SVC and Spectrum Virtualize Hyperswap SAN Design Best Practices

David Green
IBM SAN Central

Jordan Fincher
Spectrum Virtualize Level 3 support

Accelerate with IBM Storage Webinars

The Free IBM Storage Technical Webinar Series Continues in 2020…

Washington Systems Center – Storage experts cover a variety of technical topics.

Audience: Clients who have or are considering acquiring IBM Storage solutions. Business Partners and IBMers are also welcome.

To automatically receive announcements of upcoming Accelerate with IBM Storage webinars, Clients, Business Partners and IBMers are welcome to send an email request to accelerate-join@hursley.ibm.com.

Located on the Accelerate with IBM Storage Site: https://www.ibm.com/support/pages/node/1125513

Also, check out the WSC YouTube Channel here:
https://www.youtube.com/channel/UCNuks0go01_ZrVVF1jgOD6Q

2020 Upcoming Webinars:

August 25 - Protecting your business with GDPS Logical Corruption Protection (LCP) Manager
Register Here: https://ibm.biz/Bdq8gA

September 10 – VersaStack 2020 – What's the latest?
Register Here: https://ibm.biz/BdqVQ7

September 15 – DS8900F Product Update
Register Here: https://ibm.biz/BdqAS6

Accelerate with IBM Storage Survey

Please take a moment to share your feedback with our team!

You can access this 5-question survey via Menti.com with code 91 74 40 or

Direct link https://www.menti.com/mkg7a2x6q8

Or

QR Code

What We Will Cover

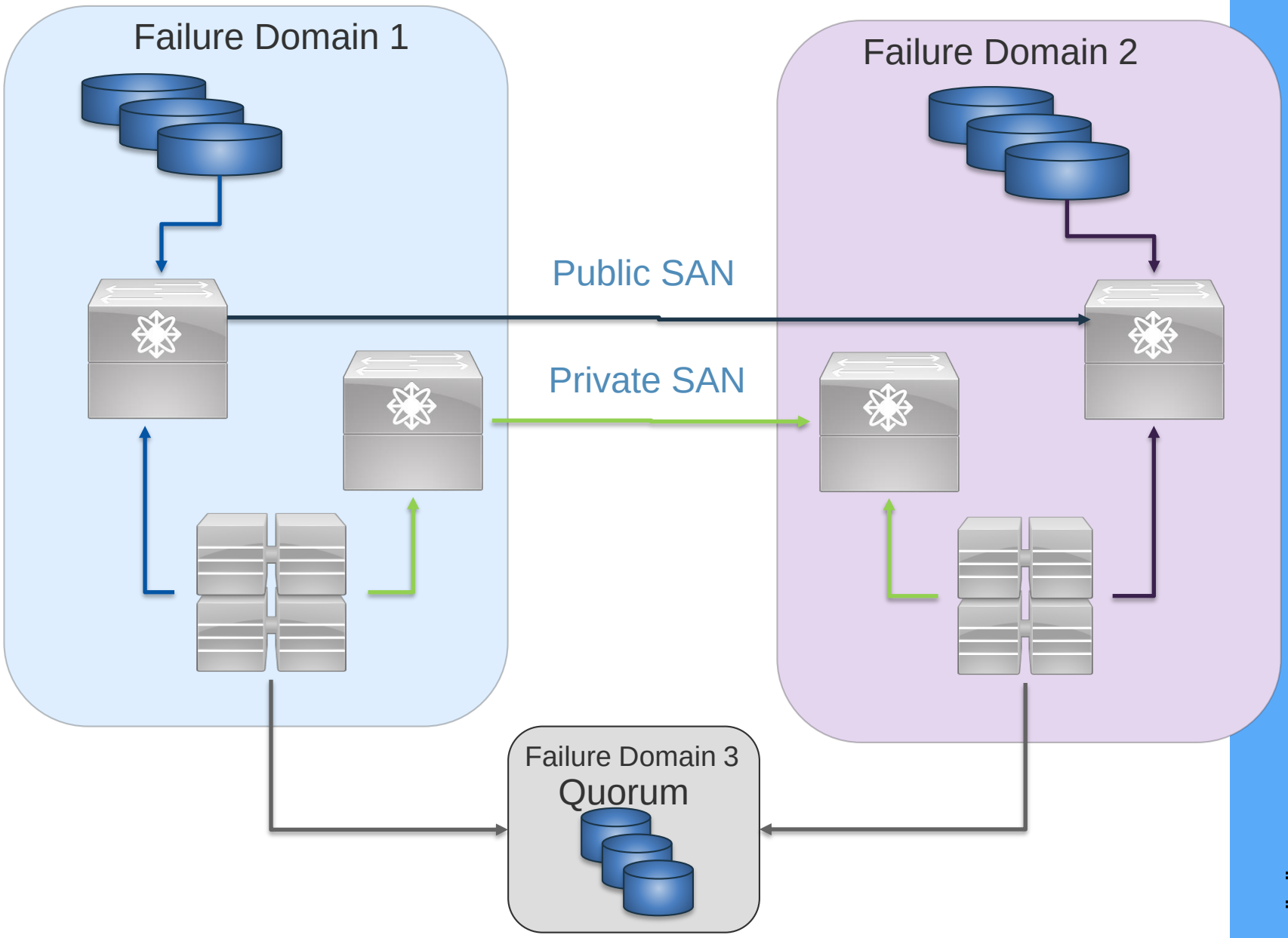
- Basic requirement for SAN Topology for HyperSwap and Stretched Clusters
- Conceptual Design
- Design Variations
- Cisco Design Considerations
- Brocade Design Considerations
- Sizing the Inter-site links

