Design and Implementations of FCoE for the DataCenter

Mike Frase, Cisco Systems

SNIA Legal Notice

Design and Implementations of FCoE for the DataCenter

This session will discuss the technologies and solutions for implementing FCoE in an end to end design for the Unified Data Center. The terms and technical requirements of the 10 GigE Network along with Multi-hop FCoE and Unified I/O Compute will be explained, allowing anyone new or old to the technology the ability to engage in the discussion. We will focus on what the Data Center Engineer can deploy in today's designs , intergrading FCoE with current Storage Area Networks (Native Fibre Channel) or complete new FCoE deployment. Attendees shall learn what new skills are required for the implementation and management of an FCoE virtualized network.

Learning Objectives:

◆ - FCoE technologies

◆ - Skills requirements for Unified I/O Networks

◆ - 10 GigE designs for FCoE

Agenda

◆ **Network, Storage, and Server Admin -** Positioning your skills

◆ **FCoE Technology**

◆ **FCoE Deployment Designs – Today**

◆ **FCoE Deployments Designs - Tomorrow**

Data Center Engineer Evolution

Network Data Engineer

- 1, 10 GigE
- Spanning-Tree Protocols
- Highly Available Networks
- Low Latency Switching
- Layer 2 & 3
- Network adaptors and drivers
- IP addressing IPV4 & IPV6
- IP Routing Protocols
- Standard Network services ARP, DHCP, PXE, LLDP
- QOS
- Security
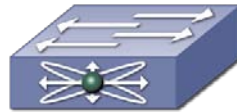
SAN Engineer

- 2, 4, 8 Gig FC
- Fibre Channel
- FSPF
- Highly Available Fabrics
- Low Latency Switching
- Fabric Login Services
- FCID
- Domain Services
- Zone Service
- Storage
- HBA's and drivers
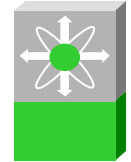
DC Engineer

◆ **Data Center Engineer**

- LAN Engineer
- SAN Engineer
- DCB (Lossless) GigE
- CNA adaptors & drivers

Technologies for LAN & SAN
   iSCSI
   FCoE
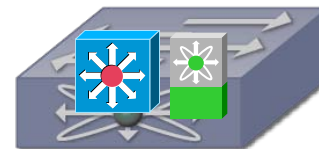
◆ Add these extras for **Advanced DC Engineer**

- Virtualization
- Server Hardware (Memory, CPU, PCIe)
- Data Center Interconnect Technologies

Learning Curve
Unified Network Technologies

Ethernet

Layer 2 & 3 protocols

VLANS

LAN

SAN A

SAN B

Fibre Channel Protocol and FC services

Zoning

SCSI data

Virtual Fabrics

Cabling and Optics

10 GigE

HBA

FCoE

DCB

Performance Monitoring

CNA

CNA

O/S & Driver support

Design and Implementations of FCoE for the DataCenter
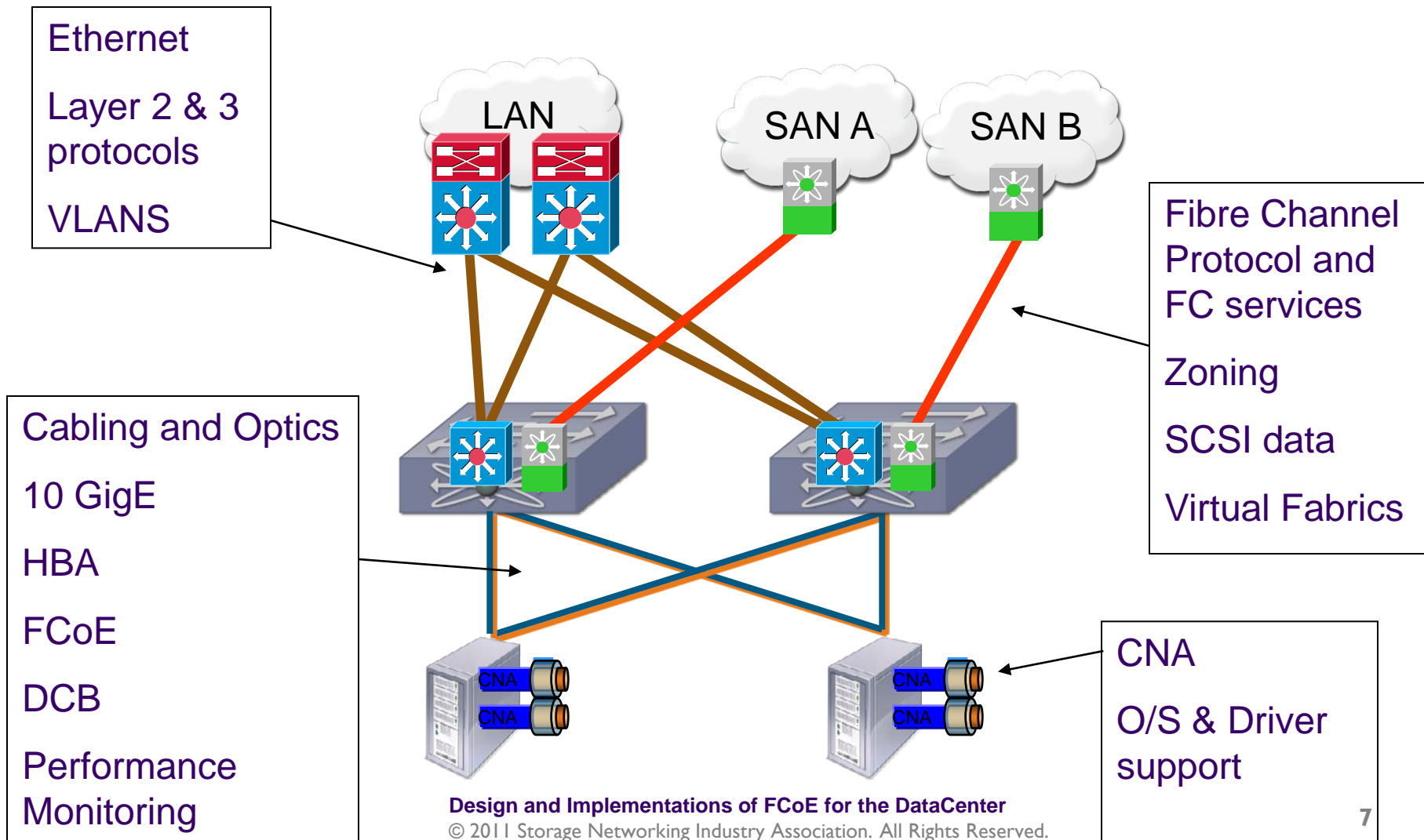
Unified I/O: Network Operations View

Server Ops

- Work Processes: Same as today
     Converged Network Adapter (CNA)
  config equivalent to NIC and HBA.
- Mgmt: No impact
- Network Mgmt System:
     Existing tools

SAN B

SAN

SAN Ops

- Work Process: Same as today
     Configure FC Interface mode, speed,
  Virtual Fabrics, Port Channels, Trunks,
  NPV etc.
     Monitor SAN specific MIBs, Syslog,
  RMON.
- Mgmt :
     CLI (Role based)
     XML API
- Network Mgmt System:
     Existing tool – SAN Fabric Managers

Converged Activities

- Work Processes:
     Switch reboot
     Image upgrade
     User roles
     Initial creation of virtual interfaces
- Mgmt: Creation of Mgmt Interface
- Day-2 System level support: Memory,
  CPU error, Call Home features, Span ports.

——	1 GbE
——	10 GbE
——	FC
——	10GbE FCOE

Network Ops

- Work Process: Same as today
     Configure Ethernet port, VLAN, STP,
  Port Channels, Trunks etc.
     Monitor LAN specific MIBs, Syslog,
  RMON.
- Mgmt :
     CLI (Role based)
     XML API
- Network Mgmt System:
     Network/Device Management

