What You Make Possible

TOMORROW
starts here.

FCoE – Design, Implementation and Management Best Practices

BRKSAN-2047

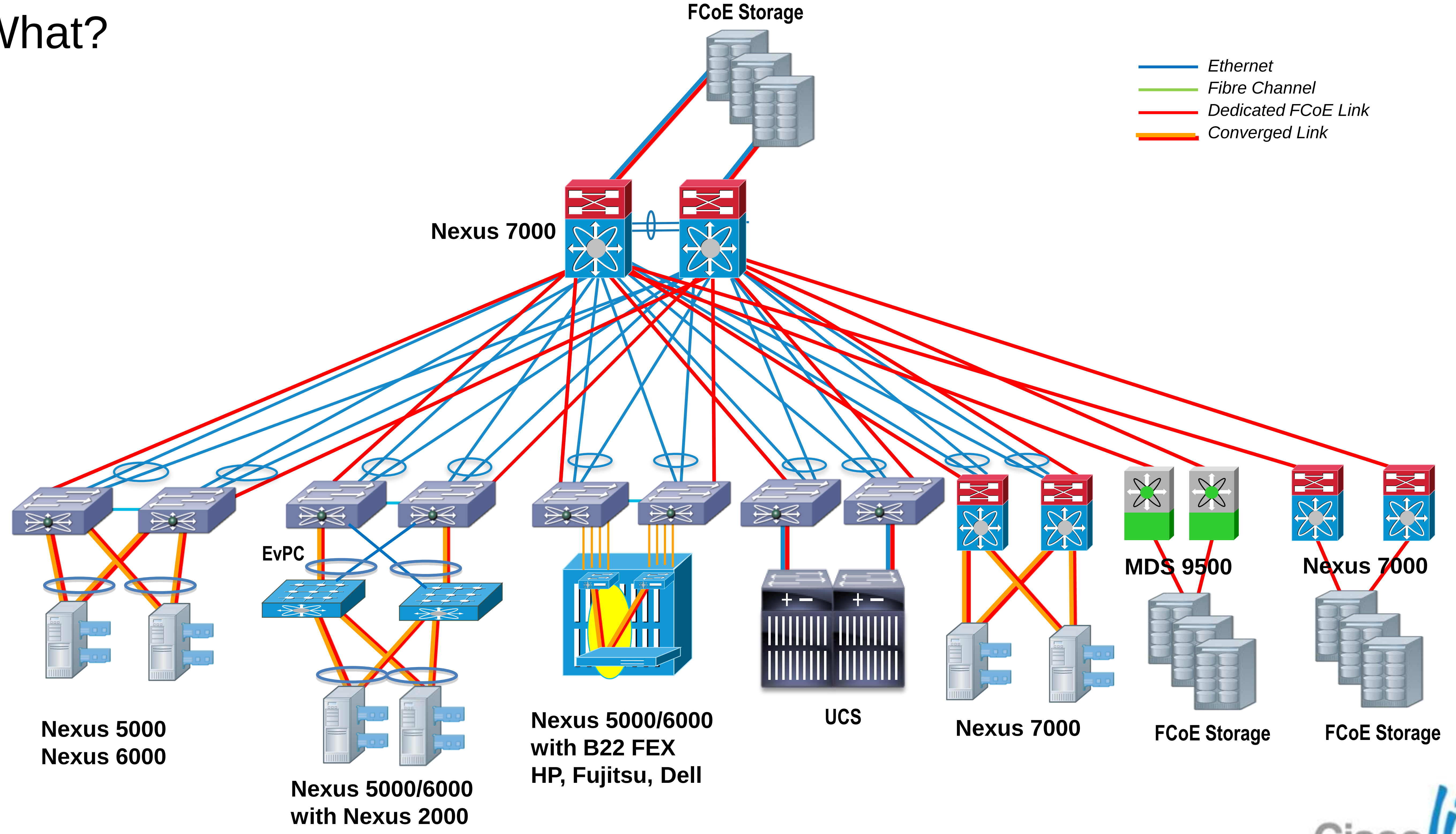
TOMORROW starts here.

Agenda

- Unified Fabric – What and Why
- FCoE Protocol Fundamentals
- Nexus FCoE Capabilities
- FCoE Network Requirements and Design Considerations
- DCB & QoS - Ethernet Enhancements
- Single Hop Design
- Multi-Hop Design
- Futures

Unified Fabric and FCoE

What?

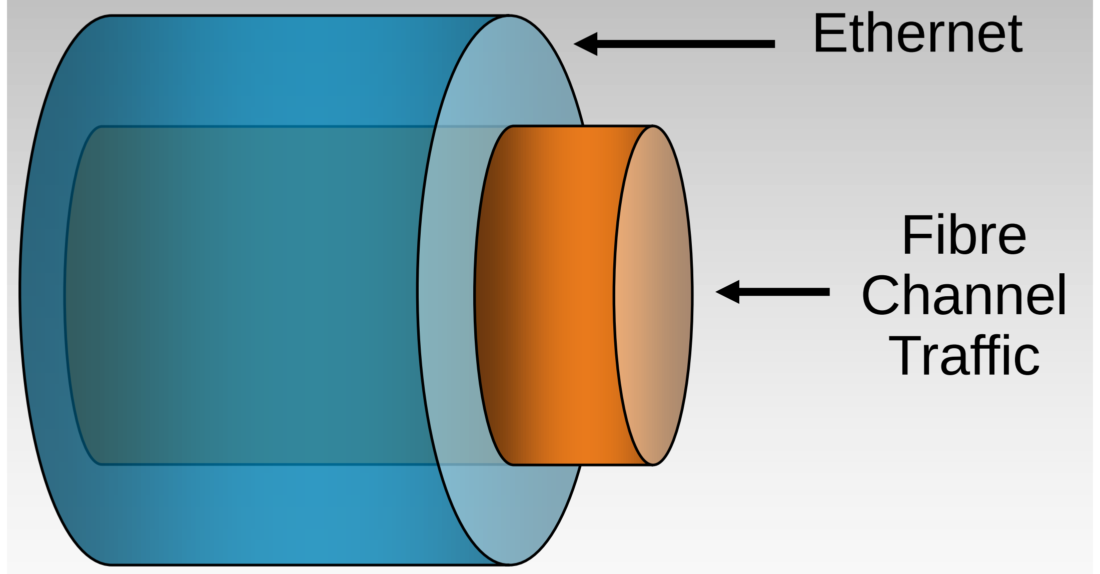


Unified Fabric & FCoE

Why?

FCoE

- Encapsulation of FC Frames over Ethernet
- Enables FC to run on a Lossless Ethernet Network



Benefits

- Infrastructure Consolidation
  - Support both FC and Ethernet switching in single fabric
- Fewer Cables
  - Both block I/O & Ethernet traffic co-exist on same cable
- Fewer adapters needed
- Overall less power
- Interoperates with existing SAN
  - Consistent SAN Management and Operations
- No Gateway

Unified Fabric

Why?

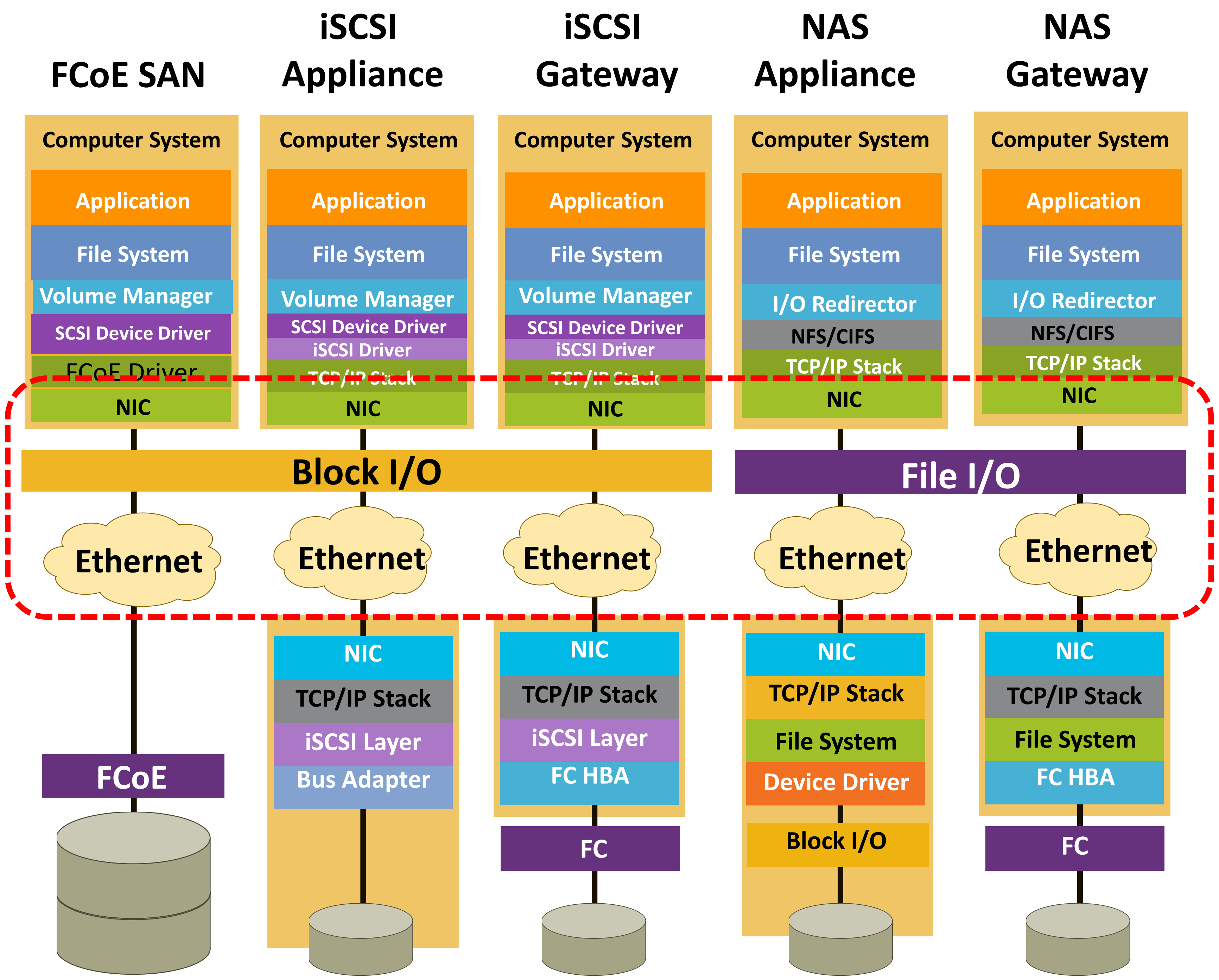
Ethernet Model has Proven Benefits

Ethernet Economic Model	**FC Economic Model**
Embedded on Motherboard	Always a stand alone Card
Integrated into O/S	Specialised Drivers
Many Suppliers	Few Suppliers
Mainstream Technology	Specialised Technology
Widely Understood	Special Expertise
Interoperability by Design	Interoperability by Test

Unified Fabric

Why?

- Ability to re-provision any compute unit to leverage any access method to the data stored on the 'spindle'
- Serialised Re-Use – (e.g. Boot from SAN and Run from NAS)
- Virtualisation requires that the Storage Fabric needs to exist everywhere the IP fabric does

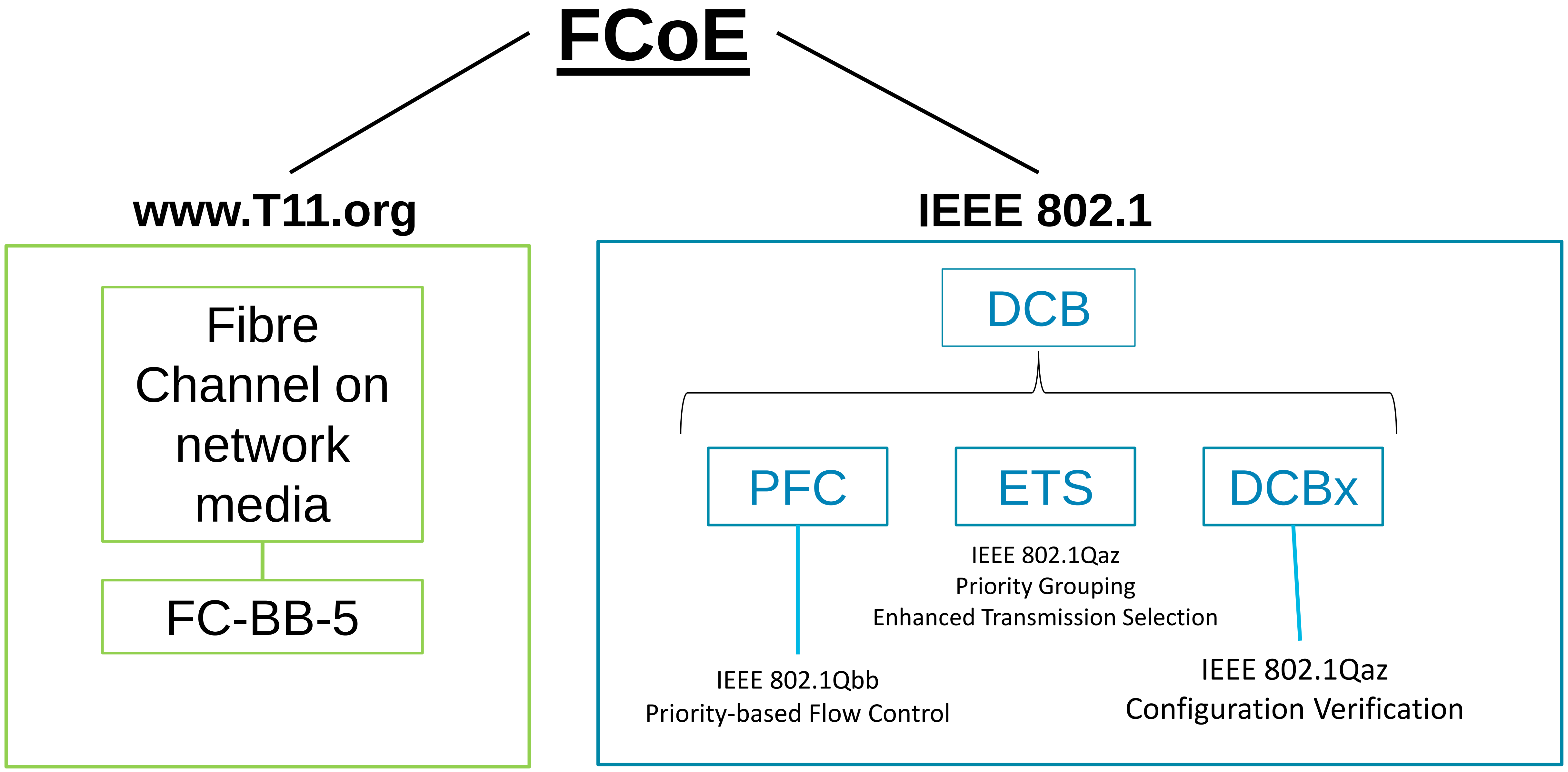


Agenda

- Unified Fabric – What and When
- FCoE Protocol Fundamentals
- Nexus FCoE Capabilities
- FCoE Network Requirements and Design Considerations
- DCB & QoS - Ethernet Enhancements
- Single Hop Design
- Multi-Hop Design
- Futures

FCoE Protocol Fundamentals

Standards for I/O Consolidation

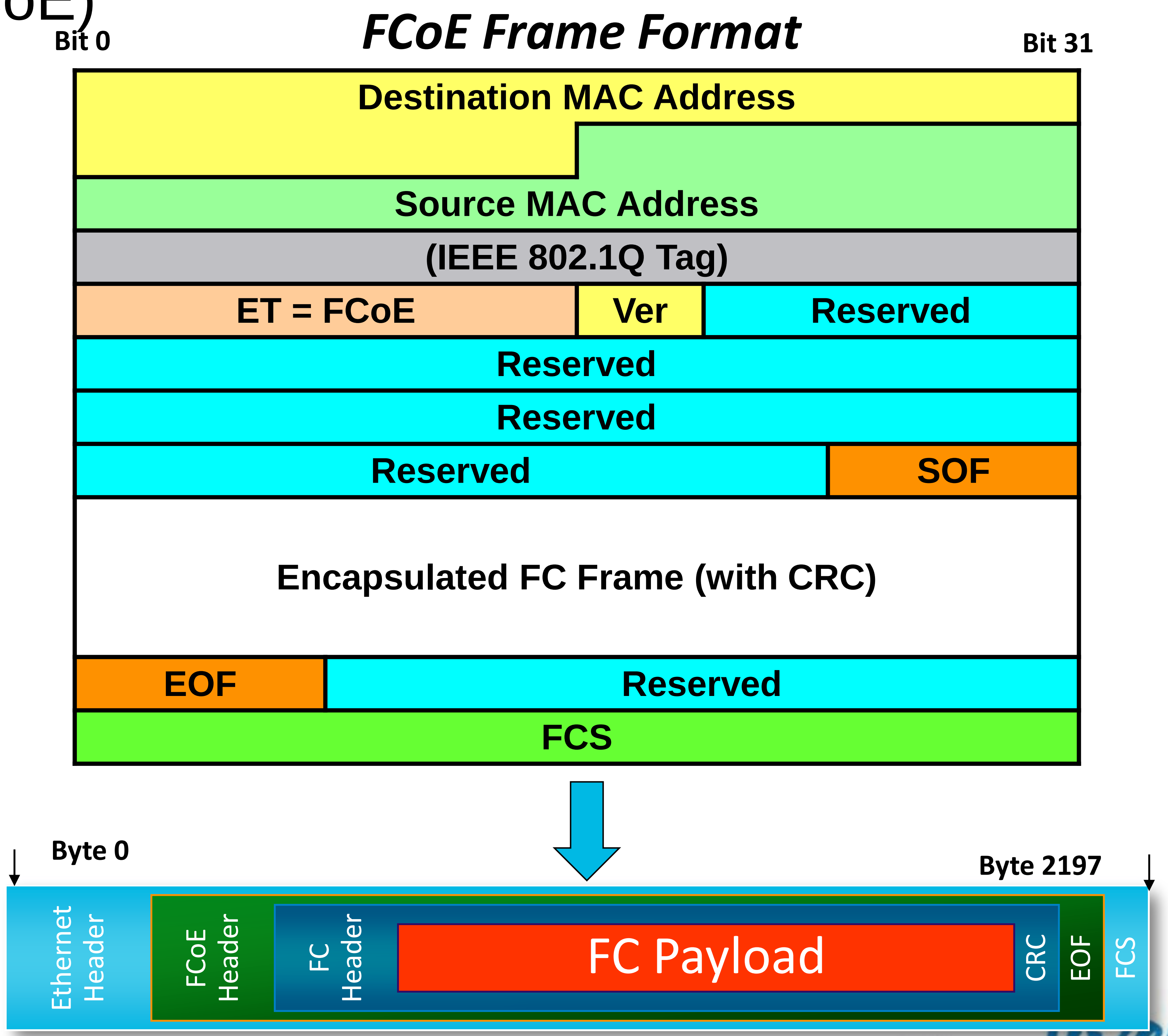


FCoE Protocol Fundamentals

incits — InterNational Committee for Information Technology Standards — Where IT all begins

Fibre Channel over Ethernet (FCoE)

- Fibre Channel over Ethernet provides a high capacity and lower cost transport option for block based storage
- Two protocols defined in the standard
  - FCoE – Data Plane Protocol
  - FIP – Control Plane Protocol
- FCoE is a standard - June 3rd 2009, the FC-BB-5 working group of T11 completed its work and unanimously approved a final standard for FCoE
- *FCoE 'is' Fibre Channel*



FCoE Protocol Fundamentals

Protocol Organisation – Data and Control Plane

FC-BB-5 defines two protocols required for an FCoE enabled Fabric

FCoE

- Data Plane
- It is used to carry most of the FC frames and all the SCSI traffic
- Uses Fabric Assigned MAC address (dynamic) : FPMA
- IEEE-assigned Ethertype for FCoE traffic is 0x8906

FIP (FCoE Initialisation Protocol)

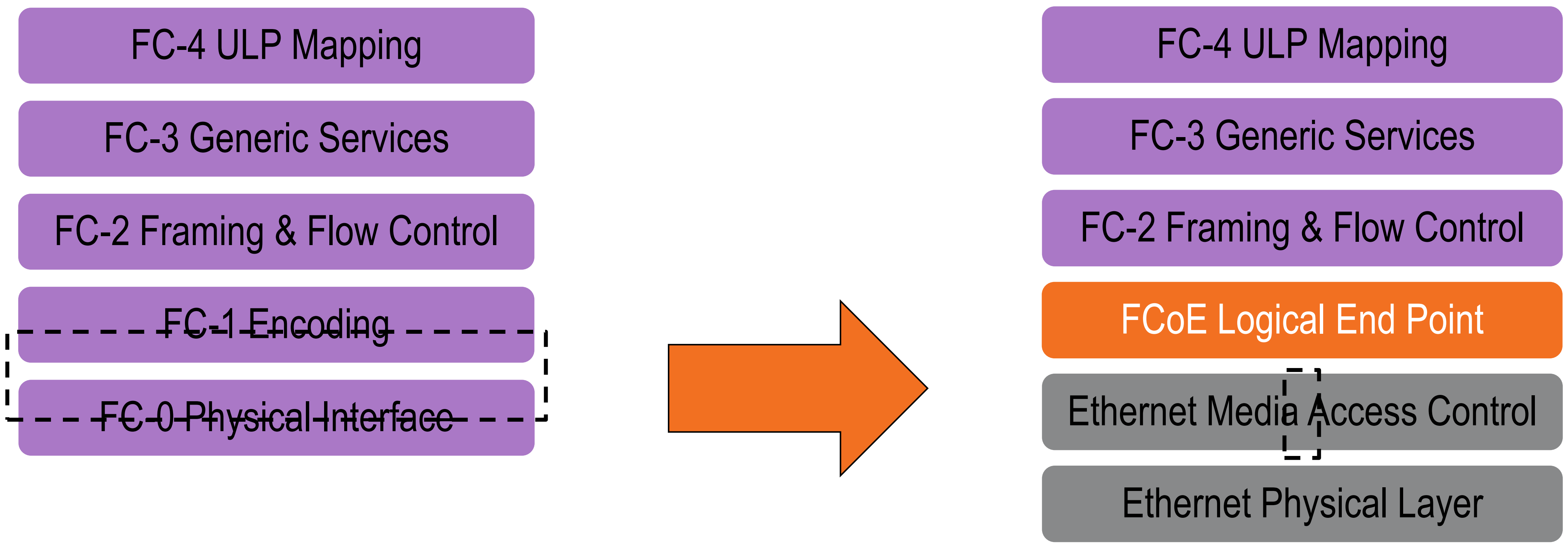
- It is the control plane protocol
- It is used to discover the FC entities connected to an Ethernet cloud
- It is also used to login to and logout from the FC fabric
- Uses unique BIA on CNA for MAC
- IEEE-assigned Ethertype for FCoE traffic is 0x8914

http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9670/white_paper_c11-560403.html

FCoE Protocol Fundamentals

It’s Fibre Channel Control Plane + FIP

- From a Fibre Channel standpoint it’s
  - –FC connectivity over a new type of cable called… Ethernet
- From an Ethernet standpoints it’s
  - –Yet another ULP (Upper Layer Protocol) to be transported

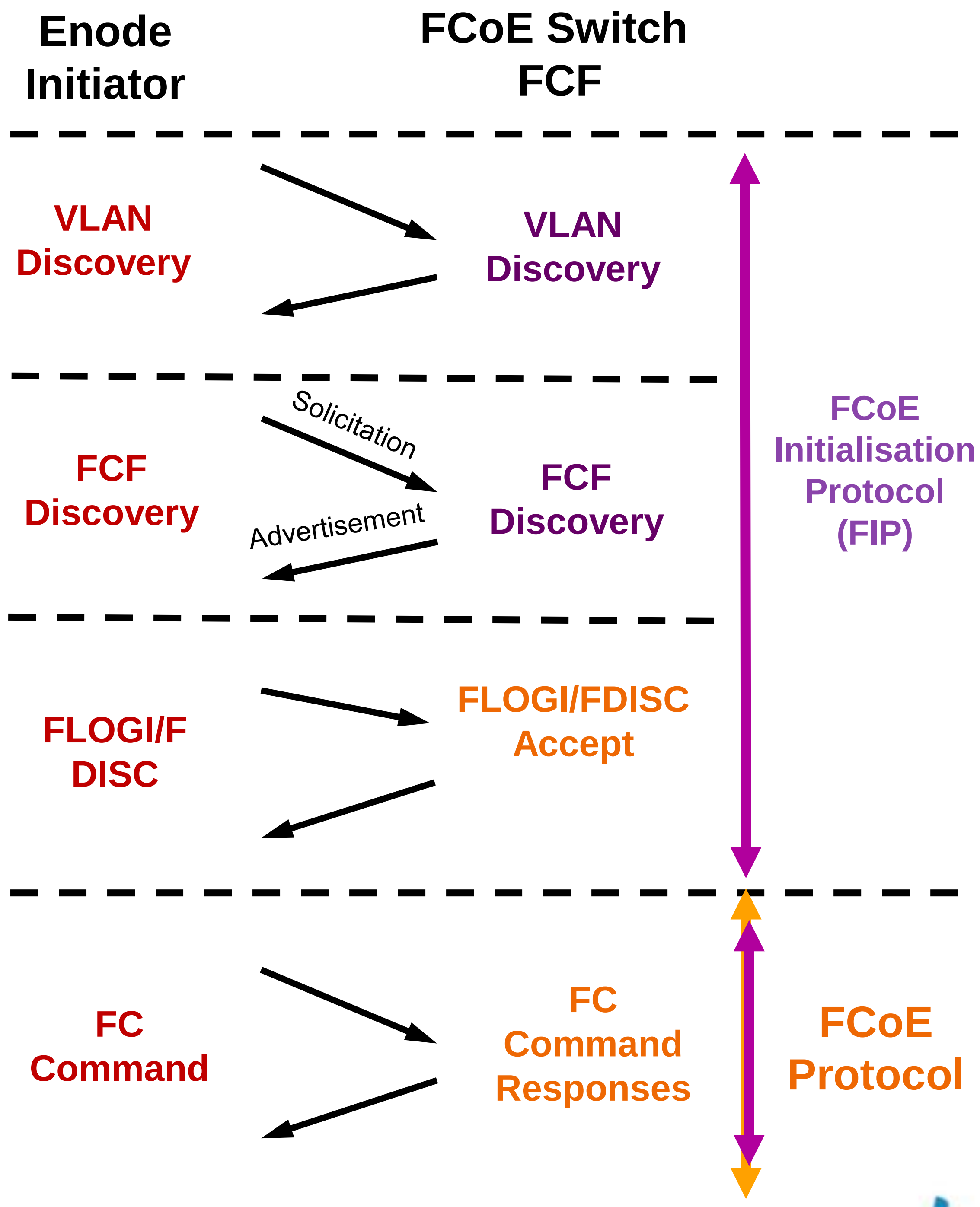


FCoE Protocol Fundamentals

FCoE Initialisation Protocol (FIP)

- **Neighbour Discovery and Configuration (VN – VF and VE to VE)**
- **Step 1: FCoE VLAN Discovery**
  - FIP sends out a multicast to ALL_FCF_MAC address looking for the FCoE VLAN
  - FIP frames use the **native VLAN**
- **Step 2: FCF Discovery**
  - FIP sends out a multicast to the ALL_FCF_MAC address on the FCoE VLAN to find the FCFs answering for that FCoE VLAN
  - FCF's responds back with their MAC address
- **Step 3: Fabric Login**
  - FIP sends a FLOGI request to the FCF_MAC found in Step 2
  - Establishes a virtual link between host and FCF

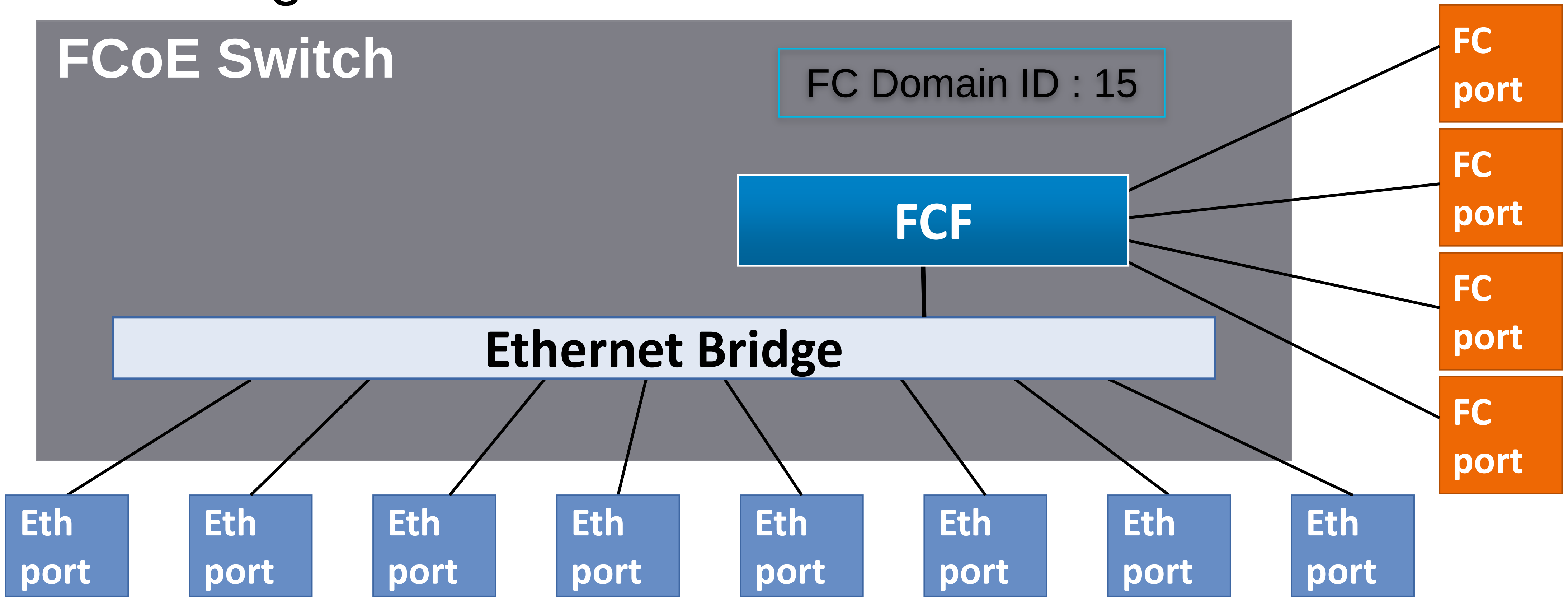
** FIP does not carry any Fibre Channel frames



FCoE Protocol Fundamentals

Fibre Channel Forwarder - FCF

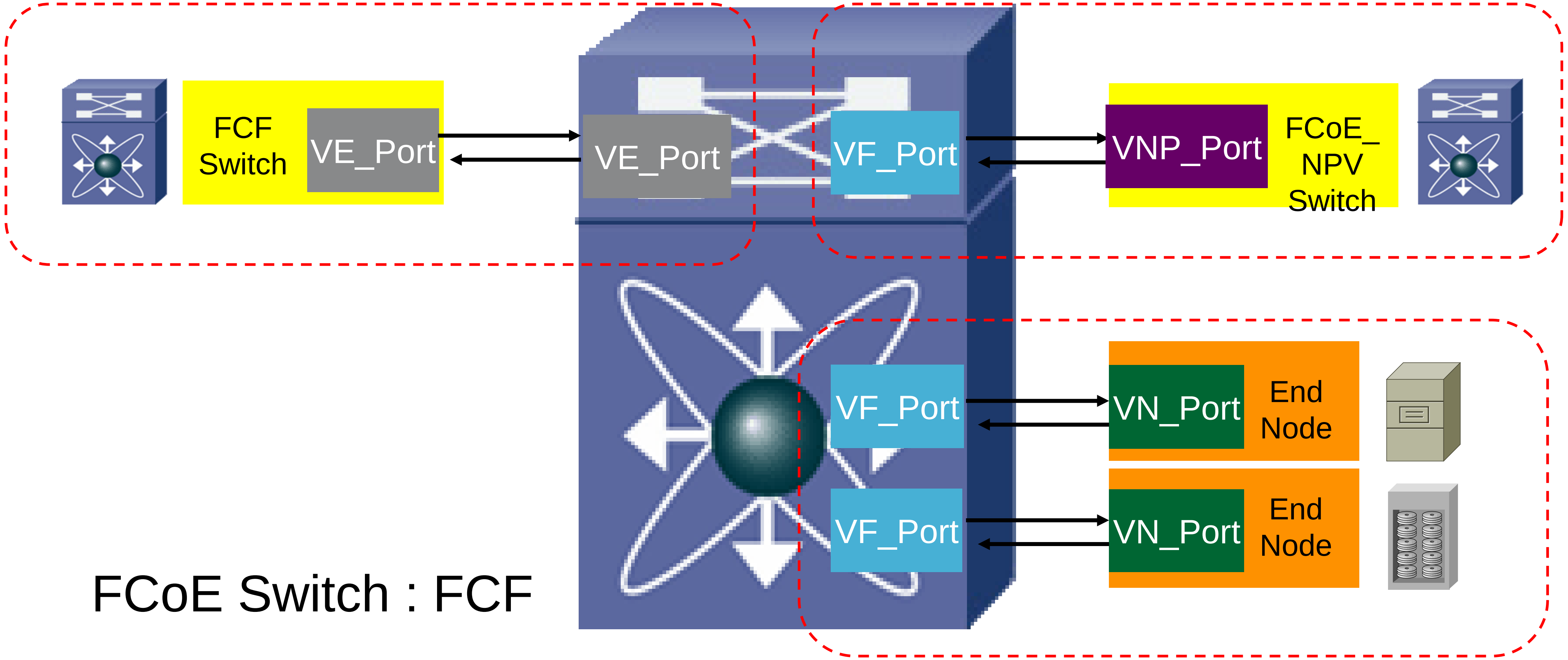
- FCF (Fibre Channel Forwarder) is the Fibre Channel switching element inside an FCoE switch
  - Fibre Channel logins (FLOGIs) happens at the FCF
  - Consumes a Domain ID
- FCoE encap/decap happens within the FCF
  - Forwarding based on FC information



FCoE Protocol Fundamentals

Explicit Roles still defined in the Fabric

- FCoE does not change the explicit port level relationships between devices (add a ‘V’ to the port type when it is an Ethernet wire)
  - Servers (VN_Ports) connect to Switches (VF_Ports)
  - Switches connect to Switches via Expansion Ports (VE_Ports)



FCoE Protocol Fundamentals

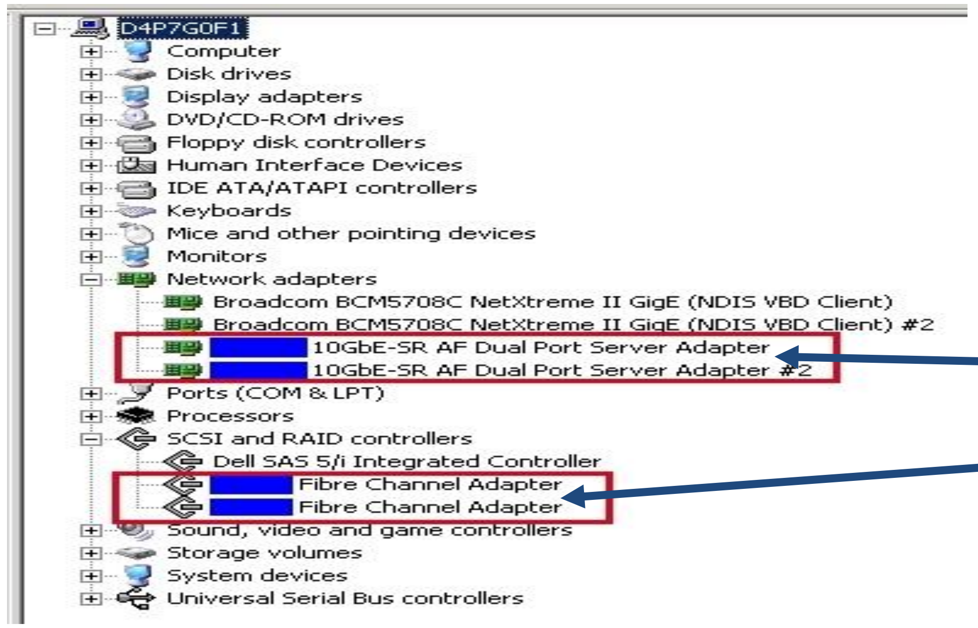
CNA: Converged Network Adapter

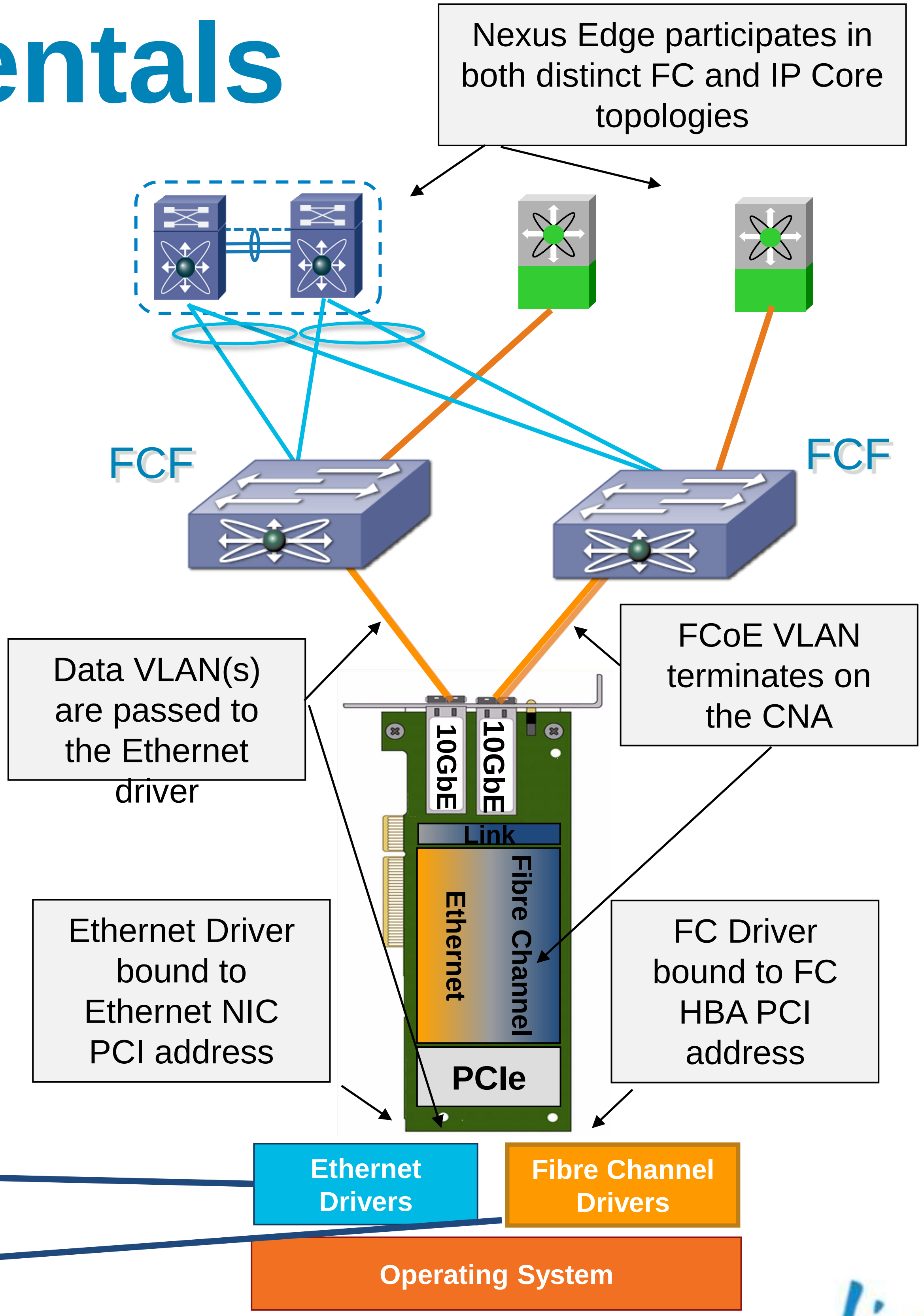
- Converged Network Adapter (CNA) presents two PCI address to the Operating System (OS)
- OS loads two unique sets of drivers and manages two unique application topologies
- Server participates in both topologies since it has two stacks and thus two views of the same 'unified wire'
  - SAN Multi-Pathing provides failover between two fabrics (SAN 'A' and SAN 'B')
  - NIC Teaming provides failover within the same fabric (VLAN)

- Operating System sees:

