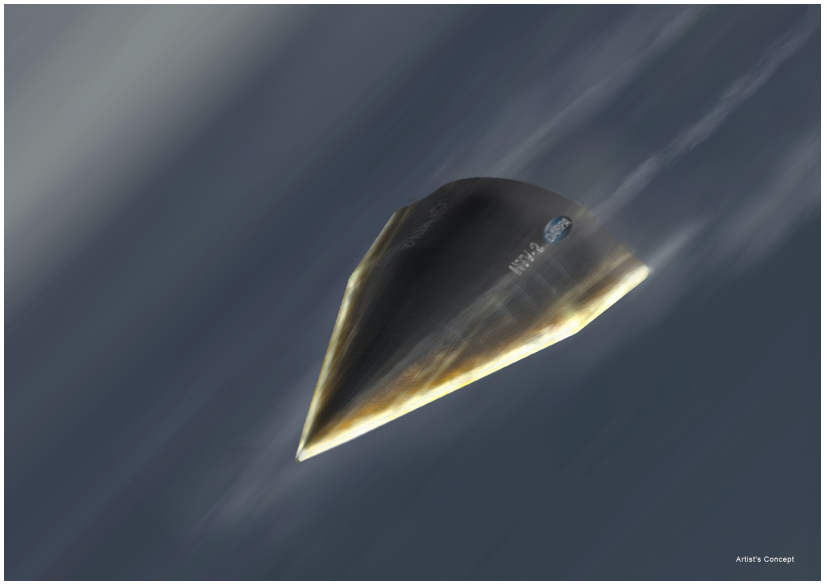
STORAGE 101

LAN to SAN in 90min

Agenda

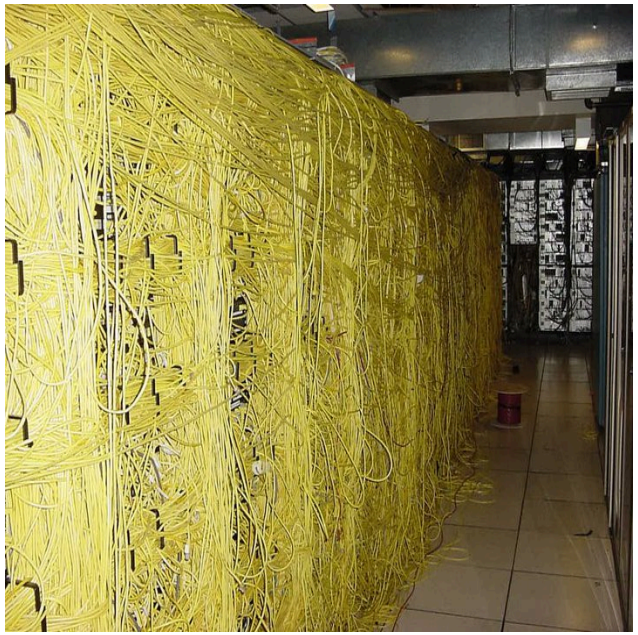
- Basics
- Open Systems Architecture Basics
  - I/O Subsystem
  - Block vs. File Storage
    - SCSI Basics
- SAN Primer for LAN Professionals
  - Basic Concepts
    - LAN vs. SAN Comparisons
    - FC Initialization Walk-Through
- Data Center Bridging and FCoE
- The Future



NEVER UNDERESTIMATE THE BANDWIDTH OF A FAST VEHICLE FULL OF TAPE

Definitions

Definitions can very from Customer to Customer



Sometimes SAN just means Fibre Channel

Sometimes SAN includes all multi-attach storage

Sometimes SAN is considered the Array

Sometimes NAS/NFS is considered a SAN

Sometimes even direct connect FC Arrays are called SANs

Safe bet is that SAN means Servers sharing Storage Arrays

STORAGE PROTOCOLS

FC/FC-AL/FC-SW

ESCON/FICON

Infiniband

CIFS

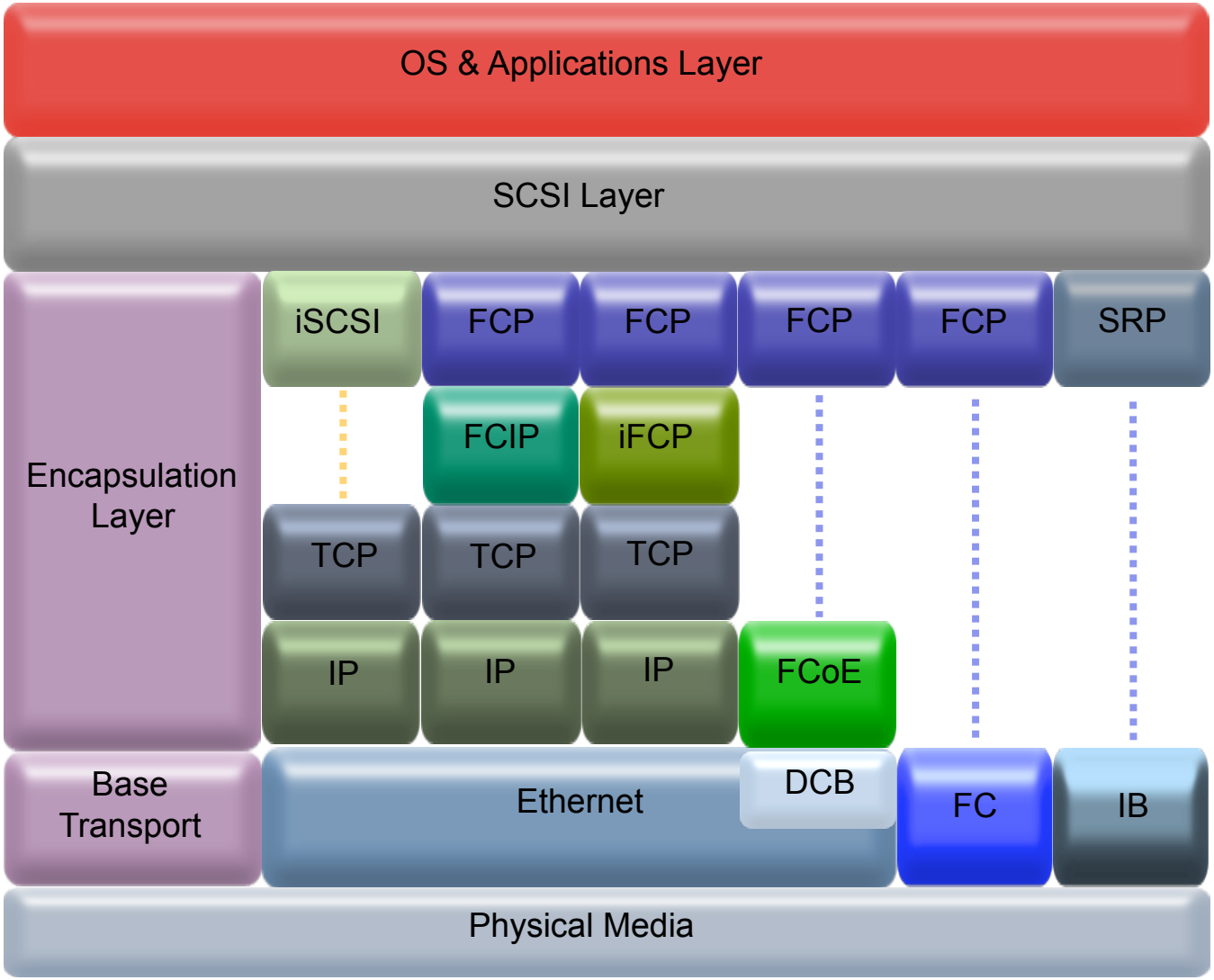
NFS*

FCoE

FCoTR

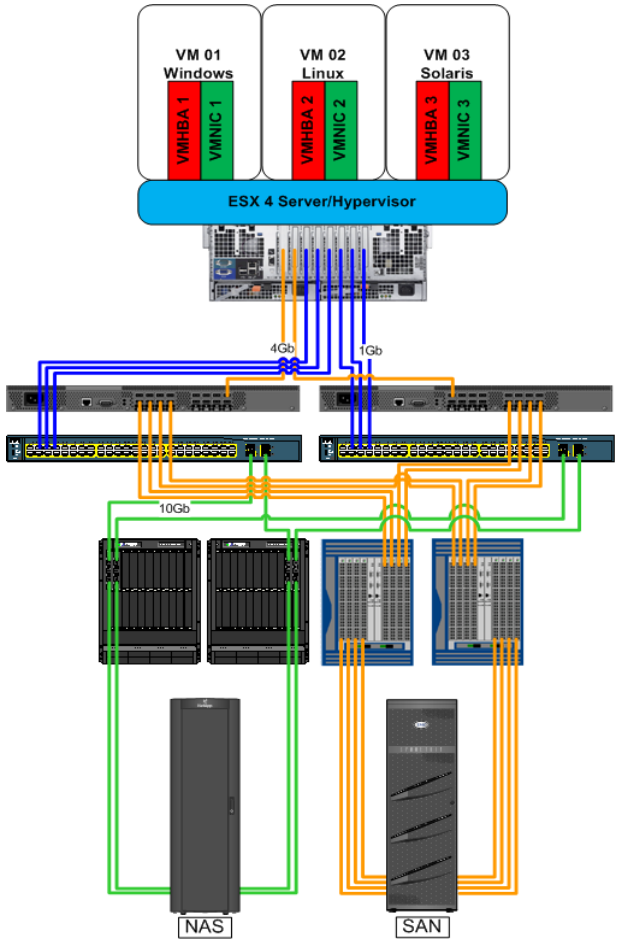
Protocol Comparisons

iSCSI, FCIP, iFCP, FCoE, FC, SRP



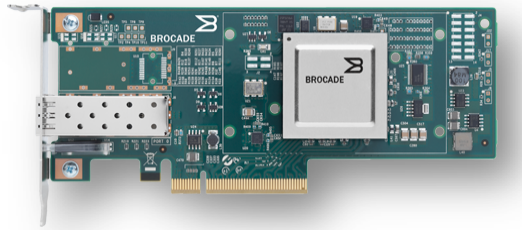
STORAGE : The Usual Suspects

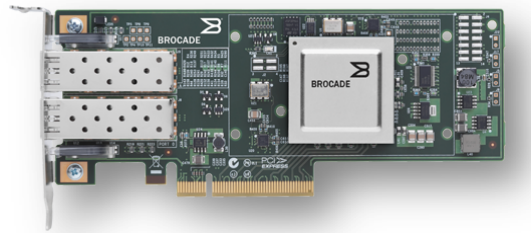
- Hosts
- HBAs, NICs, HCAs, CNAs
- Edge Switches
- Core/Director Switches
- Modular Storage
- Frame Storage



NICs, HBAs, CNAs, HCAs, Mezzanine Cards

Cards that go inside servers of all types

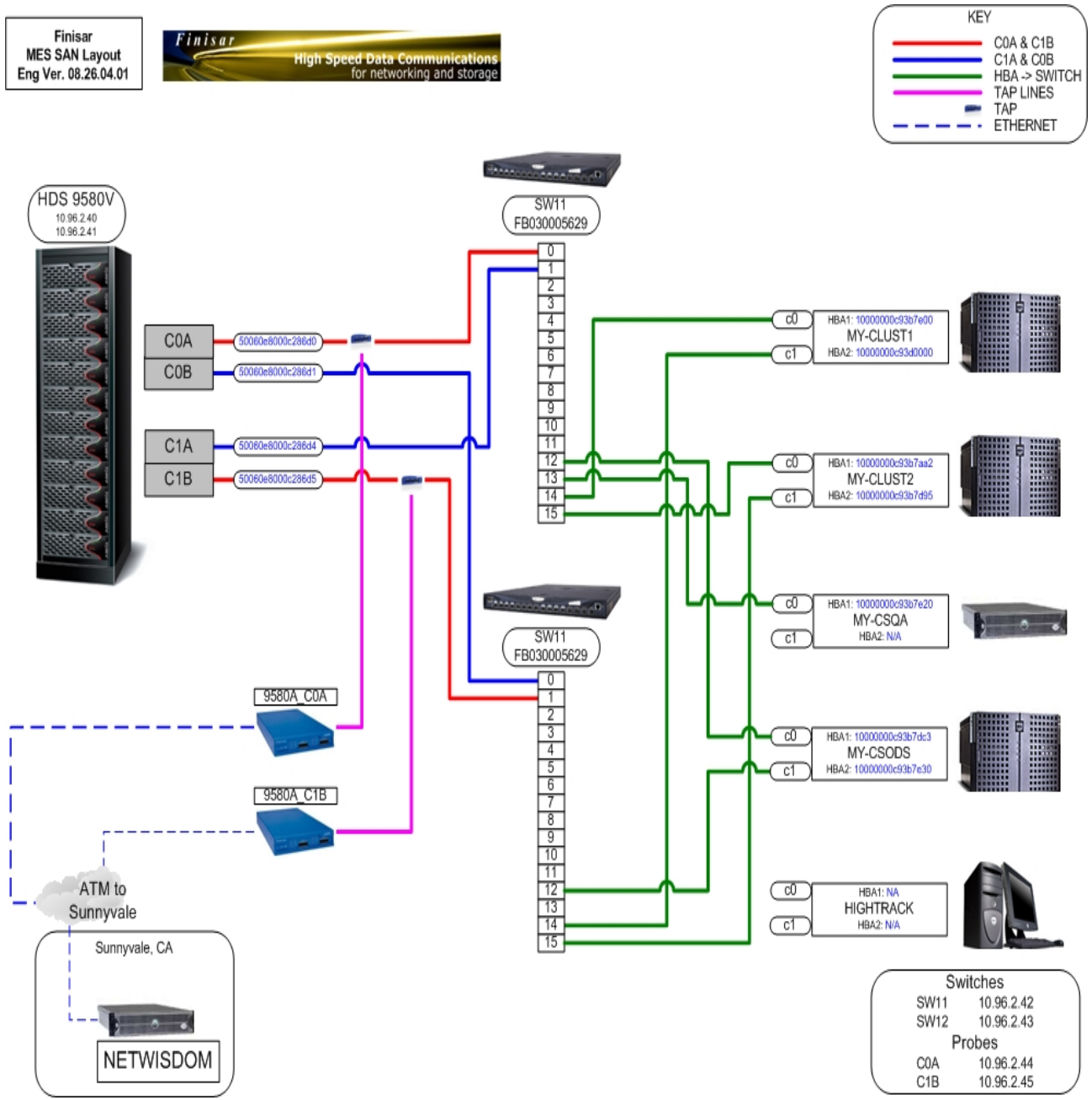


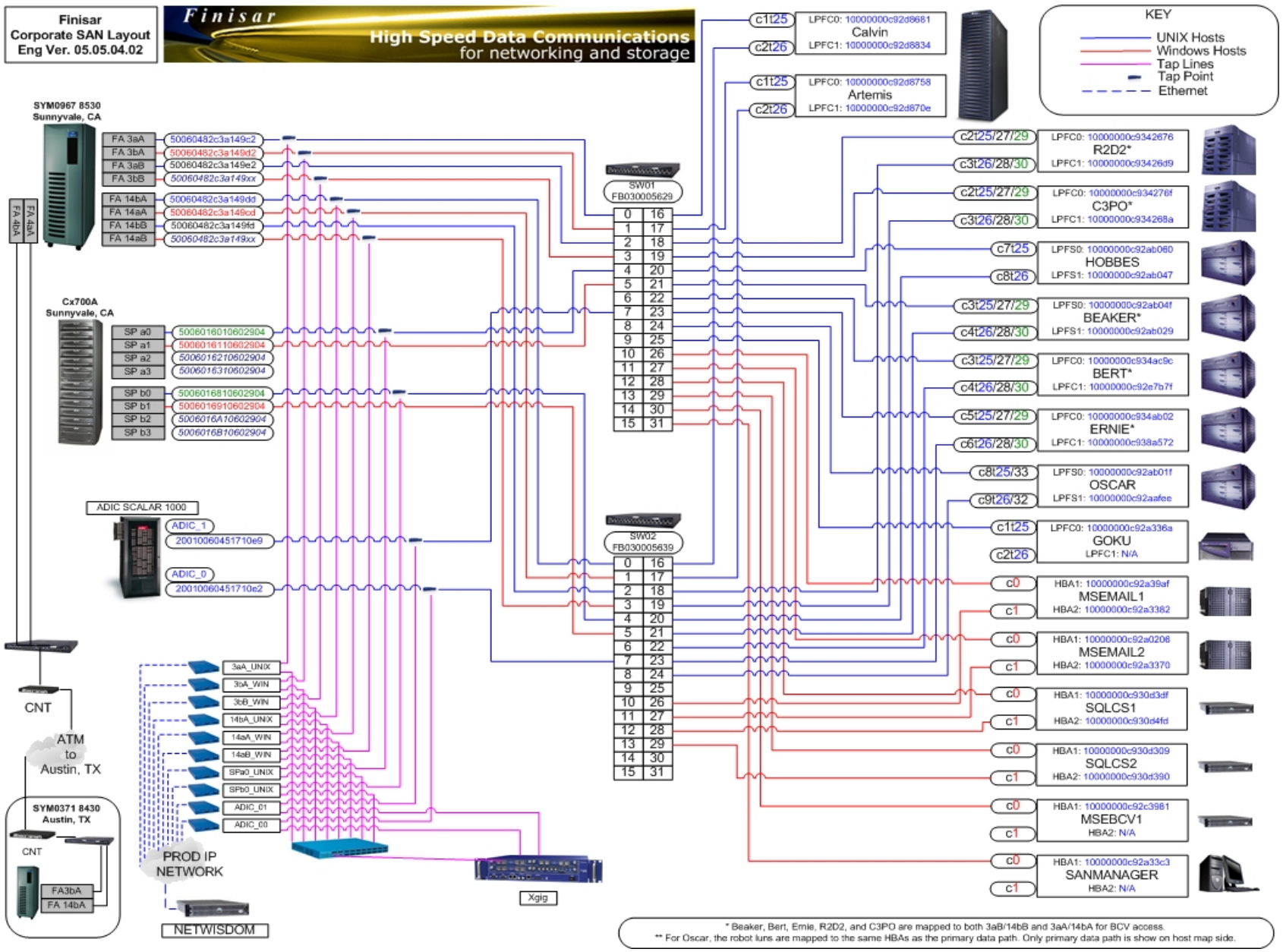


Fixed Port & Director Class Switches

Storage Arrays

Modular vs. Frame





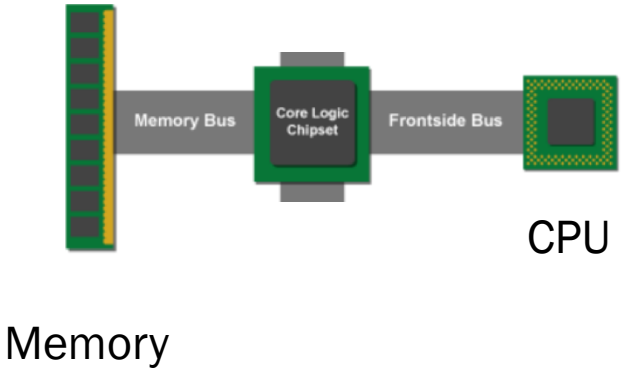
Basic System Architecture

A Simplified Look at Open Systems

Basic System Architecture

The Hardware

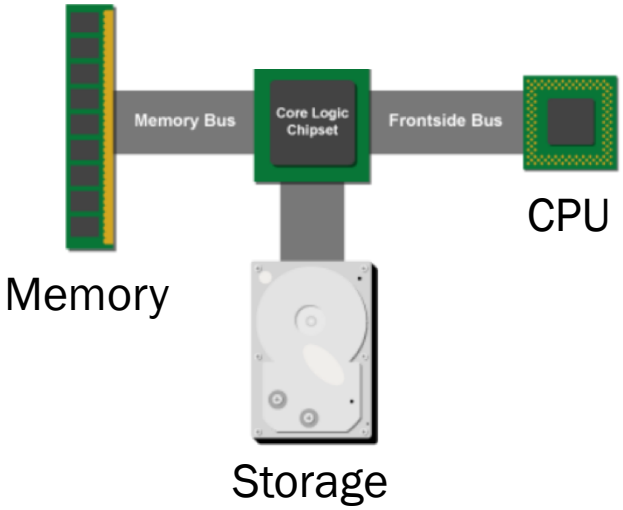
- Basic Intel Architecture
  - CPU and Memory connected via hi-speed bus
  - Core Logic Chipset divided into several tasks.
    - Manage Memory Bus and Frontside Bus
    - Manage connection point between PCIe switched bus and CPU/Memory Bus



Basic System Architecture

The Hardware

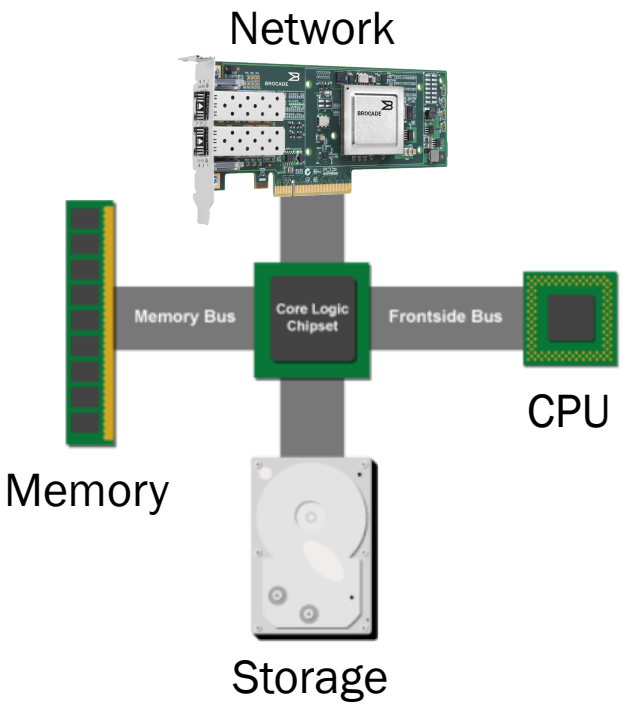
- Basic Intel Architecture
  - CPU and Memory connected via hi-speed bus
  - Core Logic Chipset divided into several tasks
    - Manage Memory Bus and Frontside Bus
    - Manage connection point between PCIe switched bus and CPU/Memory Bus
- Persistent Storage
  - Connected to system bus via hi-speed PCIe switch
    - PCIe is a switched architecture, not a shared bus like old legacy “PCI” systems
    - Each “lane” of PCIe is 2.5Gbps (An 8-lane PCIe device has 20Gbps bandwidth available.)



Basic System Architecture

The Hardware

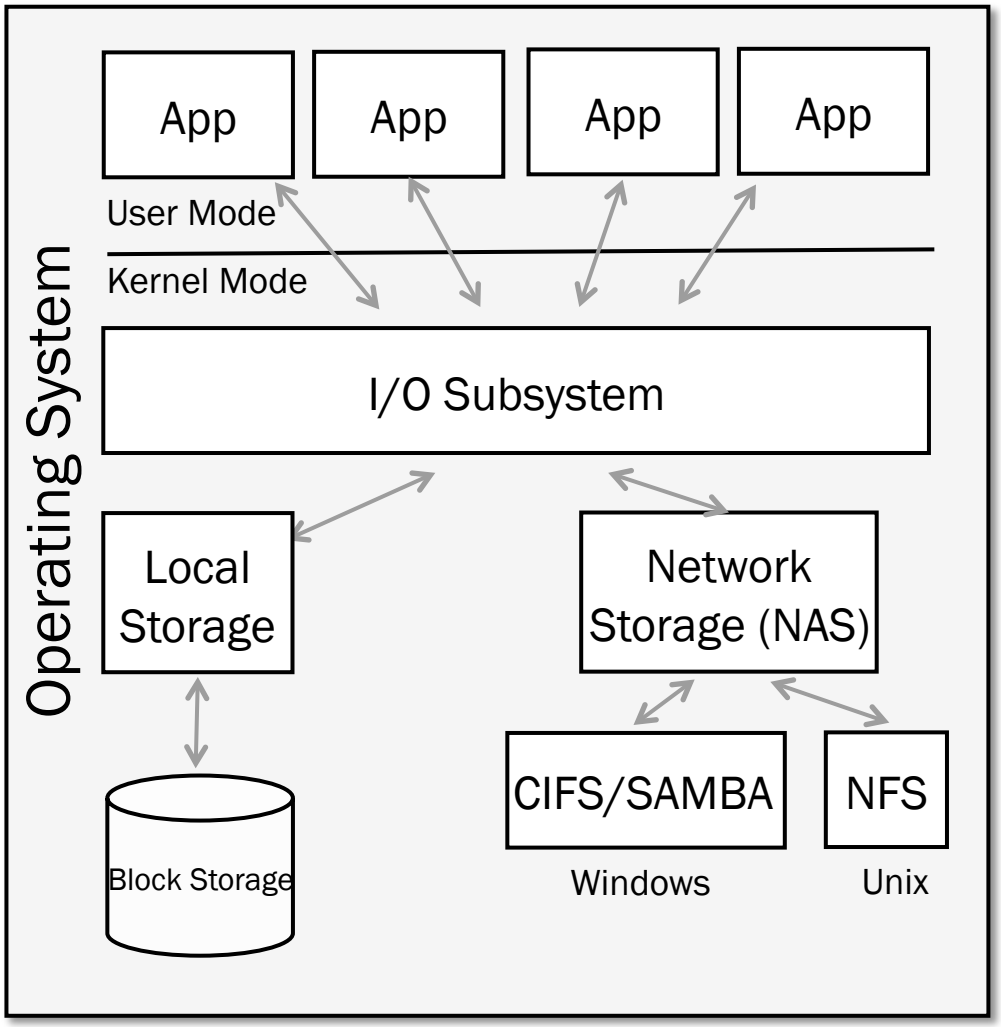
- Other adapters such as NIC cards provide i/o via the switched PCIe bus.
  - Notice that in the past, slower adapters like NICs used a slower “southbridge” connection.
  - Today all peripherals use the switched PCIe interface.