Design and Implementations of FCoE for the DataCenter

FCOE TECHNOLOGY

FCoE Behind the Scenes
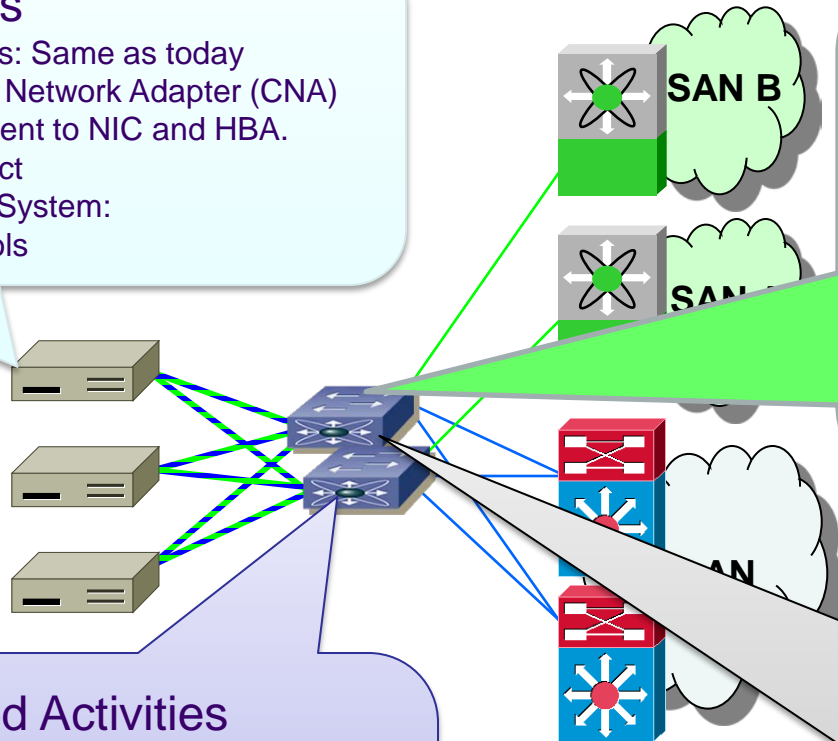
◆ **Data Center Bridging eXchange (DCBX):**

- **DCBX**, is a management protocol that extends the Link Layer Discovery Protocol (LLDP) defined in IEEE802.1AB. DCBX allows the Fibre Channel Forwarder (FCF) to provide Link Layer configuration information to the Converged Network Adaptor (CNA) and allows both the CNA and FCF to exchange status.

◆ **FCoE Protocols:**

- **FCoE   Eternet Frame Type 8906 Hex**

  FCoE data protocol requires lossless Ethernet and is typically implemented in hardware, and is used to carry most of the FC frames and all the SCSI traffic.

- **FIP (FCoE Initialization Protocol) Ethernet Frame Type 8914 Hex**

  A Discovery Protocol whose purpose is to present a multi-access Ethernet network as a set of point-to-point Virtual Links for FCoE to use.

DataCenter Bridging (DCB) for FCoE

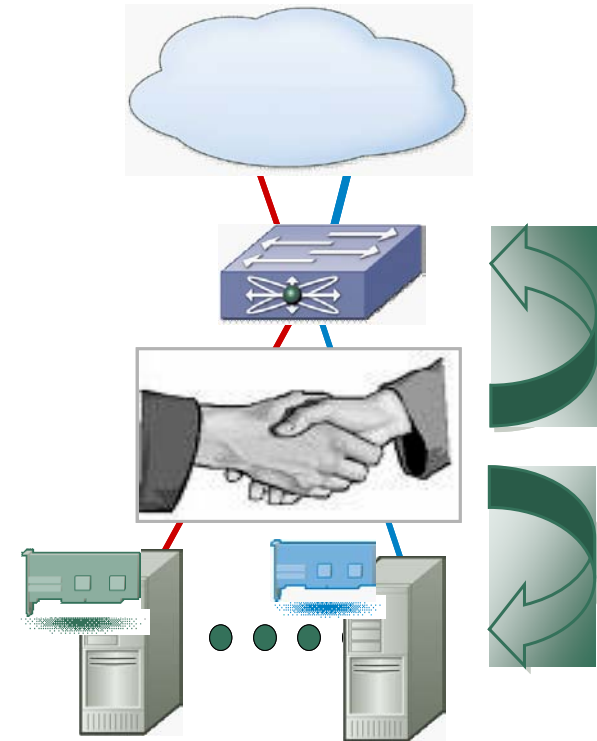◆ **Part of the IEEE 802.1 Work Group**

  ◆ Lossless Ethernet 802.1Qbb

  ◆ Priority Grouping   802.1Qaz
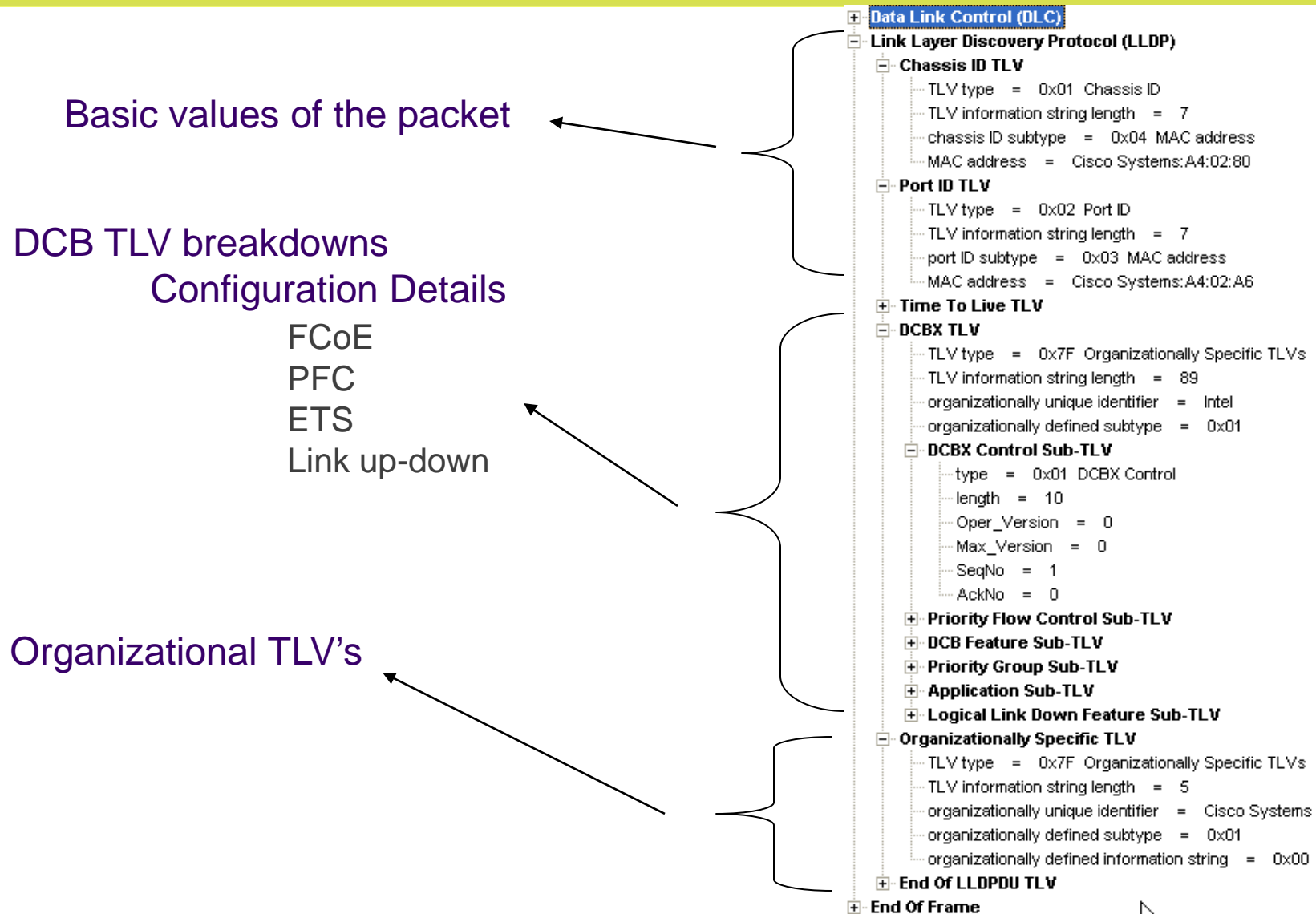
  ◆ Configuration Verification 802.1Qaz

◆ **Auto-negotiation of capability and configuration**

◆ **Allows network devices to advertise their capabilities over a link**

  ◆ Enables hosts to get proper configuration from network

  ◆ Enables switches to verify proper configuration

LLDP & DCB inspection

Basic values of the packet

DCB TLV breakdowns
      Configuration Details
            FCoE
            PFC
            ETS
            Link up-down

Organizational TLV's

```
⊞ Data Link Control (DLC)
⊟ Link Layer Discovery Protocol (LLDP)
   ⊟ Chassis ID TLV
       TLV type  = 0x01 Chassis ID
       TLV information string length = 7
       chassis ID subtype = 0x04 MAC address
       MAC address = Cisco Systems:A4:02:80
   ⊟ Port ID TLV
       TLV type = 0x02 Port ID
       TLV information string length = 7
       port ID subtype = 0x03 MAC address
       MAC address = Cisco Systems:A4:02:A6
   ⊞ Time To Live TLV
   ⊟ DCBX TLV
       TLV type = 0x7F Organizationally Specific TLVs
       TLV information string length = 89
       organizationally unique identifier = Intel
       organizationally defined subtype = 0x01
      ⊟ DCBX Control Sub-TLV
          type = 0x01 DCBX Control
          length = 10
          Oper_Version = 0
          Max_Version = 0
          SeqNo = 1
          AckNo = 0
      ⊞ Priority Flow Control Sub-TLV
      ⊞ DCB Feature Sub-TLV
      ⊞ Priority Group Sub-TLV
      ⊞ Application Sub-TLV
      ⊞ Logical Link Down Feature Sub-TLV
   ⊟ Organizationally Specific TLV
       TLV type = 0x7F Organizationally Specific TLVs
       TLV information string length = 5
       organizationally unique identifier = Cisco Systems
       organizationally defined subtype = 0x01
       organizationally defined information string = 0x00
   ⊞ End Of LLDPDU TLV
⊞ End Of Frame
```

