EUROPEAN ROADSHOW – Nov 2010

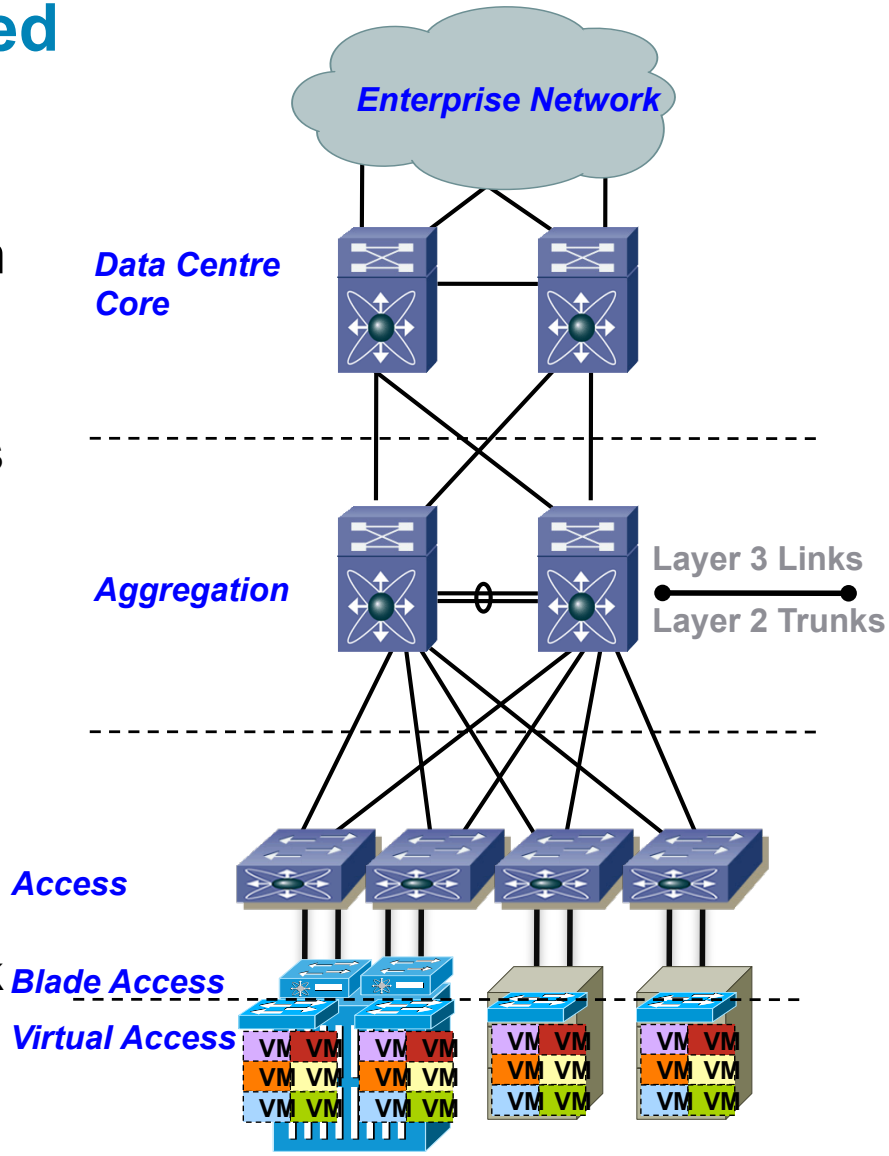
Addressing Server Access Networking Challenges

Ranga Rao (rangrao@cisco.com)

SAVBU Technical Marketing

Virtualization
A new network layer is created

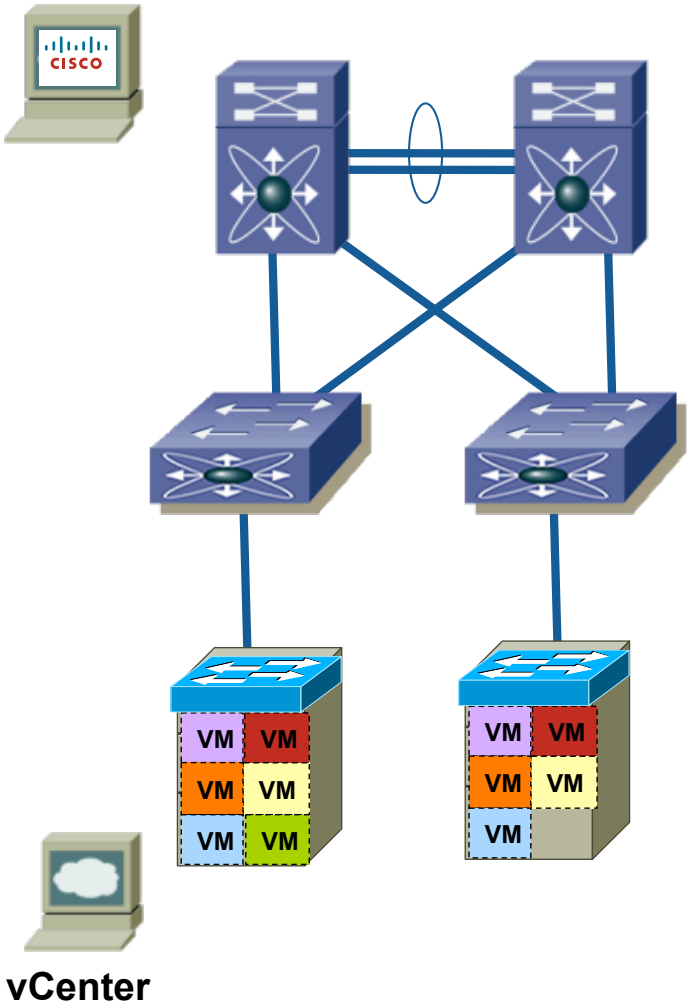
- The network no longer connects **directly** to the server (which lives in a VM)
  - The physical switchport is now a 802.1Q trunk with X number of vlans
  - The Virtual Access Layer (vSwitch) is managed only by the server team
- This brings many challenges
  - VLAN proliferation and scaling
  - Policy is no longer possible per server
  - Lack of visibility of per-VM traffic, lack of accountability
  - Complex to operate and troubleshoot



And the best of Virtualization is what most stresses the network …

- vMotion moves VMs across different switch physical ports
- The network has no visibility
- Network configuration and policy must follow
  - Vlan
  - ACLs
  - Security features, etc …
- Network admin loses track of where applications are running
- Server admin gets busy with networking issues (each ESX host = 1 vSwitch)

All these become inhibitors to Broader Virtualization Adoption

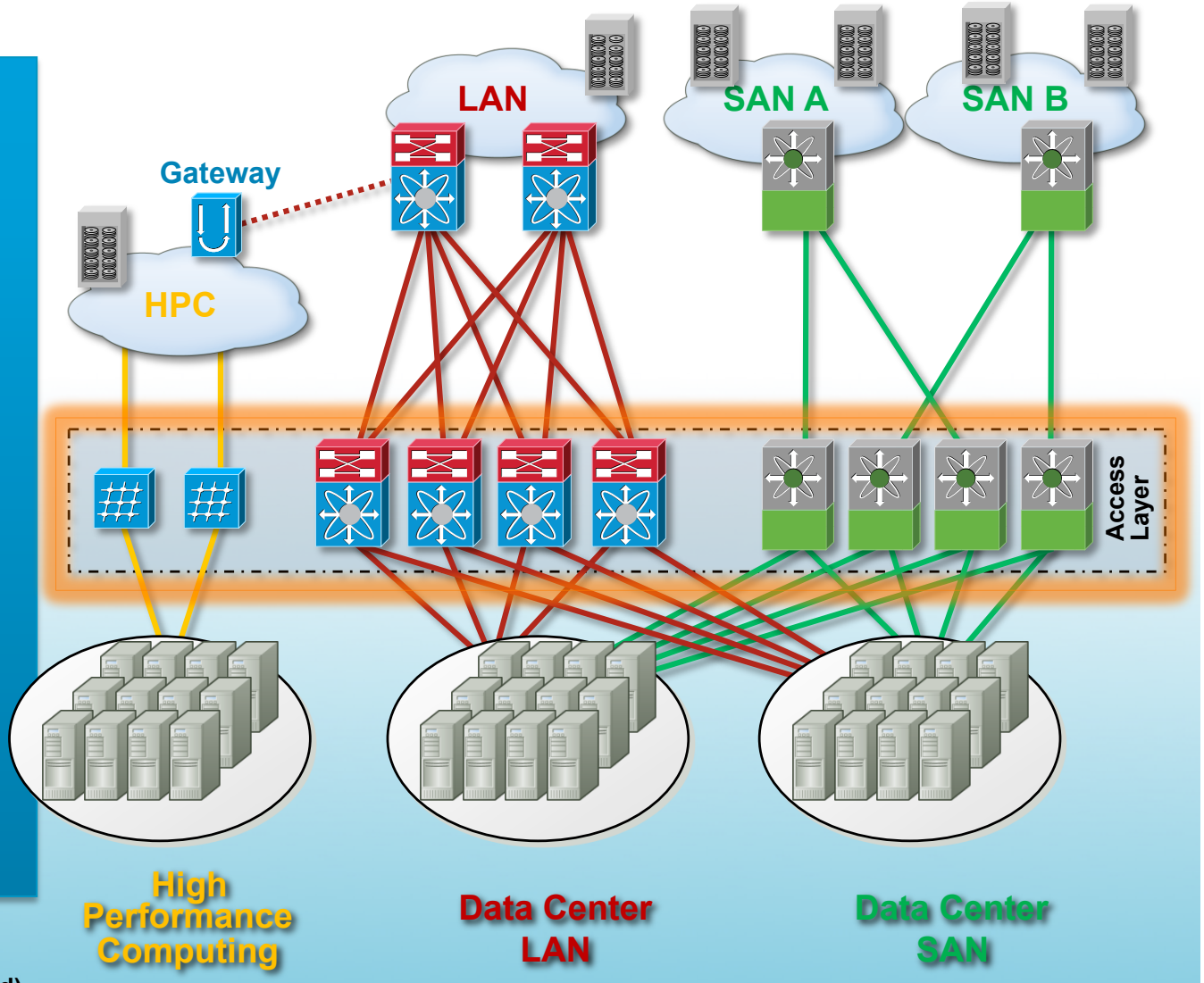


DC Access is a real challenge for everybody

Growing demands & increasing complexity

- Data centres have multiple networks
- High demand on data centre resources
- More servers
- More virtual machines per server
- More connections & cables per server
- Proliferation of network equipment

Increased power & cooling needed



Agenda

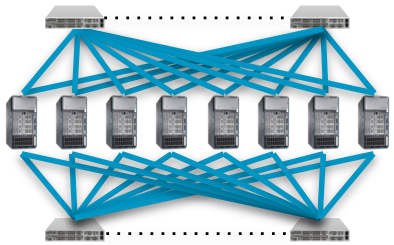
- Distributed
- Scalable
- Virtual
- Unified

Nexus Data Center Access Technologies

Distributed Access Fabric

- Fabric Extender Technology
- 100M/1G/10G Capable
- FCoE & VNTag Support

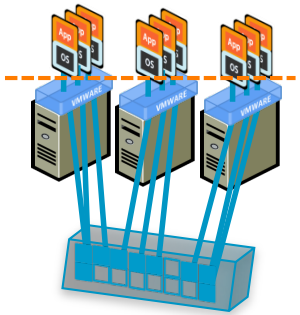
- 10:1 reduction in managed devices
- No cross-rack cabling
- Form-Factor & MGMT optimized



Scalable Access Choice

- Virtual Port Channel (vPC)
- FabricPath/Trill
- Routed Access

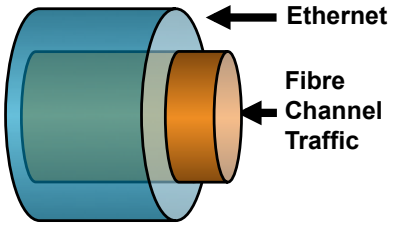
- Accelerate IO intensive applications
- Eliminate logical Layer 2 loops
- Massive bisectional bandwidth



Virtual Networking

- VN-Link
- VN-Services
- VN-Manager

- Extend networking principles to virtualized environments
- Extend network services to virtualized environments
- Enable consistent VN-wide management



Unified Fabric

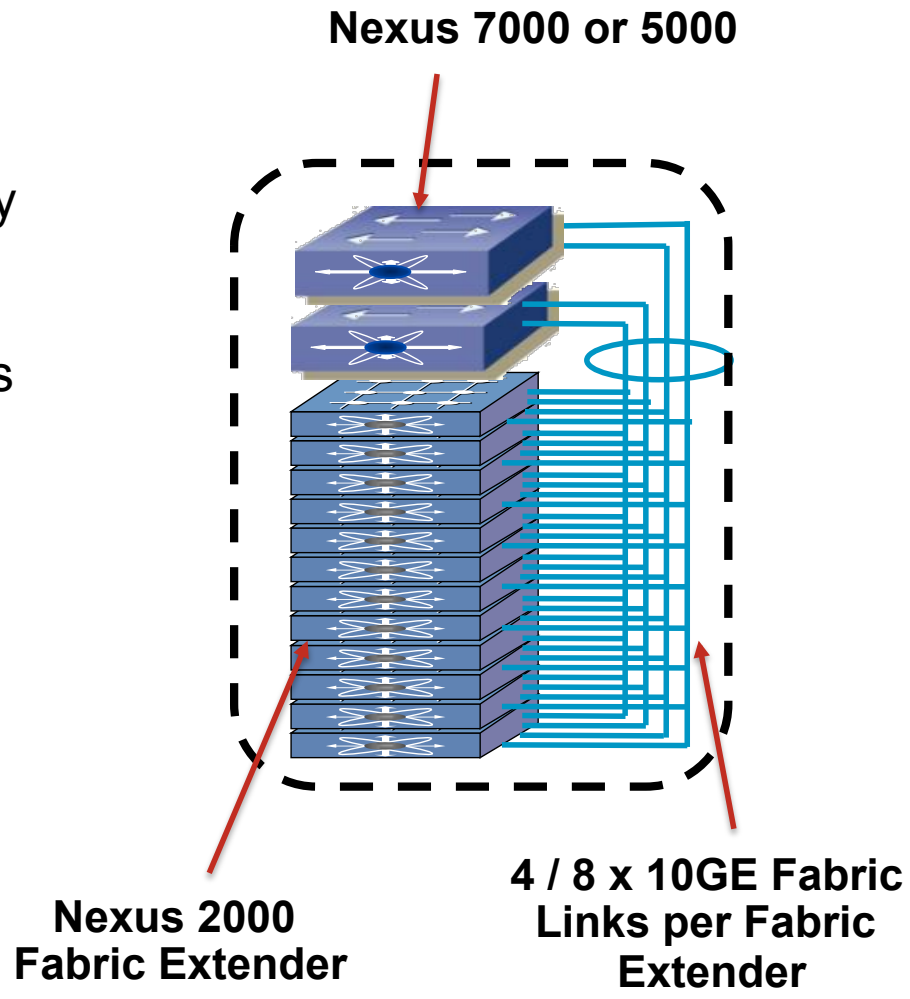
- LAN/SAN Consolidation
- Converged Network Adapters (CNAs)
- High Performance Server Interconnect

- 50% or more cable reduction
- 50% or more network/SAN device reduction
- Maintains logical LAN/ SAN mgmt model

Powered by Cisco NX-OS Data Center Operating System

Virtualized Access Switch with FEXlink

- Nexus 7000/5000/2000 Virtualized Access Switch provides a number of design options to address evolving Data Center requirements
- Fabric Extender provides for flexibility in the design of the physical topologies
- Aids in building larger layer 2 designs safely
  - 10:1 reduction in management complexity
  - Single virtual access switch (Simplifies the layer 2 design)
- Support of 16-way 10GE Etherchannel combined with vPC
- Leverages Nexus ISSU for non-stop access layer

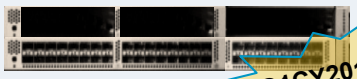


Cisco Nexus 2000 Fabric Extender (FEX)
Platform Overview

Model	Nexus 2148T	Nexus 2224TP	Nexus 2248TP	Nexus 2232PP-10G
Product Shipping	Yes (Q1CY09)	Yes (Q3CY10)	Yes (Q2CY10)	Yes (Q2CY10)
Form Factor	1 RU	1 RU	1 RU	1 RU
Uplink Ports	4 x 10GbE SFP+	2 x 10GbE SFP+	4 x 10GbE SFP+	8 x 10GbE SFP+
Uplink Transceivers Supported	Copper CX-1 (passive): 1m, 3m, 5m. Optical: FET (Nexus 2200 platforms), SR, LR [distance limited to 300m]			
Host Facing Ports	48 x 1GbE RJ45 (note: 1000BaseT only)	24 x 100/1000Base-T RJ45	48 x 100/1000Base-T RJ45	32 x SFP/SFP+ (1/10G) (note: 1GE SFP support in 4.2(1)N2(1))
FCoE	N/A	N/A	N/A	Yes
Dimensions	1.72 x 17.3 x 20.0 in	1.72 x 17.3 x 17.7in	1.72 x 17.3 x 17.7in	1.72 x 17.3 x 17.7 in
Operational Power	165W	95W	110W	270W
Supports FET	No	Yes	Yes	Yes
Multiple PortChannel member ports on a FEX	Not Supported	Yes	Yes	Yes
Scalability	576 GbE Ports per N5K (12 FEX)	384 GbE Ports per N5K (16 FEX)	768 GbE Ports per N5K (16 FEX) 1536 GbE Ports per N7K	512 1/10GbE Ports per N5K (16 FEX)

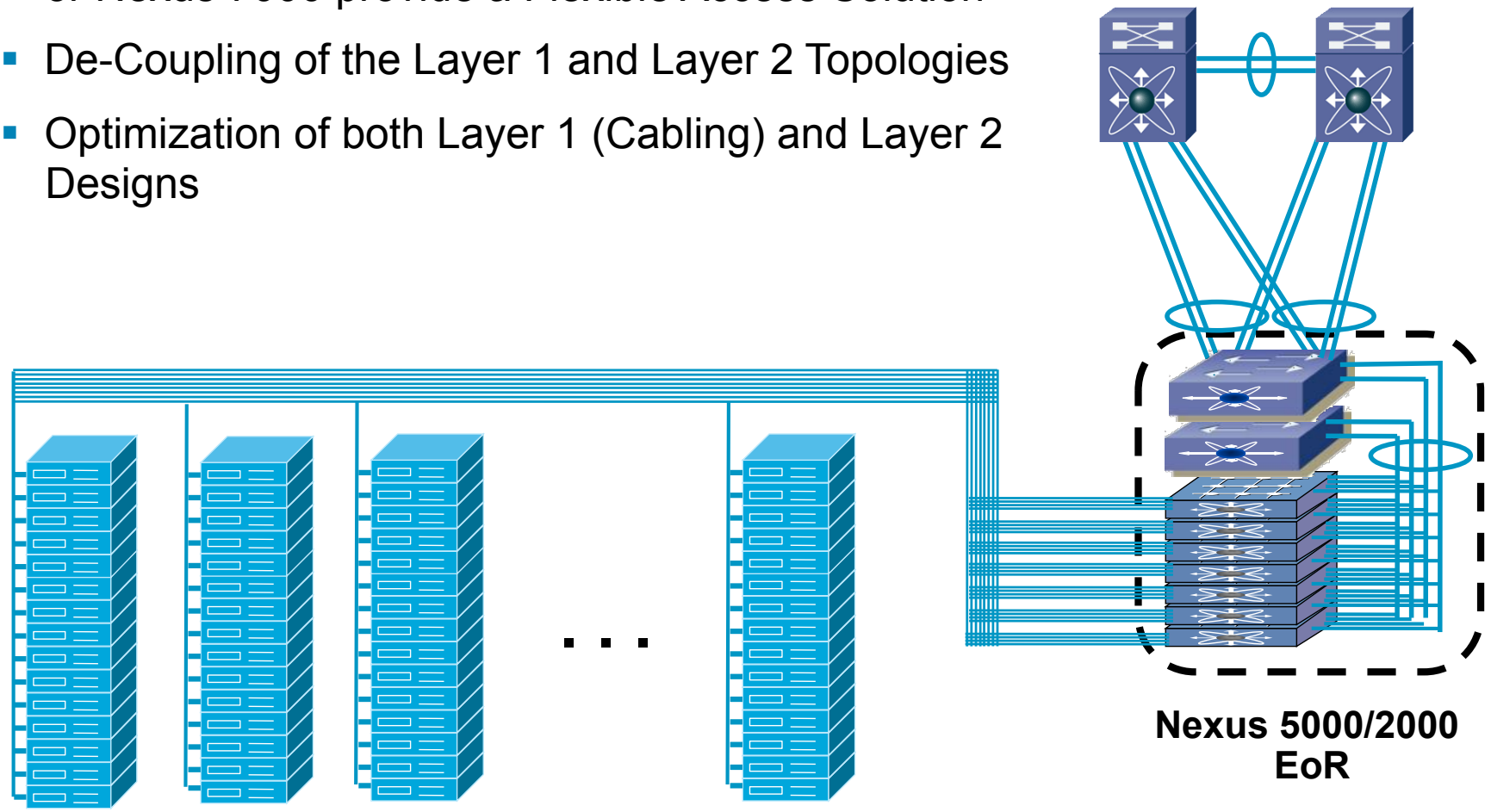
Nexus 2000 Parent Switch Options

Connects to both Nexus 5000 and Nexus 7000

Parent Switch	FEX Capable	# FEX	Optics/Transceivers Supported for FEX
Nexus 5010  Nexus 5020 	All shipping FEX models	12	Passive CX-1 SFP+ (1/3/5m) Active CX-1 SFP+ (7/10m) SR SFP+ (MMF) – OM3 300m LR SFP+ (SMF) – 300m (FCoE no drop) FET SFP+ (MMF) – OM2 20m, OM3 100m
Nexus 5548  Nexus 5596  Q1CY2011	All shipping FEX models	16	Passive CX-1 SFP+ (1/3/5m) Active CX-1 SFP+ (7/10m) SR SFP+ (MMF) – OM3 300m LR SFP+ (SMF) – 300m (FCoE no drop) FET SFP+ (MMF) – OM2 20m, OM3 100m
N7K-M132XP-12  (Nexus 7000) N7K-M132XP-12L 	Nexus 2248 Nexus 2232 (Future)	32	Passive CX-1 SFP+ (1m/3m/5m): On M132XP-12 Active CX-1 SFP+ (7m/10m) SR SFP+ (MMF): OM1 26m OM3 300m LR SFP+ (SMF): up to 10km FET SFP+ (MMF): OM2 25m OM3 100m
N7K-M108XP-12  (Nexus 7000) N7K-F132XP-15 		-	FEX not supported

FEXlink: Optimizing L1 and L2 Environments

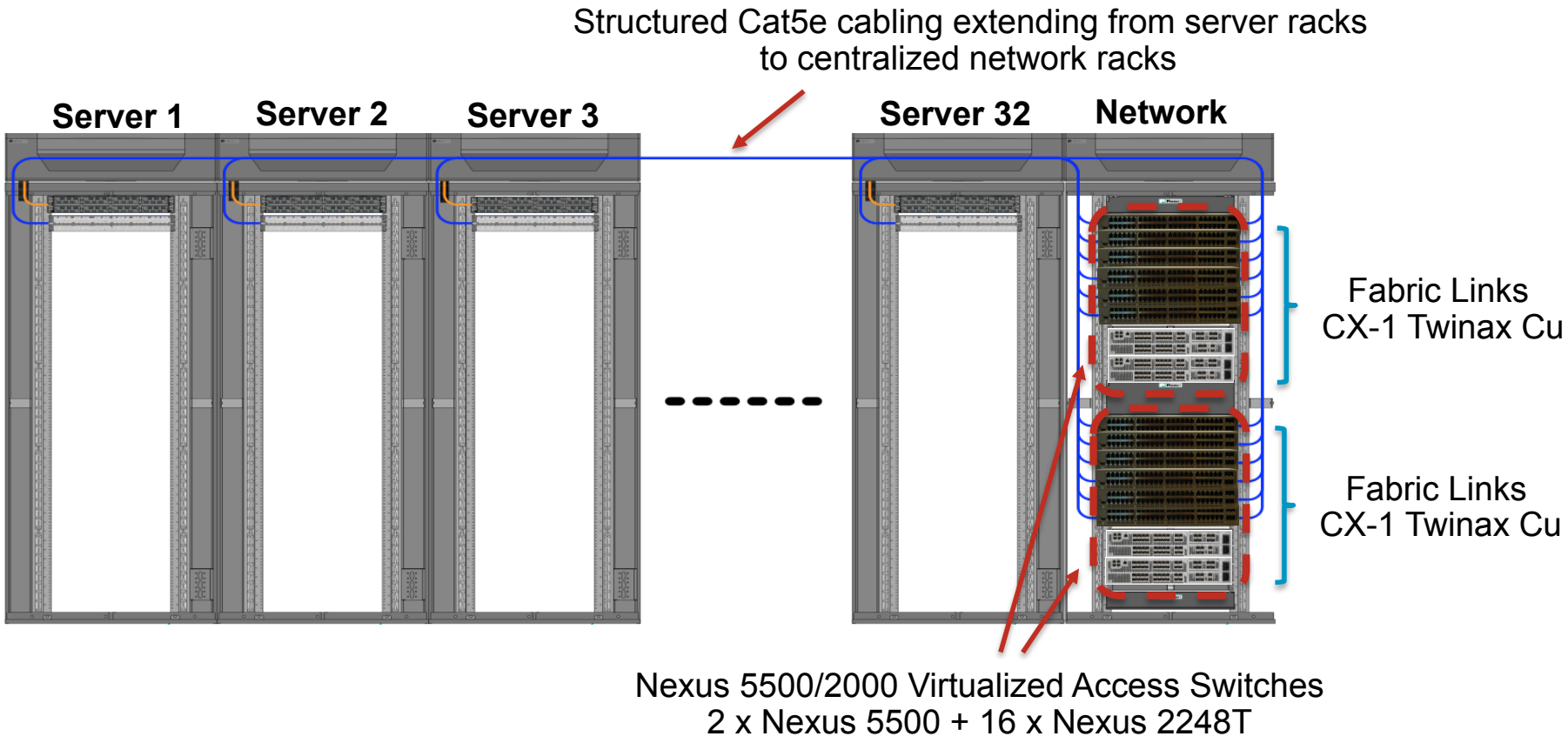
- Cisco Nexus 2000 Fabric Extender and Nexus 5000 or Nexus 7000 provide a Flexible Access Solution
- De-Coupling of the Layer 1 and Layer 2 Topologies
- Optimization of both Layer 1 (Cabling) and Layer 2 Designs



