**.

2. Click the **Tunnels** tab. This tab displays all tunnels that are configured between the discovered extension tunnel-capable switches. See Figure 9-24

Figure 9-24   Tunnels tab

The Status column values are defined as follows (see Figure 9-25 on page 162).

Configuration Status	Description
Online	Both ends of the tunnel have a status of Online.
Offline (In Progress)	If either end of the tunnel is not Online, then the tunnel status is Offline.
Inactive	The tunnel configuration is not pushed to the switch.
Degraded	One switch is degraded, and the other switch is either Online or degraded.
Disabled	If either end of the tunnel is disabled, then the tunnel status is Disabled.

Figure 9-25   Status column values

Note: One-sided tunnels (discovered from a switch) are not displayed in the Extension Tunnels configuration list. You can see those tunnels only from Inventory connections.

Note: If the list of tunnels is incomplete, the SANnav server has yet to discover all switches. The description from the first switch is displayed. If the description is unavailable for the first switch, then data from the second one is displayed. If the description is set on the SANnav server but not set on either switch, the setting on the server is displayed.

Applying the Extension Dashboard

SANnav dashboards provide you with a customizable view of your SAN environment.

SANnav provides four dashboards, one of which is devoted to the extension tunnel management.

Perform the following steps to launch and utilize the Extension Dashboard.

1. Click **Dashboard & Reports** in the navigation bar, and then click **Dashboard View** in the sub-navigation bar.

2. Click the **Select Dashboard** button in the upper-right corner of the window, select E**xtension Dashboard**, and click **OK**.

   The Extension Dashboard displays. The dashboard consists of six widgets: two showing extension tunnel data, and four showing circuit data. See Figure 9-26 on page 163.
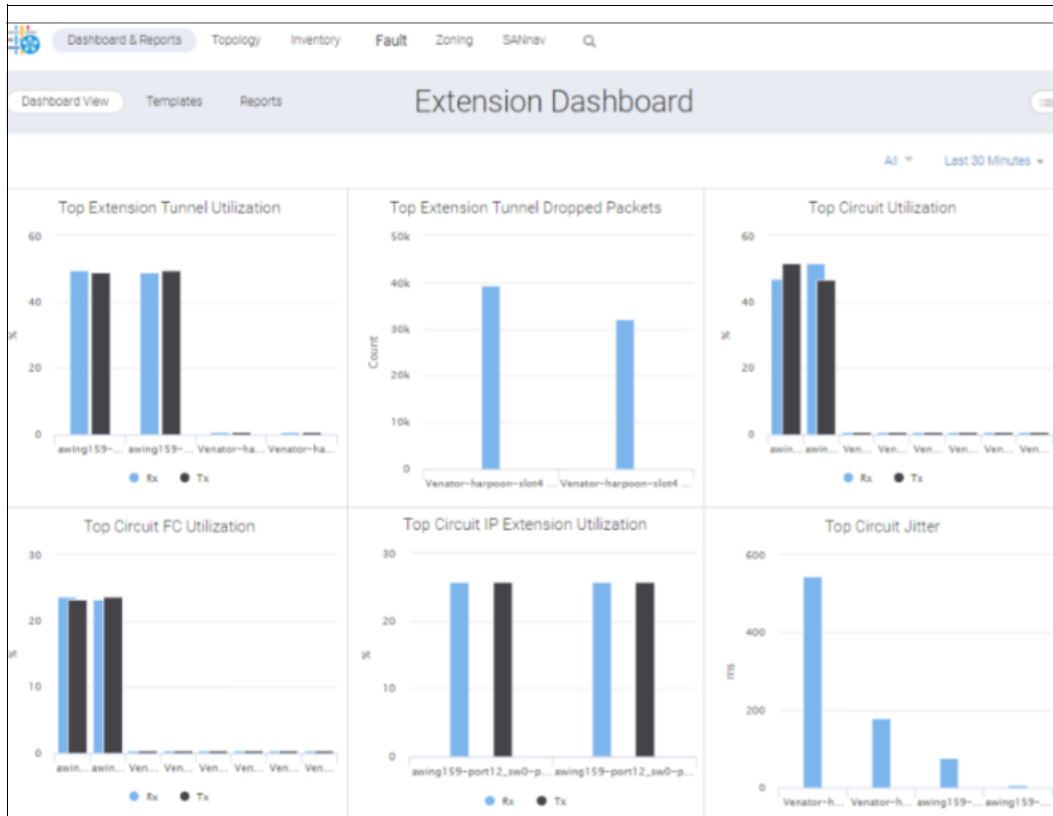
Figure 9-26   Extension Dashboard