Design and Implementations of FCoE for the DataCenter

FIP Protocols Defined

- FIP VLAN Discovery - used by the ENode to discover FCOE VLAN
- FIP Discovery Solicitations - used by the ENode to discover a FCF
- FIP Discovery Advertisement - used by FCF to advertise itself to ENodes
- FIP FLOGI, FIP LS_ACC and FIP LS_RJT - used by FCF to login and get itself an FCID
- FIP LOGO - sent by ENode to logout
- FIP NPIV FDISC - sent by VN ports to login
- FIP KeepAlives - liveliness mechanism between ENodes and FCF
- FIP Clear Virtual Link - used by the FCF to signal to the ENode to bring down the virtual link

FIP: FCoE Initialization Protocol

❖ **FCoE VLAN discovery**

  ◆ Automatic discovery of FCoE VLANs

❖ **Device discovery**

  ◆ ENodes discover VF_Port capable FCF-MACs for VN_Port to VF_Port Virtual Links

  ◆ VE_Port capable FCF-MACs discover other VE_Port capable FCF-MACs for VE_Port to VE_Port Virtual Links

  ◆ Verifies the Ethernet network supports the required Max FCoE Size along with many other settings like FC map value using Descriptor TLV's
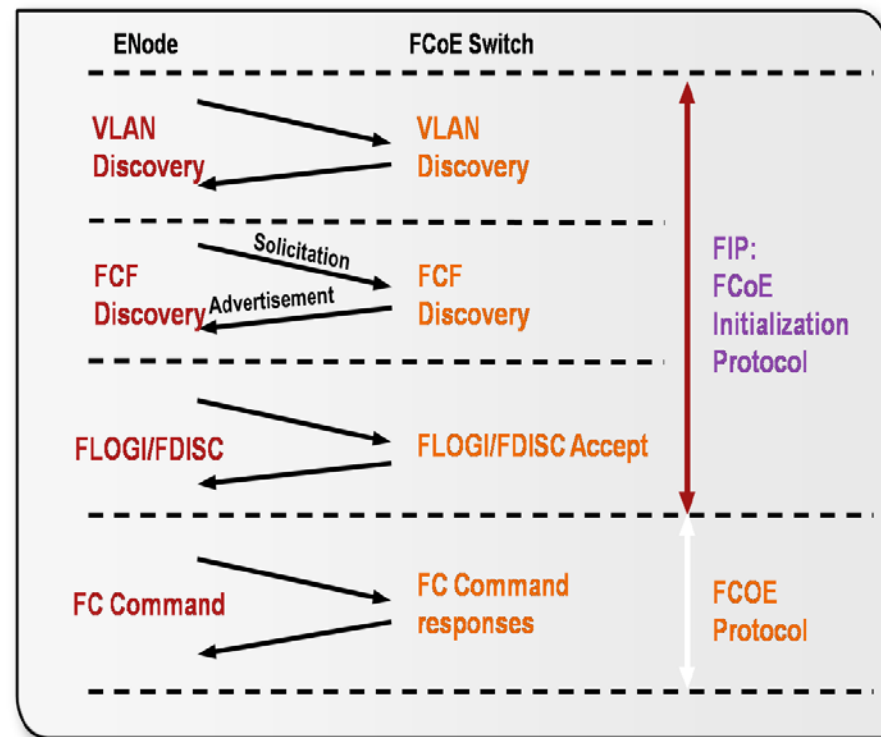

Virtual Link instantiation

  • Builds on the existing Fibre Channel Login process

    • Fabric Provided MAC Address (FPMA),

Virtual Links maintenance

  • Timer based

14

FCoE, Same Model as FC

- ❖ **A-B SAN separation** – picture showing one side
- ❖ **Same host to target communication**
  - Host has 2 CNA's (one per fabric)
  - Target has multiple ports to connect to fabric
- ❖ **Connect to a DCB capable switch**
  - Port Type Negotiation (FC port type will be handled by FIP)
  - Speed Negotiation
  - DCB Negotiation
- ❖ **Access switch is a Fibre Channel Forwarder (FCF)**
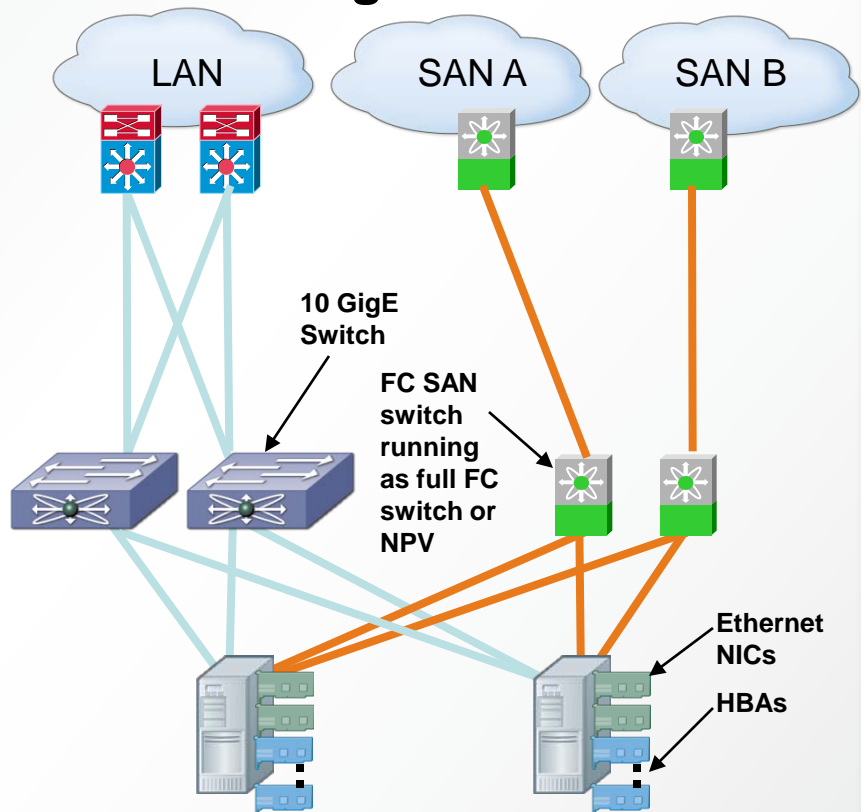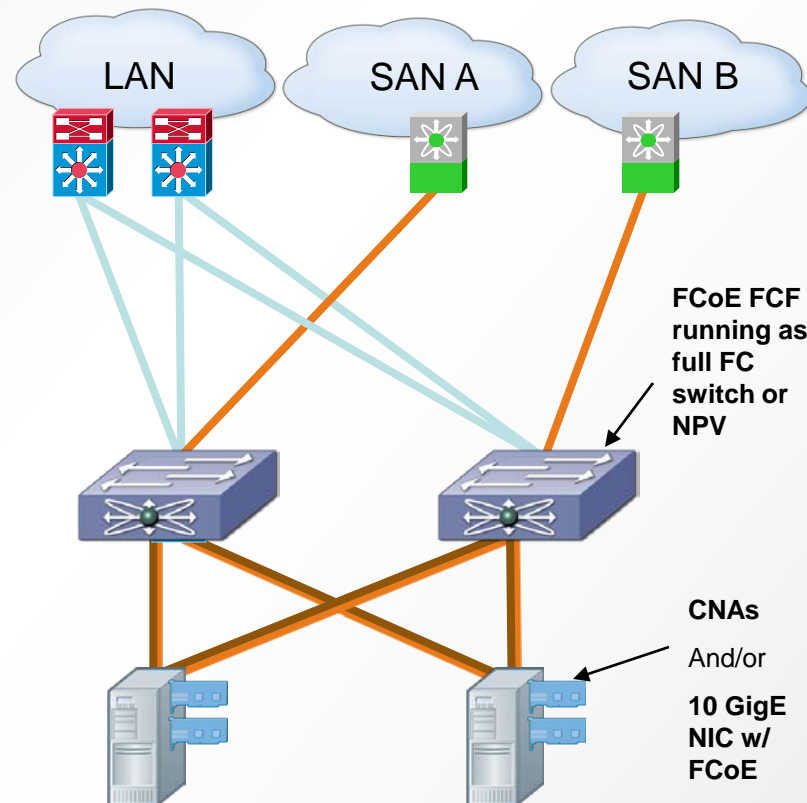- ❖ **Dual fabrics are still deployed for redundancy**