1GE Attached Servers - Maintain Existing Cat5e Server Wiring Infrastructure with EoR topology

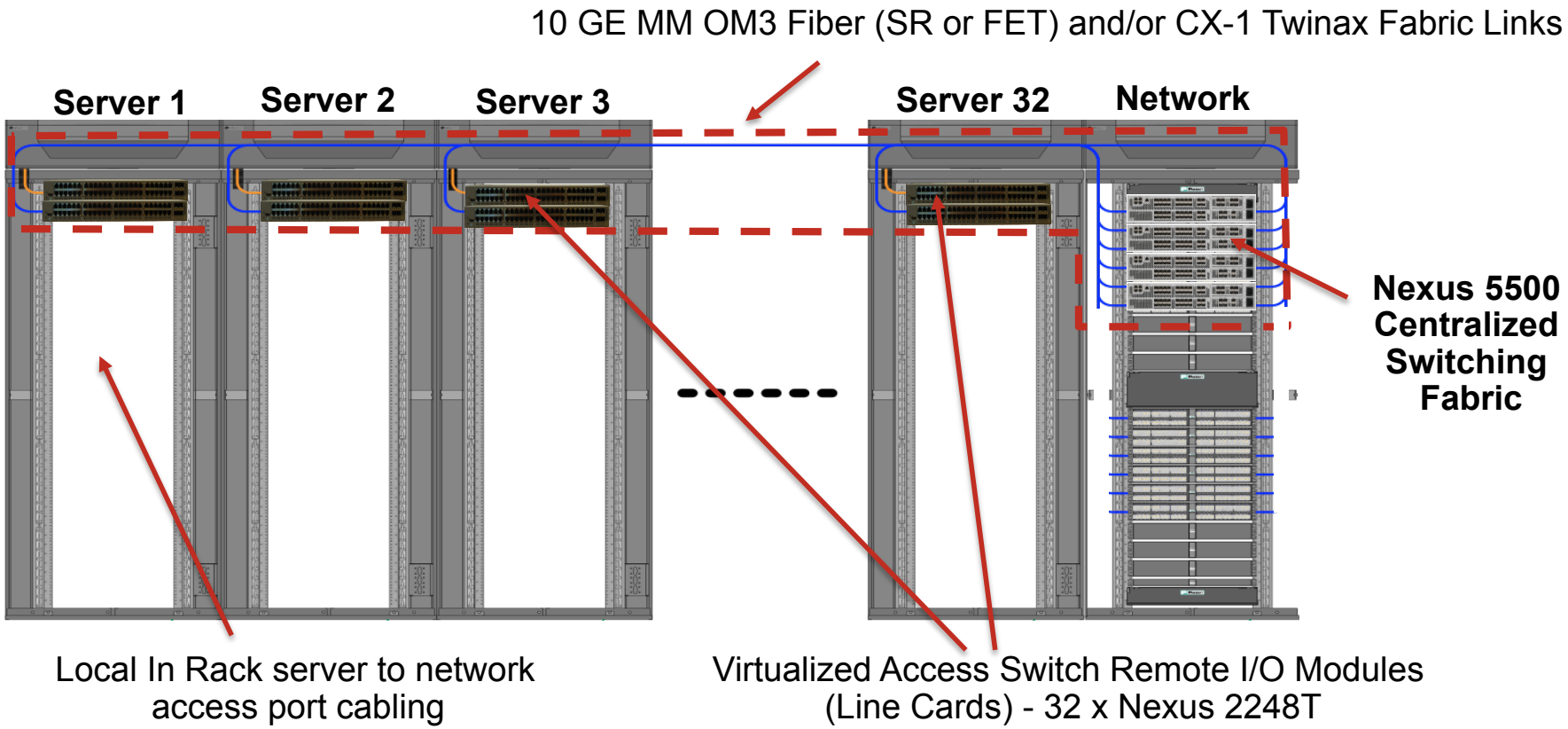
Physical Pod (end of row) Topology

- 768 Server ports per Virtualized Access Switch – One managed node
- Dense: Only 18 Rack Units required to support 768 x 1GE + 96 x 10GE ports (Fabric ports will be allocated out of the 96 x 10GE ports)
- Architecture accommodates End of Row centralized cabling where required



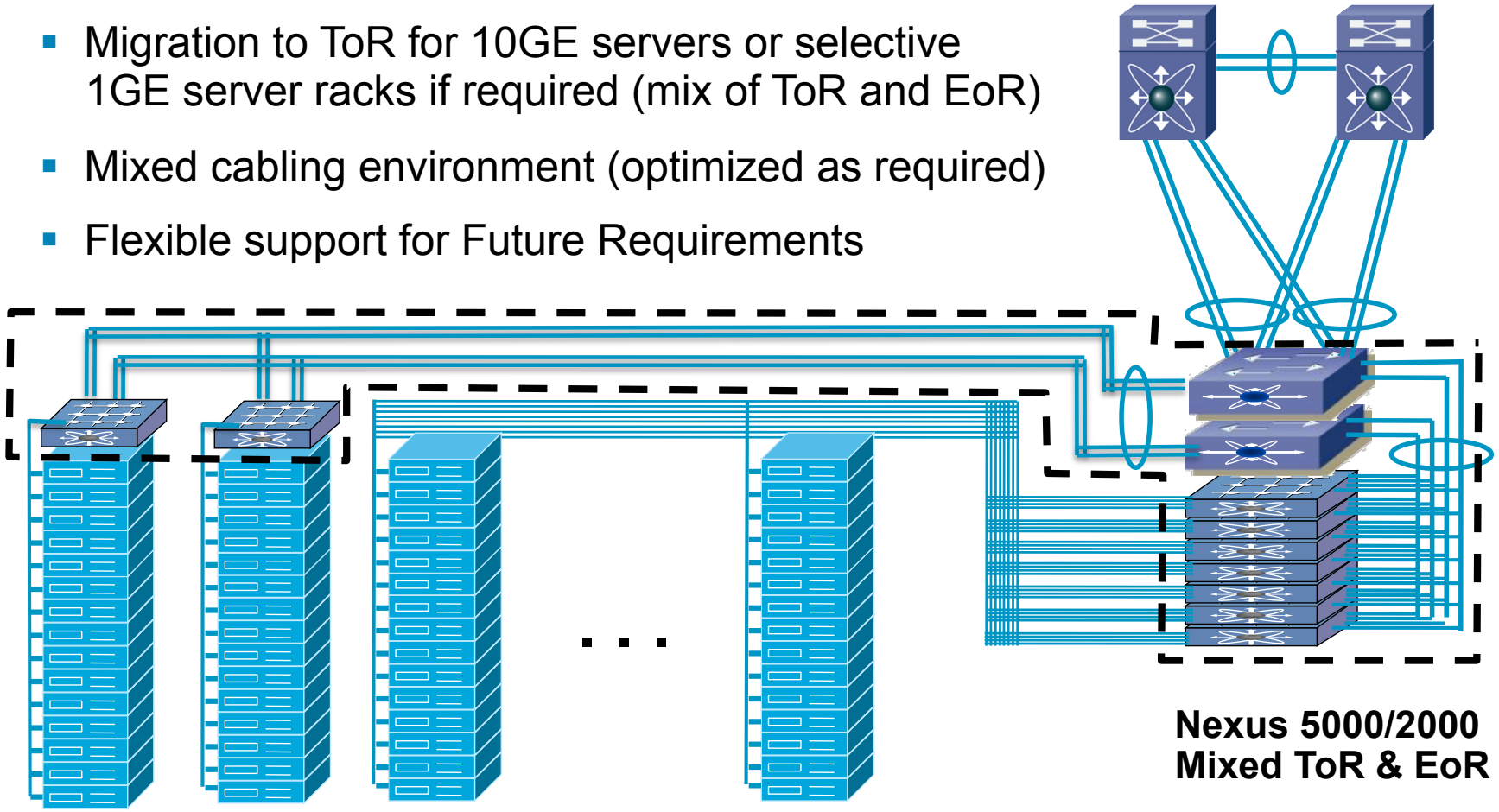
Physical PoD (top of rack) Topology

- Virtualized Access Switch architecture supports ToR cabling plant where required
- A single managed device (one node and one set of uplinks to the aggregation)
- L2 redundancy does not use spanning tree



N5K/N2K Advantages – Flexible Cabling

- Cisco Nexus Fabric Extender (FEX) and Nexus 5000 provide a Flexible Access Solution
- Migration to ToR for 10GE servers or selective 1GE server racks if required (mix of ToR and EoR)
- Mixed cabling environment (optimized as required)
- Flexible support for Future Requirements

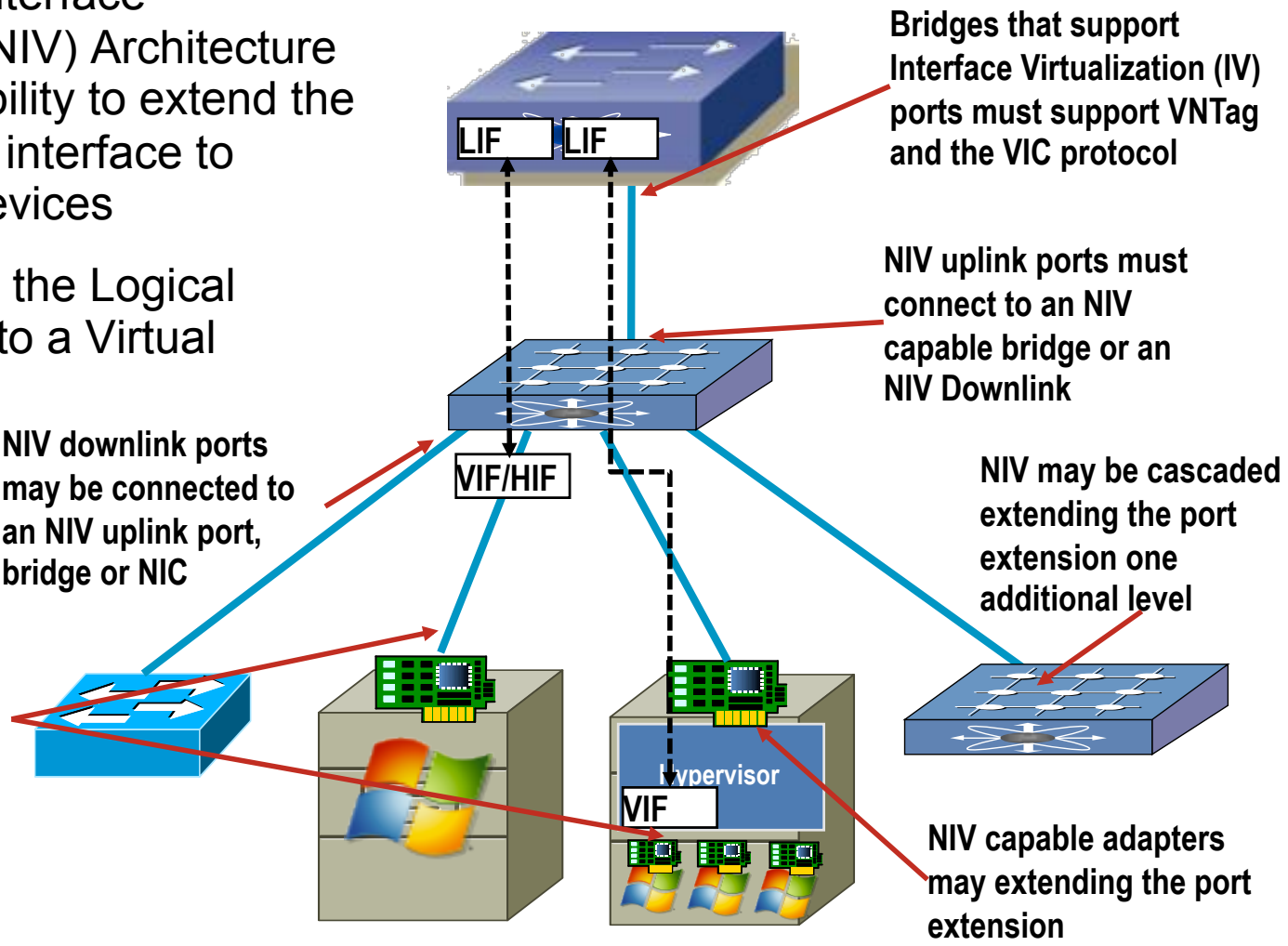


Combination of EoR and ToR cabling

Cisco FEXLink

Network Interface Virtualization Architecture (NIV)

- The Network Interface Virtualization (NIV) Architecture provides the ability to extend the bridge (switch) interface to downstream devices
- NIV associates the Logical Interface (LIF) to a Virtual Interface (VIF)

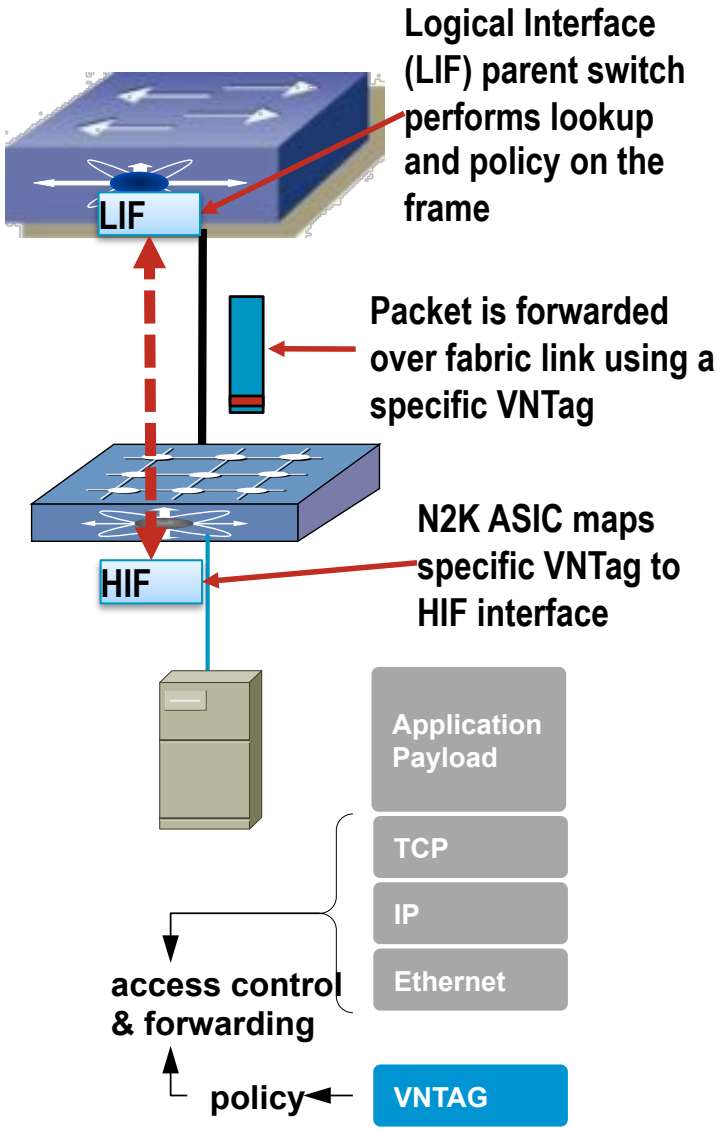


Note: Not All Designs Supported in the NIV Architecture Are Currently Implemented

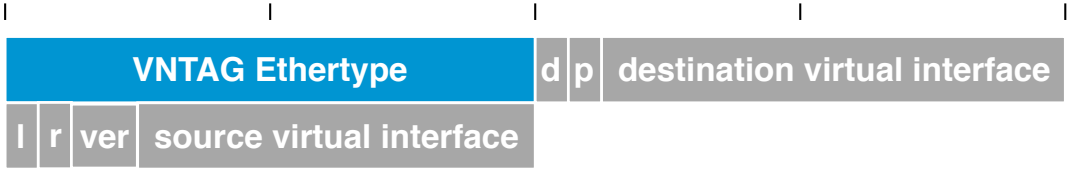
Cisco FEXLink

Network Interface Virtualization - Port Extension

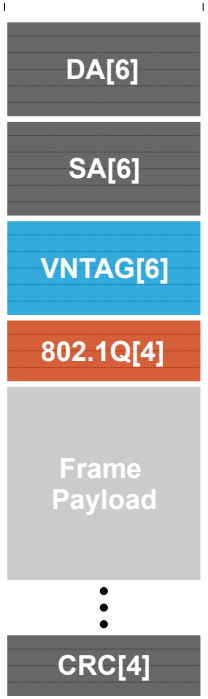
- **Control and Management:** Nexus 2000 Fabric Extender operates as a remote line card (local CPU with protocol offload)
- **Data Plane:** Forwarding is performed on the Parent Switch ASIC's
- VNTag is a Network Interface Virtualization (NIV) technology that 'extends' the Parent switch port (Logical Interface = LIF) down to the Nexus 2000 (Host Interface = HIF)
  - VNTag is added to the packet between Fabric Extender and Parent Switch
  - Link Local State only – Internal switch header
- Port Extension allows the Fabric Extender to act as a data path extension of Parent Switch



VNTAG Overview



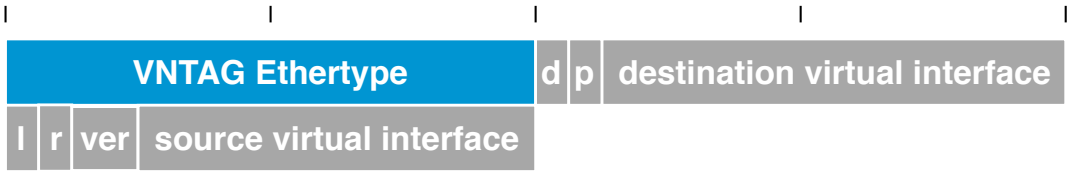
- direction indicates to/from adapter
- source virtual interface indicates frame source
  - looped indicates frame came back to source adapter
- destination virtual interface dictates forwarding
  - pointer helps pick specific destination vNIC or vNIC list
- Link local scope
  - Rooted at Virtual Interface Switch
  - 4096 virtual interfaces
  - 16,384 Virtual interface lists
- Coexists with VLAN (802.1Q) tag
  - 802.1Q tag is **mandatory** to signal data path priority

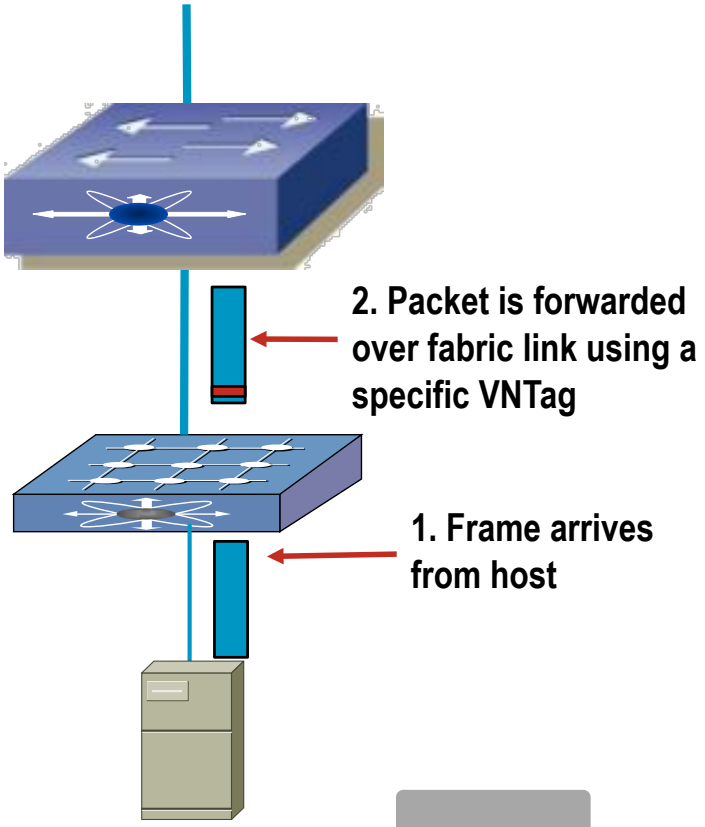


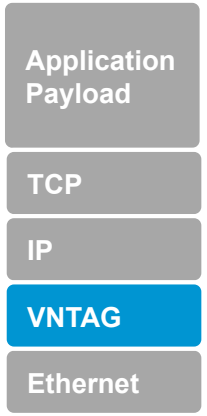
Nexus 2000 Fabric Extender

Host-to-Network Forwarding Part 1

- Nexus 2000 adds VNTAG
  - Unique VNTag for each Nexus 2000 Host Interface (HIF)
- VNTag field values
  - Direction bit is set to 0 indicating host to network forwarding
  - Source Virtual Interface is set based on the ingress HIF
  - p (pointer), l (looped), and destination virtual interface are undefined (0)
- Frame is unconditionally sent to the Nexus 5000



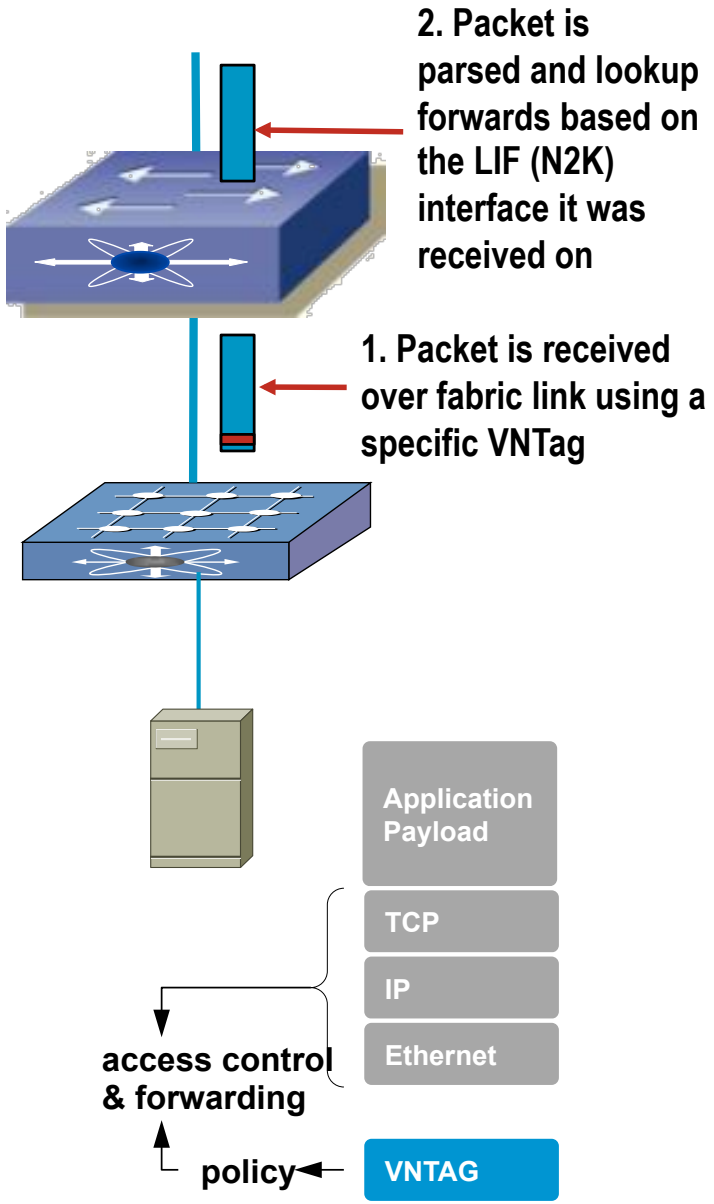




Nexus 2000 Fabric Extender

Host-to-Network Forwarding Part 2

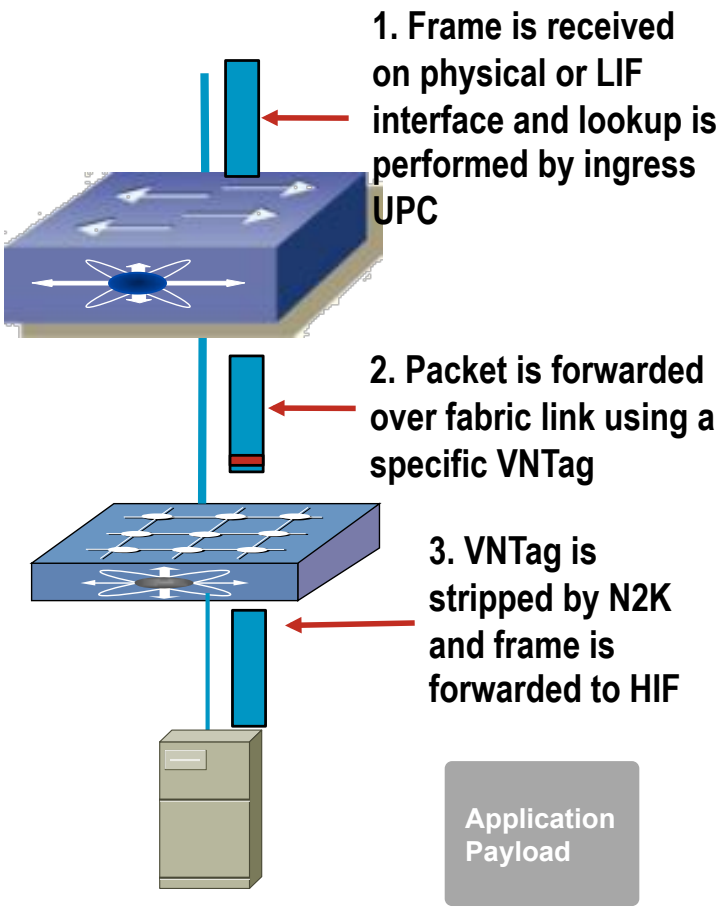
- Nexus 5000 ingress processing on fabric ports
- UPC extracts VNTAG which identifies the Logical Interface (LIF) corresponding to the physical HIF on the actual Nexus 2000
- Ingress policy based on physical Nexus 5000 port and LIF
  - Access control and forwarding based on frame fields and virtual interface (LIF) policy
  - Physical link level properties are based on the Nexus 5000 port
- Forwarding selects destination port(s) and/or destination virtual interface(s)