  Dual port 10 Gigabit Ethernet adapter

  Dual Port Fibre Channel HBAs

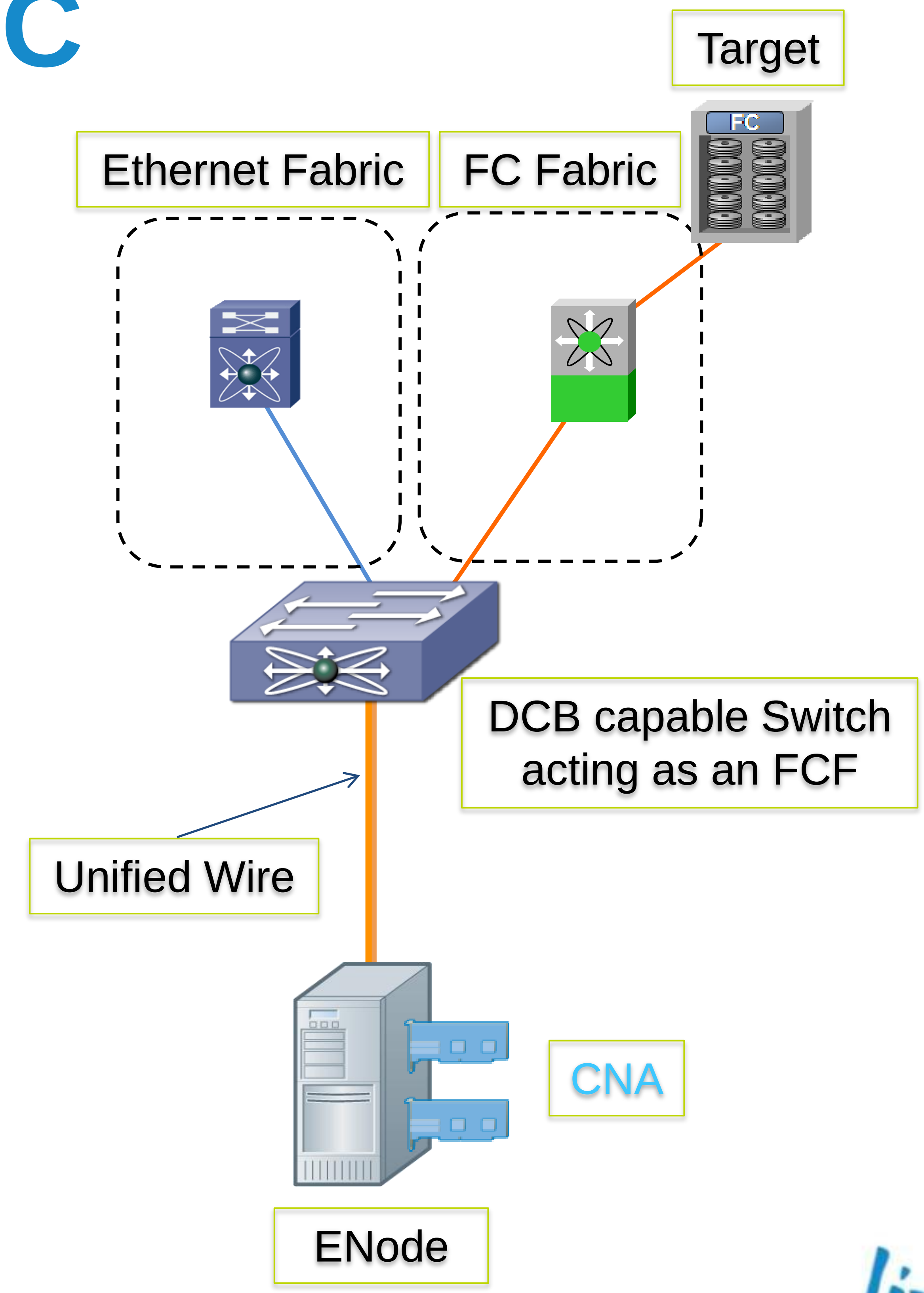




FCoE, Same Model as FC

Connecting to the Fabric

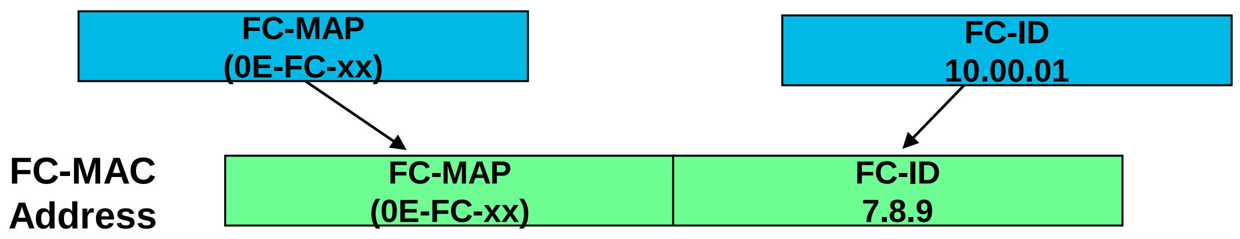
- Same host to target communication
  - Host has 2 CNA’s (one per fabric)
  - Target has multiple ports to connect to fabric
- Connect to a DCB capable switch
  - Port Type Negotiation (FC port type will be handled by FIP)
  - Speed Negotiation
  - DCBX Negotiation
- Access switch is a Fibre Channel Forwarder (FCF)
- Dual fabrics are still deployed for redundancy

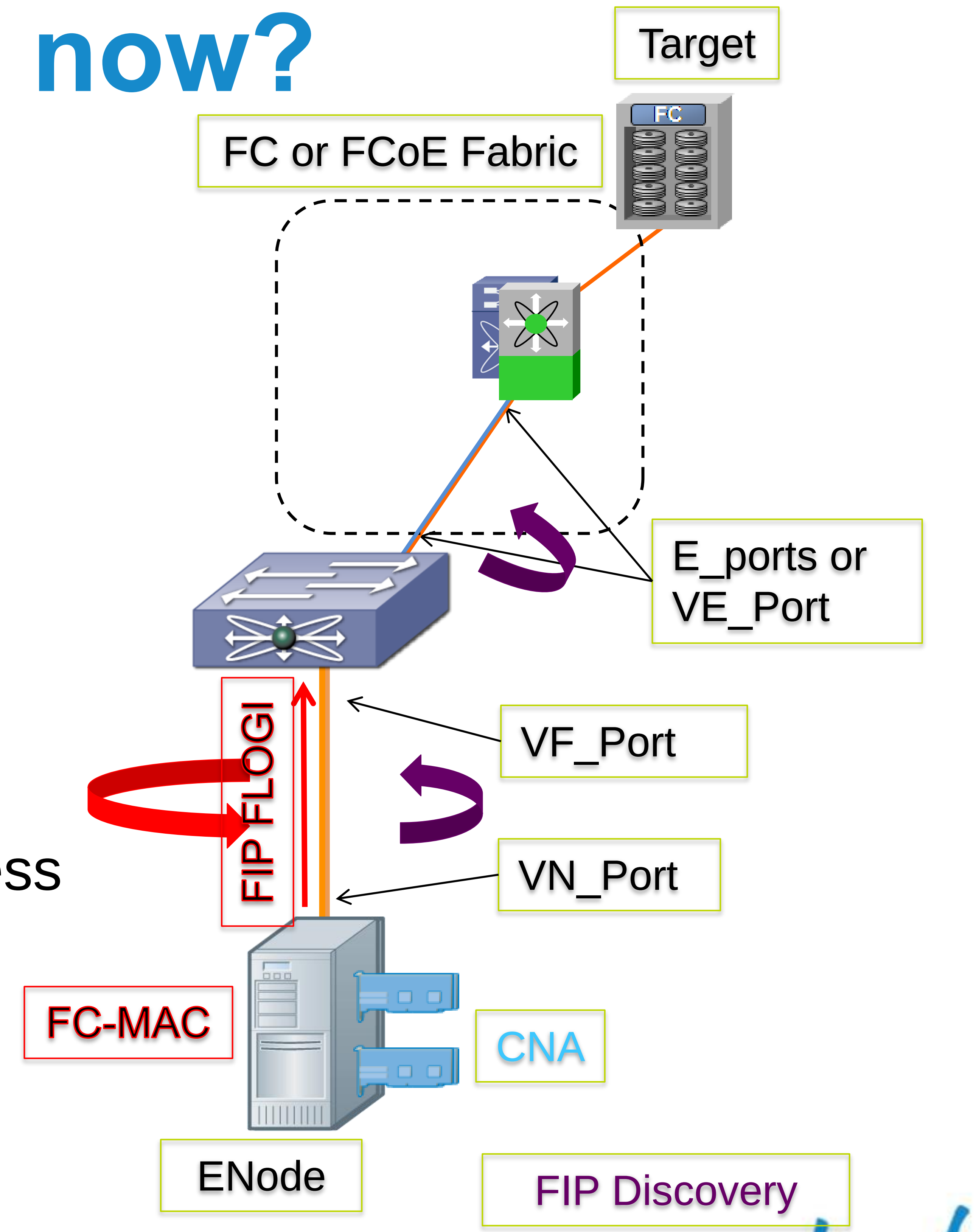


My port is up…can I talk now?

FIP and FCoE Login Process

- Step 1: FIP Discovery Process
  - FCoE VLAN Discovery
  - FCF Discovery
  - Verifies Lossless Ethernet is capable of FCoE transmission
- Step 2: FIP Login Process
  - Similar to existing Fibre Channel Login process - sends FLOGI to upstream FCF
  - FCF assigns the host a Enode MAC address to be used for FCoE forwarding (Fabric Provided MAC Address - FPMA)

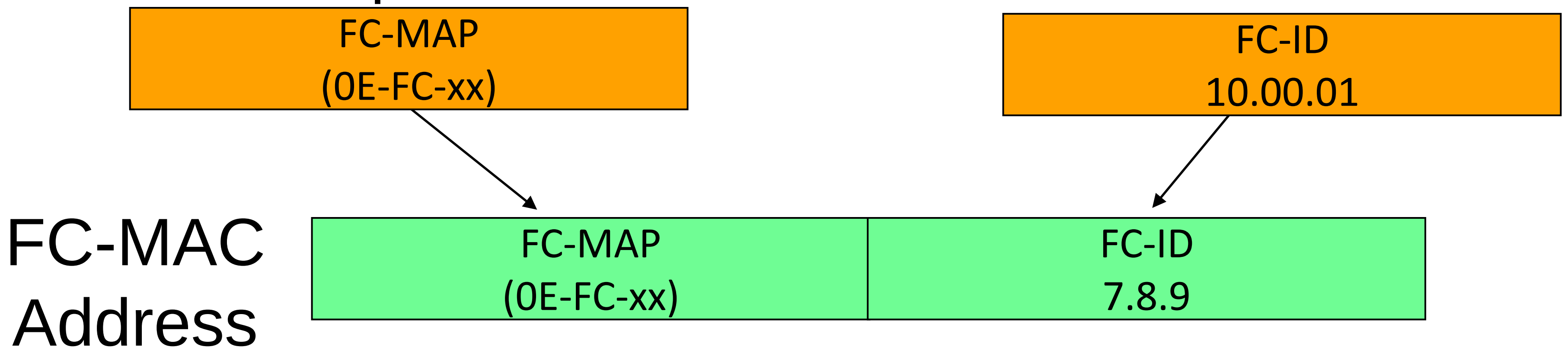




FCoE Protocol Fundamentals

Fibre Channel over Ethernet Addressing Scheme

- Enode FCoE MAC assigned for each FCID
- Enode FCoE MAC composed of a FC-MAP and FCID
  - FC-MAP is the upper 24 bits of the Enode's FCoE MAC
  - FCID is the lower 24 bits of the Enode's MAC
- FCoE forwarding decisions still made based on FSPF and the FCID within the Enode MAC
- For different physical networks the FC-MAP is used as a fabric identifier
  - FIP snooping will use this as a mechanism in realising the ACLs put in place to prevent data corruption



My port is up…can I talk now?

FIP and FCoE Login Process

- The FCoE VLAN is manually configured on the Nexus 5K

```
tme-n5k-2# conf t
Enter configuration commands, one per line.  End with CNTL/Z.
tme-n5k-2(config)# vlan 2
tme-n5k-2(config-vlan)# fcoe vsan 2
tme-n5k-2(config-vlan)# show vlan fcoe
VLAN       VSAN       Status
--------   --------   --------
2          2          Operational
```

- The FCF-MAC address is configured on the Nexus 5K by default once `feature fcoe` has been configured
  - This is the MAC address returned in step 2 of the FIP exchange
  - This MAC is used by the host to login to the FCoE fabric

```
tme-n5k-2# show fcoe
Global FCF details
        FCF-MAC is 00:0d:ec:df:5f:80
        FC-MAP is 0e:fc:00
        FCF Priority is 128
        FKA Advertisement period for FCF is 8 seconds
tme-n5k-2#
```

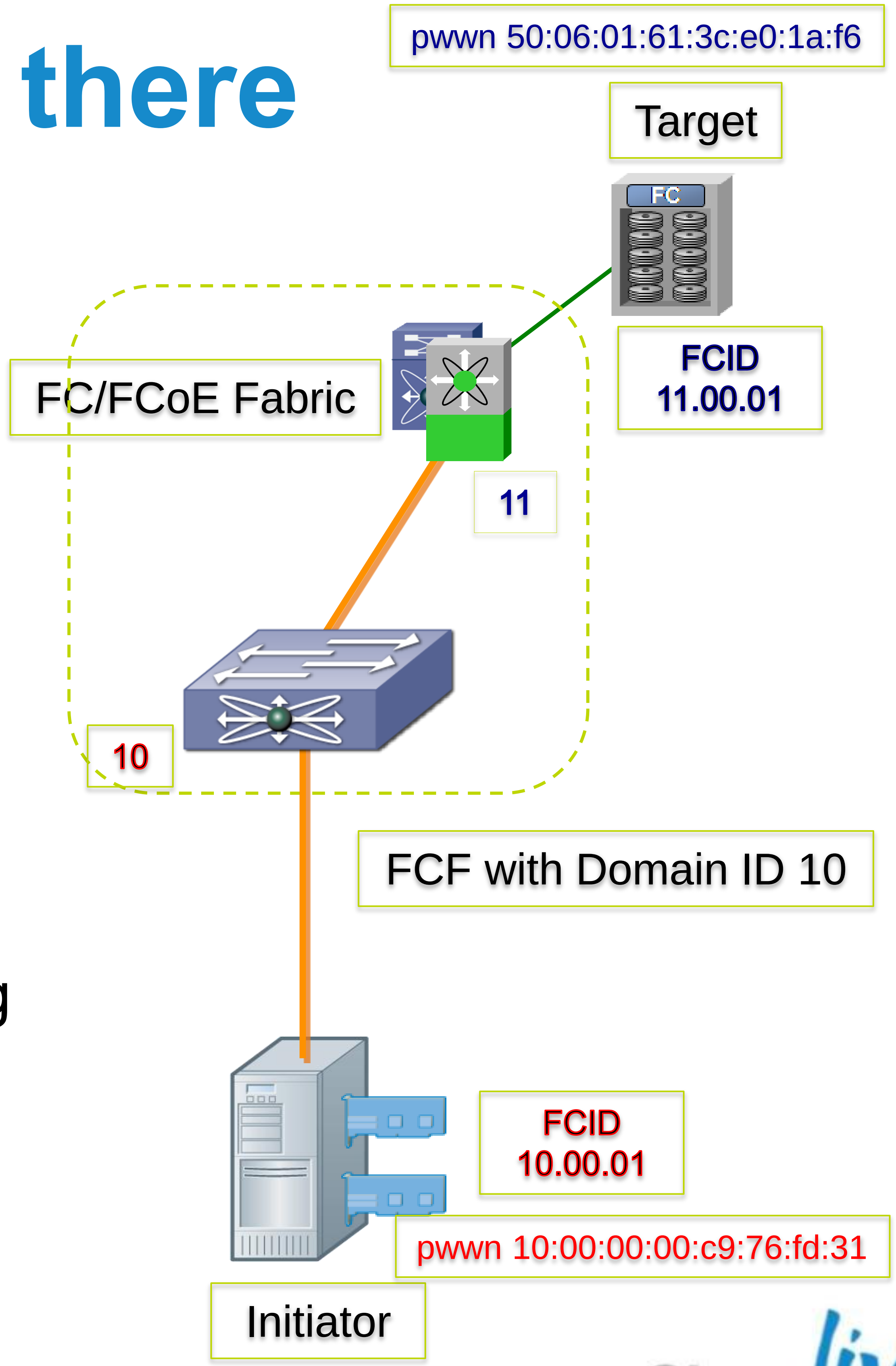
** FIP does not carry any Fibre Channel frames

Login complete…almost there

Fabric Zoning

- Zoning is a feature of the fabric and is independent of Ethernet transport
- Zoning can be configured on the Nexus 5000/7000 using the CLI or Fabric Manager
- If Nexus 5000 is in NPV mode, zoning will be configured on the upstream core switch and pushed to the Nexus 5000
- Devices acting as Fibre Channel Forwarders participate in the Fibre Channel security (Zoning) control
- DCB ‘only’ bridges do not participate in zoning and require additional security mechanisms (ACL applied along the forwarding path on a per FLOGI level of granularity)

```
fcid 0x10.00.01 [pwwn 10:00:00:00:c9:76:fd:31] [tnitiator]
fcid 0x11.00.01 [pwwn 50:06:01:61:3c:e0:1a:f6] [target]
```



Login complete

Flogi and FCoE Databases are populated

- Login process: show flogi database and show fcoe database show the logins and associated FCIDs, xWWNs and FCoE MAC addresses

```
tme-n5k-2# show flogi database
--------------------------------------------------------------------------------
INTERFACE        VSAN    FCID           PORT NAME               NODE NAME
--------------------------------------------------------------------------------
vfc1             2      0xb00000  21:00:00:c0:dd:11:29:1d 20:00:00:c0:dd:11:29:1d
vfc2             2      0xb00001  21:00:00:c0:dd:11:2c:61 20:00:00:c0:dd:11:2c:61
vfc14            2      0xb00004  21:00:00:c0:dd:12:13:8f 20:00:00:c0:dd:12:13:8f
vfc15            2      0xb00005  21:00:00:c0:dd:12:13:b3 20:00:00:c0:dd:12:13:b3
vfc16            2      0xb00006  21:00:00:c0:dd:12:14:23 20:00:00:c0:dd:12:14:23
vfc25            2      0xb00008  50:0a:09:83:87:d9:6e:b7 50:0a:09:80:87:d9:6e:b7
vfc26            2      0xb00009  50:0a:09:87:87:d9:6e:b7 50:0a:09:80:87:d9:6e:b7
                                  [netapp_fcoe1]
vfc30            2      0xb00007  50:0a:09:85:87:d9:6e:b7 50:0a:09:80:87:d9:6e:b7

Total number of flogi = 8.

tme-n5k-2# show fcoe database

-------------------------------------------------------------------------------
INTERFACE        FCID           PORT NAME               MAC ADDRESS
-------------------------------------------------------------------------------
vfc1             0xb00000       21:00:00:c0:dd:11:29:1d 00:c0:dd:11:29:1d
vfc2             0xb00001       21:00:00:c0:dd:11:2c:61 00:c0:dd:11:2c:61
vfc14            0xb00004       21:00:00:c0:dd:12:13:8f 00:c0:dd:12:13:8f
vfc15            0xb00005       21:00:00:c0:dd:12:13:b3 00:c0:dd:12:13:b3
vfc16            0xb00006       21:00:00:c0:dd:12:14:23 00:c0:dd:12:14:23
vfc25            0xb00008       50:0a:09:83:87:d9:6e:b7 00:c0:dd:0a:b7:82
vfc26            0xb00009       50:0a:09:87:87:d9:6e:b7 00:c0:dd:11:41:21
vfc30            0xb00007       50:0a:09:85:87:d9:6e:b7 00:c0:dd:11:0d:89
tme-n5k-2#
```

For Your Reference

FCoE Protocol Fundamentals

Summary of Terminology

- **CE -** Classical Ethernet (non lossless)
- **DCB & DCBx** - Data Centre Bridging, Data Centre Bridging Exchange
- **FCF** - Fibre Channel Forwarder (Nexus 5000, Nexus 7000, MDS 9000)
- **FIP** – FCoE Initialisation Protocol
- **Enode**: a Fibre Channel end node that is able to transmit FCoE frames using one or more Enode MACs.
- **FIP snooping Bridge**
- **FCoE-NPV** - Fibre Channel over IP N_Port Virtualisation
- **Single hop FCoE** : running FCoE between the host and the first hop access level switch
- **Multi-hop FCoE** : the extension of FCoE beyond a single hop into the Aggregation and Core layers of the Data Centre Network
- **Zoning -** Security method used in Storage Area Networks
- **FPMA –** Fabric Provided Management Address

Agenda

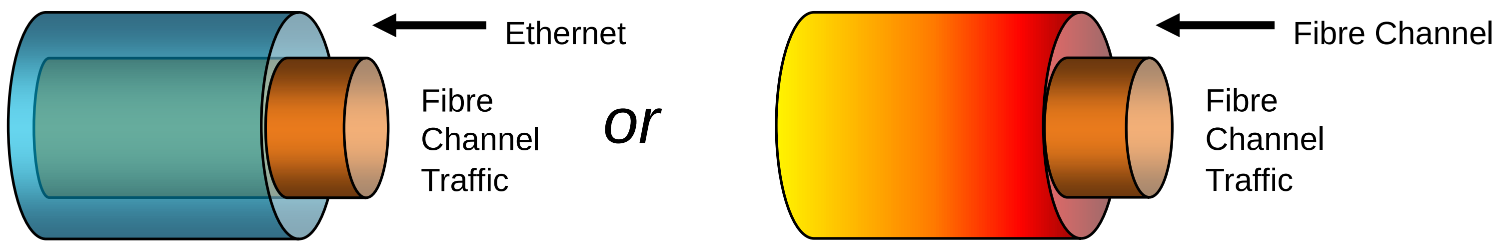
- Unified Fabric – What and Why
- FCoE Protocol Fundamentals
- Nexus FCoE Capabilities
- FCoE Network Requirements and Design Considerations
- DCB & QoS - Ethernet Enhancements
- Single Hop Design
- Multi-Hop Design
- Futures

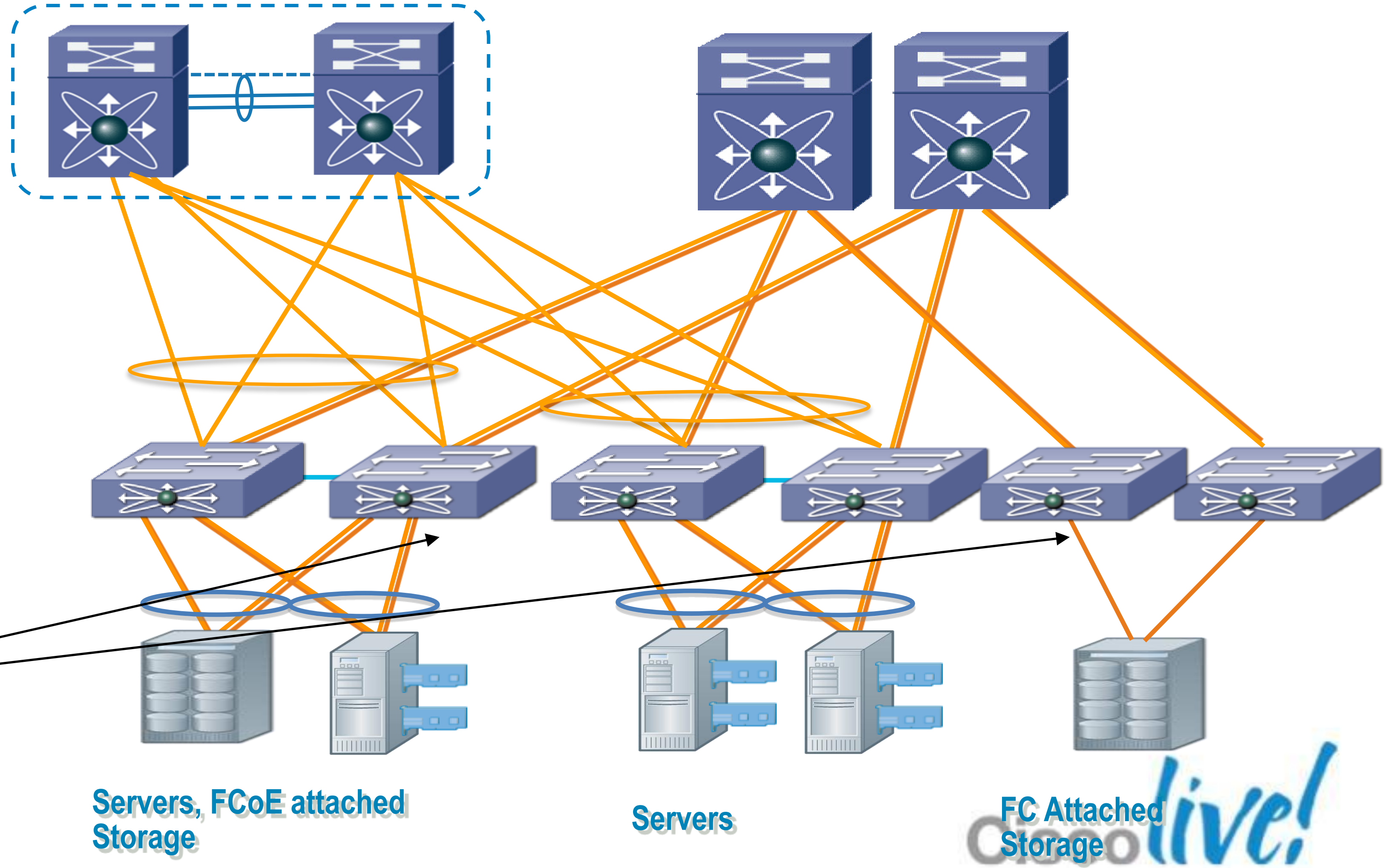
Nexus 5500 Series

Fibre Channel, FCoE and Unified Ports

- Nexus 5000 and 5500 are full feature Fibre Channel fabric switches
  - No support for IVR, FCIP, DMM
- Unified Port supports multiple transceiver types
  - 1G Ethernet Copper/Fibre
  - 10G Ethernet Copper/Fibre
  - 10G DCB/FCoE Copper/Fibre
  - 1/2/4/8G Fibre Channel
- Change the transceiver and connect evolving end devices,
  - Server 1G to 10G NIC migration
  - FC to FCoE migration
  - FC to NAS migration

Any Unified Port can be configured as





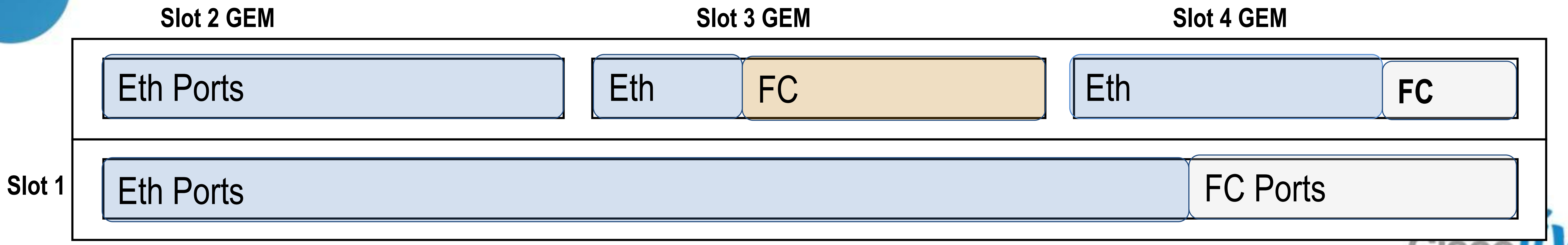
Unified Port – 'Any' device in any rack connected to the same edge infrastructure

Nexus 5500 Series

5548UP/5596UP – UPC (Gen-2) and Unified Ports

- With the 5.0(3)N1 and later releases each module can define any number of ports as Fibre Channel (1/2/4/8 G) or Ethernet (either 1G or 10G)
- Initial SW releases supports only a continuous set of ports configured as Ethernet or FC within each 'slot'
  - Eth ports have to be the first set and they have to be one contiguous range
  - FC ports have to be second set and they have to be contiguous as well
- Future SW release will support per port dynamic configuration

```
n5k(config)# slot <slot-num>
n5k(config-slot)# port <port-range> type <fc | ethernet>
```



Nexus 6004

N6000 only supports FCoE, no native FC at FCS

High Performance

- **Line rate L2 and L3** with all ports and all features and all frame sizes
- **1 us port to port latency** with all frame sizes
- 40Gbps flow
- **40Gbps FCoE**
- Cut-through switching for 40GE and 10GE
- 25MB buffer per 3xQSFP interfaces

High Scalability**

- **96x40GE in 4RU**
- **384x10GE in 4RU**
- **Up to 256K MAC**
- **Up to 128K ARP**
- 32k LPM
- 16K Bridge Domain
- 31 Bi-dir SPAN
- 4K VRF

Feature Rich

- L2 and L3 feature
- FEXlink
- vPC FabricPath TRILL
- FabricPath with segment ID*
- Leaf, spine and border leaf node*
- Adapter-FEX /VM-FEX
- NAT(2K entries)*
- VDC*

Visibility & Analytics

- Line rate SPAN
- Sampled Netflow*
- Buffer Monitoring*
- Latency monitoring*
- Conditional SPAN- SPAN on drop/ SPAN on higher latency*
- Micro-burst monitoring*

Nexus 2000 Series

FCoE Support

N2232PP

32 Port 1/10G FCoE Host Interfaces
8 x 10G Uplinks

N2232TM/N2232TM-E

32 Port 1/10GBASE-T Host Interfaces
8 x 10G Uplinks (Module)

- 32 server facing 10Gig/FCoE ports

  T11 standard based FIP/FCoE support on all ports
- 8 10Gig/FCoE uplink ports for connections to the Nexus 5K, 6K
- Support for DCBx
- N7K will support FCoE on 2232 (Future)

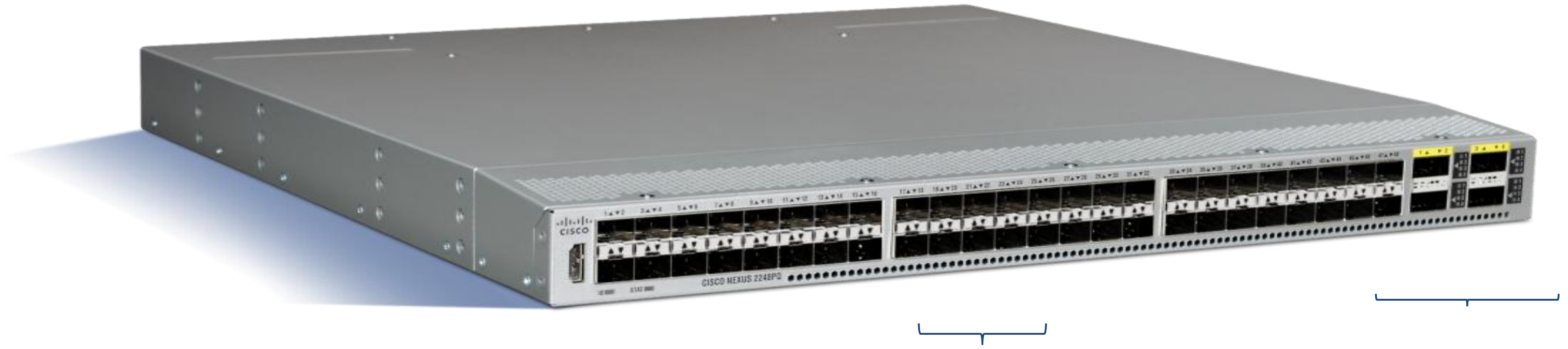
FCoE on 10GBASE-T

- BER characteristics improving with newer generations of PHYs
  - 40nm PHYs (2012) better than 65nm PHYs (2011)
- FCoE support need ~10^{-15} – No single standard
- Working with the ecosystem to define requirement and test
  - Adapter vendors: QLogic, Emulex, Intel, Broadcom
  - Storage vendors: EMC, NetApp
- FCoE **not** supported on Nexus 2232TM
- BER testing underway for Nexus 2232TM-E no FCoE support at FCS
  - Targeting Harbord software release for up to 30M distance
  - Targeting later release for up to 100M distance

Fabric Extender

Nexus 2248PQ-10GE

- 1G/10G SFP+ Fabric Extender
  - 48x 1/10GE SFP+ host interfaces
  - 4x QSFP (16x10GE SFP+) on network interfaces
  - Front-to-back airflow and back-to-front airflow
  - Additional uplink buffers (2x16MB)
- Design scenarios:
  - High Density 10GE SFP+ ToR
  - Connectivity Flexibility
  - Virtualised environments
  - Storage consolidation (FCoE, iSCSI, NFS…)
  - Predictable Low latency across large number of ports

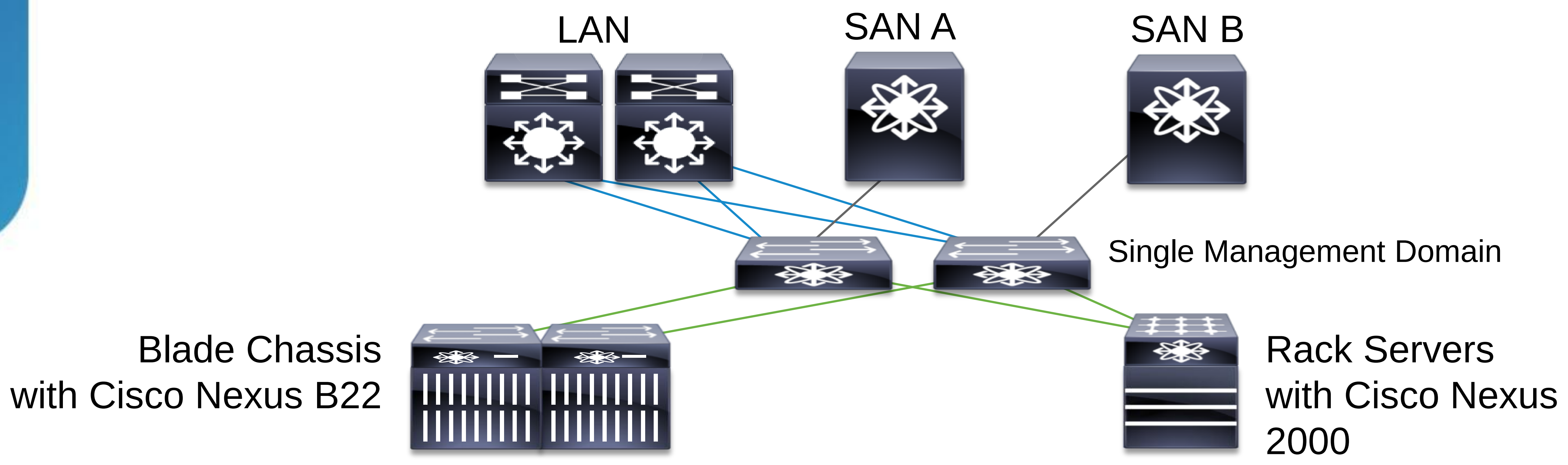


Cisco Nexus B22 Use Case

Legacy Blade and Rack Server Footprint

Customer Desires a Cisco Unified Fabric

- Consolidation of switch modules and cabling
- Network management point consolidation and consistency with rack servers
- Nexus Fabric Visibility within Blade Chassis
- Require end-to-end FCoE and/or FabricPath

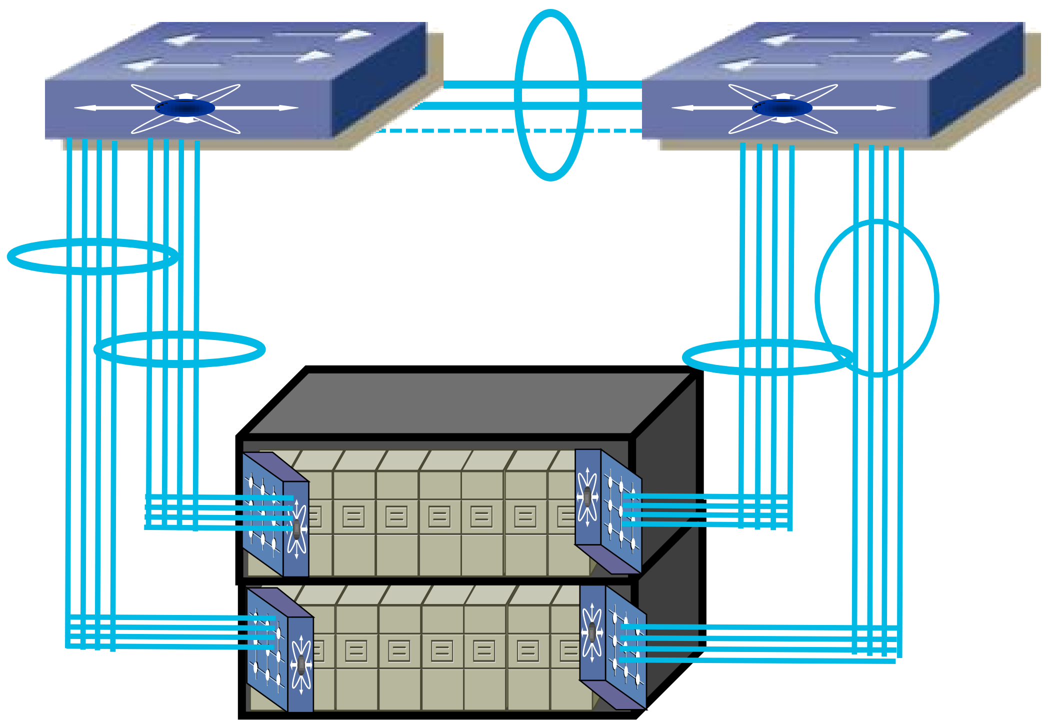


DC Design Details – Blade Chassis

Nexus B22 Series Fabric Extender

- B22 extends FEX connectivity into the HP blade chassis
- Cisco Nexus 5000 Switch is a single management point for all the blade chassis I/O modules
- 66% decrease in blade management points*
- Blade & rack networking consistency
- Interoperable with Nexus 2000 Fabric Extenders in the same Nexus parent switch
- *End-to-end FCoE support*
- Support for 1G & 10G, LOM and Mez
- Dell supports Pass-Thru as an alternative option to directly attaching Blade Servers to FEX ports

Cisco Nexus B22 Series Blade FEX



Nexus 5500 + B22 (HP FEX)

Cisco Nexus B22 Fabric Extenders

FEX Connectivity for the Blade Server Ecosystem

FEATURES
- Extends FEX connectivity into blade chassis
- Cisco Nexus Switch is a single management point for all the blade chassis I/O modules
- End-to-end FCoE support

BENEFITS:
- 50% decrease in blade chassis I/O modules
- 66% decrease in blade management points
  - Blade & rack networking consistency
- Increased network resiliency

Cisco Nexus B22 H

Cisco Nexus B22 F

Cisco Nexus B22 D

Cisco Nexus B22F Fabric Extender Overview

- 16 x 10 GE server interfaces
- 8 x 10 GE network interfaces
- Host vPC (virtual Port-Channel)
- DCB and FCoE in 10G mode
- Upstream Nexus 5000 supports FEX mix & match
- 8 QoS queues (6 configurable)

- Fabric link interconnects:
  - Fabric Extender Transceiver (FET)

    1/3/5M Twinax, 7/10M active Twinax, SR, LR, ER
  - NX-OS version 5.2(1)N1(1) or greater for the

    Nexus 5000/5500

Fujitsu Blade Enclosures: Primergy BX900/BX400

Requires upstream Nexus 5000/55xx/600x

Cisco Nexus B22 DELL Fabric Extender Overview

- 16 x 10 GE server interfaces
- 8 x 10 GE network interfaces
- Host vPC (virtual Port-Channel)
- DCB and FCoE in 10G mode
- 8 QoS queues (6 configurable)
- Fabric link interconnects:
  – Fabric Extender Transceiver (FET)
    1/3/5M Twinax, 7/10M active Twinax, SR, LR, ER
- NX-OS version 5.2(1)N2 or greater for the Nexus 5500
- Upstream Nexus 5500 platform supports FEX mix & match
- Supported with the PowerEdge M1000e Blade Enclosures

Requires upstream Nexus 5500/600X Platform

Nexus 7000 F-Series SFP+ Module

FCoE Support

- 32 & 48 port 1/10 GbE for Server Access and Aggregation
- F1 Supports FCoE
- F2 support for FCoE targeted
  - FEX + FCoE support – 2HCY12
- 10 Gbps Ethernet supporting Multiprotocol Storage Connectivity
  - Supports FCoE, iSCSI and NAS
  - Loss-Less Ethernet: DCBX, PFC, ETS
- Enables Cisco FabricPath for increased bisectional bandwidth for iSCSI and NAS traffic
- FCoE License (N7K-FCOEF132XP)
  - One license per F1/F2 module
- SAN Enterprise (N7K-SAN1K9)
  - One license per chassis
  - IVR, VSAN Based Access Control, Fabric Binding
- Supervisor 2/2E required to enable FCoE on F2 modules

