#SHAREorg

SHARE

The Inner Workings and Hidden Mechanisms of FICON

David Lytle, BCAF
Principal Engineer / Global Solutions Architect
Brocade Communications
dlytle@brocade.com

March 13, 2014

Session Number 14375

SHARE in Anaheim

Legal Disclaimer

All or some of the products detailed in this presentation may still be under development and certain specifications, including but not limited to, release dates, prices, and product features, may change. The products may not function as intended and a production version of the products may never be released. Even if a production version is released, it may be materially different from the pre-release version discussed in this presentation.

Nothing in this presentation shall be deemed to create a warranty of any kind, either express or implied, statutory or otherwise, including but not limited to, any implied warranties of merchantability, fitness for a particular purpose, or non-infringement of third-party rights with respect to any products and services referenced herein.

OF MAINFRAME DOMINANCE

Happy Birthday System z!

- Mainframes are in use in 90% of the *Fortune* 1000 and a large percentage of midmarket companies"
  - IBM and the Computer & Communications Industry Association

…more than 70 percent of all corporate data resides on mainframes. Legacy systems represent *trillions of dollars* in assets…

- z/Journal

- About 25 % of IBM's annual revenue comes from the sale of mainframes and associated products like storage systems, software and services!
  - NY Times

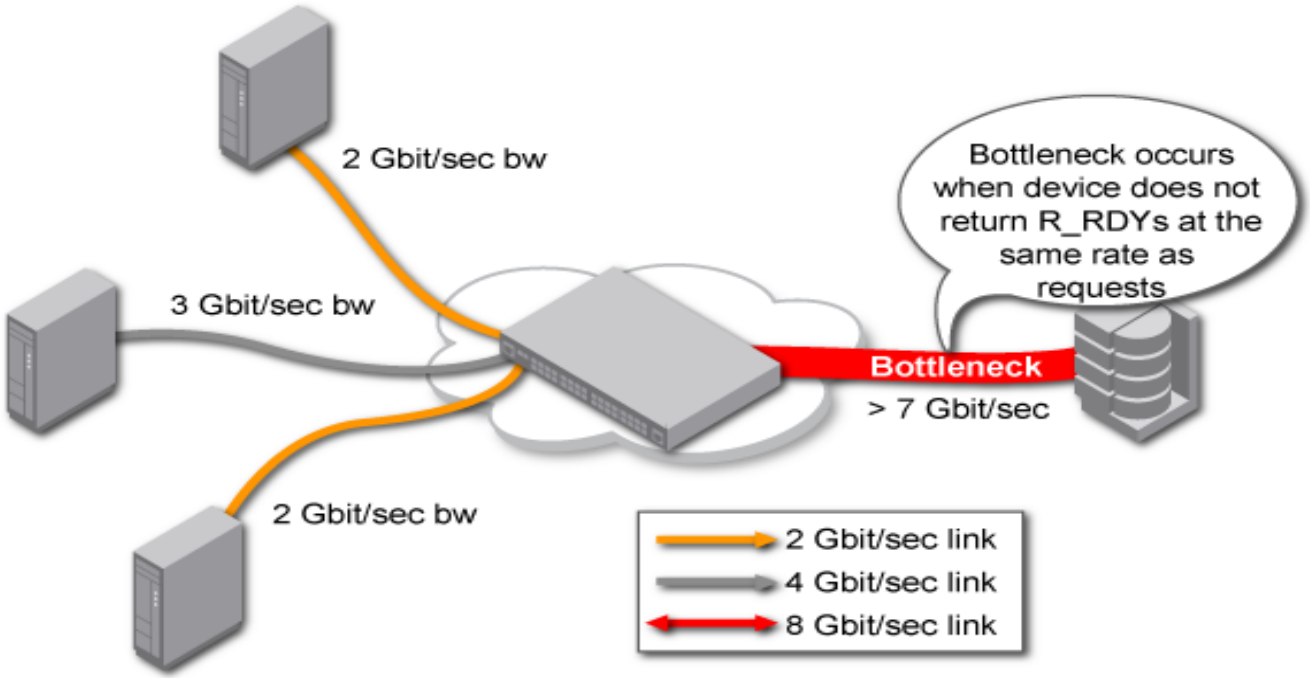
Now – Let‘s Look Inside FICON I/O Processing

- There have been many, significant improvements in how I/O is handled by the mainframe.
- **FI**bre **CON**nection, introduced in 2000, has undergone a vast number of improvements since then.
- It has been necessary to match the improvements of FICON on the mainframe with improvements of FICON flowing along a Fibre Channel storage network.
- This is an overview of some of those storage networking improvements.

Bottleneck Detection

Identify Slow Draining Devices and Storage Network Congestion

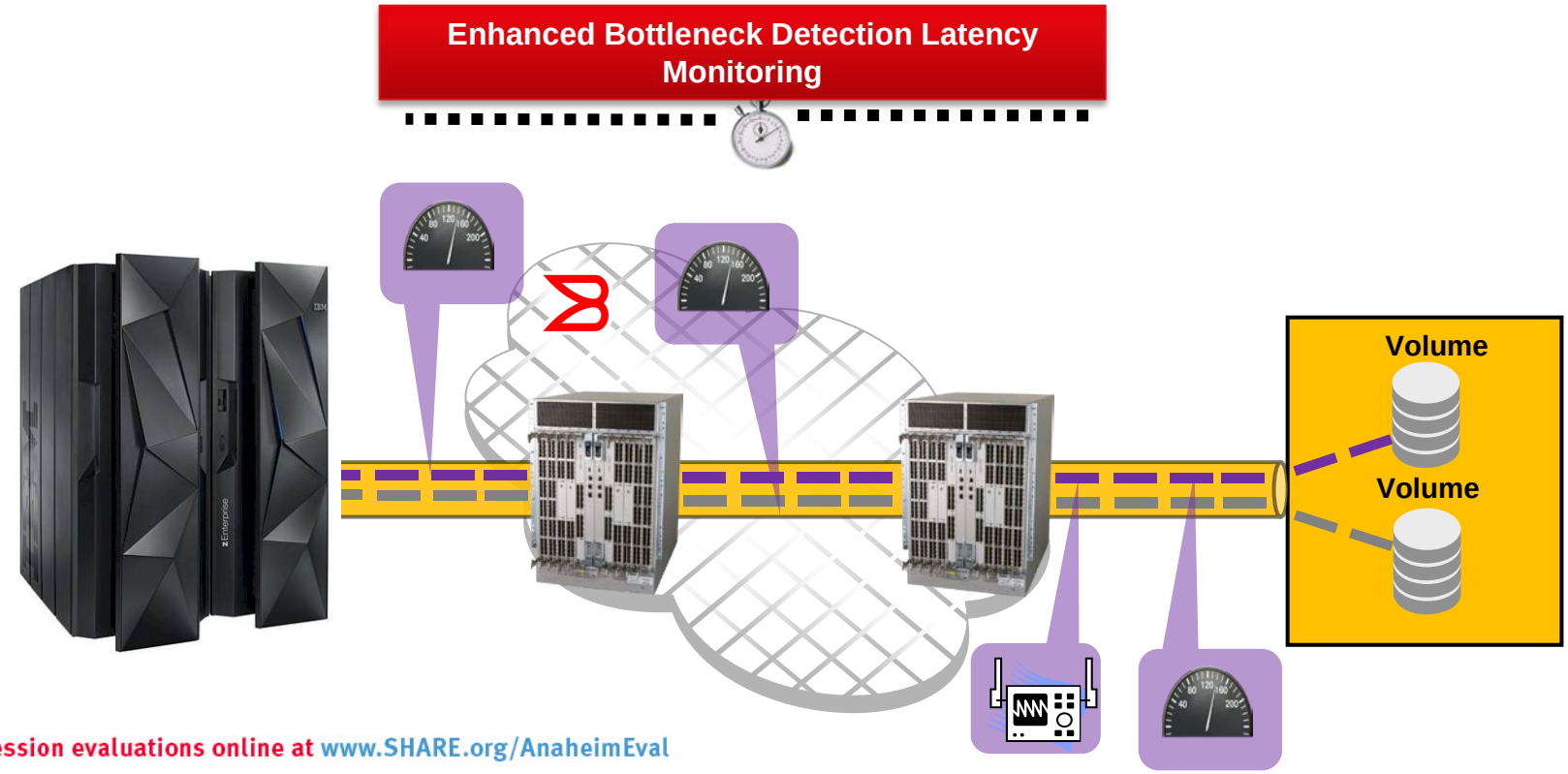
- The bottleneck detection feature identifies ports where the offered load is greater than the achieved egress throughput
- Bottleneck detection enables a user to:
  - Prevent degradation of throughput in the fabric
  - Reduce the time it takes to troubleshoot network problems