Any recommendations made or SAN diagrams pictured in this session apply to SVC, Spectrum Virtualize and FlashSystem clusters in a HyperSwap or Stretched configuration unless otherwise noted.

The Requirement Stated Simply:

A cluster configured as a Stretched or Hyperswap cluster must have a dedicated means of communication between the nodes in the cluster

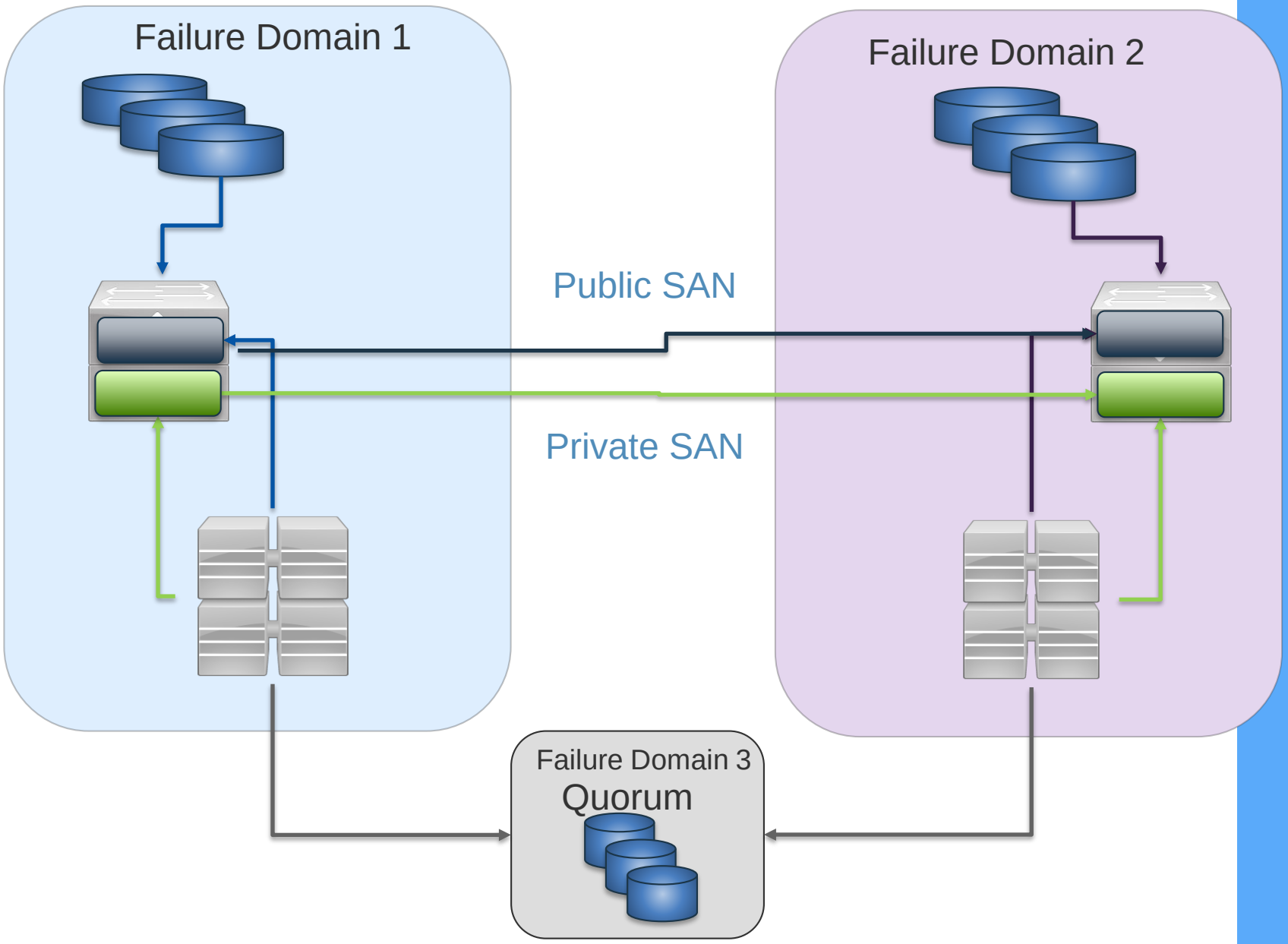
Hyperswap SAN Design



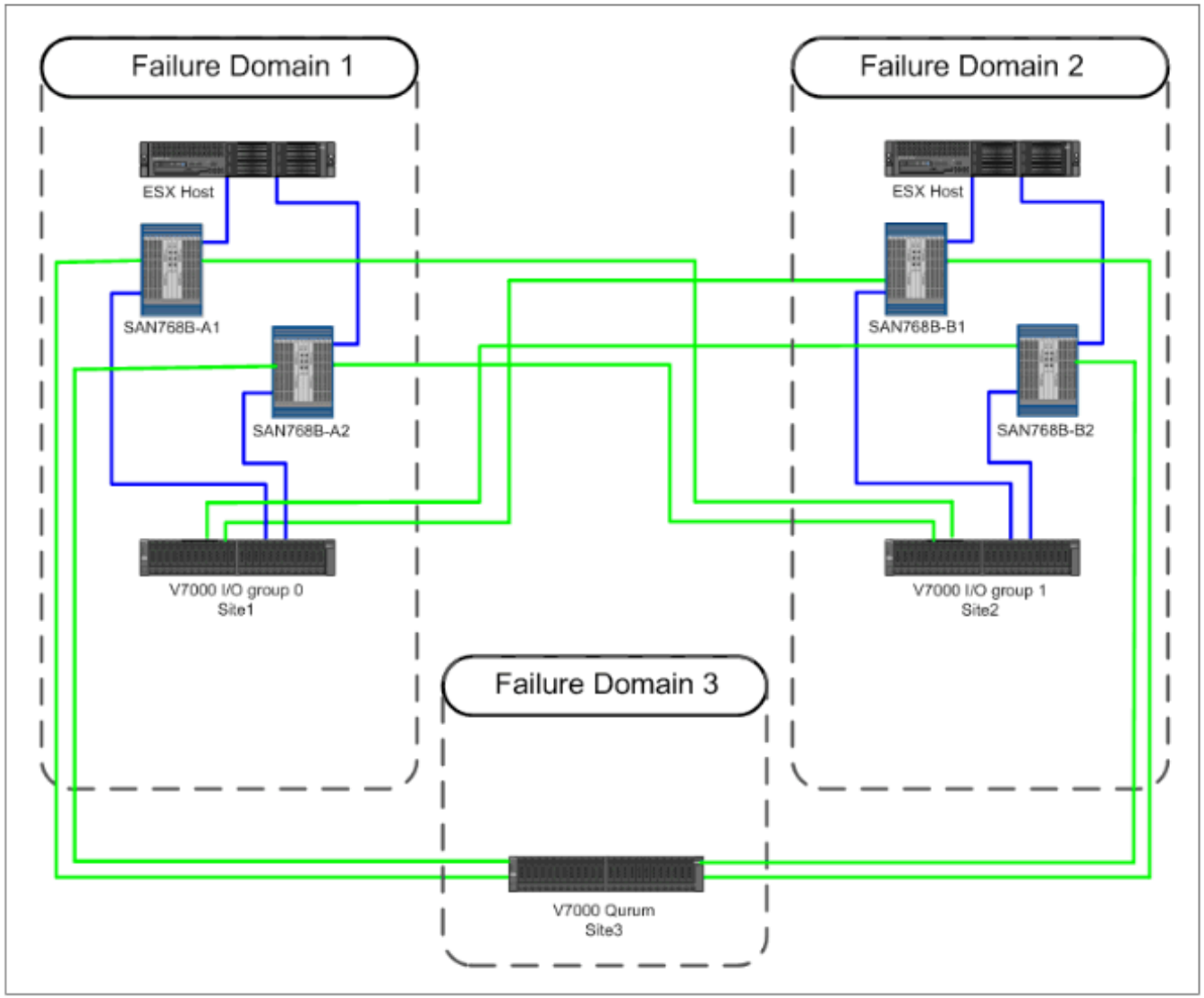
Hyperswap SAN Design with ISLs

- Private fabrics must be completely private – including implementing separate ISLs from public fabrics
- Private fabrics can be virtual (Brocade virtual fabrics or Cisco VSANs)

Hyperswap with Virtual Switches



Hyperswap SAN Design with No ISLs



Best Practice For ISLs

- ISLs should always be trunked
- Where possible use multiple providers for inter-site links