Ethernet Fabric   FC Fabric   Target

FC

DCB capable Switch acting as an FCF

Unified Wire

CNA

ENode
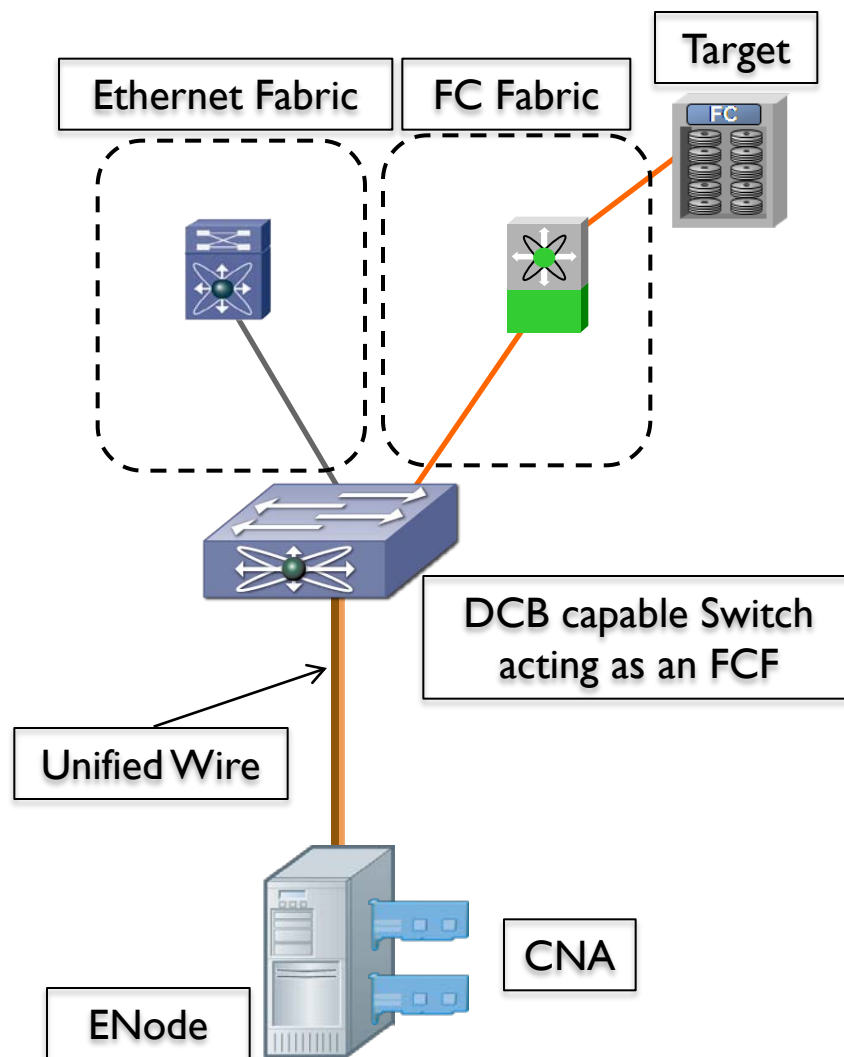
FCOE NETWORK DEPLOYMENTS

Basic Network Study

Enhanced Ethernet and FCoE   **Ethernet**   **FC**

10 GigE Ethernet

LAN   SAN A   SAN B

10 GigE Switch

FC SAN switch running as full FC switch or NPV

Ethernet NICs

HBAs

I/O Consolidation FCoE

LAN   SAN A   SAN B

FCoE FCF running as full FC switch or NPV

CNAs

And/or

10 GigE NIC w/ FCoE
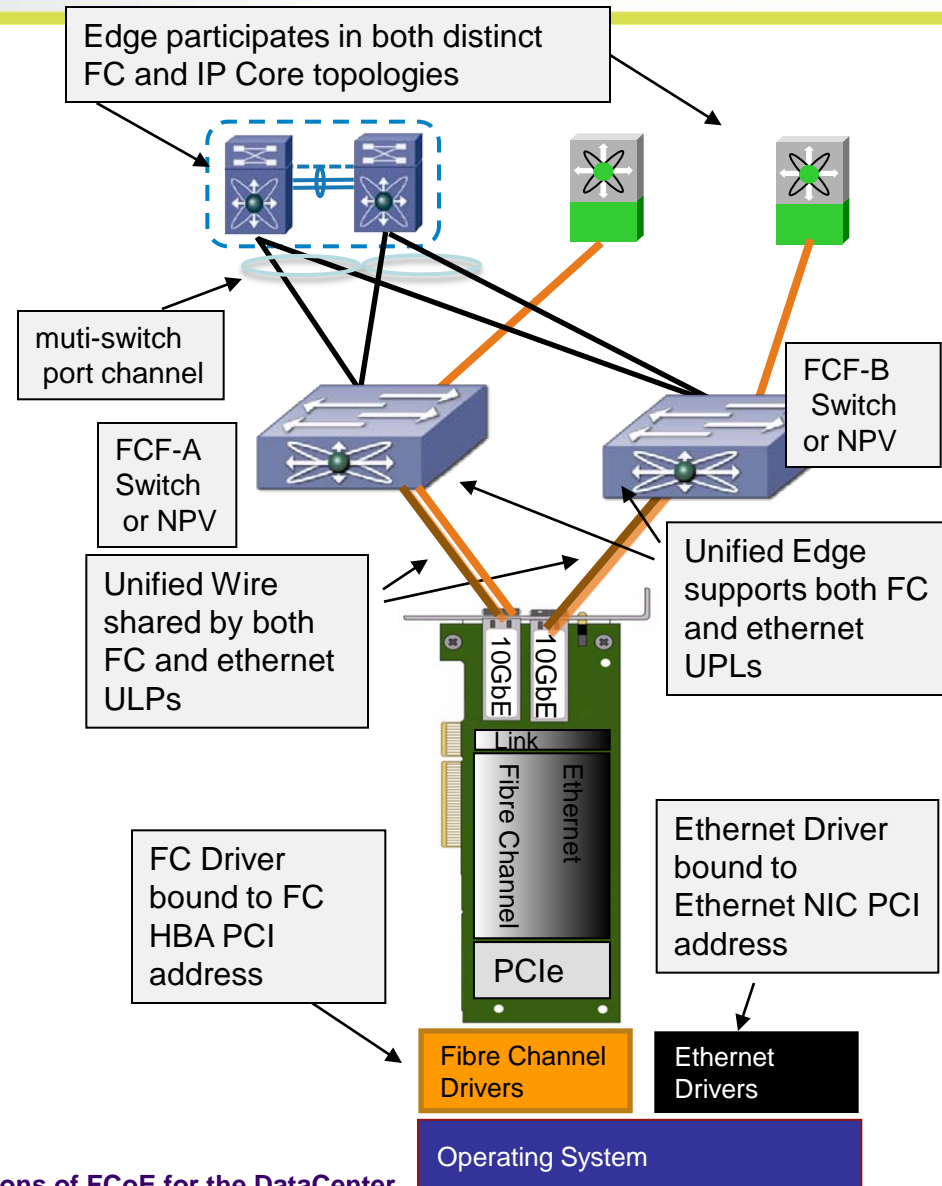
Deployments with FCoE

- Converged Network Adapter (CNA) presents two PCI address to the Operating System (OS)

- OS loads two unique sets of drivers and manages two unique application topologies

- Server participates in both topologies separately

- Two stacks and thus two different views of the same 'unified wire'

  - SAN Multi-Pathing provides failover between two fabrics (SAN 'A' and SAN 'B')