Bottleneck Detection

Quickly Identify and Resolve FICON Performance Degradation

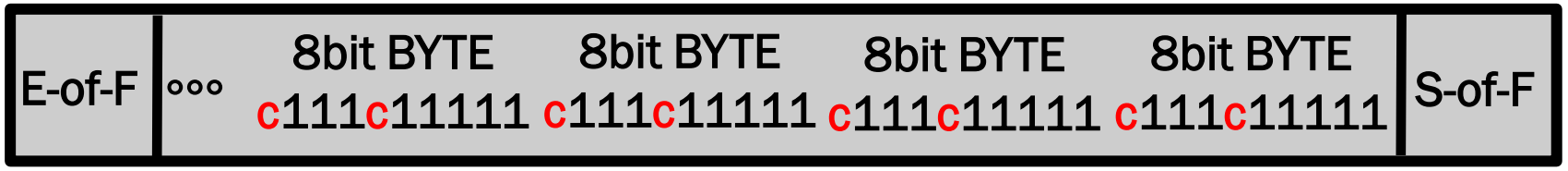
- Monitor for latency in the I/O fabric with Bottleneck Detection and receive notification when problems are detected
- Monitor resource contention, congestion, and other issues impacting user application performance and FICON performance
- Alert the user when thresholds are exceeded and display data on a Dashboard



Significant Change In Data Encoding

8b/10b compared to 64b/66b

8b/10b (since 1950s but patented in 1983)

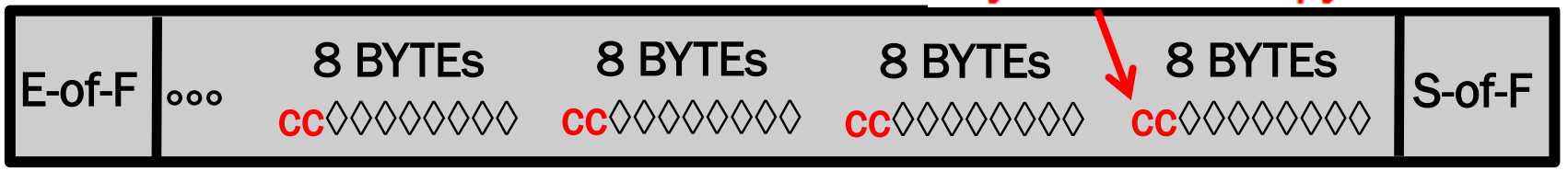


Payload Area of Frame – up to 2112 bytes of data

8b/10b: Each 8 bit Byte becomes a 10 bit Byte – 20% overhead

64b/66b (available since 2003)

Sync Header – copy/save hi-order bit



Payload Area of Frame – up to 2112 bytes of data

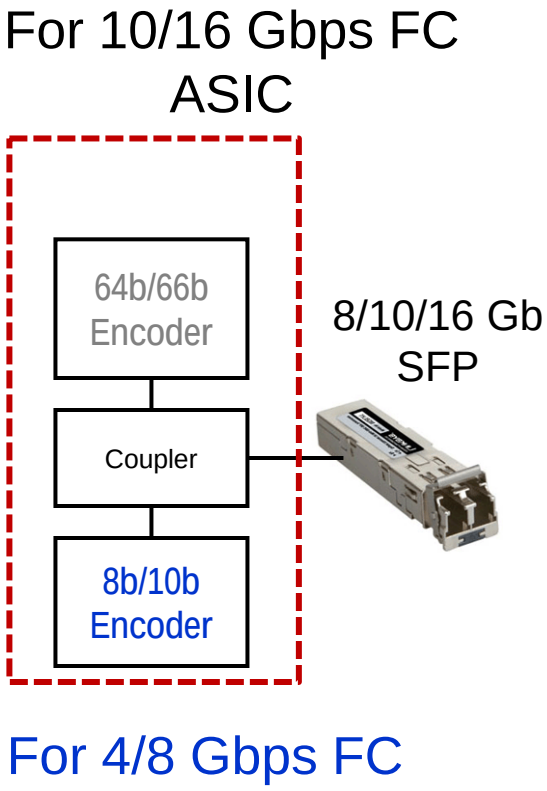
64b/66b: Two check bits are added after every 8 Bytes – 2% overhead

- At the end of every 32, eight byte groups, we have collected 32 hi-order ck bits
- This is a 32 bit check sum to enable Forward Error Correction to clean up links

- 1/2/4 and 8Gbps will always use 8b/10b data encoding
- 10Gbps and 16Gbps will always use 64b/66b data encoding

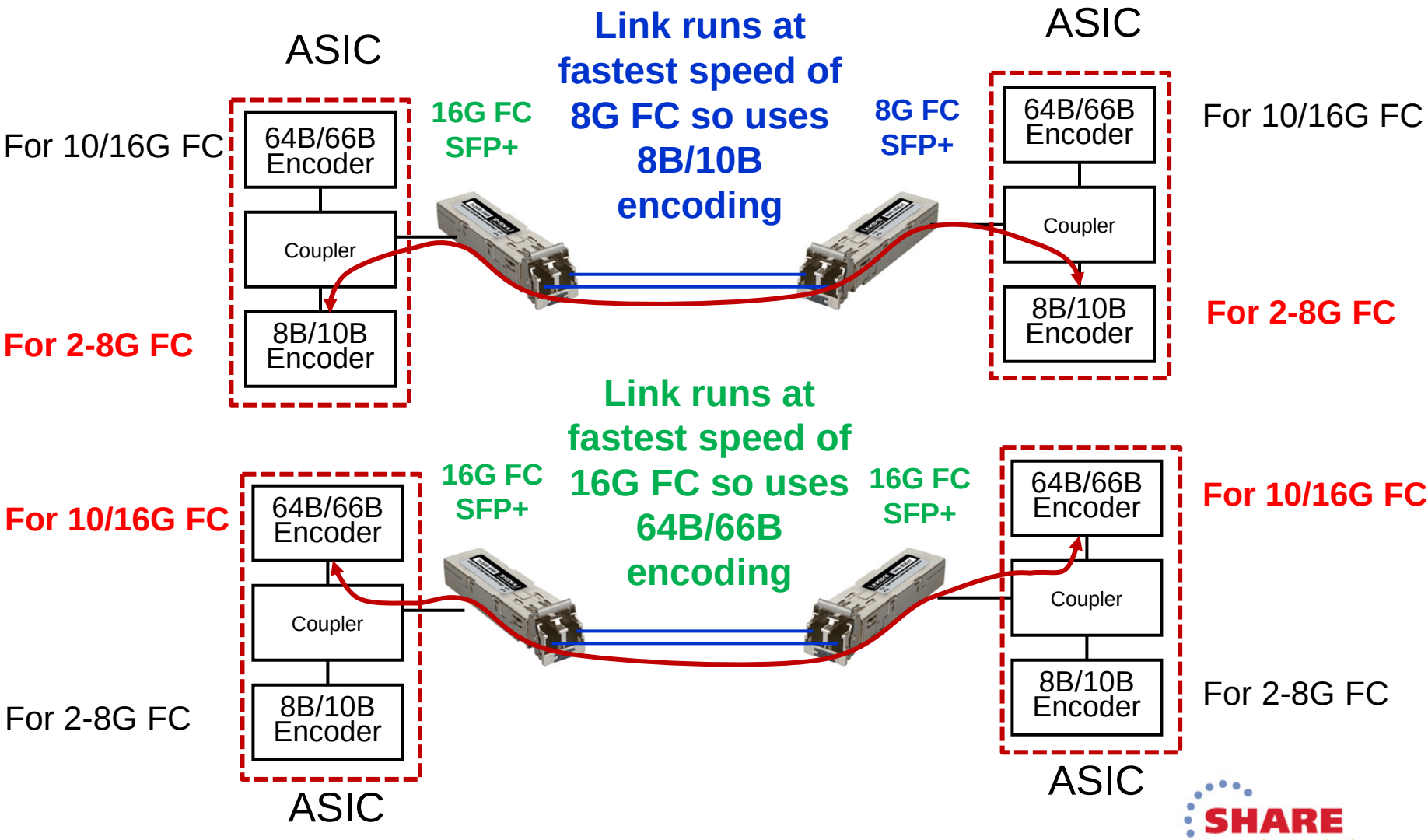
Multi-speed Encoders Along With A Significant Change In Data Encoding