- If multiple providers are used, put each public/private fabric pair on a separate provider rather than both fabrics on both providers

Port Masking and Zoning Considerations

- As with standard topology clusters, port masking should be implemented
- Properly configured dedicated fabrics remove the zoning requirements for zoning inter-node ports away from hosts and storage
- The NPIV feature for hosts should still be implemented

Port Masking and Zoning Considerations

	4 port	8 port	12 port	16 port	SAN Fabric
Adapter 1 Port 1	Host+Storage	Host+Storage	Host+Storage	Host+Storage	A
Adapter 1 Port 2	Host+Storage	Host+Storage	Host+Storage	Host+Storage	B
Adapter 1 Port 3	Intracluster+Replication	Intracluster	Intracluster	Intracluster	A
Adapter 1 Port 4	Intracluster+Replication	Intracluster	Intracluster	Intracluster	B
Adapter 2 Port 1		Host+Storage	Host+Storage	Host+Storage	A
Adapter 2 Port 2		Host+Storage	Host+Storage	Host+Storage	B
Adapter 2 Port 3		Intracluster or Replication	Replication or Host+Storage	Replication or Host+Storage	A
Adapter 2 Port 4		Intracluster or Replication	Replication or Host+Storage	Replication or Host+Storage	B
Adapter 3 Port 1			Host+Storage	Host+Storage	A
Adapter 3 Port 2			Host+Storage	Host+Storage	B
Adapter 3 Port 3			Intracluster	Intracluster	A
Adapter 3 Port 4			Intracluster	Intracluster	B
Adapter 4 Port 1				Host+Storage	A
Adapter 4 Port 2				Host+Storage	B
Adapter 4 Port 3				Replication or Host+Storage	A
Adapter 4 Port 4				Replication or Host+Storage	B
localfcportmask	1100	11001100 OR 00001100	110000001100	0000110000001100	
remotefcportmask	1100	00000000 OR 11000000	000011000000	1100000011000000	

Host refers to host objects defined in the system.