32-port F1 Series

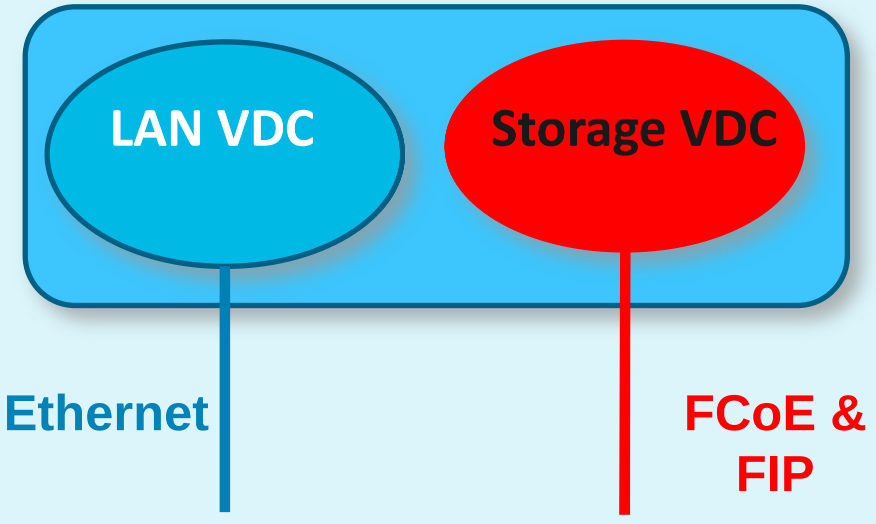
48-port F2 Series

Storage VDC on the Nexus 7000

Supported VDC models

- Separate VDC running ONLY storage related protocols
- Storage VDC: a *virtual* MDS FC switch
- Running only FC related processes
- Only one such VDC can be created
- Provides control plane separation

Dedicated for VE_Ports

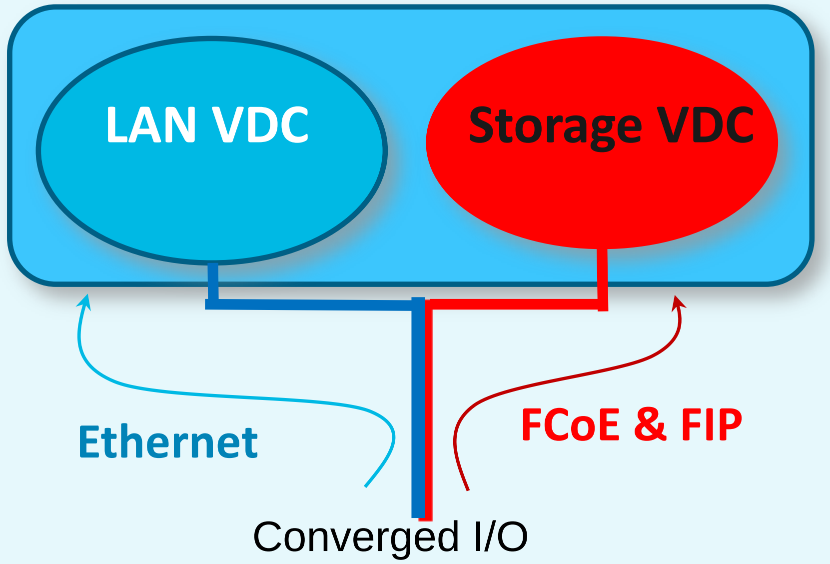


Shared Converged Port

Model for host/target interfaces, not VE_Port

Ingress Ethernet traffic is split based on frame ether-type

FCoE traffic is processed in the context of the Storage VDC



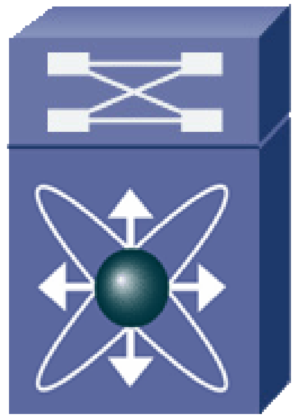
Creating the Storage VDC

- Create VDC of type storage and allocate non-shared interfaces:

```
N7K-50(config)# vdc fcoe id 2 type storage
N7K-50(config-vdc)# allocate interface Ethernet4/1-16, Ethernet4/19-22
```

- Allocate FCoE vlan range from the Owner VDC to the Storage VDC. This is a necessary step for sharing interfaces to avoid vlan overlap between the Owner VDC and the Storage VDC

```
N7K-50(config) vdc fcoe id 2
N7K-50(config-vdc)# allocate fcoe-vlan-range 10-100 from vdcs n7k-50
```



N7K only

- Allocated the shared interfaces:

```
N7K-50(config-vdc)# allocate shared interface Ethernet4/17-18
```

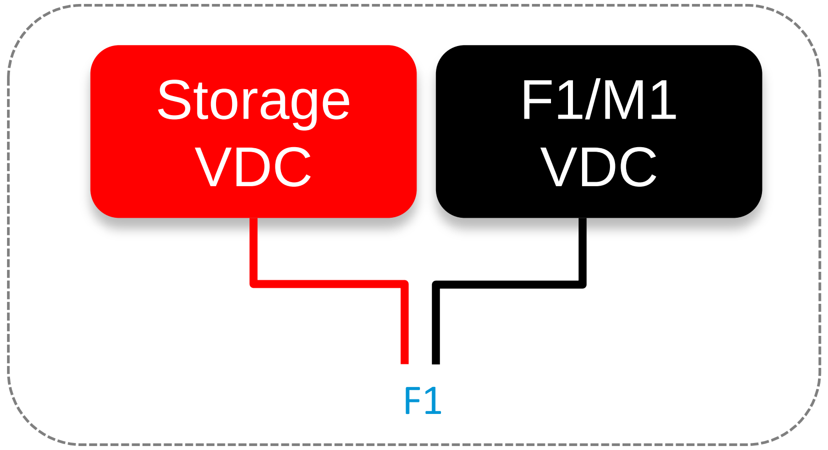
- Install the license for the FCoE Module.

```
n7k-50(config)# license fcoe module 4
```

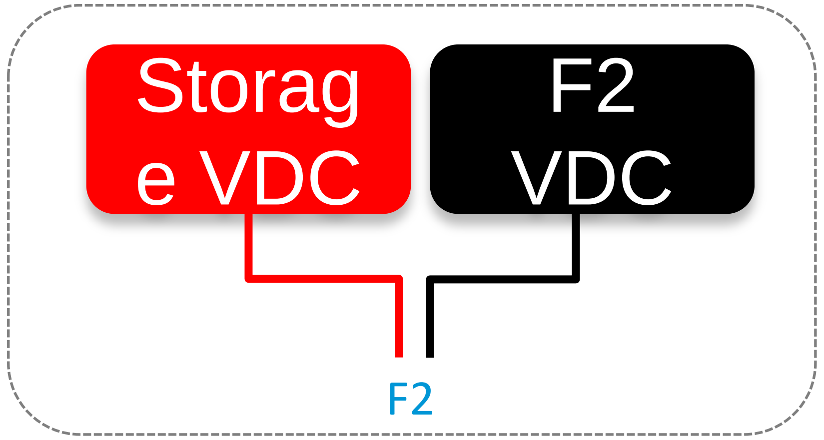
Storage VDC
F2 line cards

- Some restrictions when using mixed line cards (F1/F2/M1)
  - F2 ports need to be in a dedicated VDC if using ‘shared ports’

Shared Ports

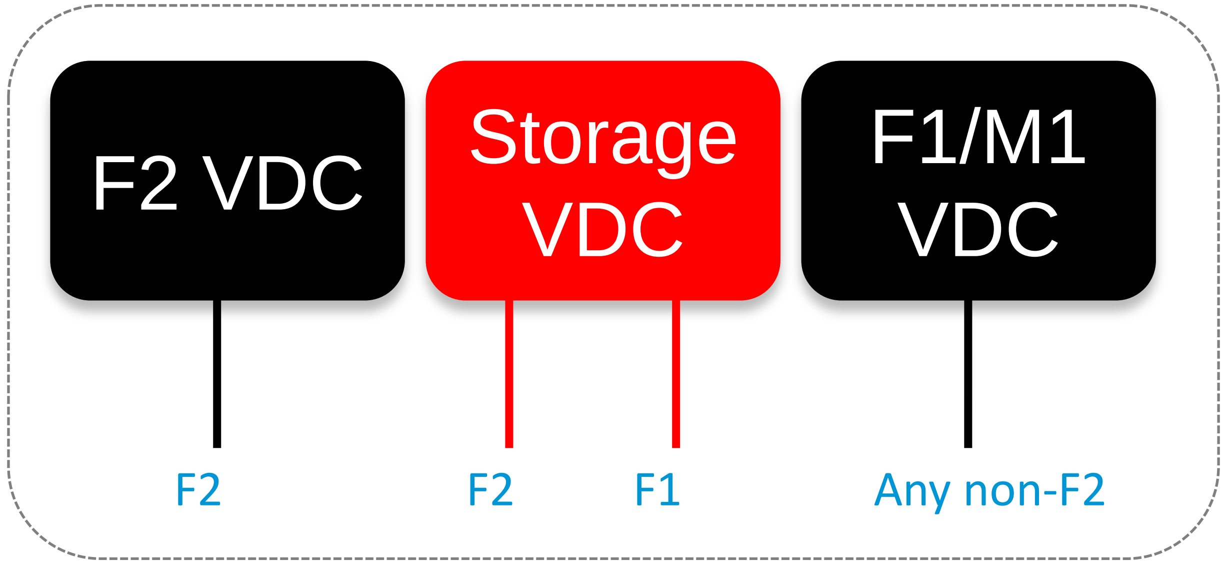


NX-OS 5.2



NX-OS 6.1

Dedicated Ports



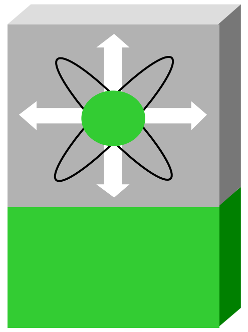
NX-OS 6.1

MDS 9000 8-Port 10G FCoE Module

FCoE Support

- Enables integration of existing FC infrastructure into Unified Fabric
  - 8 FCoE ports at 10GE full rate in MDS 9506, 9509, 9513
  - No FCoE License Required
- Standard Support
  - T11 FCoE
  - IEEE DCBX, PFC, ETS
- Connectivity – FCoE Only, No LAN
  - VE to Nexus 5000/6000, Nexus 7000, MDS 9500
  - VF to FCoE Targets
- Optics Support
  - SFP+ SR/LR, SFP+ 1/3/5m Passive, 7/10m Active CX-1 (TwinAx)
- Requirements
  - SUP2A
  - Fabric 2 modules for the backplane (applicable to 9513 only)

MDS 9500



MDS 9000 8-Port 10G FCoE Module

FCoE Support

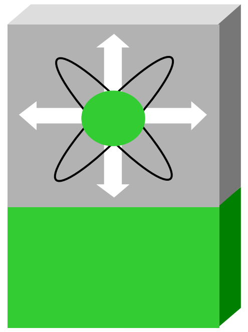
There is no need to enable FCoE explicitly on the MDS switch. The following features will be enabled once an FCoE capable linecard is detected.

Install feature-set fcoe
feature-set fcoe
feature lldp
feature vlan-vsan-mapping

Create VSAN and VLAN, Map VLAN to VSAN for FCoE

```
pod3-9513-71(config)# vsan database
pod3-9513-71(config-vsan-db)# vsan 50
pod3-9513-71(config-vsan-db)# vlan 50
pod3-9513-71(config-vlan)# fcoe vsan 50
```

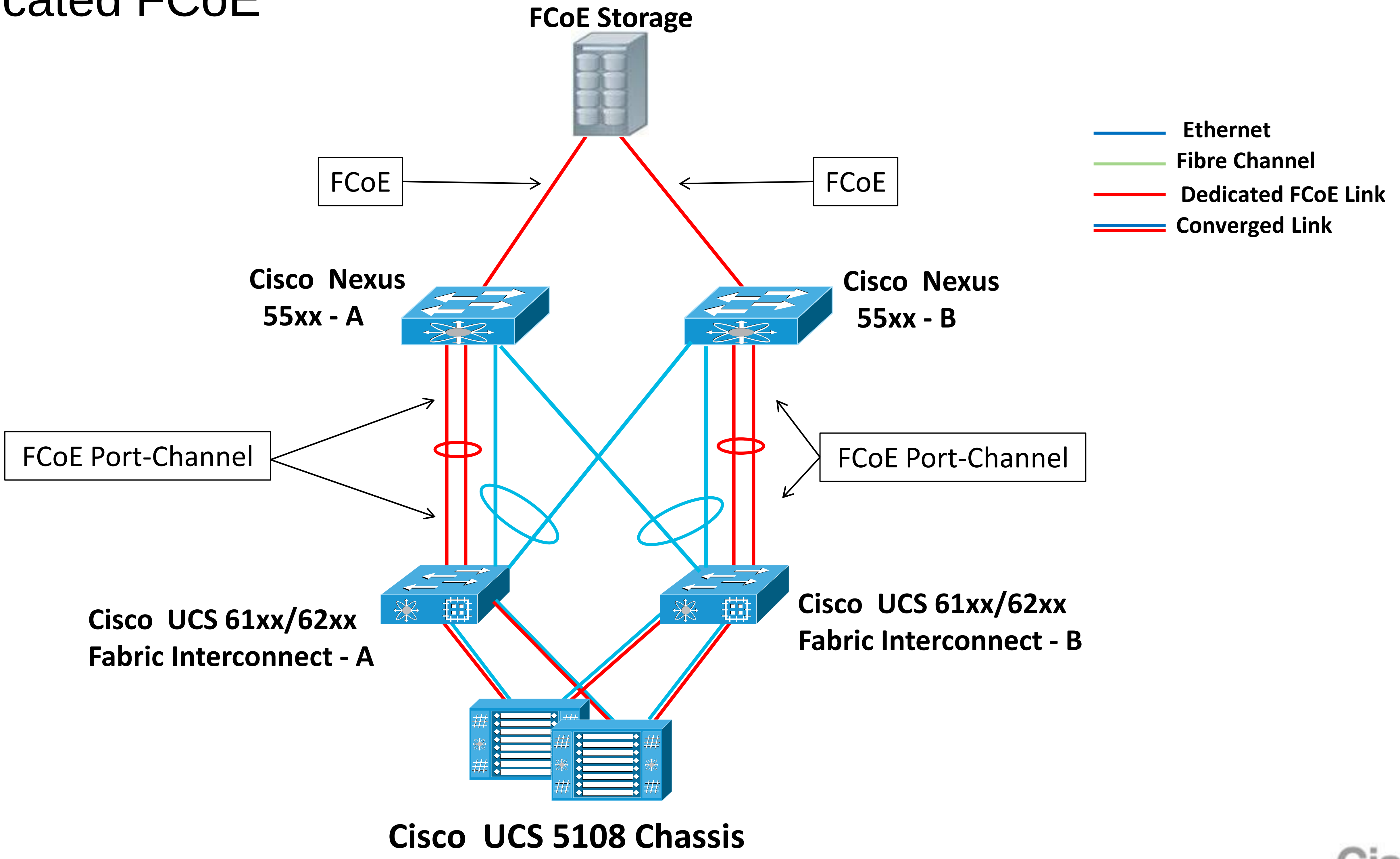
Build the LACP Port Channel on the MDS
Create VE port and assign to the LACP Port-channel



```
pod3-9513-71(config-if-range)# interface vfc-port-channel 501
pod3-9513-71(config-if)# switchport mode e
pod3-9513-71(config-if)# switchport trunk allowed vsan 50
pod3-9513-71(config-if)# no shut
```

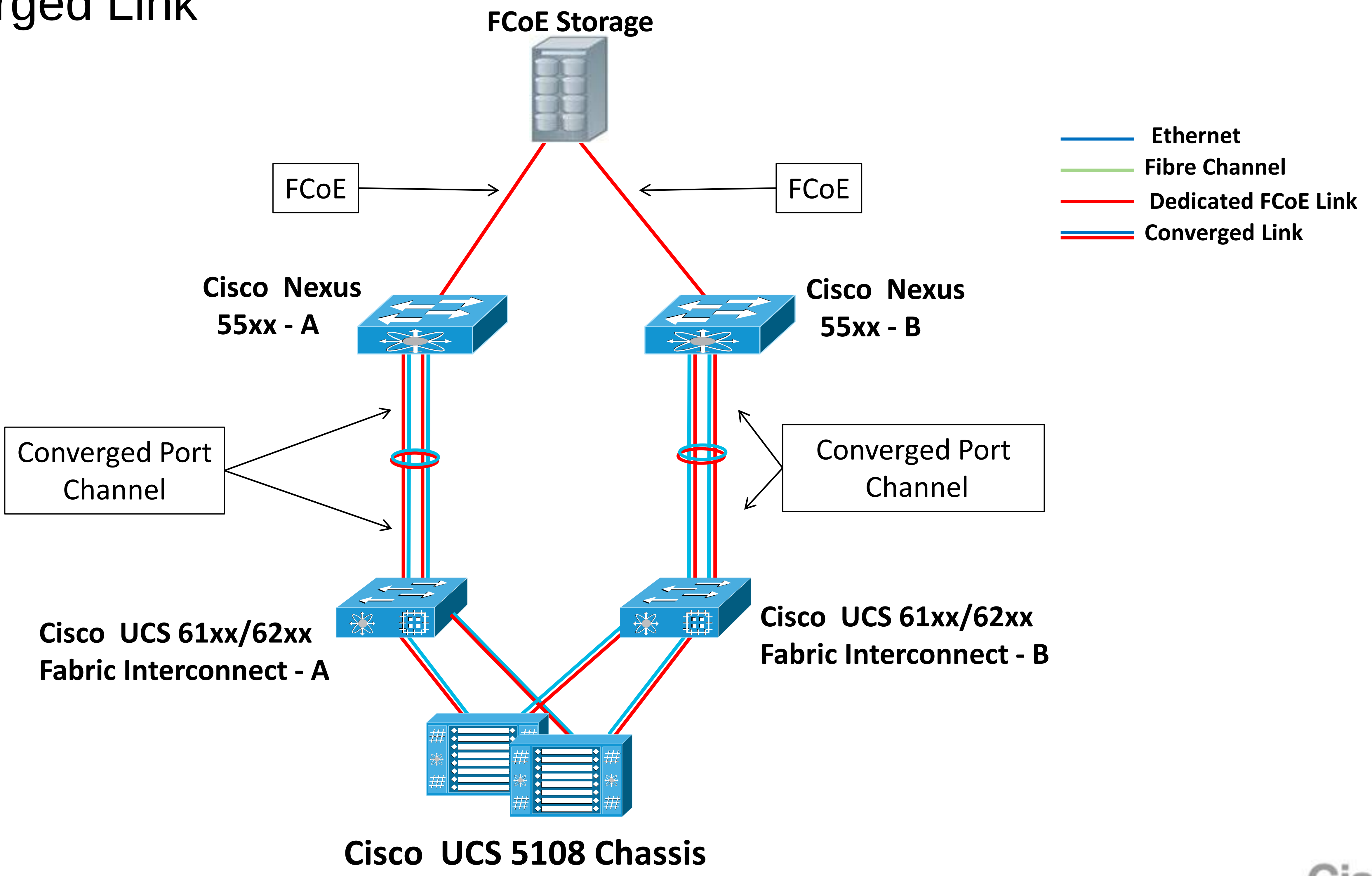
UCS Design Converged Multihop

Dedicated FCoE



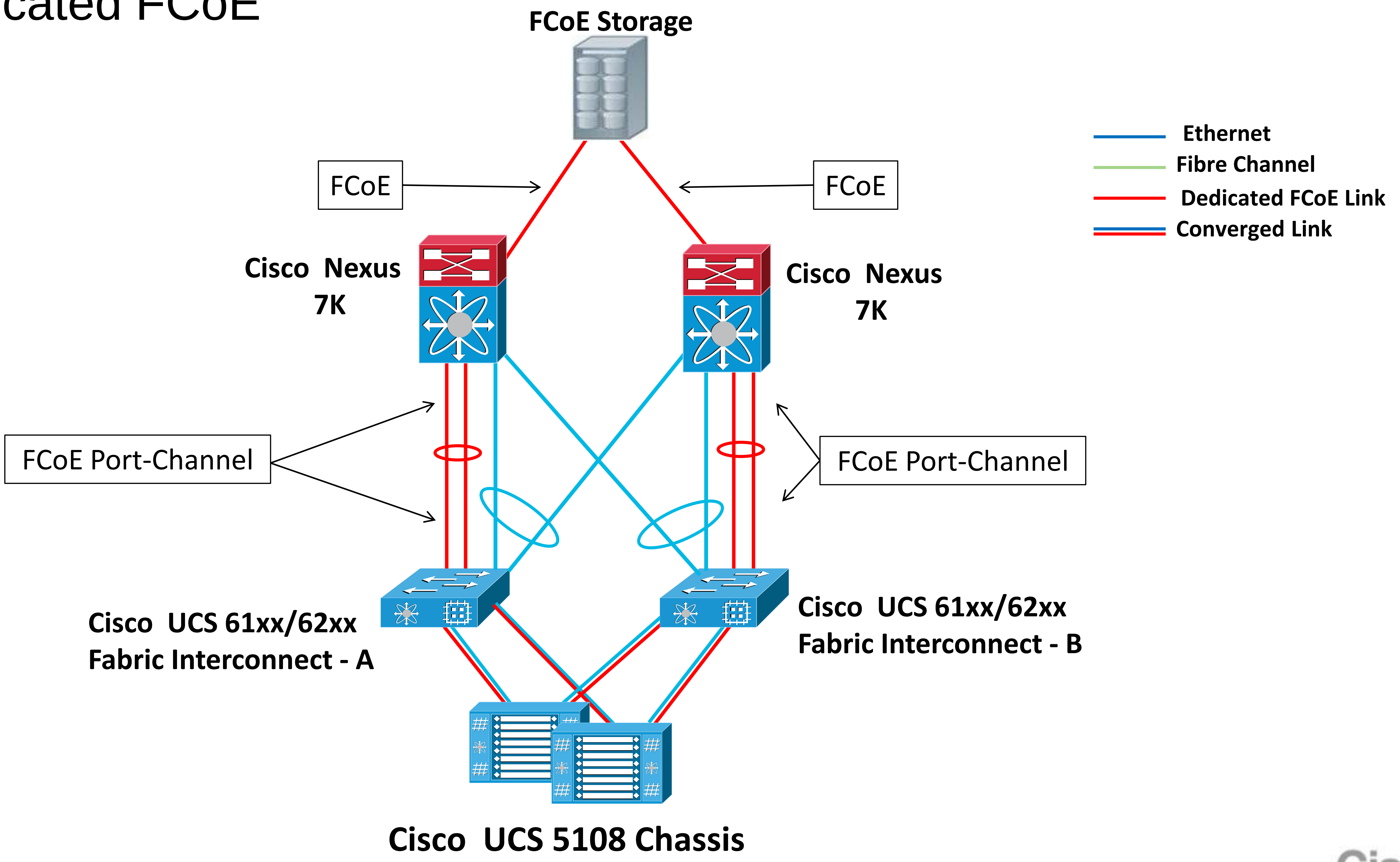
UCS Design Converged Multihop

Converged Link



UCS Design N7k Converged Multihop

Dedicated FCoE



Agenda

- Unified Fabric – What and Why
- FCoE Protocol Fundamentals
- Nexus FCoE Capabilities
- FCoE Network Requirements and Design Considerations
- DCB & QoS - Ethernet Enhancements
- Single Hop Design
- Multi-Hop Design
- Futures

Network vs. Fabric

Differences & Similarities

- Ethernet is non-deterministic.
  - Flow control is destination-based
  - Relies on TCP drop-retransmission / sliding window
- Fibre Channel is deterministic.
  - Flow control is source-based (B2B credits)
  - Services are fabric integrated (no loop concept)

Networks

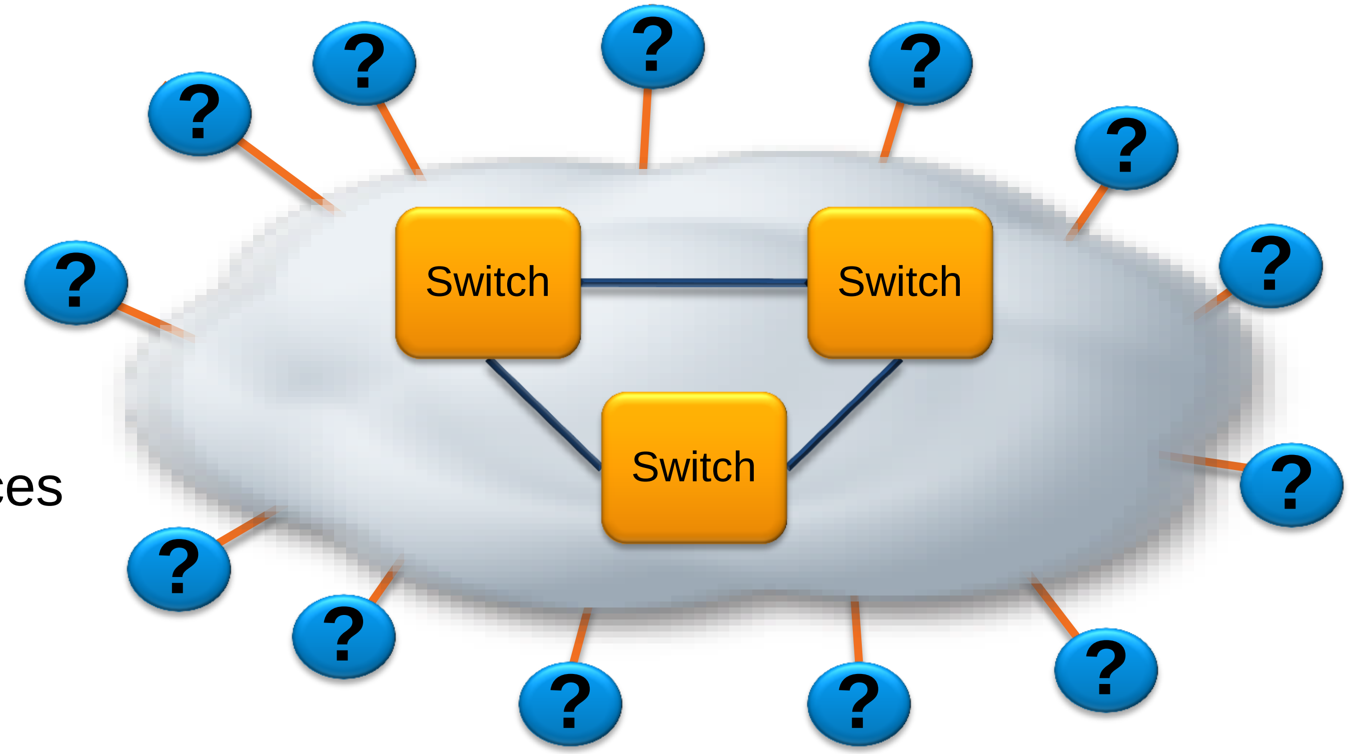
- Connectionless
- Logical circuits
- Unreliable transfers
- High connectivity
- Higher latency
- Longer distance
- Software intense

Channels

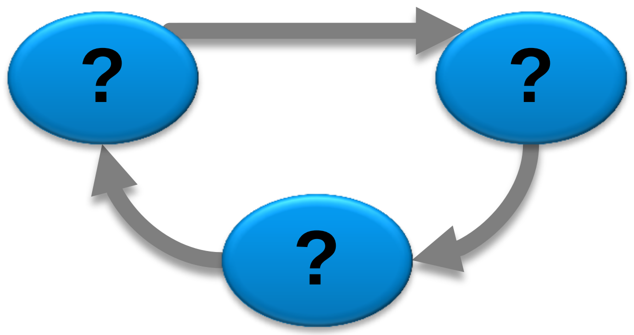
- Connection service
- Physical circuits
- Reliable transfers
- High speed
- Low latency
- Short distance
- Hardware intense

Network vs. Fabric
Classical Ethernet

- Ethernet/IP
  - –**Goal : provide any-to-any connectivity**
  - –Unaware of packet loss (“lossy”) – relies on ULPs for retransmission and windowing
  - –Provides the transport without worrying about the services - ***services provided by upper layers***
  - –East-west vs north-south traffic ratios are undefined
- Network design has been optimised for:
  - –High Availability from a transport perspective by connecting nodes in mesh architectures
  - –Service HA is implemented separately
  - –Takes in to account control protocol interaction (STP, OSPF, EIGRP, L2/L3 boundary, etc…)



Fabric topology and traffic flows are highly flexible

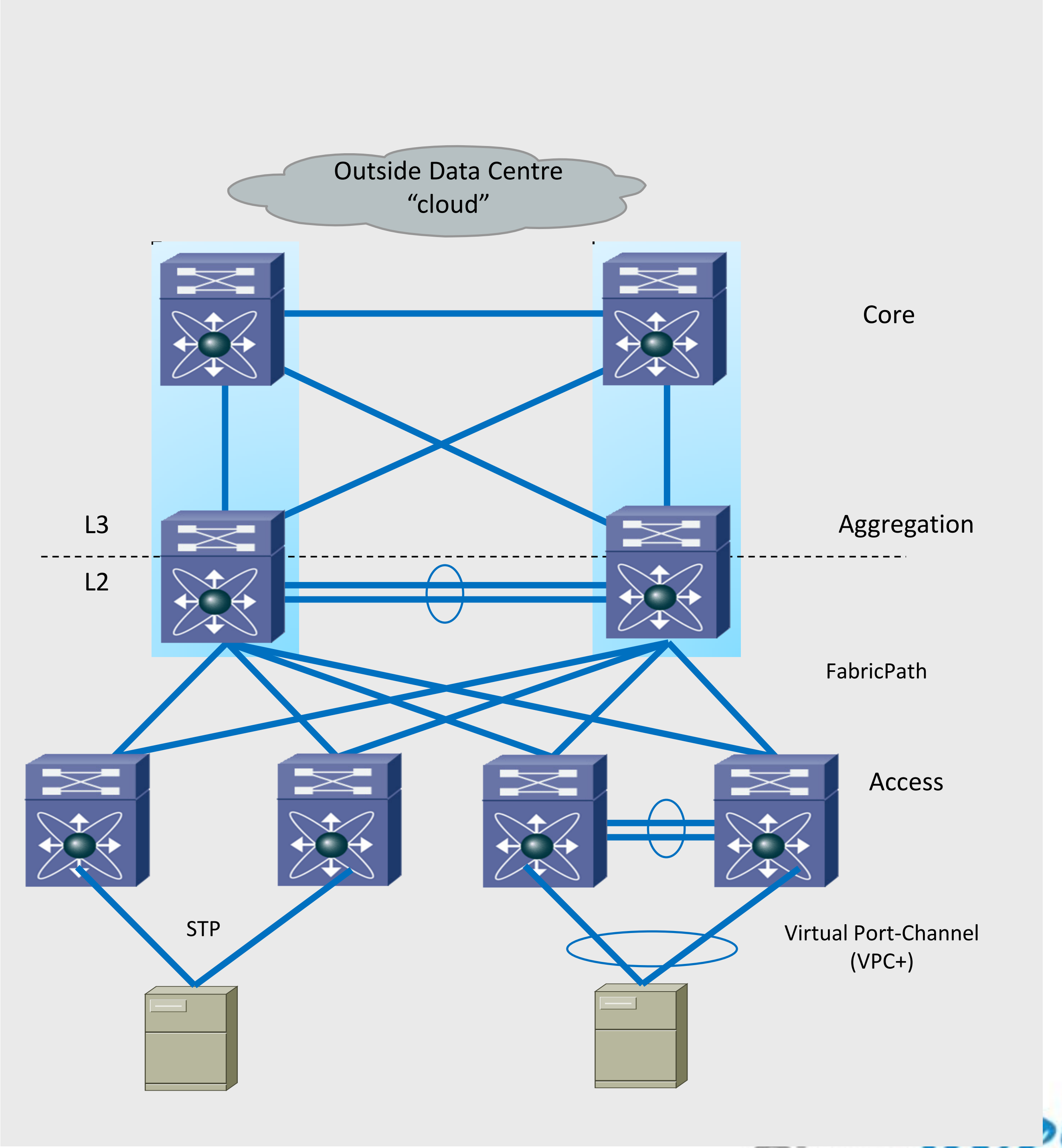


Client/Server Relationships are not pre-defined

Network vs. Fabric

LAN Design – Access/Aggregation/Core

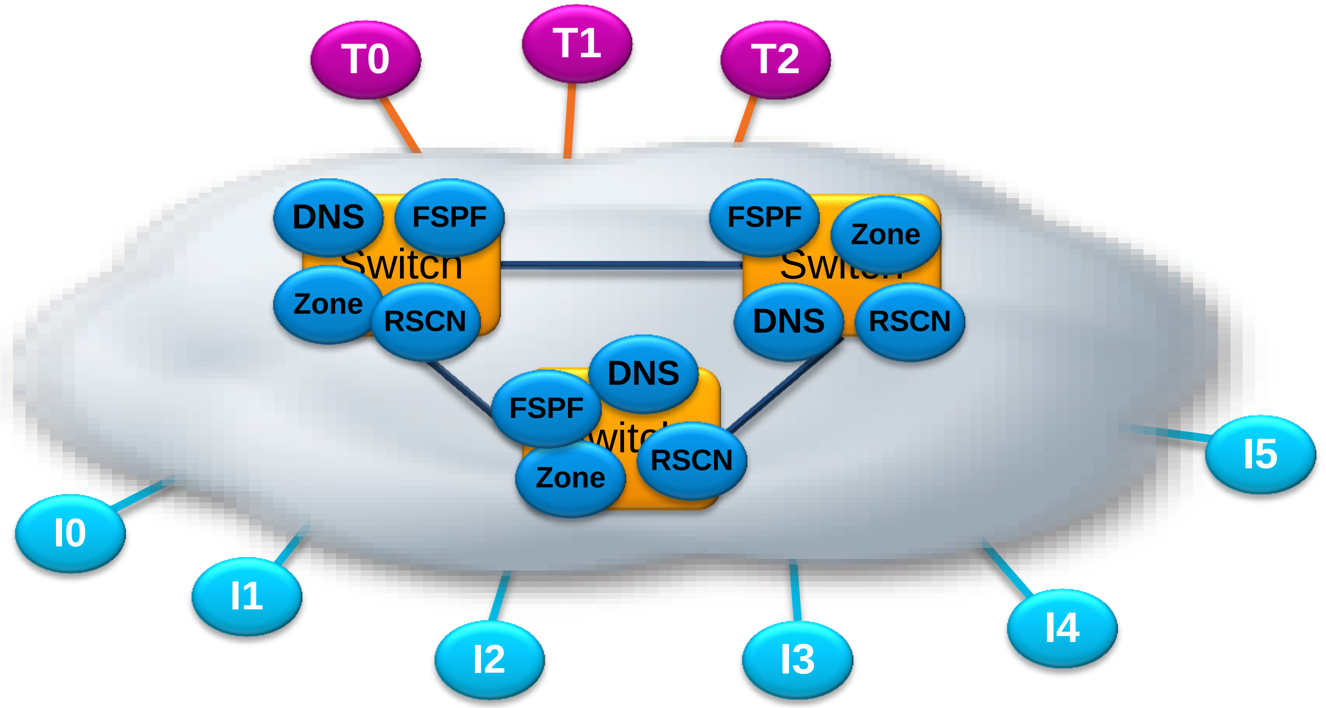
- Servers typically dual homed to two or more access switches
- LAN switches have redundant connections to the next layer
- Distribution and Core can be collapsed into a single box
- L2/L3 boundary typically deployed in the aggregation layer
  - –Spanning tree or advanced L2 technologies (vPC/FabricPath) used to prevent loops within the L2 boundary
  - –L3 routes are summarised to the core
- Services deployed in the L2/L3 boundary of the network (load-balancing, firewall, NAM, etc)



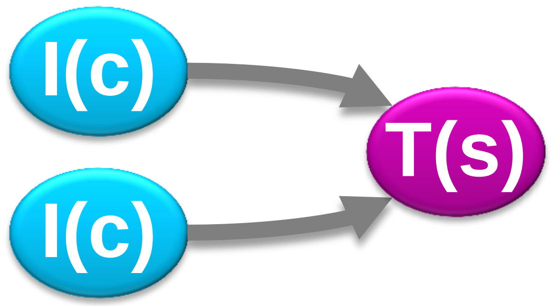
Network vs. Fabric

Classical Fibre Channel

- Fibre Channel SAN
  - –Transport and Services are on the **same layer** in the same devices
  - –Well defined end device relationships (initiators and targets)
  - –Does not tolerate packet drop – requires **lossless** transport
  - –Only north-south traffic, east-west traffic mostly irrelevant
- Network designs optimised for Scale and Availability
  - –High availability of network services provided through dual fabric architecture
  - –Edge/Core vs Edge/Core/Edge
  - –Service deployment



Fabric topology, services and traffic flows are structured

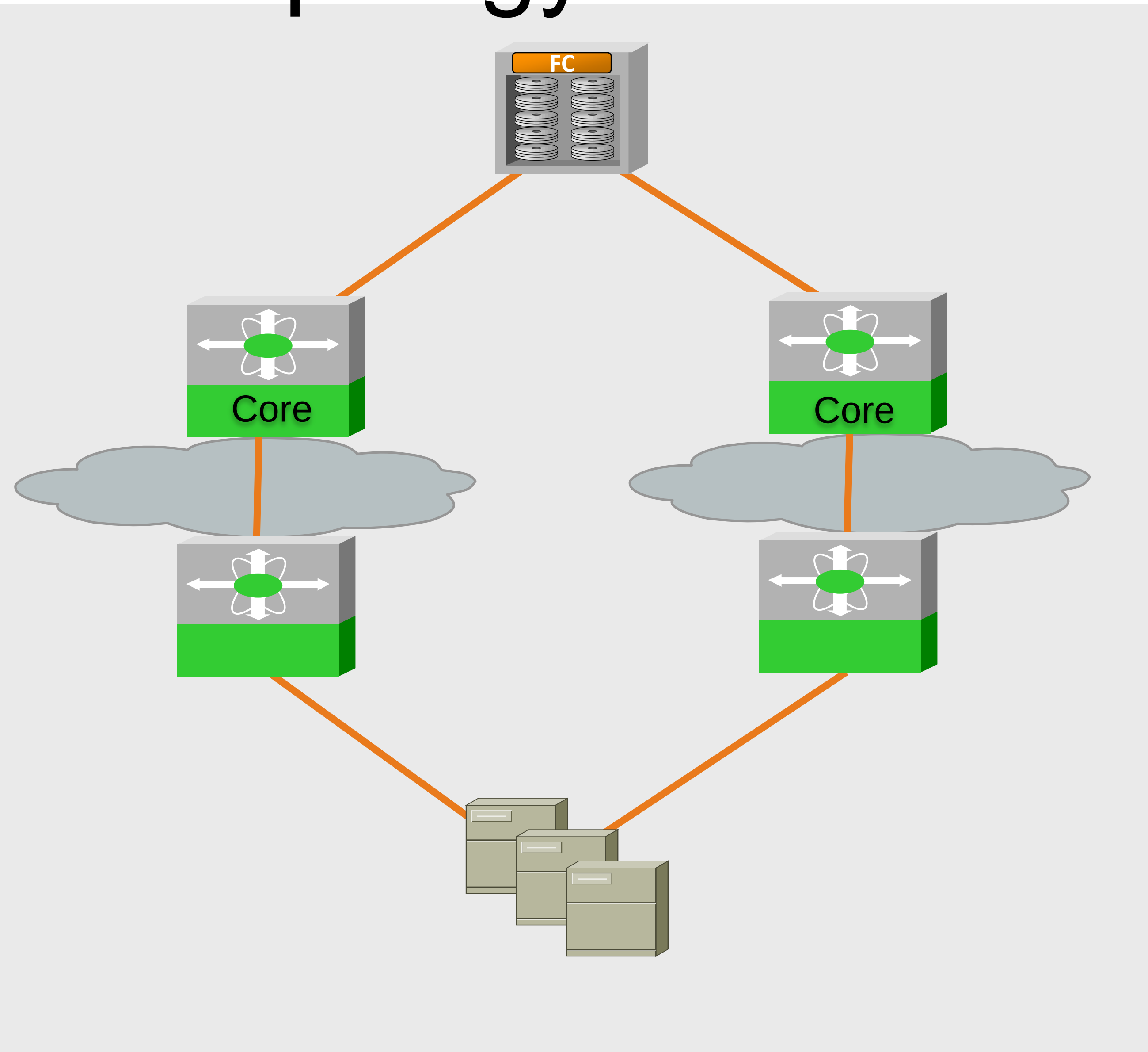


Client/Server Relationships are pre-defined

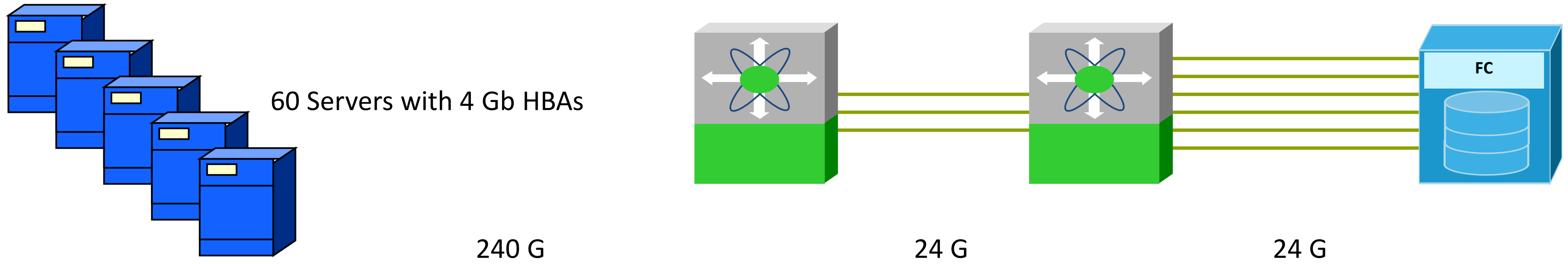
Network vs. Fabric

SAN Design – Two ‘or’ Three Tier Topology

- “Edge-Core” or “Edge-Core-Edge” Topology
- Servers connect to the edge switches
- Storage devices connect to one or more core switches
- HA achieved in two physically separate, but identical, redundant SAN fabric
- Very low oversubscription in the fabric (1:1 to 12:1)
- FLOGI Scaling Considerations