Basic System Architecture

The Operating System

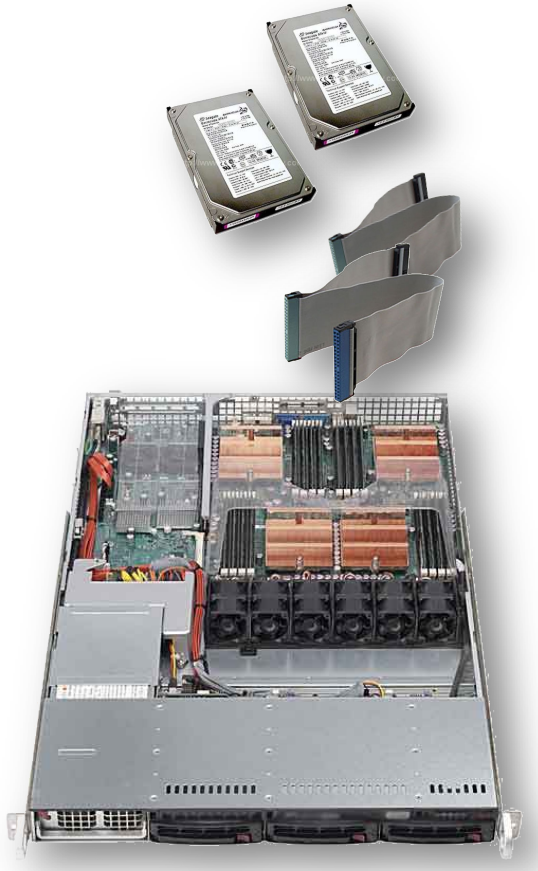
- Applications send and receive i/o via operating systems' i/o subsystem
  - I/O subsystem abstracts complexities of where to locate resources from apps.
- Network Storage (NAS)
  - File Based, not Block
  - Slower speed, higher latencies due to network packet loss, latency, and heavy protocol stack (CIFS/NFS, TCP/IP.)
- Local Block Storage
  - Very low latency, high speed channel directly to block resources.
  - SCSI protocol used



Local Storage

SCSI

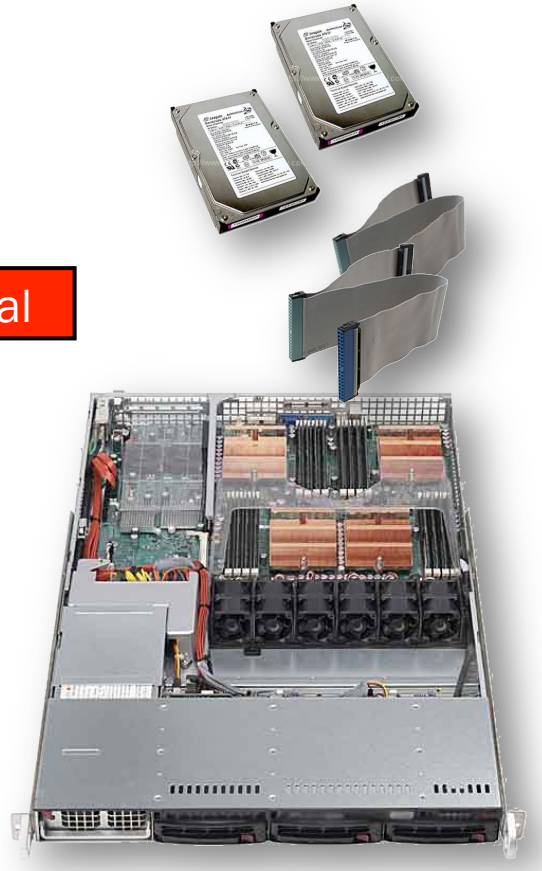
- Local disks use Small Computer System Interface (SCSI)
- SCSI Divided into two main layers…



Local Storage
SCSI

- #1. Physical Interface
  - Ribbon cable, parallel data transmission
  - Lots of rules
    - Only a certain number of devices allowed
    - Cable can only be a certain length
    - No more than two hosts can use a single SCSI channel
    - …more

SCSI Physical

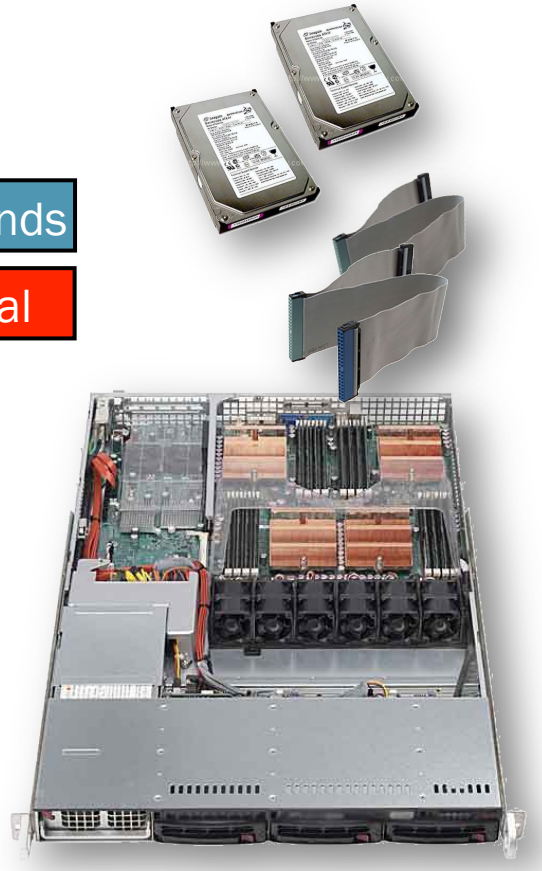


Local Storage
SCSI

- #2 Software Command Set
  - Very mature, very stable
    - SCSI Read
    - SCSI Write
    - Etc.

SCSI Commands

SCSI Physical



A Word About Block Storage

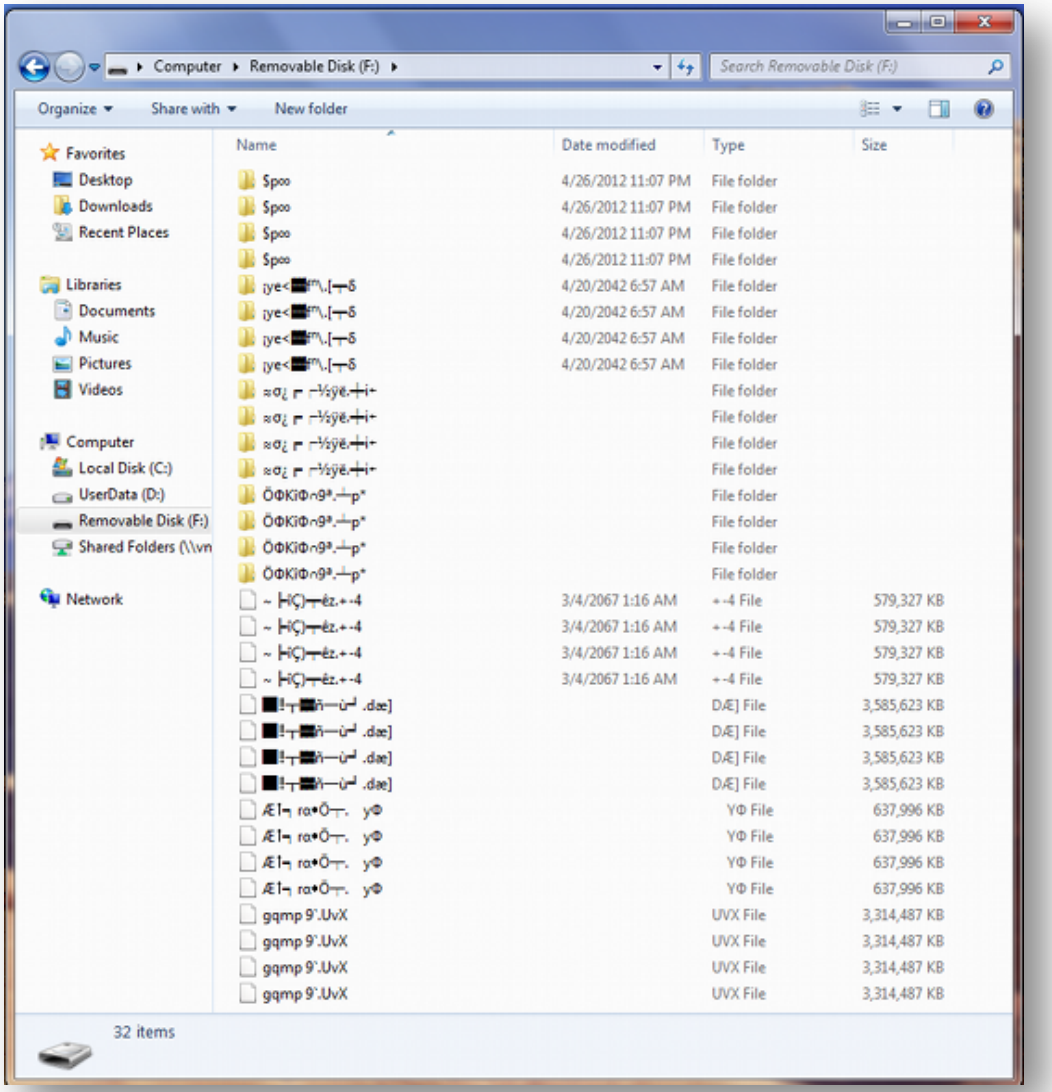
Remove Gracefully

- It can be very bad when block storage suddenly disappears.
  - Operating systems cache writes before flushing them to disk (this speeds things up).
  - Some filesystems try to minimize risk by journaling.
  - Sudden removal can mean corrupt application data or corrupt filesystem.
- You've seen this if you've ever improperly removed a USB drive from a Unix/Linux machine.

Ramification

Corrupt Filesystem

- This is corruption because an application's write operation appeared complete.
- Storage then removed without gracefully unmounting and flushing all pending writes out of the Operating System's write buffer.



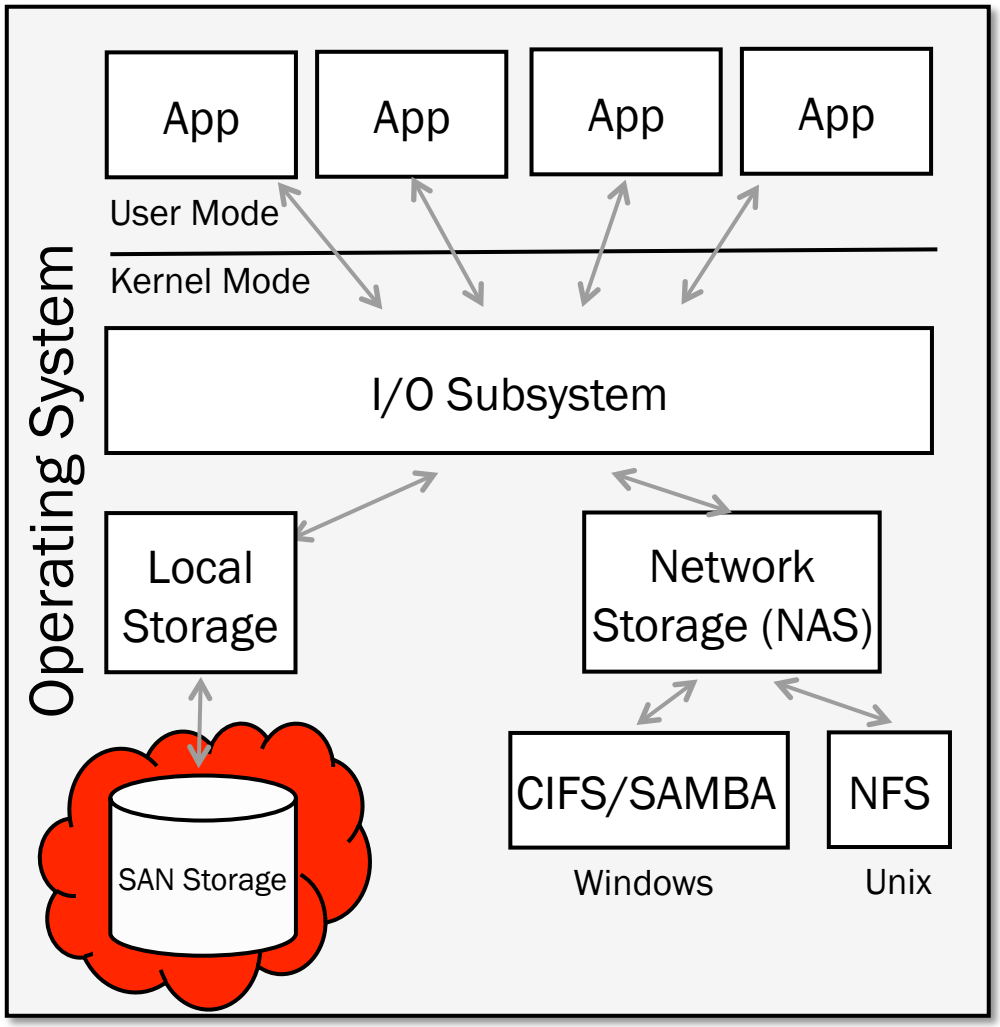
SAN Primer for LAN Professionals

Essential Understanding of SAN Concepts

Block Storage Networking

Enables Consolidation, Optimization

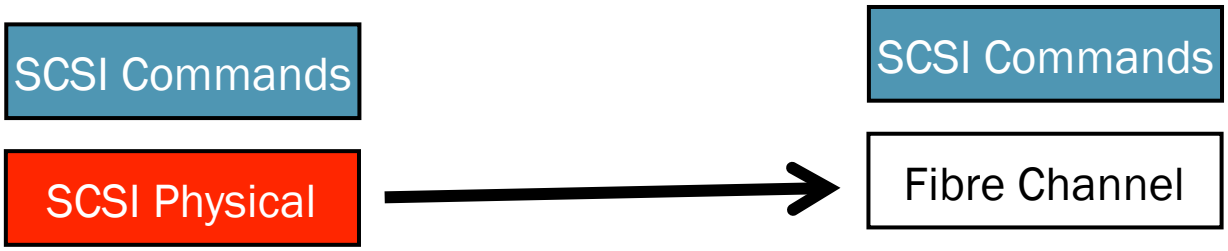
- SAN technologies are designed to make the server think it's using local block storage when those resources are actually networked and external to the server.
- SANs allow companies to reduce cost.
  - Storage consolidated, centralized, and replicated for business-critical initiatives like BC/DR, etc.



Networking SCSI

SCSI Commands Get Serialized Network Protocol – Fibre Channel

- Classic SCSI
- SAN



- This network has to be as reliable as the old SCSI cable…
  - No dropped frames,
  - Data delivered in-order.
  - Highest level of HA and reliability

Block Storage Networking

Enables Consolidation, Optimization

- A little over a decade ago, almost all servers had dedicated, internal hard drives.
- Many servers had excess storage capacity, while other servers were running out of space.
- It was very difficult to reallocate resources.

Block Storage Networking

Enables Consolidation, Optimization

- With dedicated storage, it is was not possible to take snapshots of disks, or rapidly deploy operating system images.
- If a server died, it was a very painful exercise to rebuild the OS on a new server and get apps online.
- Replication and BC/DR was extremely challenging.

Block Storage Networking

Enables Consolidation, Optimization

- By moving the drives out of the servers to a consolidated, high-performance, purpose-built storage array, resources can be managed much more efficiently.

Block Storage Networking

Enables Consolidation, Optimization

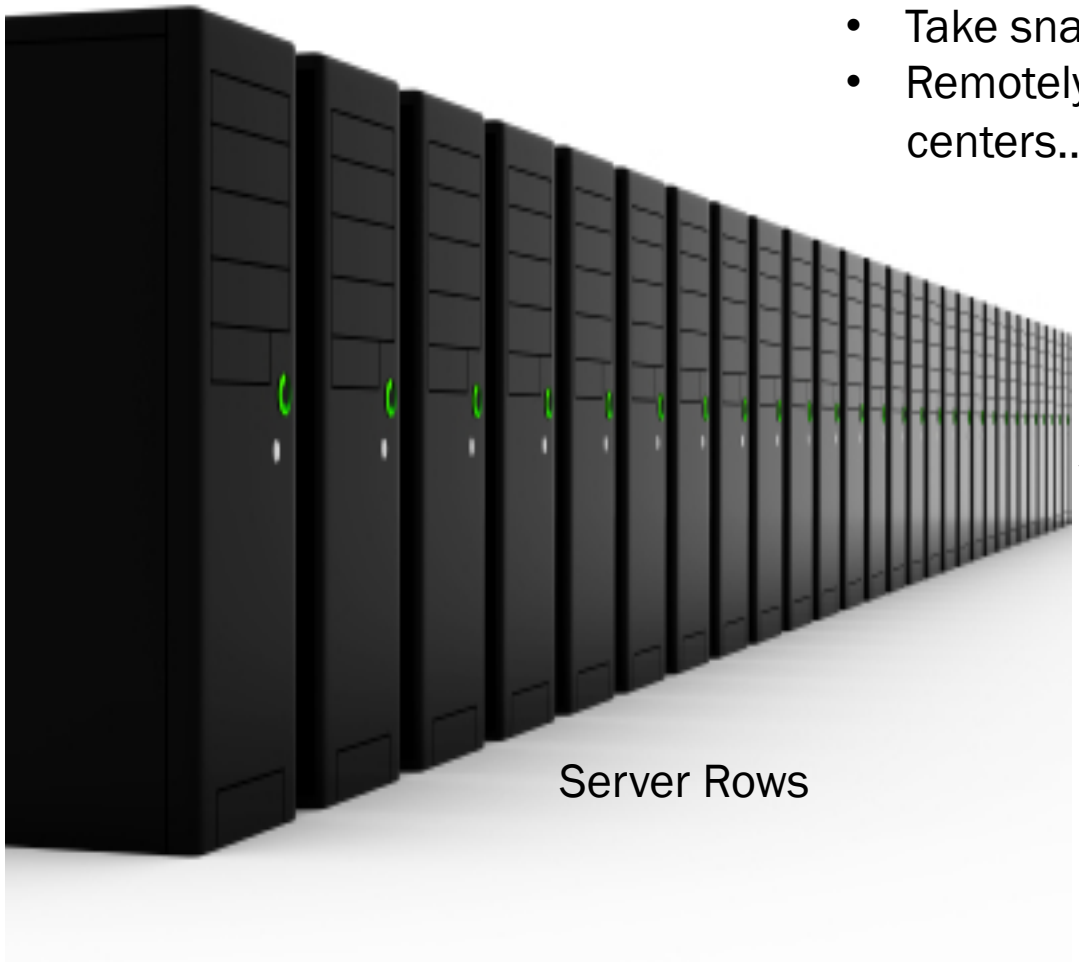
Storage Array

Block Storage Networking

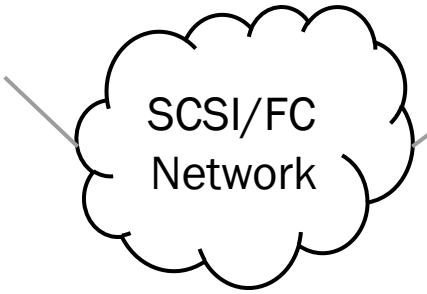
Enables Consolidation, Optimization

Now that storage is consolidated, we can:

- Give servers exactly what they need
- Take snapshots
- Remotely mirror drives to other data centers…much more!



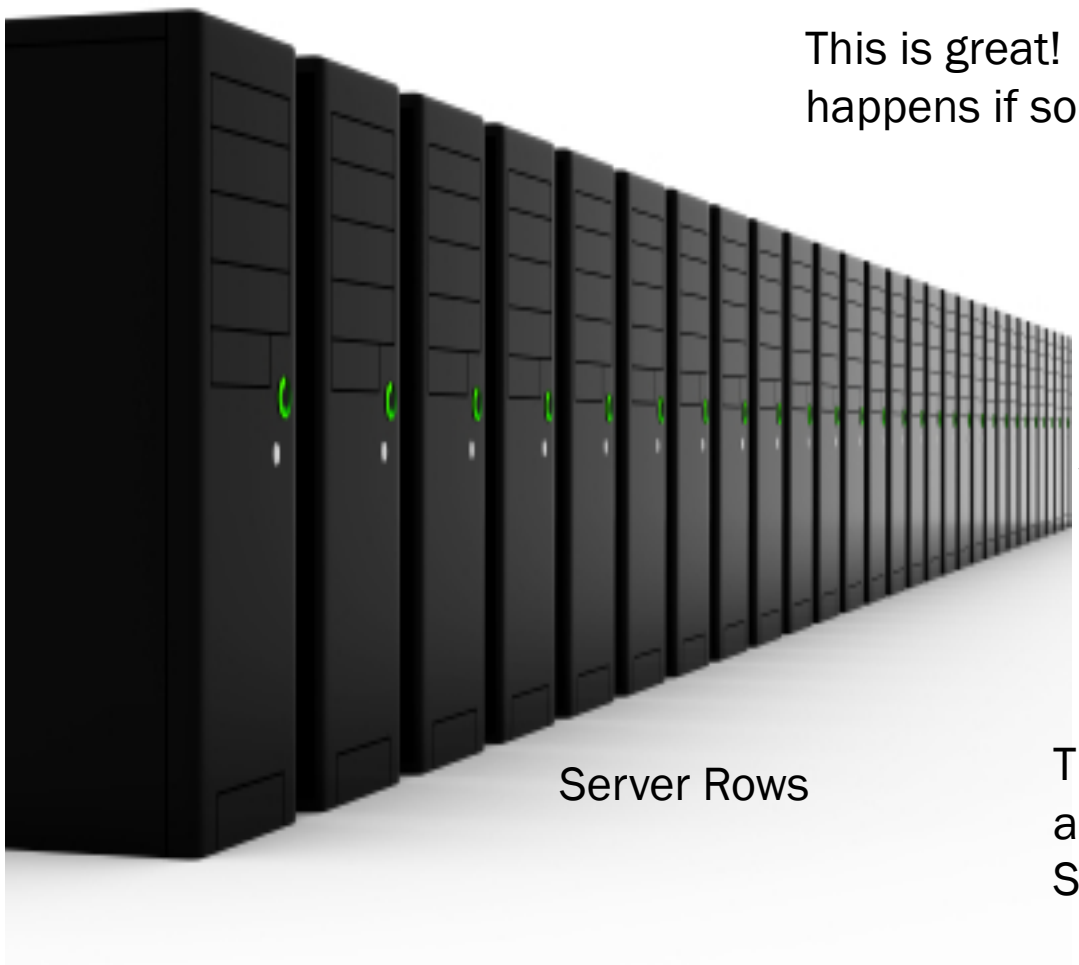
Server Rows



Storage Array

Block Storage Networking

Enables Consolidation, Optimization



This is great! But, wait… What happens if something goes wrong?

Server Rows

SCSI/FC Network

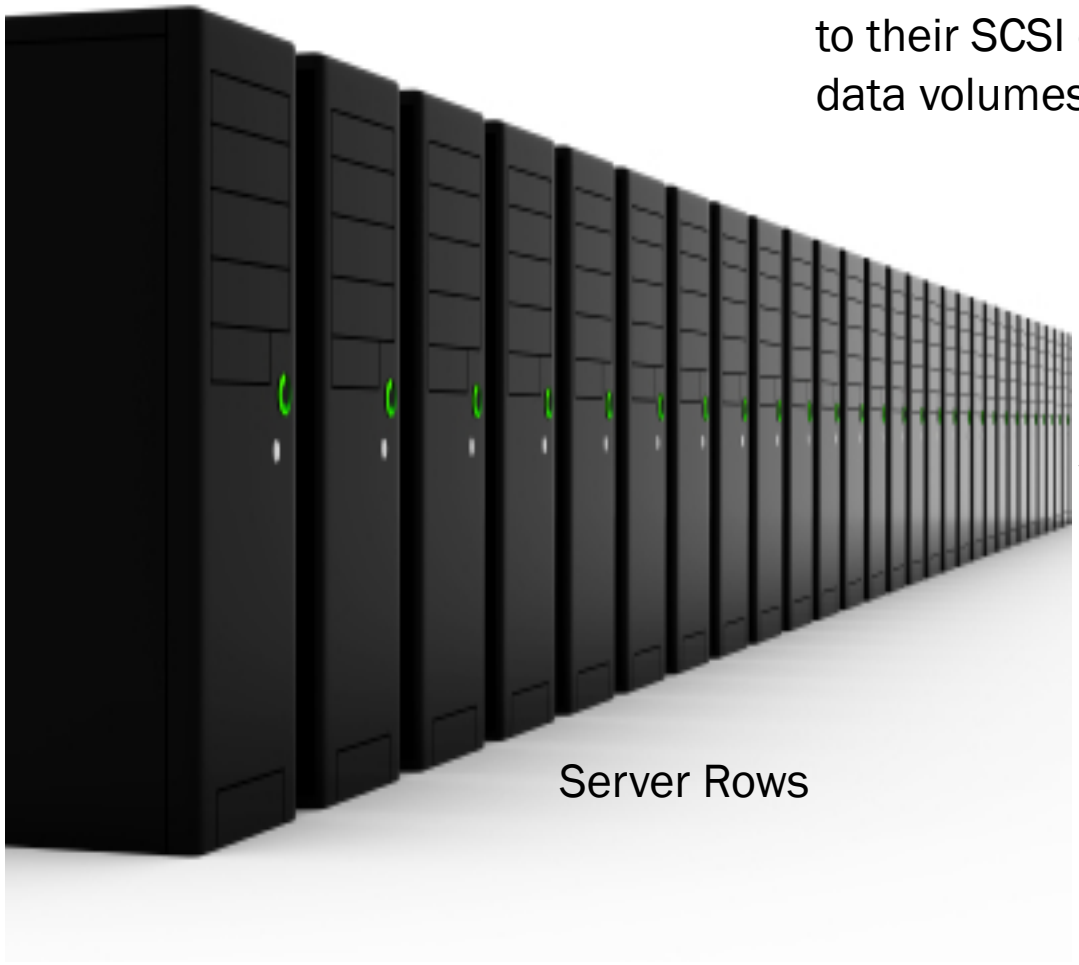
Think of this network as a big, fast, feature-rich SCSI cable

Storage Array

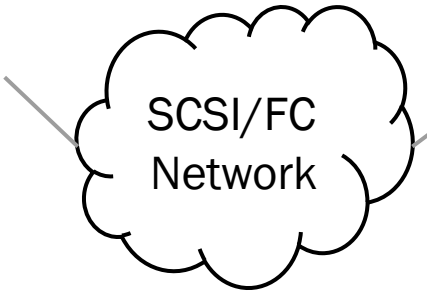
Block Storage Networking

Enables Consolidation, Optimization

An outage on the SAN can cause all servers to simultaneously lose access to their SCSI devices (i.e. boot drives, data volumes, etc).