Storage refers to controller objects defined in the system if external storage is being used.

Replication refers to nodes which are part of a different cluster.

Intracluster refers to nodes within the same cluster.

The "+" indicates that both types are should to be used

The word "or" indicates that one of the options must be selected. If using replication, preference should be given to replication.

Cisco and C-Type Fabric Design Considerations

- Trunked ISLs on a Cisco or C-Type fabric are called Port-Channels
- Cisco Trunking means an ISL or Port-Channel is configured to allow multiple VSANs traversing the ISL (TE) ports
- The default for Cisco is to create Port-Channels in TE mode and allow traffic for all VSANs
- ISLs on the private fabric must NOT be configured as TE ports, or the VSANs allowed statement must only include the private VSAN

Cisco and C-Type Implementation Example

```
show interface port-channel1
.......
port-channel1 is up
Hardware is IPStorage
Port WWN is 24:01:00:2a:6a:b4:71:e0
Admin port mode is auto, trunk mode is auto
Port vsan is 1
Trunk vsans (admin allowed and active) (1,5)
```

Trunk mode should be turned off for Port-Channels on the private VSAN
And allowed VSANs should only include the private VSAN

```
sc9222ia# conf t
sc9222ia# int port-channel1
sc9222ia# switchport trunk mode off
sc9222ia# switchport trunk allowed vsan 1
```

IBM

Brocade and B-Type Design Considerations

- XISLs allow multiple virtual fabrics to traverse the same physical ISLs
- Private fabrics must not be allowed to travers the shared XSLs
- Separate ISLs must be implemented in the virtual fabrics being used for the public and private fabrics
- Not all Brocade switches support Virtual Fabrics. For smaller Brocade switches the fabric design must include separate physical switches for the public and private fabrics

Inter-site Link Sizing

- Some baseline numbers. A link is considered fully utilized at 80% used.
- Based on the above, for each Gigabit of speed, you can achieve 100MBps throughput, instead of the theoretical 125MBps when doing a direct conversion

Speed in Gbps	Max Data Rate in MBps
8	800
16	1600
32	3200

- The above numbers assume the physical link is optimal
- Factors such as distance and the health of the link can lower the maximum throughput

Inter-site Link Sizing Considerations

HyperSwap volumes mirror all writes between sites. Additionally, the cluster may (in the event of a back-end storage failure) may require the forwarding of reads across the private inter-site link. As such the following needs to be considered when sizing the inter-site link:

- Average Write Data Rate of all volumes
- Peak Write Data Rate of all volumes
- Performance expectations during failure scenarios

In order to assess this, you will want to review long term historical performance data of your existing storage environment. As a general rule, a **minimum of 1 month's data** should be used to determine the requirements. The goal is to ensure an accurate workload profile of all phases of production are captured in the assessment.

Pro Tip: IBM Storage Insights or Spectrum Control can be used to gather the long term performance data. If using Storage Insights, IBM will automagically have this data as well!

Inter-site Link Sizing - Examples

Assuming a local throughput demand of 2GB/s on the volumes...

16Gbps = 1600MB/s
2GB/s = 2048MB/s
2048 / 1600 = 1.28

So easy math says we need 1.28 16Gb links.

This is absurd as you can't have a fractional port, so this is really 2 16Gb links which ultimately provide a peak bandwidth of 3200MB/s

If we do the same math with 8Gb ports...
2048 / 800 = 2.56 =roundup= 3 8Gb ports for peaks up to 2400MB/s

The same math with 32Gb ports...
2048 / 3200 = 0.64 =roundup= 1 32Gb port for peaks up to 3200MB/s

Inter-site Link Sizing Additional Considerations

- Simply meeting the minimum in the previous example may not provide adequate bandwidth for workload growth or peak demands

- Ideally each fabric has links sized to carry the entire load if one of the fabrics fails
- The easy math to factor in redundancy is 2x the calculated capacity