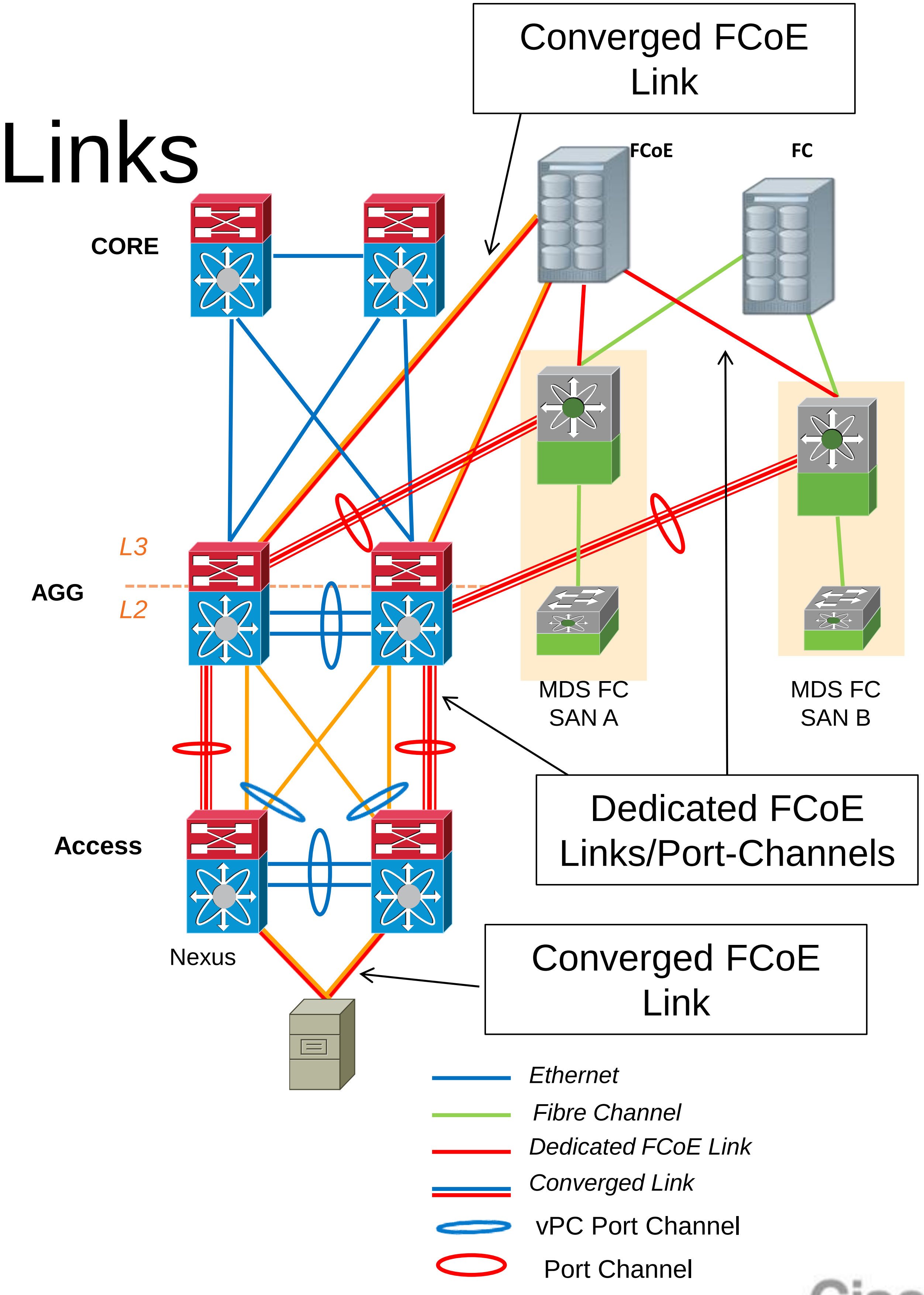


Example: 10:1 O/S ratio



Network vs. Fabric
Converged and Dedicated Links

- **Converged Link** to the access switch
  - Cost savings in the reduction of required equipment
  - “cable once” for all servers to have access to both LAN and SAN networks
- **Dedicated Link** from access to aggregation
  - Separate links for SAN and LAN traffic - both links are same I/O (10GE)
  - Advanced Ethernet features can be applied to the LAN links
  - Maintains fabric isolation



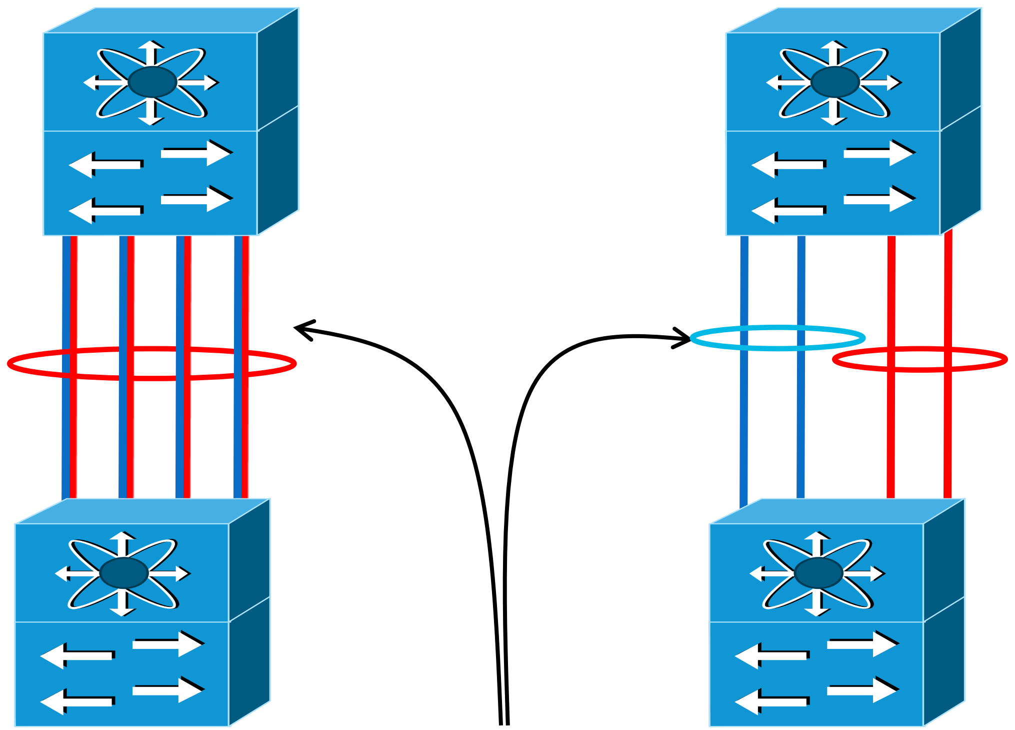
Dedicated vs. Converged ISLs

Why support dedicated ISLs as oppose to Converged?

Converged

- ✓One wire for all traffic types
- ✓ETS: QoS output feature guarantees minimum bandwidth allocation
- ✓No Clear Port ownership
- ✓Desirable for DCI Connections

Available on Nexus 5x00
Nexus 7000 Support Under Consideration



Agg BW: **40G**
FCoE: **20G**
Ethernet: **20G**
HA: 4 Links Available

Dedicated

- ✓Dedicated wire for a traffic type
- ✓No Extra output feature processing
- ✓Distinct Port ownership
- ✓Complete Storage Traffic Separation

Available on Nexus 5x00
Nexus 7000 Supported at NX-OS 5.2(1)

Different methods, Producing the ***same*** *aggregate bandwidth*

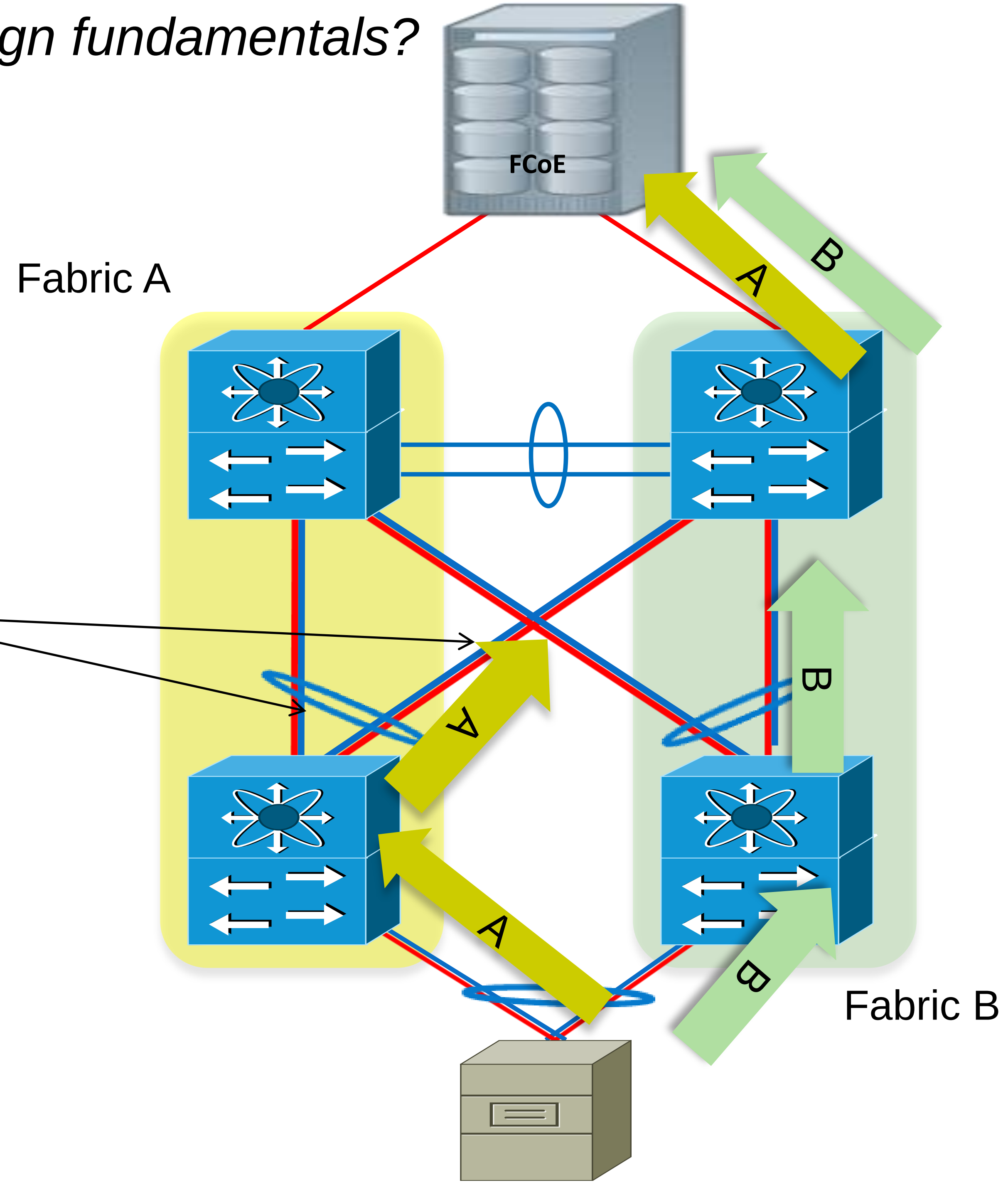
Dedicated Links *provide additional isolation of Storage Traffic*

Converged Links and vPC

Shared wire and VPC – does it break basic SAN design fundamentals?

Now that I have Converged Link Support.
Can I deploy vPC for my Storage Traffic?

- vPC with Converged Links provides an Active-Active connection for FCoE traffic
- Seemingly more bandwidth to the Core…
- Ethernet forwarding behaviour can break SAN A/B separation

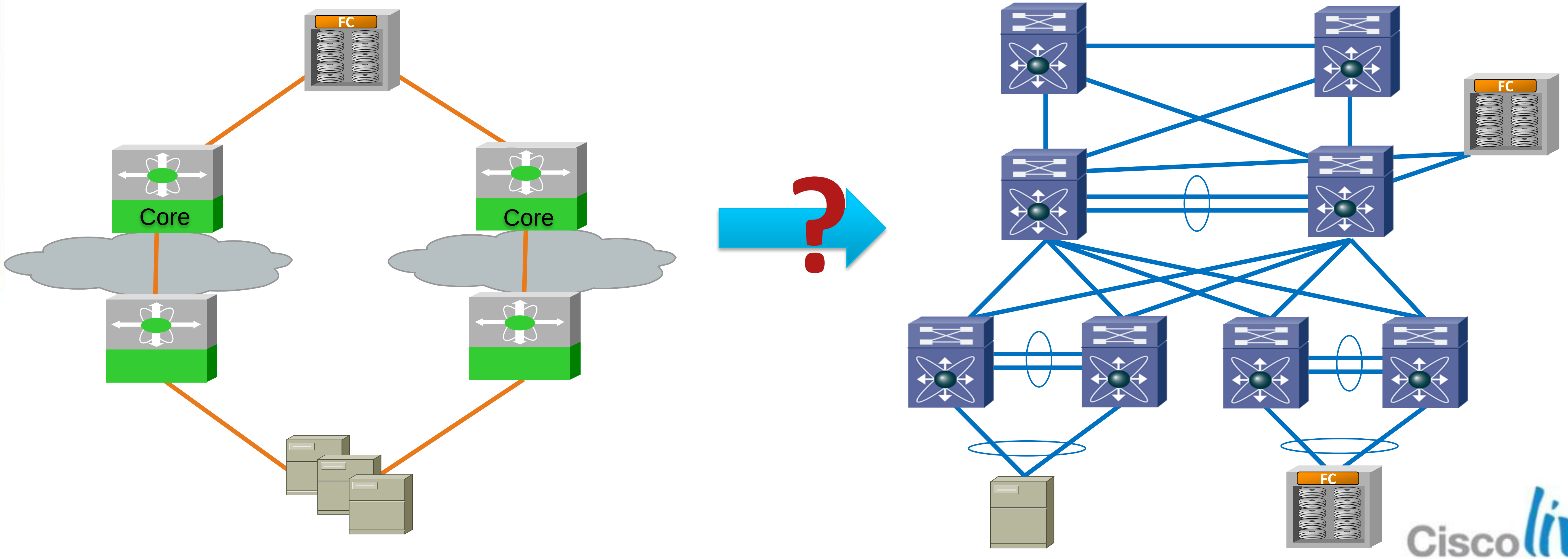


Currently Not supported on Nexus Switches (exception is the dual homed FEX - EVPC)

“Fabric vs. Network” or “Fabric & Network”

SAN Dual Fabric Design

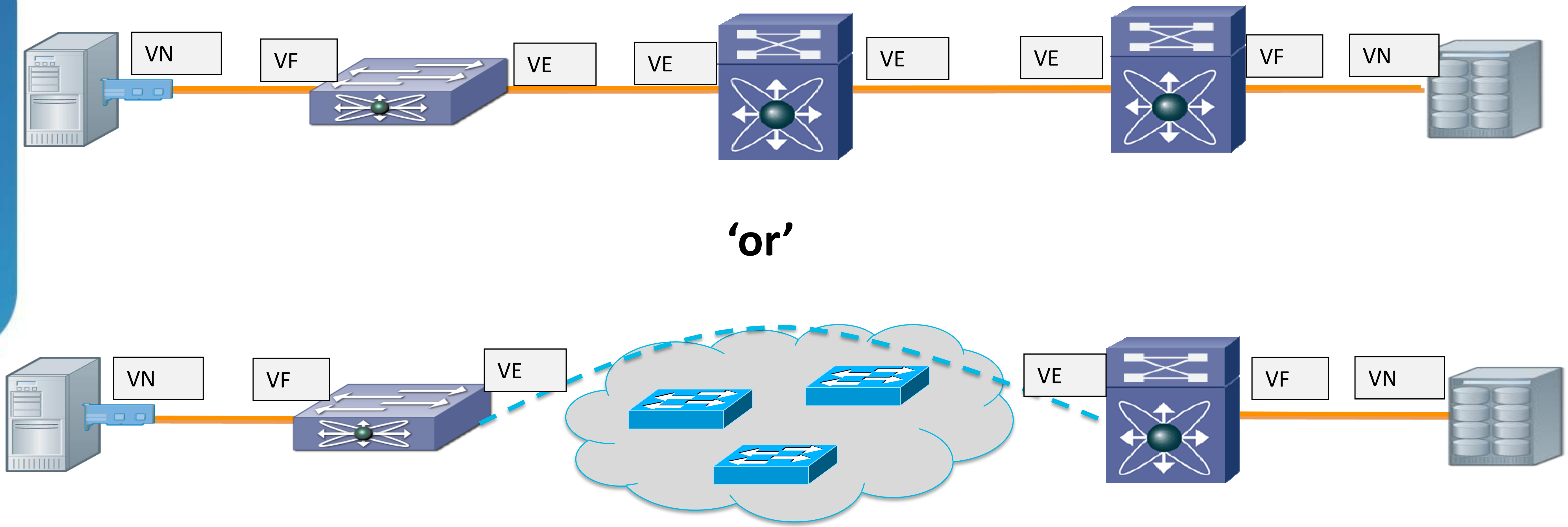
- Will you migrate the SAN dual fabric HA model into the LAN full meshed HA model
  - Is data plane isolation required? (traffic engineering)
  - Is control plane isolation required? (VDC, VSAN)



“Fabric vs. Network” or “Fabric & Network”

Hop by Hop or Transparent Forwarding Model

- A number of big design questions for you
  - Do you want a ‘routed’ topology or a ‘bridged’ topology
  - Is FCoE a layer 2 overlay or integrated topology (ships in the night)



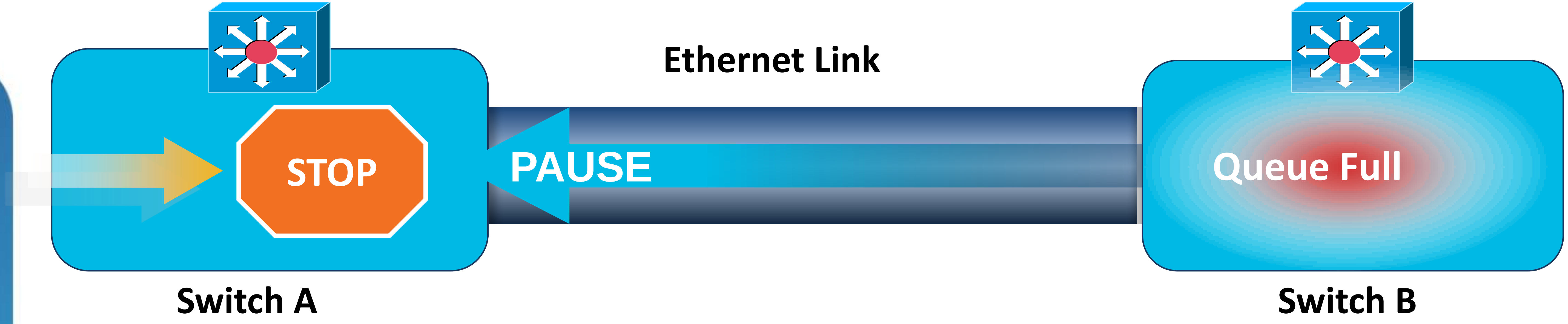
Agenda

- Unified Fabric – What and Why
- FCoE Protocol Fundamentals
- Nexus FCoE Capabilities
- FCoE Network Requirements and Design Considerations
- DCB & QoS - Ethernet Enhancements
- Single Hop Design
- Multi-Hop Design
- Futures

Ethernet Enhancements

Can Ethernet Be Lossless?

- Yes, with Ethernet PAUSE Frame



- Defined in IEEE 802.3—Annex 31B
- The PAUSE operation is used to inhibit transmission of data frames for a specified period of time
- Ethernet PAUSE transforms Ethernet into a lossless fabric, a requirement for FCoE

Ethernet Enhancements

IEEE DCB

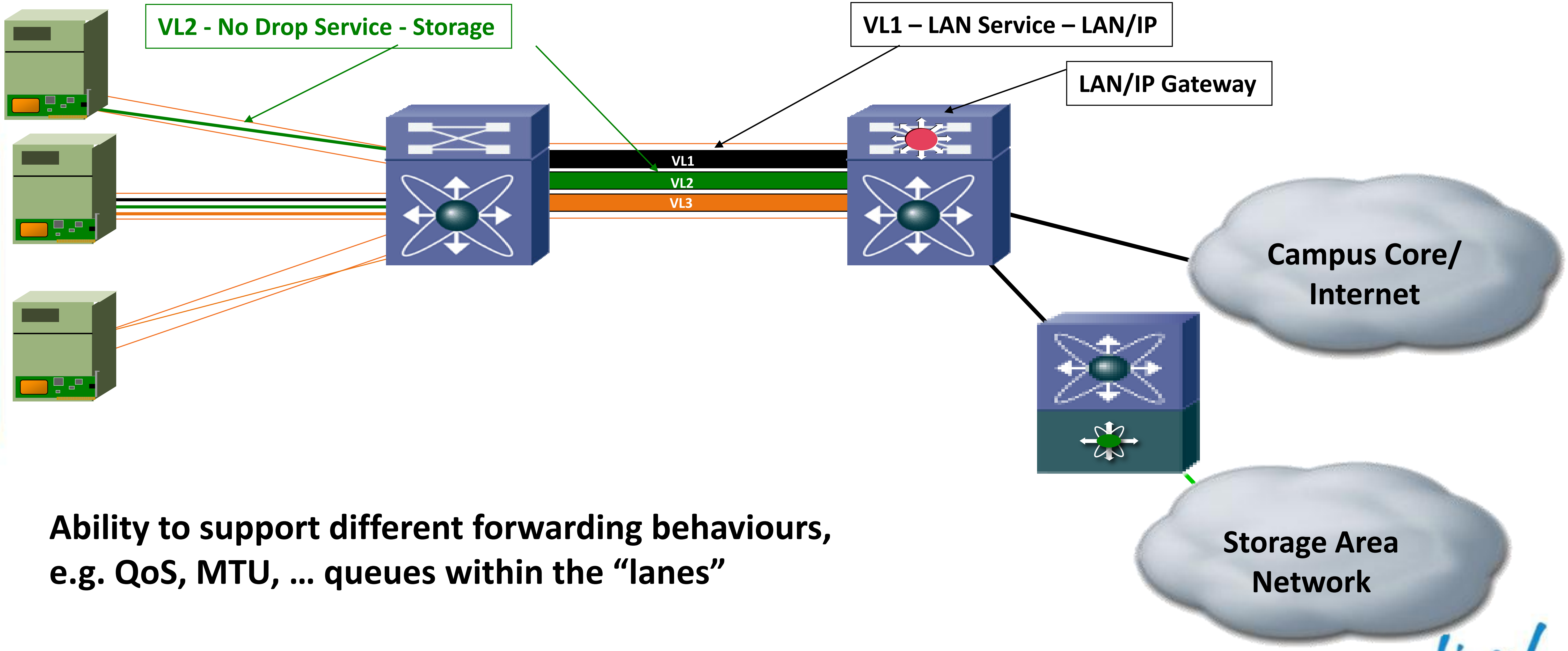
- **Developed by IEEE 802.1 Data Centre Bridging Task Group (DCB)**
- **All Standards Complete**

Standard / Feature	Status of the Standard
IEEE 802.1Qbb Priority-based Flow Control (PFC)	Completed
IEEE 802.3bd Frame Format for PFC	Completed
IEEE 802.1Qaz Enhanced Transmission Selection (ETS) and Data Centre Bridging eXchange (DCBX)	Completed
IEEE 802.1Qau Congestion Notification	Complete, published March 2010
IEEE 802.1Qbh Port Extender	In its first task group ballot

CEE (Converged Enhanced Ethernet) is an informal group of companies that submitted initial inputs to the DCB WGs.

Ethernet Enhancements

DCB "Virtual Links"

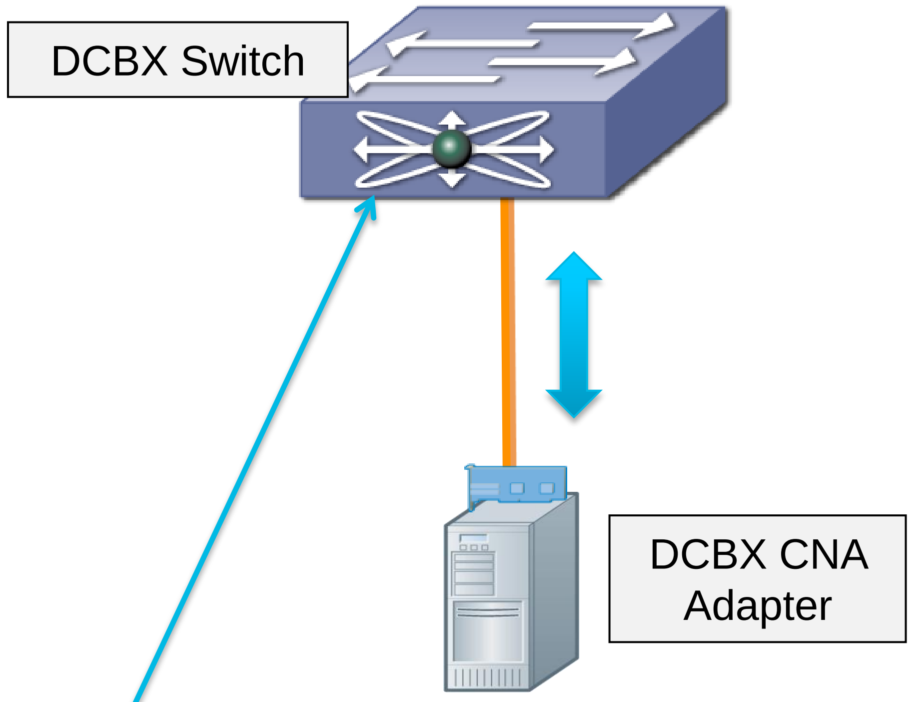


Ability to support different forwarding behaviours, e.g. QoS, MTU, … queues within the "lanes"

Data Centre Bridging Control Protocol

DCBX Overview - 802.1Qaz

- Negotiates Ethernet capability's : PFC, ETS, CoS values between DCB capable peer devices
- Simplifies Management : allows for configuration and distribution of parameters from one node to another
- Responsible for Logical Link Up/Down signalling of Ethernet and Fibre Channel
- DCBX is LLDP with new TLV fields
- The original pre-standard CIN (Cisco, Intel, Nuova) DCBX utilised additional TLV's
- DCBX negotiation failures result in:
  - per-priority-pause not enabled on CoS values
  - vfc not coming up – when DCBX is being used in FCoE environment



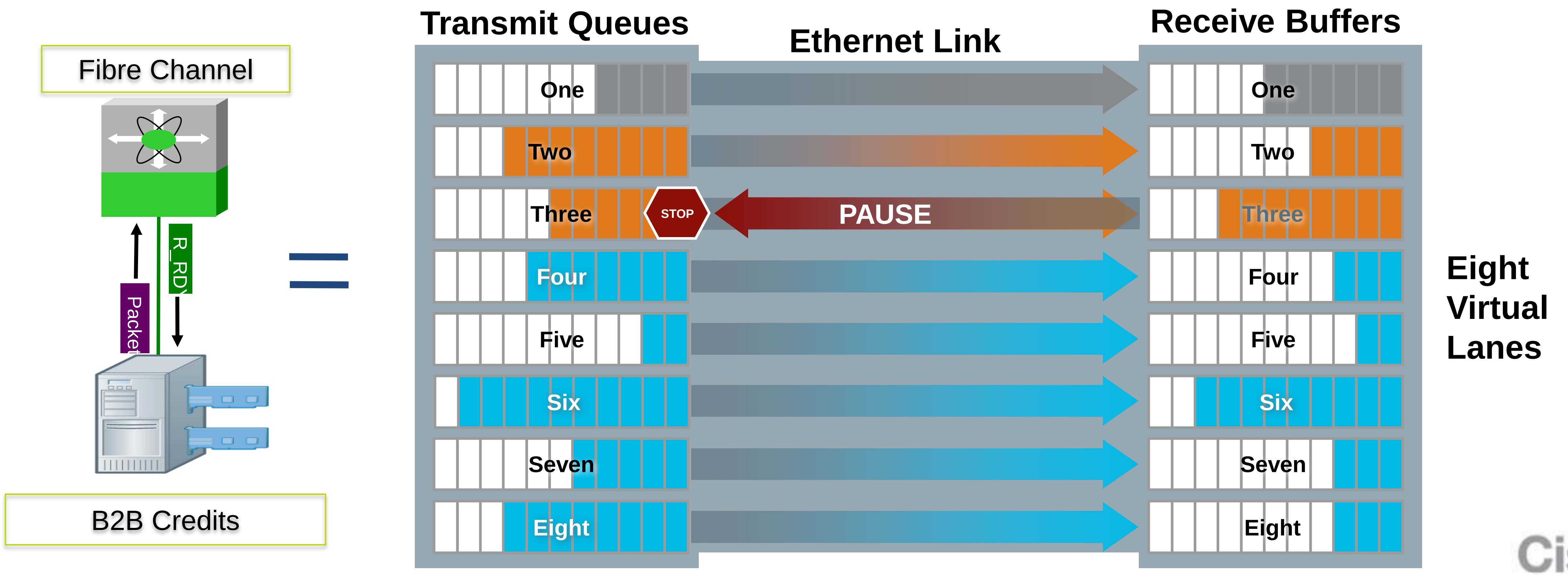

```
dc11-5020-3# sh lldp dcbx interface eth 1/40

Local DCBXP Control information:
Operation version: 00  Max version: 00  Seq no: 7  Ack no: 0
Type/
Subtype    Version    En/Will/Adv Config
006/000     000        Y/N/Y      00
<snip>
```

http://www.cisco.com/en/US/netsol/ns783/index.html

Priority Flow Control

FCoE Flow Control Mechanism – 802.1Qbb

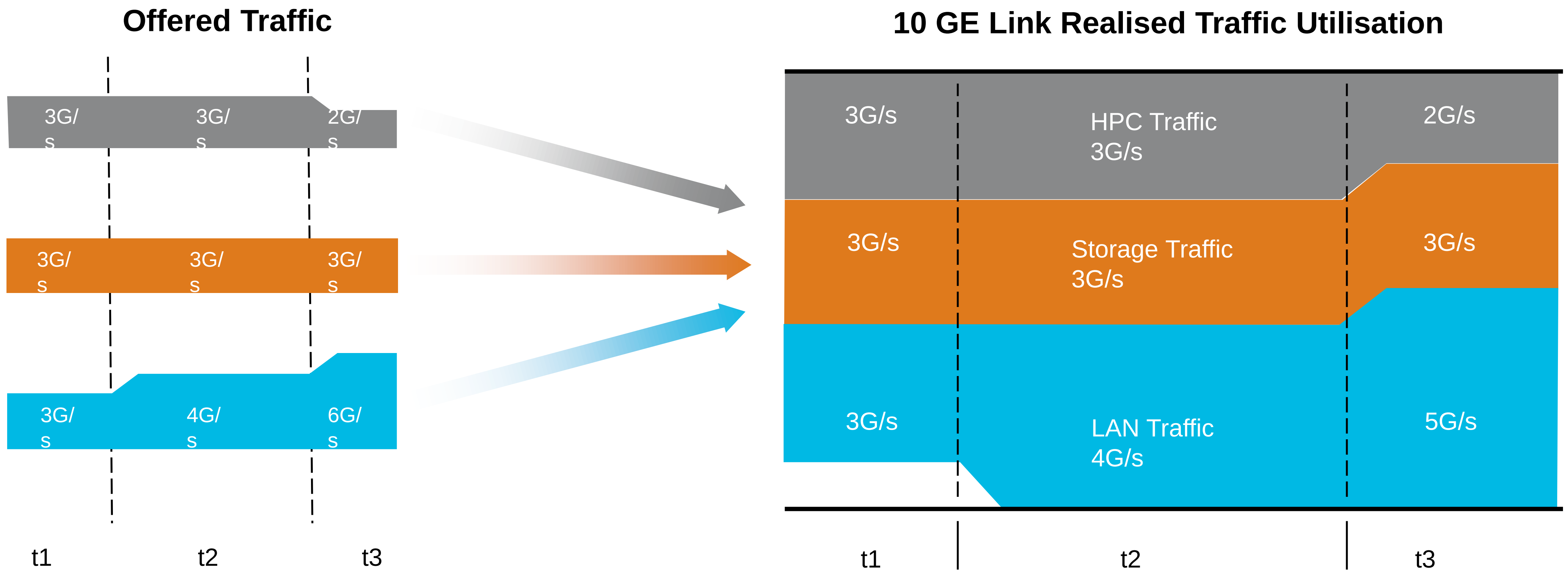
- Enables lossless Ethernet using PAUSE based on a COS as defined in 802.1p
- When link is congested, CoS assigned to “no-drop” will be PAUSED
- Other traffic assigned to other CoS values will continue to transmit and rely on upper layer protocols for retransmission
- Not only for FCoE traffic



Enhanced Transmission Selection (ETS)

Bandwidth Management – 802.1Qaz

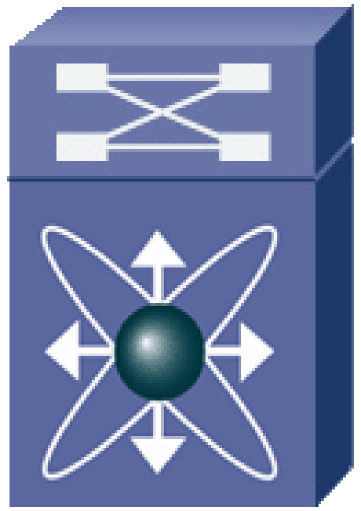
- Prevents a single traffic class of "hogging" all the bandwidth and starving other classes
- When a given load doesn't fully utilise its allocated bandwidth, it is available to other classes
- Helps accommodate for classes of a "bursty" nature



Nexus QoS

QoS Policy Types

- There are three QoS policy types used to define system behaviour (qos, queuing, network-qos)
- There are three policy attachment points to apply these policies to
  - Ingress interface
  - System as a whole (defines global behaviour)
  - Egress interface



Policy Type	Function	Attach Point
qos	**Define traffic classification rules**	**system qos ingress Interface**
queuing	**Strict Priority queue Deficit Weight Round Robin**	**system qos egress Interface ingress Interface**
network-qos	**System class characteristics (drop or no-drop, MTU), Buffer size, Marking**	**system qos**

Configuring QoS on the Nexus 5500/6000

Create New System Class

Step 1 Define qos Class-Map

```
N5k(config)# ip access-list acl-1
N5k(config-acl)# permit ip 100.1.1.0/24 any
N5k(config-acl)# exit
N5k(config)# ip access-list acl-2
N5k(config-acl)# permit ip 200.1.1.0/24 any
N5k(config)# class-map type qos class-1
N5k(config-cmap-qos)# match access-group name acl-1
N5k(config-cmap-qos)# class-map type qos class-2
N5k(config-cmap-qos)# match access-group name acl-2
N5k(config-cmap-qos)#
```

- Create two system classes for traffic with different source address range
- Supported matching criteria

```
N5k(config)# class-map type qos class-1
N5k(config-cmap-qos)# match ?
  access-group        Access group
  cos                 IEEE 802.1Q class of service
  dscp                DSCP in IP(v4) and IPv6 packets
  ip                  IP
  precedence          Precedence in IP(v4) and IPv6 packets
  protocol            Protocol

N5k(config-cmap-qos)# match
```

Step 2 Define qos Policy-Map

```
N5k(config)# policy-map type qos policy-qos
N5k(config-pmap-qos)# class type qos class-1
N5k(config-pmap-c-qos)# set qos-group 2
N5k(config-pmap-c-qos)# class type qos class-2
N5k(config-pmap-c-qos)# set qos-group 3
```

- Qos-group range for user-configured system class is 2-5

Step 3 Apply qos Policy-Map under "system qos" or interface

```
N5k(config)# system qos
N5k(config-sys-qos)# service-policy type qos input policy-qos
```

```
N5k(config)# interface e1/1-10
N5k(config-sys-qos)# service-policy type qos input policy-qos
```

- Policy under *system qos* applied to all interfaces
- Policy under interface is preferred if same type of policy is applied under both *system qos* and interface

Configuring QoS on the Nexus 5500/6000

Create New System Class(Continued)

Step 4 Define network-qos Class-Map

```
N5k(config)# class-map type network-qos class-1
N5k(config-cmap-nq)# match qos-group 2
N5k(config-cmap-nq)# class-map type network-qos class-2
N5k(config-cmap-nq)# match qos-group 3
```

- Match qos-group is the only option for network-qos class-map
- Qos-group value is set by qos policy-map in previous slide

Step 5 Define network-qos Policy-

```
N5k(config)# policy-map type network-qos policy-nq
N5k(config-pmap-nq)# class type network-qos class-1
N5k(config-pmap-nq-c)# class type network-qos class-2
```

- No action tied to this class indicates default network-qos parameters.
- Policy-map type *network-qos* will be used to configure no-drop class, MTU, ingress buffer size and 802.1p marking
- Default network-qos parameters are listed in the table below

Step 6 Apply network-qos policy-map under *system qos* context

```
N5k(config-pmap-nq-c)# system qos
N5k(config-sys-qos)# service-policy type network-qos policy-nq
N5k(config-sys-qos)#
```

Network-QoS Parameters	Default Value
Class Type	Drop class
MTU	1538
Ingress Buffer Size	20.4KB
Marking	No marking

Configuring QoS on the Nexus 5500/6000

Strict Priority and Bandwidth Sharing

- Create new system class by using policy-map *qos* and *network-qos*(Previous two slides)
- Then Define and apply policy-map type *queuing* to configure strict priority and bandwidth sharing
- Checking the queuing or bandwidth allocating with command *show queuing interface*

```
N5k(config)# class-map type queuing class-1
N5k(config-cmap-que)# match qos-group 2
N5k(config-cmap-que)# class-map type queuing class-2
N5k(config-cmap-que)# match qos-group 3
N5k(config-cmap-que)# exit

N5k(config)# policy-map type queuing policy-BW
N5k(config-pmap-que)# class type queuing class-1
N5k(config-pmap-c-que)# priority
N5k(config-pmap-c-que)# class type queuing class-2
N5k(config-pmap-c-que)# bandwidth percent  40
N5k(config-pmap-c-que)# class type queuing class-fcoe
N5k(config-pmap-c-que)# bandwidth percent 40
N5k(config-pmap-c-que)# class type queuing class-default
N5k(config-pmap-c-que)# bandwidth percent 20

N5k(config-pmap-c-que)#  system qos
N5k(config-sys-qos)# service-policy type queuing output policy-BW
N5k(config-sys-qos)#
```