- To improve the efficiency of the FC protocols, 10G and 16G optics use only 64b/66b coding which is 98% efficient
  - Encoding is done by the ASIC and depends on which optic is being utilized in that port
  - 8b/10b coding is used for 1/2/4/8G FC and is 80% efficient
  - 16G signals cannot use the 8b/10b encoders
- To be backward compatible with 4/8G FC the 16G FC **ASICs** must support both 8b/10b and 64b/66b coder/decoders (codec) on each link
  - 1Gbps and 2Gbps are not supported at 16Gbps
- During speed negotiation, the transmitter and receiver switch back and forth between the speeds (and the corresponding codecs) until the fastest speed is reached for a given link



Multi-speed Encoders – Example

This example assumes the connection of 2 switches



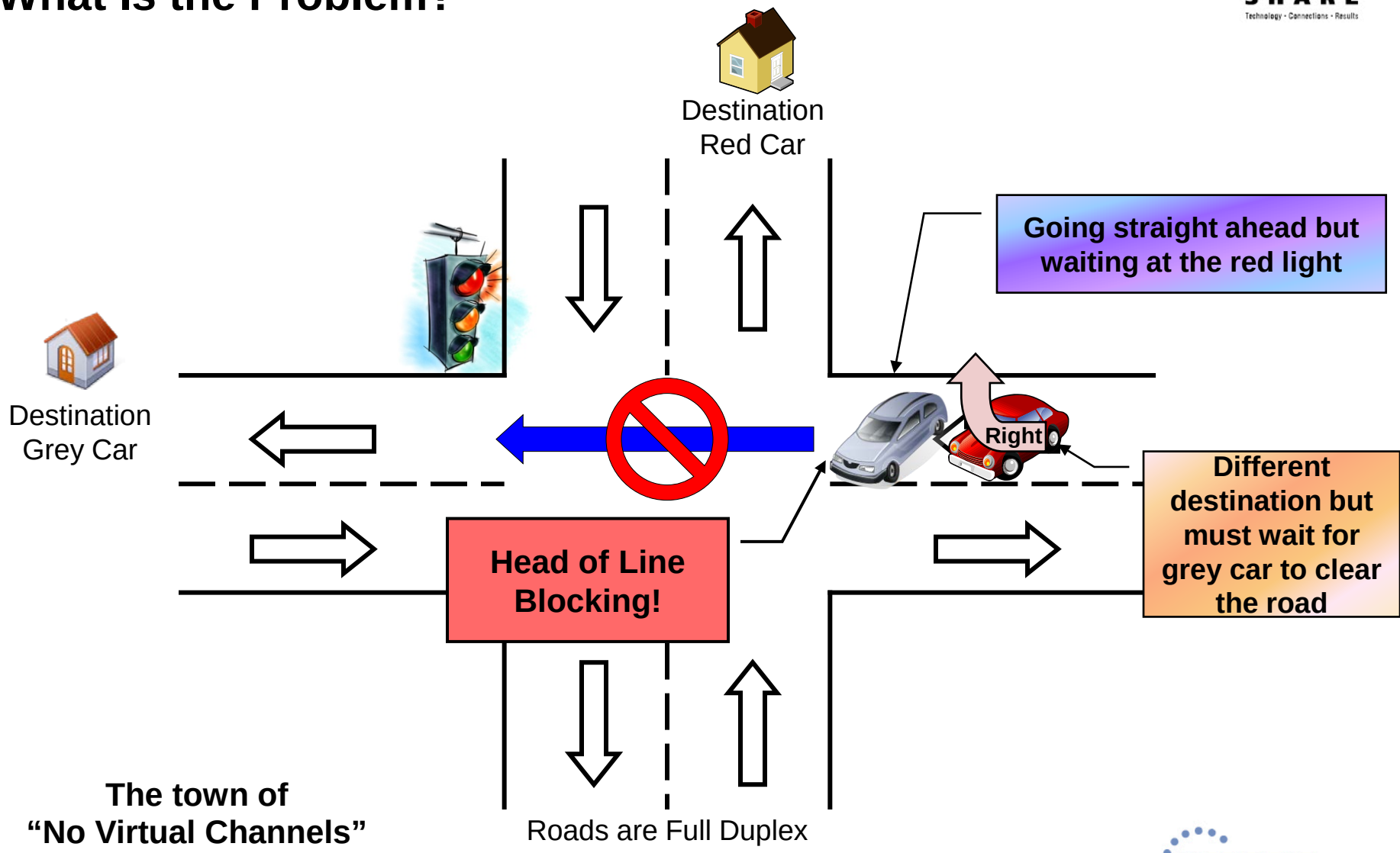
Forward Error Correction

- ASIC-based functionality that is enabled by by default and allows the hardware to fix bit errors in a 10G or 16Gbps data stream
  - Works on Frames and on Primitives
  - The high-order bits collected from the 64b/66b data encoding help correct transmission bit errors
  - Corrects up to 11 bit errors per each 2,112 bits in a payload transmission
    - *11 bit corrections per 264 bytes of payload*
- Requires hardware that is capable of performing this function for its E_Ports
- Significantly enhances reliability of frame transmissions across an I/O network