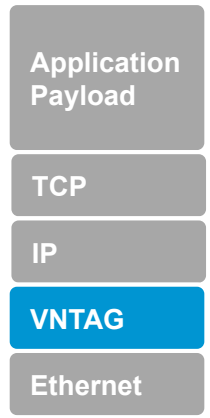


Nexus 2000 Fabric Extender

Network to Host Forwarding

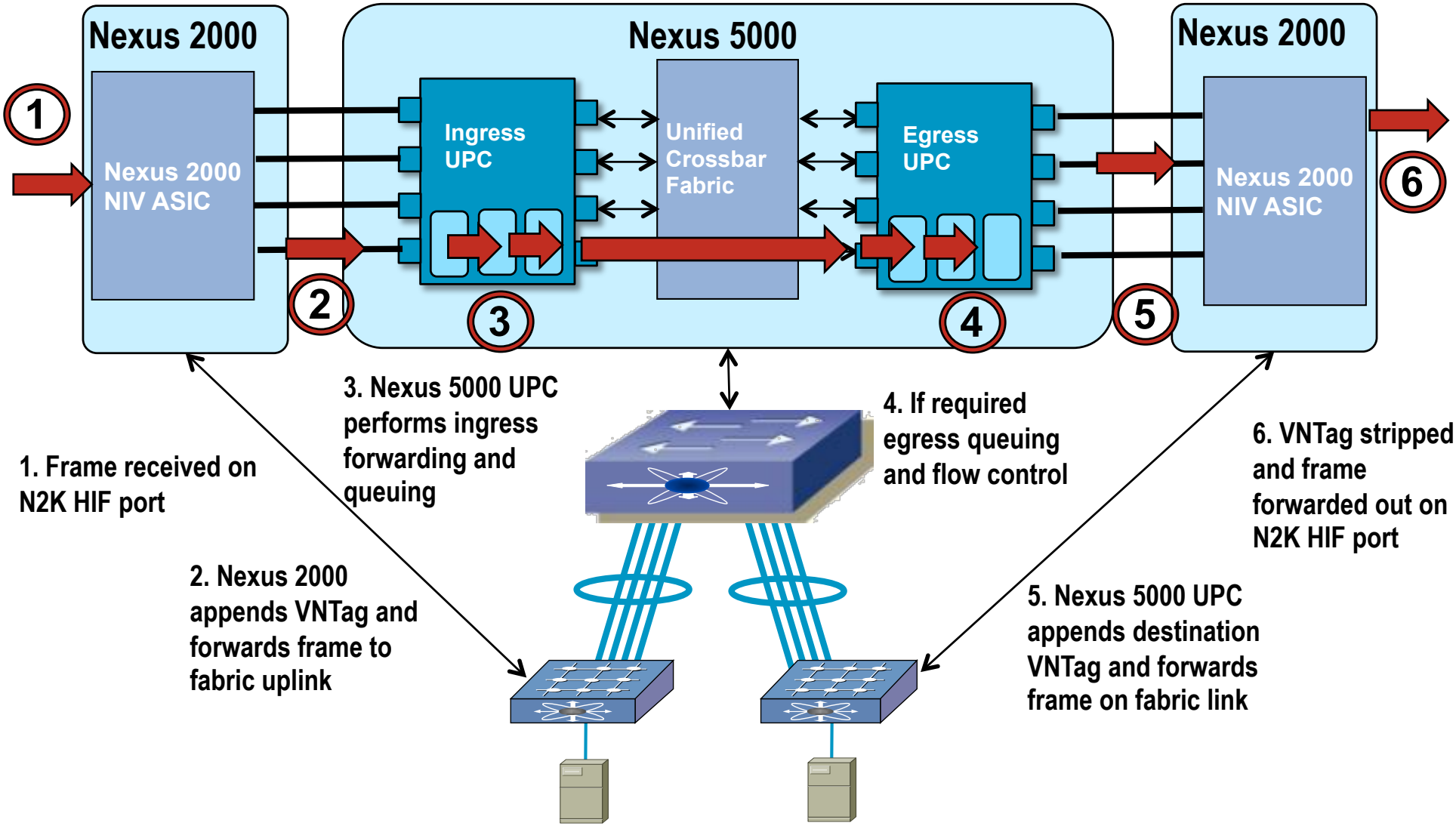
- Nexus 5000/7000 performs standard lookup and policy processing, when the egress port is determined to be an LIF (Nexus 2000) port
  - Insert VNTAG with direction is set to 1 (network to host)
  - Destination virtual interface is set to be the Nexus 2000 port VNTag
  - Source virtual interface is set if packet was sourced from an N2K port
  - L bit (looped) filter set if sending back to a source N2K
  - P bit is set if this is a multicast frame and requires egress replication





Nexus 5000 and 2000

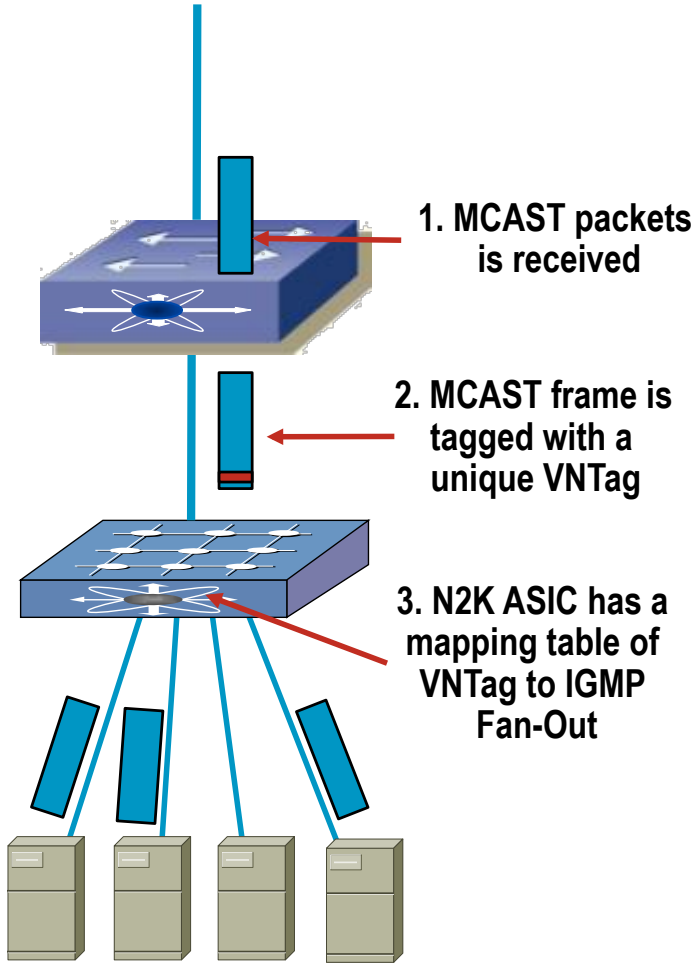
Packet Forwarding Overview



Nexus Virtualized Access Switch

Nexus 2000 Multicast Forwarding

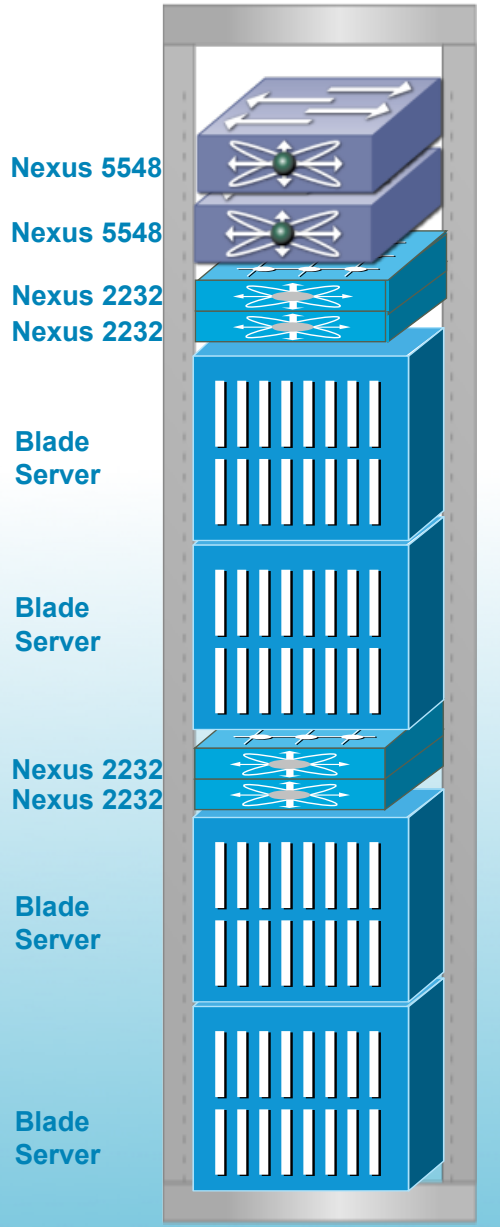
- Nexus 2000 supports egress based Multicast replication
- A single copy of the multicast packet is sent down the fabric link
- Each fabric link has a list of VNTag's associated with each Multicast group
- A single copy of each multicast frame is sent down the fabric links to the Nexus 2000
- Extended Multicast VNTag has an associated flooding fan-out on the Nexus 2000 built via IGMP Snooping
- Nexus 2000 replicates and floods the multicast packet to the required interfaces



Nexus 2232 + Pass-Thru Architecture

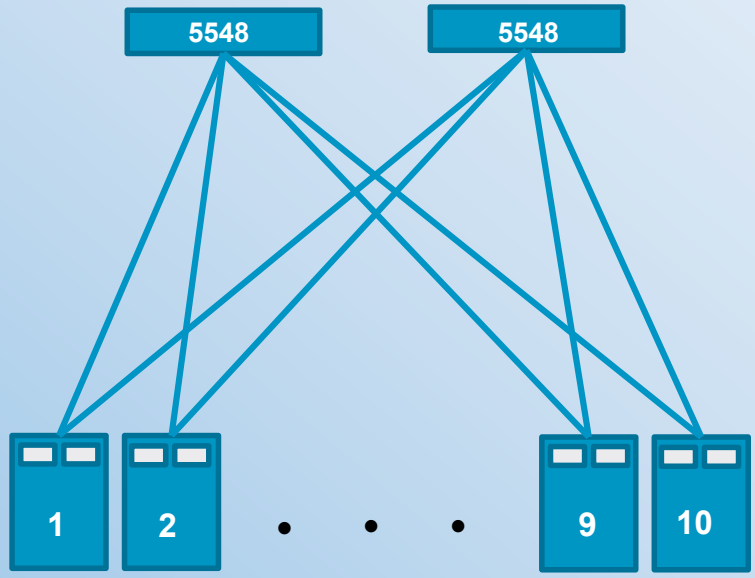
Simple & Scalable Blade Server Environments

- **Simplicity**
  - Zero configuration within blade chassis; No Server Admin training
  - Dramatically fewer network elements to manage
  - Network visibility to the server/VM & Superior trouble shooting
- **Flexibility**
  - Ethernet, Unified Fabric & VM Server access (& phased migration)
  - Consistent server access architecture for blade & rack mount servers across DC and with single server POD
  - TOR or EOR access model & benefits of both
- **Advanced Network Capability**
  - Field validated FCoE & DCB support
  - Transparent support for server virtualization via VNLink
  - Data Center-class NX-OS design & In-Service Software Upgrades
  - Comprehensive L2 feature set including vPC



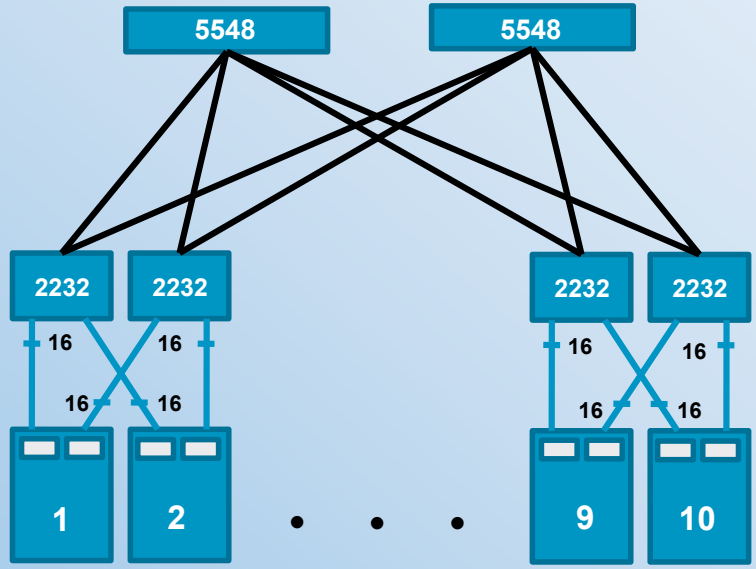
Fewer Network Elements - Ethernet

Blade Switch Topology



10 Chassis/160 Servers/4:1 Oversubscription

Pass Through to Nexus 2232 Topology

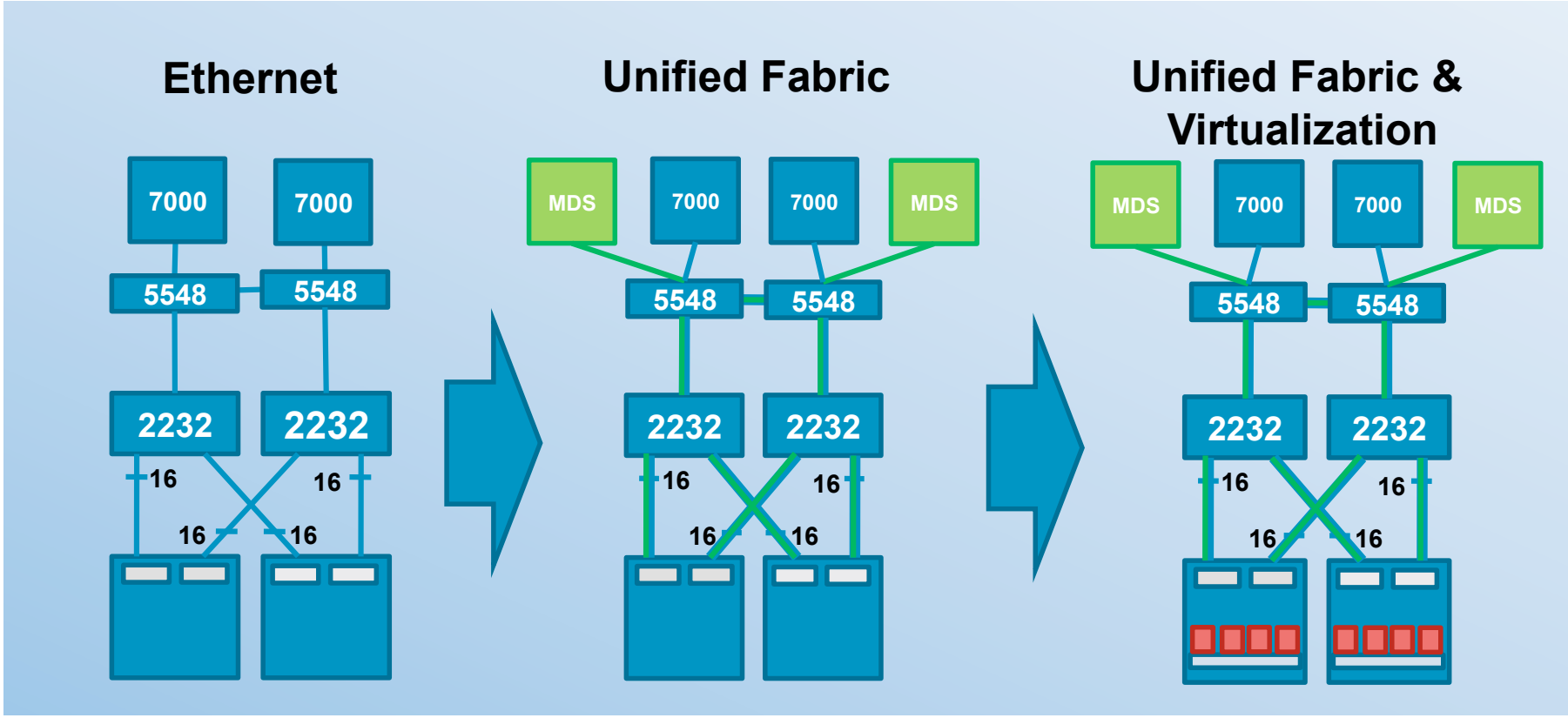


10 Chassis/160 Servers/4:1 Oversubscription

	Ethernet Switches
Pass-Thru	2
Blade Switches	22

- No Blade Switch supports ISSU
- Upgrading the network with FEX +PT design takes no downtime!

Seamless Migration Path



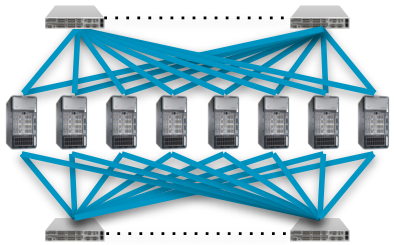
	10Gb Pass Through	Blade Switch Alternative
1/10GbE Support	Yes	Yes (some vendors)
Unified Fabric Support	Yes	HW Upgrade Required
VNLink SW Support	Yes	Minimal
VNLink HW Support	Yes	HW Upgrade Required

Nexus Data Center Access Technologies

Distributed Access Fabric

- Fabric Extender Technology
- 100M/1G/10G Capable
- FCoE & VNTag Support

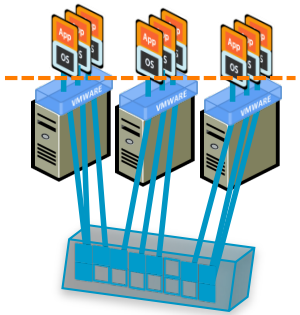
- 10:1 reduction in managed devices
- No cross-rack cabling
- Form-Factor & MGMT optimized



Scalable Access Choice

- Virtual Port Channel (vPC)
- FabricPath/Trill
- Routed Access

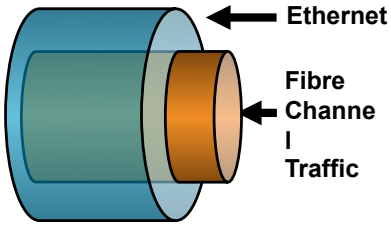
- Accelerate IO intensive applications
- Eliminate logical Layer 2 loops
- Massive bisectional bandwidth



Virtual Networking

- VN-Link
- VN-Services
- VN-Manager

- Extend networking principles to virtualized environments
- Extend network services to virtualized environments
- Enable consistent VN-wide management



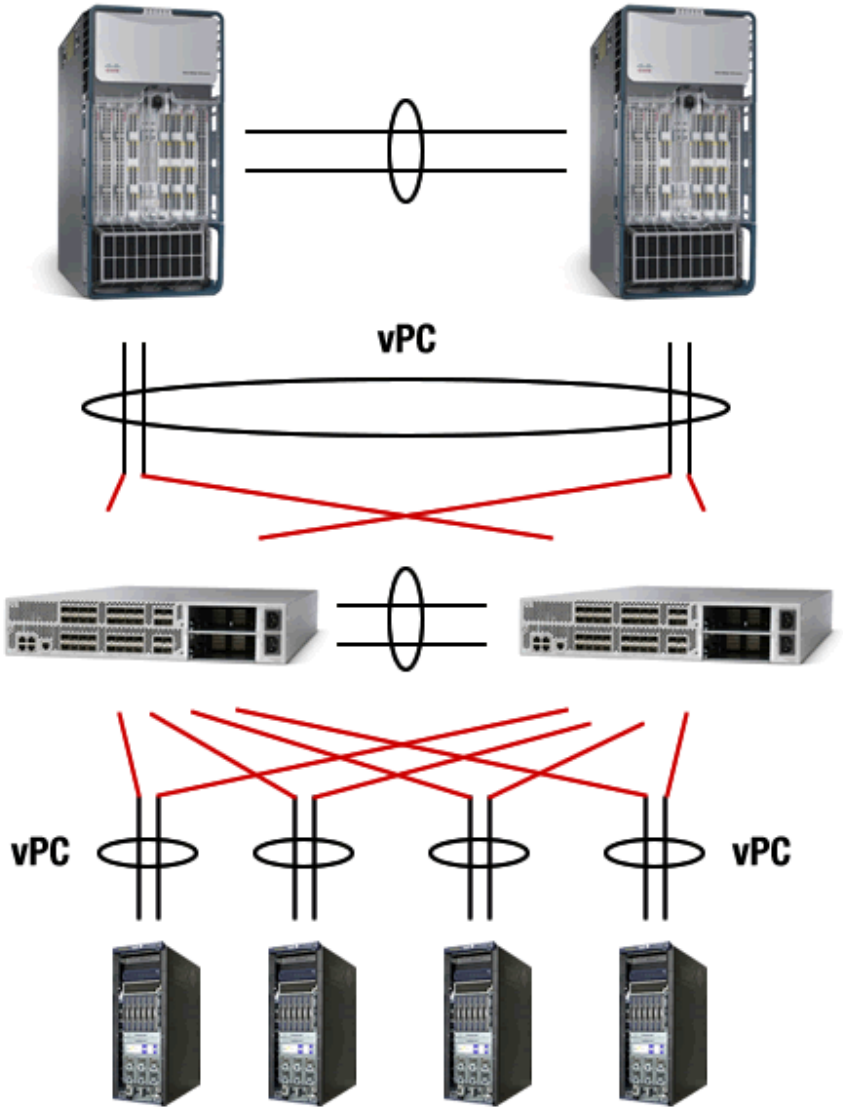
Unified Fabric

- LAN/SAN Consolidation
- Converged Network Adapters (CNAs)
- High Performance Server Interconnect

- 50% or more cable reduction
- 50% or more network/SAN device reduction
- Maintains logical LAN/ SAN mgmt model

Powered by Cisco NX-OS Data Center Operating System

L2 Redundancy Options – VPC



Virtual Port Channels

Virtual Port Channels (vPC allows for the creation of Port channels across multiple chassis

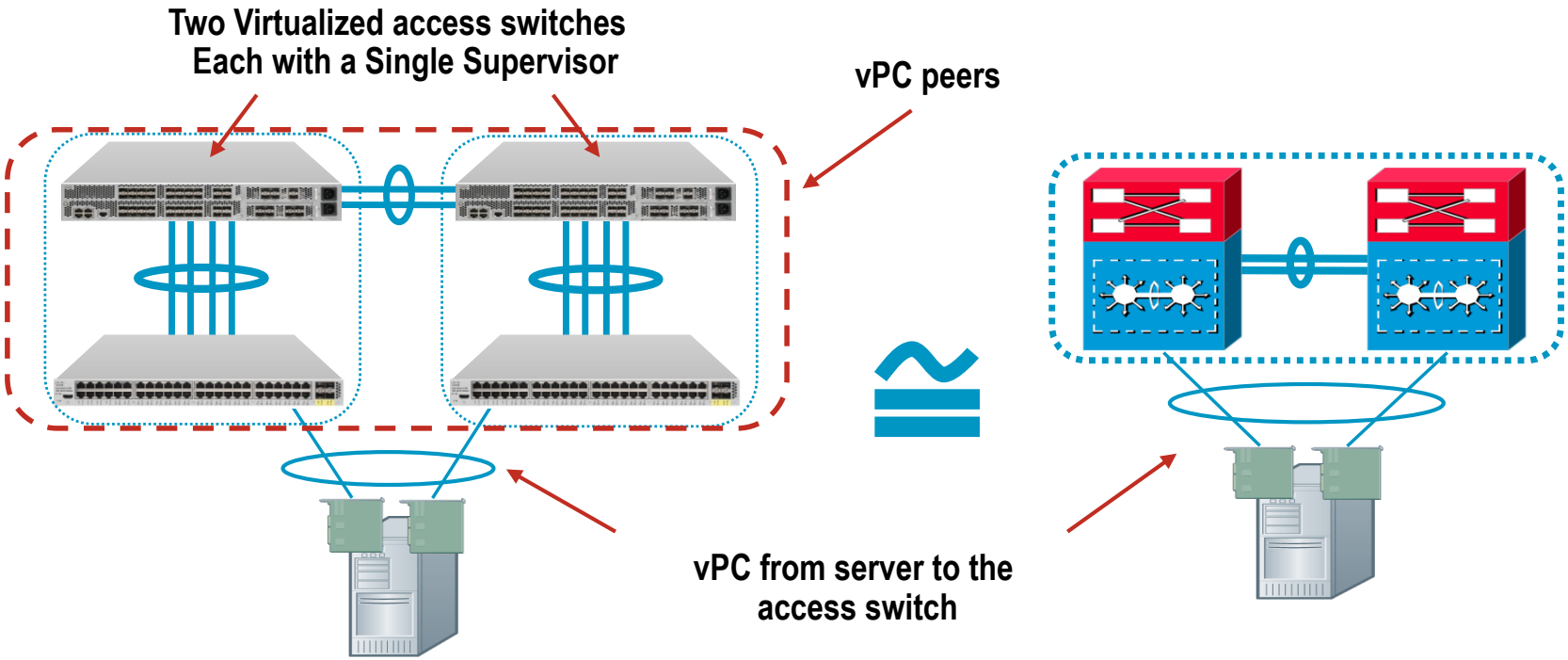
LACP or static configuration may be leveraged to form the vPC