3. If you want to examine tunnel or circuit utilization, click a bar in one of the utilization widgets, and then select **I**nvestigate from the list. See Figure 9-27 on page 164.
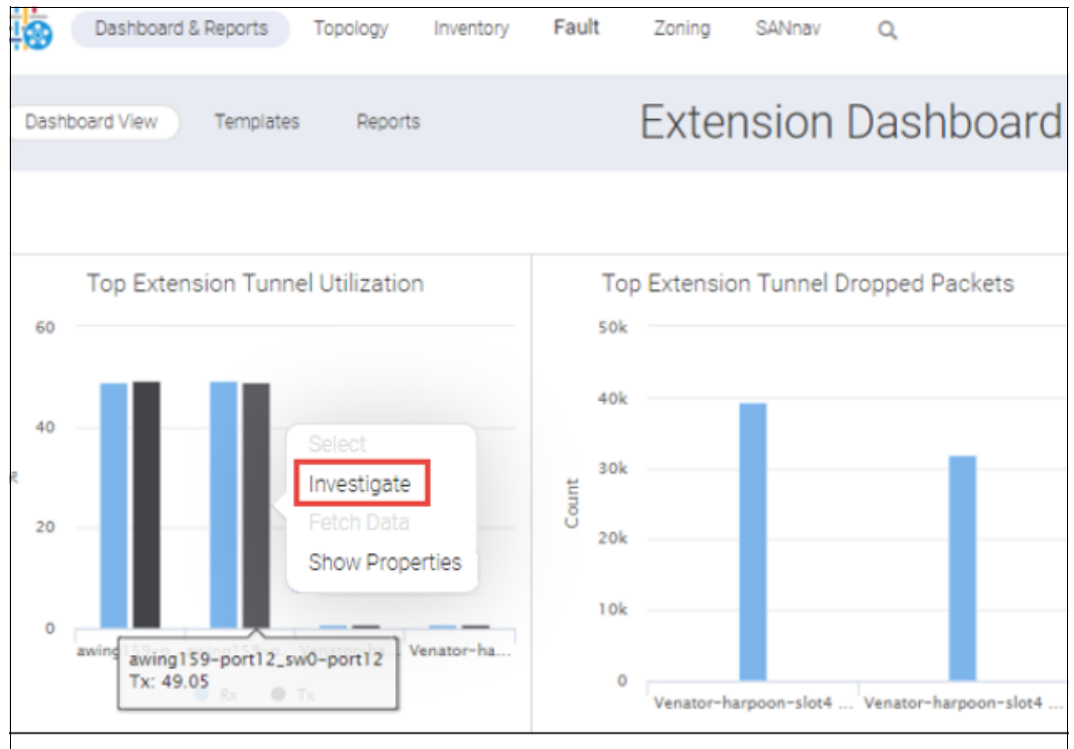
Figure 9-27   Tunnel or circuit utilization