Server Rows

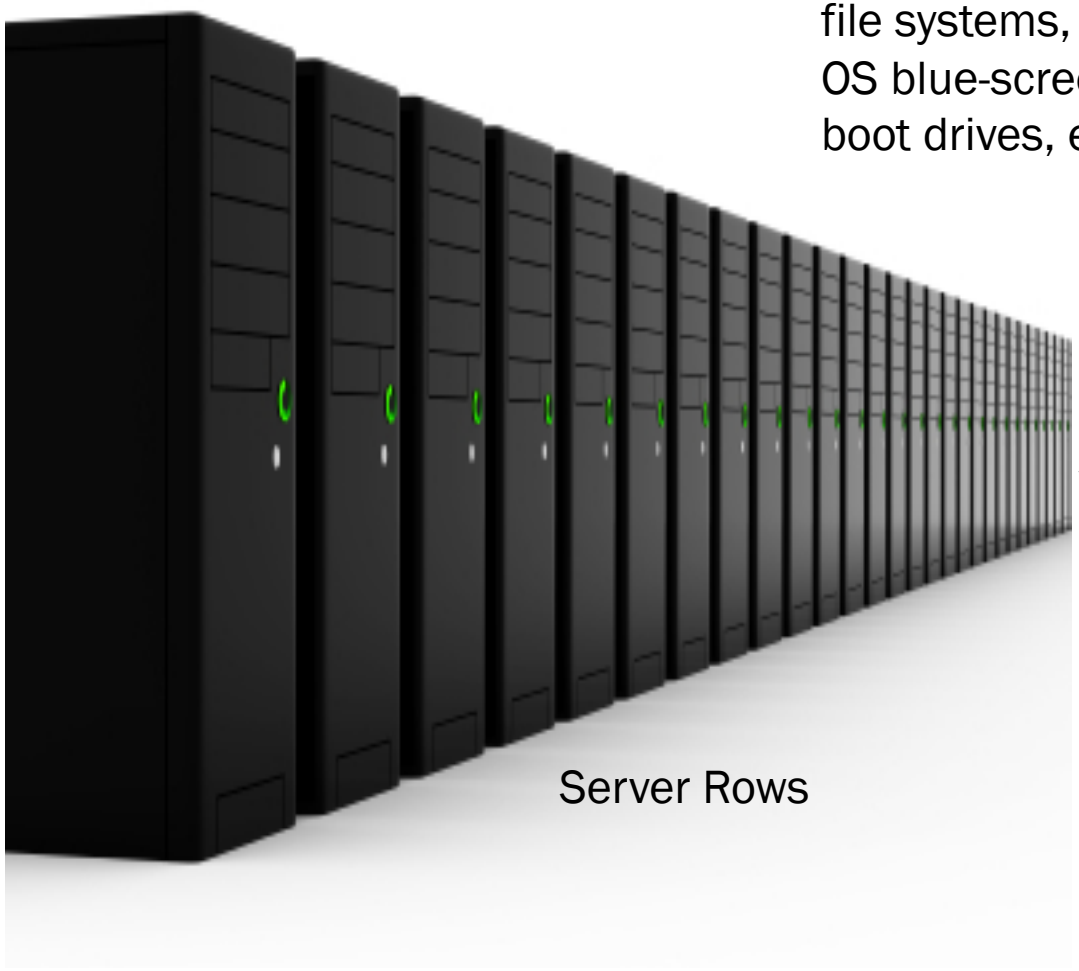


Storage Array

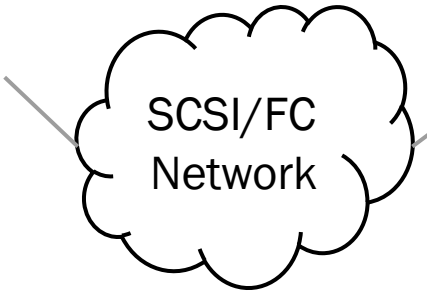
Block Storage Networking

Enables Consolidation, Optimization

There is potential for data corruption on a wide scale - corrupt files, corrupt file systems, corrupt application data, OS blue-screens, core dumps, loss of boot drives, etc.



Server Rows



Storage Array

Block Storage Networking

Enables Consolidation, Optimization

Consider that all VM's boot from SAN storage resources.

Server Rows

SCSI/FC Network

Storage Array

SAN Basics

LAN / SAN Comparison

LAN

- If the LAN goes down, everyone is generally happy when the network comes back online.
  - “Hey my phone works again, and I can get to Google! This is great!”

SAN

- If the SAN network goes down, no one is happy when it comes back online.
- SAN network outages can cause data center-wide corruption and can require many hours (or days) to restore once the SAN network comes back online. This is why we don’t build one SAN network – we always build two.

SAN Basics

LAN / SAN Comparison

LAN

- LANs get most of their best features when they are a single network.
  - Active/Active NIC-teaming from servers
  - LAG across switches
  - etc.

SAN

- SANs are always fully redundant (air gap) for enterprise applications…always.
  - Gets the benefits of consolidated storage while mitigating the risk of widespread outage due to code errors, hardware problems, or human errors.
  - Allows maintenance windows

SAN Basics

LAN / SAN Comparison

Regardless of how resilient a single network can be with multiple paths, etc, it is still a single network.

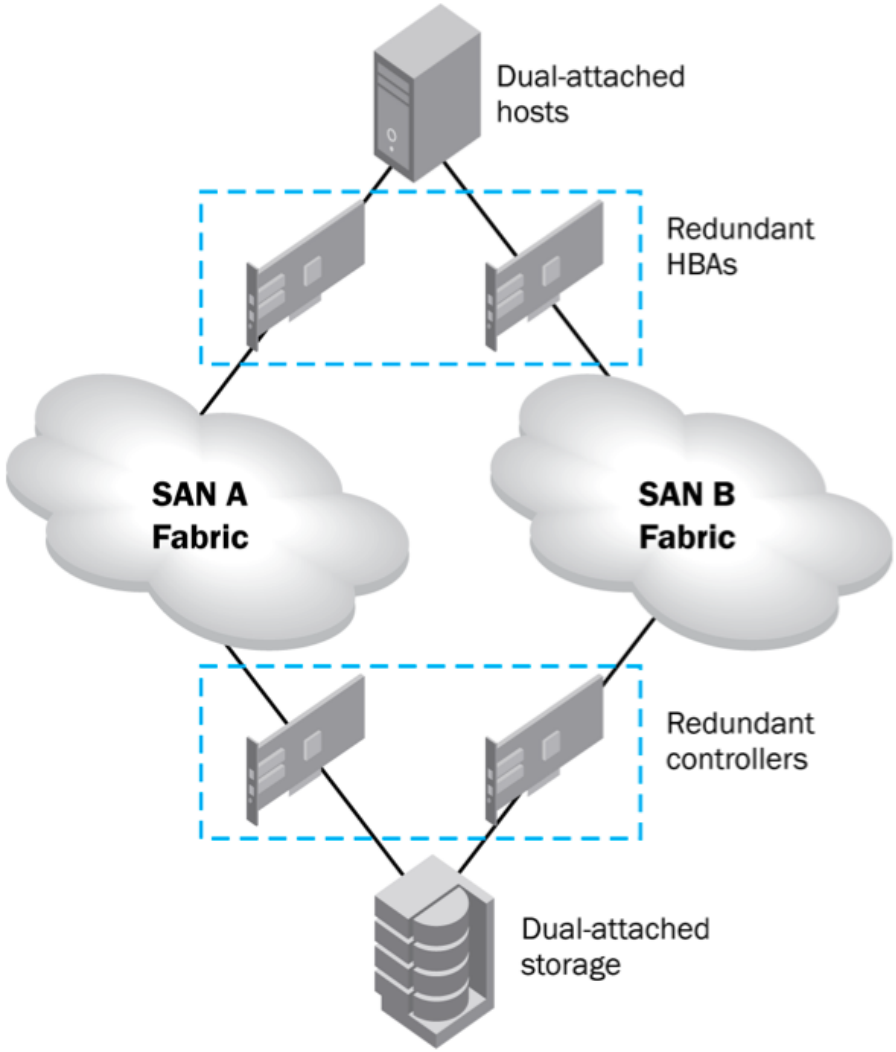
There is no virtual or logical partitioning that can protect the network completely.

If electrons flow between switches, they can bring each other down. This is why we use completely redundant networks with air-gaps for protection in the SAN.

SAN Design Principles

Redundancy

- Redundant HBA's plugged into
- Redundant PCIe busses
- Redundant Fabrics with absolutely no physical interconnects between fabrics (air gap)
- LUNS presented to redundant controllers, redundant cache, redundant disks all electrically separated within the array



How Does An Operating System Address Storage?

Classic Device Hierarchy

- In Unix, physical devices are mapped to a device path
- For example, a disk (LUN) might be: c2t0l5 (or for Solaris – c2t0d5 – the d is for ‘disk’.)

Controller 2 (HBA 2), Target 2 (SCSI Target 2), LUN 5 (Logical Unit Number 5 – think ‘partition’)

How Does An Operating System Address Storage?

Classic Device Hierarchy

- In Unix, physical devices are mapped to a device path
- For example, a disk (LUN) might be: c2t0l5

Controller 2 (HBA 2), Target 2 (SCSI Target 2), LUN 5 (Logical Unit Number 5 – think 'partition')

However, there is a very important problem that must be overcome:

- If the server "sees" a LUN down two separate paths, it will think there are two LUNs instead of one.
- Multipath I/O drivers MUST be installed on the host to correct this "double-vision" to avoid corruption.
- Most modern O/S's have MPIO by default.

SAN Basics

LAN / SAN Comparison

LAN

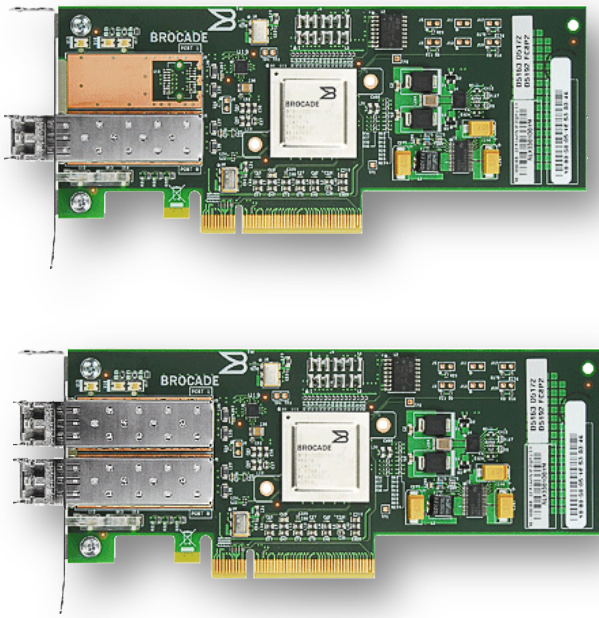
- Nodes use Network Interface Cards (NIC) with 48-bit MAC Addresses.

SAN

- Nodes use Host Bus Adapters (HBA) with 64-bit World Wide Names (WWN).

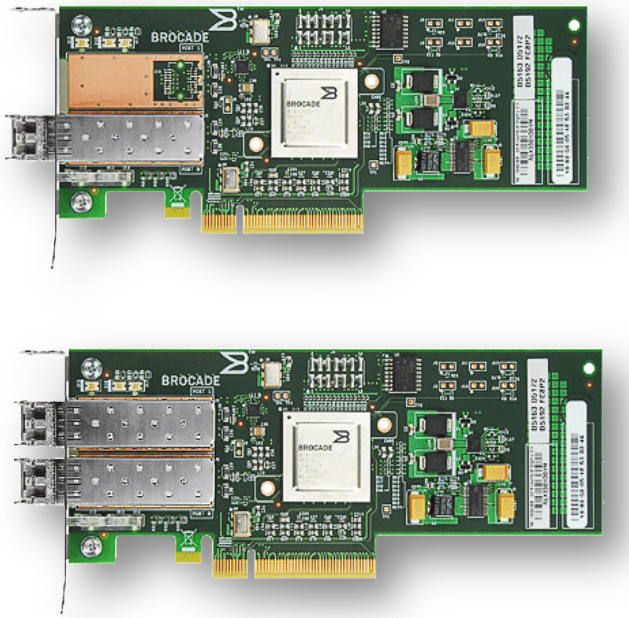
Fibre Channel Host Bus Adapters

- Provides an interface between the server or workstation internal bus (e.g. PCIe) and the Fibre Channel network
- HBA software driver provides the storage information required by the operating system
  - Handles I/O and control requests
  - Copper/Optical media support (may be dual port cards)
- Looks like a SCSI adapter to the host OS



Fibre Channel Host Bus Adapters

- Provides an interface between the server or workstation internal bus (e.g. PCIe) and the Fibre Channel network
- HBA software driver provides the storage information required by the operating system
  - Handles I/O and control requests
  - Copper/Optical media support (may be dual port cards)
- **Looks like a SCSI adapter to the host OS**



Fibre Channel Addressing

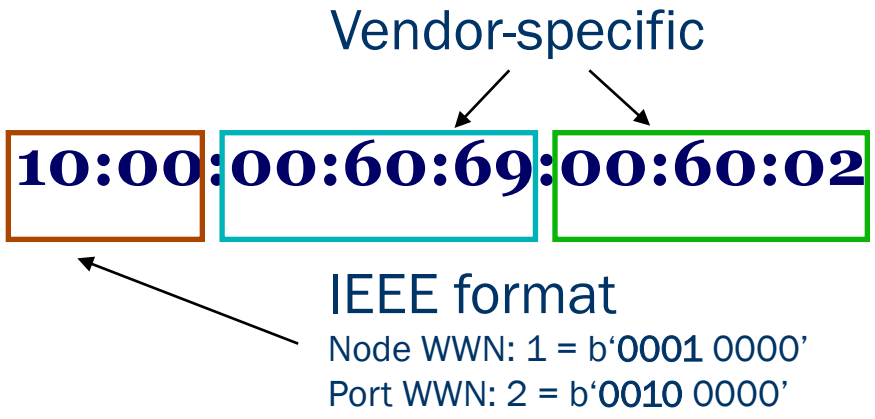
WWN's and Port ID's

- FC has two types of addressing at Layer 2
  - Fixed – World Wide Name Burned in at factory
  - Dynamic – Port ID. A layer 2, 24-bit address that is assigned at fabric login.
- Note: There is no concept of ISO layer 3 or Layer 4 in FC.
  - BTW - "FC routing" is equivalent to L2 NAT, not a true ISO L3 protocol.

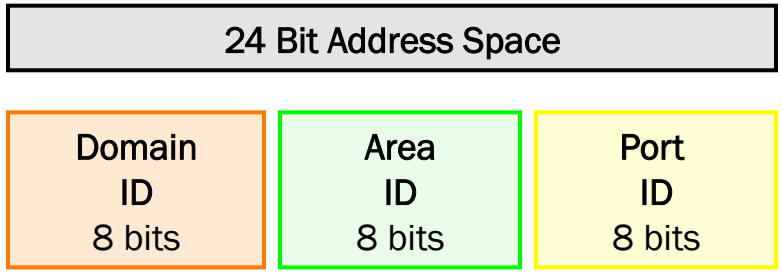
<table>
<tr><th>OSI Model</th><th>Ethernet & TCP/IP</th><th>Fibre Channel</th></tr>
<tr><td>Application</td><td rowspan="3">Application Layers
(POP3, SMTP, DNS,
DHCP,FTP,
WWW protocols)</td><td>Upper Layer Protocols (ULP)
[FCP=SCSI] [FICON]</td></tr>
<tr><td>Presentation</td><td rowspan="2">FC-4: ULP Mapping</td></tr>
<tr><td>Session</td></tr>
<tr><td>Transport</td><td>TCP / UDP</td><td rowspan="2">Not Applicable</td></tr>
<tr><td>Network</td><td>Dynamic IP Address
10.77.77.77</td></tr>
<tr><td rowspan="2">Data Link</td><td rowspan="2">Fixed MAC Address
x'00-00-0E-21-17-6B'</td><td>FC-3: Common Services</td></tr>
<tr><td>Dynamic Native Address (8/24-bit)
Fixed World-Wide Name (64-bit)</td></tr>
<tr><td>Physical</td><td>Physical Interface</td><td>FC-1: 8b/10b or 64b/66b Encoding
FC-0: Physical Interface</td></tr>
</table>

Fixed & Dynamic Address Formatting

WWN's and Port ID's (PID's)



Every fabric device (HBA, switch, director, storage device) has one or more 64-bit WWN addresses.
Uses an IEEE-assigned addressing scheme.



Dynamic address (24-bit)
Assigned dynamically when logging into the Fibre Channel network

24-bit = 16 million fabric addresses

N_Port/F_Port usable range:
x'010000' to x'EFEFFF'

SAN Basics

LAN and SAN Comparisons

LAN

- In a LAN, it is possible to have shared media

SAN

- All connections in a SAN are point-to-point.

SAN Basics

LAN / SAN Comparison

LAN

- L2 Ethernet does not guaranty delivery of frames. Frame drop can happen by congested end devices, or switches.

SAN

- FC is considered “lossless”. This is achieved by careful flow control.
- Receiving device always in charge of flow.
- Flow control is based on credits. “If I give you 4 credits, you may send me 4 frames.”

SAN Basics

LAN / SAN Comparison

LAN

- Nodes communicate with nodes.

SAN

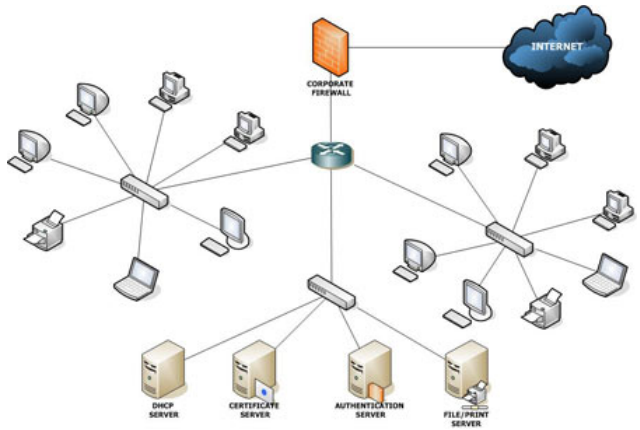
- Nodes distinctly categorized into two groups:
  - Host (initiator)
  - Storage (target)
- Hosts do not communicate with other hosts on a SAN…they only communicate with storage targets.

SAN Basics

LAN / SAN Comparison

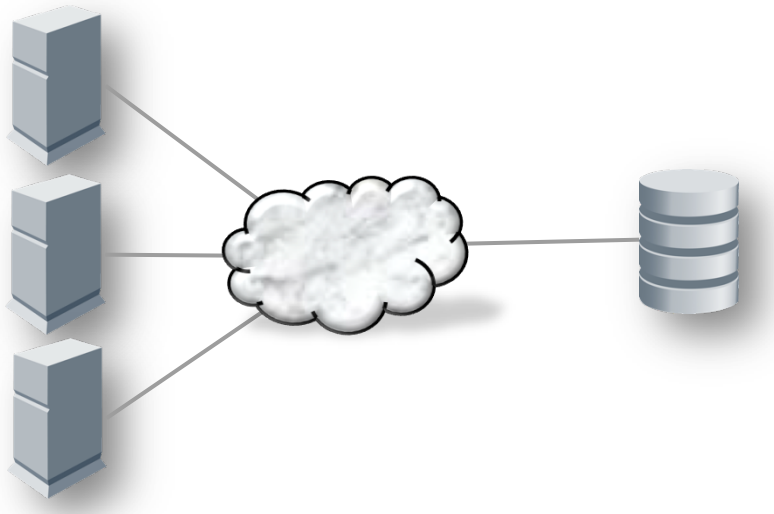
LAN

- Networks provide any-to-any connectivity



SAN

- Networks connect many-to-few.

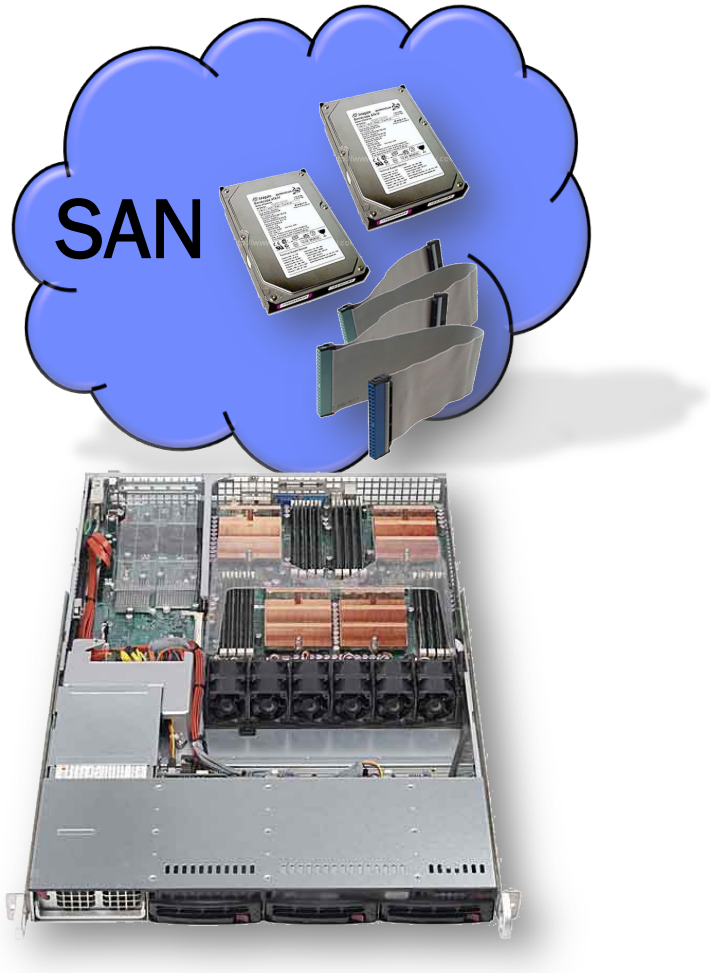


SAN

The Basics

SAN

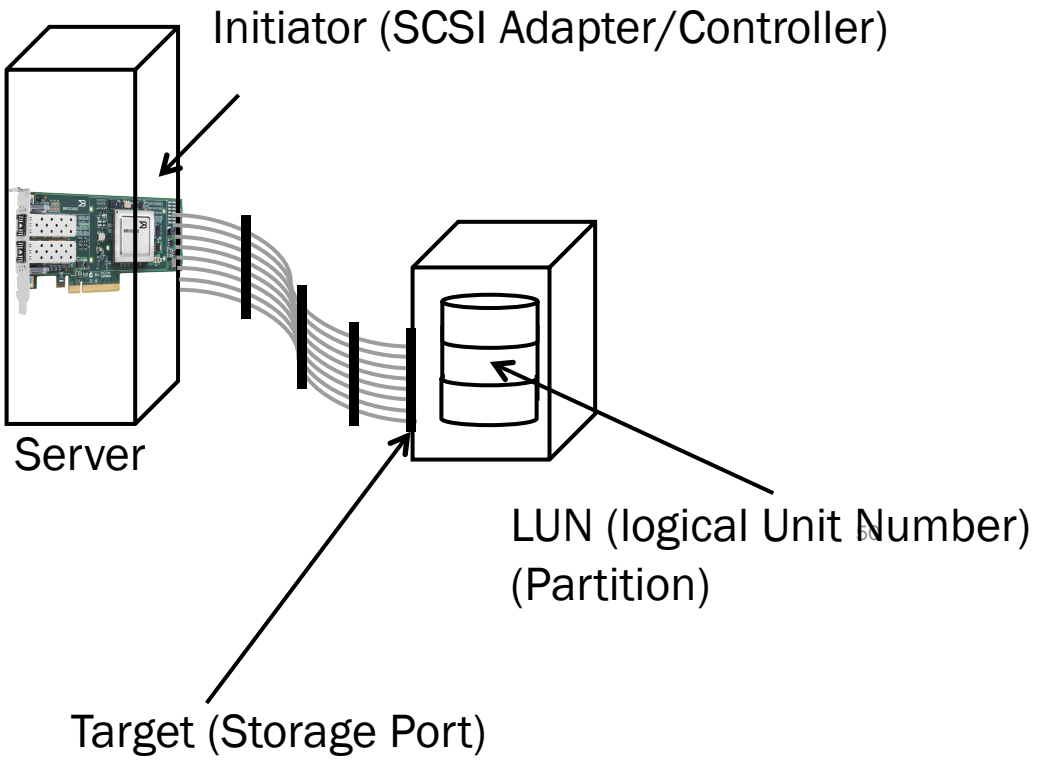
- Important zoning and masking tools in the SAN and target systems make certain that each host only sees what it thinks is a simple SCSI channel with a small number of attached drives (LUNs).