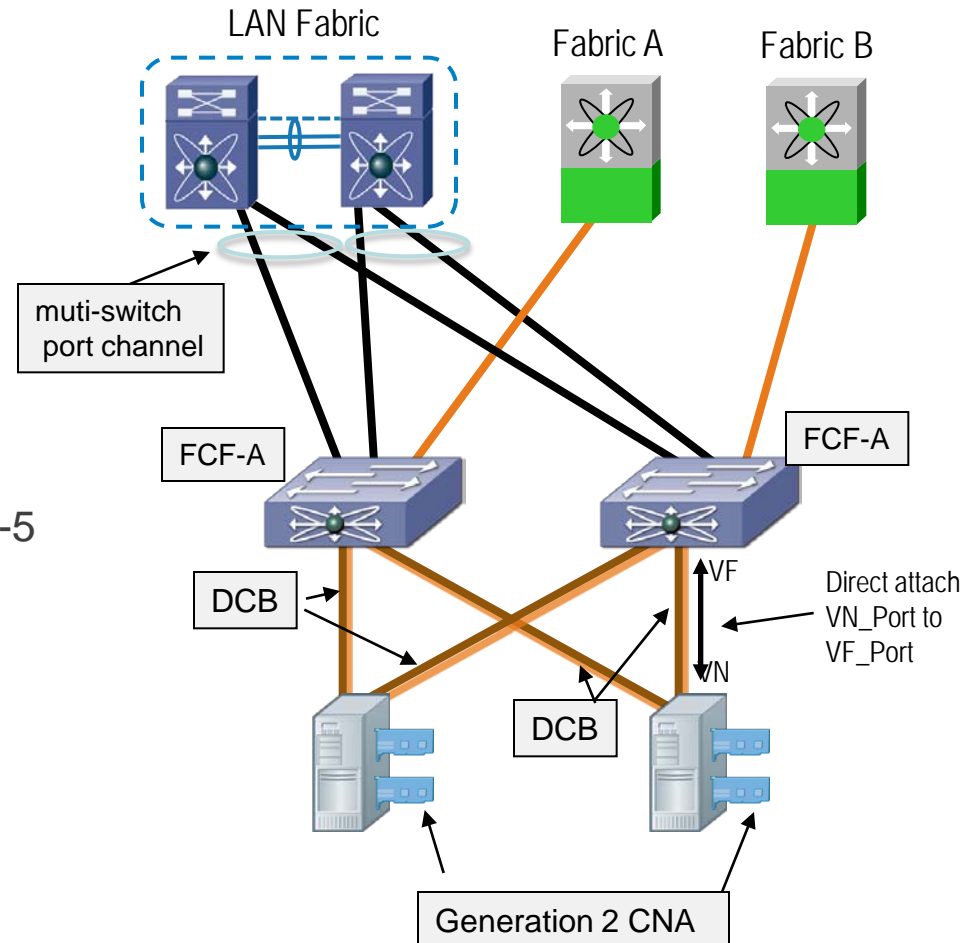 - NIC Teaming provides failover within the same fabric (VLAN)

Edge participates in both distinct FC and IP Core topologies

muti-switch port channel

FCF-B Switch or NPV

FCF-A Switch or NPV

Unified Edge supports both FC and ethernet UPLs

Unified Wire shared by both FC and ethernet ULPs

10GbE

10GbE

Link

Fibre Channel

Ethernet

PCIe

FC Driver bound to FC HBA PCI address

Ethernet Driver bound to Ethernet NIC PCI address

Fibre Channel Drivers

Ethernet Drivers

Operating System

What is being deployed Single Hop Design
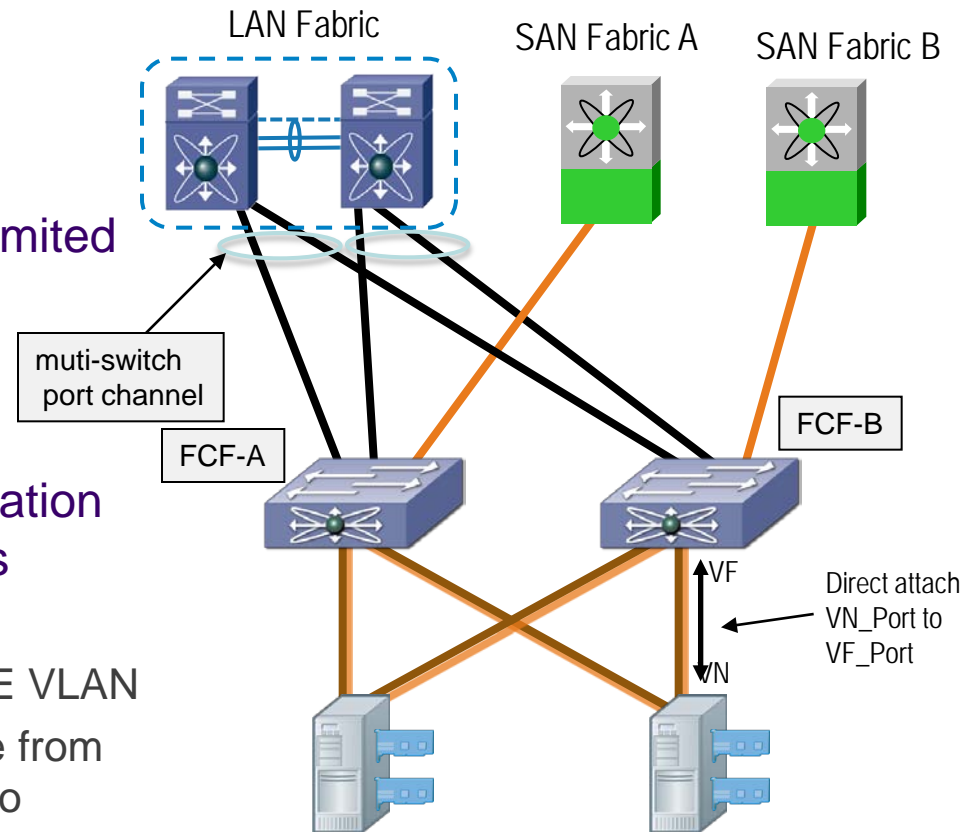
❖ **Generation 2 CNA**

   ➢ Utilizes Converged Enhanced
      Ethernet & Data Center Bridging
      Exchange protocol (DCB)

   ◆ Utilizes FCoE Initialization Protocol
      (FIP) as defined by the T.11 FC-BB-5
      specification

   ◆ Supports both direct and multi-hop
      attachment *(i.e. through a FIP
      Snooping Bridge or over VE-Ports)*

LAN Fabric

Fabric A   Fabric B

muti-switch
port channel

FCF-A   FCF-A

DCB

VF

Direct attach
VN_Port to
VF_Port

VN

DCB

Generation 2 CNA

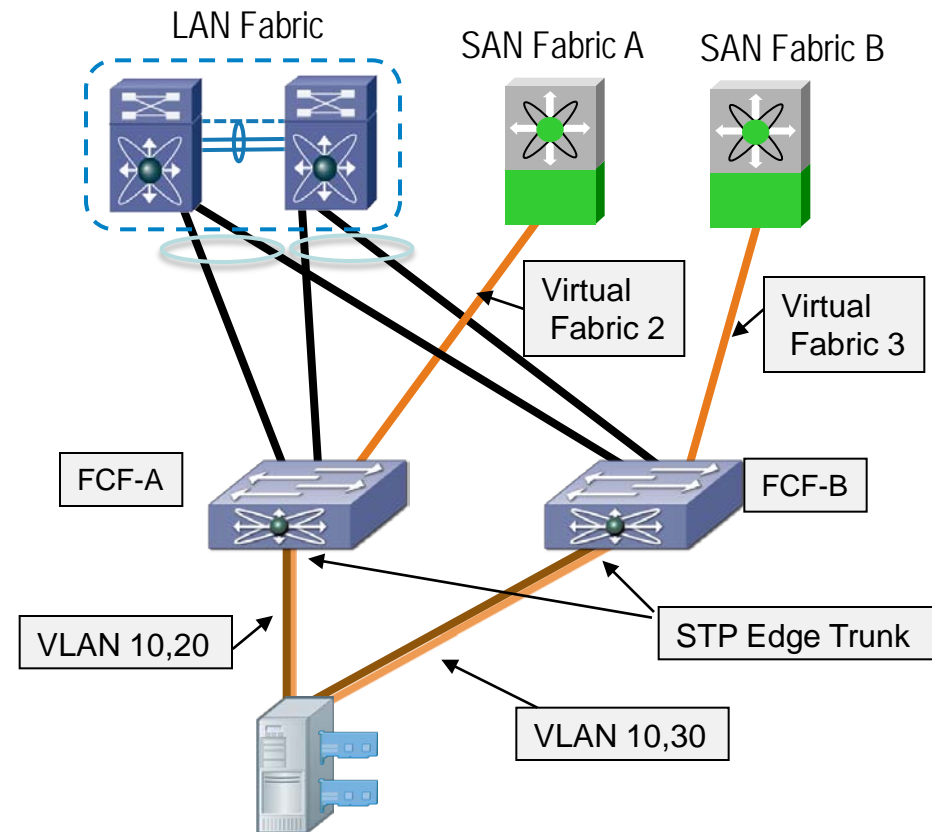# What is being deployed Initiator connection

- ❯ **Physical link is brought online** (today requires 10GE)

- ❯ **DCBX negotiation** – discovers DCB capable devices, including but not limited to lossless Ethernet.