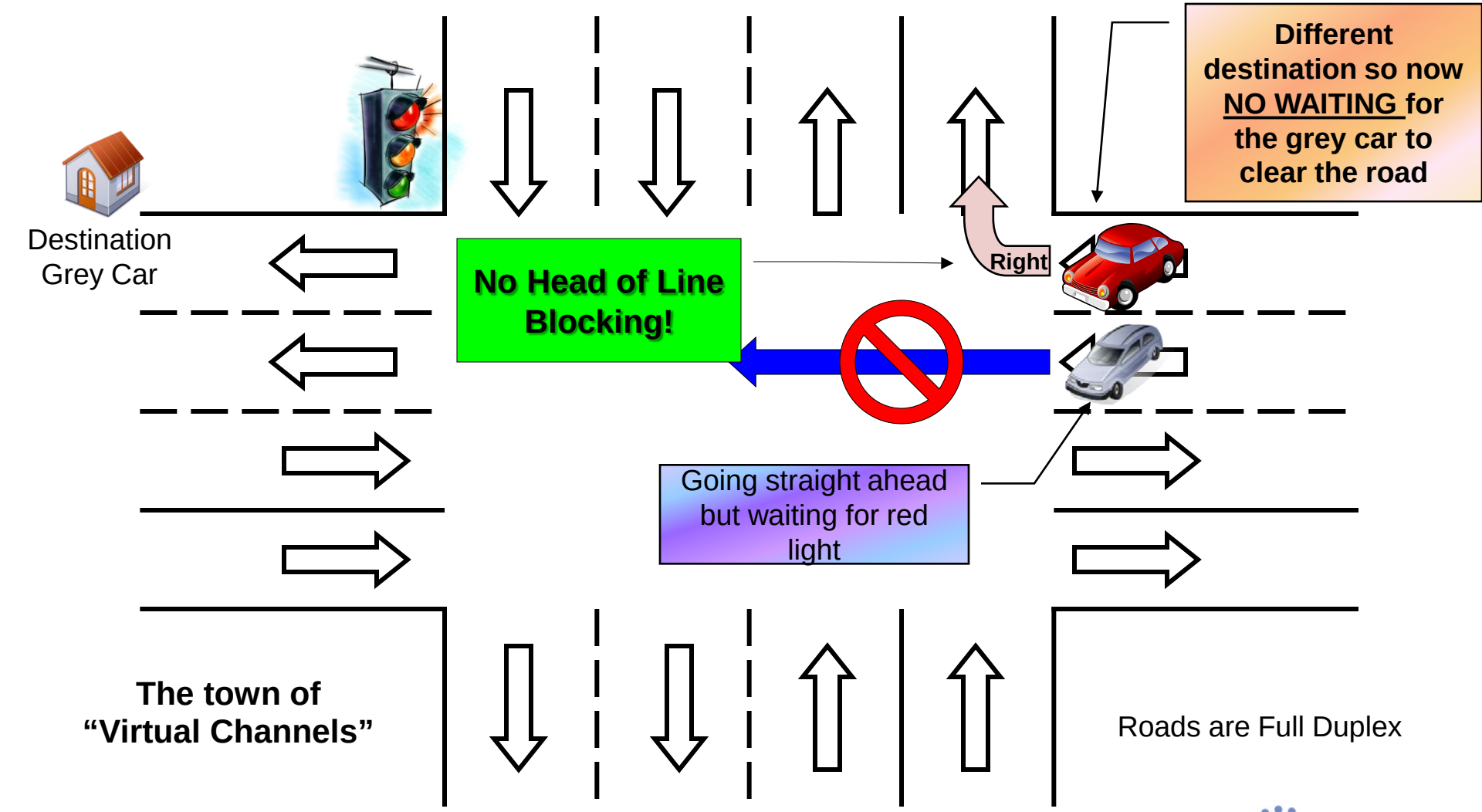
Storage Network Super Highways

- Even when ISL links are being used very well, **Head of Line Blocking** on an ISL link can create congestion and performance problems much like toll booths on a Super Highway can cause congestion and slow travel.

Head of Line Blocking
What Is the Problem?

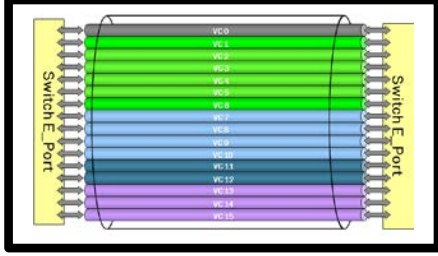


Head of Line Blocking
What Is the Problem?



Virtual Channels (VCs)

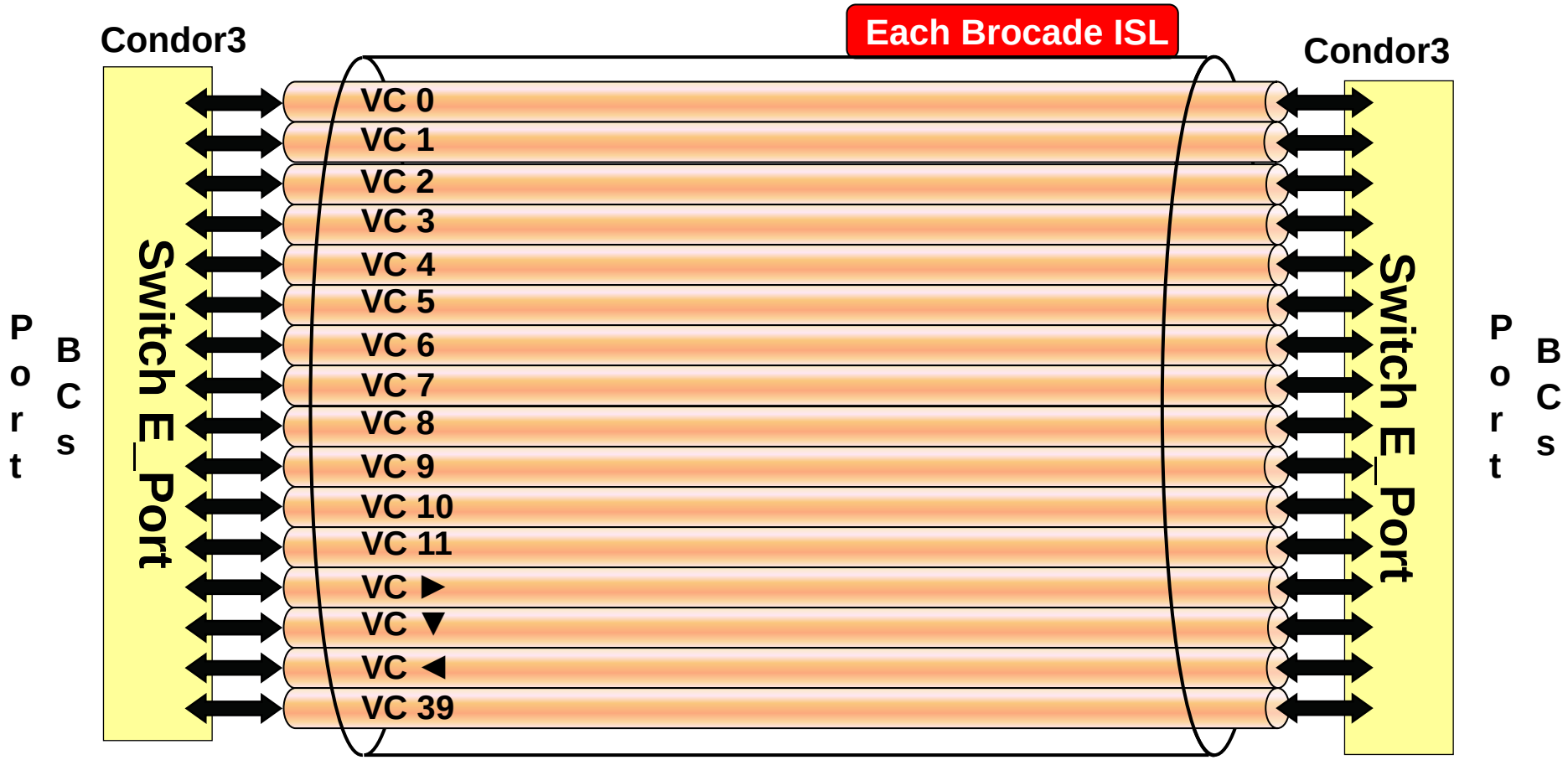
Buffer Credit Pool segments established on each ISL link



- Virtual Channels is a unique feature available on every Brocade switch/Director.
- VC technology logically partitions the buffer credits on each ISL link into many different BC pools (virtual channels) and then prioritizes traffic to optimize performance and prevent head of line blocking.
- In addition, each VC has its own queues.
- *Of course an ISL is still just one fibre link so only a single lane (a single frame) of I/O traffic is passing across it, in each direction, physically at a time*:
  - A VC is really just a smaller allotment segment of the total buffer credit pool that becomes dedicated to servicing a specific VC number.
  - Segmenting the single physical buffer credit pool into multiple pools allows frames to avoid HoLB and keeps the ISL link optimized

Virtual Channels help create a SAN Super Highway

Buffer Credit (BC) Pool segments provide multiple lanes of traffic on 1 link



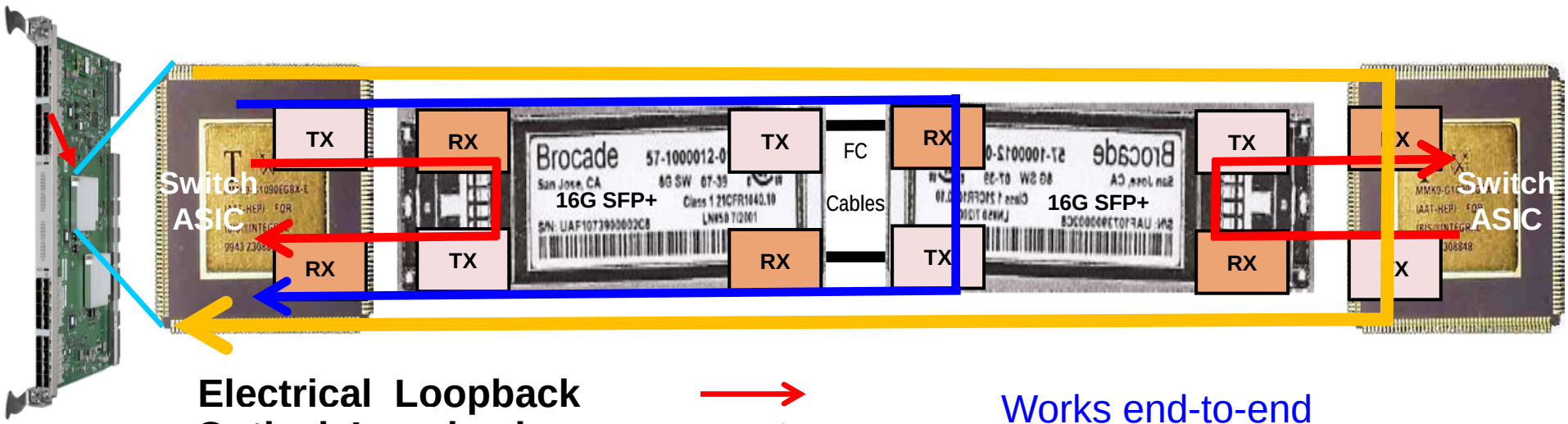
The Port BC pool is logically segmented into groups of virtual BCs (virtual channels) and then ingress ports can be assigned to a virtual channel segment if an ISL is required to be used.