SAN

The Basics

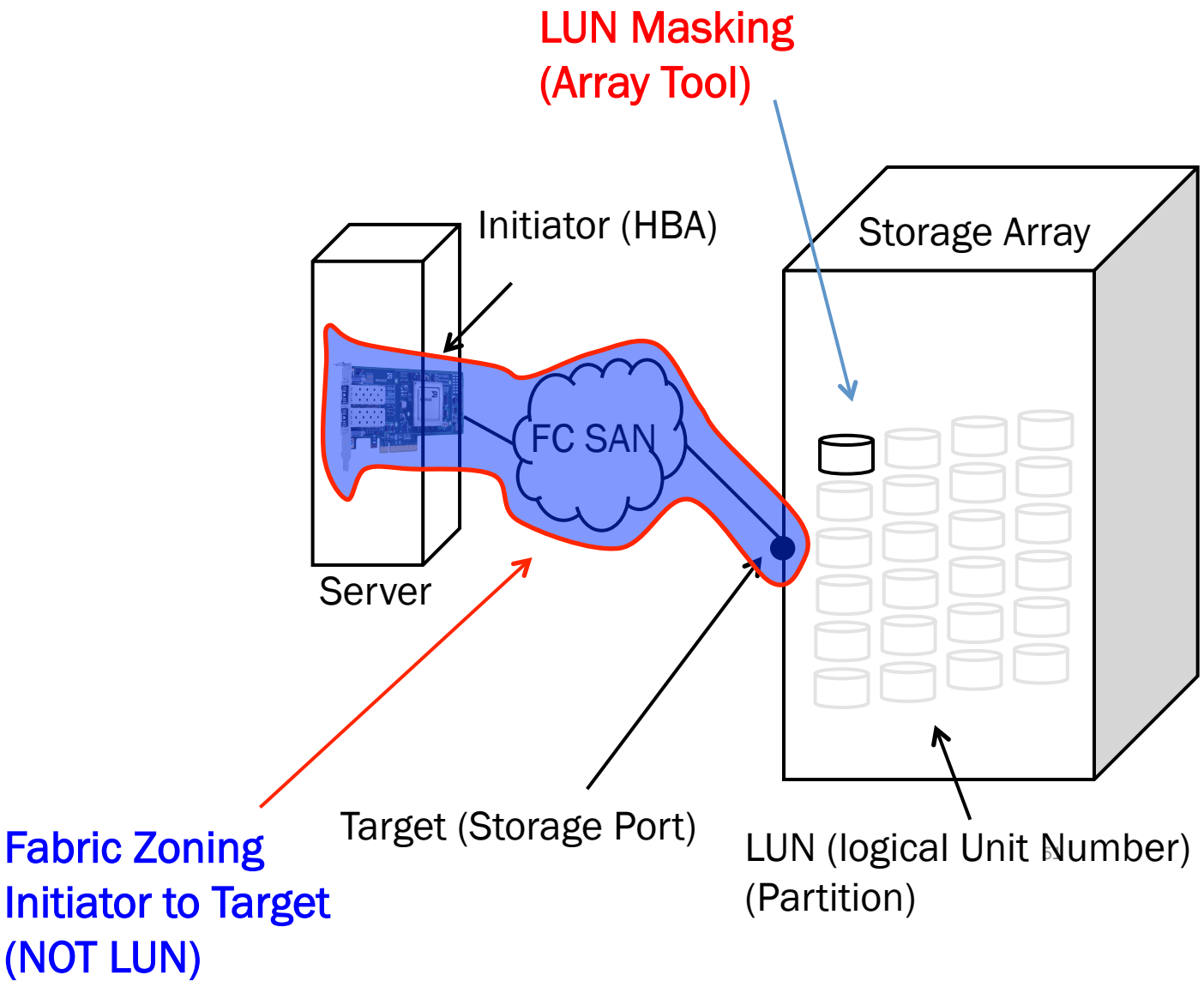
SCSI Terminology



SAN

The Basics

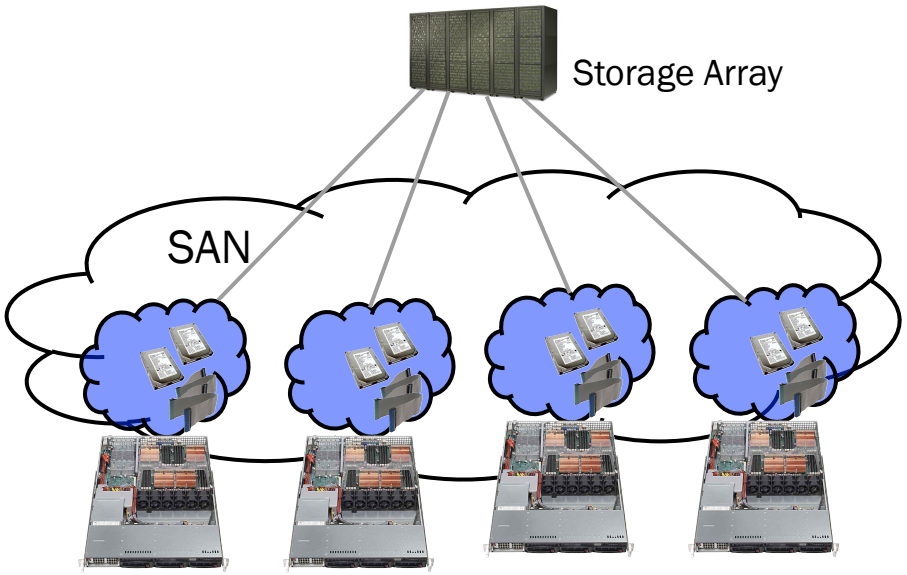
How does it look on a SAN?



SAN
The Basics

SAN

- Although a SAN may have many hundreds of initiators and targets, each host must only see its own storage, not other hosts or other systems' storage.
  - An exception is when certain server clustering tools are being use. In this case multiple servers may see the same storage pool.



SAN Basics

LAN / SAN Comparison

LAN

- For the most part, LAN's are painstakingly, manually configured…port by port.

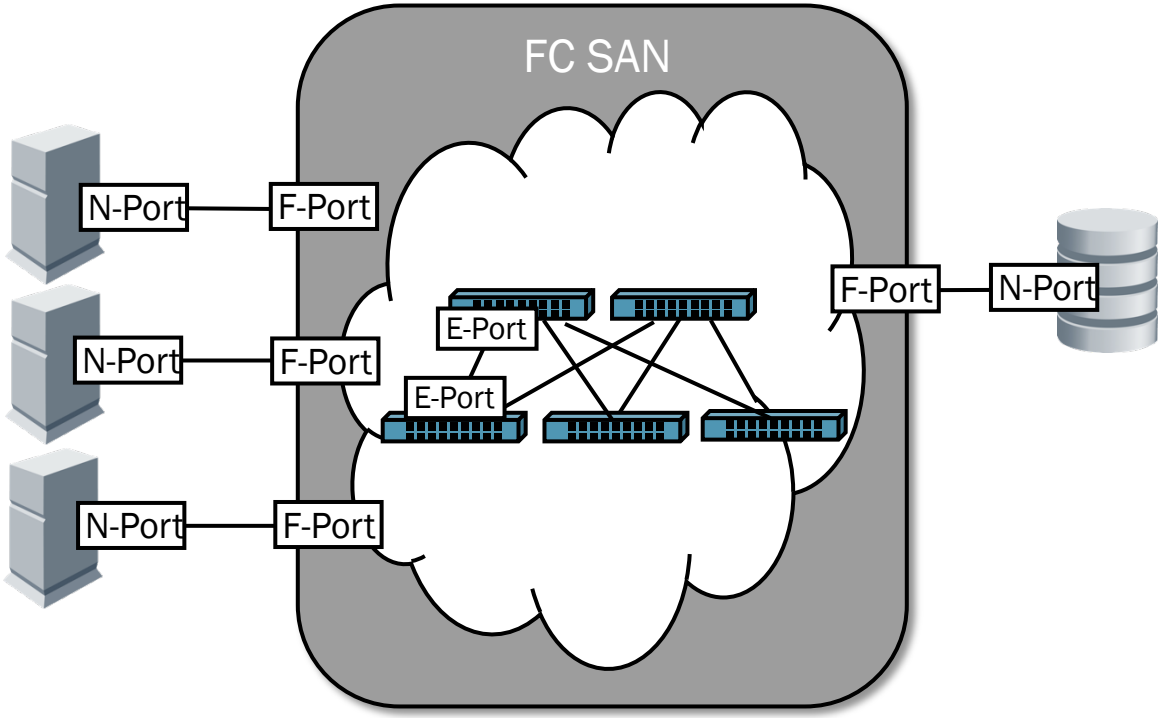
SAN

- In a Brocade SAN, the network ports configure themselves. It is plug and play.

Fibre Channel Port Types

Understanding the Basics

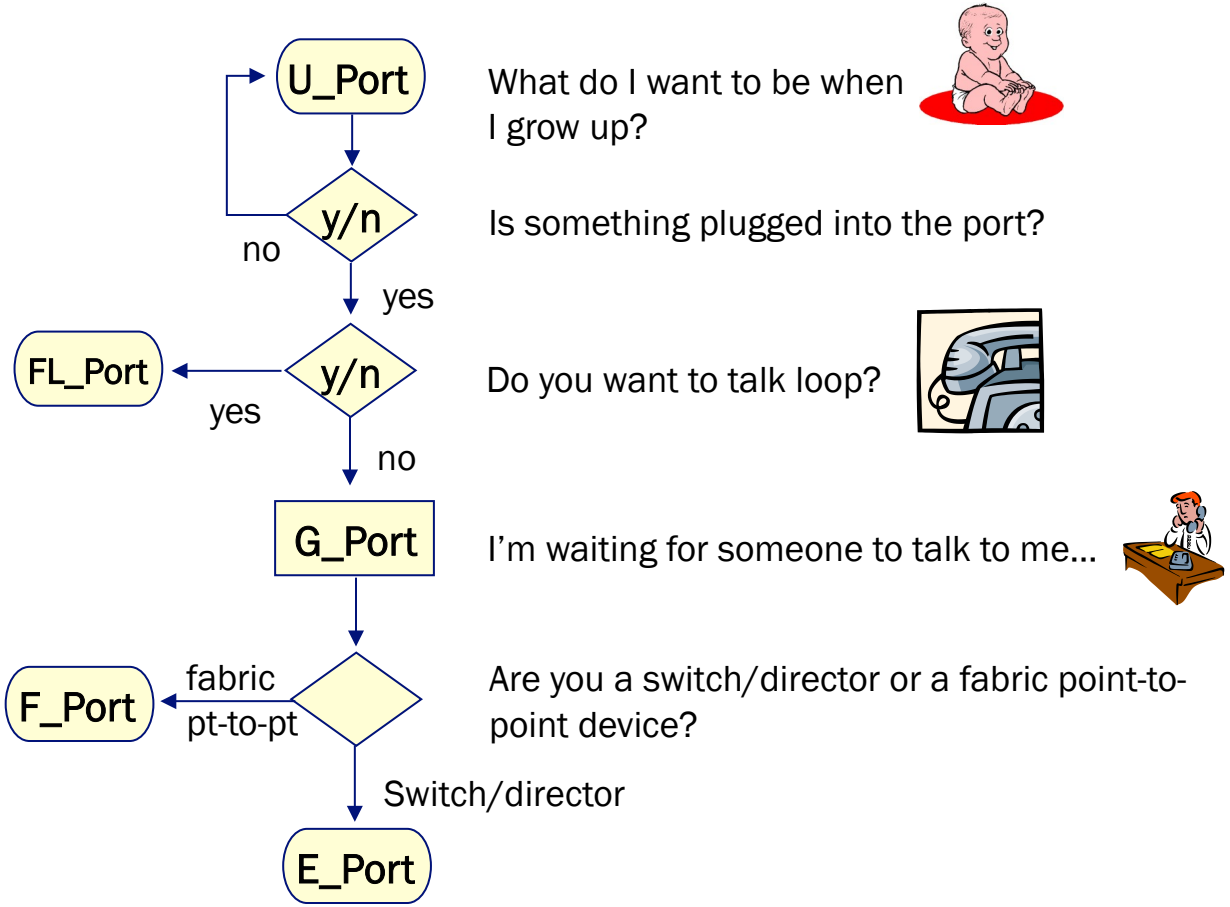
- Fibre Channel SANs have several different port types.
  - N-Port “Node Port”
  - F-Port “Fabric Port”
  - E-Port “Extension Port”
  - N-Ports connect to F-Ports
  - E-Ports connect to other E-Ports
  - More types (Ex, M, D, FL, L)



Fabric Port Initialization

Self-Configuring Ports

- After speed negotiation, a switch port will figure out what device is plugged in (i.e. a host, or another switch).
- It will then automatically configure the appropriate port type to accommodate that device.



SAN Basics

LAN / SAN Comparison

LAN

- A Layer 2 LAN uses broadcasts and must deal with unknown destination addresses.
- Must listen to traffic to learn MAC addresses

SAN

- There are no broadcasts.
- Nodes must login to the fabric and register themselves before any traffic may flow, ergo – the SAN knows where everything is located.

SAN Basics

LAN / SAN Comparison

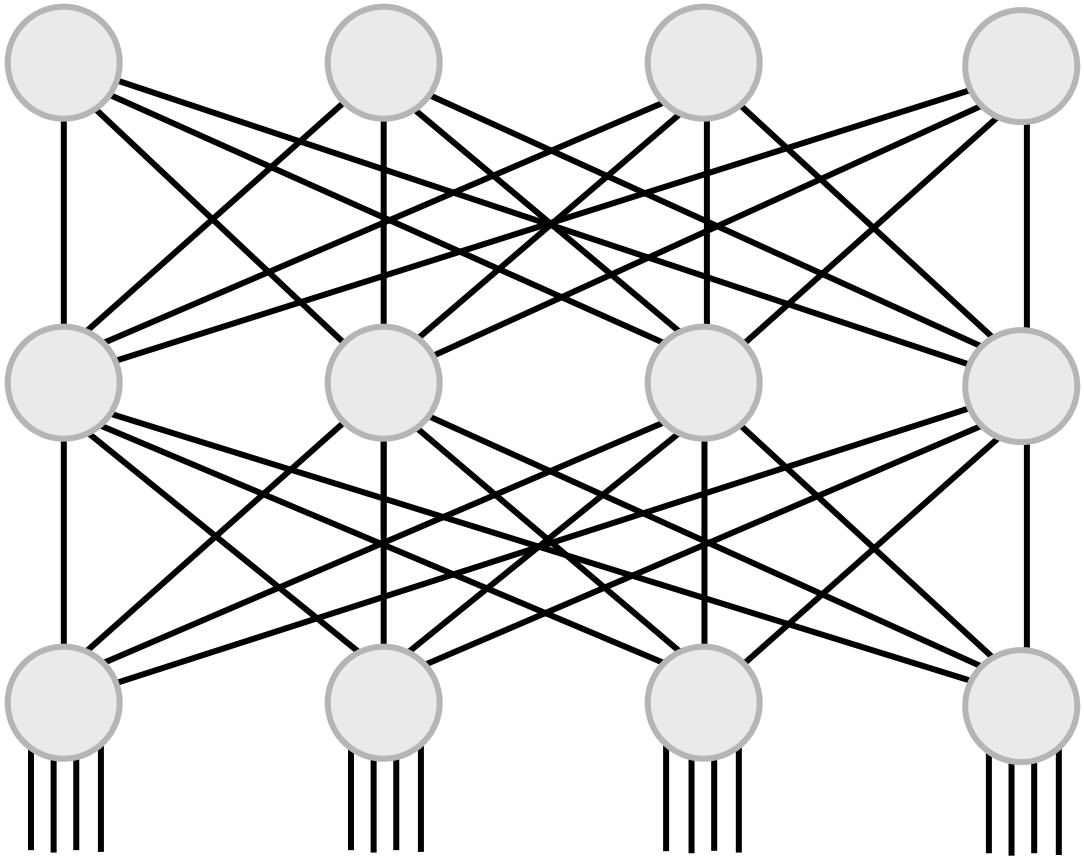
LAN

- In a layer 2 LAN, you have to worry about loops.

SAN

- Since its inception 14 years ago, Fibre Channel has incorporated layer 3 style intelligence at layer 2 – ergo, no loop concerns at all.
- A protocol similar to OSPF is used in layer 2 FC called “FSPF”, or “Fabric Shortest Path First” and it is an open standard (ANSI T11 fc-sw-5).
- Effortless L2 equal-cost multi-pathing

FC SAN = All Links Active And Forwarding



FSPF

Determining Paths

Four Main Components of FSPF:

#1. Hello Protocol

- Establish connectivity with a neighbor Switch
- Establish the identity of the neighbor Switch
- Exchange FSPF parameters and capabilities;

#2. Replicated Link State Database

- Has protocols and mechanisms to keep the databases synchronized across the Fabric;

#3. Path Computation Algorithm

#4. Routing Table Update

FSPF

Basics

The Link State Database is central to the operation of FSPF.

- It is a replicated database where all Switches in the Fabric have the same exact copy of database at all times
- The database consists of a collection of Link State Records (LSRs).

Path computation is local

- The results of the computation are not distributed to other Switches, only topology information is distributed. This is a characteristic of link-state path selection protocols.

FSPF

Determining Paths

Operation	Starting Condition	Process	Ending Condition
1. Perform Initial HELLO Exchange	The Switch originating the HELLO has a valid Domain_ID.	HLO SW_ISL frames are exchanged on the link until each Switch has received a HELLO with a valid neighbor Domain field.	Two way communication has been established
2. Perform Initial Database Exchange	Two way communication has been established.	LSU SW_ISL frames are exchanged containing the Initial database.	Link State Databases have been exchanged.
3. Running State	Initial Database Exchange completed.	Routes are calculated and set up within each Switch. Links are maintained by sending HELLOs every Hello_Interval. Link databases are maintained by flooding link updates as appropriate.	FSPF routes are fully functional.

Principal Switch; Principal ISL's

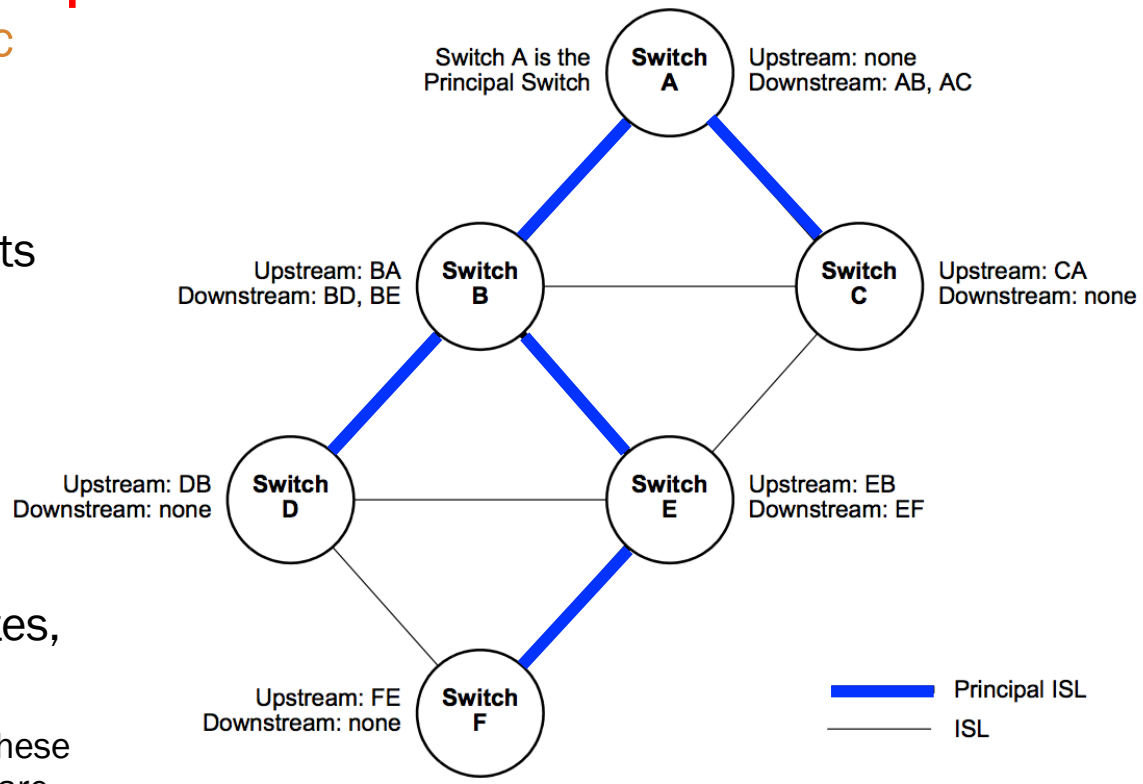
Special Paths for Class F Traffic

Principal Switch:

- Ensures No Domain ID Conflicts
- Time Sync, etc.