- ❯ **FIP Process** – discovery and negotiation of FCoE devices and characteristics
  - ◆ FCoE VLAN Discovery
  - ◆ FCF Discovery on the specific FCoE VLAN
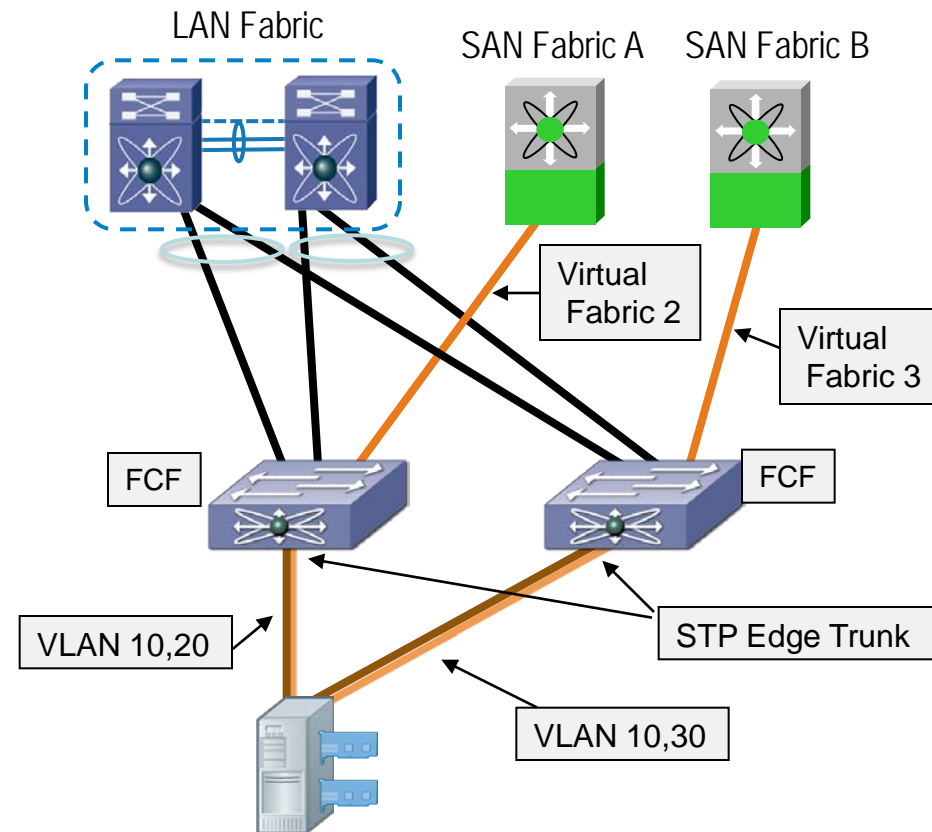 - ◆ Fabric Login - builds the logical wire from the end node to the FCF (VN_port to VF_port)

LAN Fabric   SAN Fabric A   SAN Fabric B

muti-switch port channel

FCF-A   FCF-B

VF
Direct attach VN_Port to VF_Port
VN

Direct Attach Topologies

Single Hop Design FCoE VLAN

- A VLAN can be dedicated for every Virtual Fabric in the SAN
- FIP discovers the FCoE VLAN and signals it to the hosts
- Trunking is not required on the host driver – all FCoE frames are tagged by the CNA
- FCoE VLANs can be pruned from Ethernet links that are not designate for FCoE
- Maintains isolated edge switches for SAN 'A' and 'B' and separate LAN switches for NIC 1 and NIC 2 (standard NIC teaming or Link Aggregation)

LAN Fabric

SAN Fabric A SAN Fabric B

Virtual Fabric 2

Virtual Fabric 3

FCF-A

FCF-B

VLAN 10,20

STP Edge Trunk

VLAN 10,30

Single Hop Design FCoE VLAN Separation

- In order to maintain the integrity of FC forwarding over FCoE, FCoE VLANs should be treated differently than LAN VLANs
  - No flooding, MAC learning, broadcasts, etc.
- The FCoE VLAN should be configured as a native VLAN
  - FIP uses native VLAN
- Separate FCoE VLANs are used for FCoE in SAN-A and SAN-B
- Unified Wires can be configured as trunk ports and STP edge ports

LAN Fabric

SAN Fabric A   SAN Fabric B

Virtual Fabric 2

Virtual Fabric 3

FCF

FCF

VLAN 10,20

STP Edge Trunk

VLAN 10,30

Single Hop Design
Unified Wires and multi-switch port channels

- ❖ Optimal layer 2 LAN design often leverages muti-switch port channel

- ❖ A muti-switch port channel can provide network based load sharing and redundancy and some implementations can do this without introducing layer 2 loops in the topology

- ❖ Muti-switch port channel results in diverging LAN and SAN high availability topologies
  - FC maintains separate SAN 'A' and SAN 'B' topologies
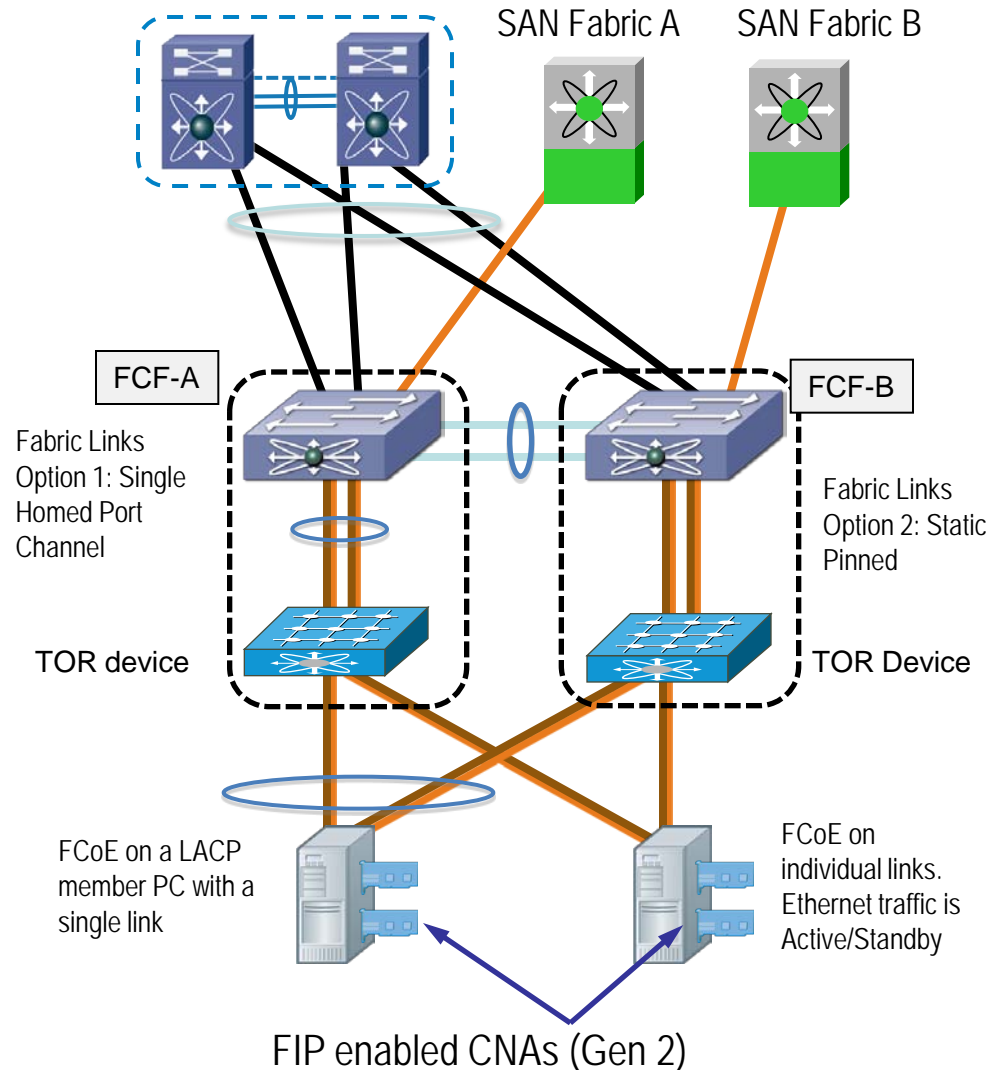 - LAN utilizes a single logical topology

LAN Fabric   SAN Fabric A   SAN Fabric B

FCF-A   FCF-B

muti-switch port channel
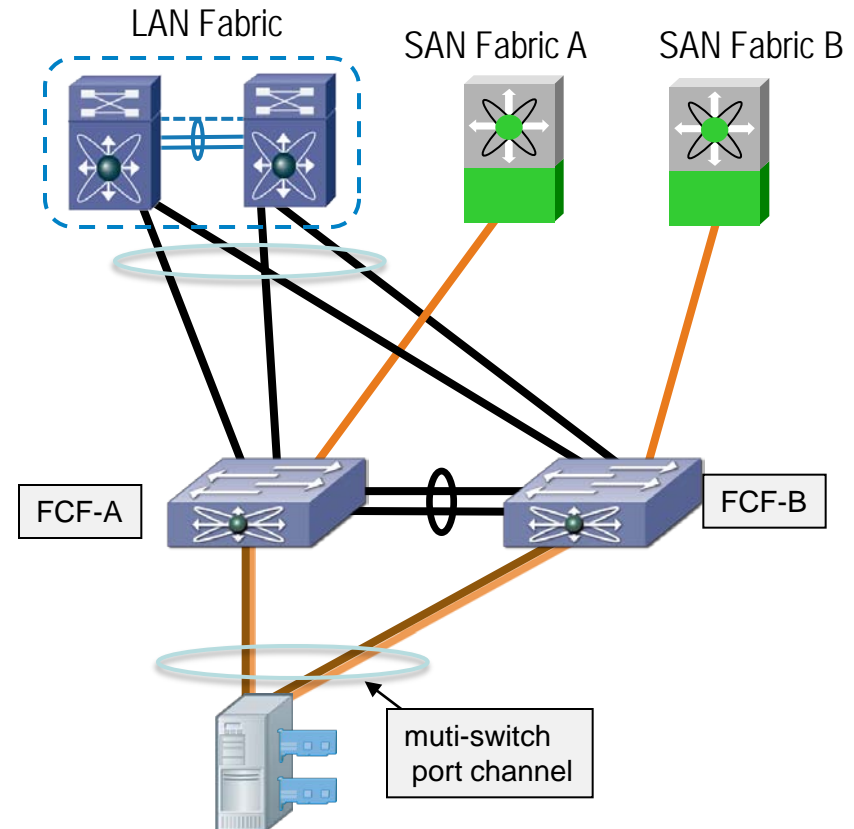
Direct Host Dual Attached Topology