← Define queuing class-map

← Define queuing policy-map

← Apply queuing policy under *system qos* or egress interface

Configuring QoS on the Nexus 5500/6000

Check System Classes

```
N5k# show queuing interface ethernet 1/1

Interface Ethernet1/1 TX Queuing
qos-group  sched-type  oper-bandwidth
 0     WRR         20
 1     WRR         40
 2     priority       0
 3     WRR         40

Interface Ethernet1/1 RX Queuing
qos-group  0:
   q-size: 163840, MTU: 1538
   drop-type: drop, xon: 0, xoff: 1024
   Statistics:
      Pkts received over the port             : 9802
      Ucast pkts sent to the cross-bar        : 0
      Mcast pkts sent to the cross-bar        : 9802
      Ucast pkts received from the cross-bar  : 0
      Pkts sent to the port                   : 18558
      Pkts discarded on ingress               : 0
      Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

qos-group  1:
   q-size: 76800, MTU: 2240
   drop-type: no-drop, xon: 128, xoff: 240
   Statistics:
      Pkts received over the port             : 0
      Ucast pkts sent to the cross-bar        : 0
      Mcast pkts sent to the cross-bar        : 0
      Ucast pkts received from the cross-bar  : 0
      Pkts sent to the port                   : 0
      Pkts discarded on ingress               : 0
      Per-priority-pause status               : Rx (Inactive), Tx (Inactive)
Continue…
```

- Strict priority and WRR configuration
- qos-group 0: class-default
- Packet counter for each class
- Drop counter for each class
- qos-group 1: *class-fcoe*
- Current PFC status

```
qos-group  2:
   q-size: 20480, MTU: 1538
   drop-type: drop, xon: 0, xoff: 128
   Statistics:
      Pkts received over the port             : 0
      Ucast pkts sent to the cross-bar        : 0
      Mcast pkts sent to the cross-bar        : 0
      Ucast pkts received from the cross-bar  : 0
      Pkts sent to the port                   : 0
      Pkts discarded on ingress               : 0
      Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

qos-group  3:
   q-size: 20480, MTU: 1538
   drop-type: drop, xon: 0, xoff: 128
   Statistics:
      Pkts received over the port             : 0
      Ucast pkts sent to the cross-bar        : 0
      Mcast pkts sent to the cross-bar        : 0
      Ucast pkts received from the cross-bar  : 0
      Pkts sent to the port                   : 0
      Pkts discarded on ingress               : 0
      Per-priority-pause status               : Rx (Inactive), Tx (Inactive)

Total Multicast crossbar statistics:
   Mcast pkts received from the cross-bar     : 18558
N5k#
```

- qos-group 2: User-configured system class: *class-1*
- qos-group 3: User-configured system class: *class-2*

Priority Flow Control – Nexus 5000/5500/6000

Operations Configuration – Switch Level

- On Nexus 5000 once **feature fcoe** is configured, 2 classes are made **by default**

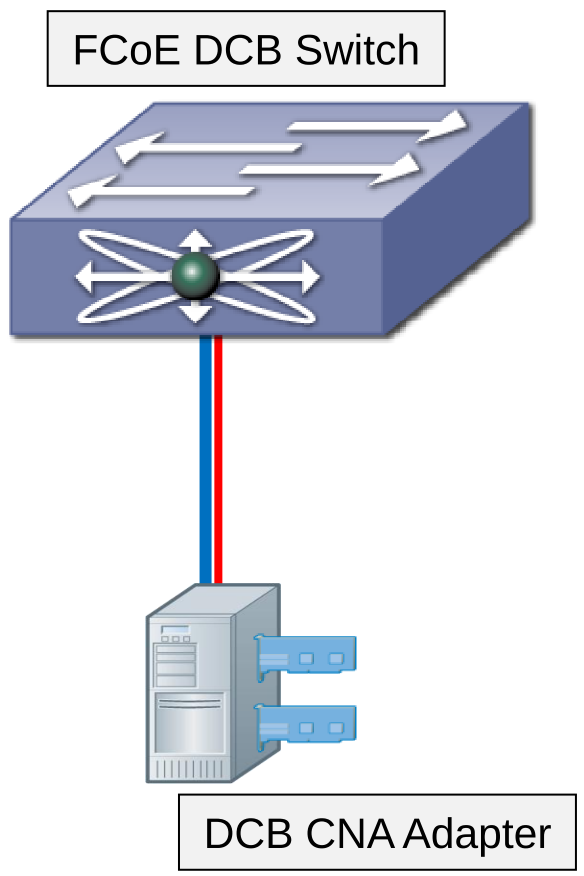
```
policy-map type qos default-in-policy
   class type qos class-fcoe
     set qos-group 1
   class type qos class-default
     set qos-group 0
```

- **class-fcoe** is configured to be **no-drop** with an MTU of 2158

```
policy-map type network-qos default-nq-policy
   class type network-qos class-fcoe
     pause no-drop
     mtu 2158
```

- Enabling the FCoE feature on Nexus 5548/96 does '***not'*** create no-drop policies automatically as on Nexus 5010/20
- Must add policies under system QOS:

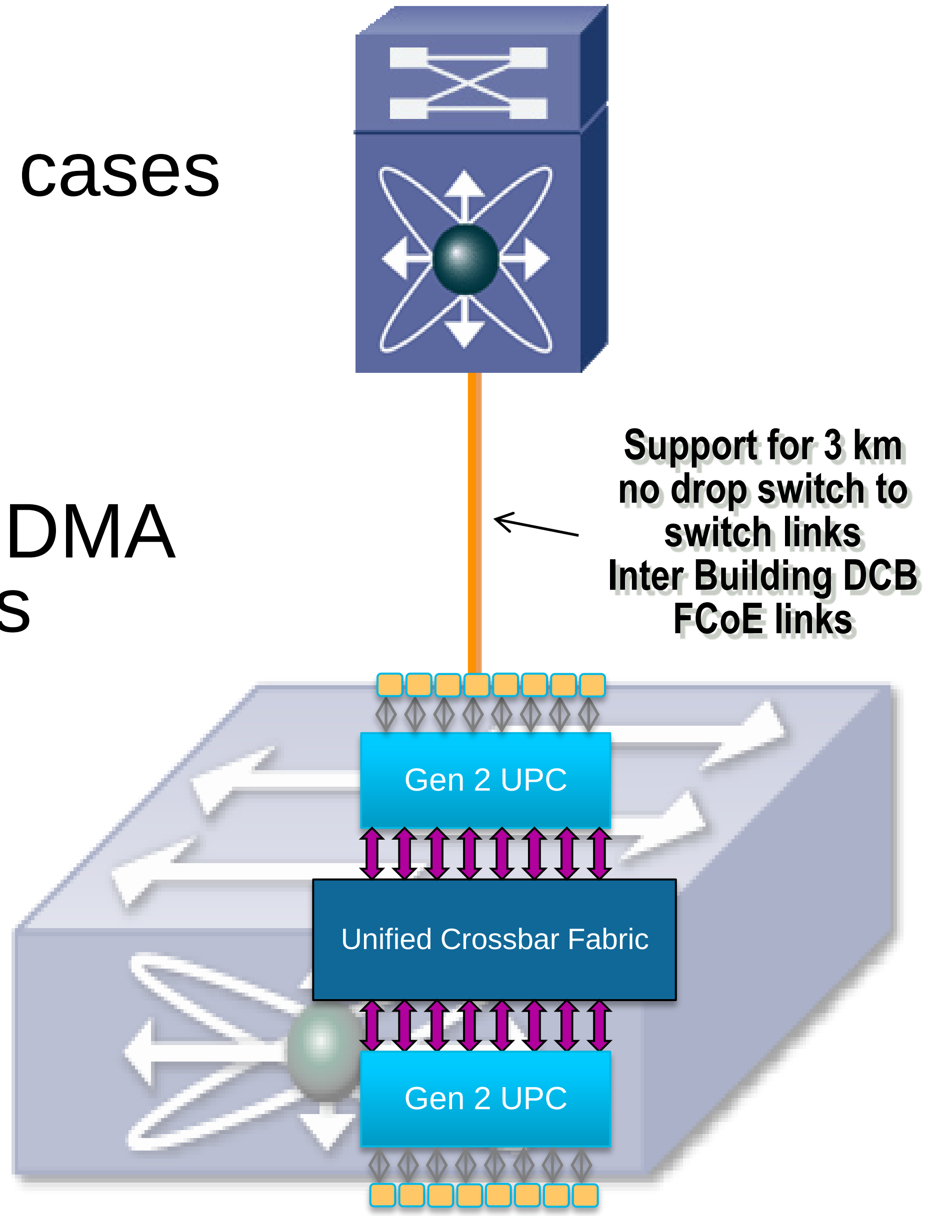
```
system qos
  service-policy type qos input fcoe-default-in-policy
  service-policy type queuing input fcoe-default-in-policy
  service-policy type queuing output fcoe-default-out-policy
  service-policy type network-qos fcoe-default-nq-policy
```



Nexus 5000/5500/6000 QoS

Priority Flow Control and No-Drop Queues

- Tuning of the lossless queues to support a variety of use cases
- Extended switch to switch no drop traffic lanes
  - Support for 3km for Nexus 5000/5500/6000
  - Increased number of no drop services lanes (4) for RDMA and other multi-queue HPC and compute applications



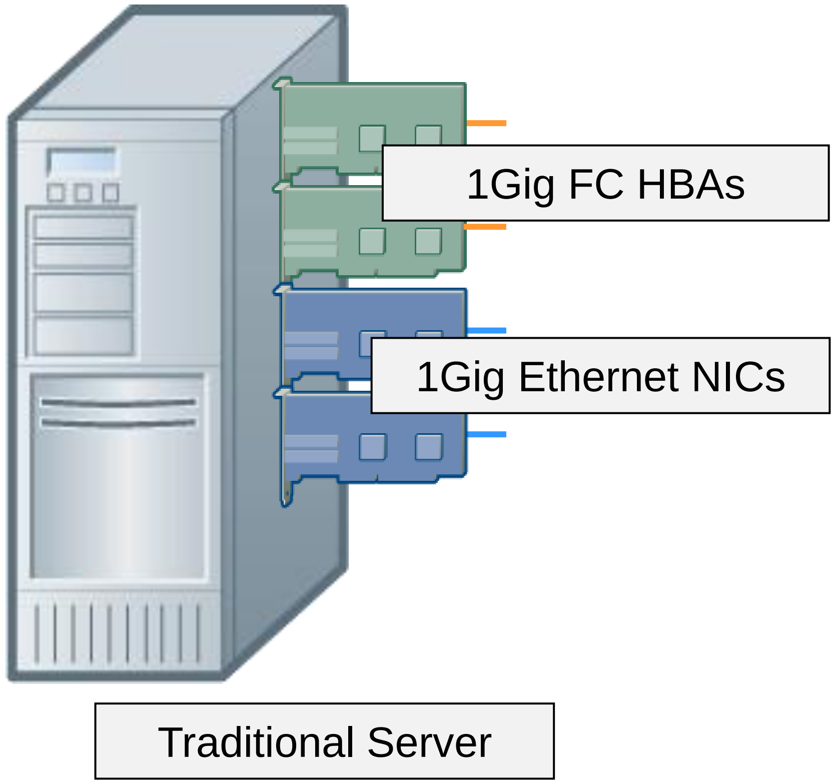
Configs for 3000m no-drop class	Buffer size	Pause Threshold (XOFF)	Resume Threshold (XON)
N5020	143680 bytes	58860 bytes	38400 bytes
N5548	152000 bytes	103360 bytes	83520 bytes
N600X	152000 bytes	103360 bytes	83520 bytes

```
5548-FCoE(config)# policy-map type network-qos 3km-FCoE
5548-FCoE(config-pmap-nq)# class type network-qos 3km-FCoE
5548-FCoE(config-pmap-nq-c)# pause no-drop buffer-size 152000 pause-threshold 103360
resume-threshold 83520
```

Enhanced Transmission Selection - N5K

Bandwidth Management

- When configuring FCoE by default, each class is given **50%** of the available bandwidth
- Can be changed through QoS settings when higher demands for certain traffic exist (i.e. HPC traffic, more Ethernet NICs)




```
N5k-1# show queuing interface ethernet 1/18
Ethernet1/18 queuing information:
  TX Queuing
    qos-group  sched-type  oper-bandwidth
        0       WRR             50
        1       WRR             50
```

- Best Practice: Tune FCoE queue to provide equivalent capacity to the HBA that would have been used (1G, 2G, …)

Priority Flow Control – Nexus 7K & MDS

Operations Configuration – Switch Level

N7K-50(config)# *system qos*
N7K-50(config-sys-qos)# *service-policy type network-qos default-nq-7e-policy*

- **No-Drop PFC w/ MTU 2K set for Fibre Channel**

```
show policy-map system
  Type network-qos policy-maps
  =====================================
  policy-map type network-qos default-nq-7e-policy
    class type network-qos c-nq-7e-drop
      match cos 0-2,4-7
      congestion-control tail-drop
      mtu 1500
    class type network-qos c-nq-7e-ndrop-fcoe
      match cos 3
      match protocol fcoe
      pause
      mtu 2112
```

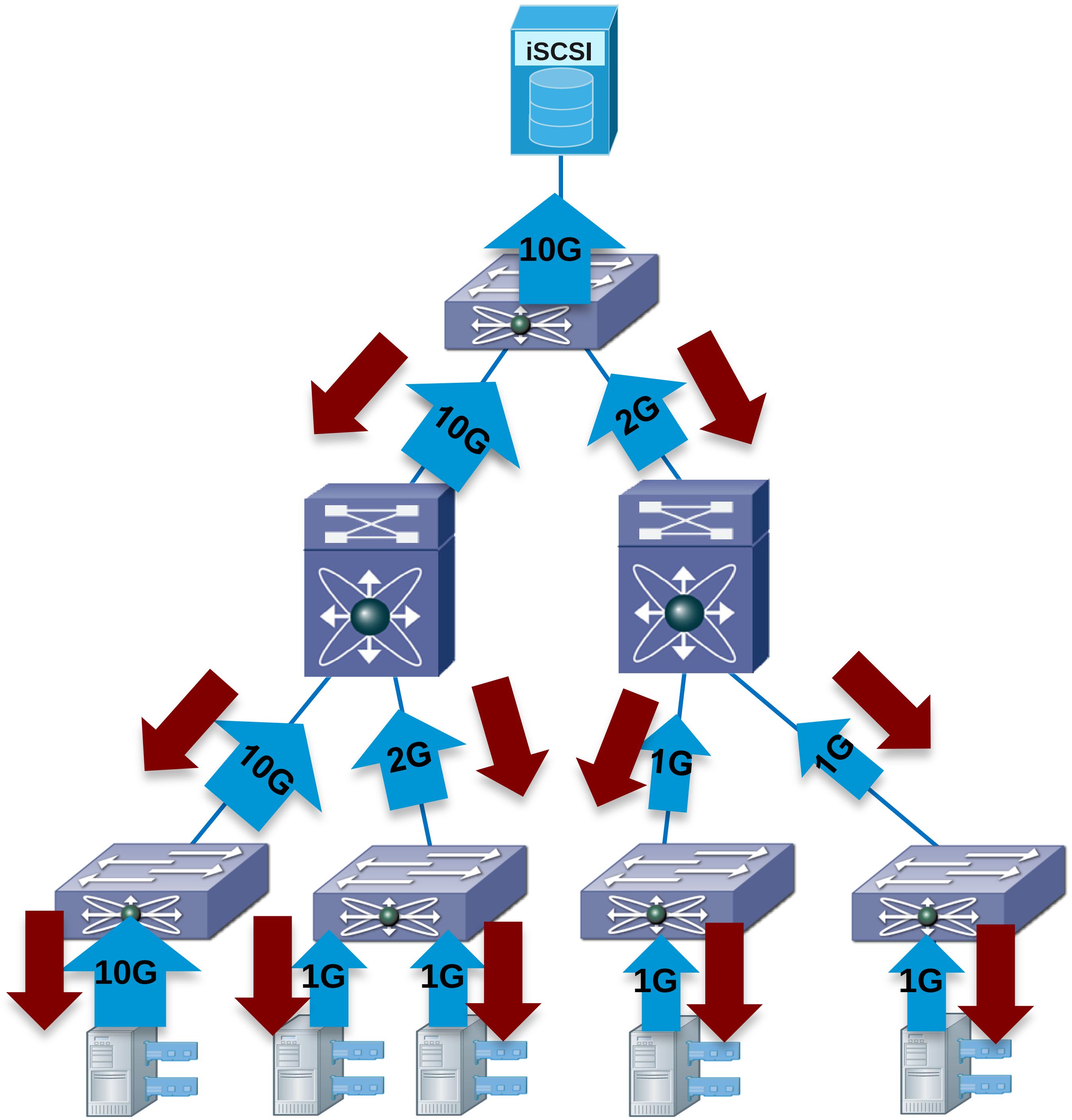
```
show class-map type network-qos c-nq-7e-ndrop-fcoe

  Type network-qos class-maps
  =============================================
  class-map type network-qos match-any c-nq-7e-ndrop-fcoe
    Description: 7E No-Drop FCoE CoS map
    match cos 3
    match protocol fcoe
```

Policy Template choices

Template	Drop CoS	(Priority)	NoDrop CoS	(Priority)
default-nq-8e-policy	0,1,2,3,4,5,6,7	5,6,7	-	-
default-nq-7e-policy	0,1,2,4,5,6,7	5,6,7	3	-
default-nq-6e-policy	0,1,2,5,6,7	5,6,7	3,4	4
default-nq-4e-policy	0,5,6,7	5,6,7	1,2,3,4	4

DC Design Details

No Drop Storage Considerations



1. Steady state traffic is within end to end network capacity
2. Burst traffic from a source
3. ‘No Drop’ traffic is queued
4. Buffers begin to fill and PFC flow control initiated
5. All sources are eventually flow controlled
   - TCP not invoked immediately as frames are queued not dropped
   - Is the optimal behaviour for your oversubscription?

DC Design Details

HOLB is also a fundamental part of Fibre Channel SAN design

- Blocking - Impact on Design Performance
- Performance can be adversely affected across an entire multiswitch FC Fabric by a single blocking port
  - HOL is a transitory event (until some BB_Credits are returned on the blocked port)
- To help alleviate the blocking problem and enhance the design performance
  - Virtual Output Queuing (VoQ) on all ports

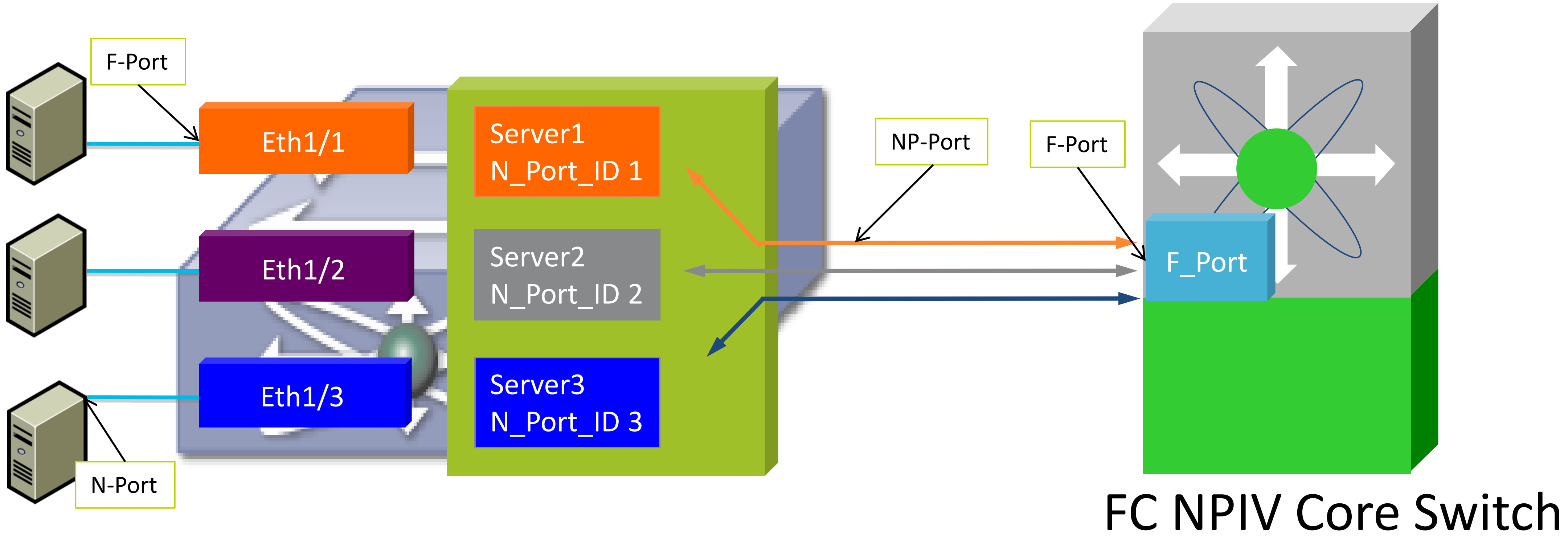
Agenda

- Unified Fabric – What and Why
- FCoE Protocol Fundamentals
- Nexus FCoE Capabilities
- FCoE Network Requirements and Design Considerations
- DCB & QoS - Ethernet Enhancements
- Single Hop Design
- Multi-Hop Design
- Futures

FCoE Edge

N-Port Virtualiser (NPV)

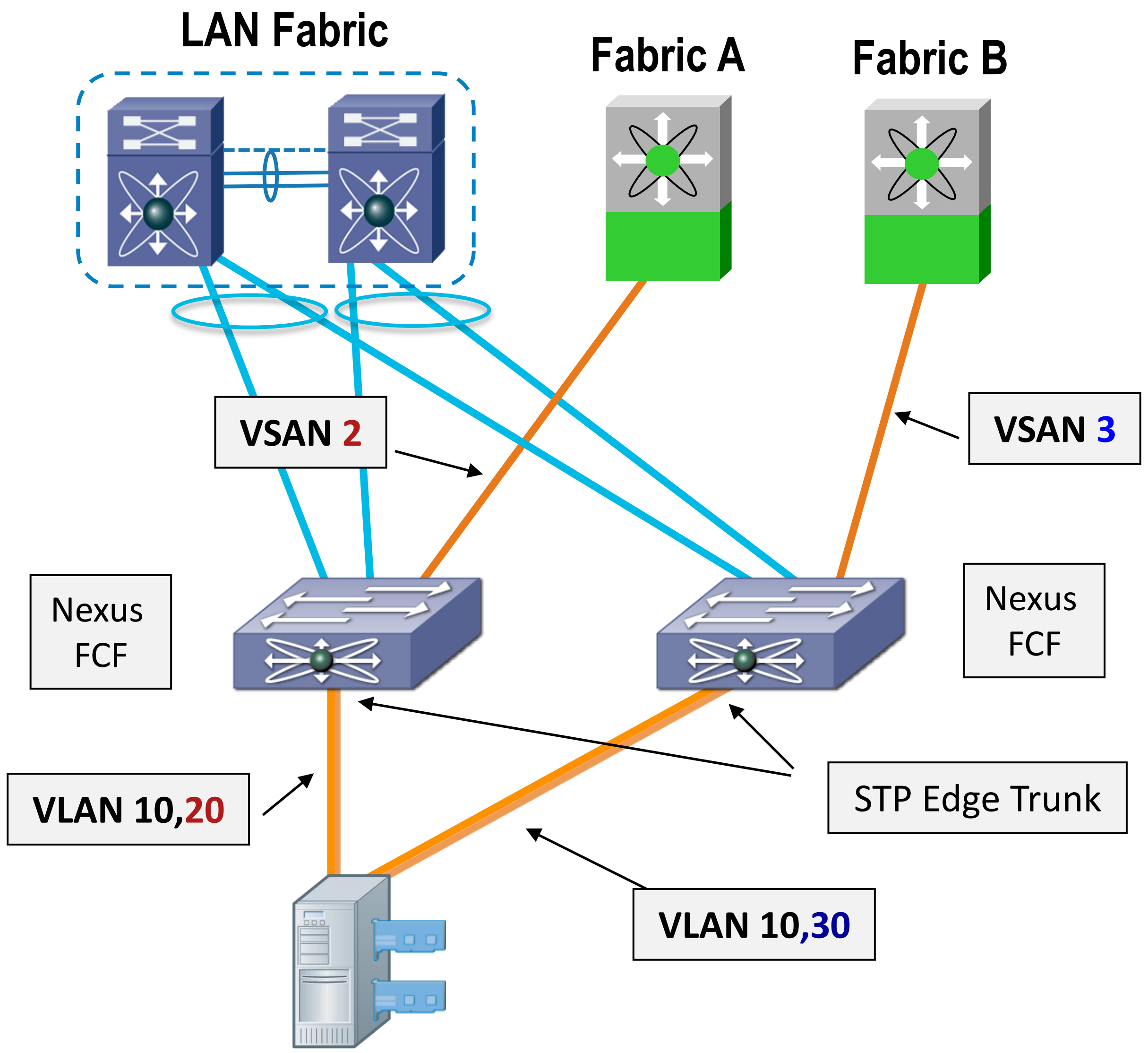
- N-Port Virtualiser (NPV) utilises NPIV functionality to allow a “switch” to act like a server performing multiple logins through a single physical link
- Physical servers connected to the NPV switch login to the upstream NPIV core switch
  - Physical uplink from NPV switch to FC NPIV core switch does actual “FLOGI”
  - Subsequent logins are converted (proxy) to “FDISC” to login to upstream FC switch
- No local switching is done on an FC switch in NPV mode
- FC edge switch in NPV mode does not take up a domain ID



Unified Fabric Design

The FCoE VLAN

- Each FCoE VLAN and VSAN count as a VLAN HW resource – therefore a VLAN/VSAN mapping accounts for TWO VLAN resources
- FCoE VLANs are treated differently than native Ethernet VLANs: no flooding, broadcast, MAC learning, etc.
- *BEST PRACTICE*: use different FCoE VLANs/VSANs for SAN A and SAN B
- The FCoE VLAN must not be configured as a native VLAN
- Shared Wires connecting to HOSTS must be configured as trunk ports and STP edge ports
- *Note:* STP does not run on FCoE vlans between FCFs (VE_Ports) but does run on FCoE VLANs towards the host (VF_Ports)

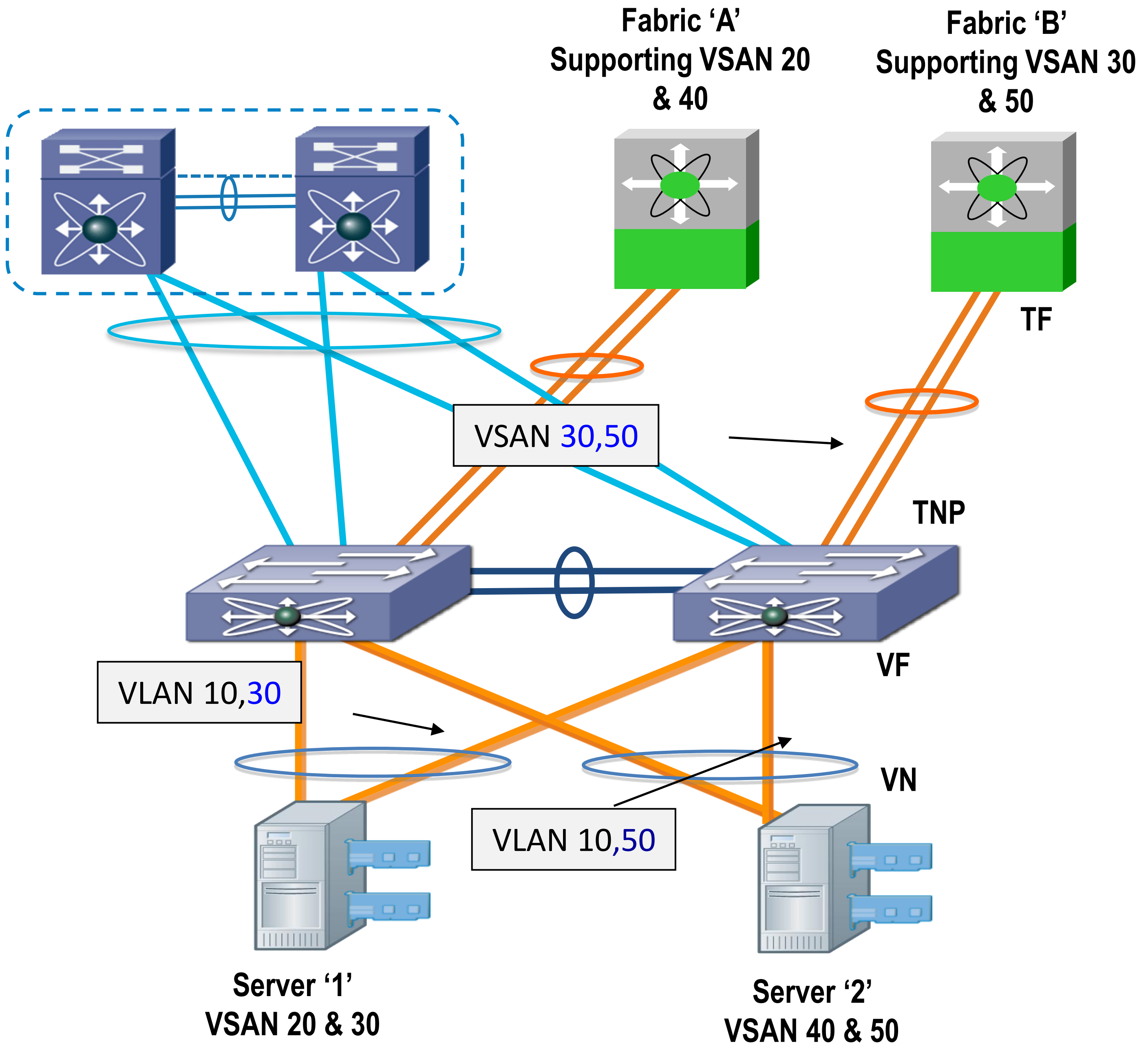



```
! VLAN 20 is dedicated for VSAN 2 FCoE traffic
(config)#  vlan 20
(config-vlan)# fcoe vsan 2
```

Unified Fabric Design

F_Port Trunking and Channelling

- Nexus 5000/5500/6000 supports F-Port Trunking and Channelling
- VSAN Trunking and Port-Channel on the links between an NPV device and upstream FC switch (NP port -> F port)
- F_Port Trunking: Better multiplexing of traffic using shared links (multiple VSANs on a common link)
- F_Port Channelling: Better resiliency between NPV edge and Director Core (avoids tearing down all FLOGIs on a failing link)
- Simplifies FC topology (single uplink from NPV device to FC director)

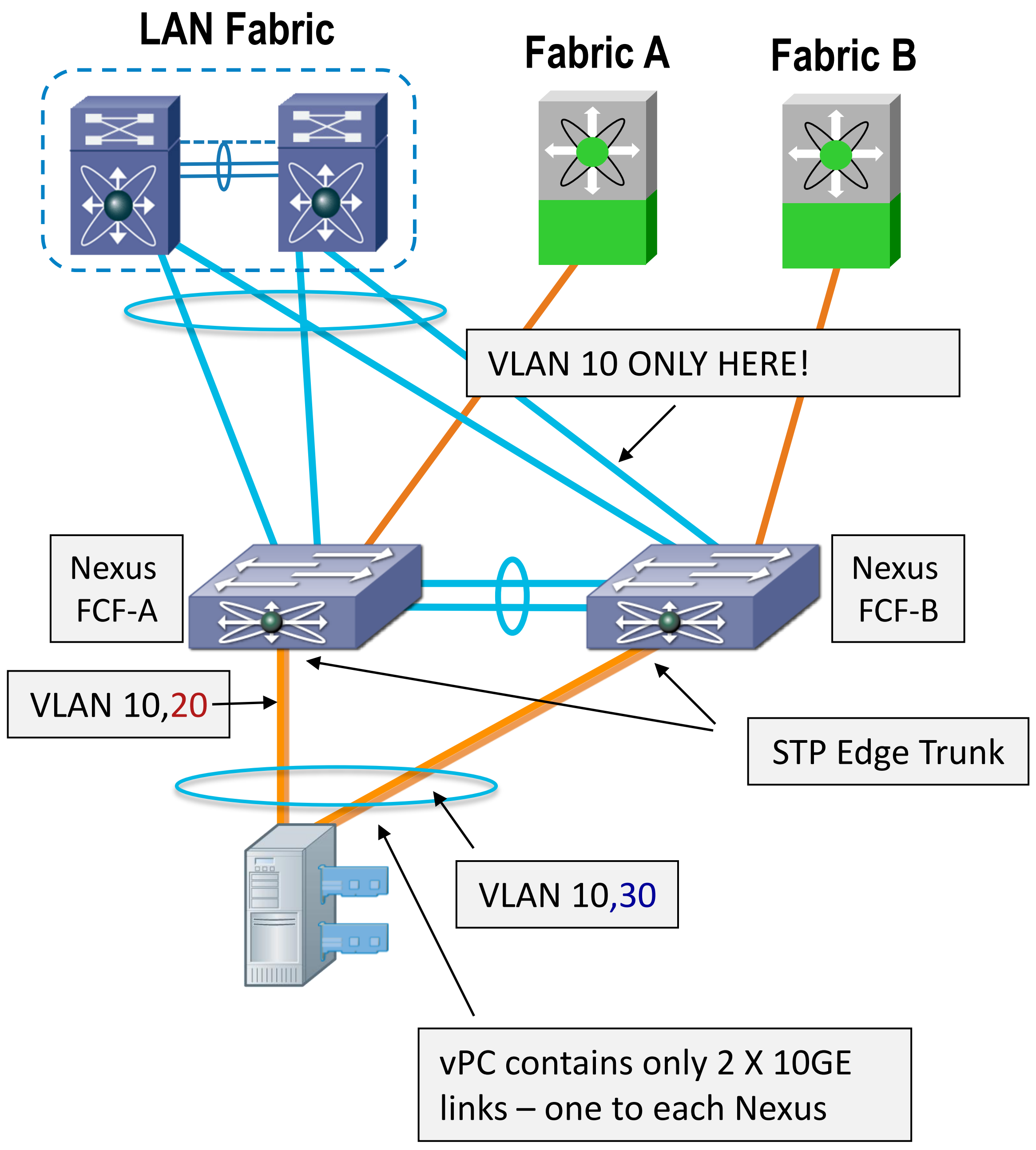


F Port Trunking & Channelling

Unified Fabric Design

FCoE and vPC together

- vPC with FCoE are ONLY supported between hosts and N5k or N5k/2232 pairs…AND they must follow specific rules
  - A ‘vfc’ interface can only be associated with a single-port port-channel
  - While the port-channel configurations are the same on N5K-1 and N5K-2, the FCoE VLANs are different
- FCoE VLANs are ‘not’ carried on the vPC peer-link (automatically pruned)
  - FCoE and FIP ethertypes are ‘not’ forwarded over the vPC peer link either
- vPC carrying FCoE between two FCF’s is NOT supported
- *Best Practice:* Use static port channel configuration rather than LACP with vPC and Boot from SAN (this will change with future releases)

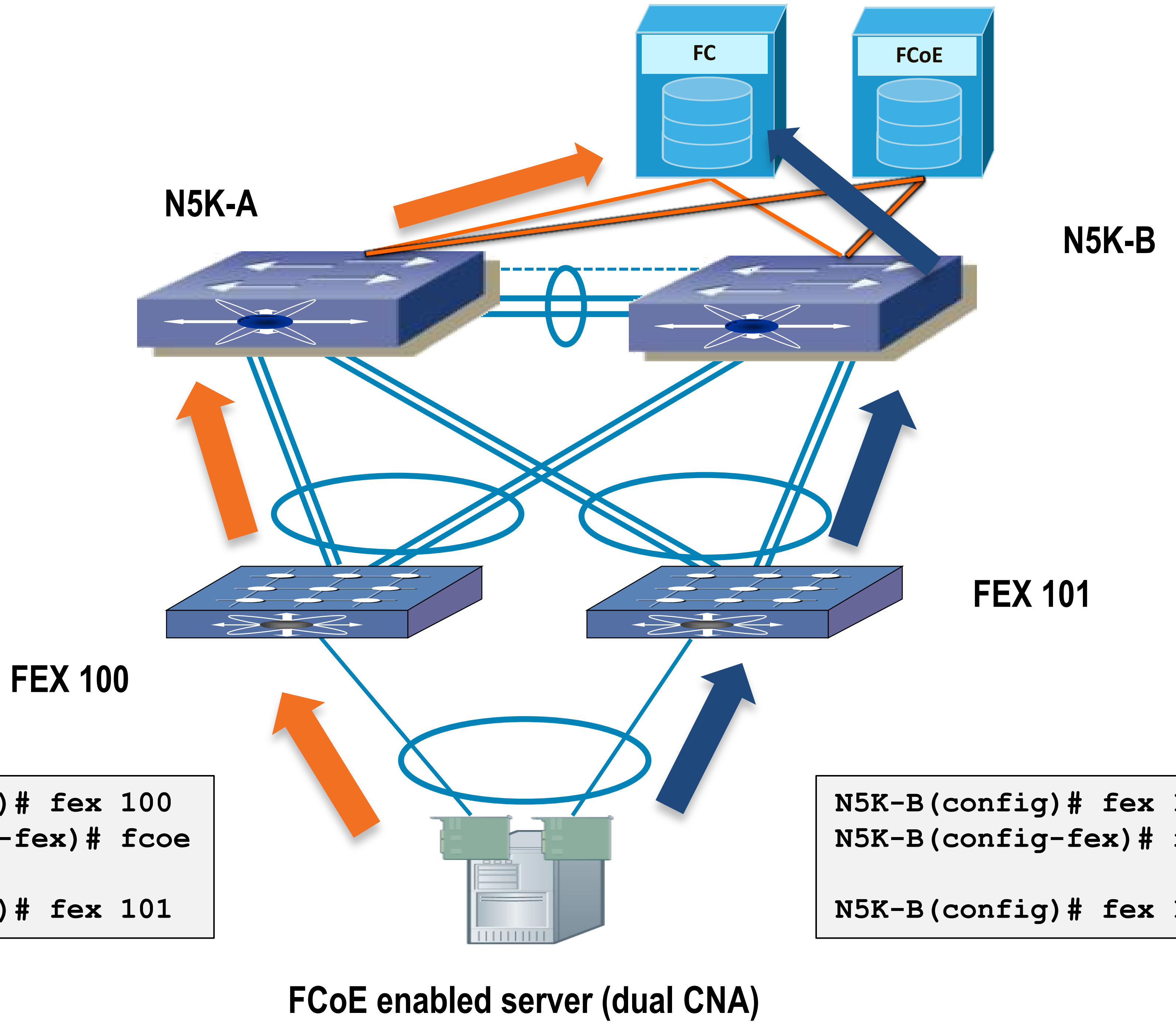


Direct Attach vPC Topology

EvPC & FEX

Nexus 5550 Topologies starting with NX-OS 5.1(3)N1

- In an Enhanced vPC (EvPC) SAN ‘A/B’ isolation is configured by associating each FEX with either SAN ‘A’ or SAN ‘B’ Nexus 5500
- FCoE & FIP traffic is forwarded only over the links connected to the specific parent switch
- Ethernet is hashed over ‘all’ FEX fabric links
- Nexus 6000 only supports FCoE, no native FC at FCS