Principal ISL's

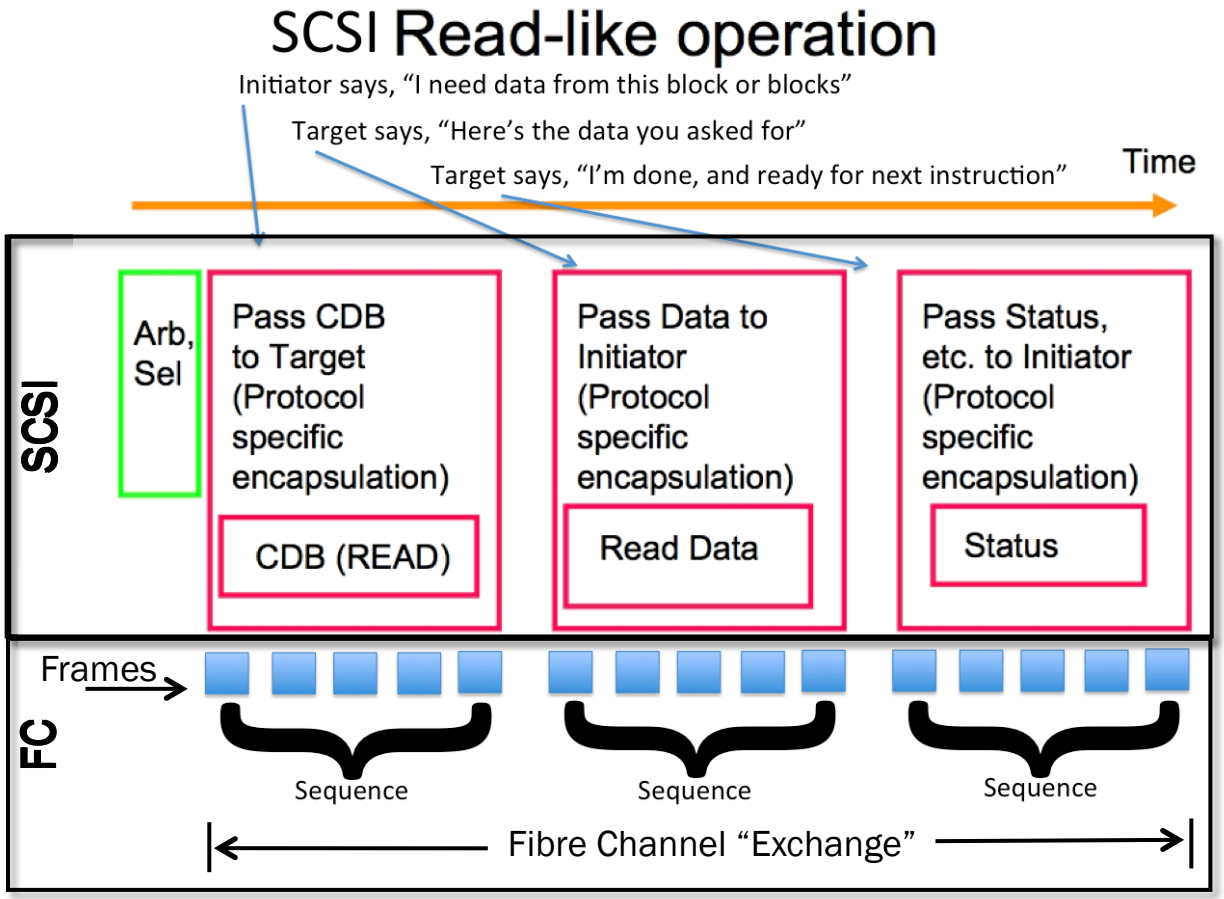
- Establish a Path to Principal Switch (Principal Links)
- Used for FSPF link-state updates, etc.
  - Note: SAN traffic can use all links, these links are only special because they are designated for fabric-stabilization traffic



Dynamic Path Selection

Sharing the Load across FSPF Paths

SCSI Commands are split into sequences of FC frames



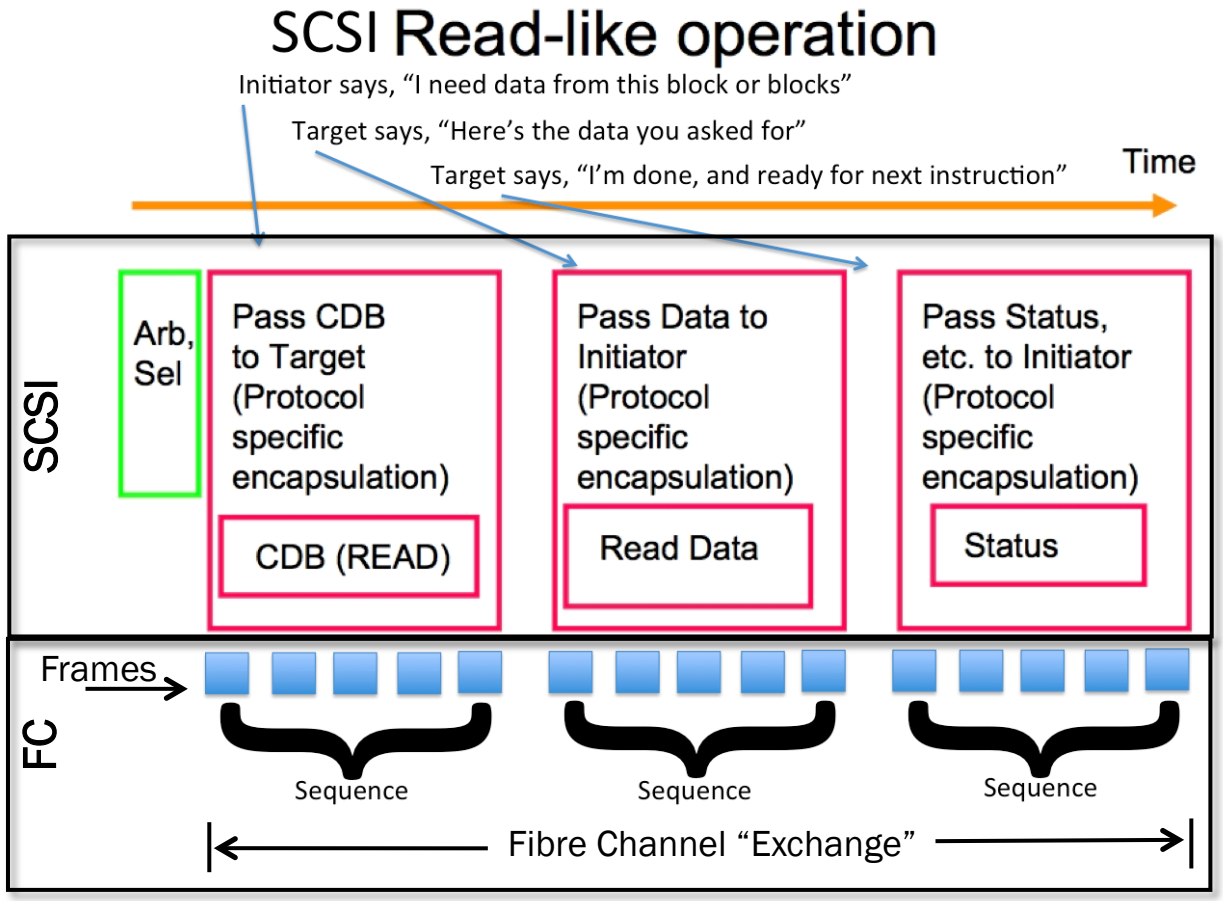
Dynamic Path Selection

Sharing the Load across FSPF Paths

SCSI Commands are split into sequences of FC frames

A complete SCSI command maps to a FC "Exchange"

FC Fabric load balances by hashing on these "Exchange ID's" (OXID) and spraying exchanges across equal-cost paths



SAN Basics

LAN / SAN Comparison

LAN

- Nodes communicate with nodes. They are unaware of switching infrastructure.

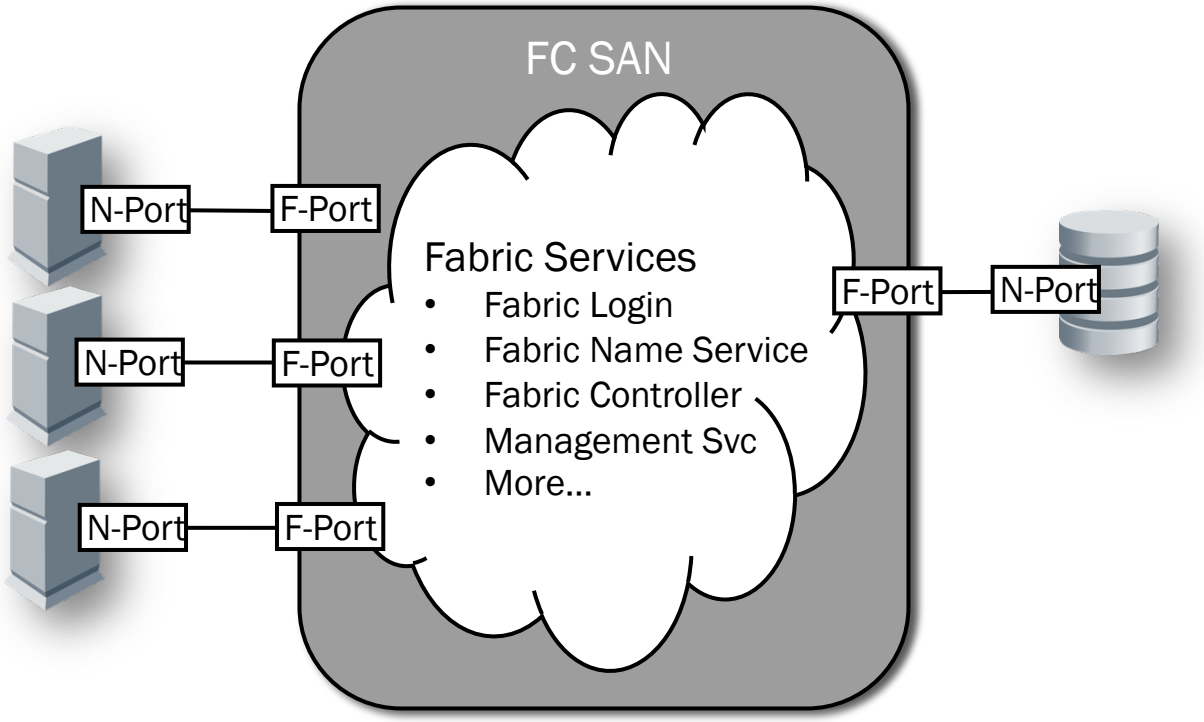
SAN

- Nodes are intrinsically aware of the network infrastructure. They have conversations with the network; they rely on the network for device discovery, change notification, etc.

Fibre Channel Services

Understanding the Basics

- Intelligent Fabric Services are distributed among all switches in a fabric.
- Nodes communicate with the “fabric”, not individual switches. If multiple switches are in a fabric, they will respond to service requests with one singular, cohesive voice.



Port ID's for Fabric Services

Fibre Channel Well-Known Addresses

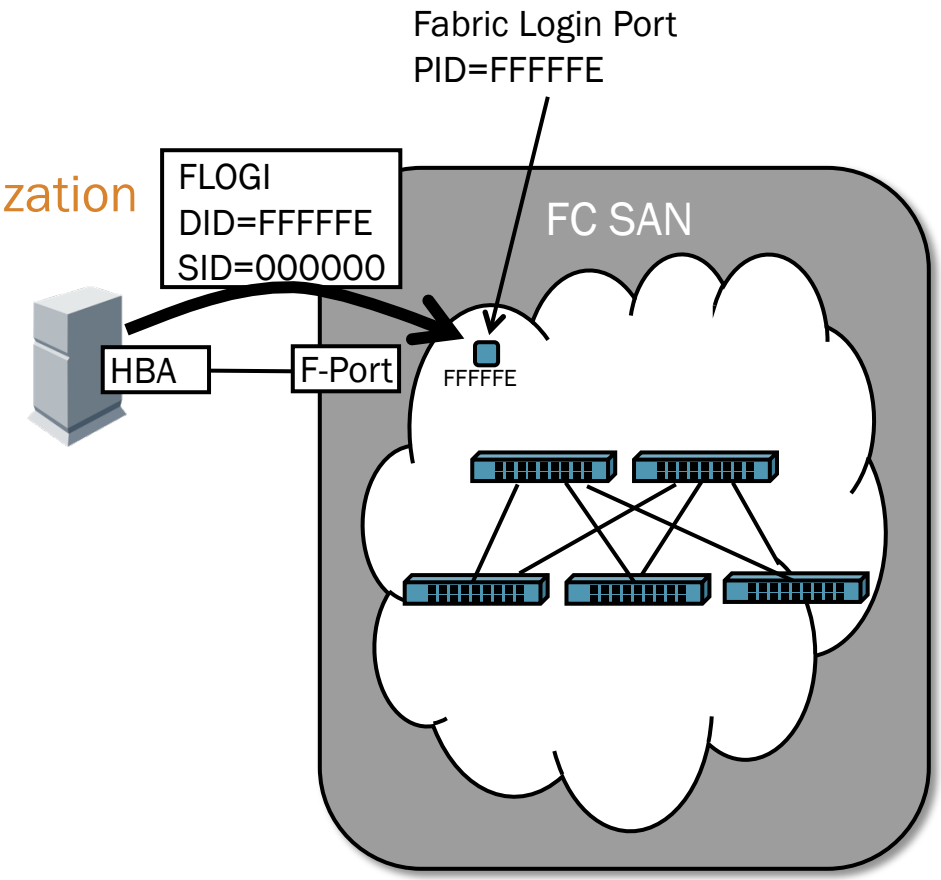
So how do initiators and targets find these services in the fabric?
…by the services' "Well Known Addresses" (PIDs), of course!

- x'FFFFFA' — Management Server
- x'FFFFFB' — Time Server
- **x'FFFFFC' — Name Server**
- **x'FFFFFD' — Fabric Controller**
- **x'FFFFFE' — Fabric Login Server**
- x'FFFFFF' — Broadcast Address (For IP over FC)
- x'000000' — unidentified N_Port
- x'FFFFF5' — Multicast Server
- x'FFFFF6' — Clock Synchronization Server
- x'FFFFF7' — Security Key Distribution Server
- x'FFFFF8' — Alias Server
- x'FFFFF9' — Quality-of-Service Facilitator

Fabric Login

The Process of Device Initialization

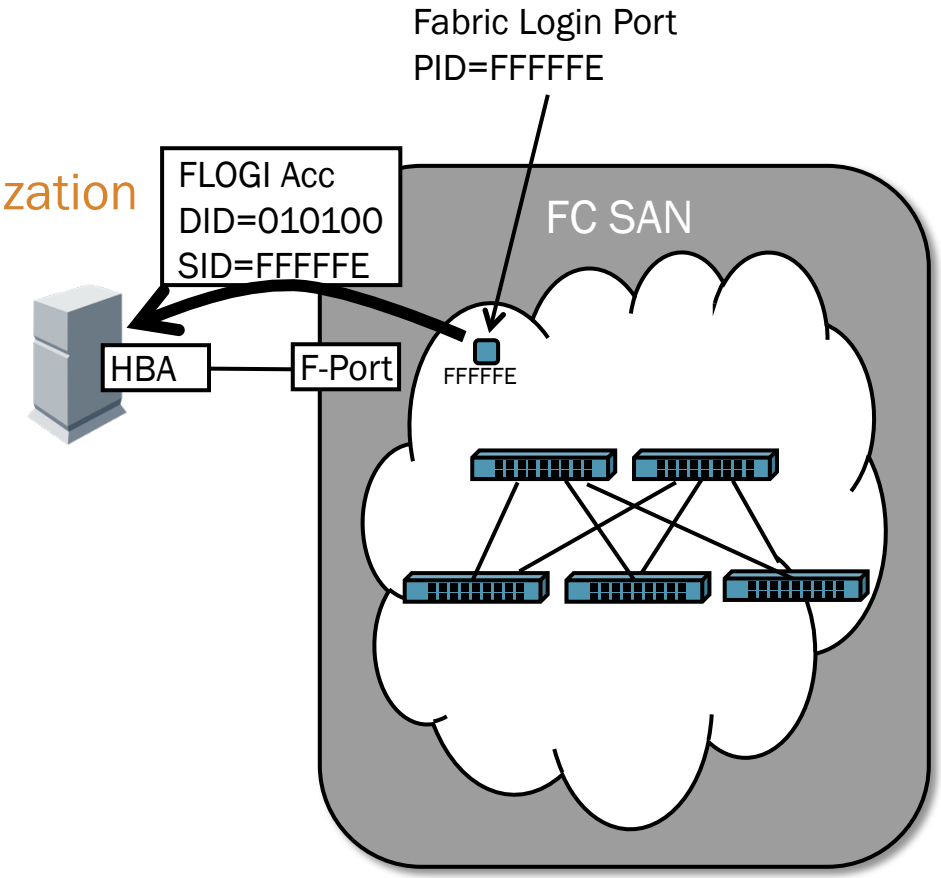
- First, the host sends a FLOGI (Fabric Login) to the Fabric Login Port's Well Known Address.
  - The destination address is the PID of the Fabric Login Port (xFFFFFE).
  - However, the source address on the FLOGI is "x000000". (The host doesn't know yet what its PID is.)



Fabric Login

The Process of Device Initialization

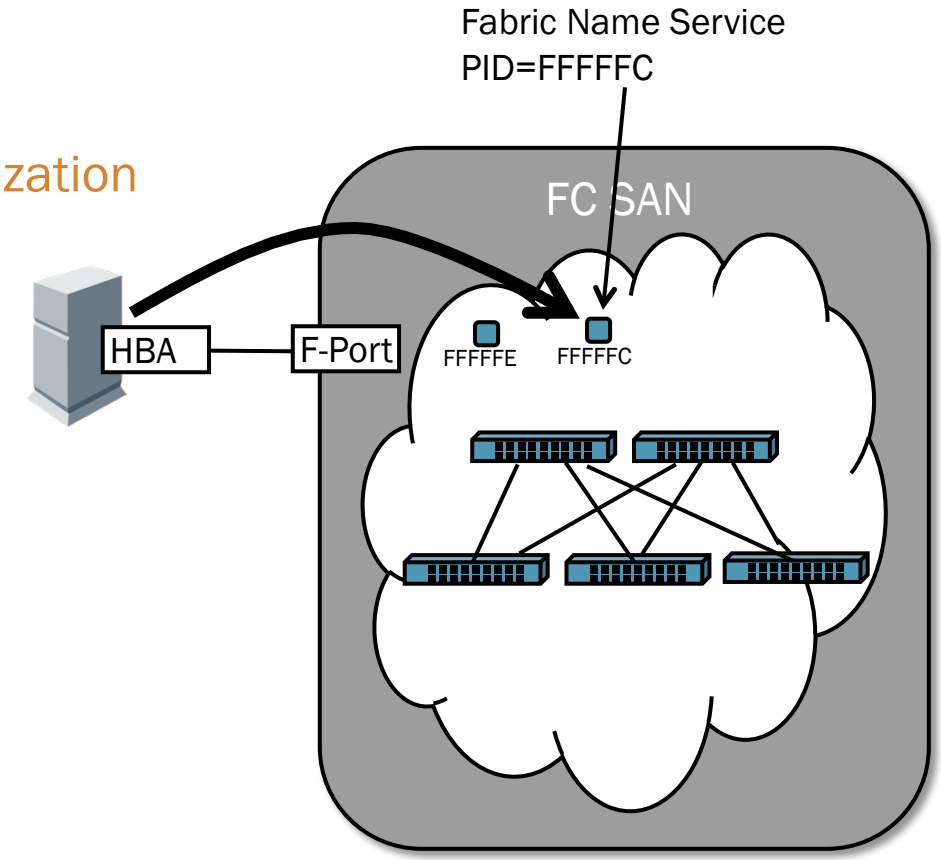
- The Fabric Login port responds with the hosts' newly-assigned PID.
  - Now the host has its PID! In this case it's x010100.
    - PID tells us:
      - Domain of the switch it's plugged into is "x01"
      - It's physical port number is "x01"
      - (Not hard and fast rule though)



Name Service

The Process of Device Initialization

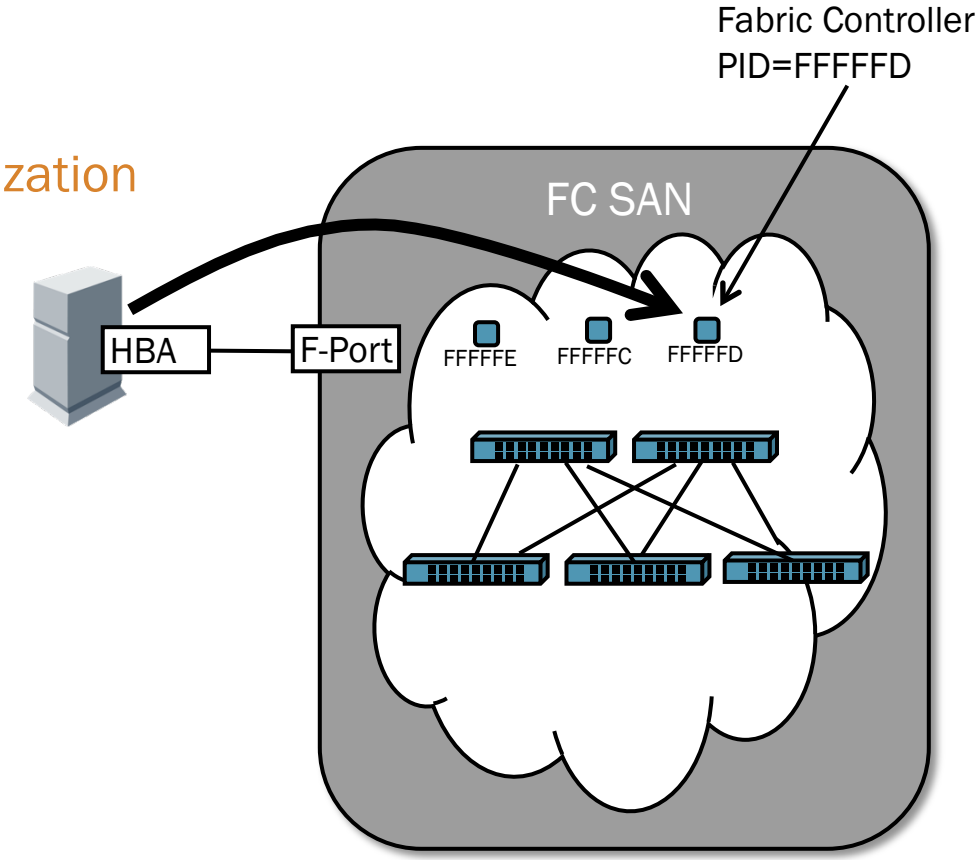
- Next, the host will login (PLOGI) to the port of the Name Service (FFFFFC).
  - It "registers" with the name service.
    - "Here's my PID, my WWNs, who made me (vendor specific information), more stuff, etc."



Fabric Controller

The Process of Device Initialization

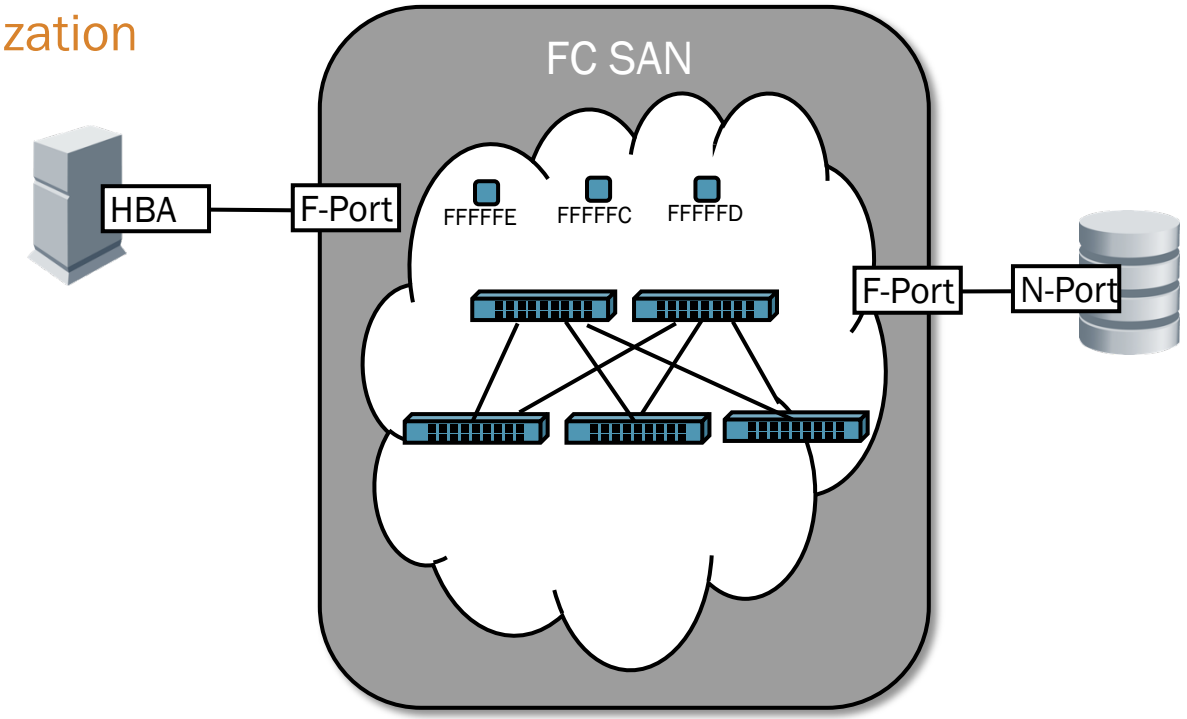
- Next, the host will PLOGI to the Fabric Controller (FFFFFD).
  - It "registers" for "State Change Notification".
    - "Mr. Fabric Controller, if anything changes in this network [that I need to know about] please notify me" I am Registering for "State Change Notification" (RSCN).



Fabric Controller

The Process of Device Initialization

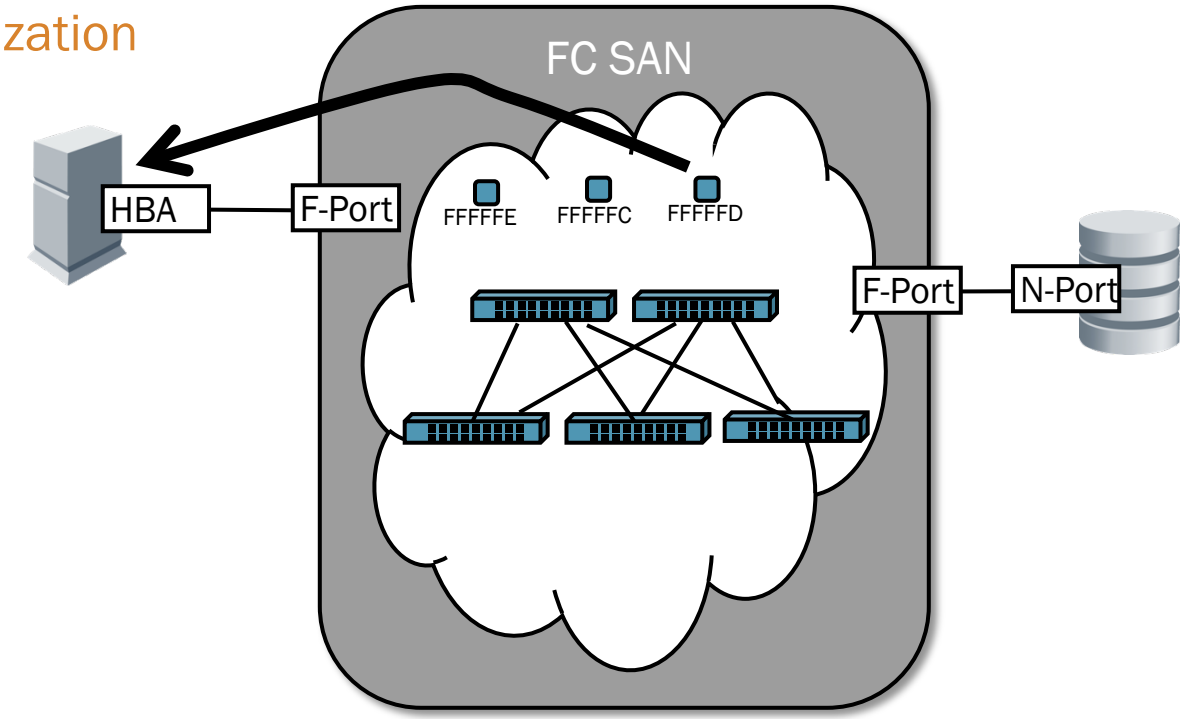
- Hey, what's that? A new storage device just came online, logged into the fabric and registered with the name service!