Single Hop Design
Extending the FCoE Edge to Top of Rack

- ❖ Server Ethernet driver connected to TOR device with NIC Teaming

- ❖ FCoE can connect over multi member port with a single link from server to TOR

- ❖ TOR device *single* homed to upstream core FCF
  - ❖ Top of Rack fabric links can be connected to FCF with individual links or a muti-switch port channel
  - ❖ Can be oversubscribed

- ❖ Consistent with separate LAN Access and SAN Edge Topologies

SAN Fabric A   SAN Fabric B

FCF-A

Fabric Links Option 1: Single Homed Port Channel

FCF-B

Fabric Links Option 2: Static Pinned

TOR device

TOR Device

FCoE on a LACP member PC with a single link

FCoE on individual links. Ethernet traffic is Active/Standby

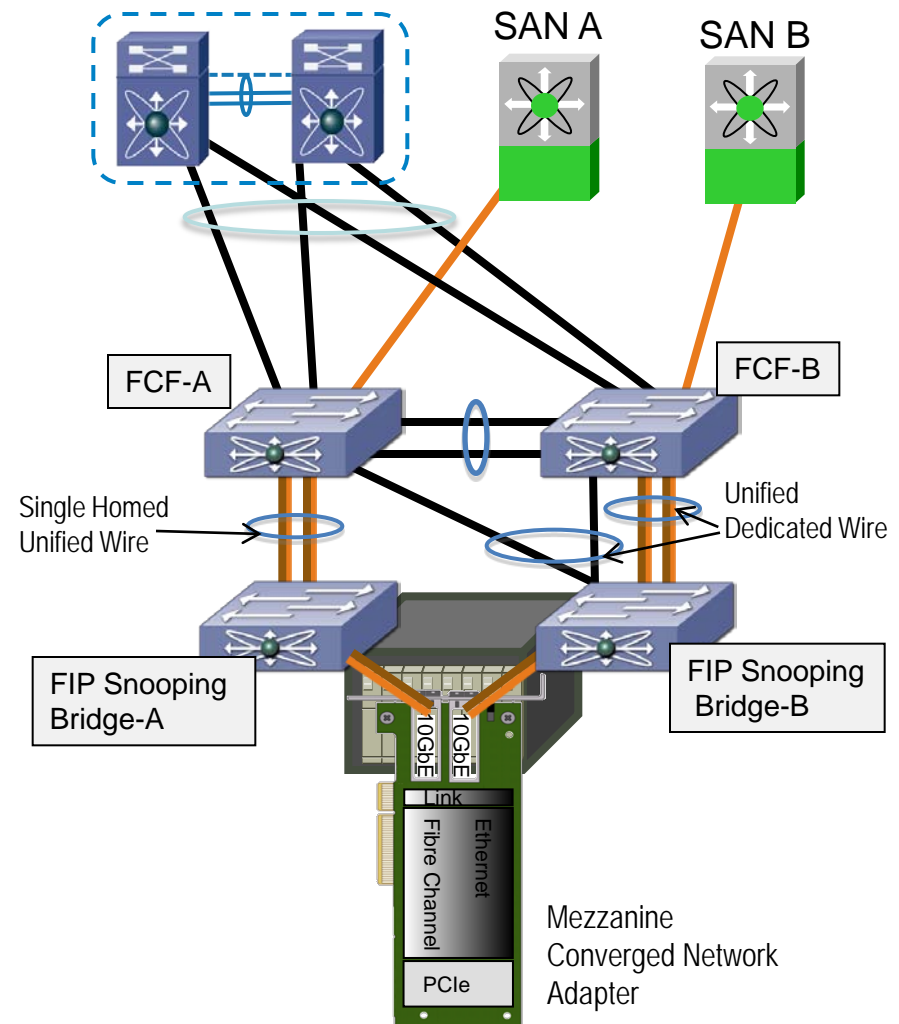FIP enabled CNAs (Gen 2)

Extending FCoE with FIP Snooping

- ### FIP snooping Bridge
  - DCB enabled

- ### Dual Topology Capable

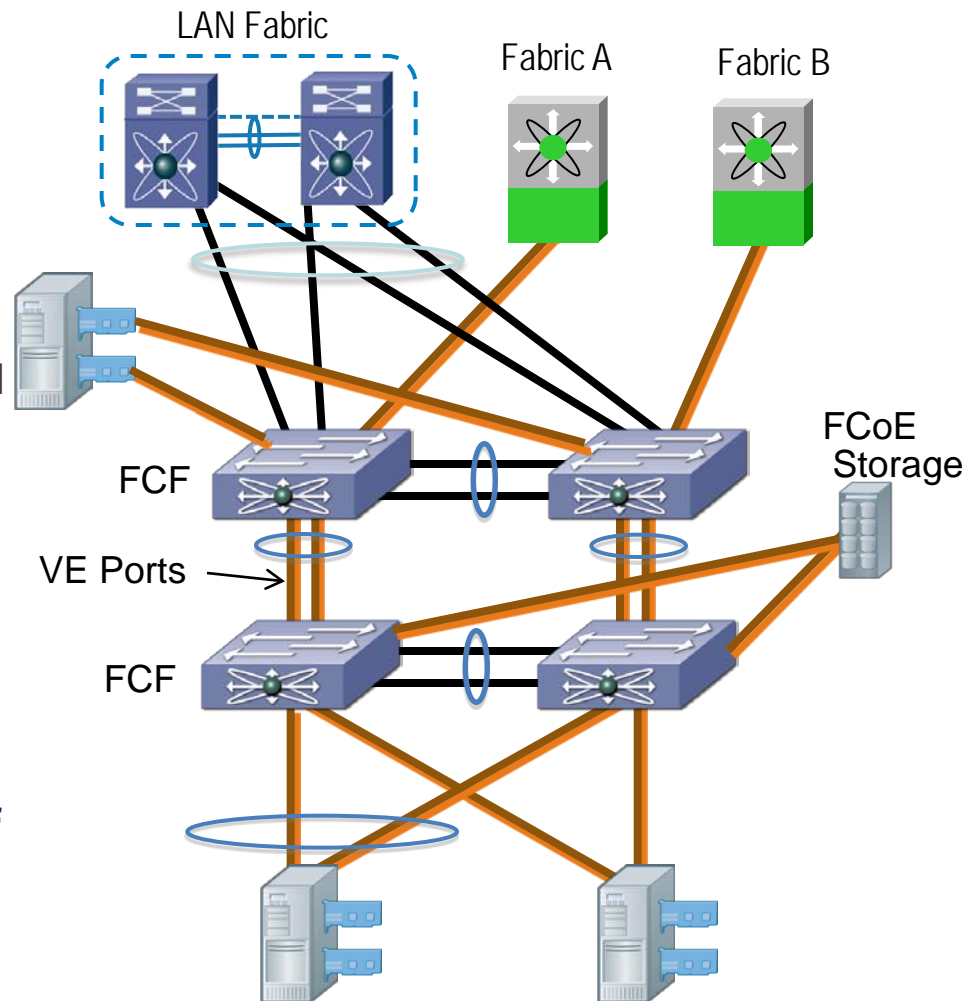- ### Still need to maintain the A-B SAN separation

  - Unified Dedicated Wires from FIP bridge to FCF
  - Single Unified Wire Port Channel from FIP Bridge to FCF