4. Using Investigation mode, you can see trends over time. Figure 9-28 on page 165 shows the Investigation Mode dialog for a tunnel. Click the **X** in the upper-right corner of the window to return to the dashboard view.
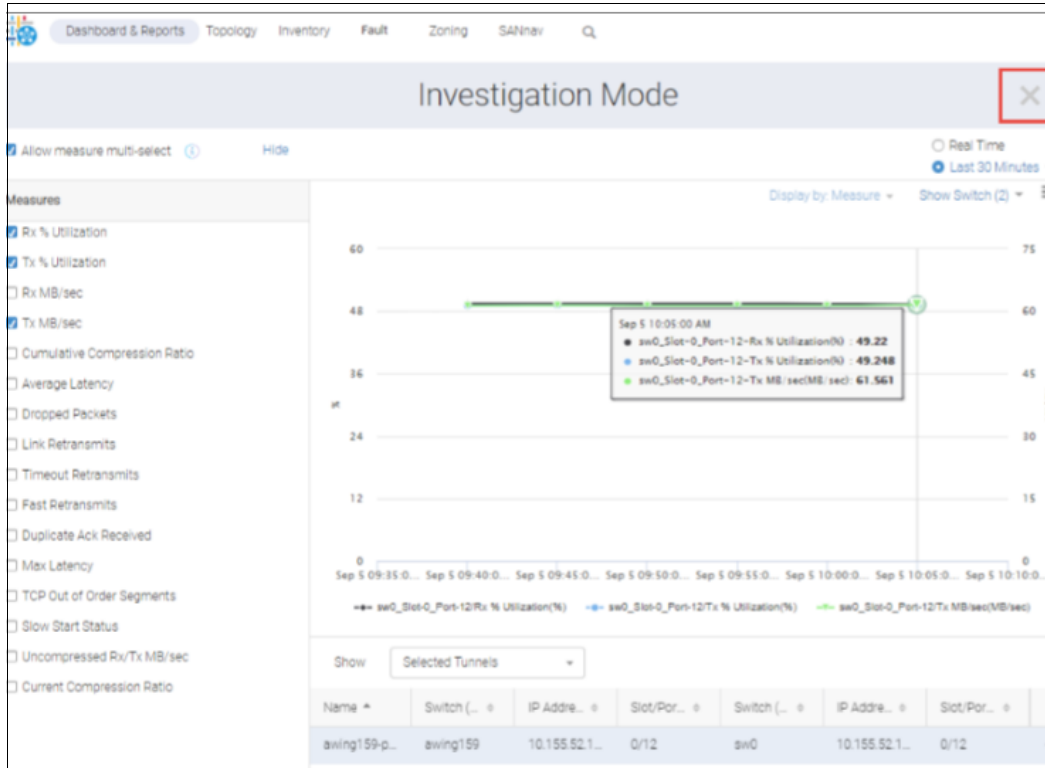
Figure 9-28   Investigation Mode dialog for a tunnel

5. If you want to display the properties of a tunnel or circuit, click a bar in one of the graphs, and then select Show Properties from the list. This displays a pop-up list with details about the tunnel or circuit. Figure 9-29 shows the properties of a tunnel.
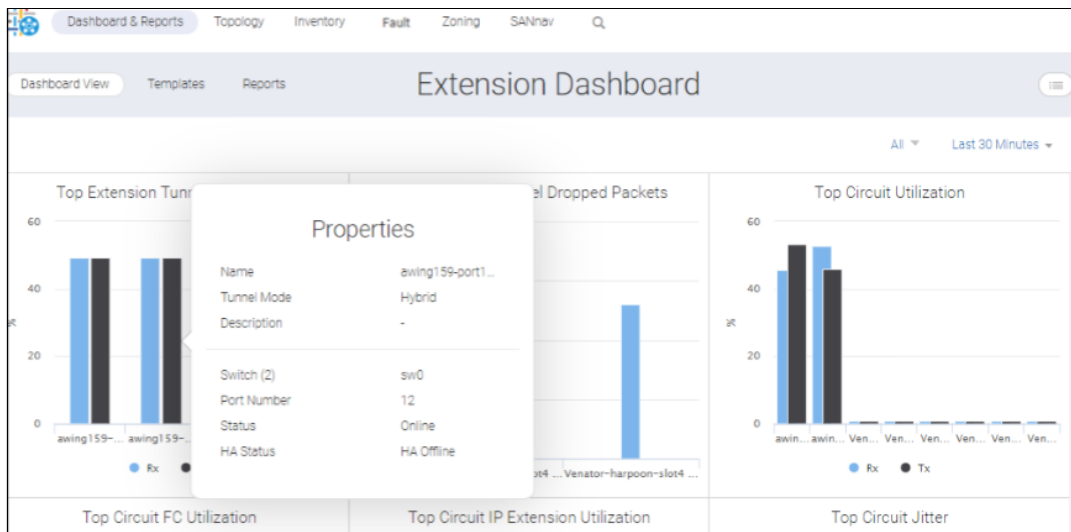