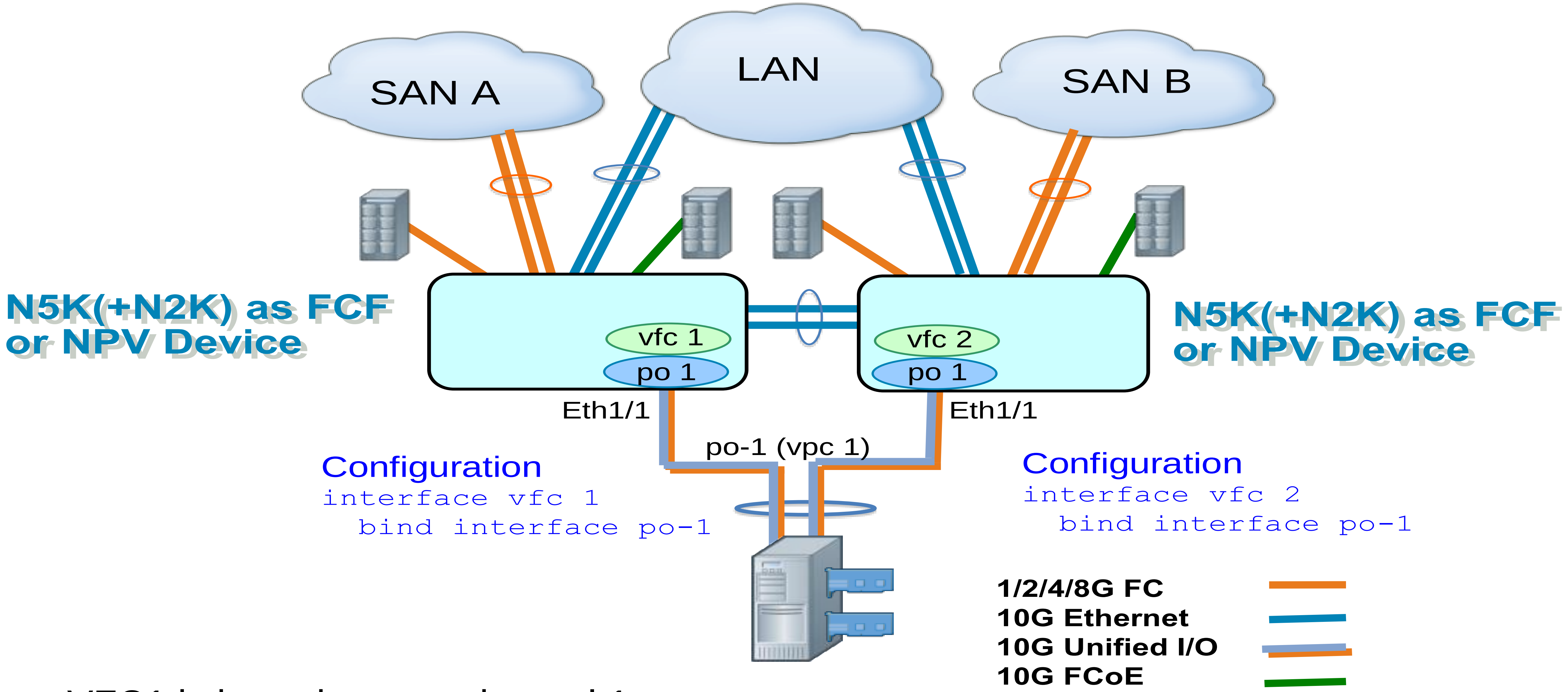



```
N5K-A(config)# fex 100
N5K-A(config-fex)# fcoe

N5K-A(config)# fex 101
```

```
N5K-B(config)# fex 101
N5K-B(config-fex)# fcoe

N5K-B(config)# fex 100
```

vPC & Boot from SAN

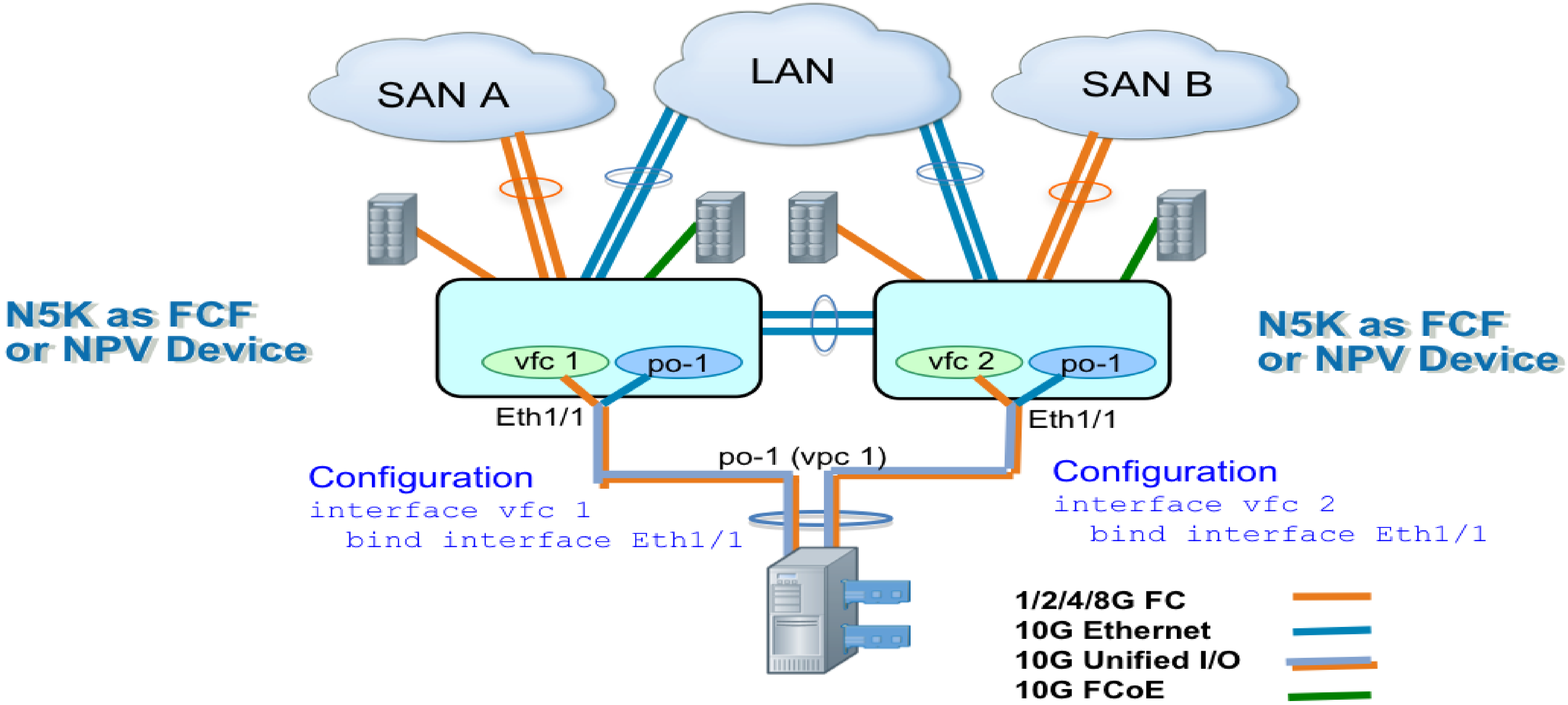
Pre 5.1(3)N1 Behaviour



- VFC1 is bound to port-channel 1
- Port-channel 1 is using LACP to negotiate with host
- The VFC/port-channel never comes up and the host isn't able to boot from SAN

vPC & Boot from SAN

5.1(3)N1 and onwards Behaviour

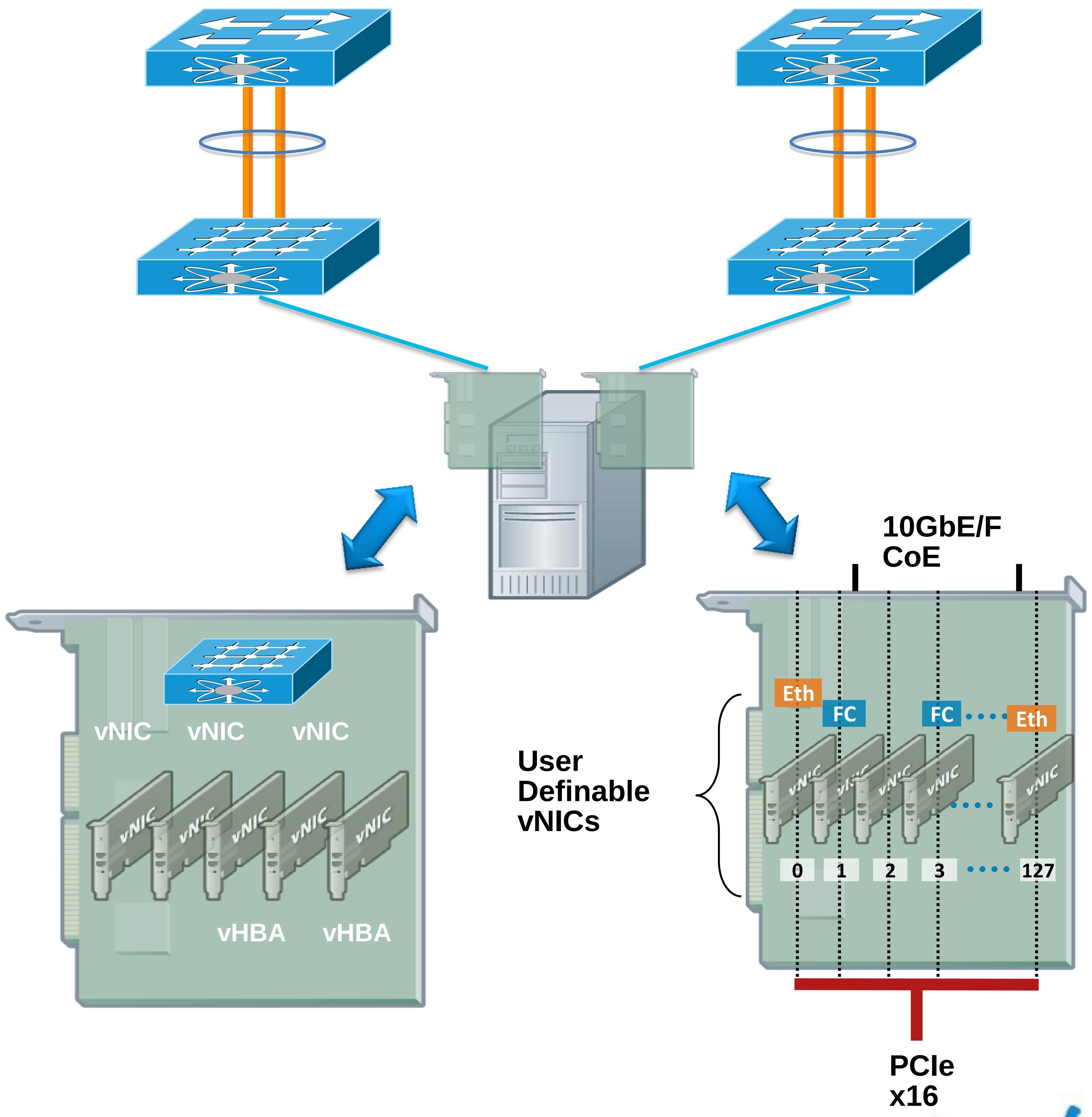


- As of NX-OS Release 5.1(3)N1(1) for N5K, new VFC binding models will be supported
- In this case, we now support VF_Port binding to a member port of a given port-channel
- Check the configuration guide and operations guide for additional VFC binding changes

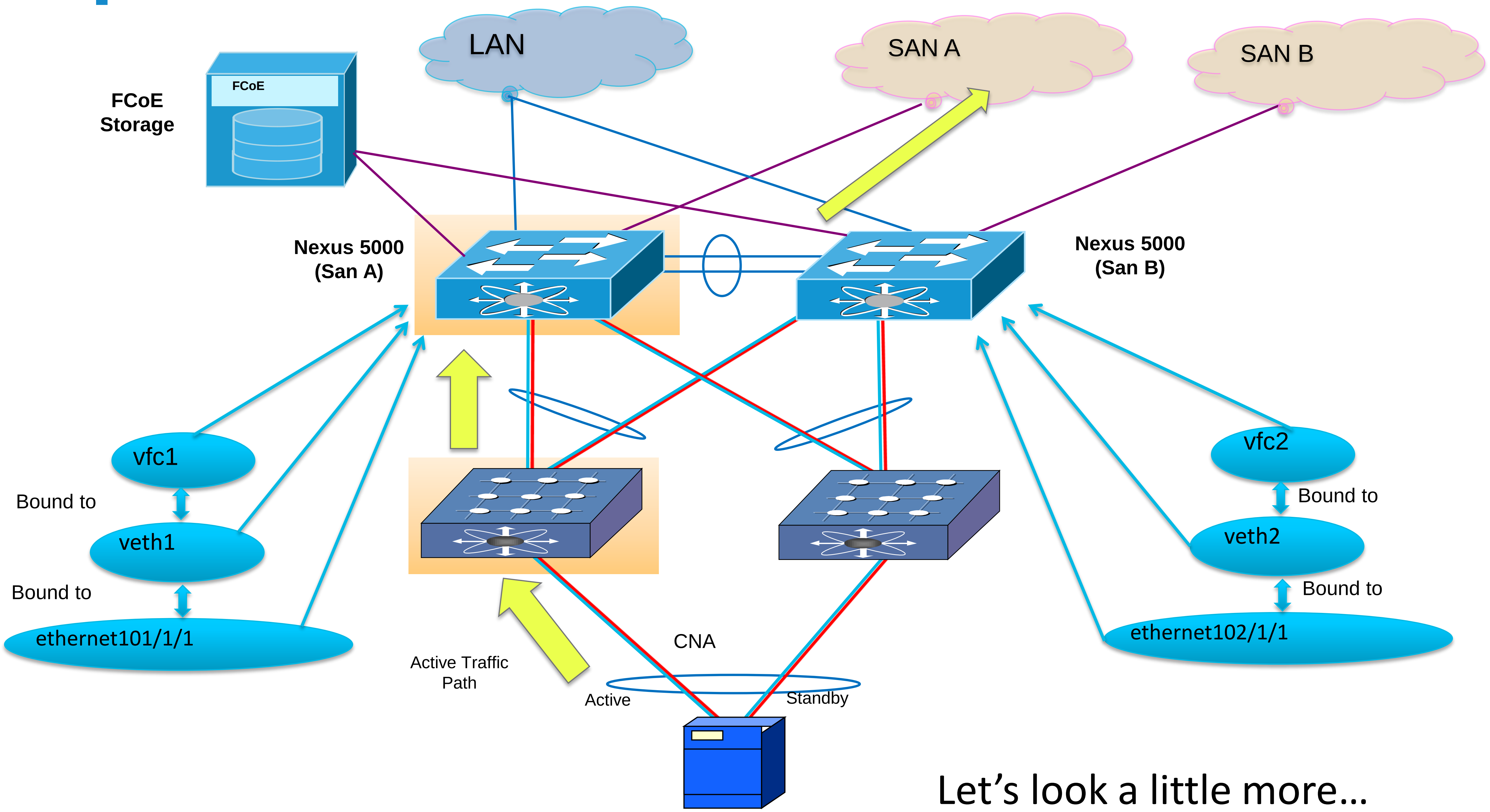
Adapter FEX

802.1BR

- Adapter-FEX presents standard PCIe virtual NICs (vNICs) to servers
- Adapter-FEX virtual NICs are configured and managed via NX-OS
- Forwarding, Queuing, and Policy enforcement for vNIC traffic by Nexus
- Adapter-FEX connected to Nexus 2000 Fabric Extender - Cascaded FEX-Link deployment
- Forwarding, Queuing, and Policy enforcement for vNIC traffic still done by Nexus



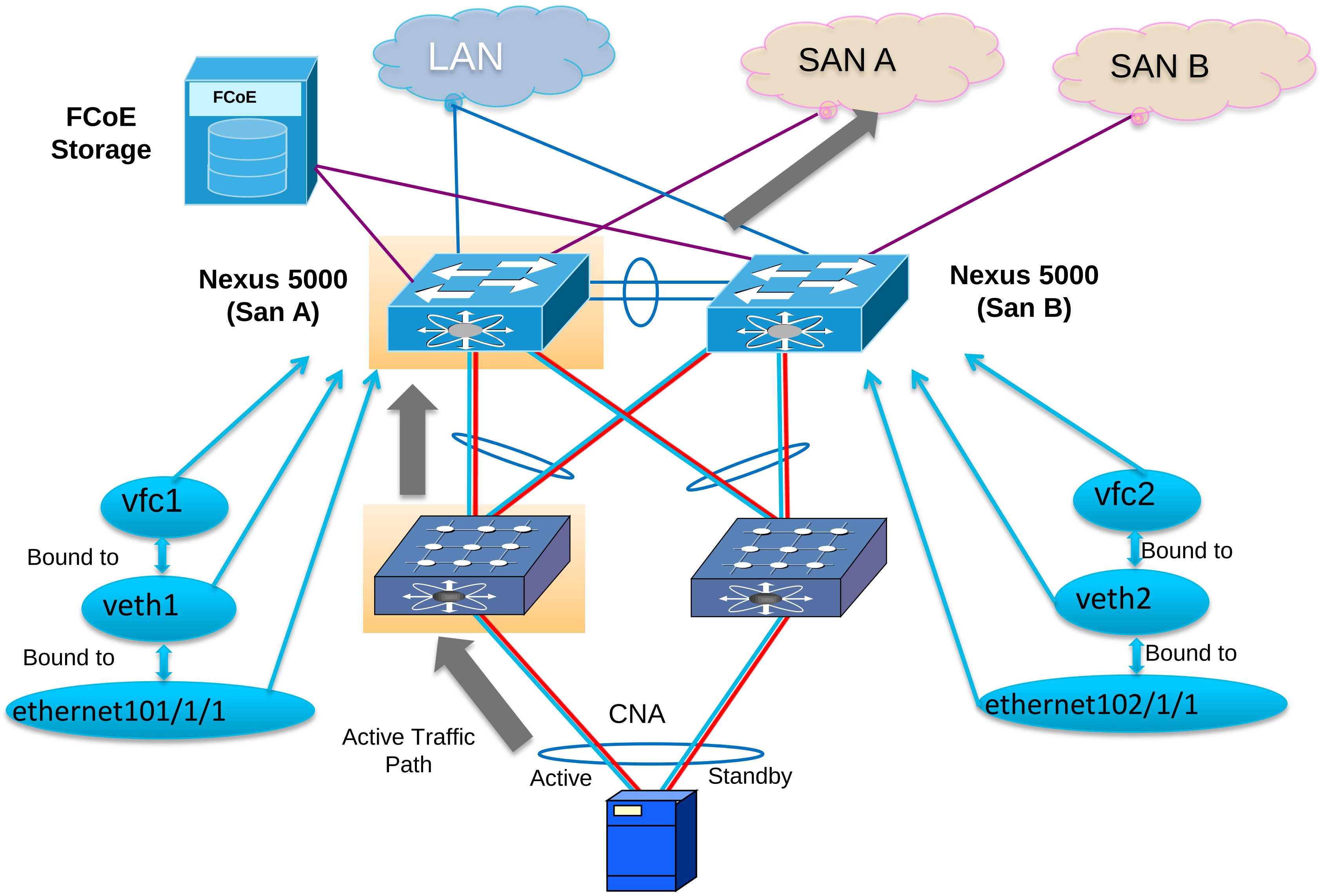
Adapter FEX & FCoE



Let's look a little more…

Adapter FEX & FCoE

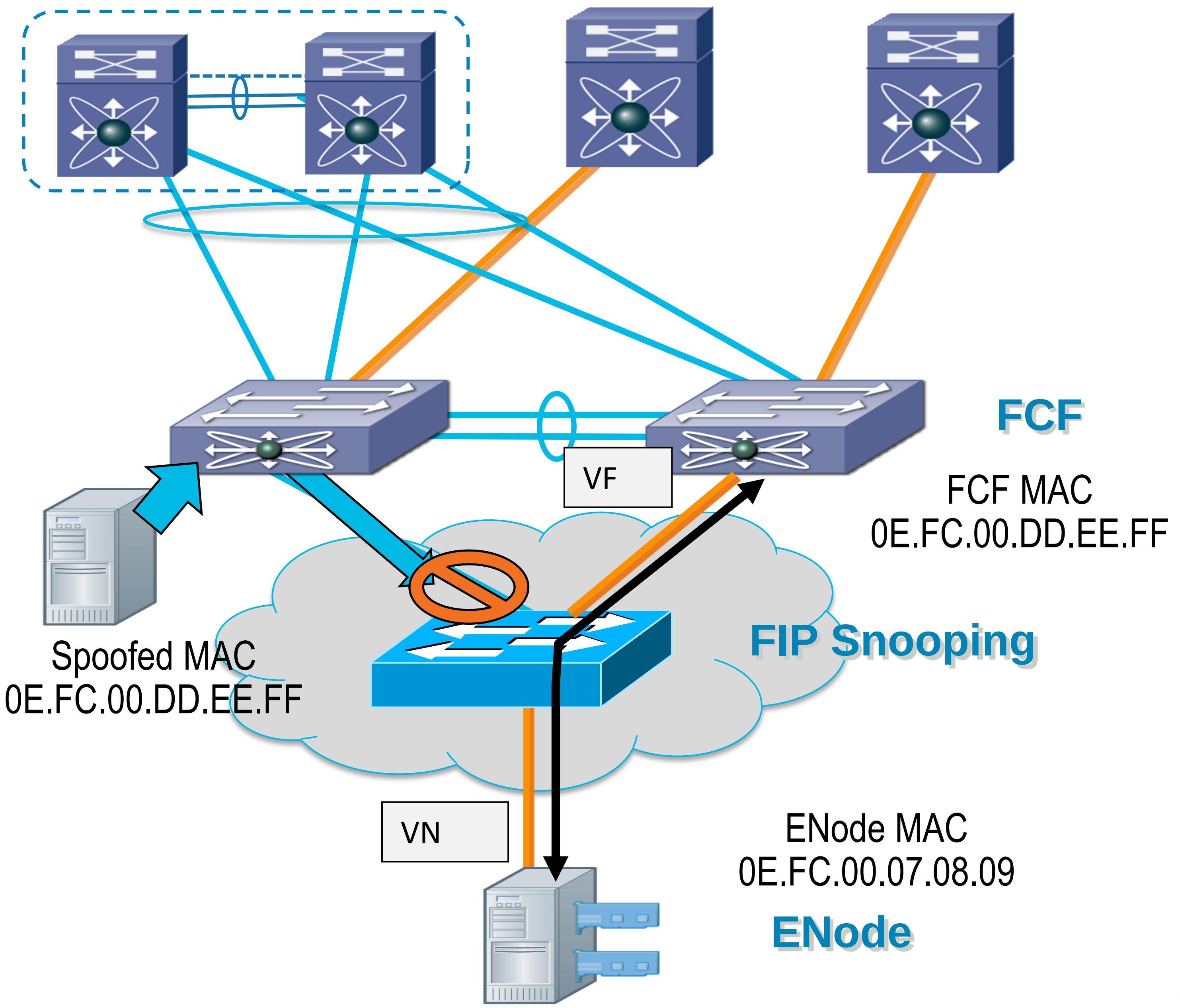
- interface vfc1
  - bind interface veth1
- interface veth1
  - bind interface eth101/1/1 channel 1
- interface eth101/1/1
  - switchport mode vntag
- And in this case, to make sure we don’t break SAN A/B separation, make sure we configure the FEX2232:
  - fex 101
  - fcoe



Transparent Bridges?

FIP Snooping

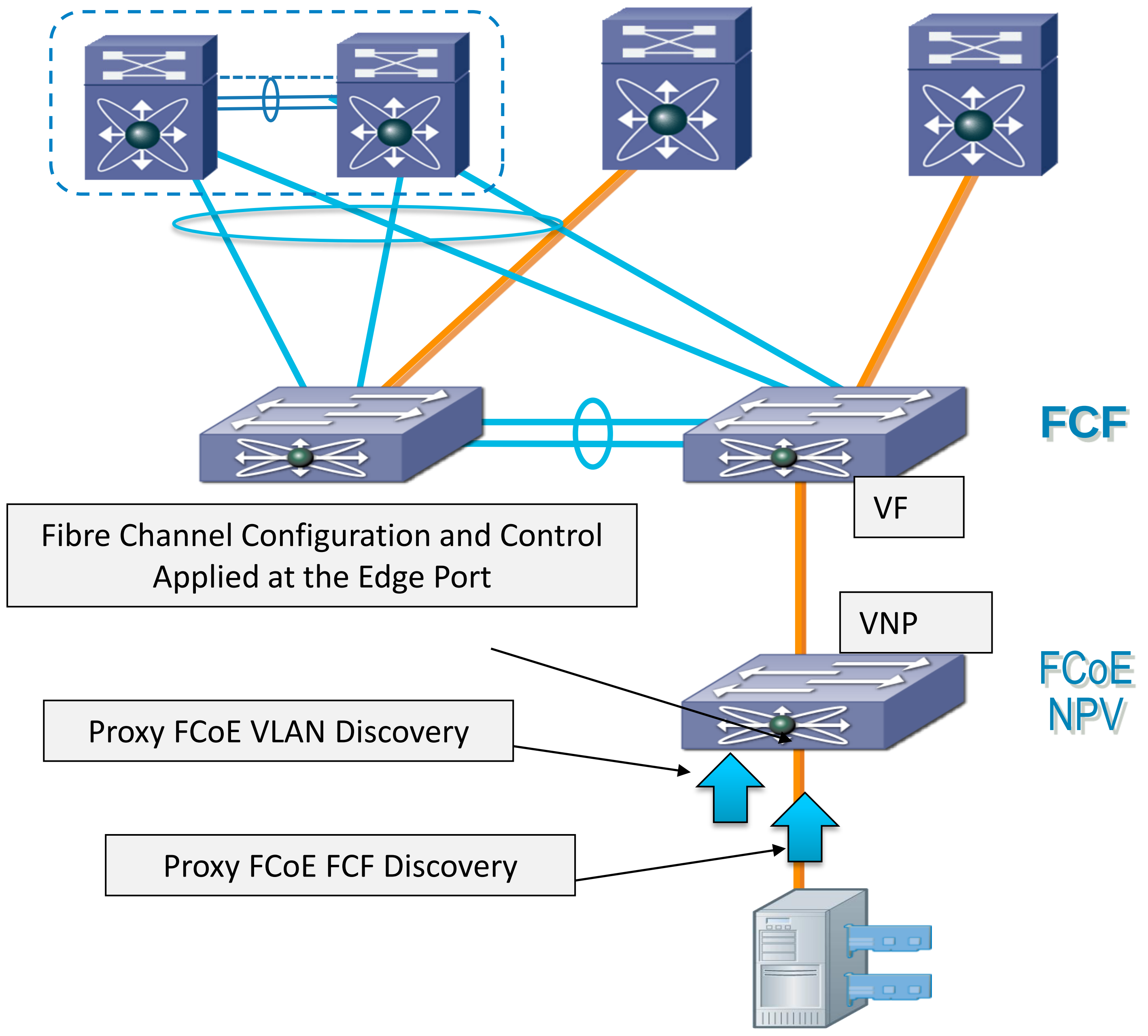
- What does a FIP Snooping device do?
  - FIP solicitations (VLAN Disc, FCF Disc and FLOGI) sent out from the CNA and FIP responses from the FCF are "snooped"
- How does a FIP Snooping device work?
  - The FIP Snooping device will be able to know which FCFs hosts are logged into
  - Will dynamically create an ACL to make sure that the host to FCF path is kept secure
- A FIP Snooping device has NO intelligence or impact on FCoE traffic/path selection/load balancing/login selection/etc
- Mentioned in the Annex of the FC-BB-5 (FCoE) standard as a way to provide security in FCoE environments
- Supported on Nexus 5000/5500 – 4.1(3)
- Supported on Nexus 7000 - 6.1(1) with F2, F1 cards



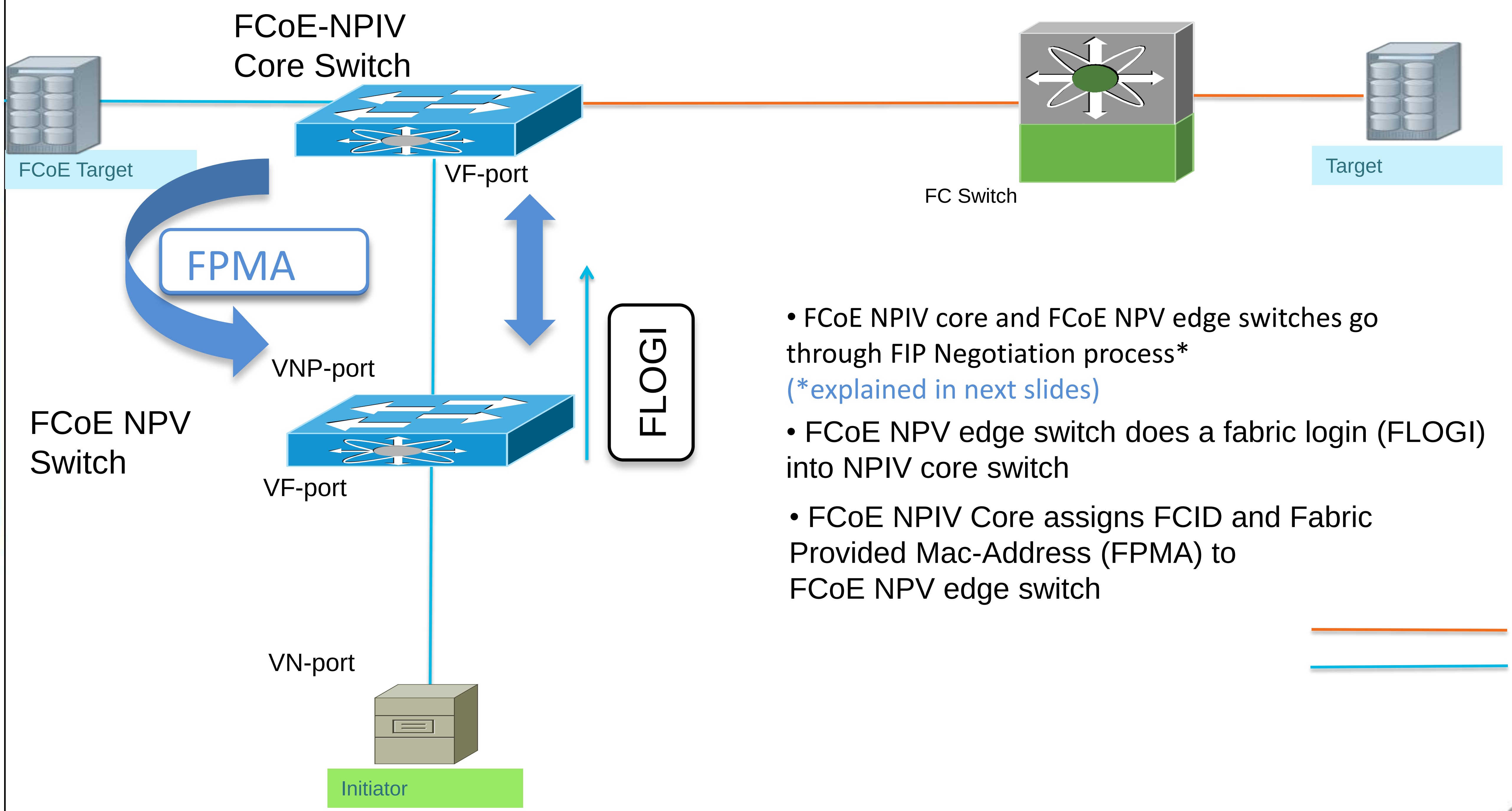
Fibre Channel Aware Device

FCoE NPV

- What does an FCoE-NPV device do?
  - ”FCoE NPV bridge" improves over a "FIP snooping bridge" by intelligently proxying FIP functions between a CNA and an FCF
- Active Fibre Channel forwarding and security element
  - FCoE-NPV load balance logins from the CNAs evenly across the available FCF uplink ports
  - FCoE NPV will take VSAN into account when mapping or ‘pinning’ logins from a CNA to an FCF uplink
- Emulates existing Fibre Channel Topology (same mgmt, security, HA, …)
- Avoids Flooded Discovery and Configuration (FIP)

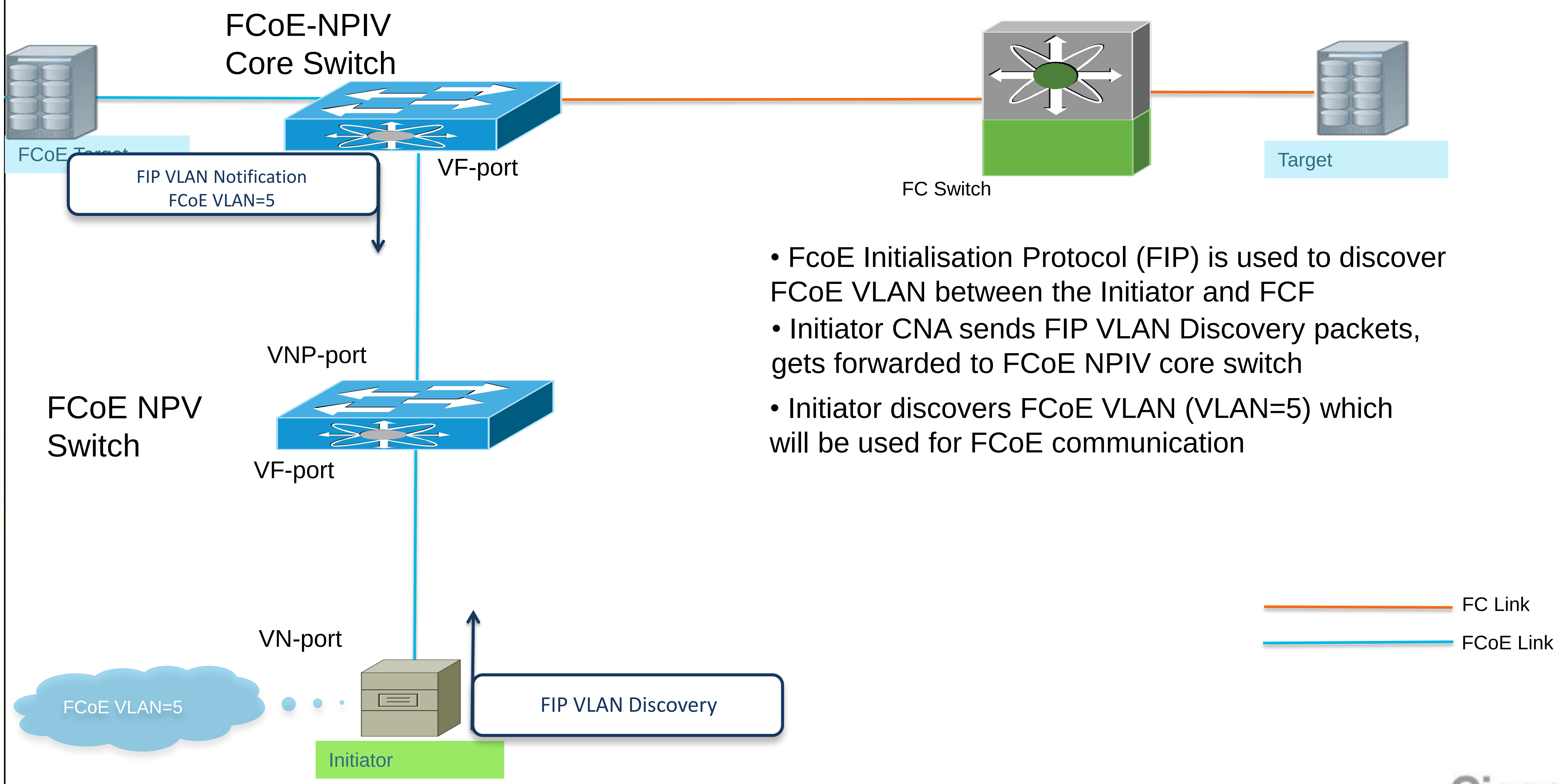


FCoE NPV: Fabric Login



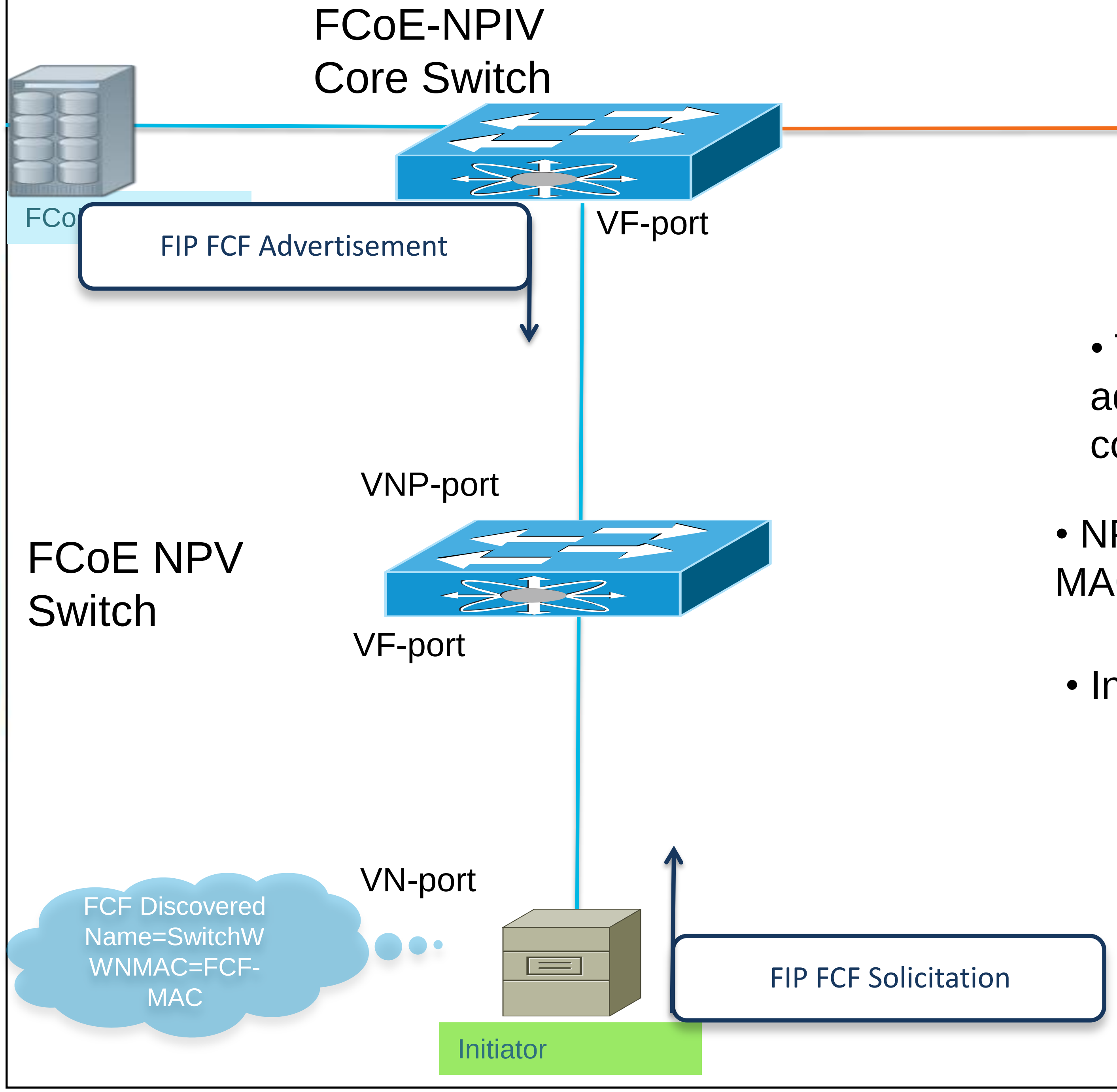
- FCoE NPIV core and FCoE NPV edge switches go through FIP Negotiation process*
  (*explained in next slides)
- FCoE NPV edge switch does a fabric login (FLOGI) into NPIV core switch
- FCoE NPIV Core assigns FCID and Fabric Provided Mac-Address (FPMA) to FCoE NPV edge switch

FCoE NPV: FIP VLAN Discovery



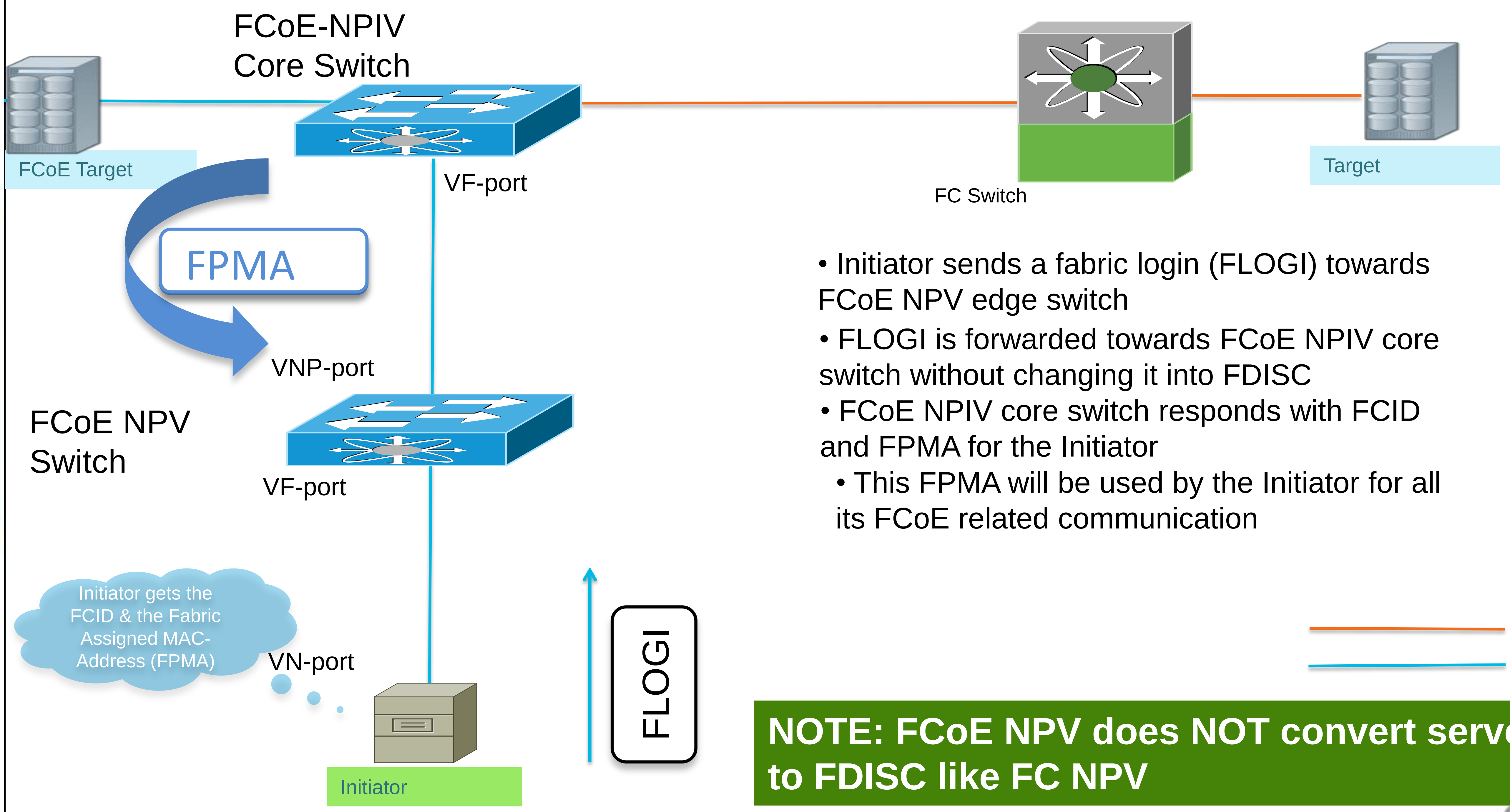
- FcoE Initialisation Protocol (FIP) is used to discover FCoE VLAN between the Initiator and FCF
- Initiator CNA sends FIP VLAN Discovery packets, gets forwarded to FCoE NPIV core switch
- Initiator discovers FCoE VLAN (VLAN=5) which will be used for FCoE communication