May be used to form Port Channels to Rack Servers or Blade chassis

Compatible with vPC on the Nexus 7000 as well as VSS on the Catalyst 6500

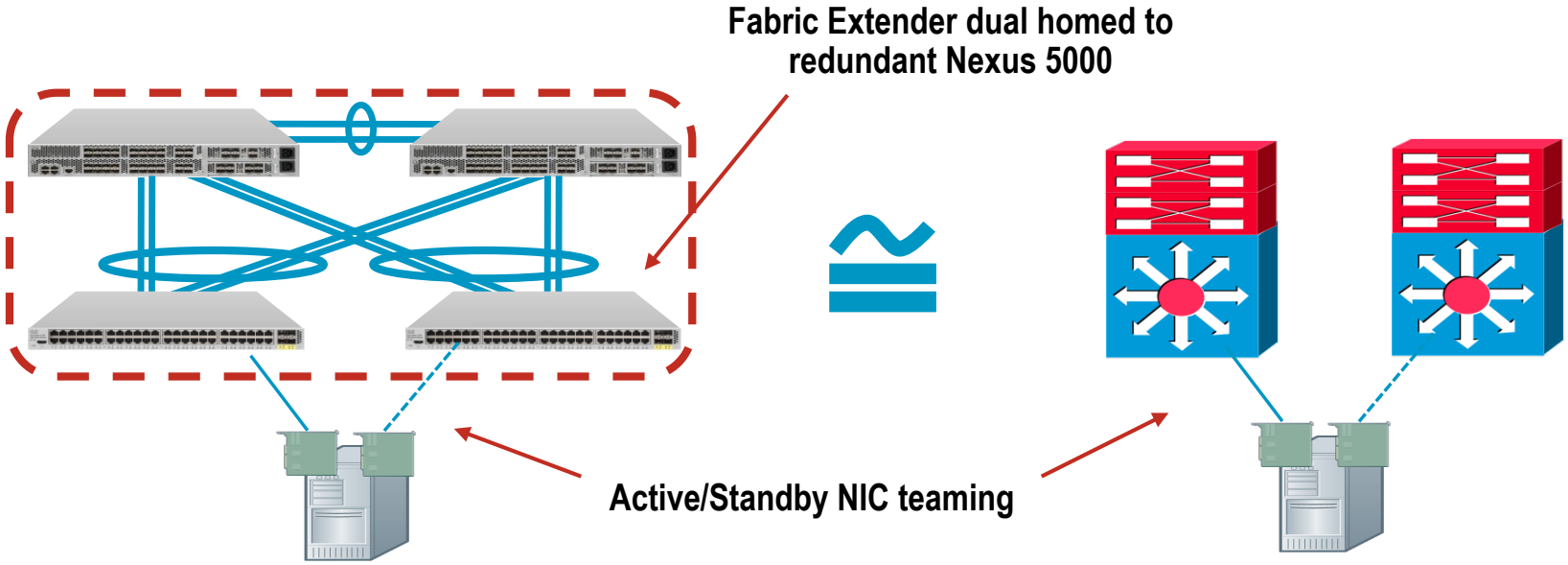
vPC Redundancy Models – Dual Chassis

- vPC provides two redundancy designs for the virtualized access switch
- Option 1 – vPC connectivity from the server
  - Two virtualized access switches bundled into a vPC pair
  - Full redundancy for supervisor, line card, cable or NIC failure
  - Logically a similar HA model to that currently provided by VSS



vPC Redundancy Models – Dual Supervisor

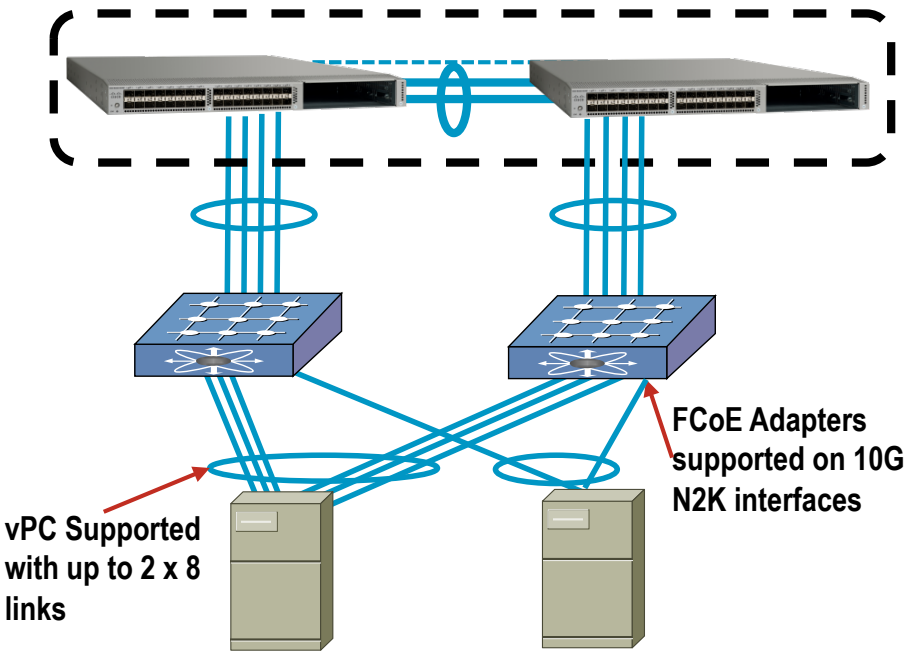
- vPC Option 2 – Fabric Extender connected to two Nexus 5000
  - From the server perspective a single access switch with each line card supported by *redundant supervisors*
  - Full redundancy for supervisor, fabric via vPC and cable or NIC failure via active/standby NIC redundancy
  - Logically a similar HA model to that currently provided by dual supervisor based modular switch



Nexus 2000 Design Options

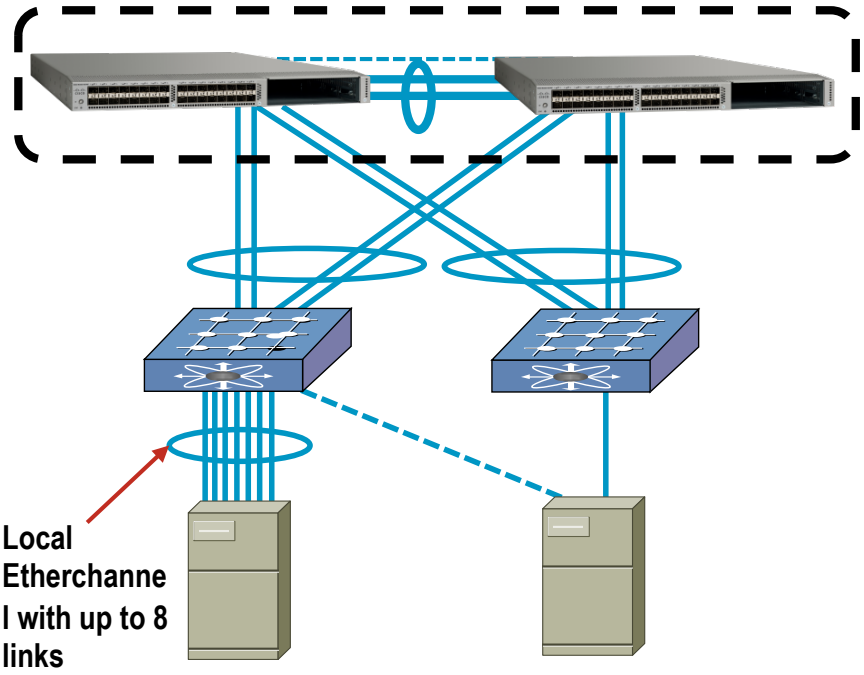
Nexus 5000 Topologies (Nexus 2148TP, 2224TP, 2248TP & 2232PP)

Straight Through



- Redundancy model – Dual Switch with redundant fabric
- Provides isolation for Storage topologies (SAN 'A' and 'B')
- Port Channel and Pinning supported for Fabric Link

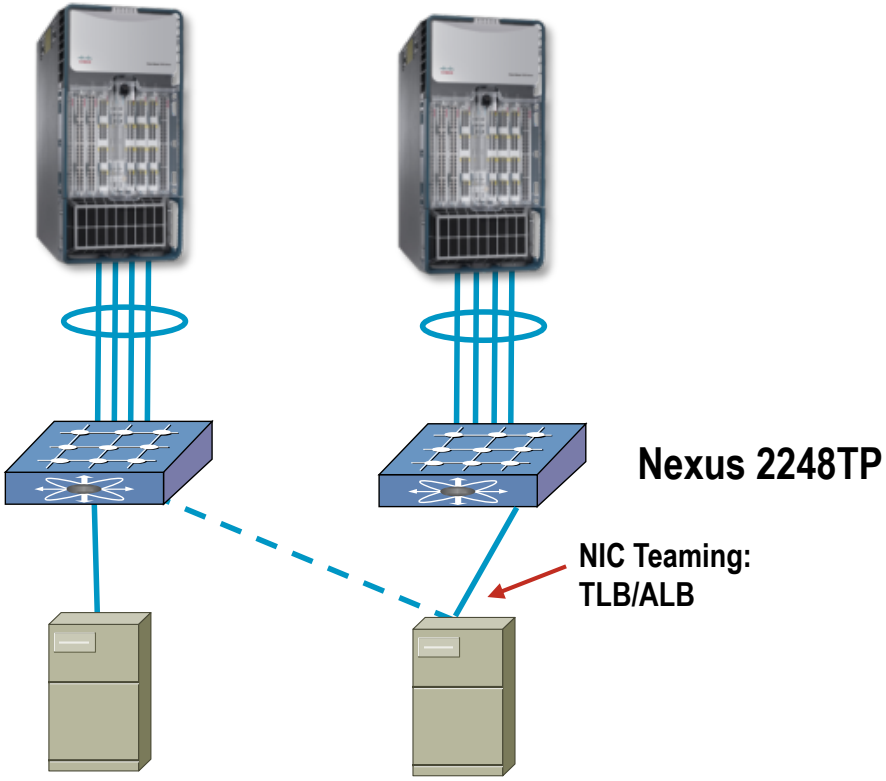
Dual Homed



- Redundancy model – Single switch with dual 'supervisor' for fabric, data control & management planes
- No SAN 'A' and 'B' isolation

Nexus 2000 Design Options

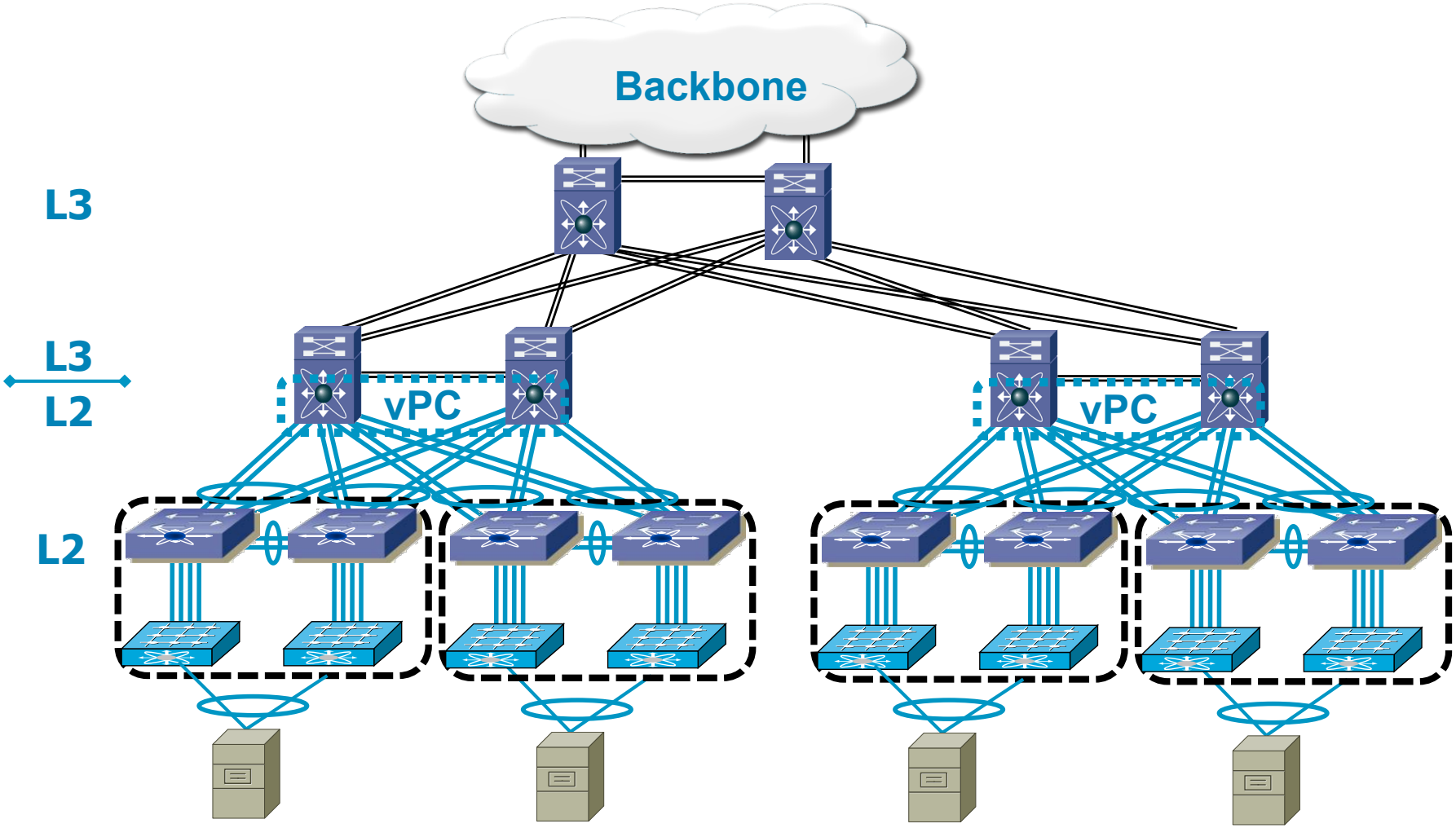
Nexus 7000 Topologies (Nexus 2248TP)



- Fabric links supported on N7K-M132XP-12 & N7K-M132XP-12L
- Port Channel only supported for Fabric Links
- Must leverage NIC Teaming to redundantly connect servers to different switches
- Allows a deployment model with a single layer for access/distribution

vPC and Fabric Extender

Loop free layer 2 topologies



Combine vPC with Fabric Extender to build larger layer 2 domains

(reduced STP logical port count, nodes, …)

L2 Redundancy Options – FabricPath

L2 Redundancy with L3 scalability

Up to 16 active spines/paths

Massive Scalability

Already supported on Nexus 7000

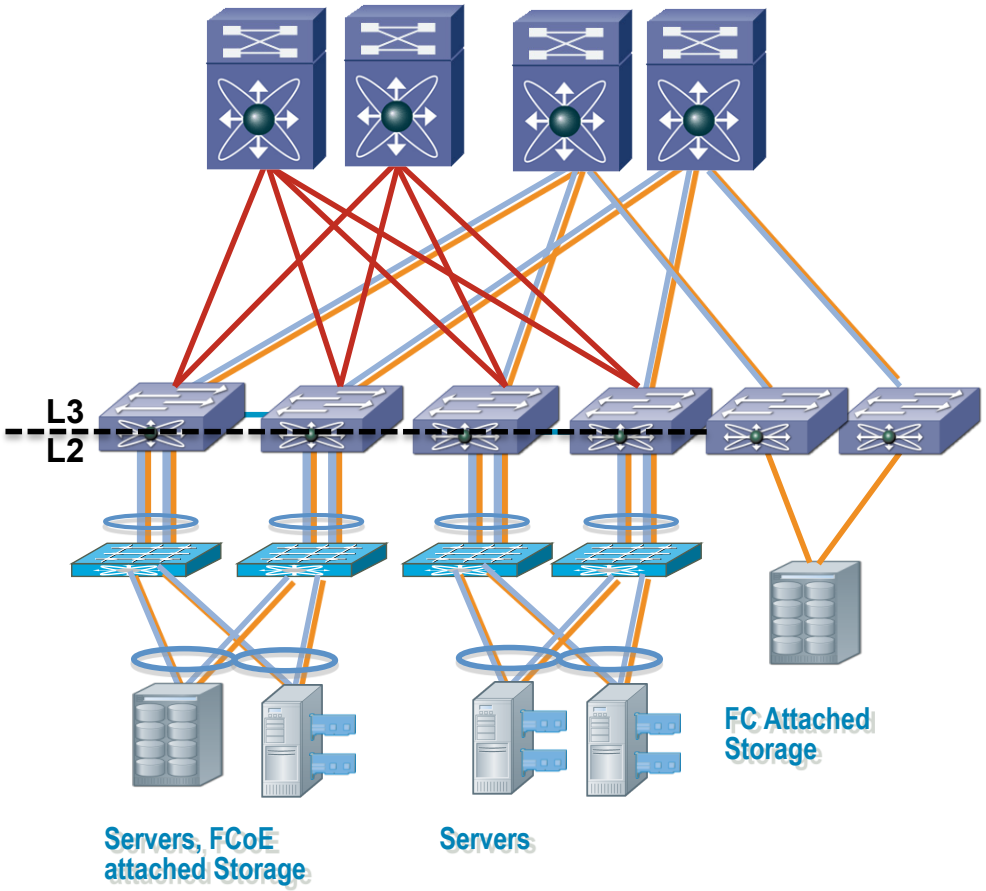
Hardware support on Nexus 5500

Can be combined with FEX

Routed Access Support

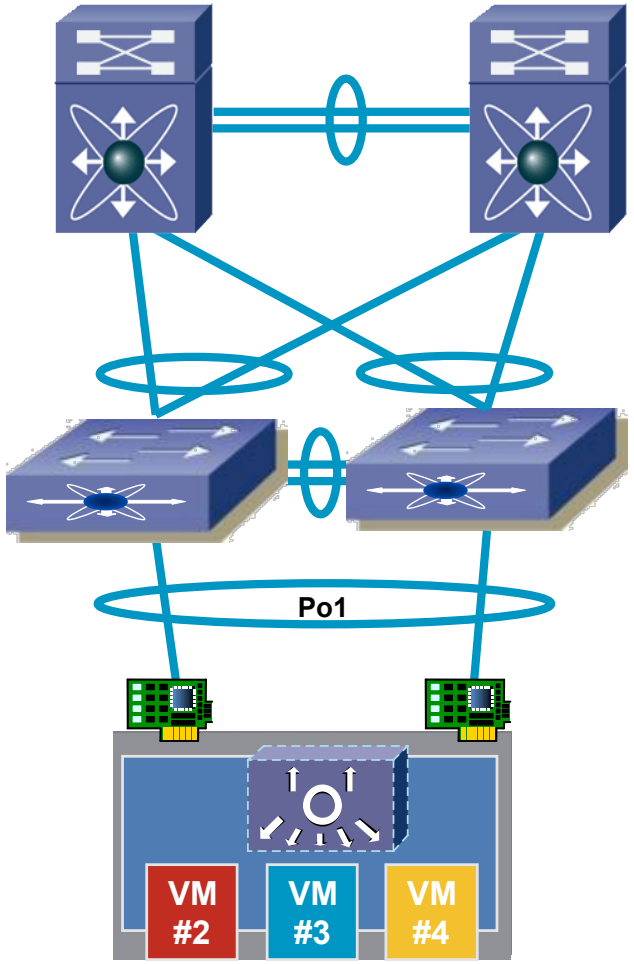
Nexus 5500 & Nexus 2000 – Routed Access (Q1CY11)

- Routed Access Topologies with Nexus 5500 & Nexus 2000 (Q1CY11)
- Supported with all Nexus 2000



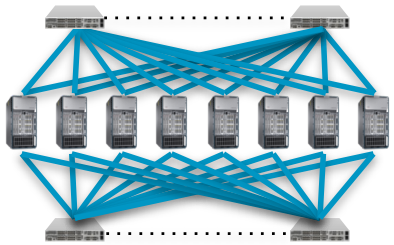
Can I design with two NICs? ESX Needs multiple NICs anyways, doesn't it?

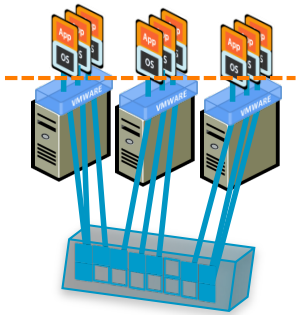
- An ESX Server does *not* need multiple NICs
  - It needs to assure bandwidth and security for different traffic types: vmkernel, vMotion, SAN (FC/FCoE, iSCSI, NFS)
    - This was usually done by having dedicated GE NICs
  - It needs a design which provides redundancy for key services
- With advent of 10GE and specifically DCB, a 2 NIC design is perfectly valid
- N1Kv can also provide QoS for vmkernel, vMotion or iSCSI

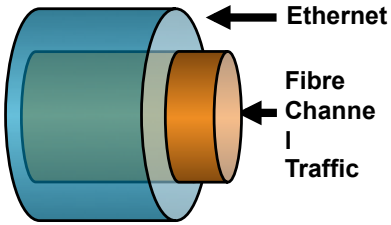


Deploying 10 Gigabit Ethernet on VMware vSphere 4.0
http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9902/white_paper_c07-607716.html

Nexus Data Center Access Technologies







Distributed Access Fabric

- Fabric Extender Technology
- 100M/1G/10G Capable
- FCoE & VNTag Support

- 10:1 reduction in managed devices
- No cross-rack cabling
- Form-Factor & MGMT optimized

Scalable Access Choice

- Virtual Port Channel (vPC)
- FabricPath/Trill
- Routed Access

- Accelerate IO intensive applications
- Eliminate logical Layer 2 loops
- Massive bisectional bandwidth

Virtual Networking

- VN-Link
- VN-Services
- VN-Manager

- Extend networking principles to virtualized environments
- Extend network services to virtualized environments
- Enable consistent VN-wide management

Unified Fabric

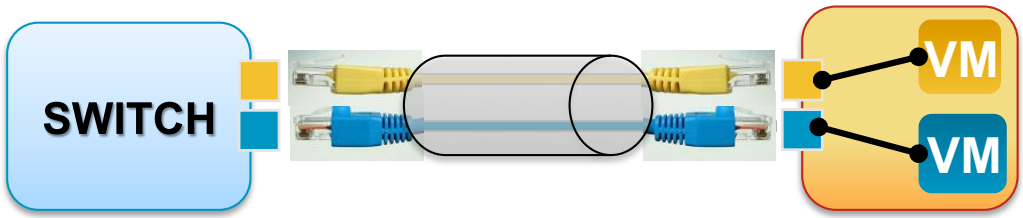
- LAN/SAN Consolidation
- Converged Network Adapters (CNAs)
- High Performance Server Interconnect

- 50% or more cable reduction
- 50% or more network/SAN device reduction
- Maintains logical LAN/SAN mgmt model

Powered by Cisco NX-OS Data Center Operating System

VN-Link – A Virtual Network Link
Removing Networking Barriers to Virtualization

- A virtual network link between the switch and the Virtual Machine
- Extends the network to the virtualization layer
- Enables:
  - Policy-Based VM Connectivity
  - Mobility of Network & Security Properties
  - Non-Disruptive Operational Model



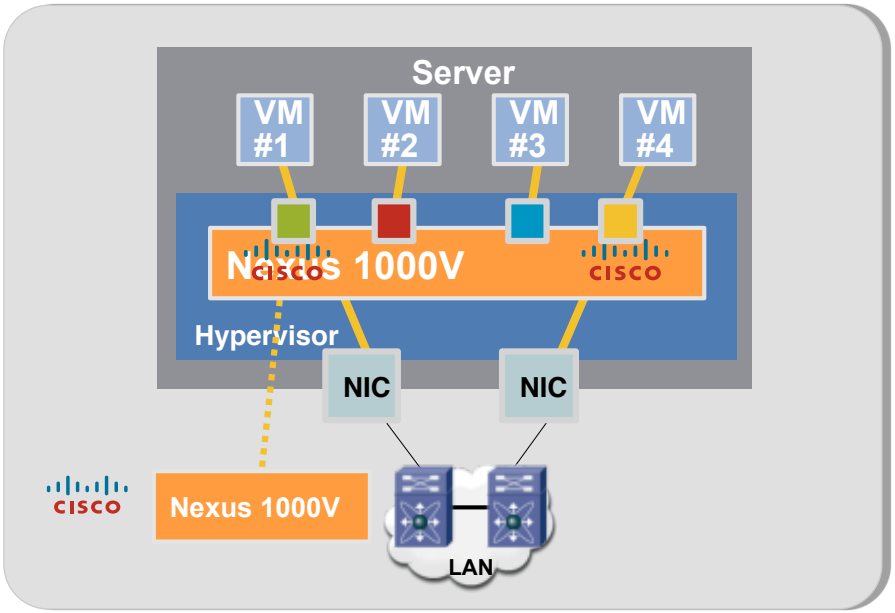
The Scope of VN-Link
Software & Hardware Implementations

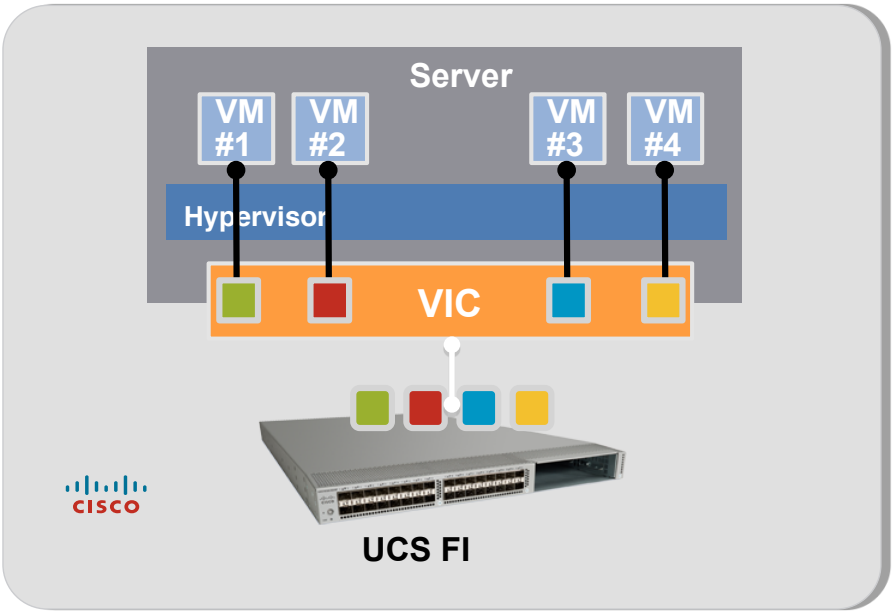
Nexus 1000V

- Software Hypervisor Switching
- Tagless (802.1Q)
- Feature set Flexibility

Nexus 55x0/UCS FI

- External Hardware Switching
- Tag-based (Pre-standard 802.1Qbh)
- Performance Consolidation