Fabric Controller

The Process of Device Initialization

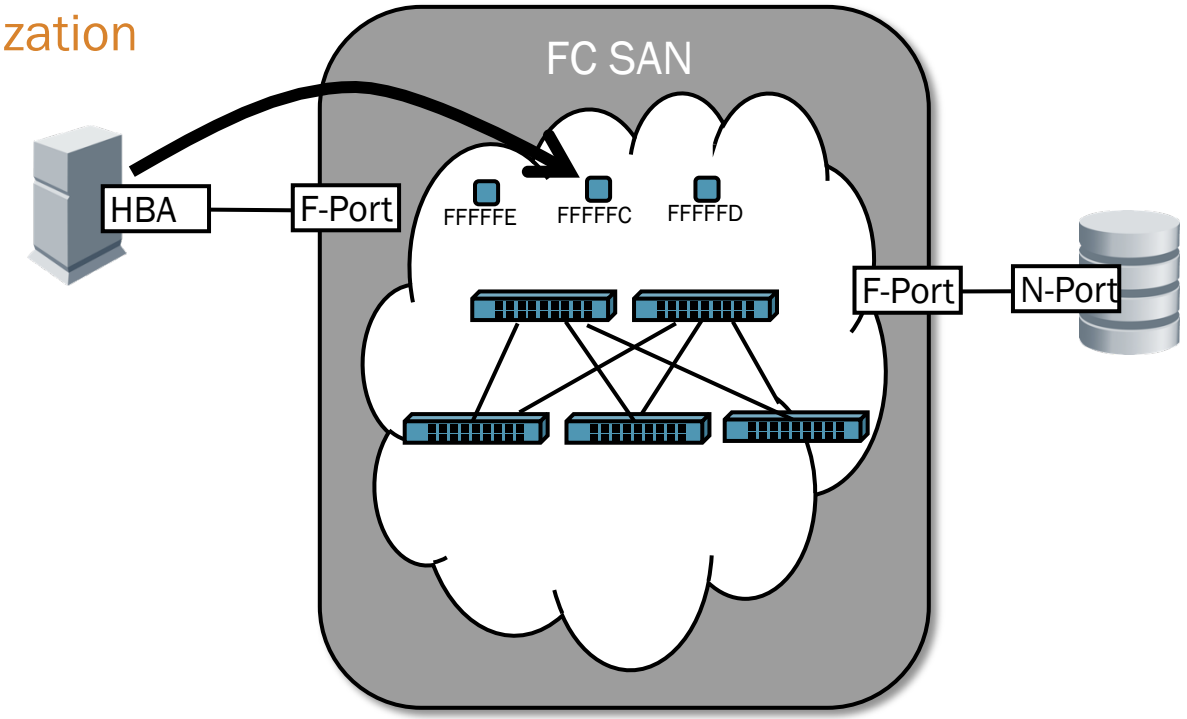
- If the new storage is in the same zone as the host, the fabric controller notifies the host of the change (State Change Notification).



Fabric Controller

The Process of Device Initialization

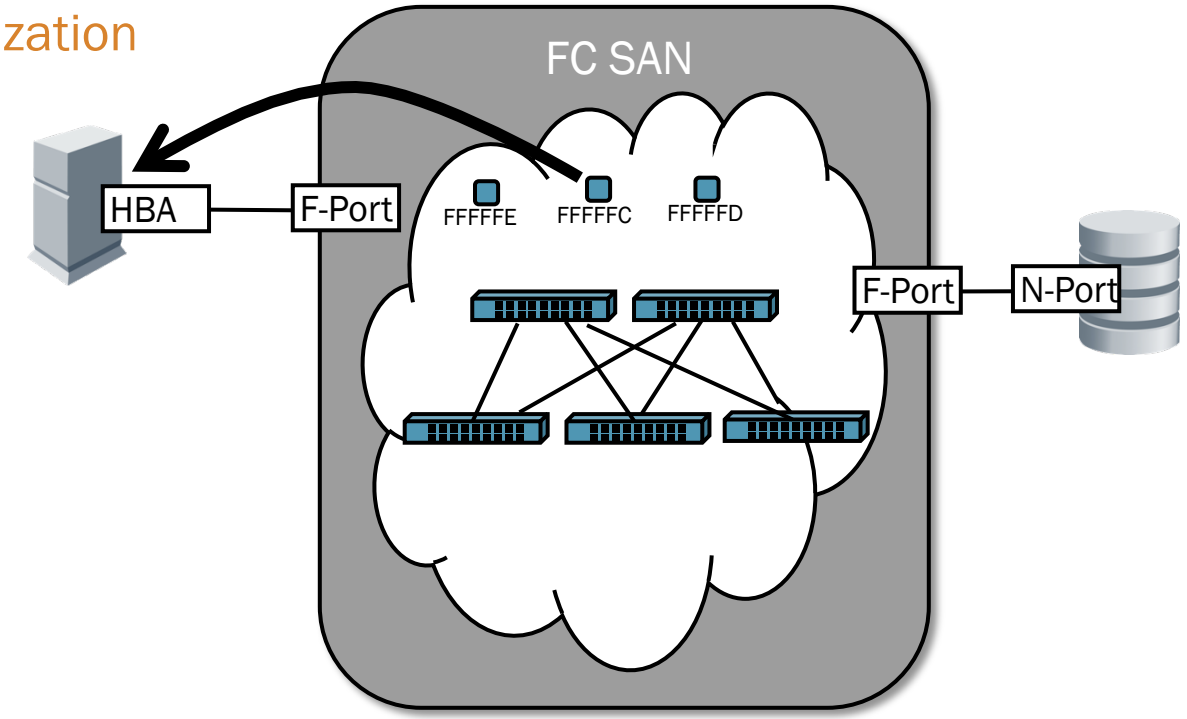
- The host PLOGI's back into the Name Service
  - Queries the NS for updated list of available devices



Fabric Controller

The Process of Device Initialization

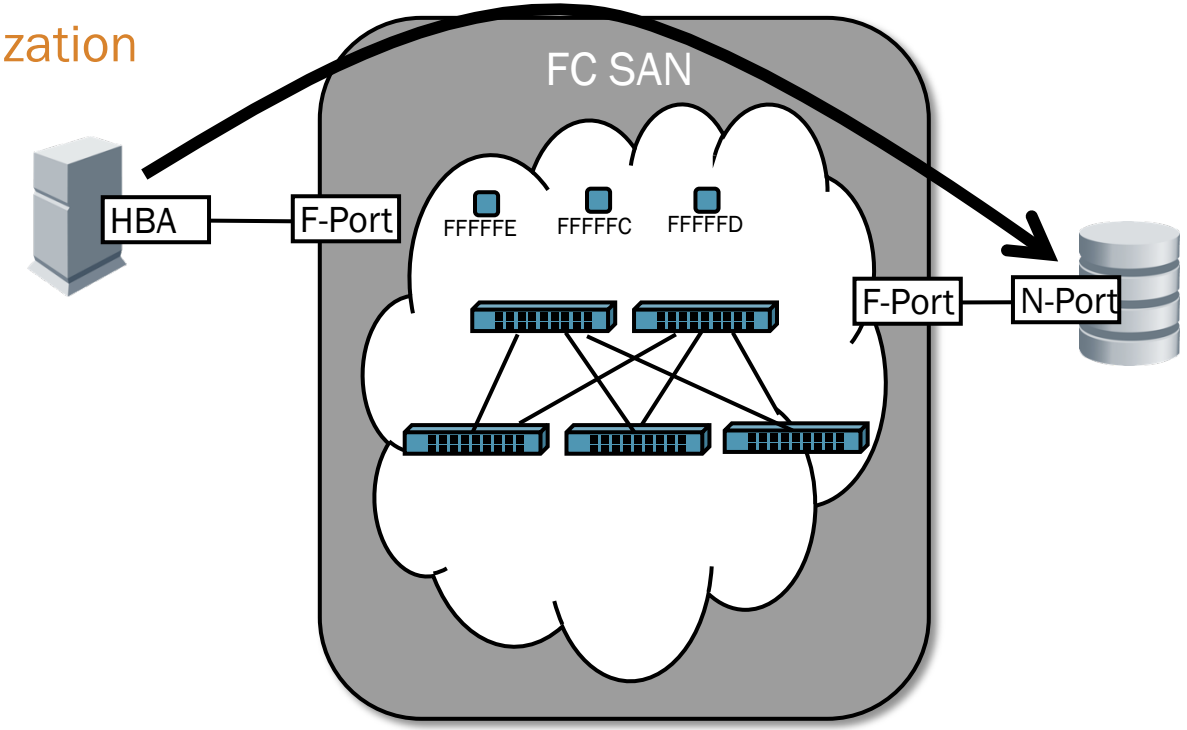
- Name service responds with a list of updated device PID's that are in the same zone as the host.



Fabric Controller

The Process of Device Initialization

- The host PLOGI's into the new storage device.
- Once logged in, the host can perform a SCSI probe to detect Logical Unit Numbers (LUN's).
- Once SCSI probe is complete, the storage can be formatted with a file system and mounted by the host's operating system.



Much More…

We've Only Scratched The Surface

- There is a lot more to it, but this is a good baseline from which to work.

DCB and Fibre Channel over Ethernet

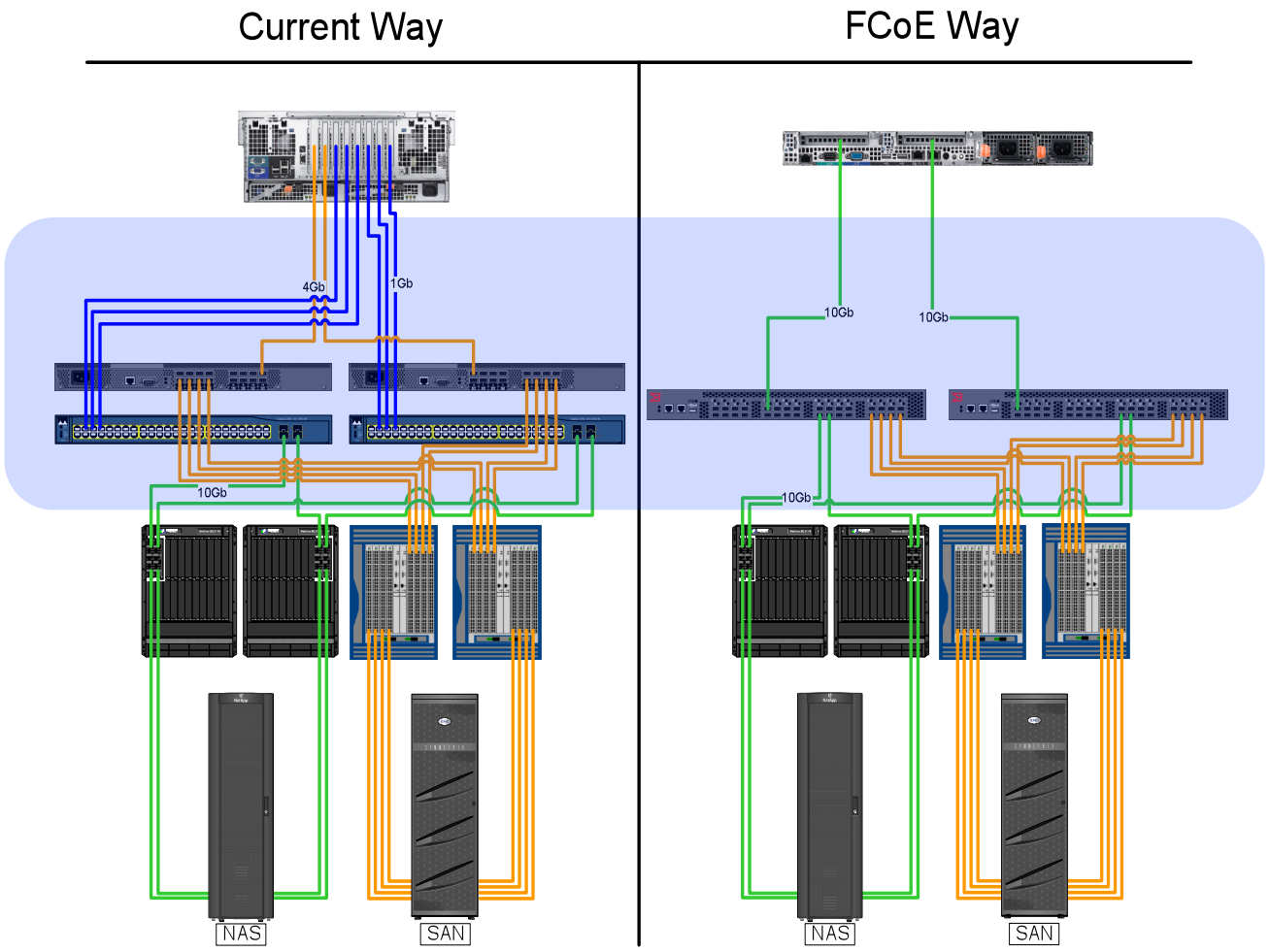
One Cable to Rule Them All, One Cable to Bind Them.....

Data Center Bridging (DCB)

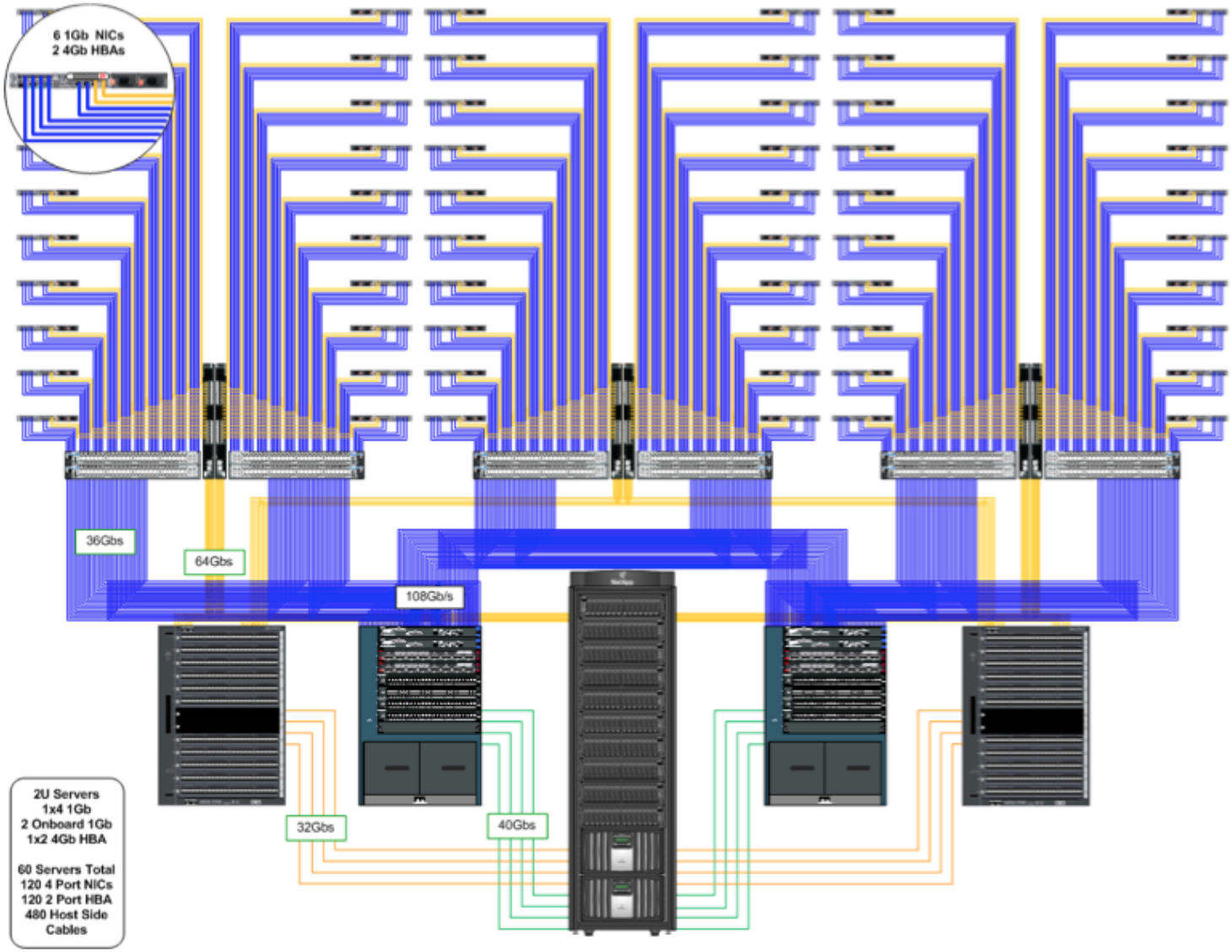
The New Sheriff In Town

- Traditional Ethernet is not suitable for transporting Fibre Channel frames.
  - Congestion
  - Latency
  - Frame Drop
- DCB is an umbrella term for Ethernet technology that has been enhanced by additional standards to meet the requirements for transporting FC frames.
  - Lossless, etc.

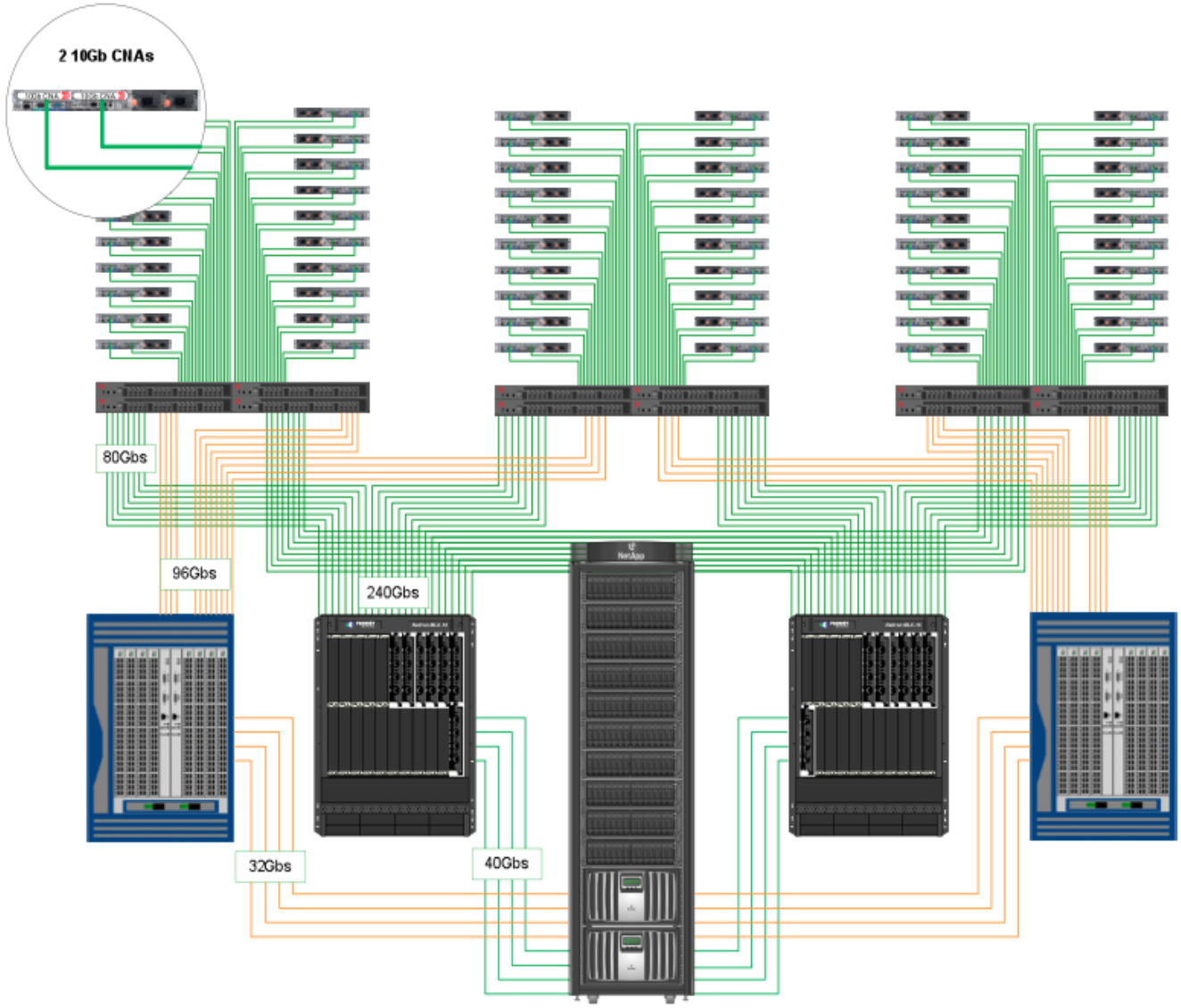
Standard SAN vs. Converged SAN



Example Environment
6 IP Connections and 2 FC per Server



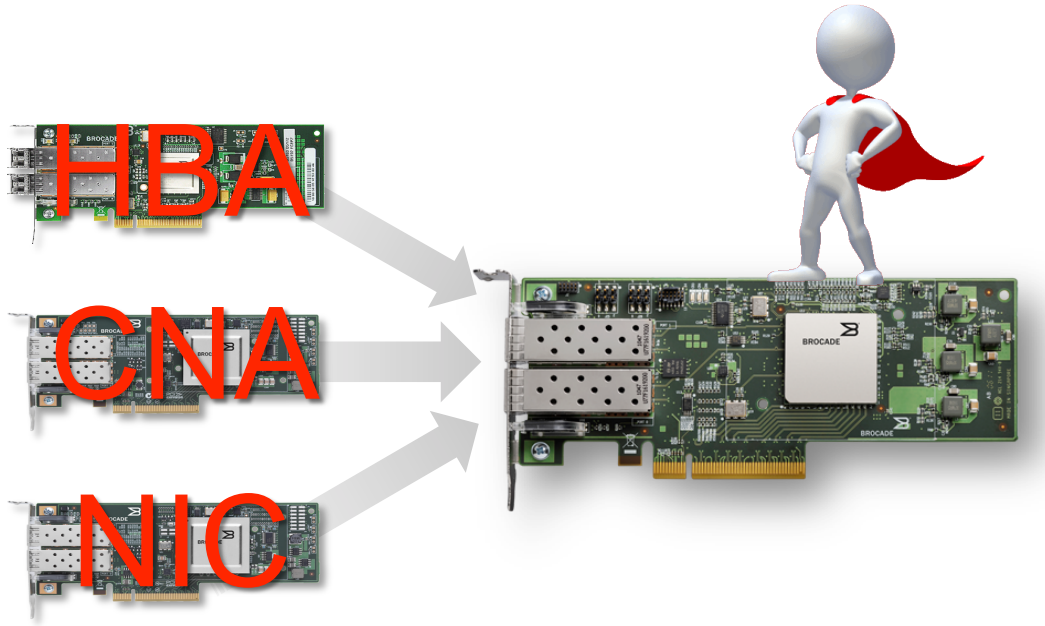
Same Example Environment
360 Host Side Cables Removed



Major DCB Enhancements to Ethernet

What's New?