FCoE NPV: FIP FCF Discovery



- The Initiator sends FCF solicitation message with destination MAC address ALL-FCF-MAC, FCoE NPV switch forwards to FCoE NPIV core switch
- NPIV core switch responds with FCF advertisement containing its own MAC address and fabric related details
- Initiator detects FCF along with FCF-MAC and switch WWN

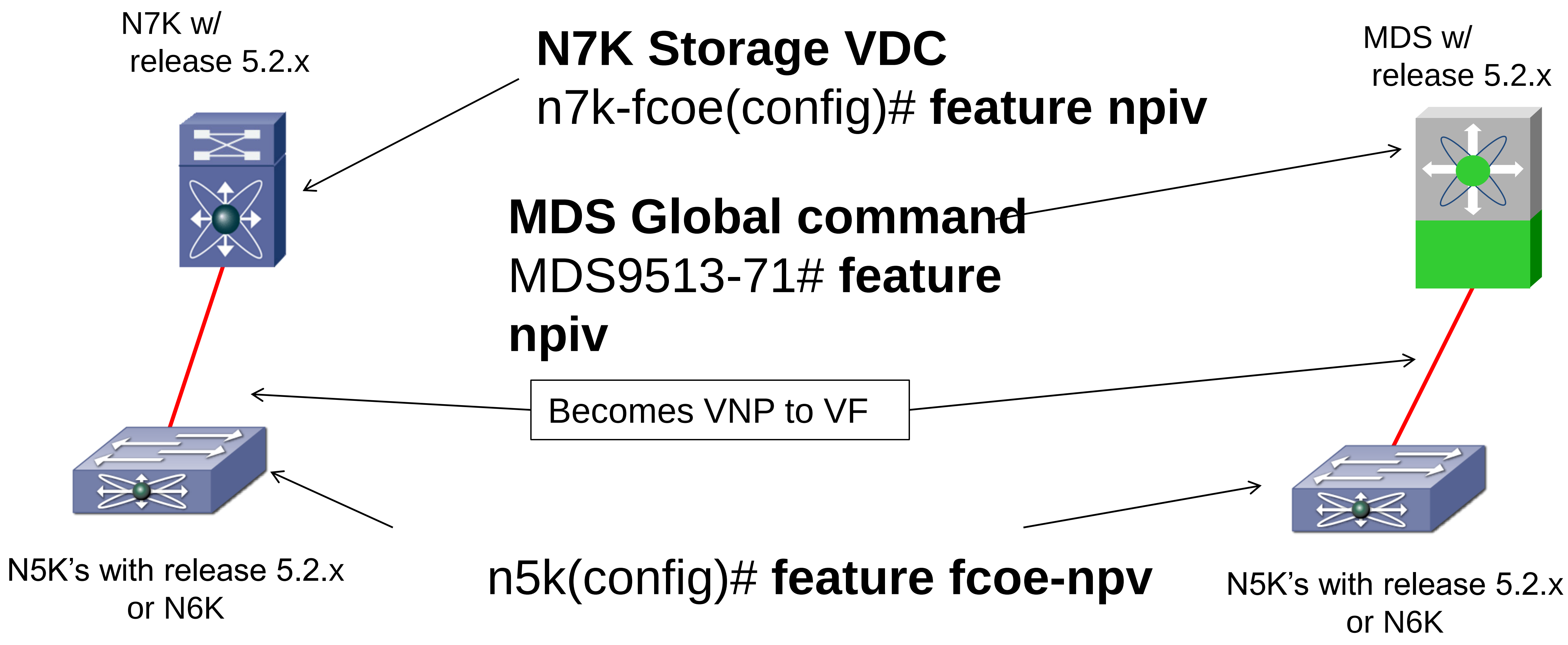
FCoE NPV: Fabric Login



- Initiator sends a fabric login (FLOGI) towards FCoE NPV edge switch
- FLOGI is forwarded towards FCoE NPIV core switch without changing it into FDISC
- FCoE NPIV core switch responds with FCID and FPMA for the Initiator
  - This FPMA will be used by the Initiator for all its FCoE related communication

NOTE: FCoE NPV does NOT convert server FLOGI to FDISC like FC NPV

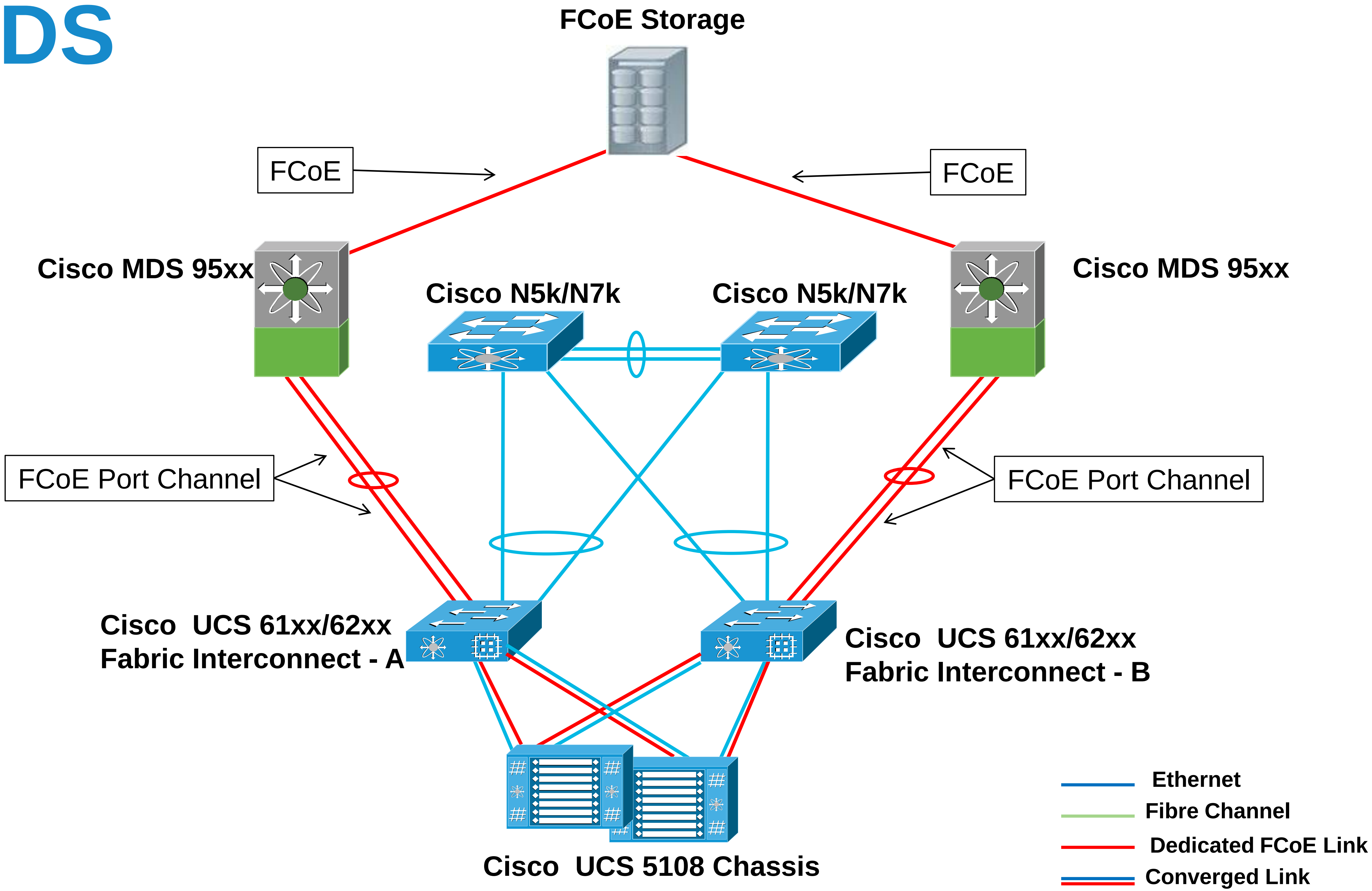
FCoE-NPV Configuration Details



Proper no drop QOS needs to be applied to all NPIV VDC's and NPV switches as shown in earlier slides

LACP Port-channels an be configured between switches for High availability.

UCS Single Hop, Direct Attach FCoE to MDS



FCoE NPV

For Your Reference

Edge Capabilities

Benefits	DCB	FIP Snooping	FCoE NPV	FCoE Switch
Scalability (Server connectivity)	✔	✔	✔	✔
Support for Lossless Ethernet	✔	✔	✔	✔
FCoE Traffic Engineering	✗	✗	✔	✔
Security (Man in the middle attack)	✗	✔	✔	✔
FC to FCoE Migration (Ease of FCoE device migration from FC fabric to FCoE network)	✗	✗	✔	✔
FCoE Traffic Load Balancing	✗	✗	✔	✔
SAN Administration (VSAN, VFC visibility for SAN Administration)	✗	✗	✔	✔

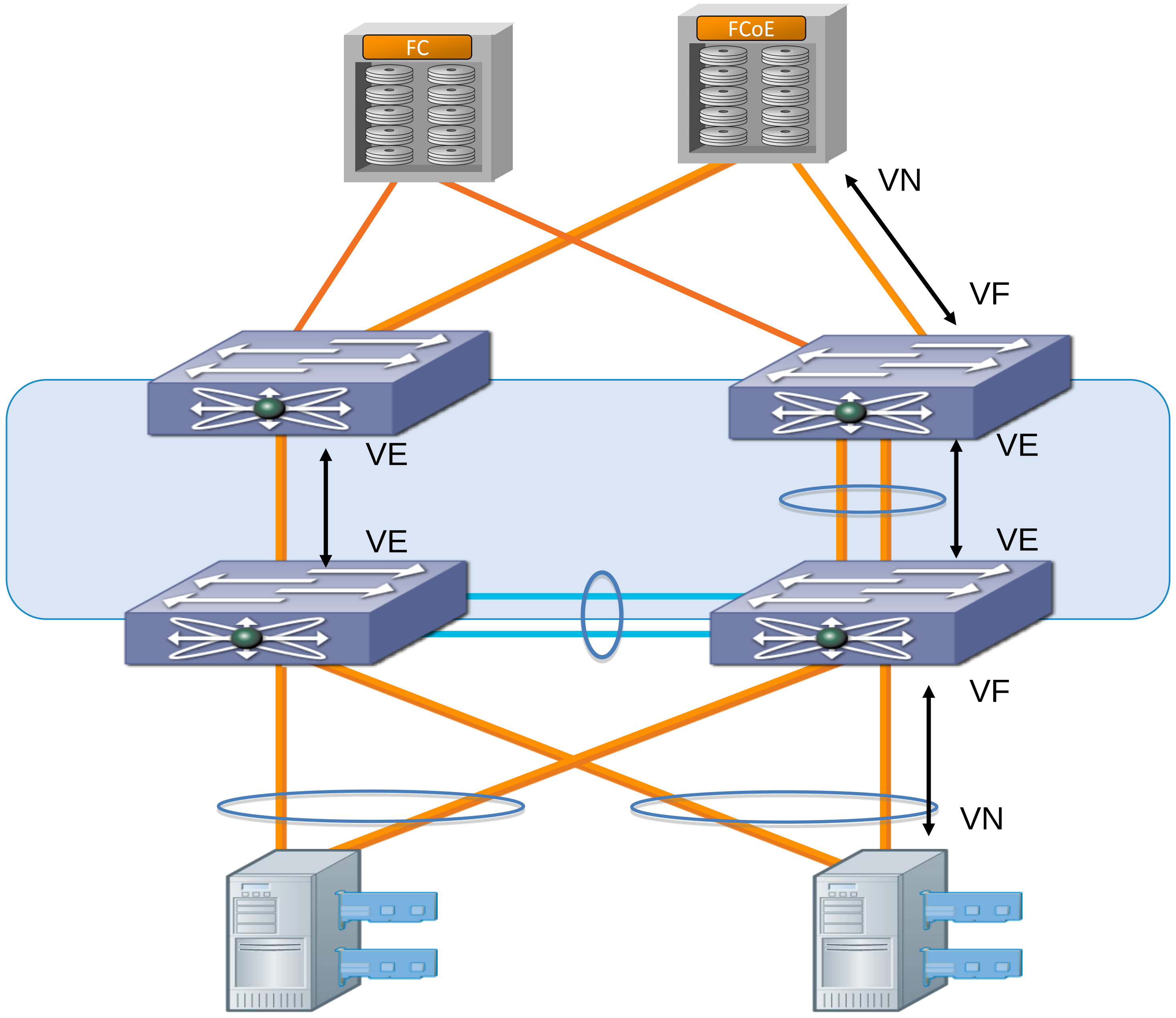
Agenda

- Unified Fabric – What and Why
- FCoE Protocol Fundamentals
- Nexus FCoE Capabilities
- FCoE Network Requirements and Design Considerations
- DCB & QoS - Ethernet Enhancements
- Single Hop Design
- Multi-Hop Design
- Futures

FCoE Multi-Tier Fabric Design

Using VE_Ports

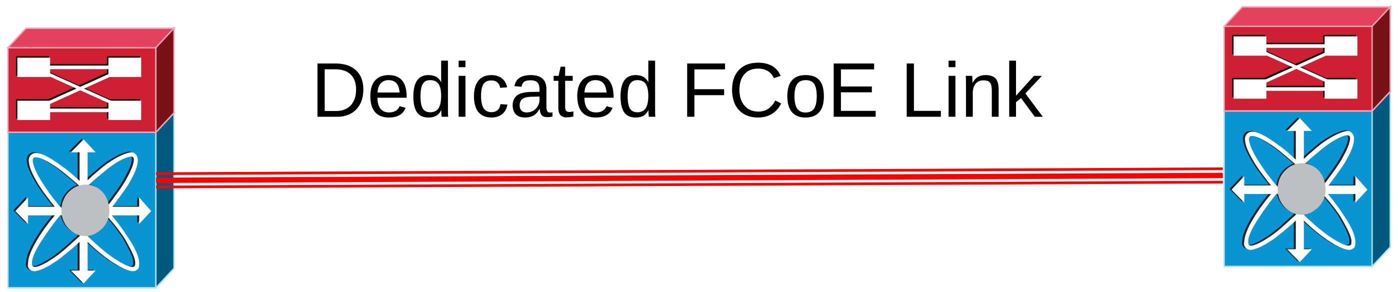
- With NX-OS 5.0(2)N2, VE_Ports are supported on/between the Nexus 5000 and Nexus 5500 Series Switches.
- Supported on Nexus 6000
- VE_Ports are run between switches acting as Fibre Channel Forwarders (FCFs)
- VE_Ports are bound to the underlying 10G infrastructure
  - VE_Ports can be bound to a single 10GE port
  - VE_Ports can be bound to a port-channel interface consisting of multiple 10GE links



All above switches are Nexus 5X00/6000 acting as an FCF

What happens when FCF' s are connected via VE_Ports

- Ethernet LACP port-channel must first be established between FCF Switches expanding the L2 ethernet network
- LLDP frames with DCBx TLVs, sourcing the MAC addresses of each switch are exchanged across the ethernet link to determine abilities.
- FIP Control exchange is done between switches
- FSPF routing established
- Fibre Channel Protocol is exchanged between the FCFs and a Fibre Channel merge of Zones is accomplished building out the FC SAN.
- You now have established a VE_Port between two DCB switches



VE_Port FIP exchange

A FIP ELP (Exchange Link Parameter)is sent on each VLAN by both switches. A FIP ACC (Accept) is sent by the switch for each VLAN.

Protocol	Summary	Source [MAC - FC]	Destination [MAC - FC]	VLAN/VSAN
FIP	Virtual Link Instantiation Request; FC4UCtl; FC FCSS; ELP;	Cisco Systems:20:A9:40 - Fabric Controller	00:26:98:0A:DF:02 - Fabric Controller	200
FIP	Virtual Link Instantiation Request; FC4UCtl; FC FCSS; ELP;	Cisco Systems:20:A9:40 - Fabric Controller	00:26:98:0A:DF:02 - Fabric Controller	100
FIP	Virtual Link Instantiation Request; FC4UCtl; FC FCSS; ELP;	Cisco Systems:20:A9:40 - Fabric Controller	00:26:98:0A:DF:02 - Fabric Controller	10
FIP	Virtual Link Instantiation Request; FC4UCtl; FC FCSS; ELP;	00:26:98:0A:DF:02 - Fabric Controller	Cisco Systems:20:A9:40 - Fabric Controller	200
FIP	Virtual Link Instantiation Request; FC4UCtl; FC FCSS; ELP;	00:26:98:0A:DF:02 - Fabric Controller	Cisco Systems:20:A9:40 - Fabric Controller	100
FIP	Virtual Link Instantiation Request; FC4UCtl; FC FCSS; ELP;	00:26:98:0A:DF:02 - Fabric Controller	Cisco Systems:20:A9:40 - Fabric Controller	10
FIP	Virtual Link Instantiation Reply; FC4SCtl; FC FCSS; Accept ELP;	Cisco Systems:20:A9:40 - Fabric Controller	00:26:98:0A:DF:02 - Fabric Controller	200
FIP	Virtual Link Instantiation Reply; FC4SCtl; FC FCSS; Accept ELP;	Cisco Systems:20:A9:40 - Fabric Controller	00:26:98:0A:DF:02 - Fabric Controller	100
FIP	Virtual Link Instantiation Reply; FC4SCtl; FC FCSS; Accept ELP;	Cisco Systems:20:A9:40 - Fabric Controller	00:26:98:0A:DF:02 - Fabric Controller	10

Discovery Solicitations & Advertisements from the FCF are sent both ways across the VE_Port, one for each FCoE mapped VLAN that is trunked on the interface.

Port	Count - Type	Count - Type	Protoc	Summary
GE Port(1,1,2)		1 - Ether Frar	LACP	LACP ver=1, A-Key=414, A-Port=271, P-Key=414, P-P
GE Port(1,1,2)		1 - Ether Frar	FIP	Discovery Solicitation;
GE Port(1,1,2)		1 - Ether Frar	FIP	Discovery Solicitation;
GE Port(1,1,2)		1 - Ether Frar	FIP	Discovery Solicitation;
GE Port(1,1,1)	1 - Ether Frar		FIP	Discovery Solicitation;
GE Port(1,1,1)	1 - Ether Frar		FIP	Discovery Solicitation;
GE Port(1,1,1)	1 - Ether Frar		FIP	Discovery Solicitation;
GE Port(1,1,2)		1 - Ether Frar	FIP	Discovery Advertisement;
GE Port(1,1,2)		1 - Ether Frar	FIP	Discovery Advertisement;
GE Port(1,1,2)		1 - Ether Frar	FIP	Discovery Advertisement;
GE Port(1,1,1)	1 - Ether Frar		SNAP	PVSTP+; Cisco;
GE Port(1,1,1)	1 - Ether Frar		SNAP	PVSTP+; Cisco;
GE Port(1,1,1)	1 - Ether Frar		SNAP	PVSTP+; Cisco;
GE Port(1,1,1)	1 - Ether Frar		FIP	Discovery Advertisement;
GE Port(1,1,1)	1 - Ether Frar		FIP	Discovery Advertisement;
GE Port(1,1,1)	1 - Ether Frar		FIP	Discovery Advertisement;
GE Port(1,1,2)		1 - Ether Frar	FIP	Discovery Advertisement;
GE Port(1,1,2)		1 - Ether Frar	FIP	Discovery Advertisement;
GE Port(1,1,2)		1 - Ether Frar	FIP	Discovery Advertisement;

FCoE VE - Fibre Channel E_Port handshake

Protocol	Summary
FC	FC4UCtl; FC FCSS; EFP;
FC	FC4UCtl; FC FCSS; BF;
FC	FC4SCtl; FC FCSS; Accept;
FC	FC4UCtl; FC FCSS; MRRA;
FC	FC4UCtl; FC FCSS; EFP;
FC	FC4UCtl; FC FCSS; EFP;
FC	FC4SCtl; FC FCSS; Accept;
FC	FC4SCtl; FC FCSS; Accept;
FC	ABTS; Basic Link Service; Abort Exchange;
FC	FC4UCtl; FC FCSS; MRRA;
FC	FC4UCtl; FC FCSS; DIA;
FC	FC4SCtl; FC FCSS; Accept;
FC	FC4UCtl; FC FCSS; RDI;
FC	FC4SCtl; FC FCSS; Accept;
FC	FC4UCtl; FC FCSS; EFP;
FC	FC4SCtl; FC FCSS; Accept;
FC	FC4UCtl; FC FCSS; MRRA;
FC	FC4SCtl; FC FCSS; Accept;
FC	FC4UCtl; FC FCSS; MR;
FC	FC4SCtl; FC FCSS; Accept;
FC	FC4SCtl; FC FCSS; Accept;
FC	FC4UCtl; FC FCSS; MR;
FC	FC4SCtl; FC FCSS; Accept;
FC	FC4UCtl; FC FCSS; HLO;
FC	FC4UCtl; FC FCSS; HLO;
FC	FC4UCtl; FC FCSS; HLO;
FC	FC4UCtl; FC FCSS; HLO;
FC	FC4UCtl; FC FCSS; LSU;
FC	FC4UCtl; FC FCSS; LSU;
FC	FC4UCtl; FC FCSS; LSA;
FC	FC4UCtl; FC FCSS; LSU;

Exchange Fabric Parameters

Build Fabric

Enhanced Zoning **M**erge **R**equest **R**esource **A**llocation

Domain ID Assign by Existing Principal Switch

Request Domain ID from New Switch

Zone Merge Request

FSPF exchanges

Differences in Trunking VSANs with FCoE VE_Ports

- In FC on the MDS, trunking is used to carry multiple VSANs over the same physical FC link. With FCoE, a physical link is replaced by a virtual link, a pair of MAC addresses.

- FCoE uses assigned MAC addresses that are unique only in the context of a single FC fabric. Carrying multiple fabrics over a single VLAN would then mean having a strong possibility for duplicated MAC addresses.

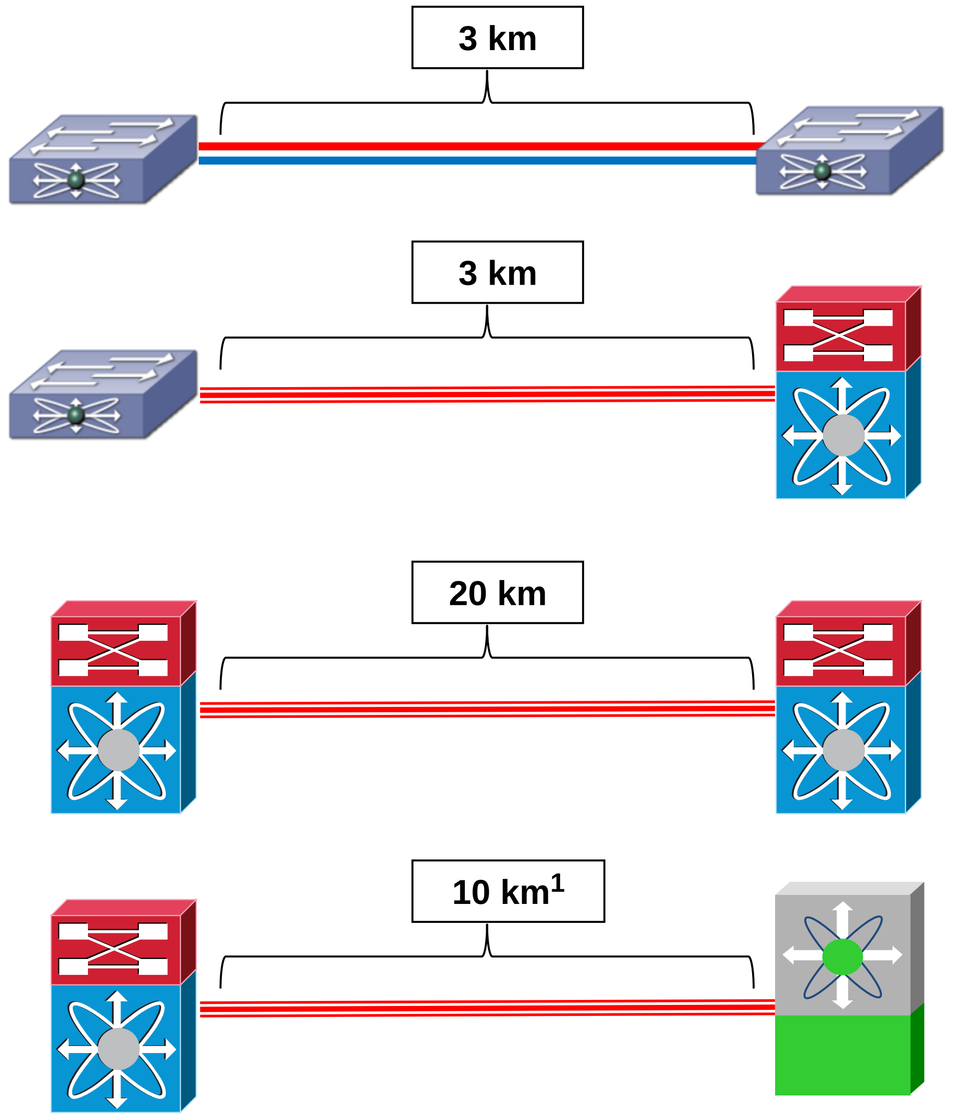
- In FCoE there cannot be more than one VSAN mapped over a VLAN.
- The net result is that trunking is done at the Ethernet level, not at the FC level.

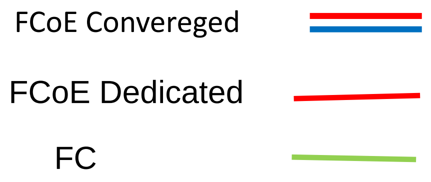
- FC trunking is not needed and the Fibre Channel Exchange Switch Capabilities(ESC) & Exchange Port Parameters (EPP) processing is not required to be performed as on the MDS

FCoE Extension Options (Nexus 5500/7000, MDS)

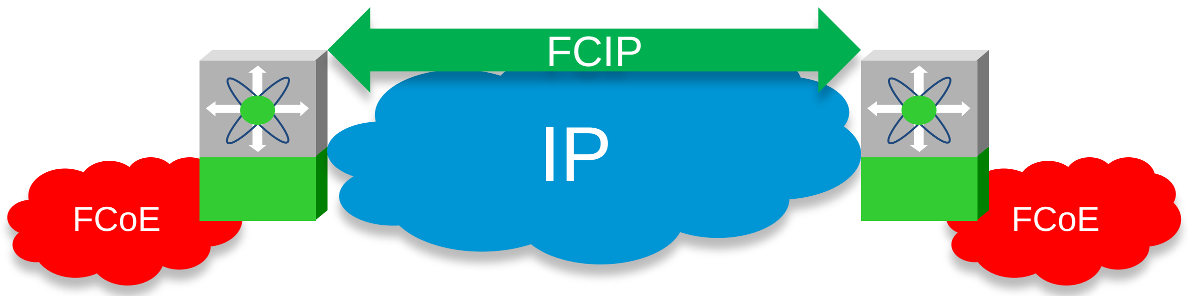
Short Distance Options

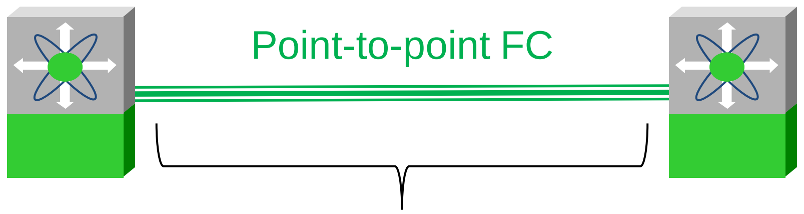
- **Requirement:** Maintain loss-less behaviour across the point-to-point link
- Supported distance is governed by the ingress buffer size available on the switch





Longer Distance Options





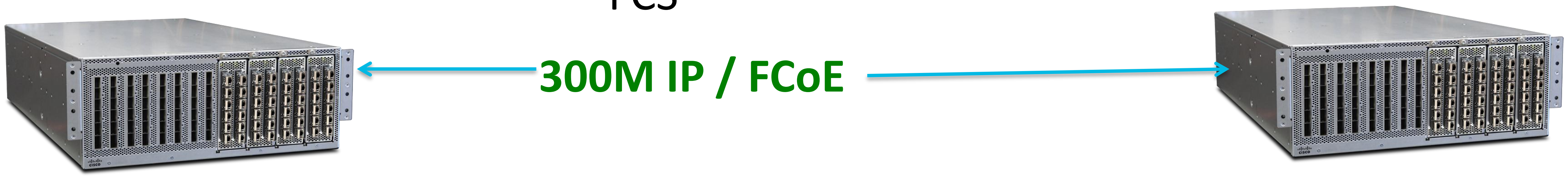
Speed (Gbps)	Max Distance (KM)
1	8000
2	4000
4	2000
8	1000
10	680

1. Limited by supported Optics

Nexus 6K Long Distance FCoE

- 300m FCoE at FCS (Optical transceiver distance limitation)

FCS



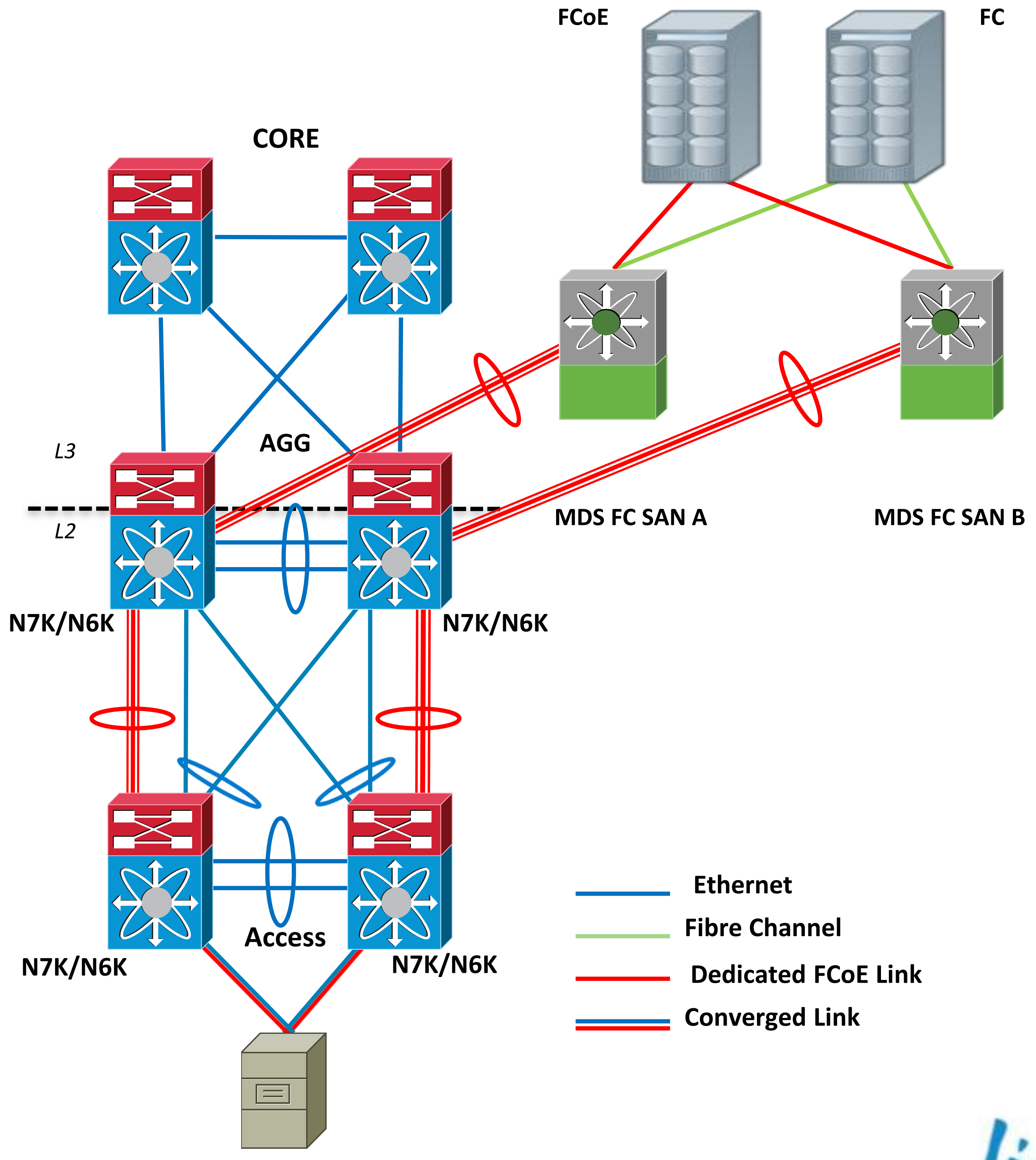
- 10KM FCoE for 10G and 40G port with global QoS policy(Roadmap)
  - Current SW implements global *network-qos* policy to tune the buffer for long distance FCoE
  - Global *network-qos* policy increase FCoE buffer for all the ports
  - >10KM can be supported with one port running long distance FCoE

Future Software Roadmap

Multi - Hop Design

Extending FCoE to MDS SAN from Aggregation

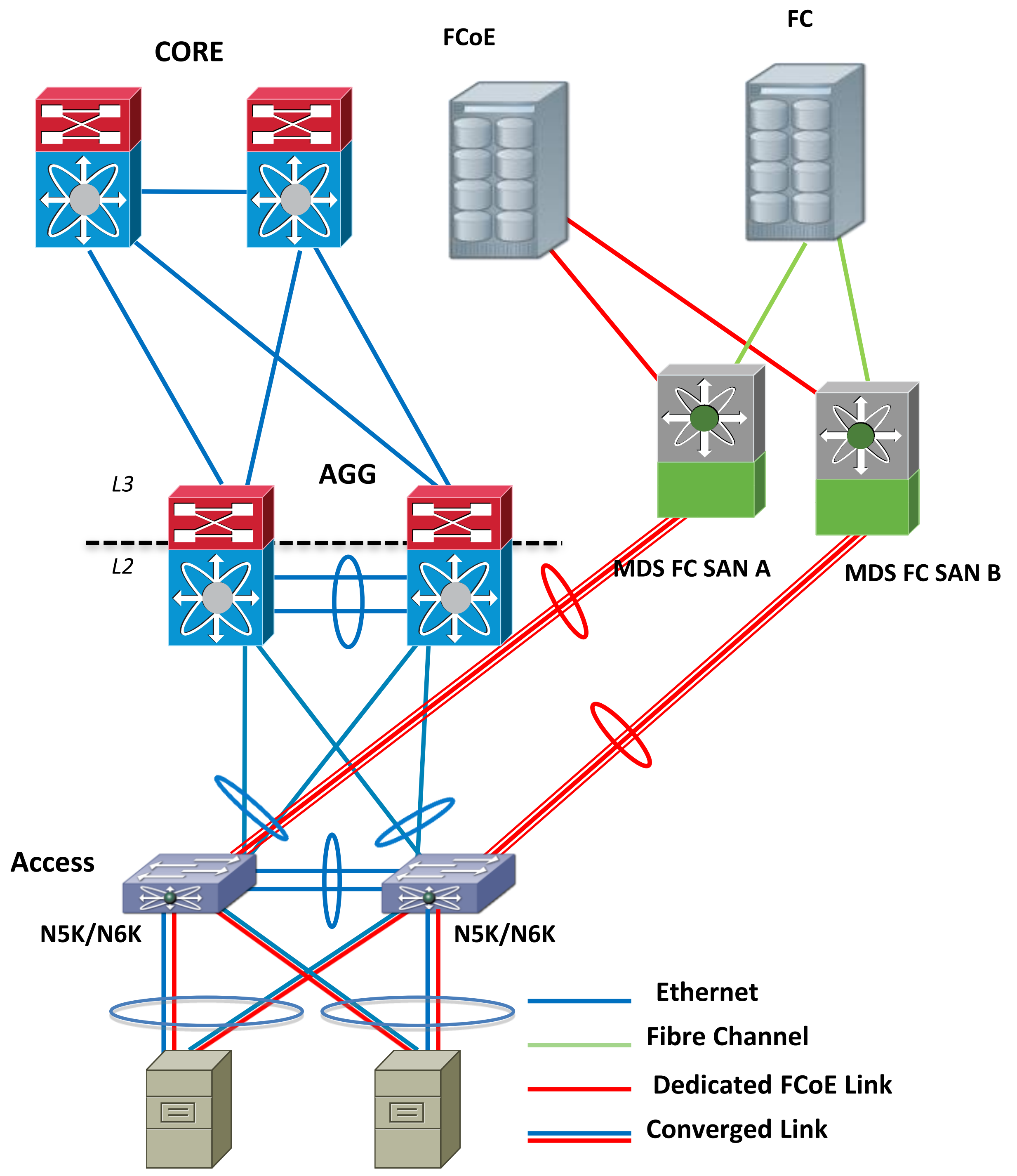
- **Converged Network to the existing SAN Core**
- Leverage FCoE wires between Fibre Channel SAN to Ethernet DCB switches in Aggregation layer using Dedicated ports
- Maintain the A – B SAN Topology with Storage VDC and Dedicated wires
- Using N7K Director Class Switches or Nexus 6000 at Access layer
- Dedicated FCoE Ports between access and Aggregation, vPC's for Data
- Zoning controlled by Core A-B SAN