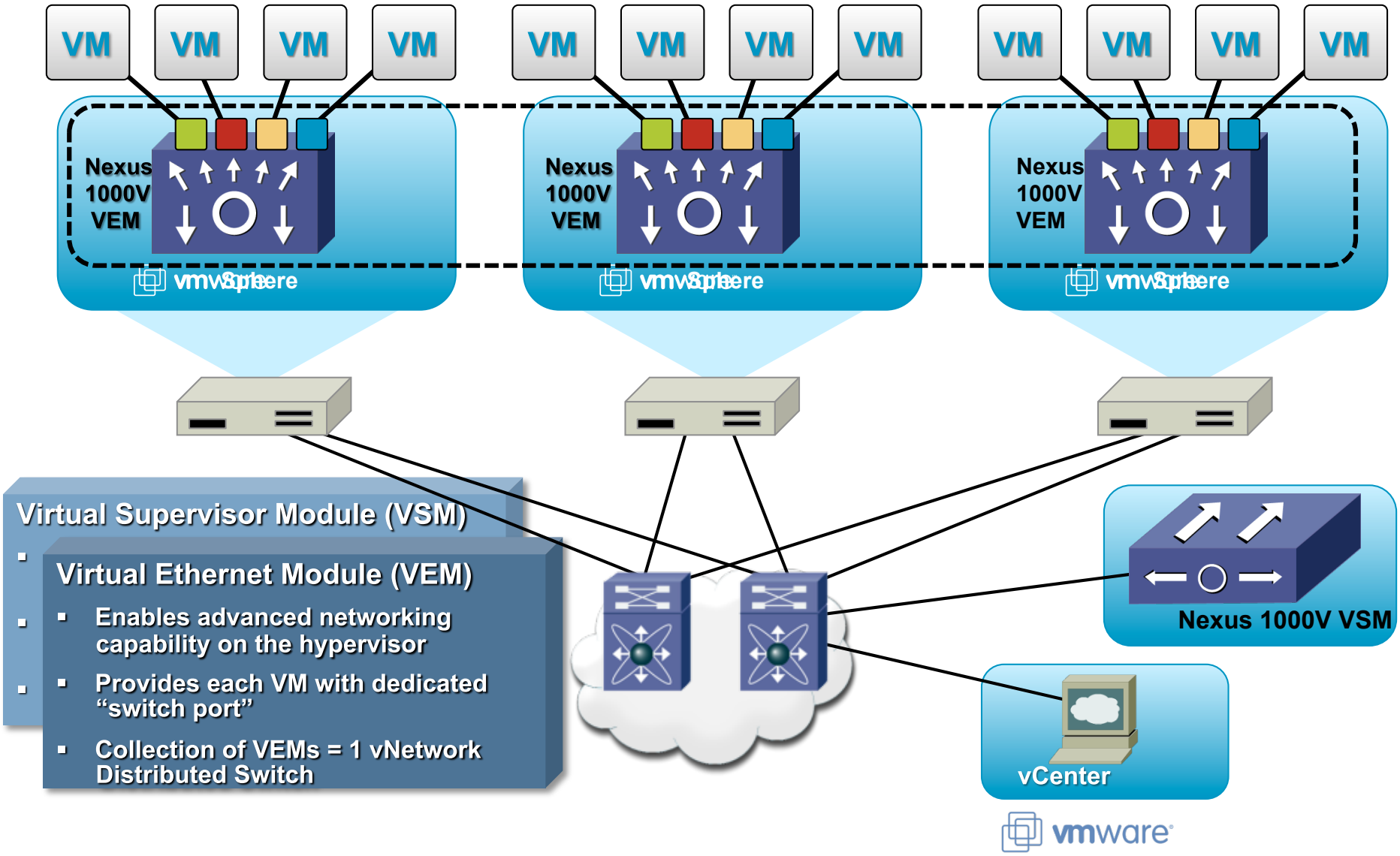




Policy-Based VM Connectivity | **Mobility of Network and Security Properties** | **Non-Disruptive Operational Model**

Data Center Architecture Evolution

VN-Link Today – Nexus 1000V



Hardware based VN-Link: UCS, Cisco VIC & Nexus 5500

Nexus 5548

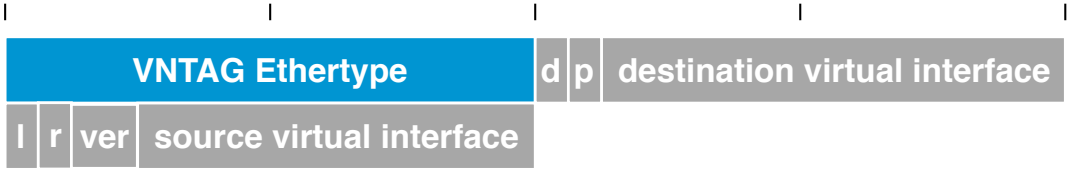
- Unified Port Concept: 1G/10G/FCoE/ DCB, 1/2/4/8 G FC, 1GE
- VN-Tag in hardware to implement Network Interface Virtualization
- NIV Initially to work only with Cisco VIC on UCS-C Series

UCS & Cisco VIC

- UCS 6100 Fabric Interconnect and Cisco VIC for NIV on UCS-B Series
- UCS 6100 Fabric Interconnect and Cisco VIC for UCS-C Series too
- VN-Tag in hardware to implement Network Interface Virtualization

VNTAG – Basis for 802.1Qbh

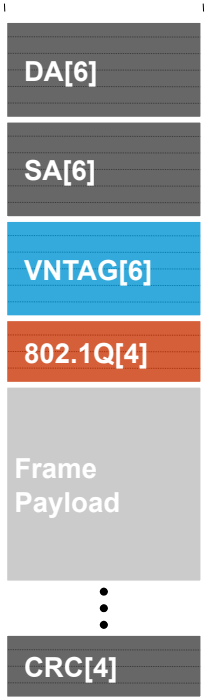
VN-Link in Hardware



- VNTAG: A tag in Layer 2 frame to address a vNIC or a list of vNICs
- Coexists with VLAN (802.1Q) tag

  802.1Q tag is mandatory to signal data path priority

  http://www.ieee802.org/1/files/public/docs2008/new-dcb-pelissier-NIV-Proposal-1108.pdf
- 802.1Qbh
  - Cisco and VMWare jointly submitted a proposal to IEEE to enable the switching with Network Interface Virtualization
  - Proposal by Joe Pelissier (Cisco Systems) and Andrew Lambeth (VMWare)

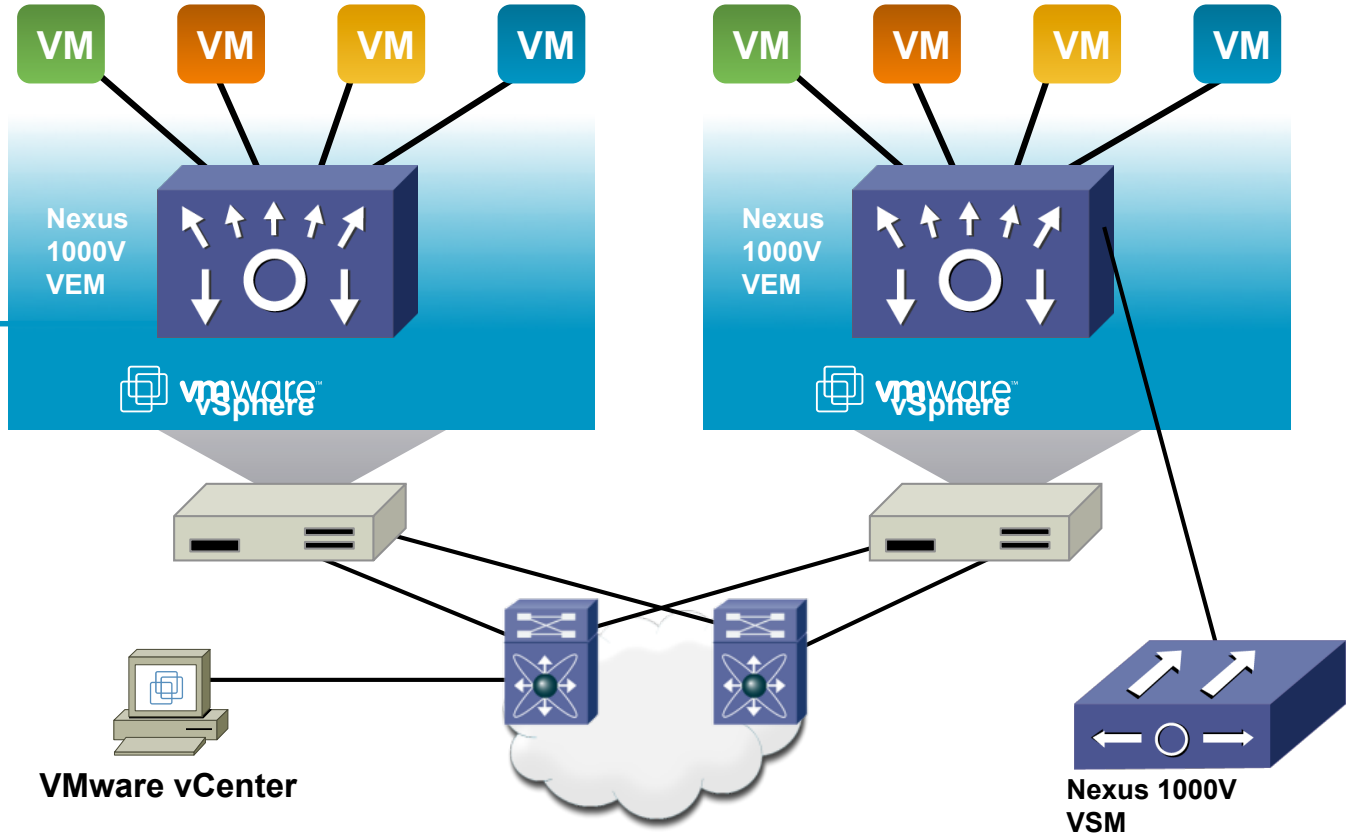


VN-Link Solution
Policy-Based VM Connectivity

Policy-Based VM Connectivity | Mobility of Network and Security Properties | Non-Disruptive Operational Model

Port Profiles
- WEB Apps
- HR
- DB
- DMZ

VM Connection Policy
- Defined in the network
- Applied in Virtual Center
- Linked to VM UUID



VSM: Virtual Supervisor Module
VEM: Virtual Ethernet Module

VN-Link Solution

Mobility of Network And Security Properties

Policy-Based VM Connectivity

Mobility of Network and Security Properties

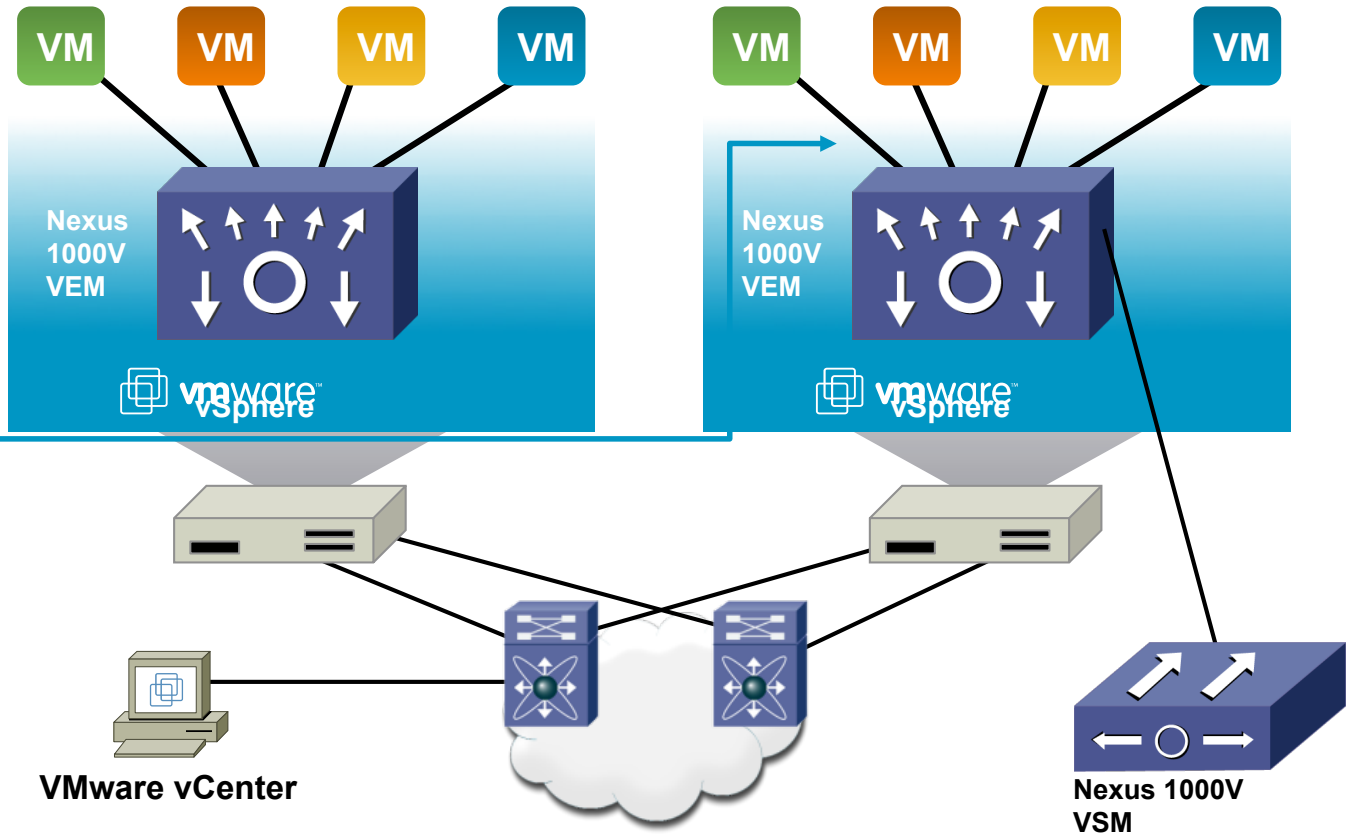
Non-Disruptive Operational Model

VMs Need to Move

- VMotion
- DRS
- SW Upgrade/Patch
- Hardware Failure

Property Mobility

- VMotion for the network
- Ensures VM security
- Maintains connection state



VN-Link Solution

Non-Disruptive Operational Model

Policy-Based VM Connectivity

Mobility of Network and Security Properties

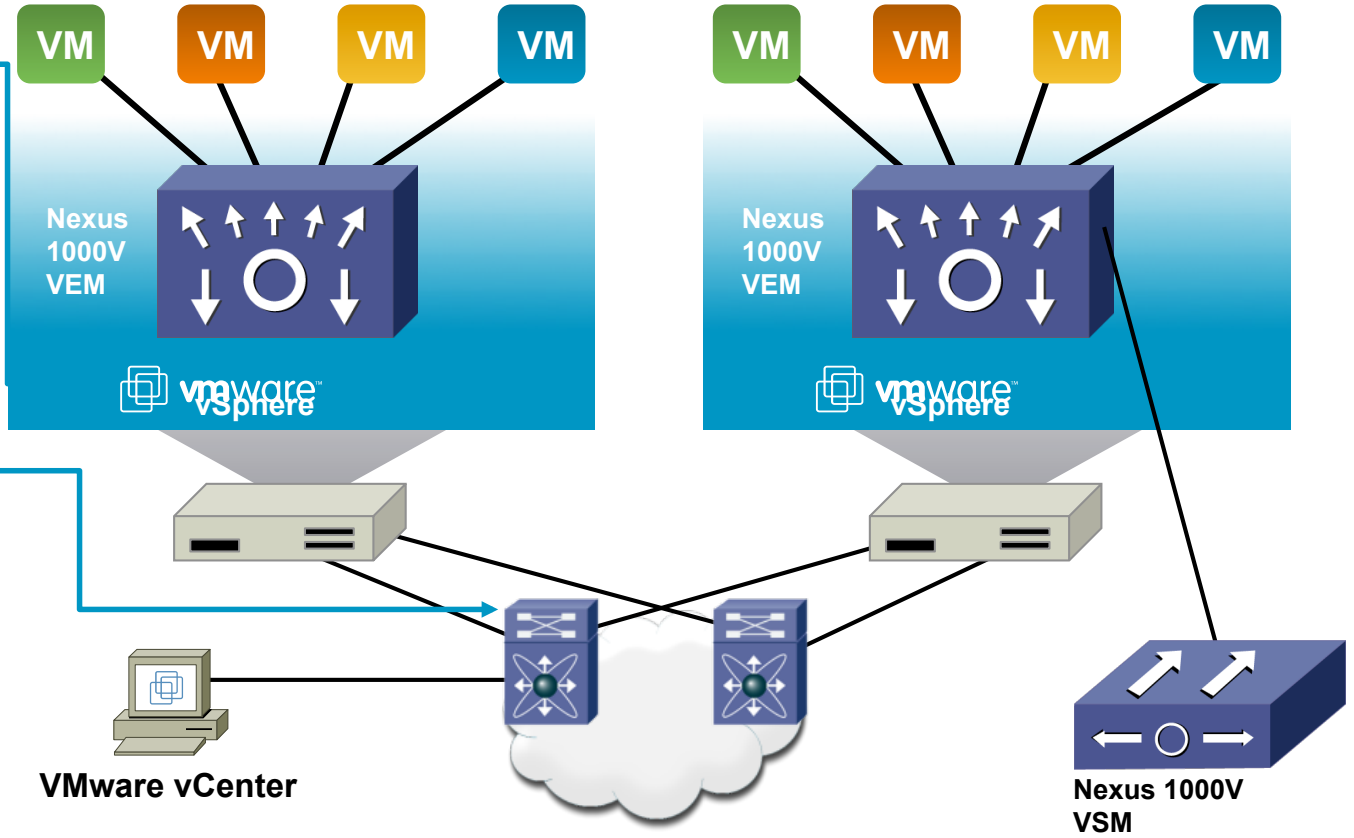
Non-Disruptive Operational Model

VI Admin Benefits

- Maintains existing VM mgmt
- Reduces deployment time
- Improves scalability
- Reduces operational workload
- Enables VM-level visibility

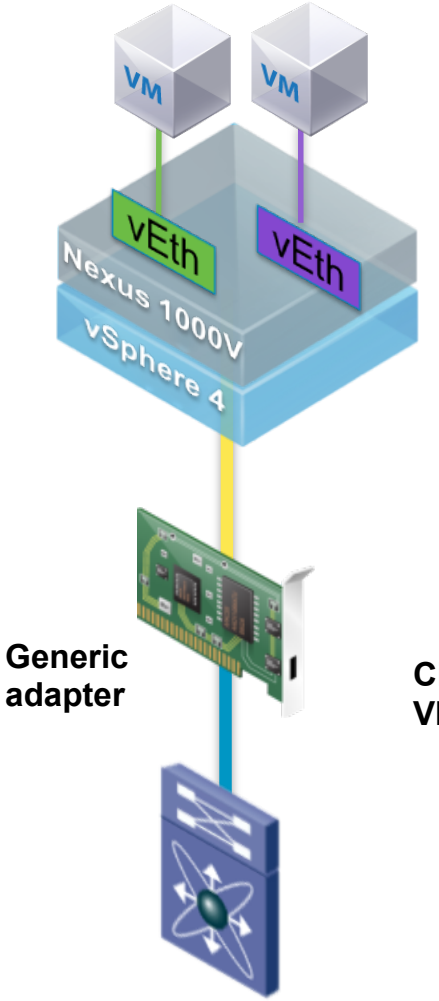
Network Admin Benefits

- Unifies network mgmt and ops
- Improves operational security
- Enhances VM network features
- Ensures policy persistence
- Enables VM-level visibility



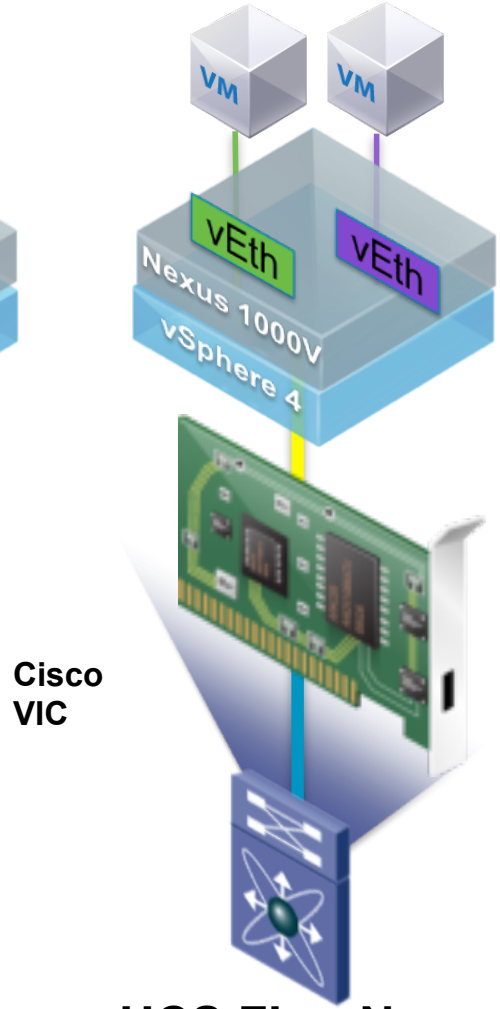
VN-Link Offering Summary

Generic Adapter
+ Nexus 1000V
Any Server



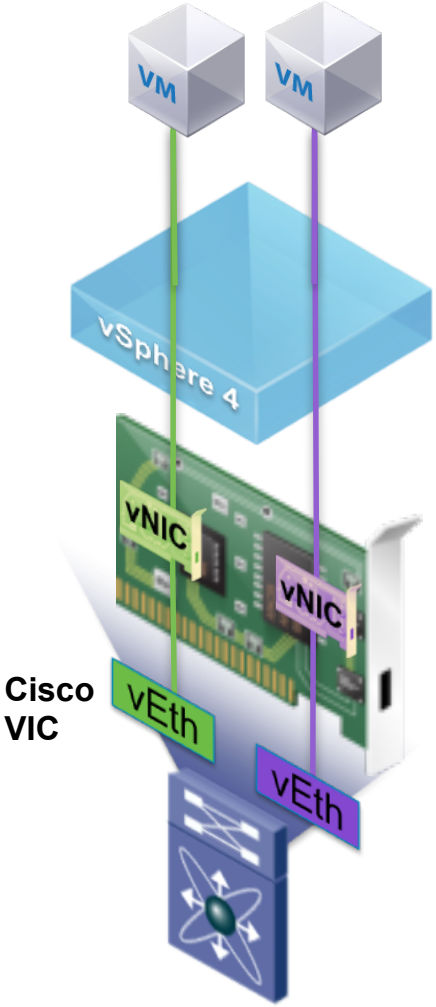
Any Switch

Cisco VIC
+ Nexus 1000V
UCS Only



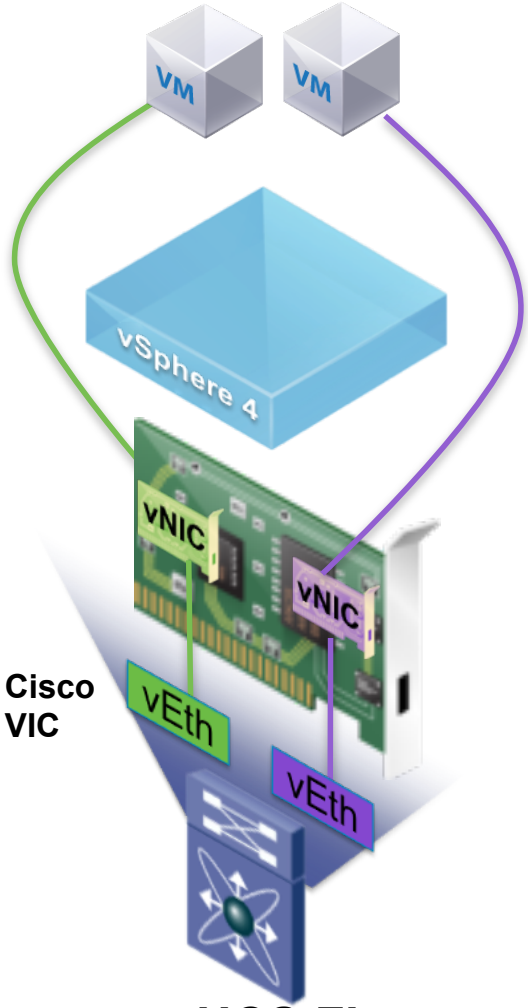
UCS FI or Nexus

Cisco VIC
+ UCS 6100
Vmware and RedHat



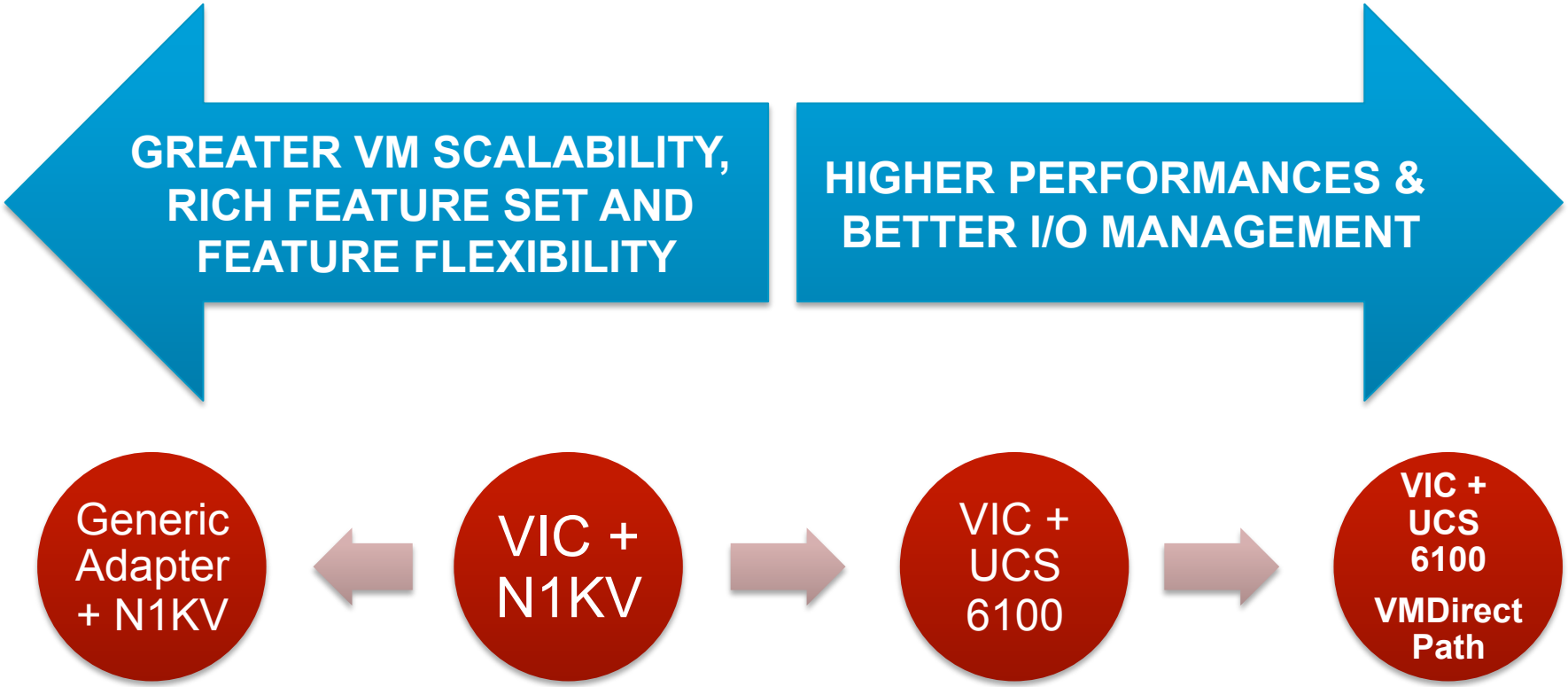
UCS FI or Nexus

Cisco VIC + UCS 6100 w/
VM DirectPath
UCS Only (future)



UCS FI

Why different models?