Testing ISL Links Before Deployment

Reduce the time it takes to successfully deploy a fabric

- Diagnostic Port will check optics and cables integrity
  - D_Port is a special port type, configured by the user to run diagnostics
  - Does not carry any FC control or data traffic
- Supported only on ISL and ICL ports

- Full support for 16G SFPs
- Partial support for 10G SFPs
- Provides the following capabilities:
  - Performs electrical loopback (16G)
  - Performs optical loopback (16G)
  - Measures link distance (10G, 16G)
  - Does link saturation testing (10G, 16G)



Electrical Loopback
Optical Loopback
Link distance/saturation

Works end-to-end
even through DWDM!

I/O Flow Generator

Better testing of ISL links before they are deployed

- Once D_Port has been executed, users might want to consider using I/O Generator to create a saturation flow across the ISL link to test for a stable link
- Generates traffic at line rate across ISL links for testing purposes.
- For System z customers, the expected use for this feature is to validate inter-switch links (ISLs), especially through DWDM and over leased lines.
- Previously, the only feature to generate traffic to validate links was the Diagnostic Port (D_Port) function but D_Port was not designed to actually stress those links.
- Flow Generator is a complimentary tool to stress links at full line rate to make sure that they are up to the task of delivering I/O successfully throughout the fabric.

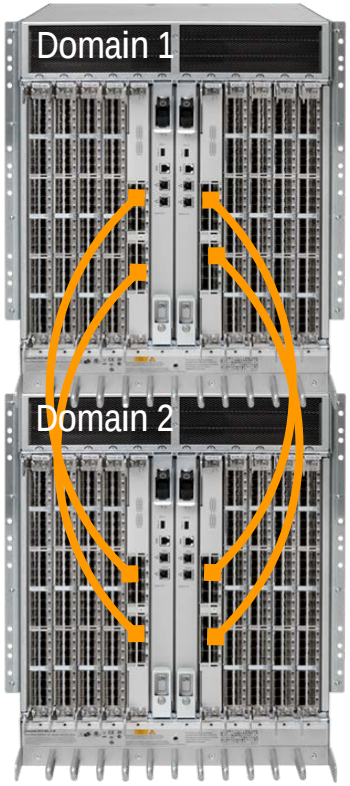
I/O Flow Monitoring

Reduce the time it takes to resolve congestion/backpressure issues

- Flow Monitoring provides performance analysis based on a specific CHPID-to-Link address path.
  - Multiple Flow Monitors can be created.
  - All Flows to a given port can be created automatically so performance at a specific port can be analyzed by channel path.
- The ability to automatically configure a Flow Monitor for all paths to a given port can be a valuable troubleshooting tool to determine where traffic to a port originates.
- For example, an HCD mistake that results in two CHPIDs from different CECs contending for the same tape port is easily found by setting up a Flow Monitor on the tape port to automatically monitor all flows.
- Using switch management software, a single mouse click brings the user right to the port information with the RNID data that articulates the CEC S/N, CHPID, and CSS.

Inter-Switch Links and Inter-Chassis Links

QSFP Connections



“Hop of no concern”

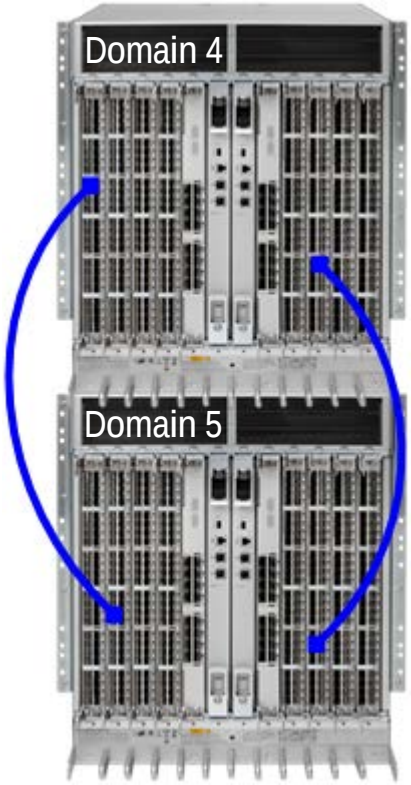
ICLs

Do not create an official hop

But

Do connect domains together

SFP Connections



Creates a FICON hop

ISLs

Do create an official hop

And

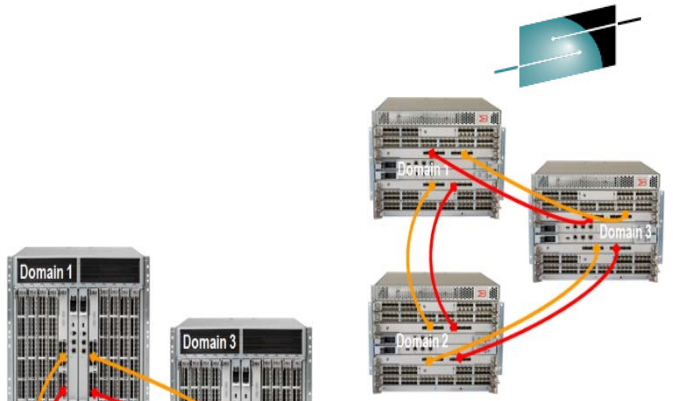
Do connect domains together

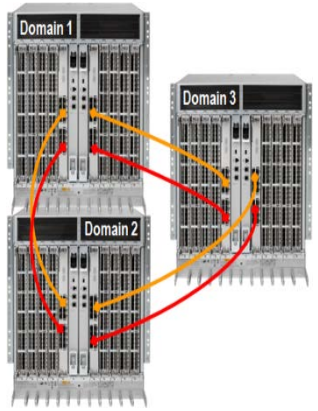
Unique Fabric Scalability

FICON Triangle Configuration

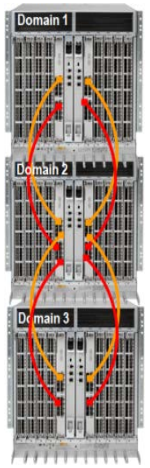
- Chassis expansion scalability through an ability to provision up to 768 ports of 16 Gbps bandwidth, within a local fabric through the use of ICLs:
  - *Three 256 port chassis can be connected together providing 768 ports with from 128 – 2048 Gbps of ICL bandwidth (three 256p chassis = 768 ports)*

- Chassis expansion scalability through an ability to provision up to 1,536 ports of 16 Gbps bandwidth, in a cascaded FICON fabric through the use of ICLs and ISLs:
  - *Three 256 port chassis can be interconnected with ICLs and then ISL connected to another pod of three 256 port chassis (6 * 256 = 1,536p)*
  - *The example to the right shows a 1,024 FICON port fabric*





Using MPO cables, can be up to 100 m so these chassis can be in widely dispersed cabinets.