- **Data Center Bridging eXchange (DCBX - IEEE 802.1Qaz)**
  - Leverages LLDP – 802.1AB
  - Used for configuring devices
- **Priority Flow Control (PFC – IEEE 802.1Qbb)**
  - Per Priority Pause Frames
  - Controlled independently for each Cos
  - Ensures Zero Loss
- **Enhanced Transmission Selection (ETS – IEEE 802.1Qaz)**
  - Bandwidth Management for different traffic flows

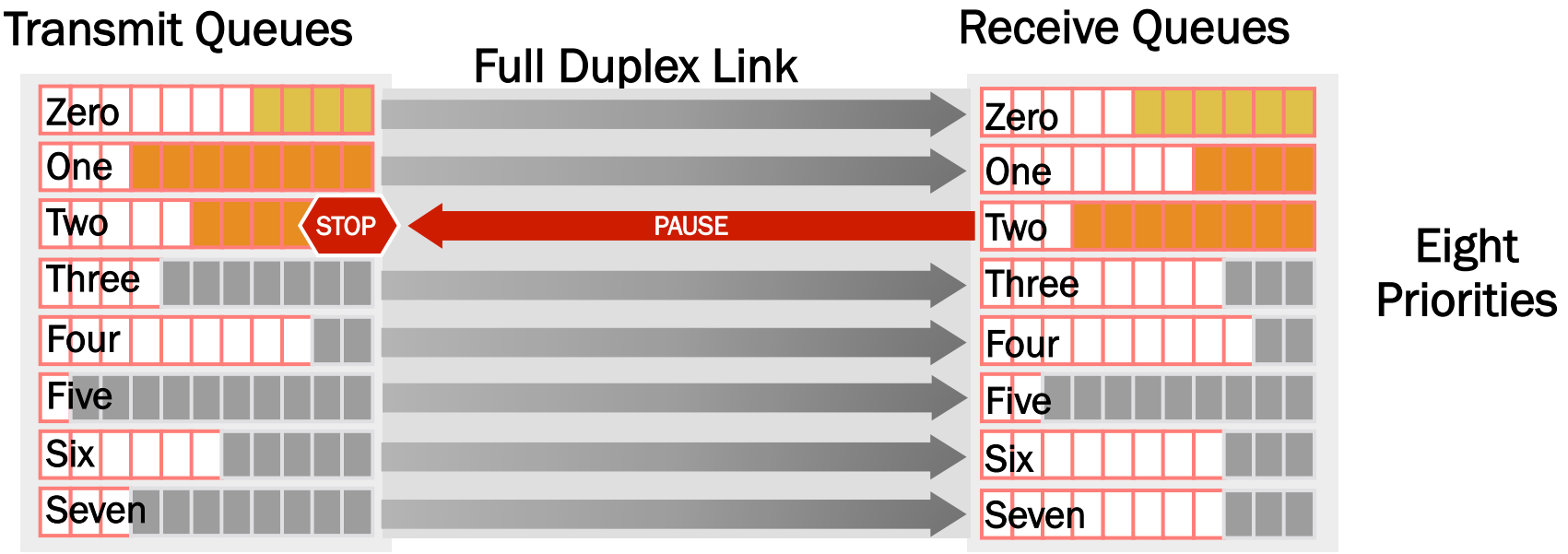


DCB Protocol Support

	Ethernet	DCB
IEEE 802.1Q VLAN Tagging	Yes	Yes
IEEE 802.1v VLAN Classification by Protocol & Port	Yes	Yes
IEEE 802.1p CoS	Yes	Yes
802.1x Network Access Control	Yes	Yes
IEEE 802.1D STP	Yes	Yes
IEEE 802.1W RSTP	Yes	Yes
IEEE 802.1s MSTP	Yes	Yes
IEEE 802.3ad LAG	Yes	Yes
IEEE 802.3x Flow Control (Link Level Pause Frames)	Yes	No
IEEE 802.1Qbb Priority Flow Control	No	Yes
IEEE 802.1AB Link Layer Discovery Protocol	Yes	Yes
IEEE 802.1Qaz DCBX and ETS	No	Yes

Priority Flow Control (PFC)

802.1Qbb

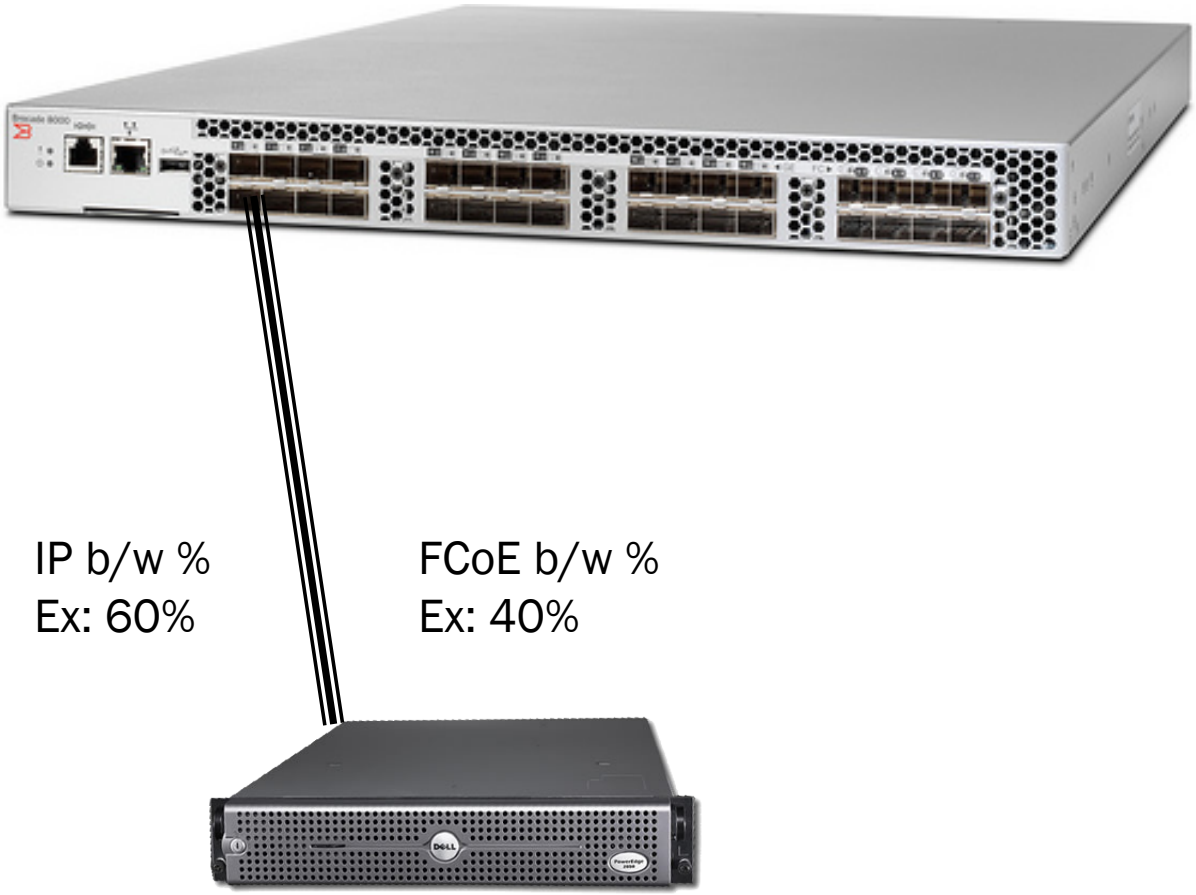


- Allow multiple kinds of traffic to be consolidated onto a single link
- Enables lossless capability for each class of service
- Network resources are partitioned between VL's
- PAUSE sent per virtual lane when buffers limit exceeded

Enhanced Transmission Selection (ETS)

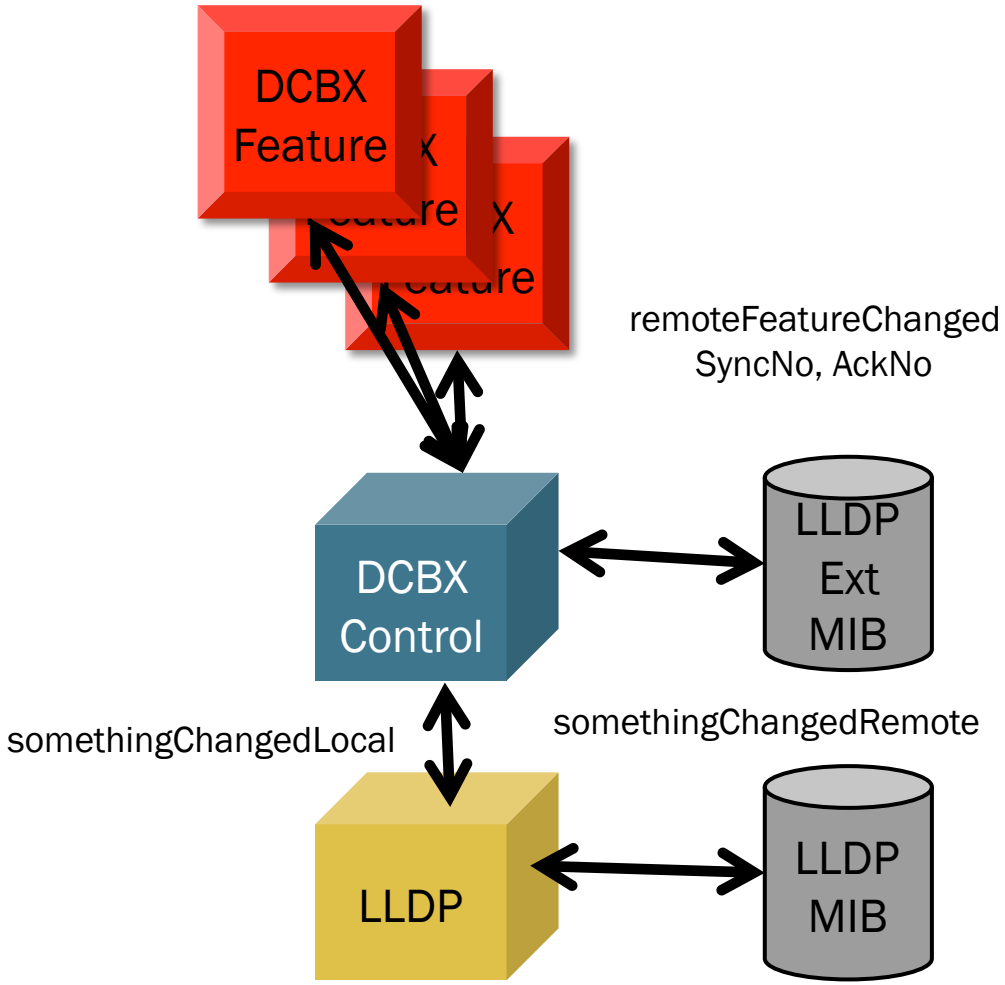
802.1Qbb

- Enhanced Transmission Selection allows multiple protocols, or traffic flows, to have different, or varying, portions of the bandwidth available on the transmission link assigned to them. Think “QoS” with priority grouping.



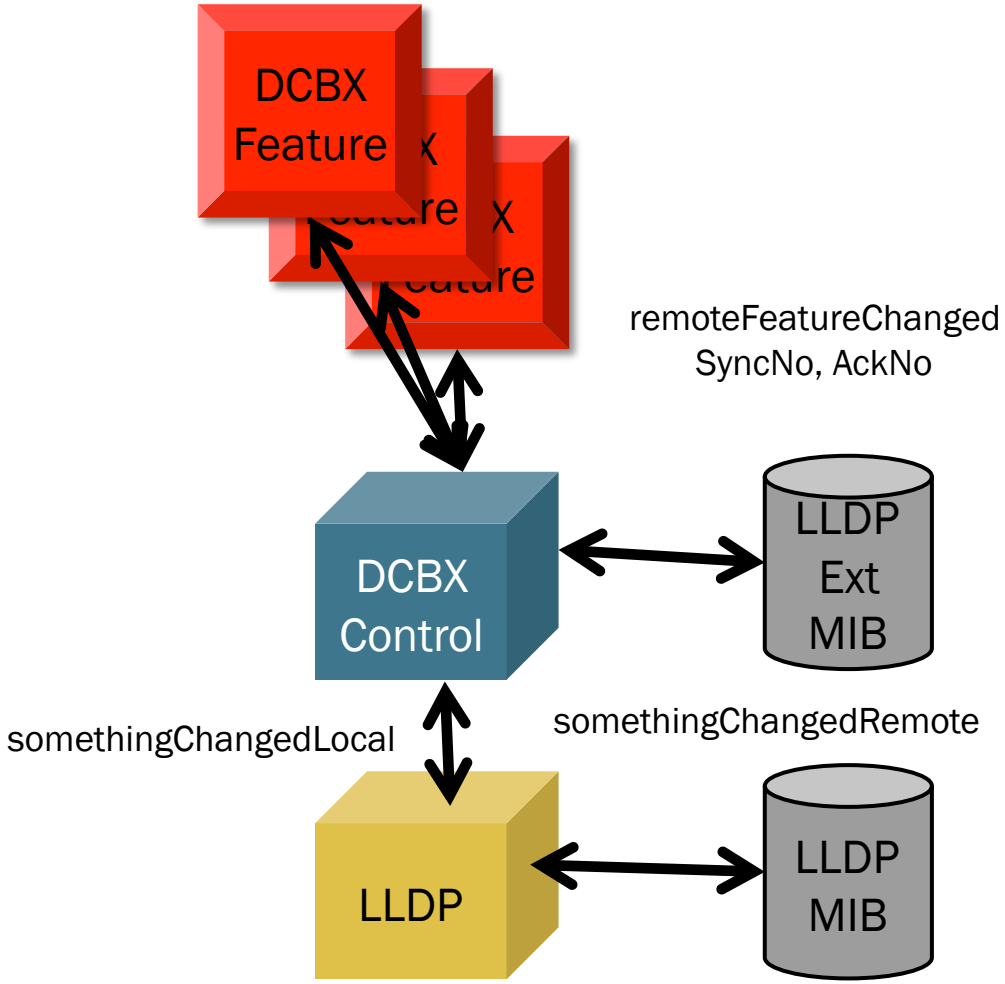
DCBX Operation

- LLDP is a one-way, unacknowledged protocol
- DCBX builds on top of LLDP to provide SeqNo and AckNo to create a reliable two-way handshake
- In addition to the LLDP TLVs, DCBX defines an additional TLV that contains ETS, PFC, and application configuration information



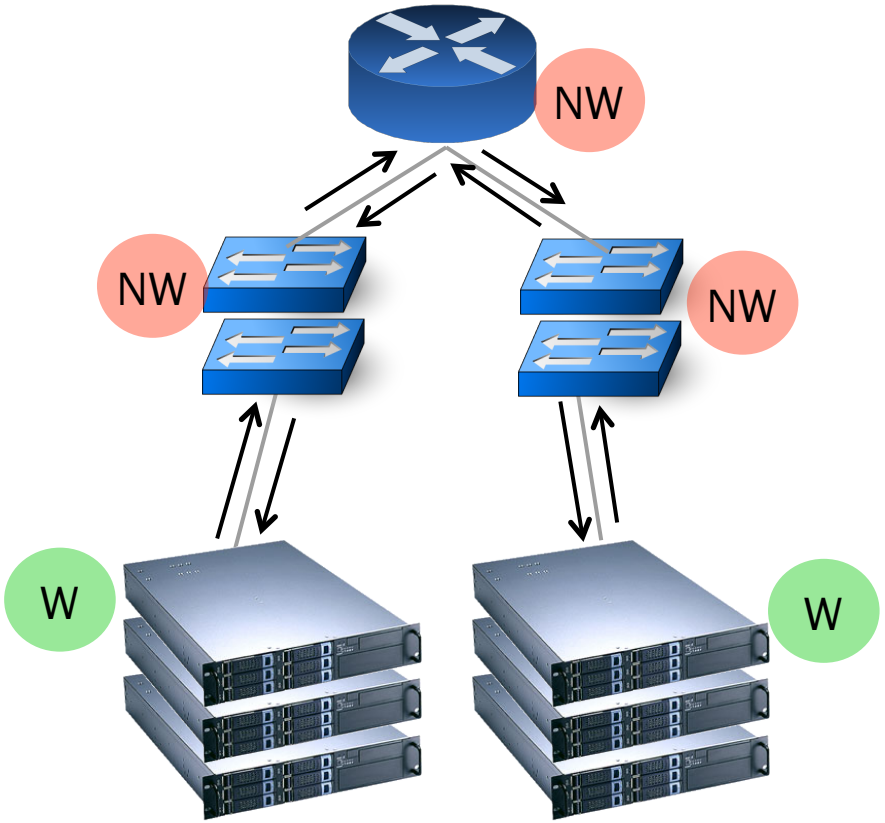
DCBX Operation

- The local system initializes and populates a seqNo in LLDP messages.
  - SeqNo is modified if the local configuration changes.
  - AckNo tells the peer the last seqNo that was received.
  - In this way, the system knows what information has been received by its peer.



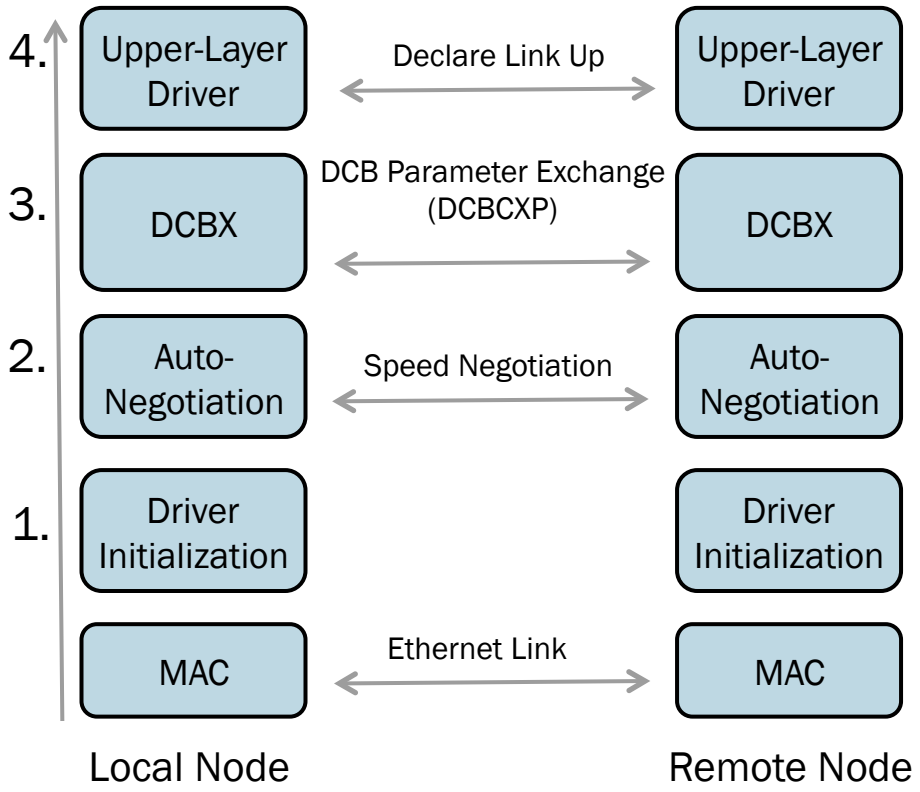
DCBX Configuration via “Willing” Option

- FCFs are normally set to “Not Willing”
  - DCBX and LLDP parameters will still be exchanged and compared
  - CNA’s should be set to “Willing” as this will allow them to accept the DCBX/LLDP and QoS configurations from the switch
    - If CNA’s are not willing, the configuration must match the FCF or a configuration mismatch will occur resulting in the CNA not being able to log into the fabrc
- Example: Brocade CNA’s are “Willing”



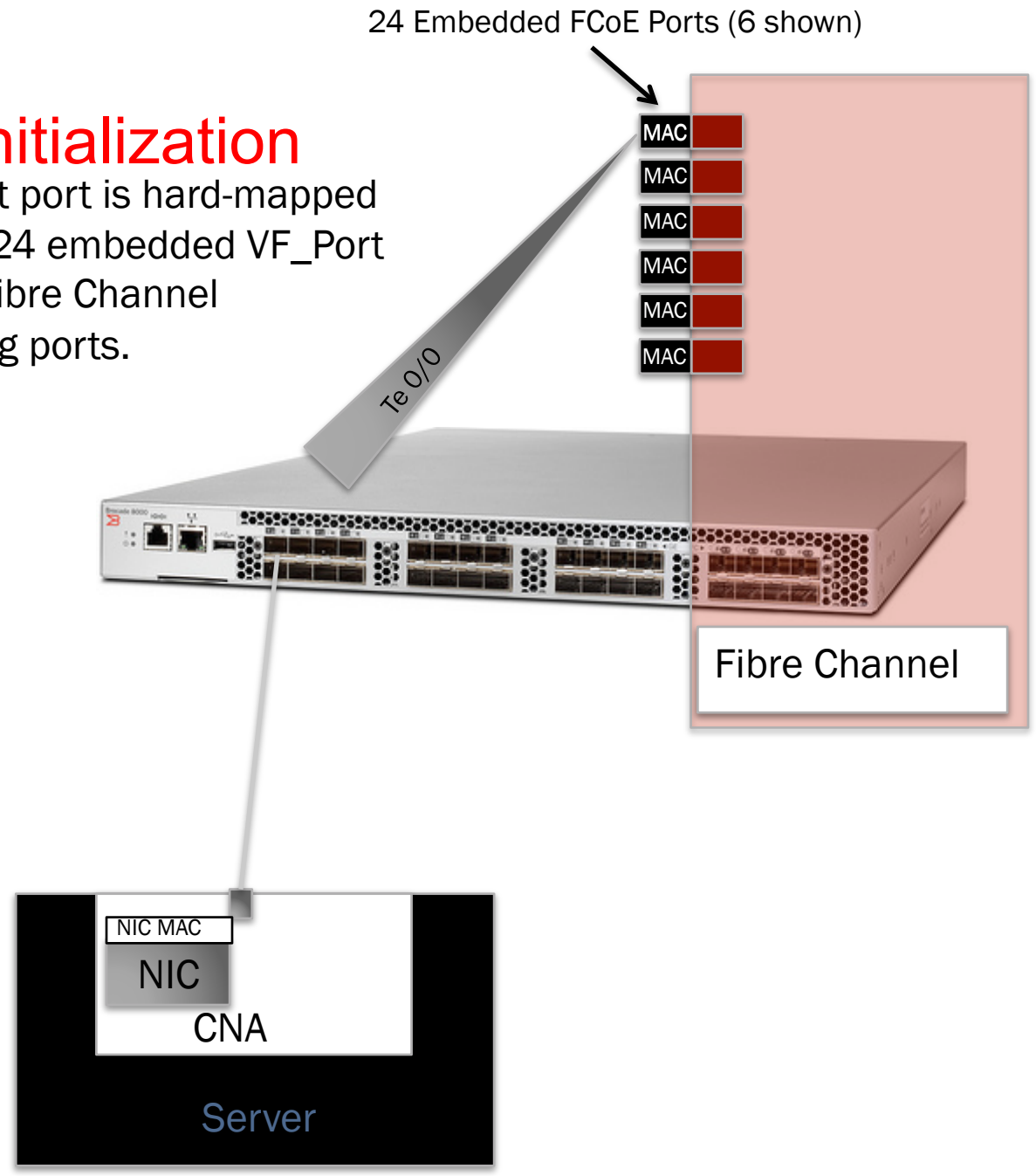
DCB Initialization

- DCBX process inserted into the CNA initialization process



FCoE Initialization

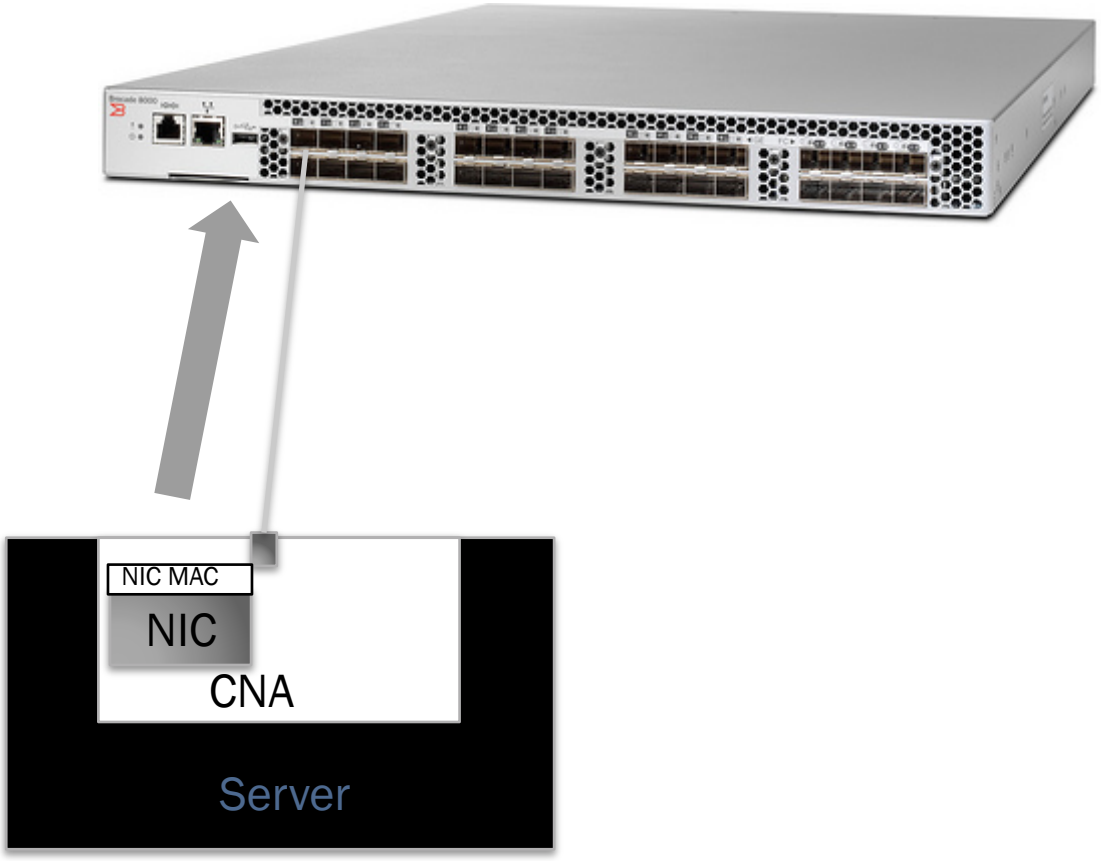
Each Ethernet port is hard-mapped to one of 24 embedded VF_Port capable Fibre Channel Forwarding ports.