Expanding VN-Tag & VN-Link Support

- With Nexus 55x0 series
  - VIC support for IO-Virtualization (with VIC on UCS C-series only)
  - VIC support with VN-Link in HW (with VIC on UCS C-series only)
- Expanding the Adapter Ecosystem with VN-Tag

- Expanding Hypervisor offering (RedHat for VN-Link in HW on UCS)

VN-Link Deployment

Nexus 1000V + Generic Adapter

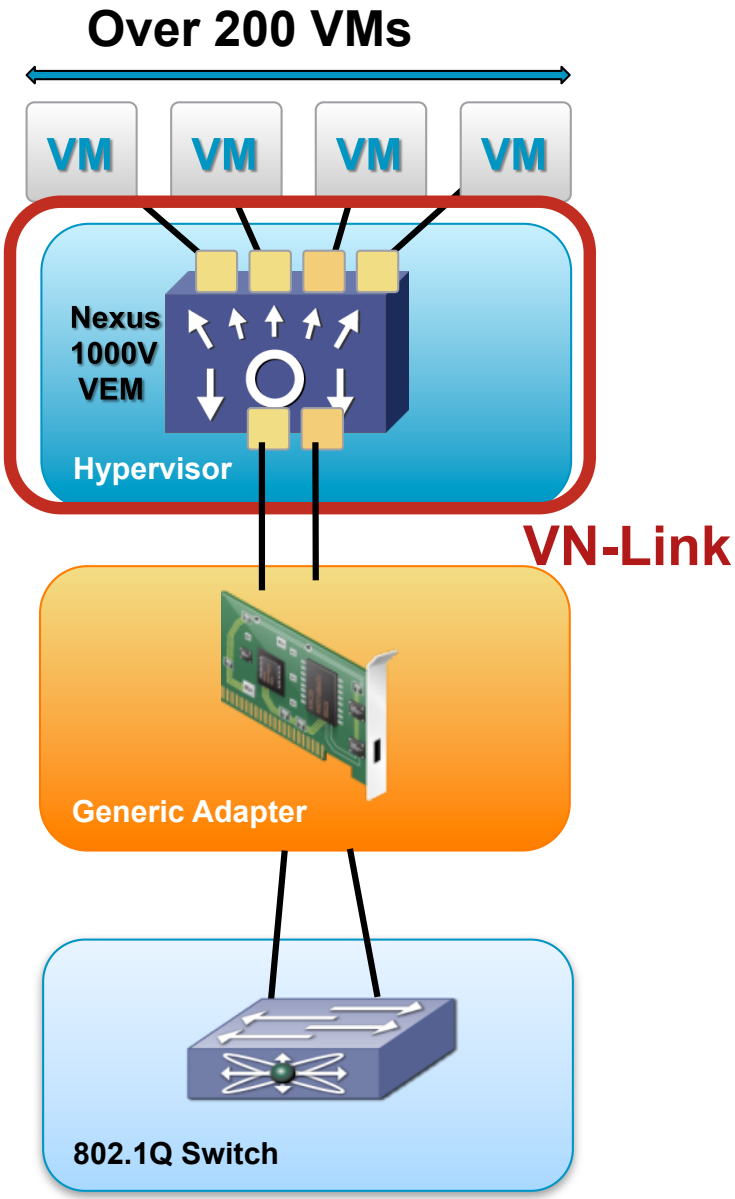
- Nexus 1000V

  Performs packets forwarding and applies advanced networking features

  Uses port security, VLAN, and ACLs, policy maps for QoS treatment for VM traffic, Console & Vmotion/Vmkernel

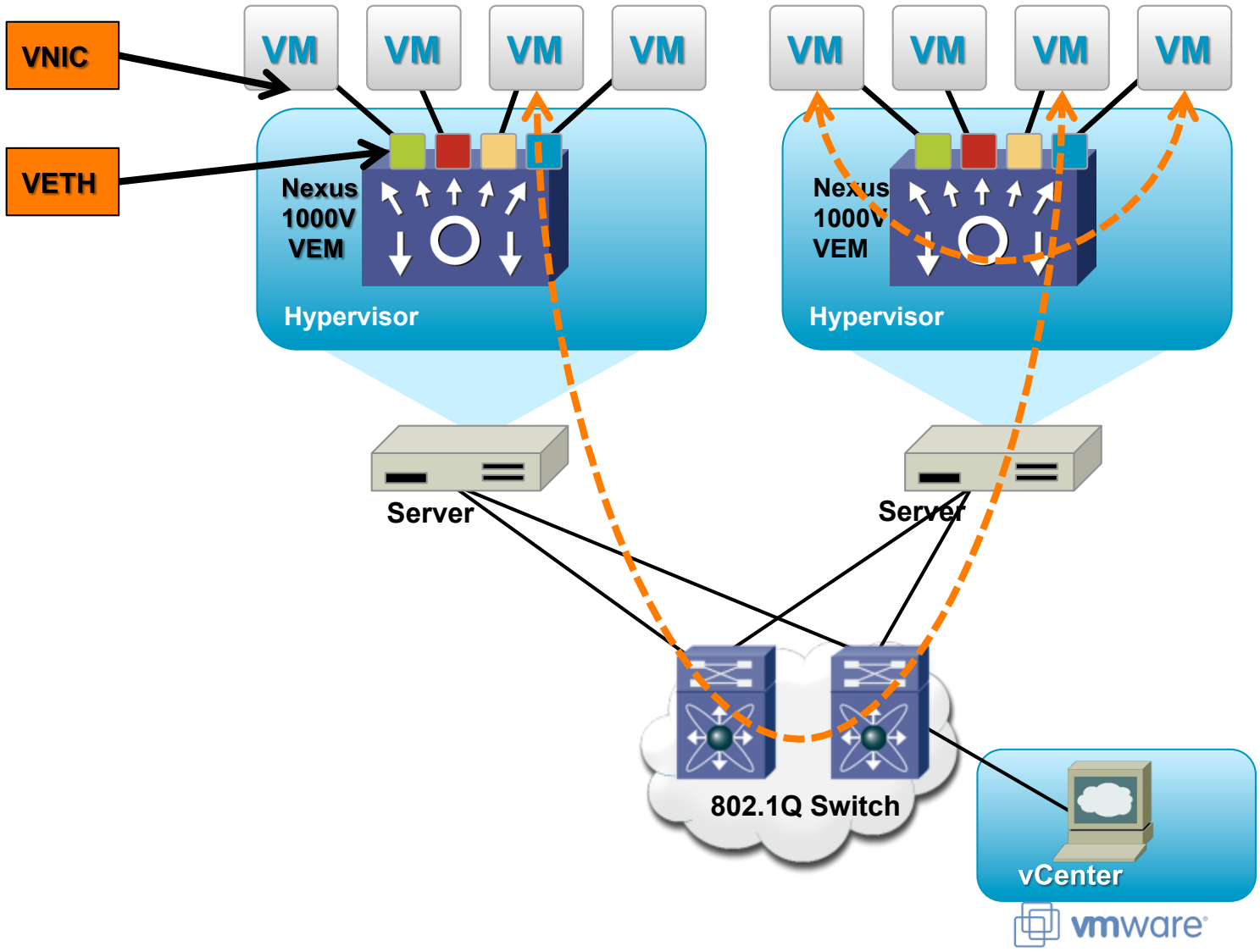
- Generic adapter on generic x86 server

- Generic Upstream Switch



VN-Link Packet Flow

Nexus 1000V + Generic Adapter



VN-Link Deployment

Nexus 1000V + Cisco VIC (UCS)

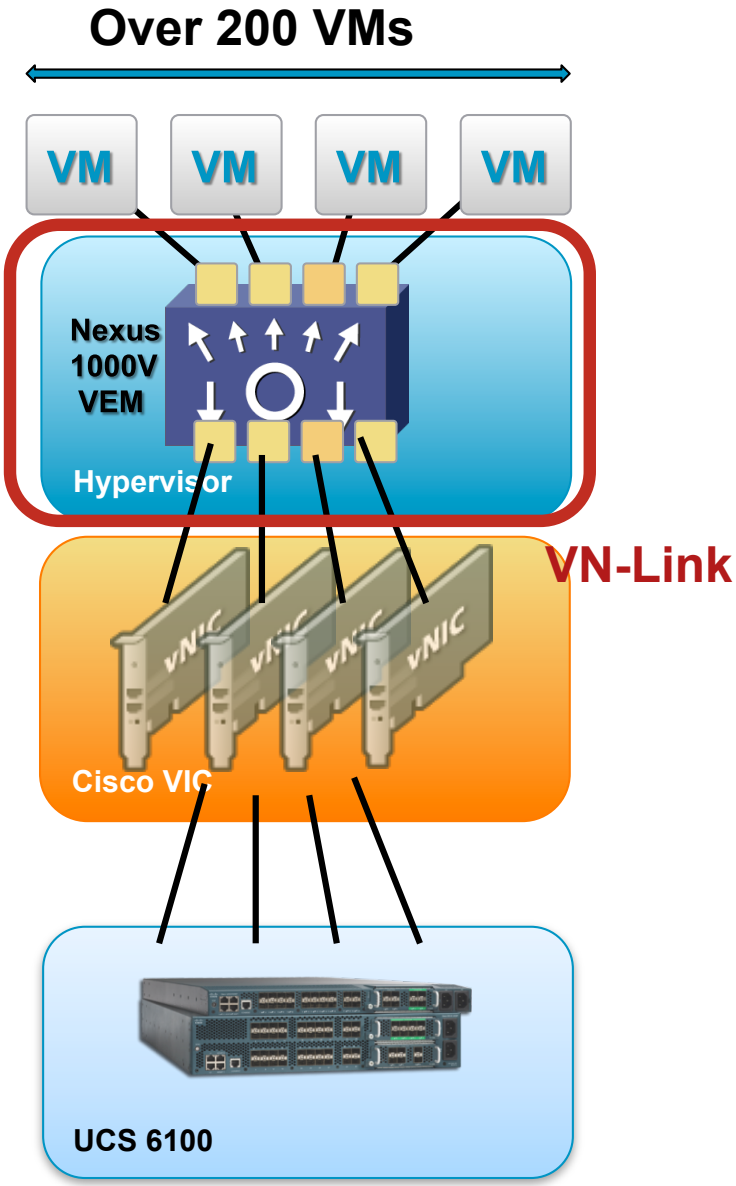
- Nexus 1000V

  Performs packets forwarding and applies advanced networking features

- Cisco VIC (UCS)

  Virtualizes PCIe devices to provide true traffic segregation in HW (VM traffic, Service Console & Vmotion/ Vmkernel)

- UCS 6100



VN-Link Deployment
VIC + UCS 6100

- Hypervisor

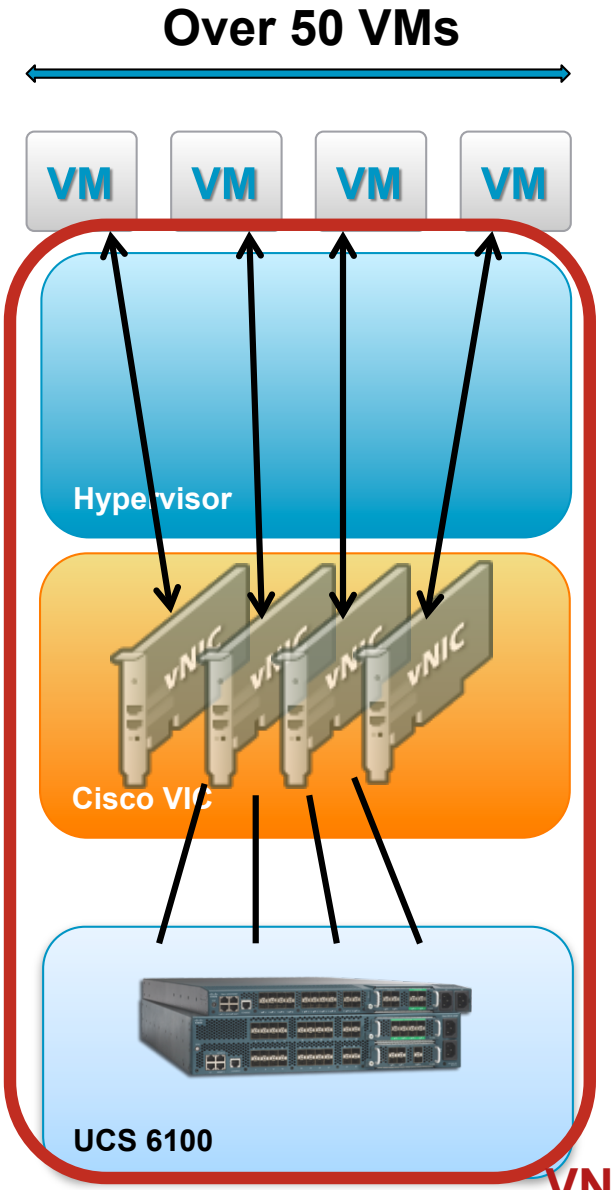
  Pass-thru within the hypervisor

- Cisco VIC

  Performance enabled – set of individual I/O queues for each vNIC

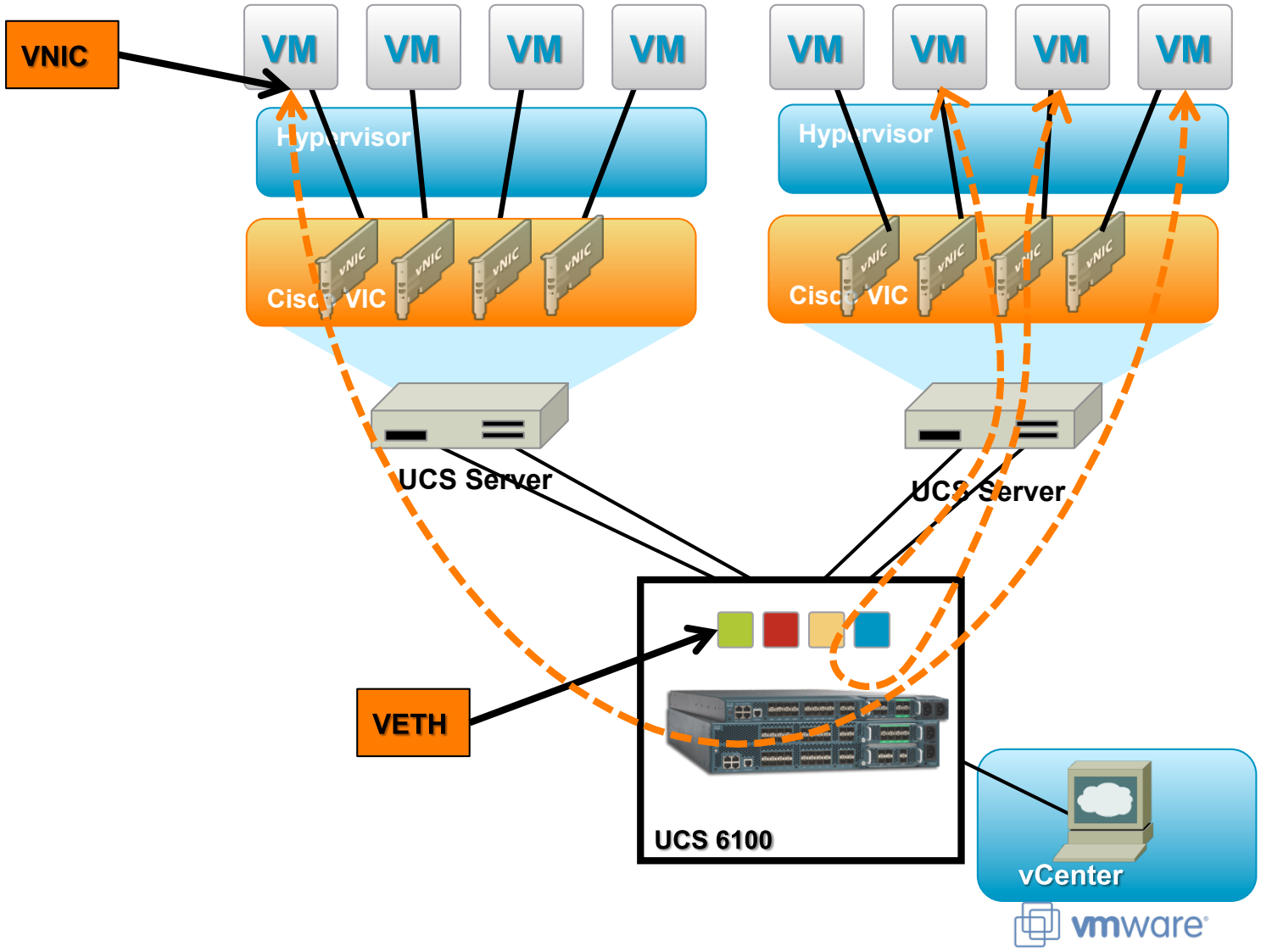
- UCS 6100

  performs packets forwarding and applies networking features



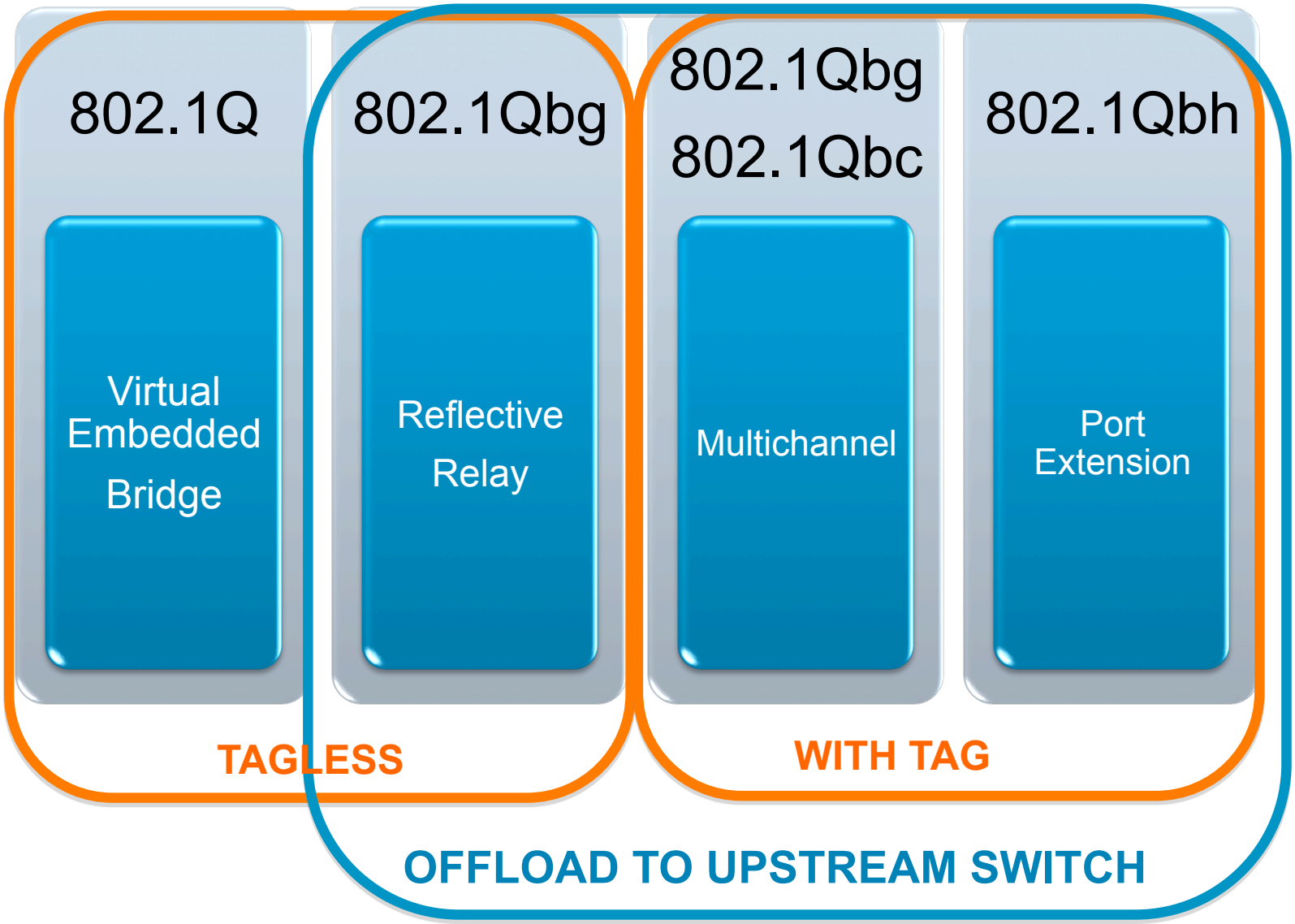
VN-Link Packet Flow

VIC + UCS 6100

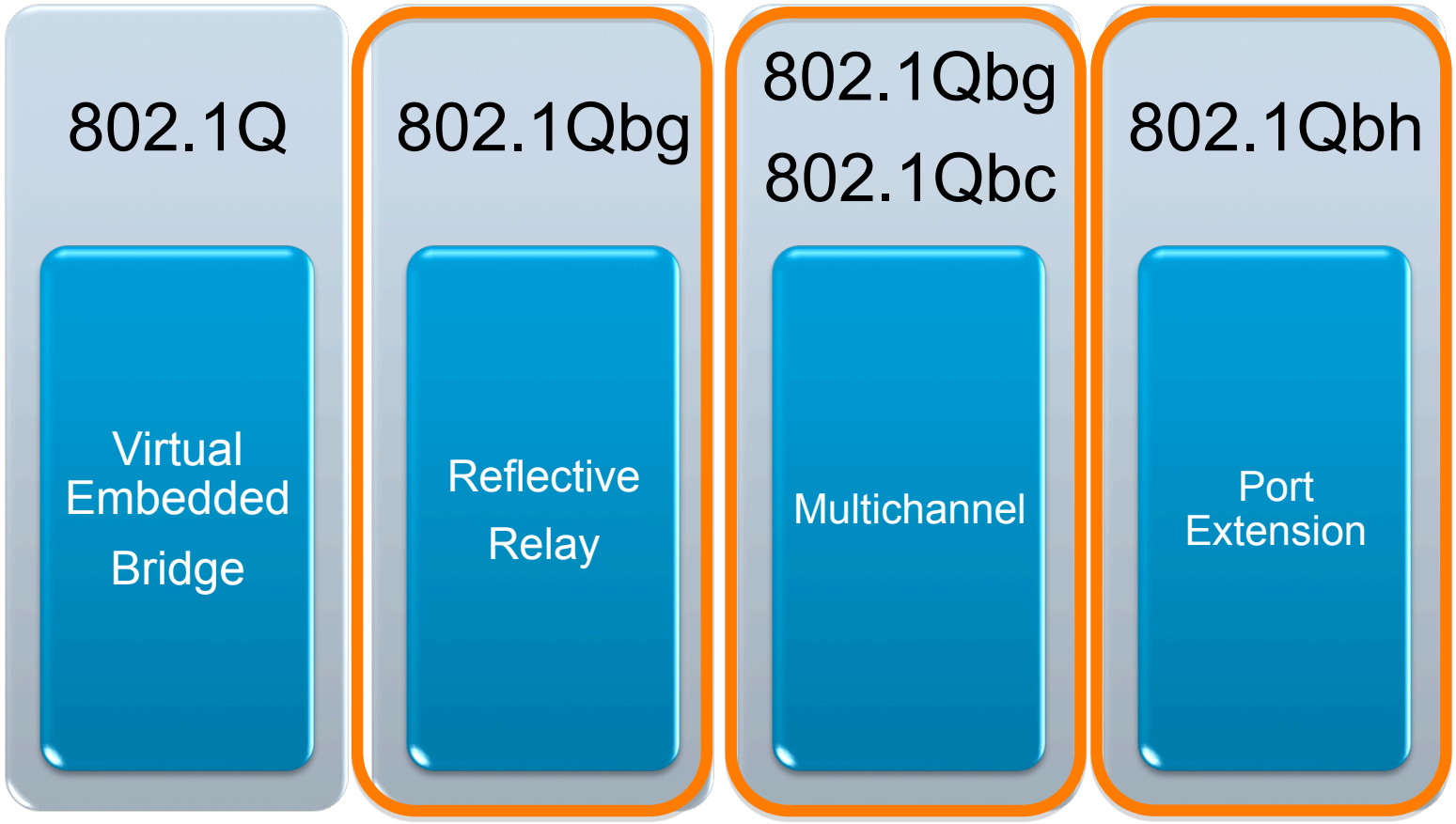


Virtual Networking Standards along the road ….