Figure 9-29   Properties of a tunnel.

6. To change the network scope and time range, click the drop-down lists in the upper-right corner of the dashboard window. The extension widgets display data for the tunnels and circuits that belong to the fabrics in the selected network scope and for the selected time range.

The Brocade SANnav offering has been completely rearchitectured with a new look and feel for the next generation of SAN infrastructure. The Brocade SANnav management tool is the only management tool that supports the latest IBM b-type SAN Extension platforms.

For more information on Brocades SANnav management tool refer to the following url:

https://www.broadcom.com/products/fibre-channel-networking/software/sannav-management-portal

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

► *IBM b-type Gen 7 Installation, Migration, and Best Practices Guide*, SG24-8497

► *IBM SANnav Management Portal v2.2.X Implementation Guide*, SG24-8534

► *IBM SAN42B-R Extension Switch and IBM b-type SX6 Extension Blade in Distance Replication Configurations (Disk and Tape)*, REDP-5404

► *Implementation Guide for IBM Storage FlashSystem and IBM SAN Volume Controller (for IBM Storage Virtualize 8.6)*, SG24-8542

► *IBM Storage DS8900F Architecture and Implementation: Updated for Release 9.3.2*, SG24-8456

► *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

► IBM Storage area network (SAN) solutions

https://www.ibm.com/storage-area-network

► Brocade Fabric OS Extension User Guide, 9.2.x

https://techdocs.broadcom.com/fabric-os-extension

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 526. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the .5" spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize(-->Hide:)>Set** . Move the changed Conditional text settings to all files in your book by opening the book file with the spine.fm still open and **File>Import>Formats** the Conditional Text Settings (ONLY!) to the book files.

Redbooks

IBM B-Type SAN Extension Platform

SG24-8553-00

ISBN DocISBN

(1.5" spine)
1.5" <-> 1.998"
789 <->1051 pages

Redbooks

IBM B-Type SAN Extension Platform

SG24-8553-00

ISBN DocISBN

(1.0" spine)
0.875" <->1.498"
460 <-> 788 pages

Redbooks

IBM B-Type SAN Extension Platform Implementation Guide

SG24-8553-00

ISBN DocISBN

(0.5" spine)
0.475" <->0.873"
250 <-> 459 pages

Redbooks

IBM B-Type SAN Extension Platform Implementation Guide

(0.2"spine)
0.17"<->0.473"
90<->249 pages

Redbooks

IBM B-Type SAN Extension Platform Implementation Guide

(0.1"spine)
0.1"<->0.169"
53<->89 pages

Draft Document for Review October 13, 2023 3:23 pm

8553spine.fm   170

IBM B-Type SAN Extension Platform

SG24-8553-00

ISBN DocISBN

(2.5" spine)
2.5"<->nnn.n"
1315<-> nnnn pages

IBM B-Type SAN Extension Platform

SG24-8553-00

ISBN DocISBN

(2.0" spine)
2.0" <-> 2.498"
1052 <-> 1314 pages

SG24-8553-00

ISBN DocISBN

Printed in U.S.A.

ibm.com/redbooks