SAN A SAN B

FCF-A FCF-B

Single Homed Unified Wire

Unified Dedicated Wire

FIP Snooping Bridge-A

FIP Snooping Bridge-B

10GbE 10GbE

Link

Ethernet

Fibre Channel

PCIe

Mezzanine Converged Network Adapter

Multi - Hop Designs (VE)
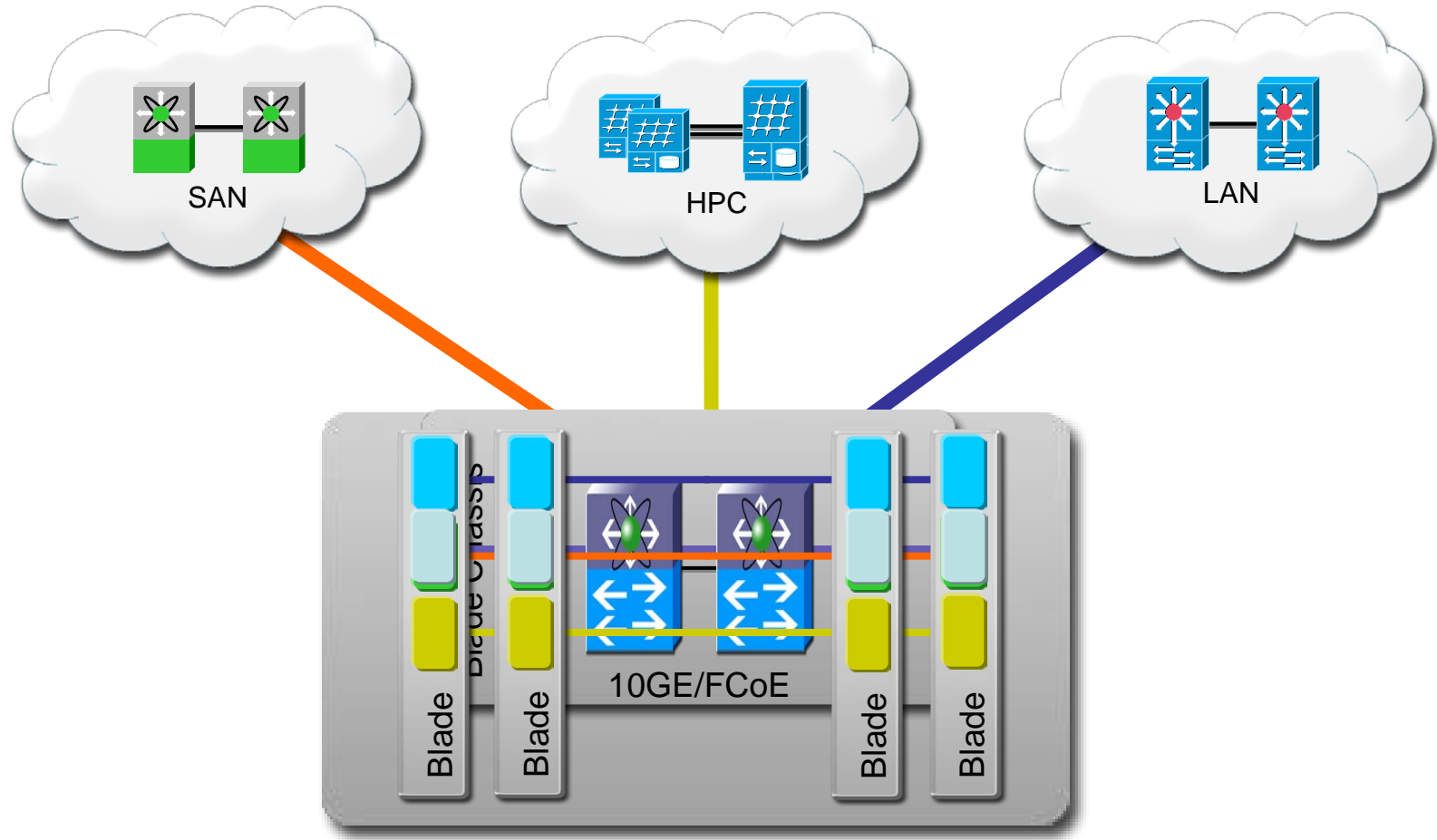## Considerations for FCoE Multi-hop

- ❖ **Virtual E_Ports (VE)**

- ❖ **Design considerations when extending FCoE beyond the Unified Edge?**
  - ◆ High Availability for both LAN and SAN
  - ◆ Oversubscription for SAN and LAN
  - ◆ Ethernet Layer 2 and STP design

- ❖ **Where does Unified Wire make sense over Unified Dedicated Wire?**

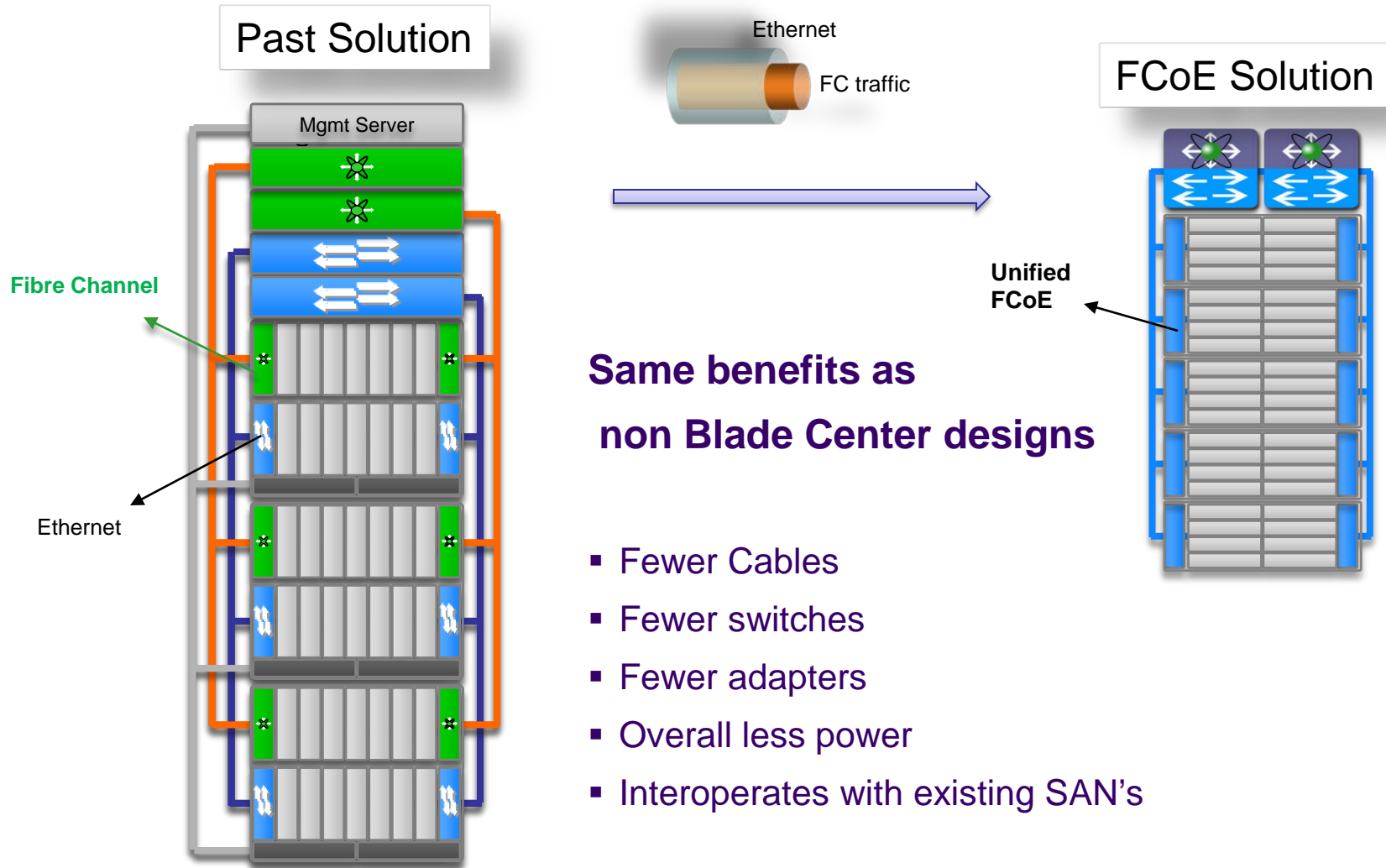- ❖ **Unified Wire provides for sharing of a single link for both FC and Ethernet traffic**

LAN Fabric

Fabric A   Fabric B

FCF

VE Ports

FCF

FCoE Storage

UNIFIED I/O IN COMPUTE BLADE CENTER

Fabric Consolidation

FCoE in the Compute Blade Center

Past Solution

Ethernet

FC traffic

FCoE Solution

Mgmt Server

Fibre Channel

Ethernet

Unified
FCoE

Same benefits as
non Blade Center designs

- Fewer Cables
- Fewer switches
- Fewer adapters
- Overall less power
- Interoperates with existing SAN's

Check Out the Hands-On Lab

Check out the Hands-On LAB at SNW:

Unified Storage Infrastructure
Unified Storage IP Solutions

Q&A / Feedback

Education
SNIA

◆ Please send any questions or comments on this presentation to SNIA: tracknetworking@snia.org

**Many thanks to the following individuals
for their contributions to this tutorial.**
- SNIA Education Committee

Simon Gordon

Joseph White