Storage on MDS

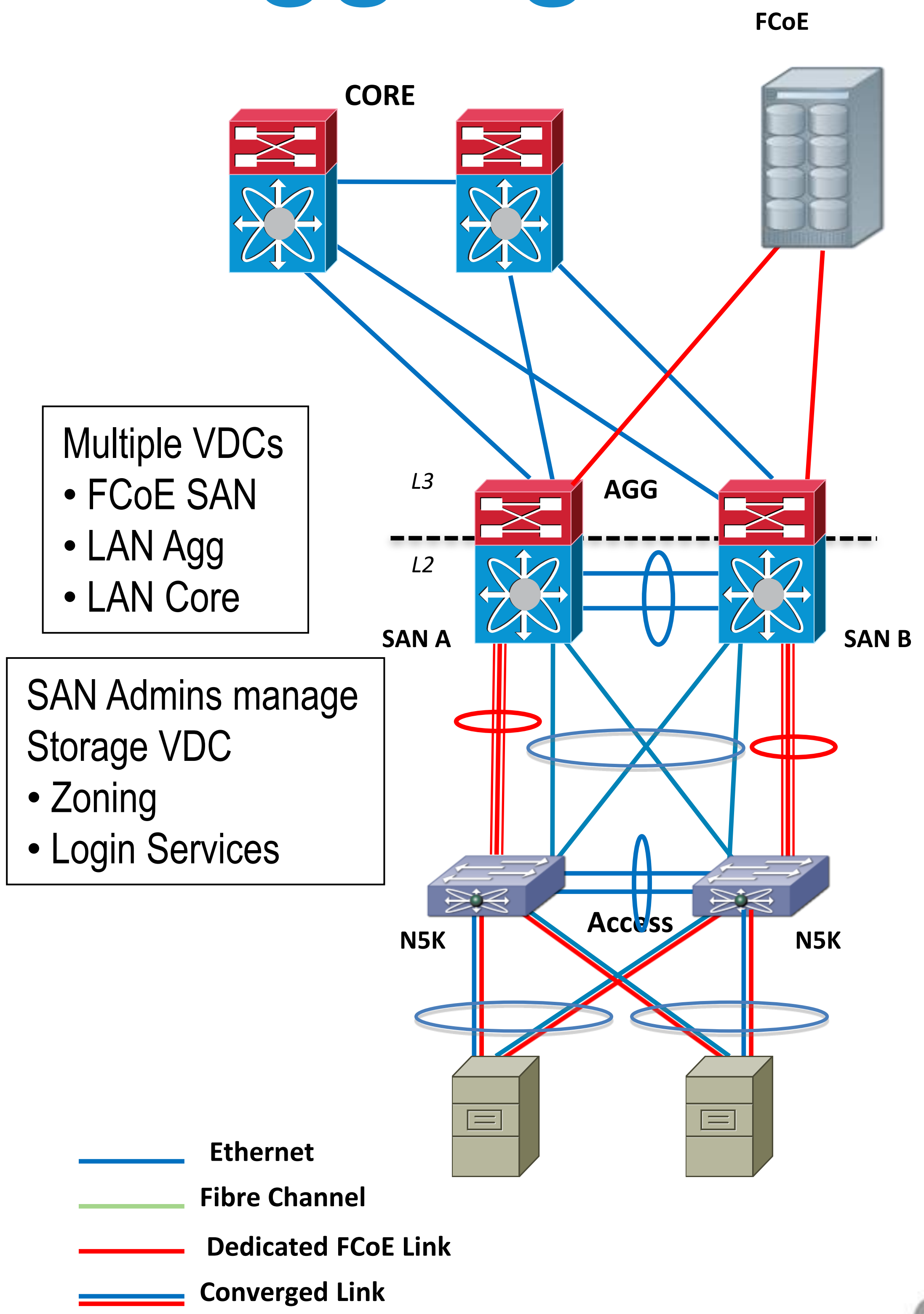
Extending FCoE to MDS SAN from Access

- Converged Network to the existing SAN Core
- Leverage FCoE wires between Fibre Channel SAN to Ethernet DCB switches (VE_Ports)
- Access switches can be in Fibre Channel switch node and assigned Domain ID , or in FCoE-NPV mode with no FC services running locally.
- Zoning controlled by Core A-B SAN



Migration of Storage to Aggregation

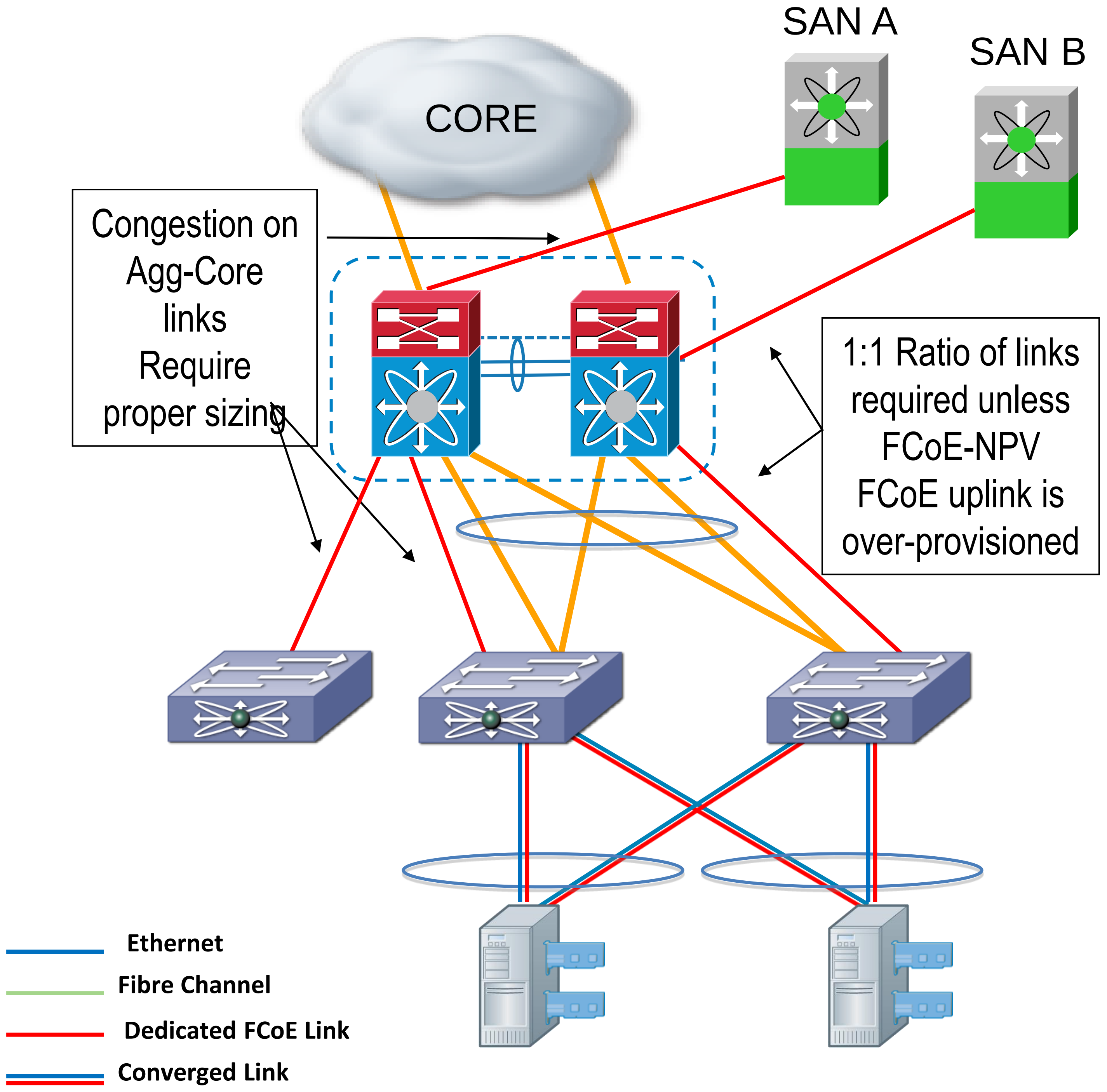
- Different requirements for LAN and SAN network designs
- Factors that will influence this use case
  - Port density
  - Operational roles and change management
  - Storage device types
- Potentially viable for smaller environments
- Larger environments will need dedicated FCoE ‘SAN’ devices providing target ports
  - Use connections to a SAN
  - Use a “storage” edge of other FCoE/DCB capable devices



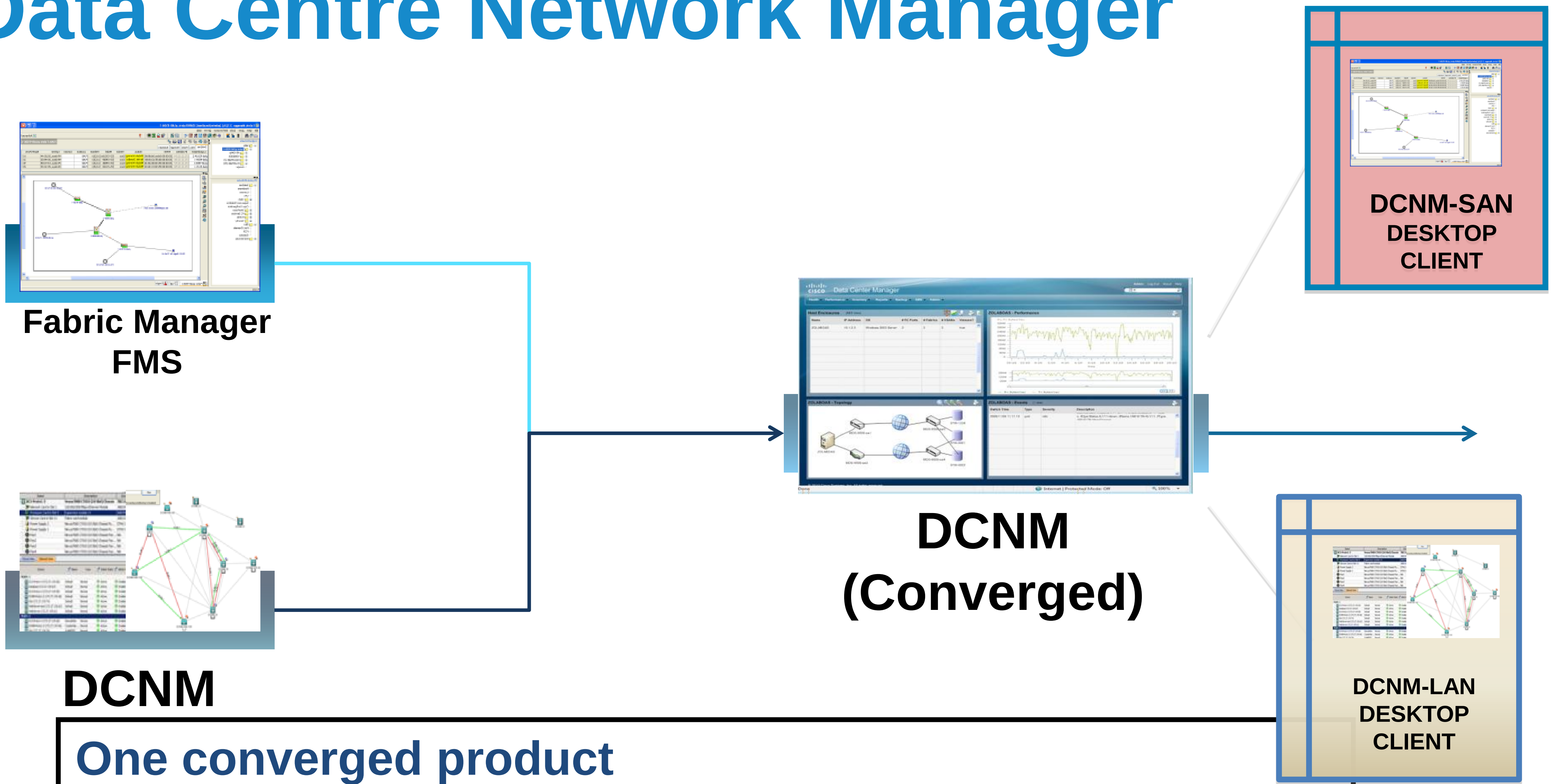
FCoE Deployment Considerations

Shared Aggregation/Core Devices

- Does passing FCoE traffic through a larger aggregation point make sense?
- Multiple links required to support the HA models
- 1:1 ratio between access to aggregation and aggregation to SAN core is required
- Need to plan for appropriate capacity in any core VE_Port link
- When is a direct Edge to Core links for FCoE are more cost effective than adding another hop?
- Smaller Edge device more likely to be able to use under-provisioned uplinks



Data Centre Network Manager



One converged product

- Single pane of glass (web 2.0 dashboard)
- Common operations (discovery, topology)
- Single installer, Role based access control
- Consistent licensing model (licenses on server)
- Integration with UCS Manager and other OSS tools

LAN/SAN Roles

Data Centre Network Manager

- Collaborative management
- Defined roles & functions
- FCoE Wizards

Nexus 7000	Tasks	Tools
LAN Admin	Storage VDC provisioning VLAN management Ethernet config (L2, network security, VPC, QoS, etc. DCB Configuration (VL, PFC, ETS Templates)	DCNM-LAN
SAN Admin	Discovery of Storage VDCs VLAN-VSAN mapping (use reserved pool) *Wizard* vFC provisioning *Wizard* Zoning	DCNM-SAN

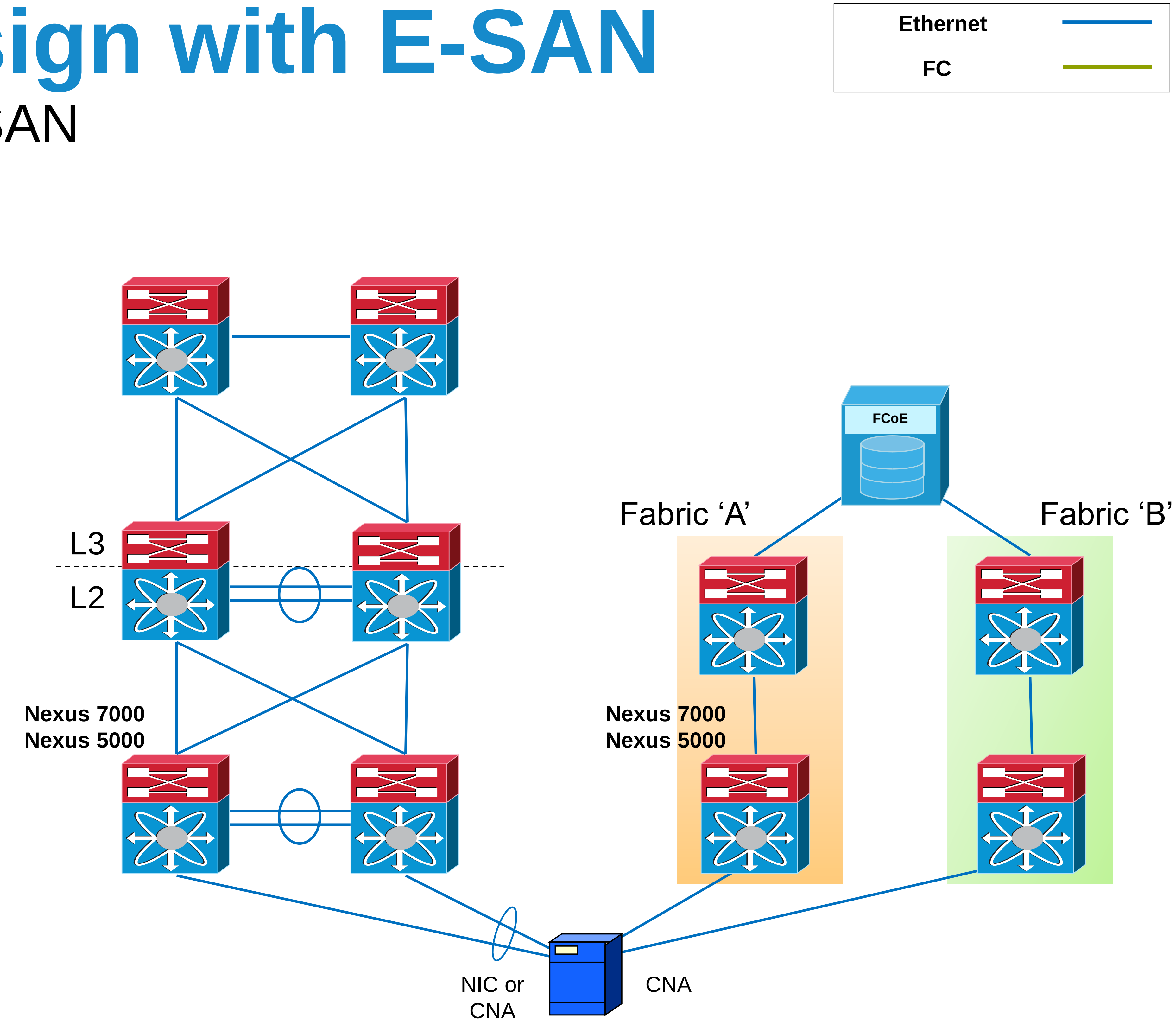
Agenda

- Unified Fabric – What and Why
- FCoE Protocol Fundamentals
- Nexus FCoE Capabilities
- FCoE Network Requirements and Design Considerations
- DCB & QoS - Ethernet Enhancements
- Single Hop Design
- Multi-Hop Design
- Futures

Data Centre Design with E-SAN

Ethernet LAN and Ethernet SAN

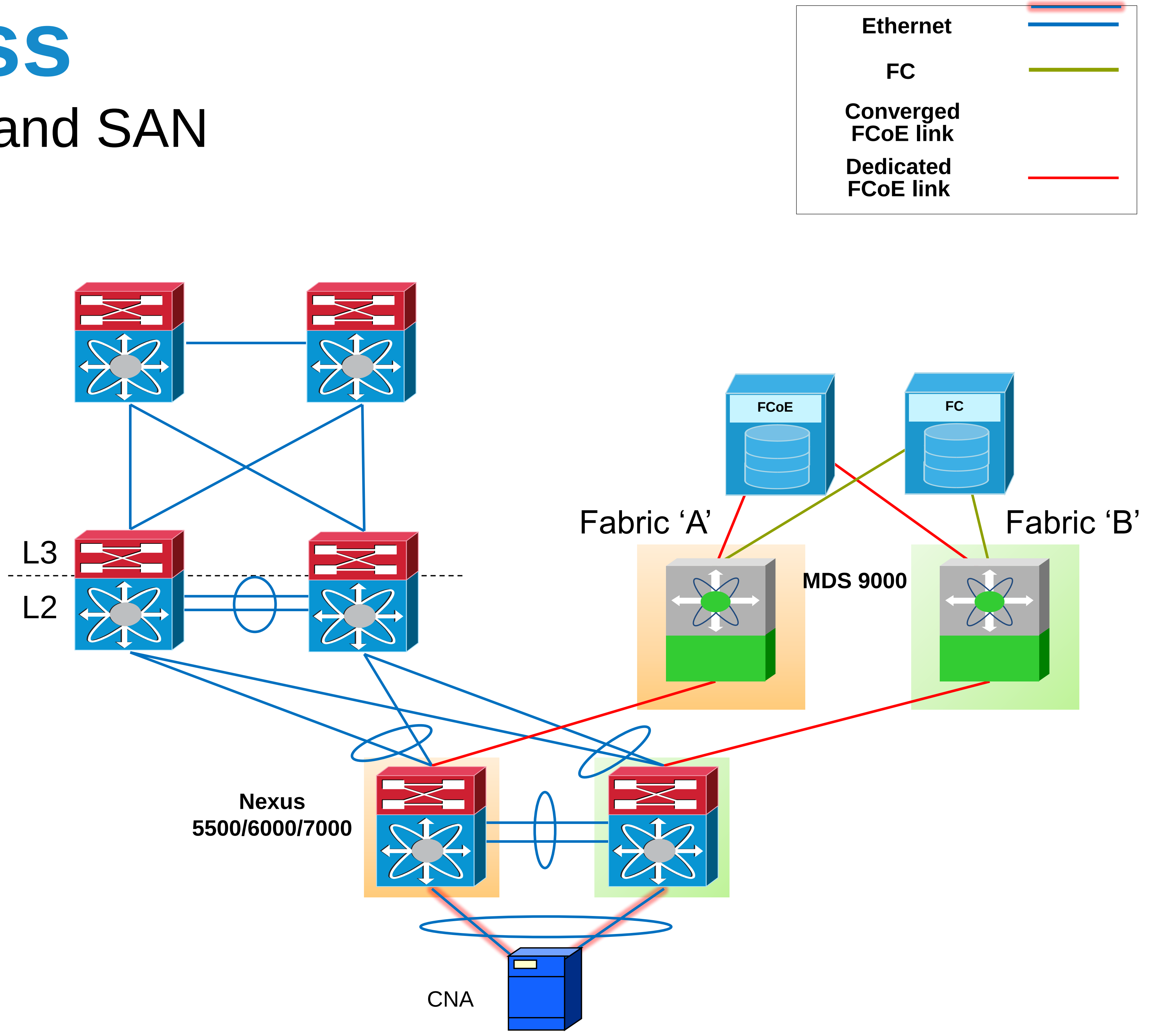
- Same topologies as existing networks, but using Nexus Unified Fabric Ethernet switches for SANs
- Physical and Logical isolation of LAN and SAN traffic
- Additional Physical and Logical separation of SAN fabrics
- Ethernet SAN Fabric carries FC/FCoE & IP based storage (iSCSI, NAS, …)
- Common components: Ethernet Capacity and Cost



Converged Access

Sharing Access Layer for LAN and SAN

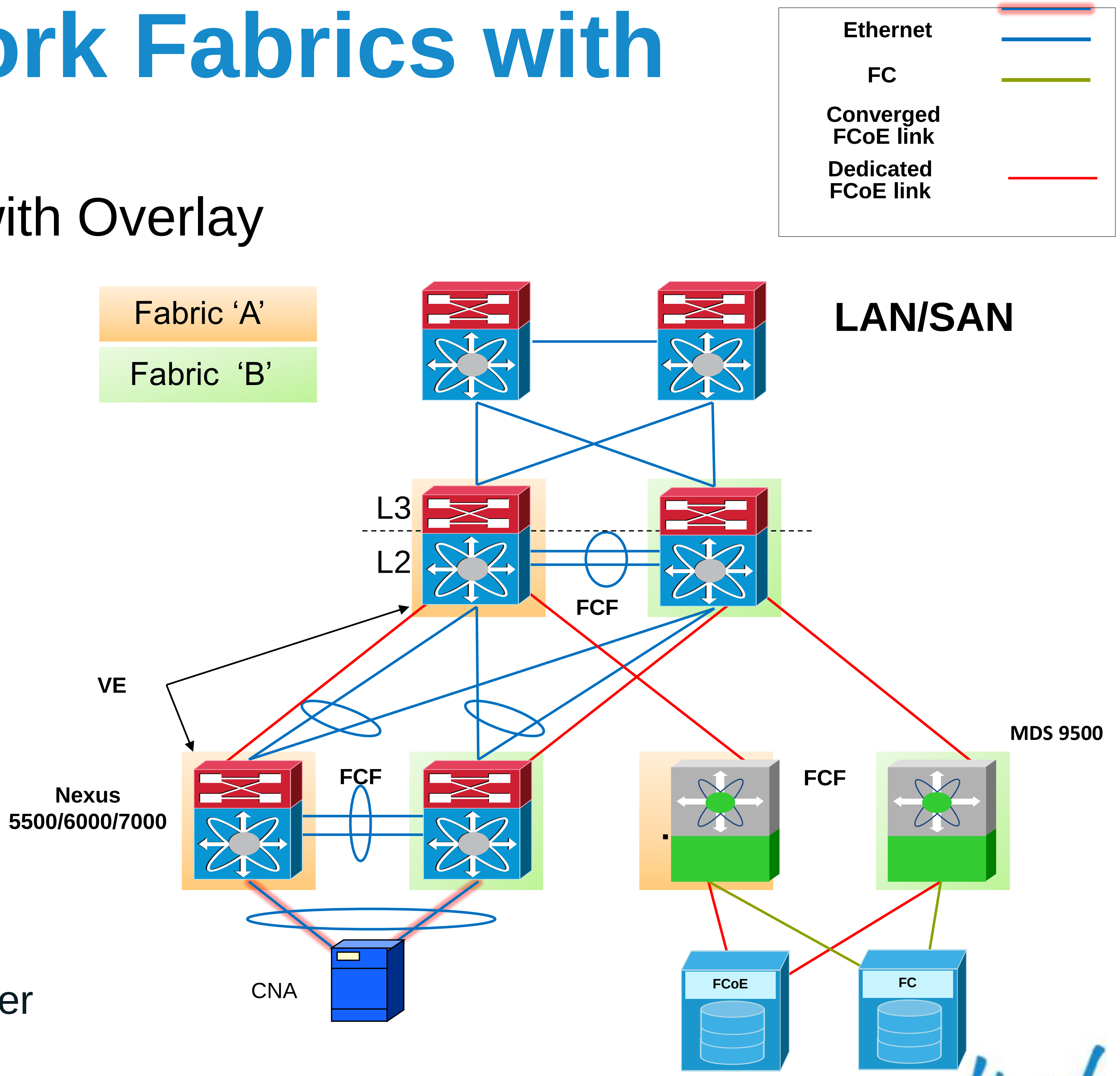
- Shared Physical, Separate Logical LAN and SAN traffic at Access Layer
- Physical and Logical separation of LAN and SAN traffic at Aggregation Layer
- Additional Physical and Logical separation of SAN fabrics
- Higher I/O, HA, fast re-convergence for host LAN traffic



Converged Network Fabrics with Dedicated Links

Maintaining Dual SAN fabrics with Overlay

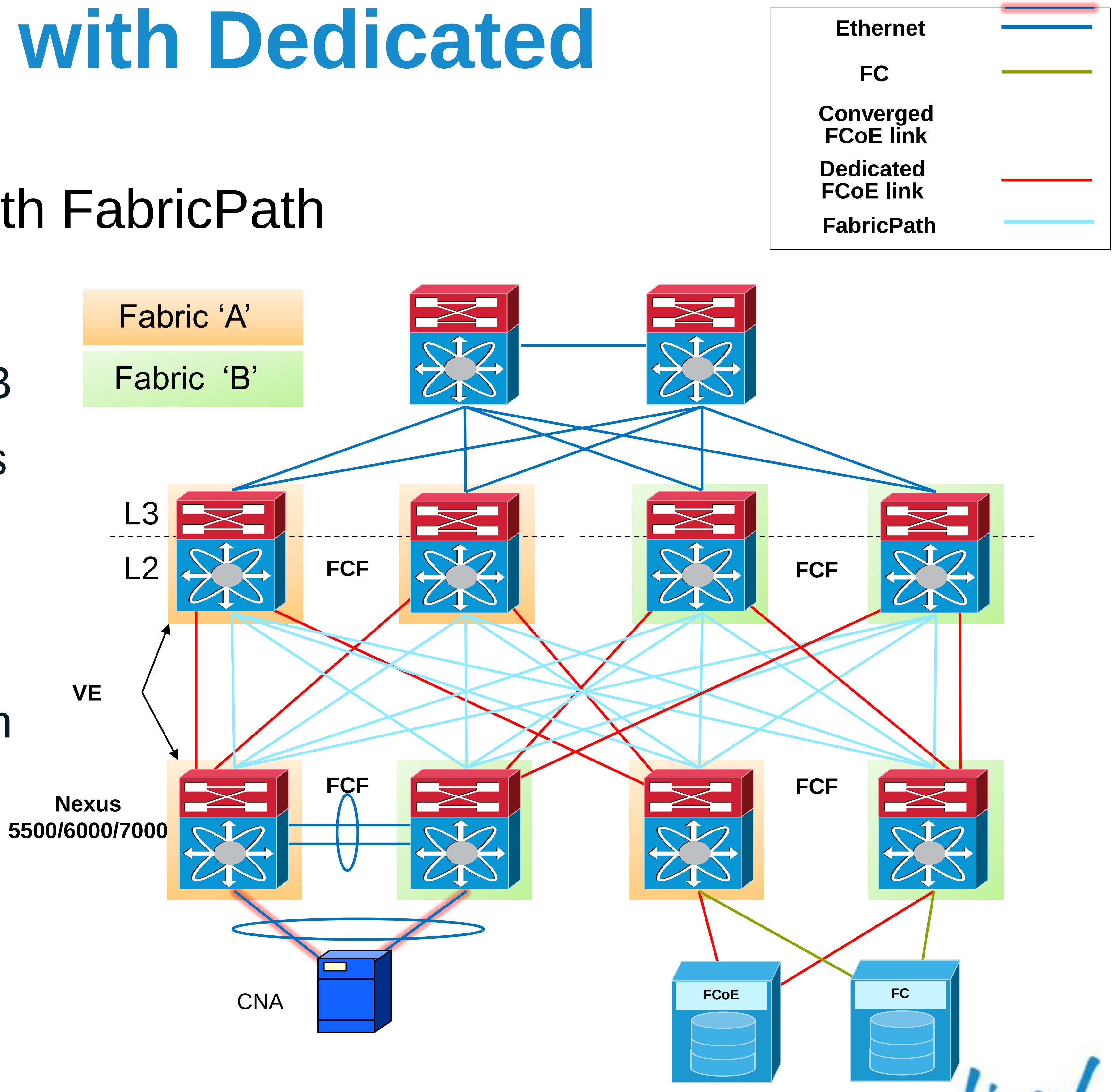
- LAN and SAN traffic share physical switches
- LAN and SAN traffic use dedicated links between switches
- All Access and Aggregation switches are FCoE FCF switches
- Dedicated links between switches are VE_Ports
- Storage VDC for additional operation separation at high function agg/core
- Improved HA, load sharing and scale for LAN vs. traditional STP topologies
- SAN can utilise higher performance, higher density, lower cost Ethernet switches for the aggregation/core



Converged Network with Dedicated Links

Maintaining Dual SAN fabrics with FabricPath

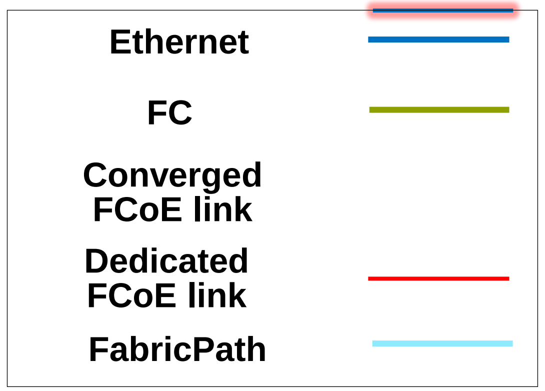
- FabricPath enabled for LAN traffic
- Dual Switch core for SAN A & SAN B
- All Access and Aggregation switches are FCoE FCF switches
- Dedicated links between switches are VE_Ports
- Storage VDC for additional operation separation at high function agg/core
- Improved HA and scale over vPC (ISIS, RPF, … and N+1 redundancy)
- SAN can utilise higher performance, higher density, lower cost Ethernet switches



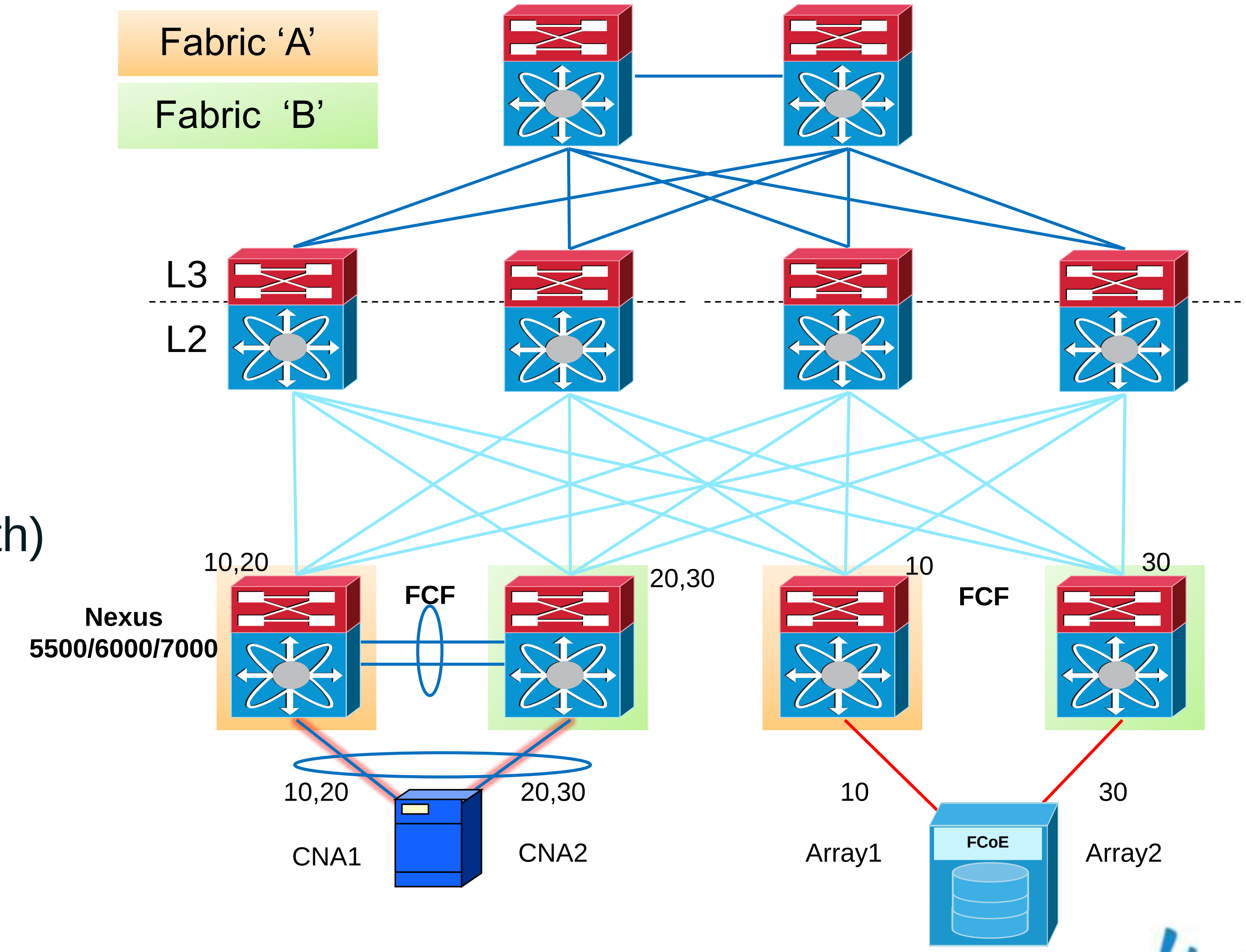
Convergence

Looking forward: Converged Network – Single Fabric

SAN Separation at the Access Switch



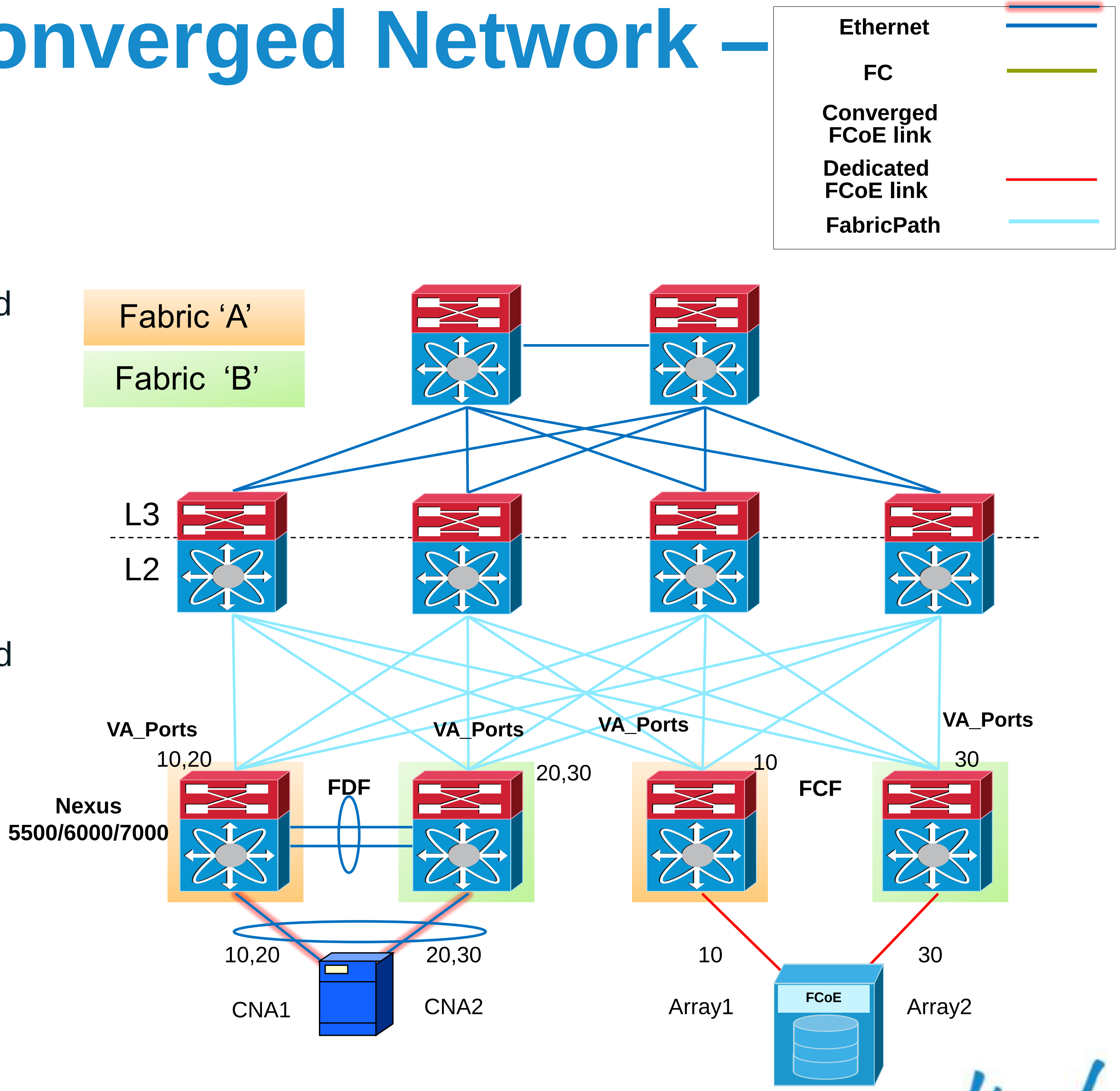
- LAN and SAN traffic share physical switches and links
- FabricPath enabled
- All Access switches are FCoE FCF switches
- VE_Ports to each neighbour Access switch
- Single process and database (FabricPath) for forwarding
- Improved (N + 1) redundancy for LAN & SAN
- Sharing links increases fabric flexibility and scalability
- Distinct SAN 'A' & 'B' for zoning isolation and multipathing redundancy



Looking forward: Converged Network – Single Fabric

FC-BB-6

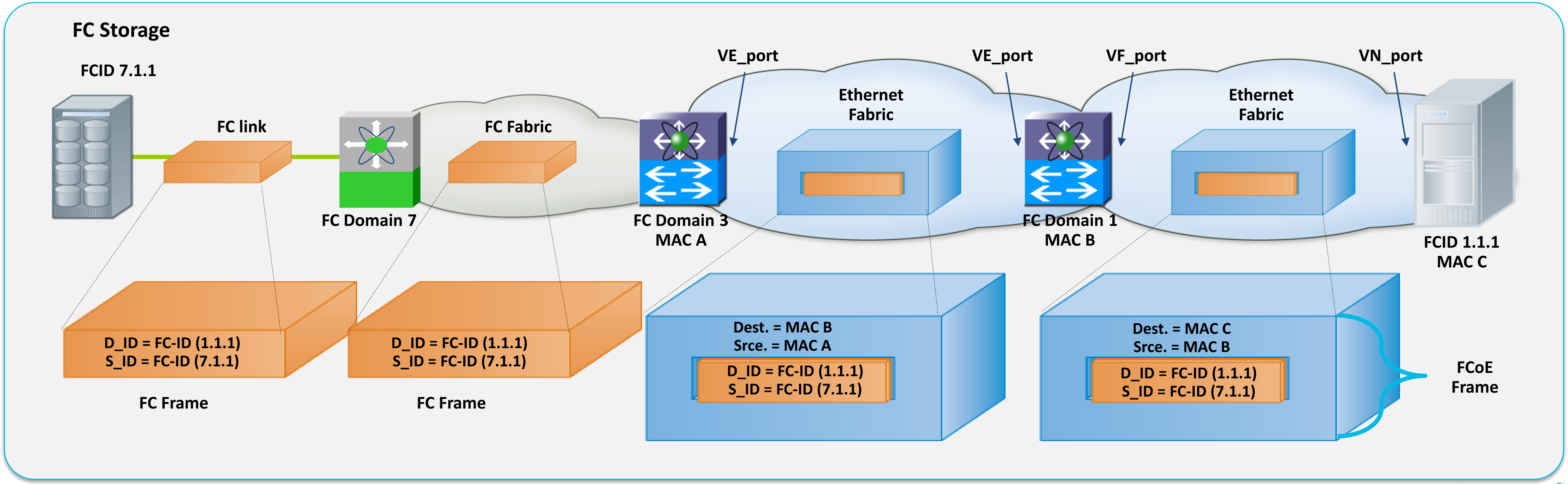
- LAN and SAN traffic share physical switches and links
- FabricPath enabled
- VA_Ports to each neighbour FCF switch
- Single Domain
- FDF to FCF transparent failover
- Single process and database Single process and database (FabricPath) for forwarding
- Improved (N + 1) redundancy for LAN & SAN
- Sharing links increases fabric flexibility and scalability
- Distinct SAN ‘A’ & ‘B’ for zoning isolation and multipathing redundancy



Unified Multi-Tier Fabric Design

Current Model

- All devices in the server storage path are Fibre Channel Aware
- Evolutionary approach to migration to Ethernet transport for Block storage
- FC-BB-6 and FabricPath (L2MP) to provide enhancements but need to be aware of your ability to evolve



Q & A

Complete Your Online Session Evaluation

Give us your feedback and receive a Cisco Live 2013 Polo Shirt!

Complete your Overall Event Survey and 5 Session Evaluations.

- Directly from your mobile device on the Cisco Live Mobile App
- By visiting the Cisco Live Mobile Site www.ciscoliveaustralia.com/mobile
- Visit any Cisco Live Internet Station located throughout the venue

Polo Shirts can be collected in the World of Solutions on Friday 8 March 12:00pm-2:00pm

Don't forget to activate your Cisco Live 365 account for access to all session material, communities, and on-demand and live activities throughout the year. Log into your Cisco Live portal and click the "Enter Cisco Live 365" button.
www.ciscoliveaustralia.com/portal/login.ww