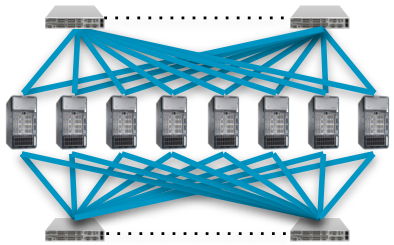
Virtual Networking Standards Components

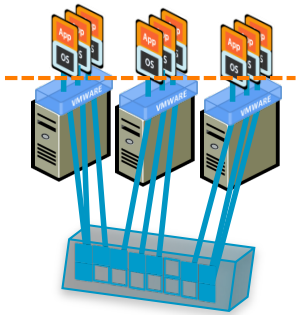


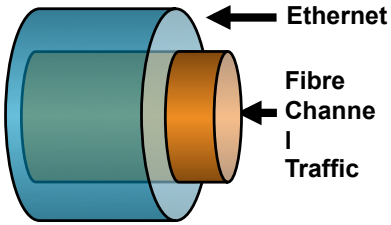
Virtual Networking Standards Components



Nexus Data Center Access Technologies







Distributed Access Fabric

- Fabric Extender Technology
- 100M/1G/10G Capable
- FCoE & VNTag Support

- 10:1 reduction in managed devices
- No cross-rack cabling
- Form-Factor & MGMT optimized

Scalable Access Choice

- Virtual Port Channel (vPC)
- FabricPath/Trill
- Routed Access

- Accelerate IO intensive applications
- Eliminate logical Layer 2 loops
- Massive bisectional bandwidth

Virtual Networking

- VN-Link
- VN-Services
- VN-Manager

- Extend networking principles to virtualized environments
- Extend network services to virtualized environments
- Enable consistent VN-wide management

Unified Fabric

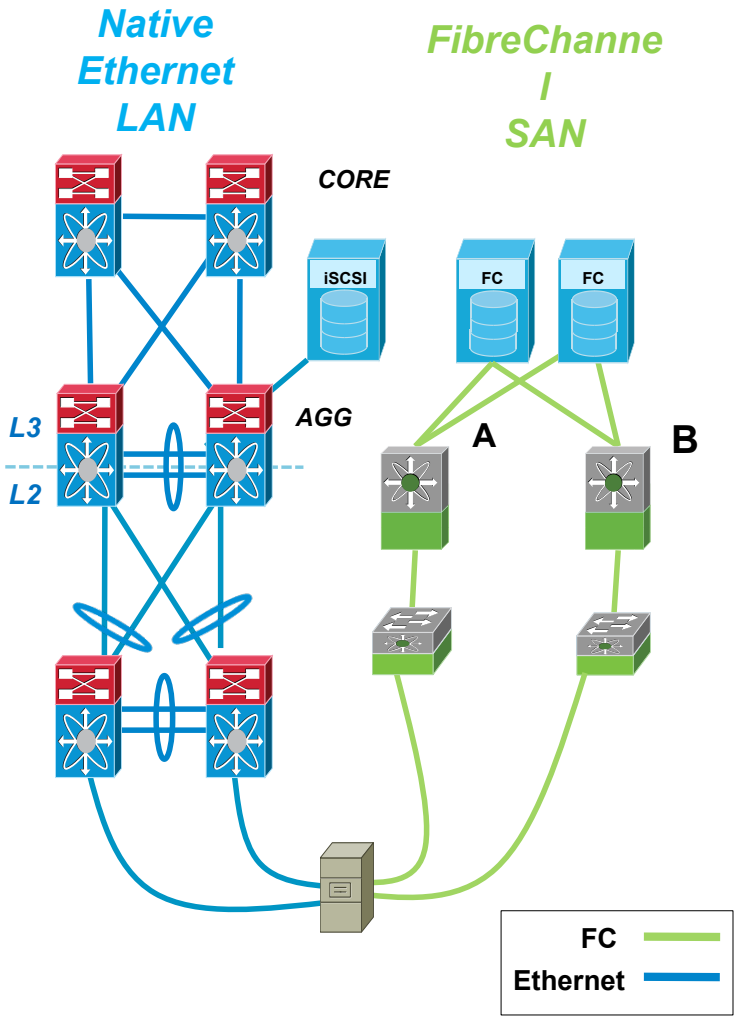
- LAN/SAN Consolidation
- Converged Network Adapters (CNAs)
- High Performance Server Interconnect

- 50% or more cable reduction
- 50% or more network/SAN device reduction
- Maintains logical LAN/ SAN mgmt model

Powered by Cisco NX-OS Data Center Operating System

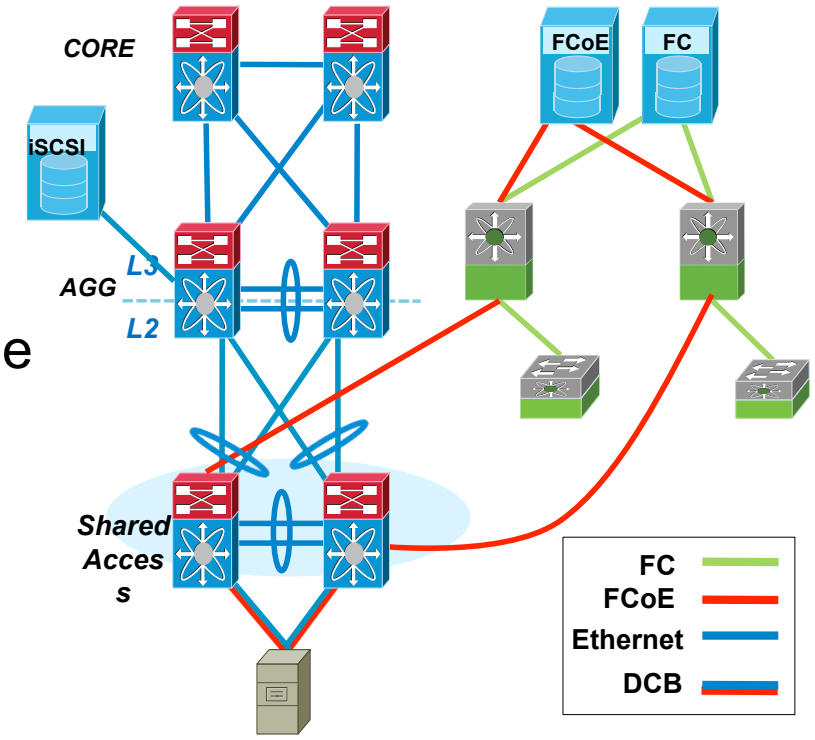
Data Center Network Today

- Segregated purpose built networks.
  - Separate management teams.
  - Limitations to scale.
- Ethernet is by far the dominant interconnection network.
  - Applications tolerate lost and out of order frame delivery.
- Fibre Channel is the standard of choice for enterprise storage networks.
  - No tolerance for lost or out of order frame delivery.
  - Credit based flow and congestion control.
  - Highly available via dual SAN topology.
- Sprawling network interfaces at the Server.
  - At a minimum 2 FC HBAs and 2 Ethernet NICs.



Shared Access-attached Storage Network

- Unifying the Access is biggest benefit
- Common SAN to access FC and FCoE storage.
  - Cable once, then repurpose any storage on an as-needed basis.
  - Facilitates end-to-end FCoE.
  - Connectivity to existing SAN.
  - Take advantage of next-gen storage without a SAN upgrade
- Maintain storage fabric segregation.
- Nexus 5000 in NPV or Switch mode.
  - NPV especially important when interoperating with another vendor



Unified Fabric

FCoE Design objectives

- I/O consolidation – Dramatic CAPEX and OPEX reduction
- Preserve SAN design best practices
  - Collapsed-core, core-edge or edge-core-edge designs
  - Oversubscription, Fan-in ratios, hop count practices honored
- Preserve SAN and LAN management models
  - Deterministic management of FC flows through all devices
  - No opaque ‘LAN’ clouds transporting SAN traffic
- SAN scalability
  - Build-up the edge, from 20% attach-rate up to 100%
  - Allow LAN and SAN to scale independently

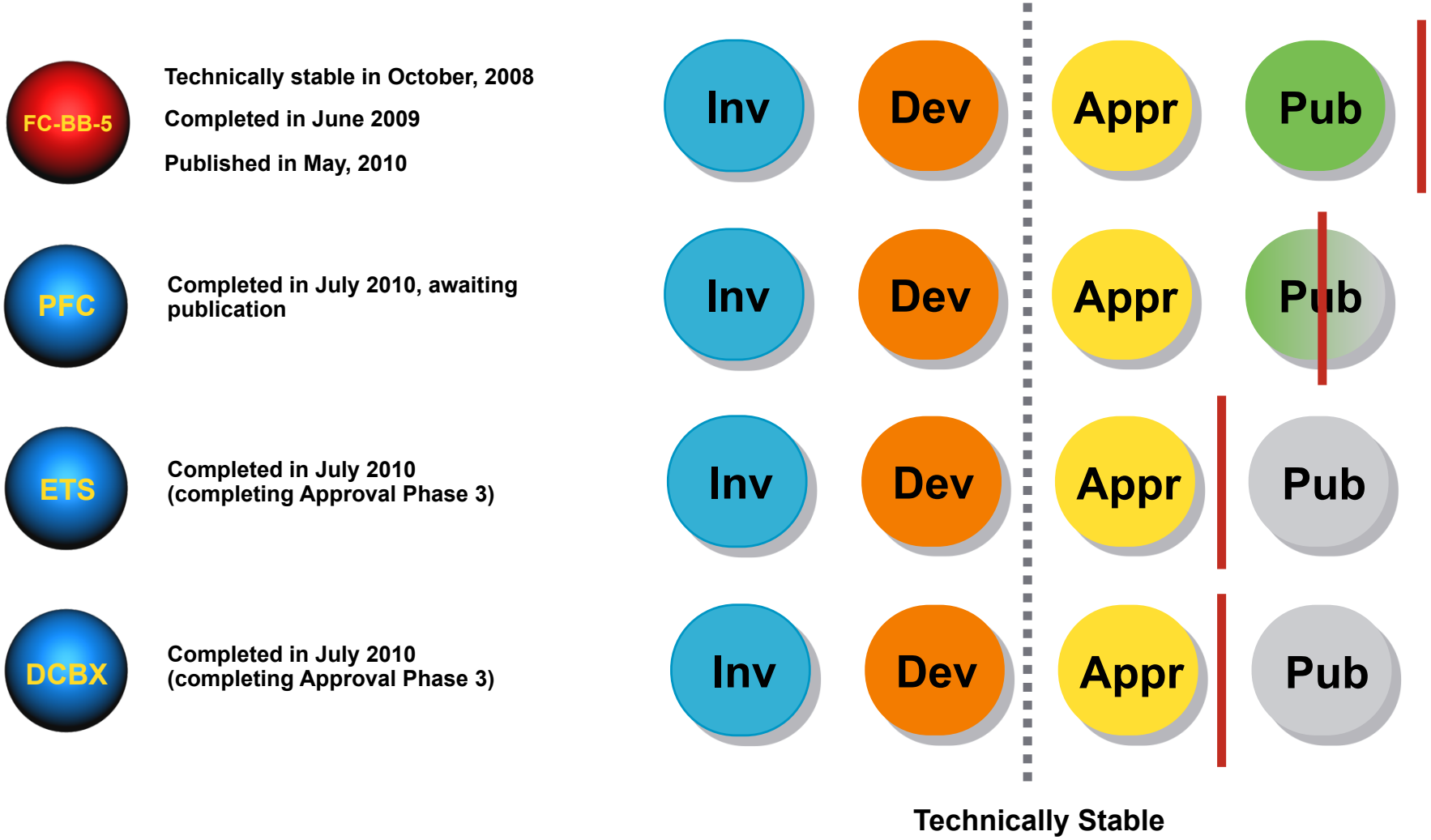
The FCoE Standard

- T11 FC-BB-5
  - Supports the full Fibre Channel fabric functionality
    - Including multi-hop FCoE
  - Working Group established in June 2007
  - Technically stable in October 2008
  - Completed (forwarded for publication) in June 2009
  - Published by ANSI in May 2010

- T11 FC-BB-6
  - Developing additional FCoE functionality
  - Not needed for current deployments
  - Working Group established in August 2009

Status of the Standards

All Standards for FCoE Are Technically Stable



Is IEEE 802.1Qau (QCN) needed?

- QCN designed to avoid spreading of congestion
- BUT there are some limitations
  - Complexity of deployment
  - Requires pervasive implementation
  - L2 only protocol
    - QCN does not propagate beyond a L2 domain.
    - QCN useless if L2 domain composed of single/pair of hops
- FC networks
  - FC does not have equivalent of QCN today and works fine in multi-tier networks
  - SCSI (the only user of FCoE) is an interlocked protocol that deals with congestion automatically

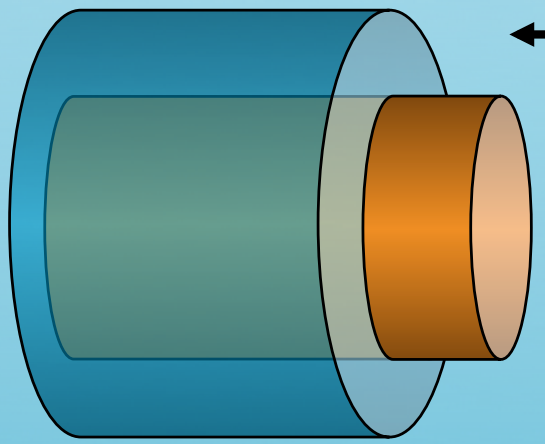
Fiber Channel over Ethernet

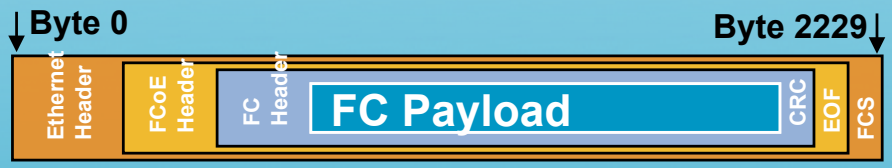
FC-BB-5 Protocol

incits InterNational Committee for Information Technology Standards
Where IT all begins

FCoE

Mapping of FC Frames over Ethernet





- Completely based on the FC model
- Same host-to-switch and switch-to-switch behavior as FC
- WWNs, FC-IDs, hard/soft zoning, DNS, RSCN

FCoE is Fibre Channel

Cisco	Dell	EMC²	EMULEX
HP	IBM	Microsoft	NetApp
Intel	QLOGIC	Redhat	VMWARE

Fiber Channel over Ethernet

Data and Control plane

FCoE is really two different protocols:

FCoE itself

- Is the data plane protocol
- It is used to carry most of the FC frames and all the SCSI traffic
- Uses Fabric Assigned MAC address (dynamic) : FPMA

FIP (FCoE Initialization 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

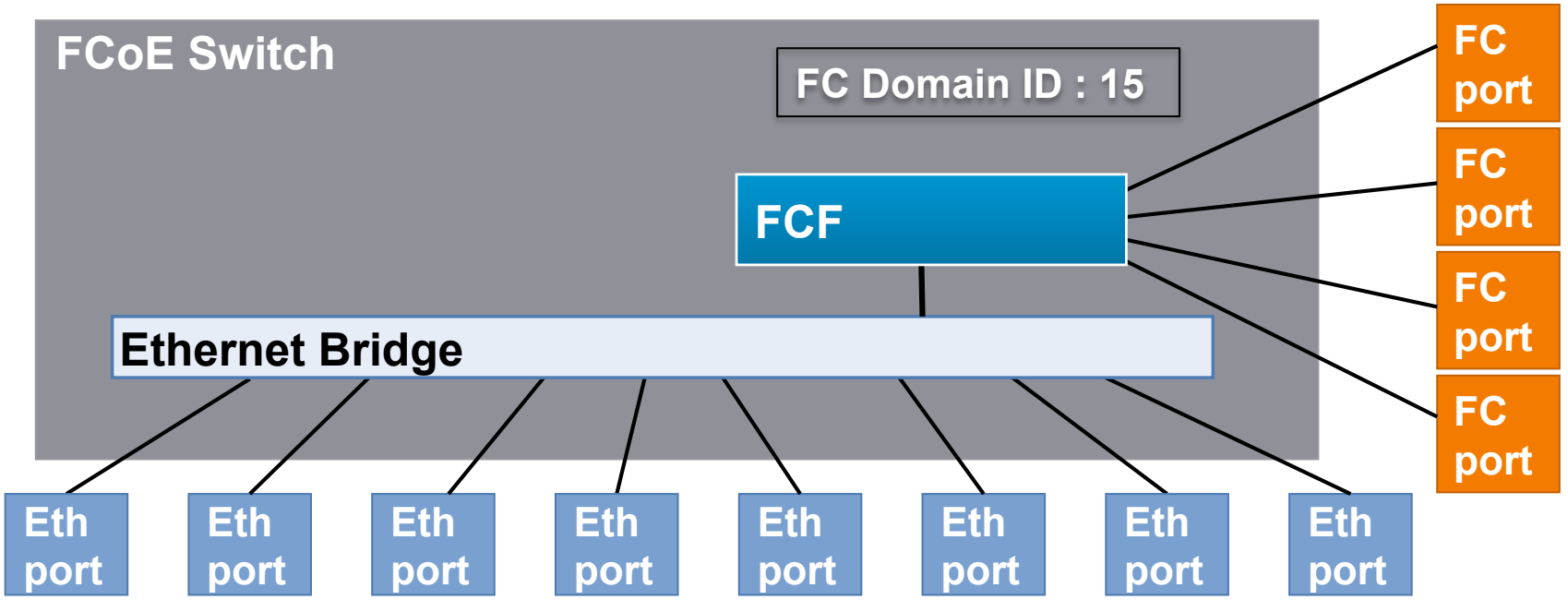
The Protocols Have…

- A different Ethertype
- A different frame format
- Both are defined in FC-BB-5

FCoE Building Blocks

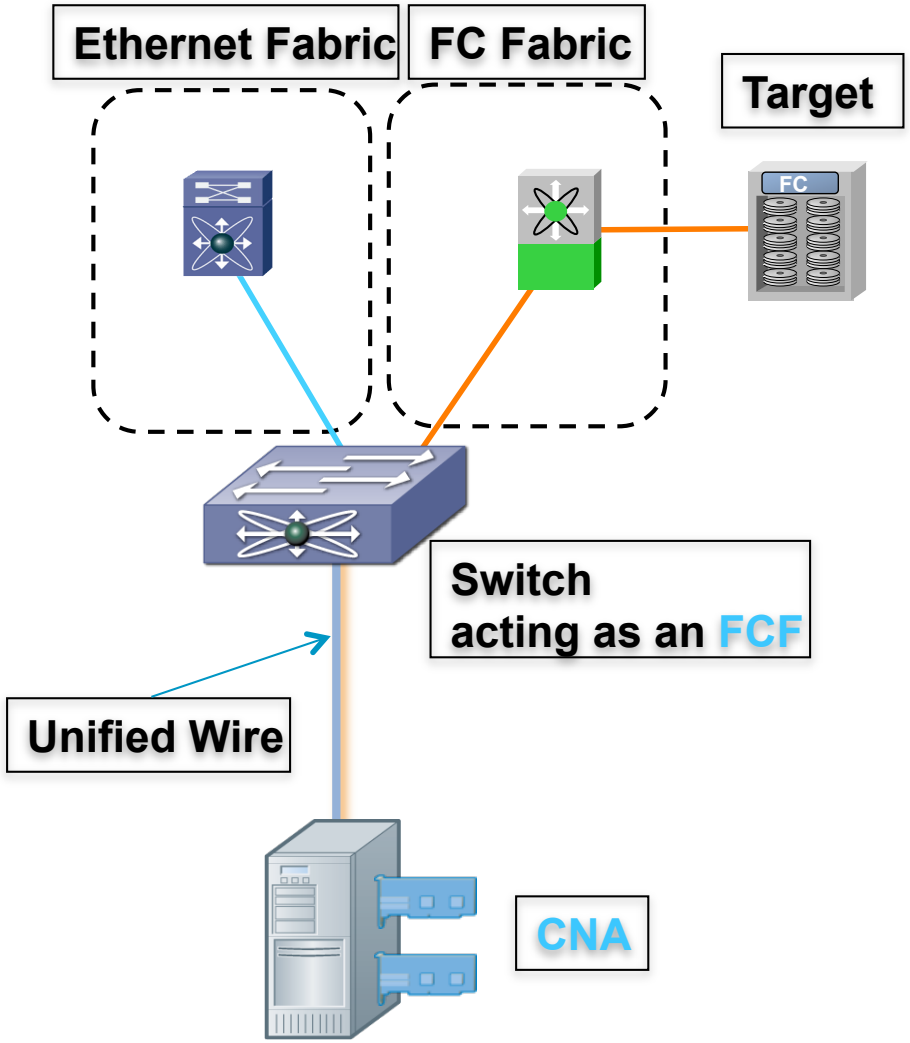
Fibre Channel Forwarder

- 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



Today's DC Solution

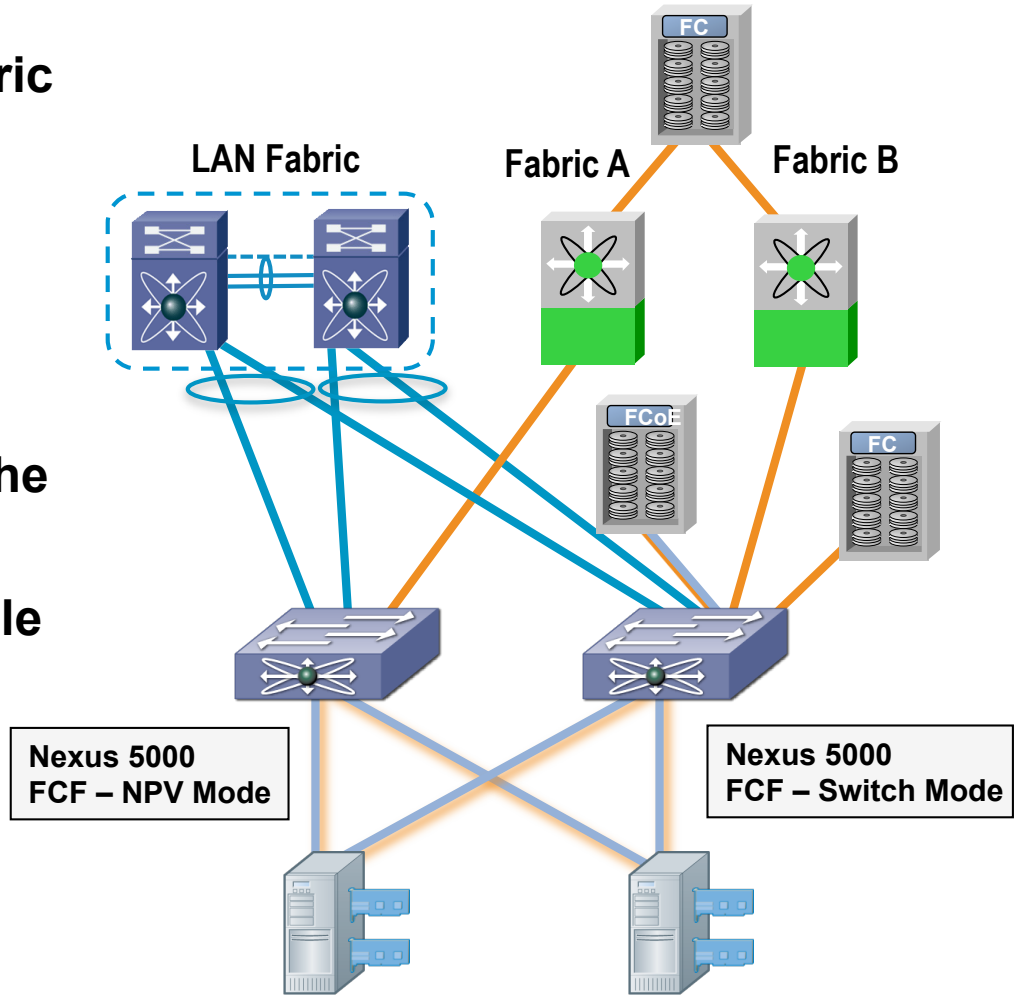
- Host connected over *unified wire* to first hop access switch
  - Access switch (Nexus 5000) is the FCF
- DCBX is used to negotiate the enhanced Ethernet capabilities
- FIP is use to negotiate the FCoE capabilities as well as the host login process
- FCoE runs from host to access switch FCF – native Ethernet and native FC break off at the access layer



Unified Fabric Design

Unified Edge

- The first phase of the Unified Fabric evolution design was focused on the fabric edge
- Unified the LAN Access and the SAN Edge by using FCoE
- Consolidated Adapters, Cabling and Switching at the first hop in the fabrics
- The Unified Edge supports multiple LAN and SAN topology options
  - Virtualized Data Center LAN designs
  - Fibre Channel edge with direct attached Initiators and Targets
  - Fibre Channel edge-core and edge-core-edge designs

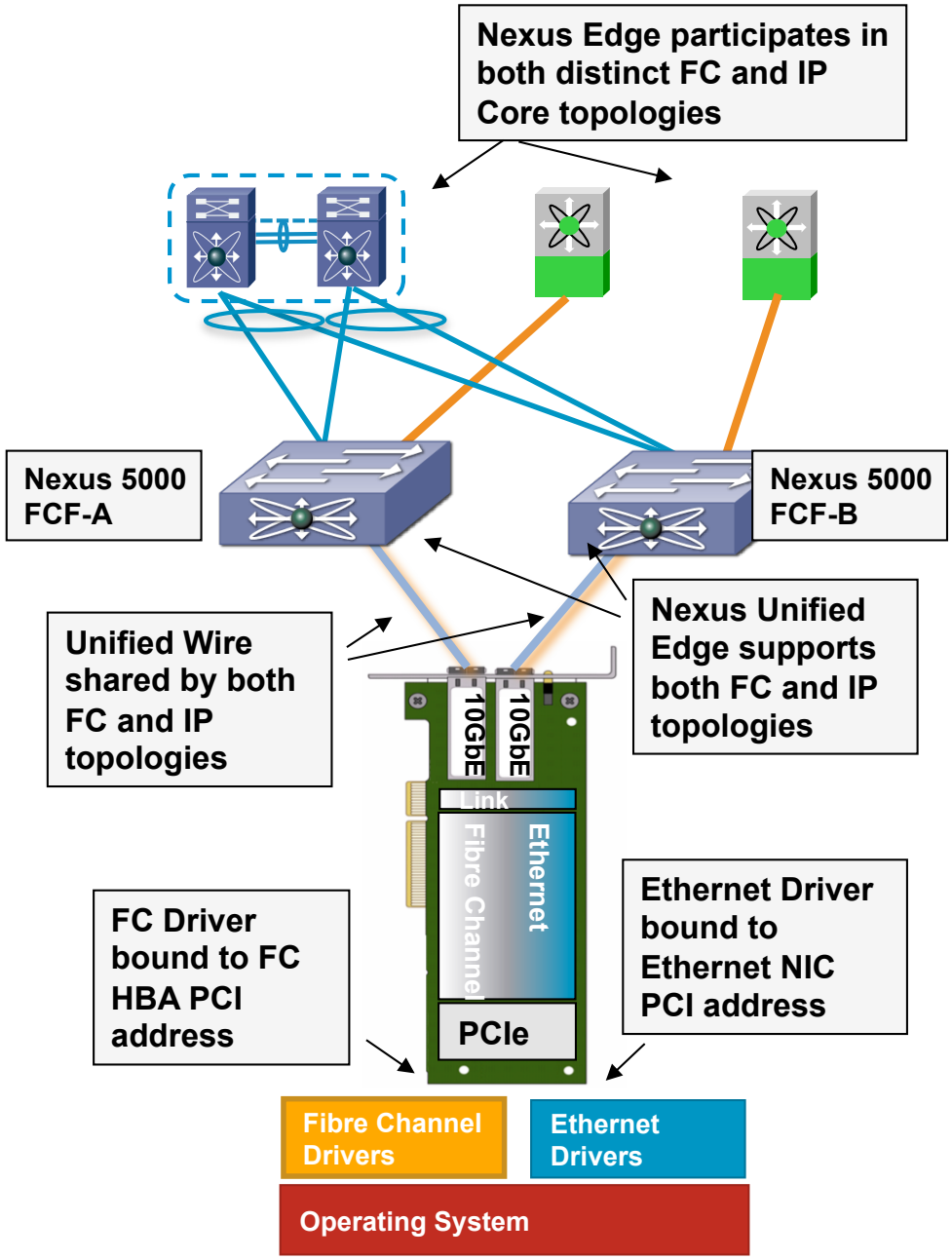


The Unified Edge

Unified Fabric Design

Unified Edge

- Converged Network Adapter (CNA) presents two PCI addresses 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)