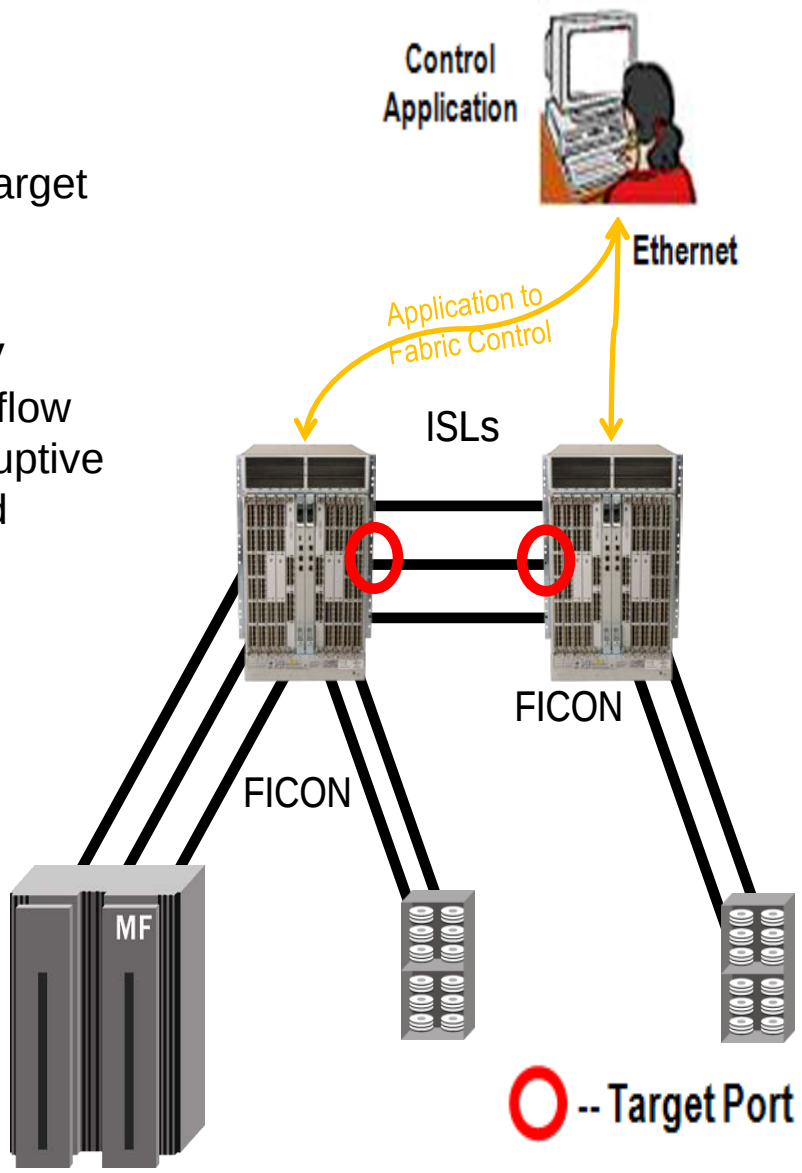


Up to
6,800 miles
11,000 km
(FCIP ISLs)

Port De/Re-commissioning for Mainframes

Improving FICON ISL Management

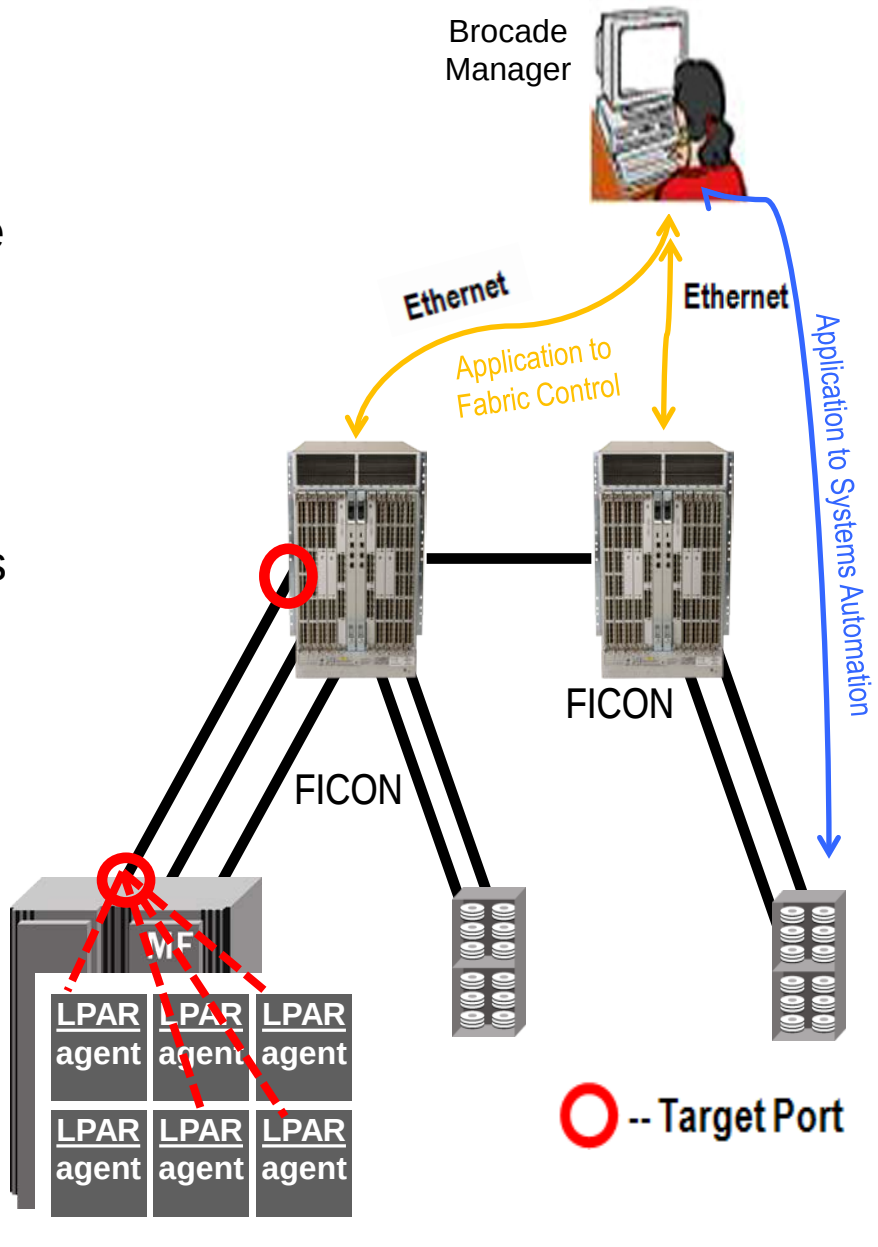
- Coordinate event with external applications
  - Switch operating system moves routes off of the target ISL before that ISL is disabled
- Mechanism to remove an ISL non-disruptively
  - Block/Disable an ISL port after moving the traffic flow to other routes so that removing it will be non-disruptive
  - Requires Lossless DLS to be enabled on both end
- Will become attached to automated processes like Port Fencing in the future



Port De/Re-commissioning for Mainframes

Improving FICON N_Port Management

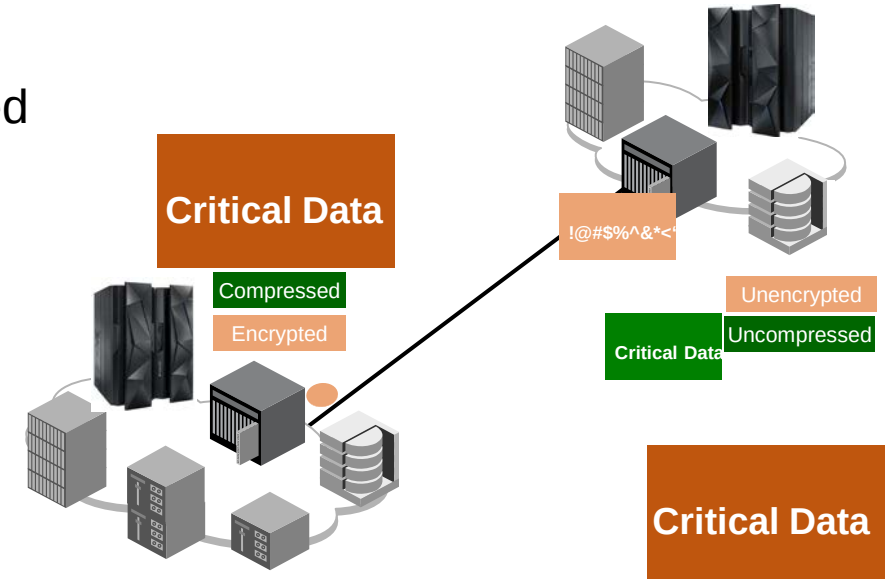
- Mechanism to remove a CHPID or Storage port non-disruptively
  - Block/Disable a device port after allowing each LPAR to quiescing the path/device so that removing it will be non-disruptive
- Coordinate event with external applications
  - Application or system manager
    - Moves workload off of a target port before that port is disabled
- Will become attached to automated processes like Port Fencing in the future
- Each LPAR would run a CIMOM agent