FIP Overview

VLAN Discovery Phase

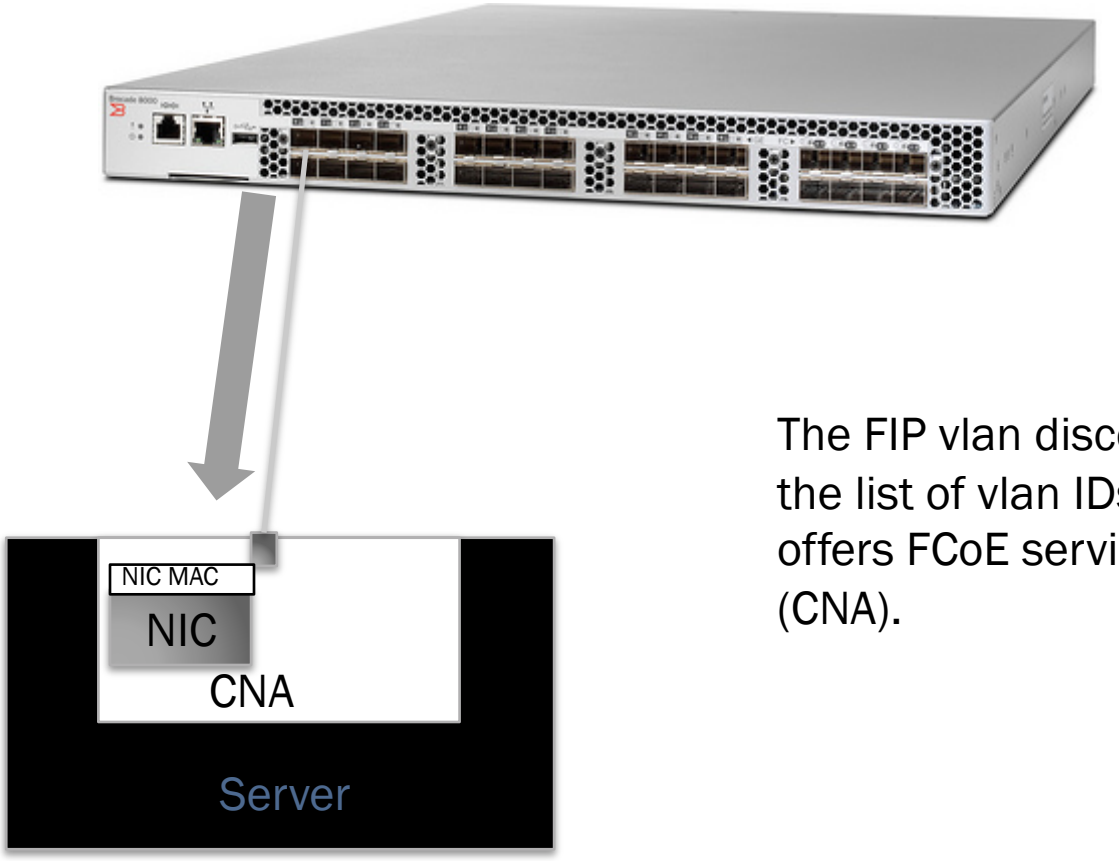
1. CNA sends VLAN Discovery Request to the All-FCoE Forwarders multicast address 01-10-18-01-00-02



FIP Overview

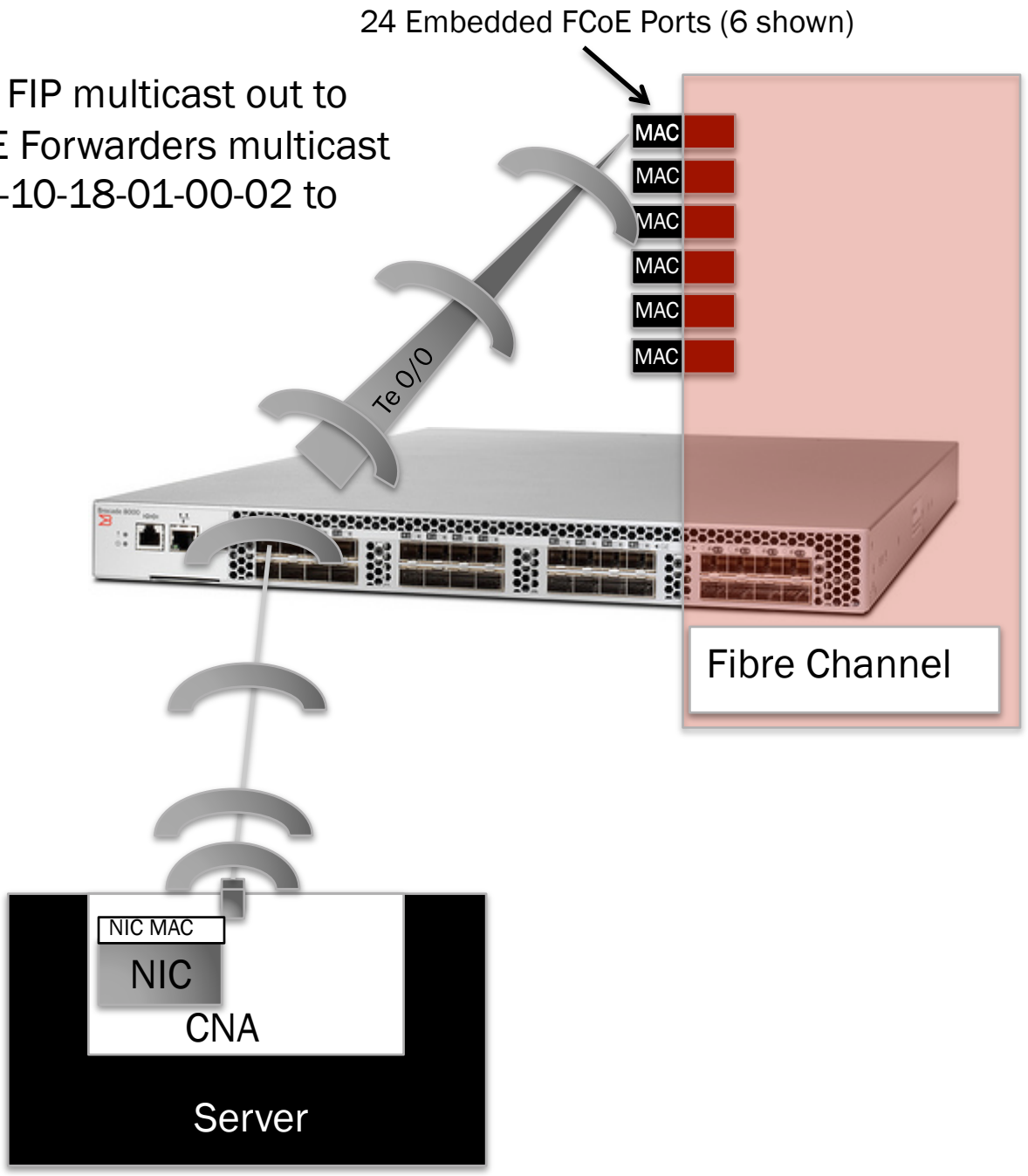
VLAN Discovery Phase

2. FCF responds with VLAN Discovery response frame with FCoE-enabled vlans.

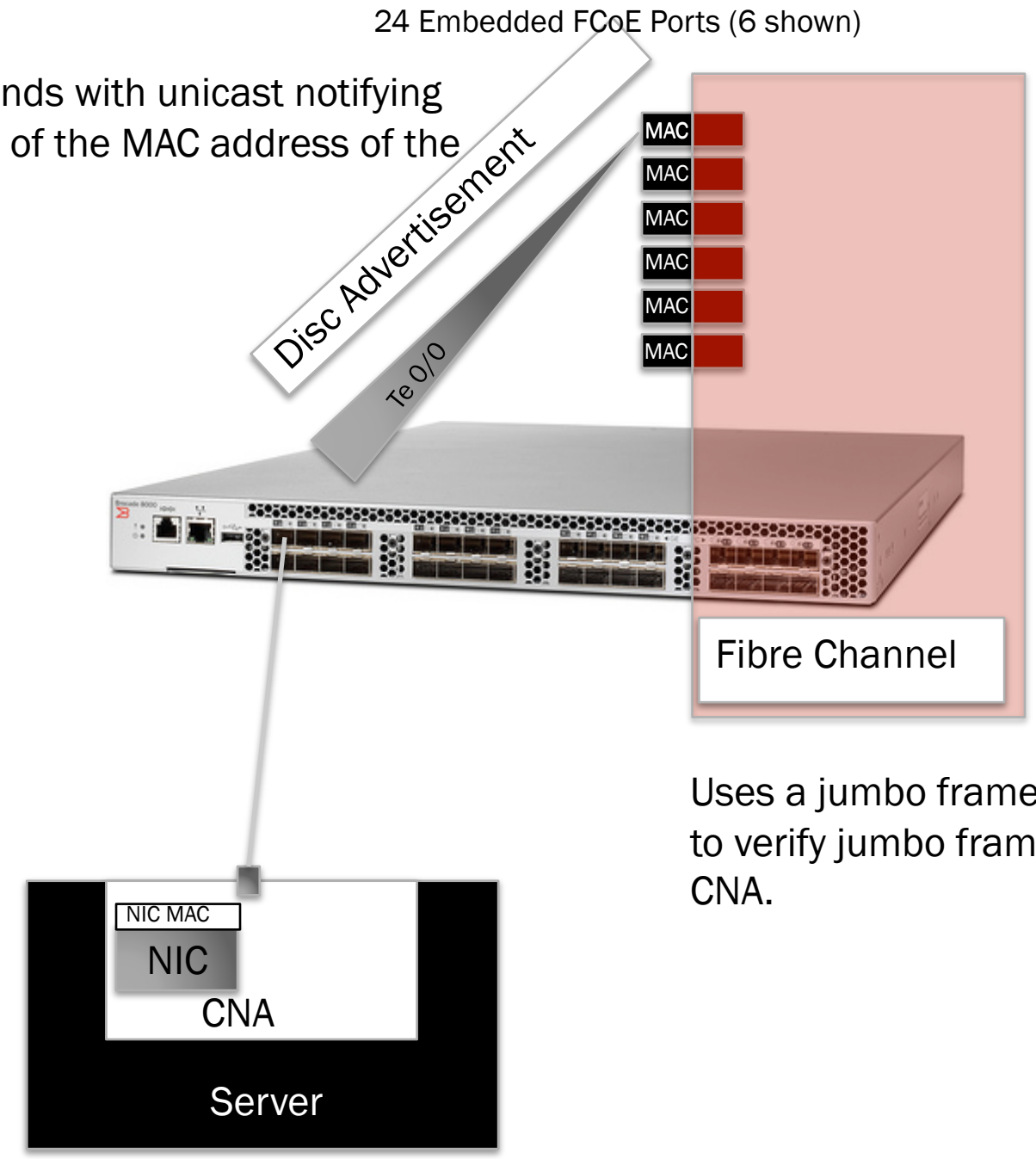


The FIP vlan discovery request carries the list of vlan IDs over which the FCF offers FCoE services to that enode (CNA).

1. Host sends FIP multicast out to the All-FCoE Forwarders multicast address 01-10-18-01-00-02 to find FCF



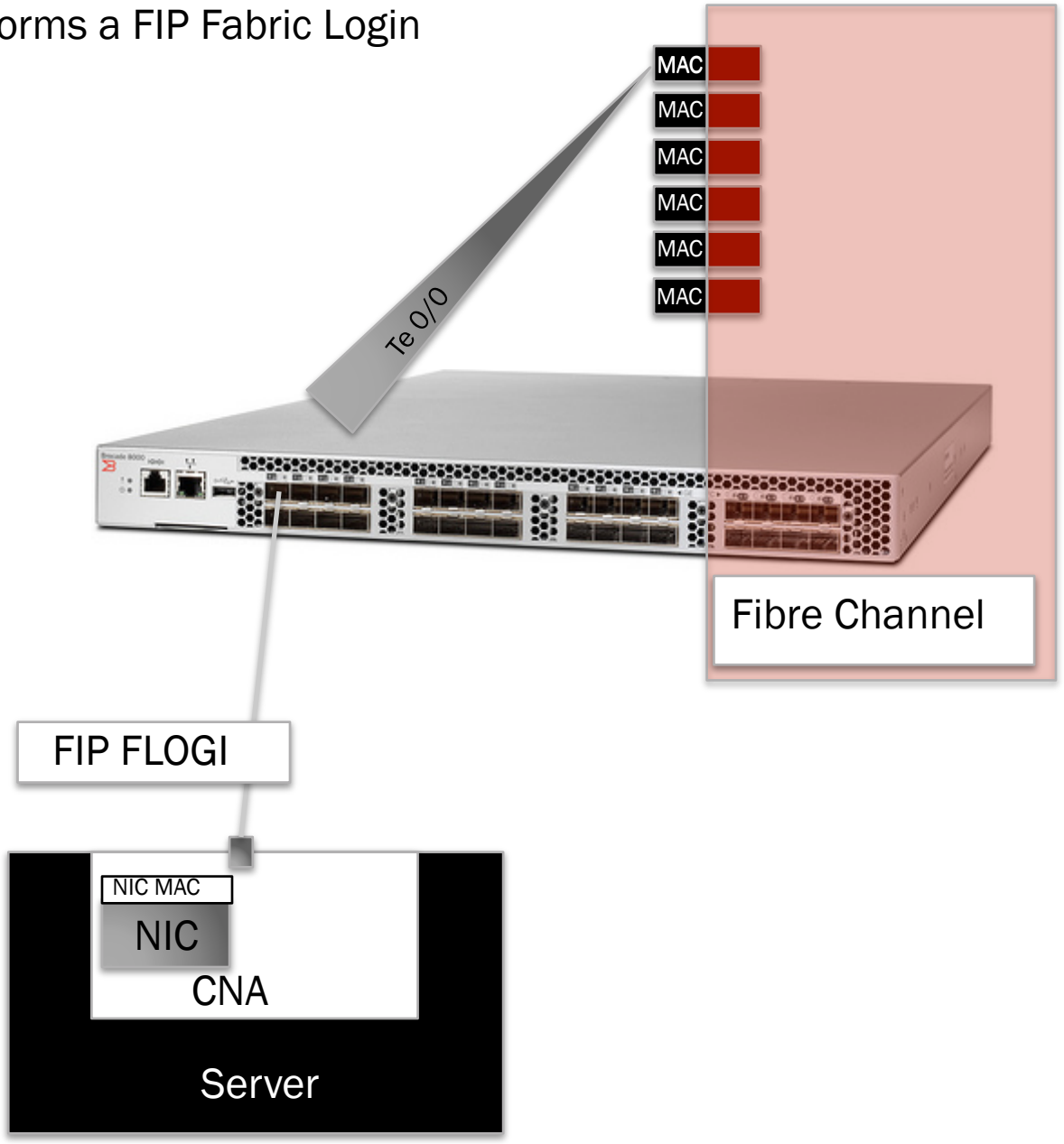
2. FCF responds with unicast notifying the Enode of the MAC address of the FCF.



Uses a jumbo frame for this response to verify jumbo frame support on the CNA.

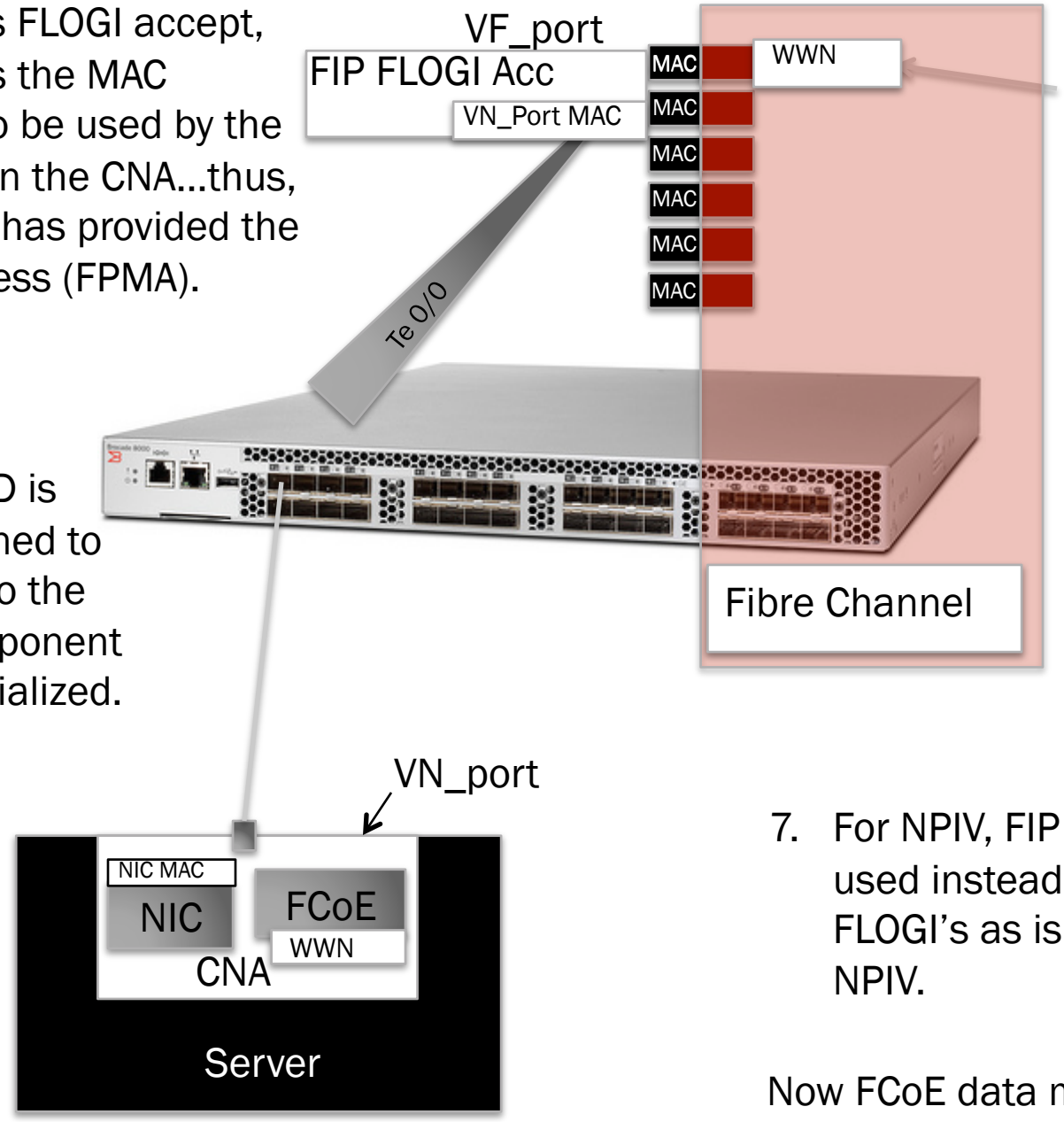
24 Embedded FCoE Ports (6 shown)

3. Enode performs a FIP Fabric Login



4. FCF sends FLOGI accept, and sends the MAC address to be used by the VN_port on the CNA…thus, the fabric has provided the mac address (FPMA).

5. The FC PID is also returned to the host so the FCoE component is now initialized.

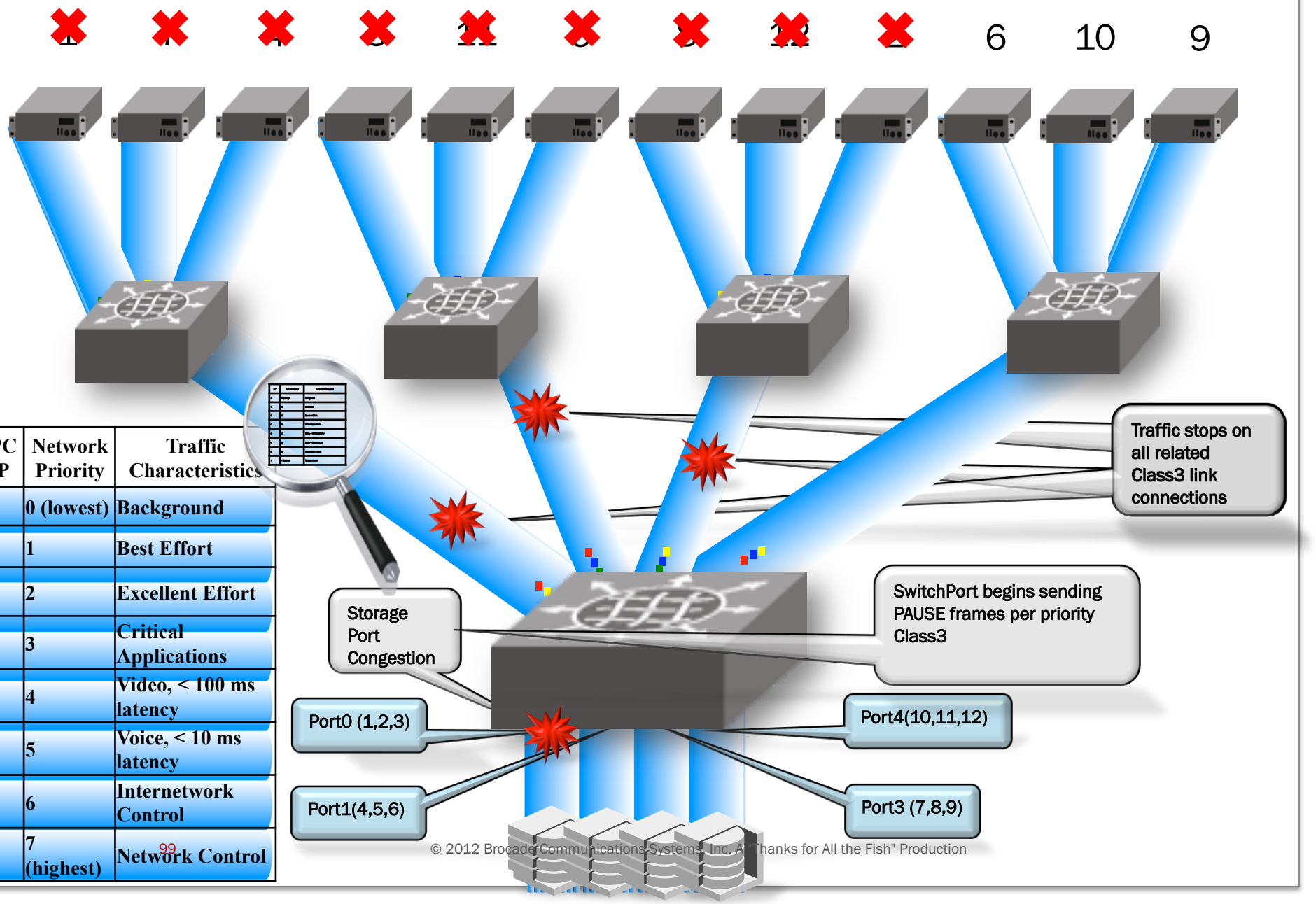


6. As usual in FC, once the FLOGI process is complete, the CNA's WWN will now be visible on the FC side. It can now be zoned to storage.

7. For NPIV, FIP FDISC can be used instead of traditional FLOGI's as is standard for NPIV.

Now FCoE data may pass.

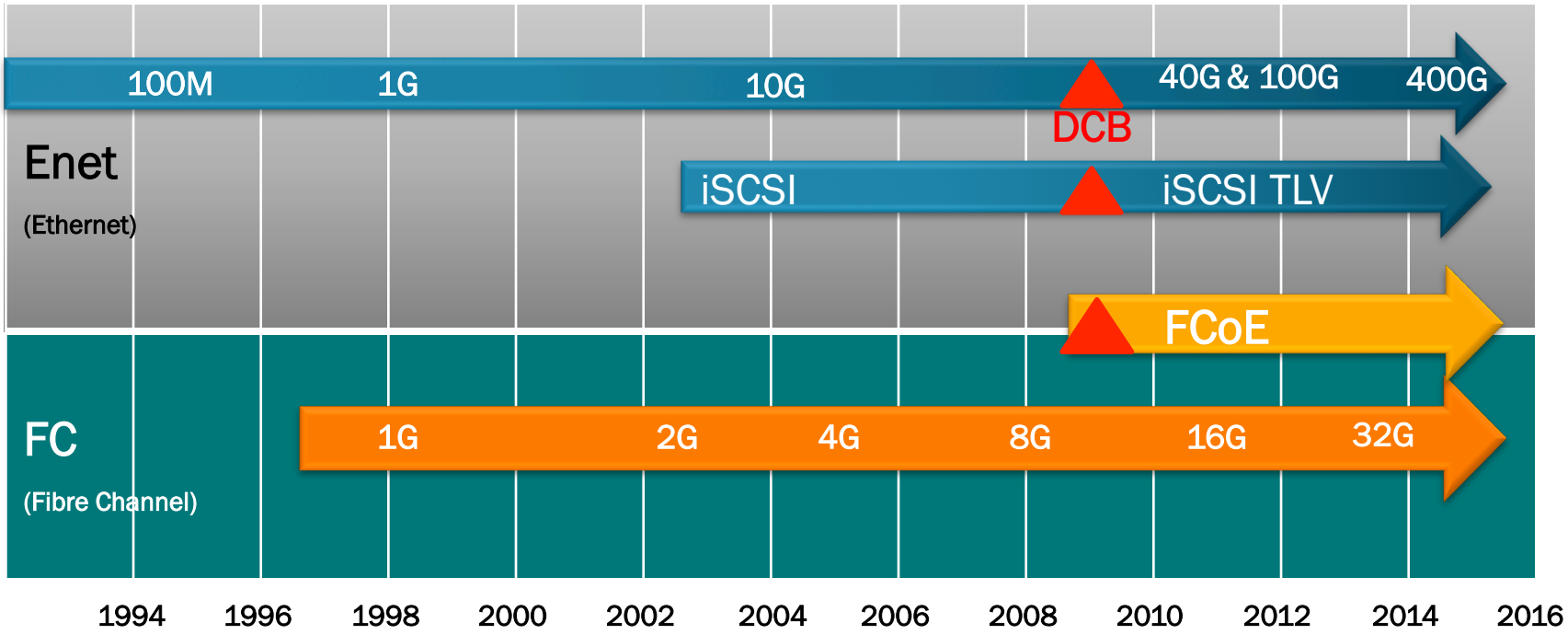
PFC: Issues with Per Priority Pause



PCP	Network Priority	Traffic Characteristics
1	0 (lowest)	Background
0	1	Best Effort
2	2	Excellent Effort
3	3	Critical Applications
4	4	Video, < 100 ms latency
5	5	Voice, < 10 ms latency
6	6	Internetwork Control
7	7 (highest)	Network Control

Ethernet and FC Roadmaps

Parallel Evolution & Potential for Convergence



- FC and Ethernet evolved in parallel paths with FC dominating storage SANs and Ethernet supporting IP networking
- Lossless Ethernet & FCoE open the door for server I/O consolidation

BROCADE

THANK YOU