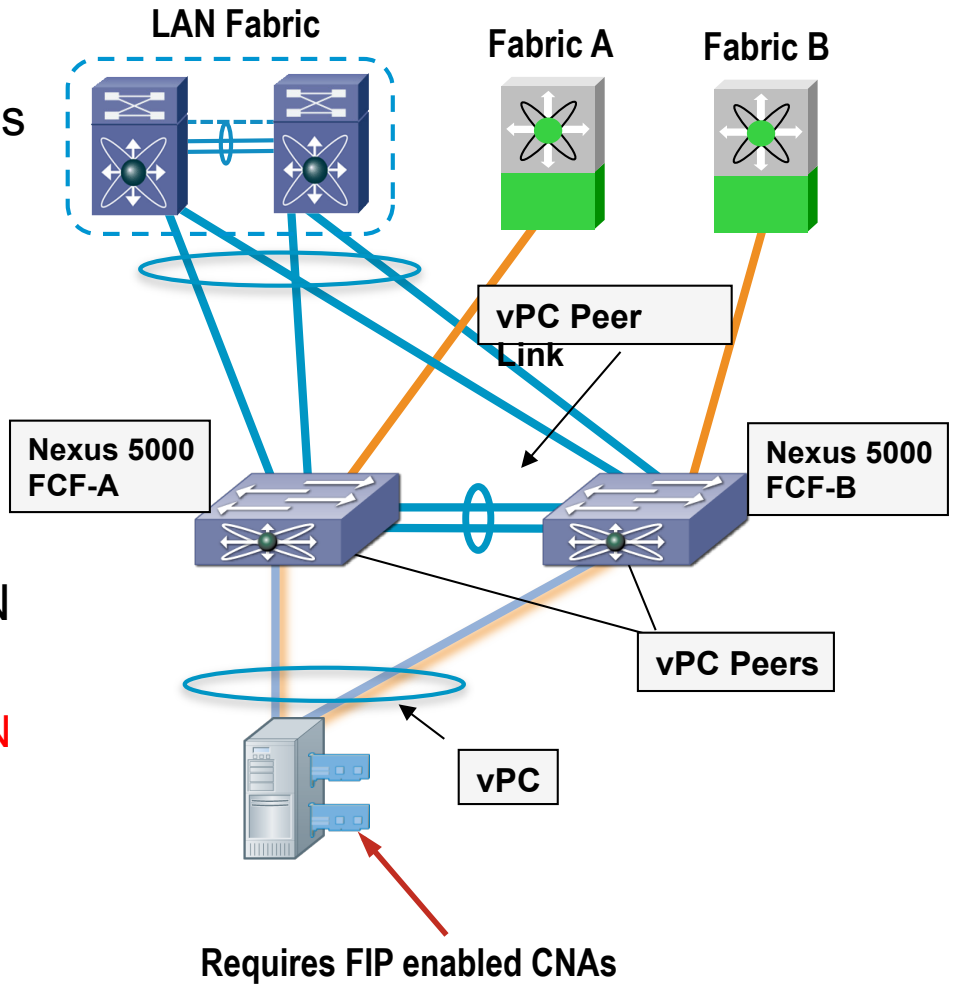
Unified Fabric Design

Unified Wires at the Edge with vPC

- Optimal layer 2 LAN design often leverages Multi-Chassis Etherchannels
- Nexus utilizes Virtual Port Channel (vPC) to 802.3ad attached servers
- vPC provides network based load sharing *and redundancy* without introducing layer 2 loops in the topology
- vPC results in diverging LAN and SAN high availability topologies

  FC maintains separate SAN 'A' and SAN 'B' topologies

  LAN utilizes a single logical topology

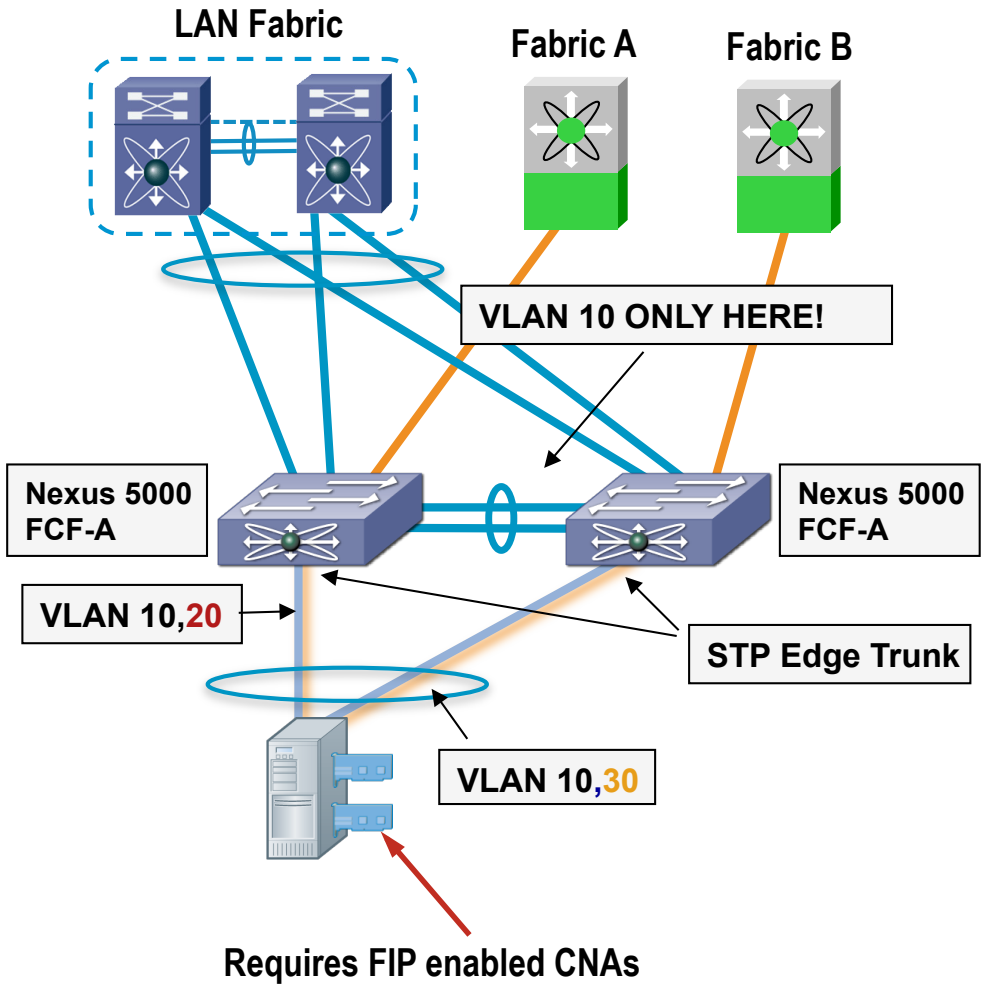


Direct Attach vPC Topology

Unified Fabric Design

Unified Wires at the Edge with vPC

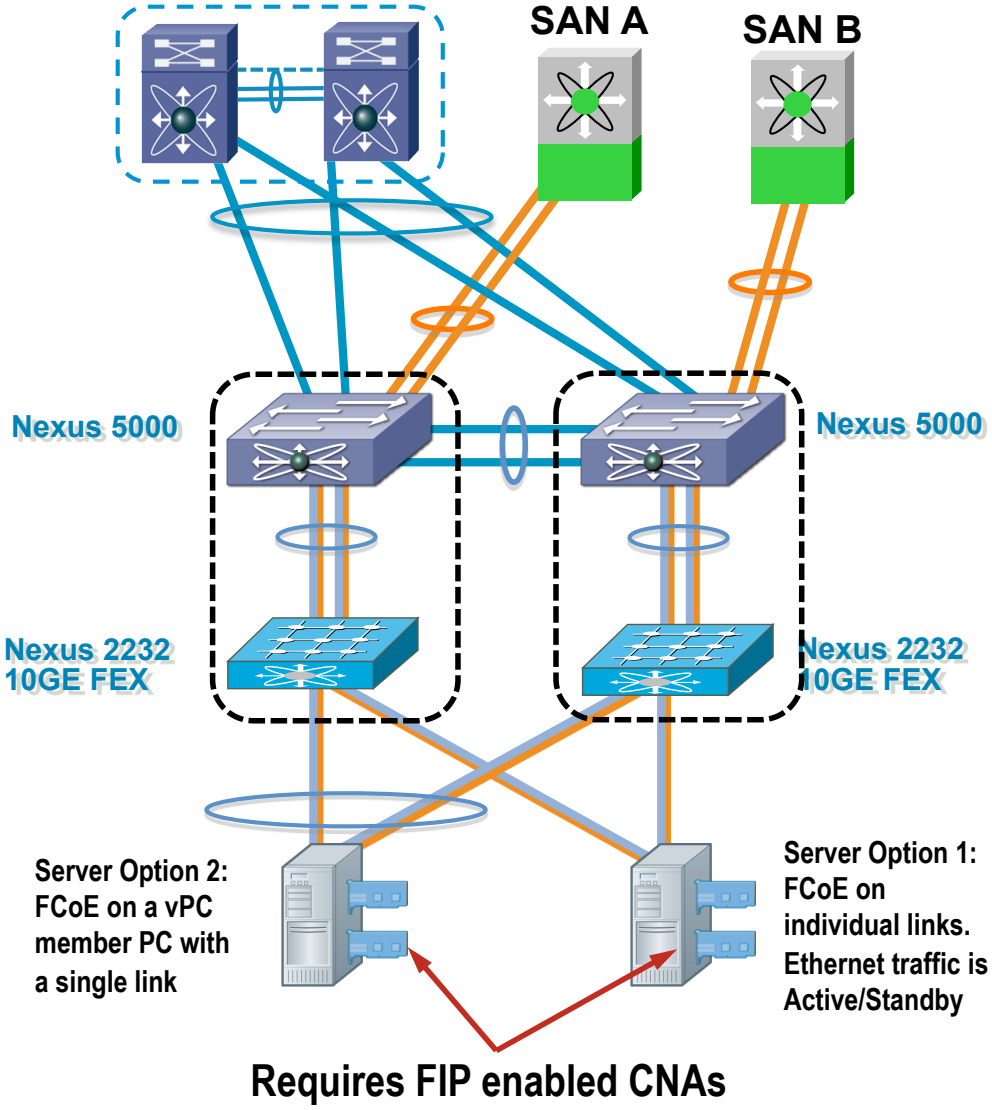
- To ensure correct forwarding behavior *for SAN traffic* in vPC enabled topologies, specific design and forwarding rules *must* be followed
- With the NX-OS 4.1(3) releases a 'vfc' interface can only be associated with a vPC etherchannel comprising ***one (1)*** CNA port attached to *each* edge switch
- While the port-channel is the same on N5K-1 and N5K-2, **the FCoE VLANs are different**
- vPC configuration works with FIP-enabled CNAs ONLY
- FCoE VLANs are *'not'* carried on the vPC peer-link
- FCoE and FIP ethertypes are *'not'* forwarded over the vPC peer link



Direct Attach vPC Topology

Virtualized Access Switch - FCoE

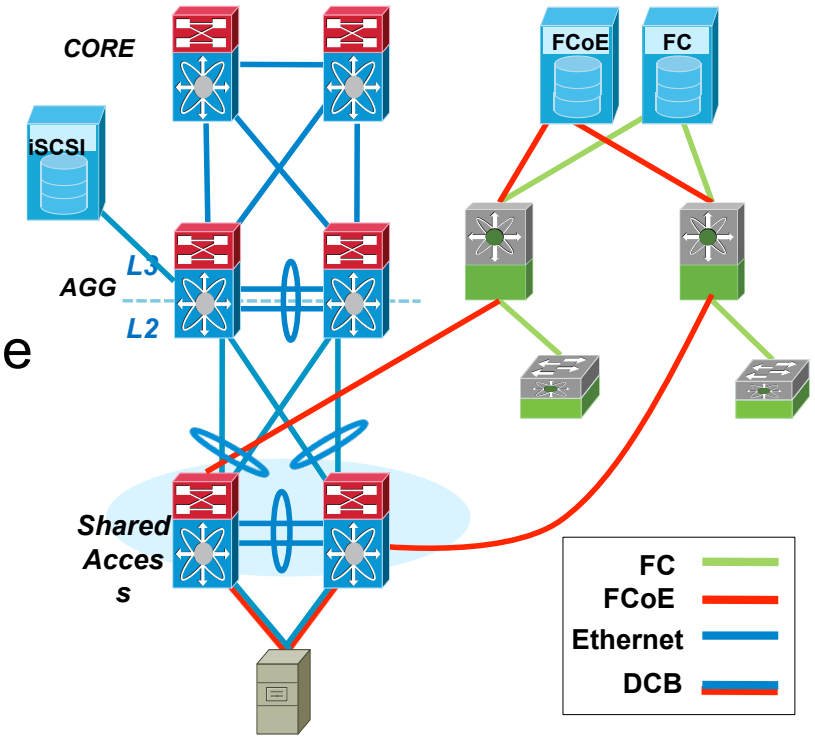
- FEX-2232 extends the reach of 10Gig Ethernet/FCoE to distributed line card (ToR)
- Support for up to 512 10Gig/FCoE attached hosts managed by a single Nexus 5500
- Nexus 5500 is the FCF
- Currently Nexus 2232 needs to be *single* homed to upstream Nexus 5500 (straight through N2K) to ensure SAN 'A' and SAN 'B' isolation
- Server Ethernet driver connected to the FEX in NIC Teaming (AFT, TLB) or with vPC (802.3ad)



Multi-hop FCOE

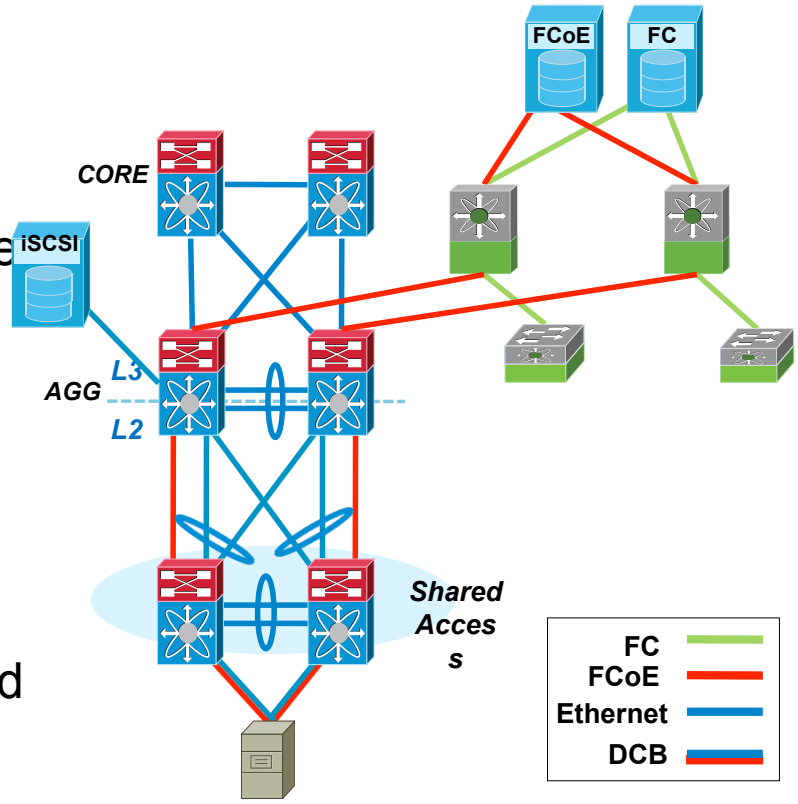
Shared Access-attached Storage Network

- Unifying the Access is biggest benefit
- Common SAN to access FC and FCoE storage.
  - Cable once, then repurpose any storage on an as-needed basis.
  - Facilitates end-to-end FCoE.
  - Connectivity to existing SAN.
  - Take advantage of next-gen storage without a SAN upgrade
- Maintain storage fabric segregation.
- Nexus 5000 in NPV or Switch mode.
  - NPV especially important when interoperating with another vendor



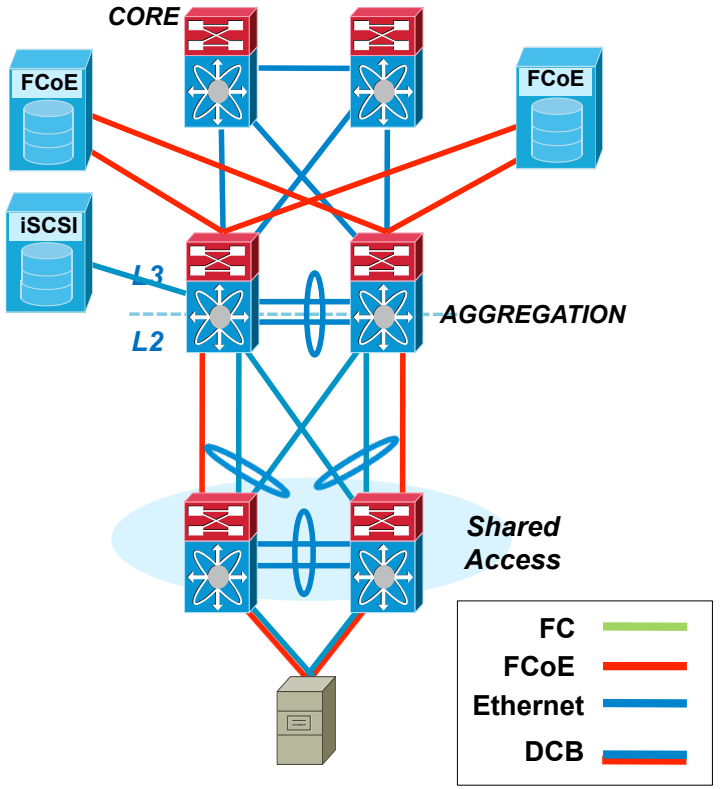
Aggregation-attached Storage Network

- Storage traffic utilizing high performance, highly available Ethernet aggregation switch.
- Increased scalability at the aggregation layer.
  - Enable larger FCoE deployments while maintaining legacy FC storage.
  - Repurpose any storage, increase storage utility and TCO for the customer.
- Maintain Isolation of Storage traffic.
  - Dedicated FCoE storage link.
  - Dedicated Virtual Device Context (VDC) for storage, segregates control traffic.
  - VDCs offer process level redundancy and fault isolation.
- Requires FCoE support on Nexus 7000 (~Q1CY11)



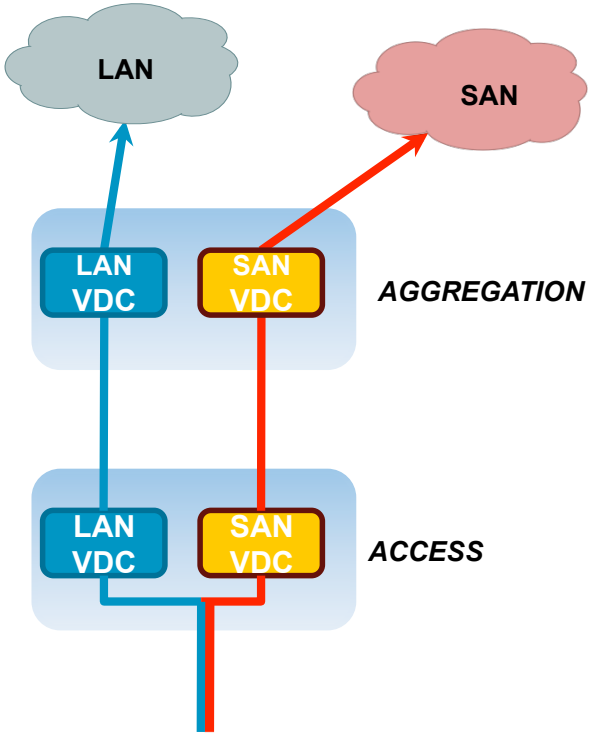
Converged Storage Network

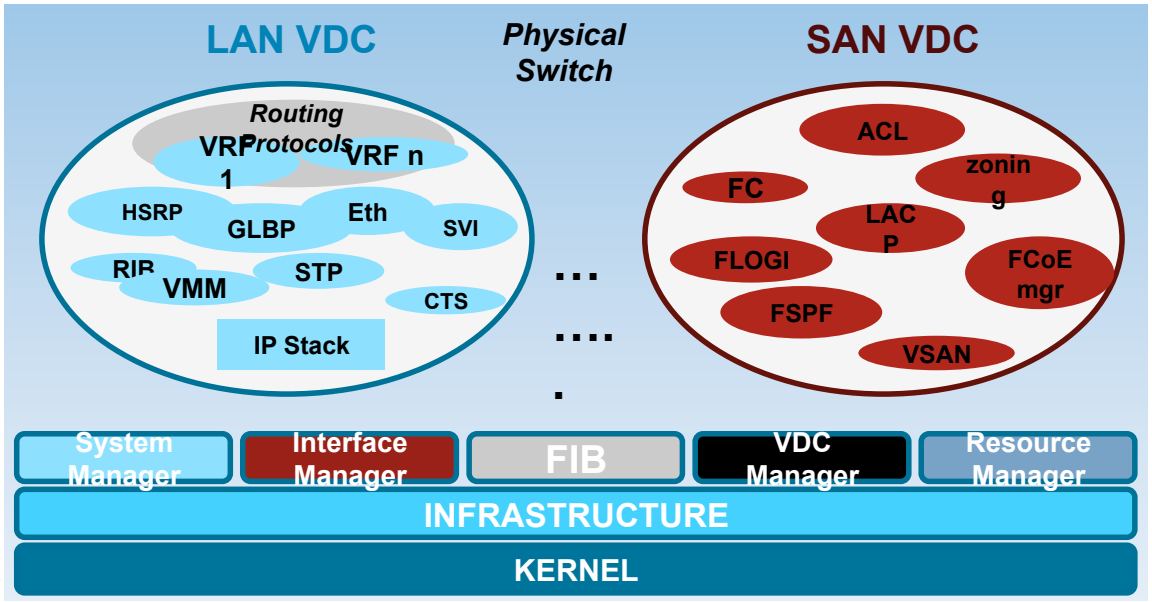
- Nexus 7000 as an Ethernet Storage director.
  - Providing High Availability at the supervisor, link and process levels.
- Maintain fabric segregation of storage traffic on a common infrastructure.
  - Without the need to deploy new hardware.
  - Control plane separation of storage traffic using VDCs.
  - Use ETS to allocate dedicated bandwidth to iSCSI and NAS as well as FCoE.
- Enables deployment of a single network for multiple different storage devices.
  - Flexibility in storage allocation and therefore increased utilization.



Storage Virtual Device Context

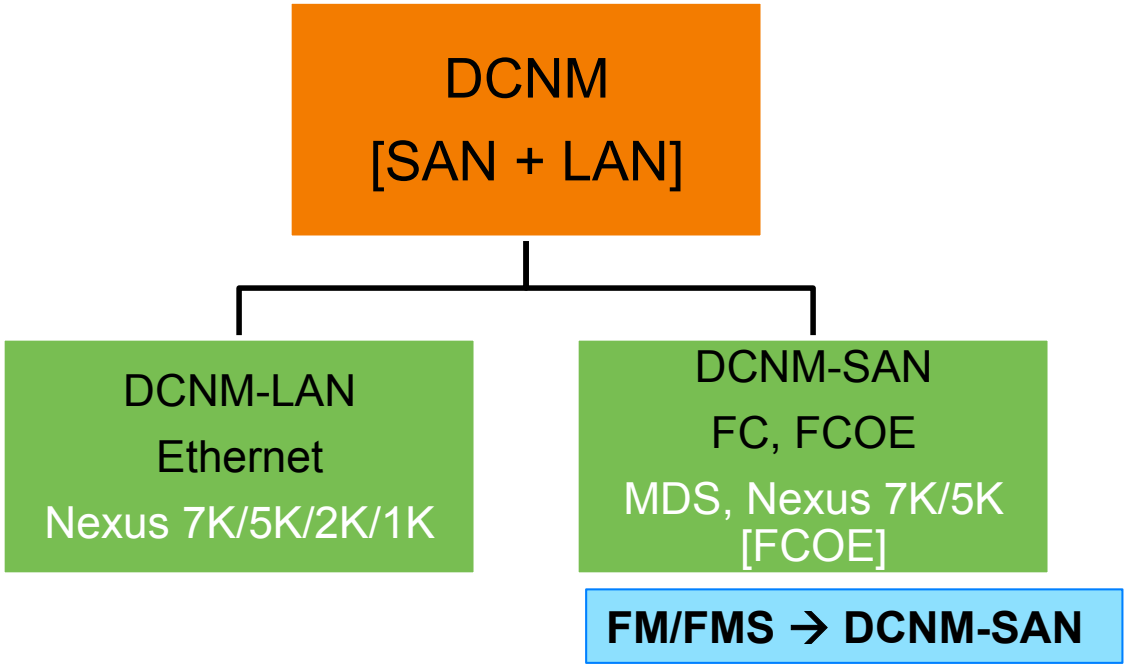
- Enhancements to Nexus7000 treatment of VDCs.
  - Allows an interface to be bound to a LAN and SAN VDCs.
  - Traffic is separated based on Ether-Type of the frame.
- VDCs offer Fault Isolation for Higher Availability.
  - Storage control traffic is unaffected by exceptions caused by other misbehaving processes.





DCNM Evolution

- DCNM-SAN & DCNM-LAN convergence.
- DCNM-SAN GUI is the familiar FM client GUI.
- FCoE wizard



Q & A