Improved ISL Security and Efficiency

ISL data traffic Compression and/or Encryption In-Flight

- **Secure Transfers**
  - Encrypts data on 4/8/16 Gbps ISLs hosted on Gen 5, 16 Gbps blades
    - Switch-to-switch encryption, not at-rest encryption
    - For Fibre Channel long-distance links
  - Uses AES-GCM algorithm for both authentication and encryption
  - Uses 256-bit encryption key



- **Maximum Network Efficiency**
  - Disk or tape traffic gets compressed on ISL and gets uncompressed at the receiving switch
  - Provides up to 2:1 compression and uses Brocade LZO algorithm
  - Provides up to 128 Gbps of compressed bandwidth per blade

Qualified Configurations:

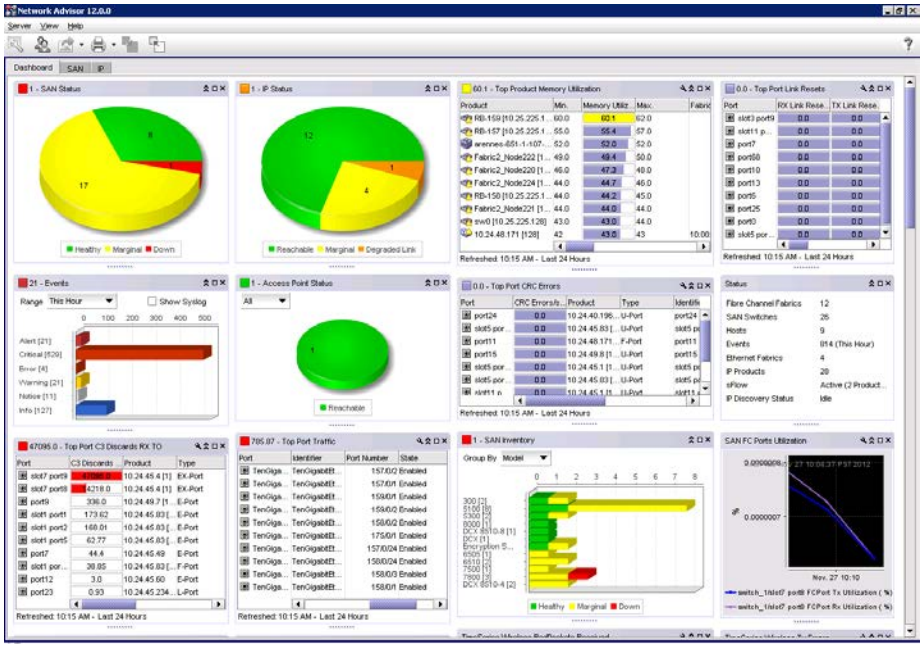
- Use neither capability
- Use both capabilities concurrently
- Use only compression
- Use only encryption

- **Requires no license and can be used in conjunction with in-flight encryption**

Storage Networking Dashboards

Proactive management of Storage Networking

- Customers are demanding Ease-of-Use as they deploy new technologies into their data centers.
- A variety of FC fabric event monitoring and displayable metrics can now provide customers with many advantages:
  - Quick check of the SAN's status
  - Quick check of the IP status
  - Early and better fault detection
  - Proactively detect deteriorating SFPs
  - Avoid down time and replace in a scheduled maintenance window.
  - Apply predefined parameters across all fabrics within an enterprise with a single mouse click

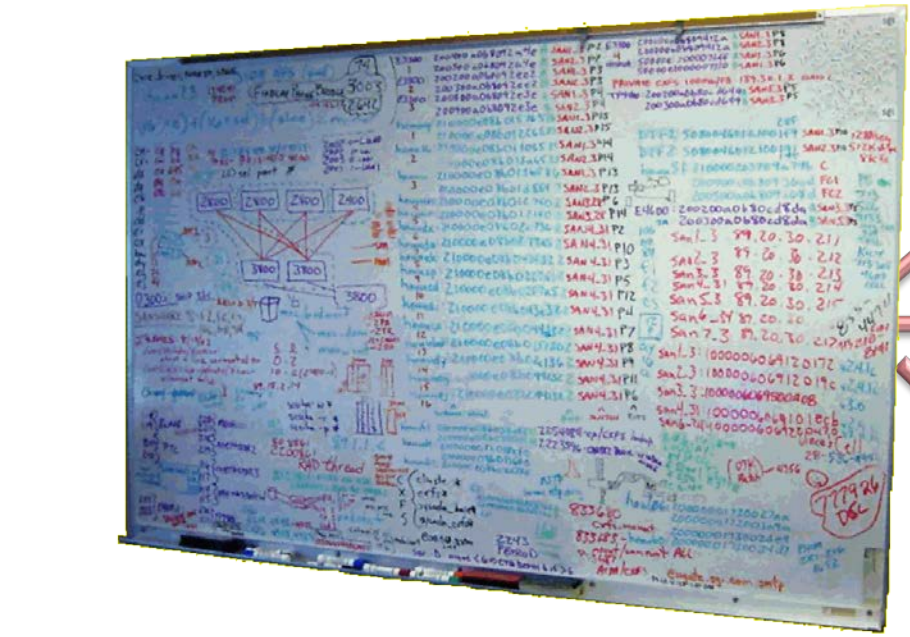


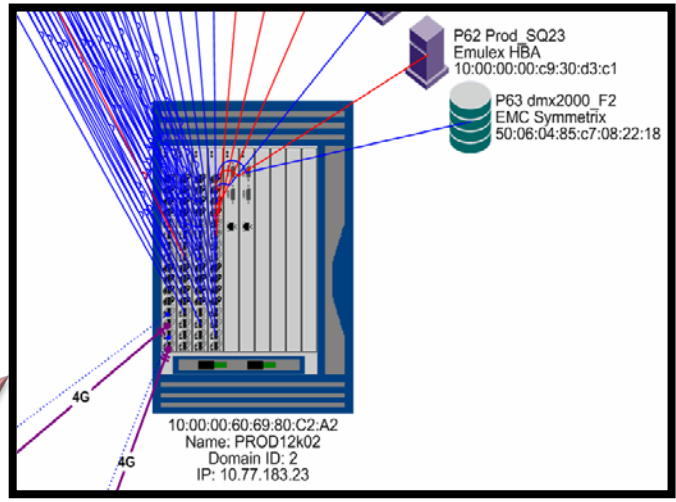
Health and Performance Dashboards provide a color coded "quick look" at many storage network metrics

Start With A SAN Health Check-up

For FCP and FICON I/O Environments

Simple and effective way to stop whiteboard SAN management





SAN SUMMARY DETAILS FOR SAN_EXAMPLE

Table Of Contents

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

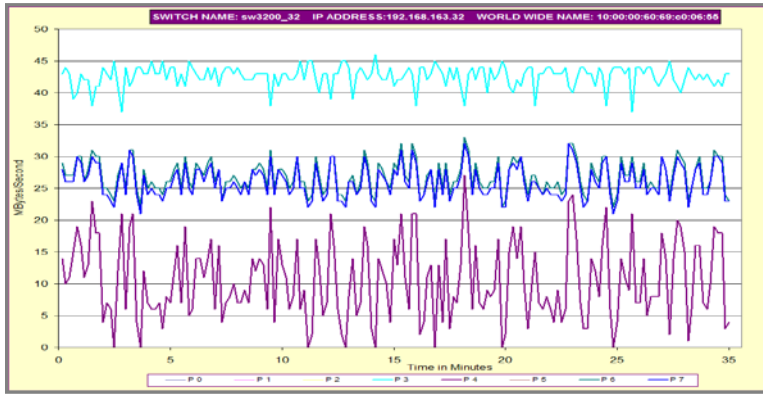
Fabric Name	Switch State		Power Supplies			Fans			Temp Sensors			Errors		SNMP		SysLog	
	Marg	OK	Bad	Marg	OK	Bad	Marg	OK	Low	OK	High	Lvl1	Lvl2	No	Yes	No	Yes
Storage_Edge	0	3	2	0	2	0	0	12	0	12	0	0	0	3	0	3	0
Server_Edge	3	2	1	0	7	0	0	19	0	17	0	0	0	5	0	5	0
TOTALS	4	5	4	0	10	0	0	34	0	35	0	0	0	9	0	9	0

PORT USE

Fabric Name	Port Use						Fan Out Ratios			Port Long Distance Modes				
	Disk	Tape	Host	ISL	Free	Total	Host:Disk	Port:ISL	Device:ISL	10km	25km	50km	100km	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	Aliases Statistics				Zone Statistics				Config Statistics			
	Database Use	Aliases	AvMem	MaxMem	Hanging	Zones	AvMem	MaxMem	Hanging	Configs	AvMem	MaxMem	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



- Topology diagrams
- Comprehensive reporting
- Performance graphs
- Health and best practice checks

SAN Health Check-up Tool

- Five Minute Process
  - One Minute to install
  - Four Minutes to audit fabrics
- Encrypted upload to secure servers located at Brocade HQ
- Detail report returned within a couple of hours via email

Brocade SAN Health

SAN Health Diagnostics Capture

This program captures diagnostic and performance data from SAN switches to aid in the generation of a SAN Health report containing inventory, topology and performance information.

Only commands that retrieve information from the switch are used.

No switch configuration settings are changed.

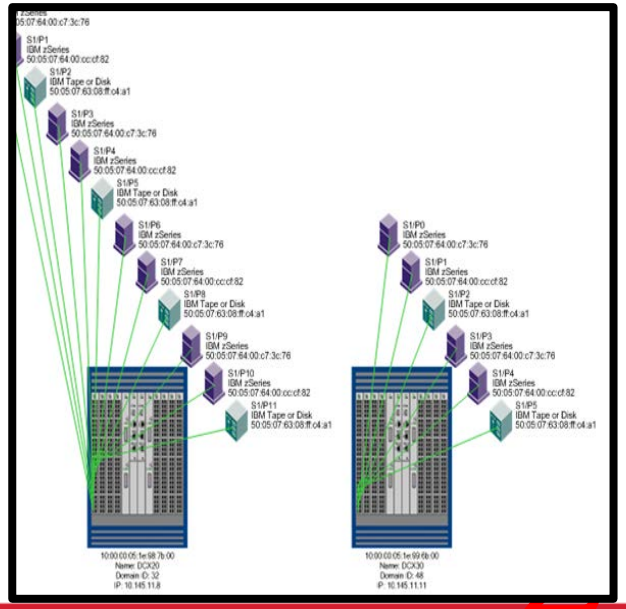
Version 3.2.2
Supported products and firmware levels SAN Health Support Matrix
Contact the SAN Health admin team on SHAdmin@brocade.com
For more information visit the SAN Health home page www.brocade.com/sanhealth

Compatible with:

- Brocade
- Cisco
- McDATA

Report Content:

- Visio Diagram
  - Detailed diagrams of all fabrics and end devices
- Excel File
  - Reports on traffic spikes and over-subscription ratios
  - Verifies Zoning and switch setting consistency
  - Flags any potential SFP and cable errors
  - Provides detailed licensing and capacity planning information



Let Me End This Session By Proudly Presenting The Industry's ONLY FICON Certification

Brocade
Certified Architect
for FICON

Mainframe-centric Classes We Teach

This FICON Certification is Unique in the Industry

Fundamentals of Brocade Mainframe Networking Seminar taught over 2 full days or 3 relaxed days

- We have been holding classes since mid-2008
- This is good for mainframers who desire to become professionally certified as FICON subject matter experts
- This seminar teaches advanced concepts and is not well suited for professionals with less than 1 year of experience

Total number of attendees at these seminars since 2008: **~550**
Total number of Brocade FICON Certifications awarded: **250+**

(There is so little FICON training available that some people attend just to get current on FICON and not for certification)

Brocade Certified Architect for FICON (BCAF)

Brocade can provide this 2 or 3 day training for YOU!

Fundamentals of Brocade Mainframe Networking Seminar

(the preparatory class for the BCAF certification)

Day 1 (8:30am to 5pm):

- Course Introduction
- Mainframe and FICON Overview
- Brocade Switching Technologies
- Design and Migration
- FICON cascading and data transmission technologies

Day 2 (8:30am to 5pm):

- Managing cascaded FICON environments
- FCIP, Data Replication and Business Continuity Networks
- FICON Implementation
- Managing and Maintaining a FICON Environment

Can be done In 3 easier days!

- Class minimum size is 8 people
- No charge if taught in the USA
- $2,500 fee for travel and living if taught overseas

- Contact David Lytle if interested: dlytle@brocade.com

SHARE in Anaheim

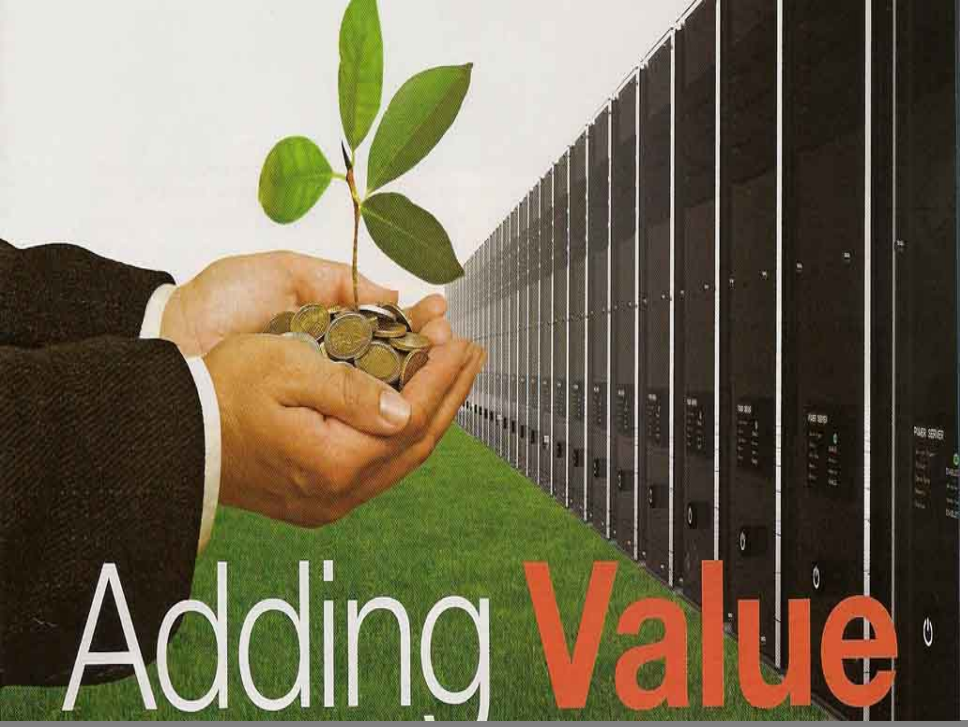
Some Mainframe Social Media Sites

Please check out the Brocade Mainframe Solutions blog today:
http://community.brocade.com/t5/Mainframe-Solutions/bg-p/MainframeSolutions

Find useful information on the Brocade Fibre Channel Communities Page today:
http://community.brocade.com/t5/Fibre-Channel-SAN/bd-p/fibre

Almost 300,000 hits!

Join us in discussing issues at the Mainframe Discussion Community today:
http://community.brocade.com/t5/Mainframe-FICON/bd-p/mainframesolutionsforum



Your Eval

My Reaction!

5 = “Aw shucks. Thanks!”
4 = “Mighty kind of you!”
3 = “Glad you enjoyed this!”
2 = “A Few Good Nuggets!”
1 = “You Got a Nice Nap!”

David Lytle, BCAF
Principal Engineer
Brocade
dlytle@brocade.com

Session 14375