Cisco Expo 2008

Data Center
Unified Fabric: the new network for Data Center

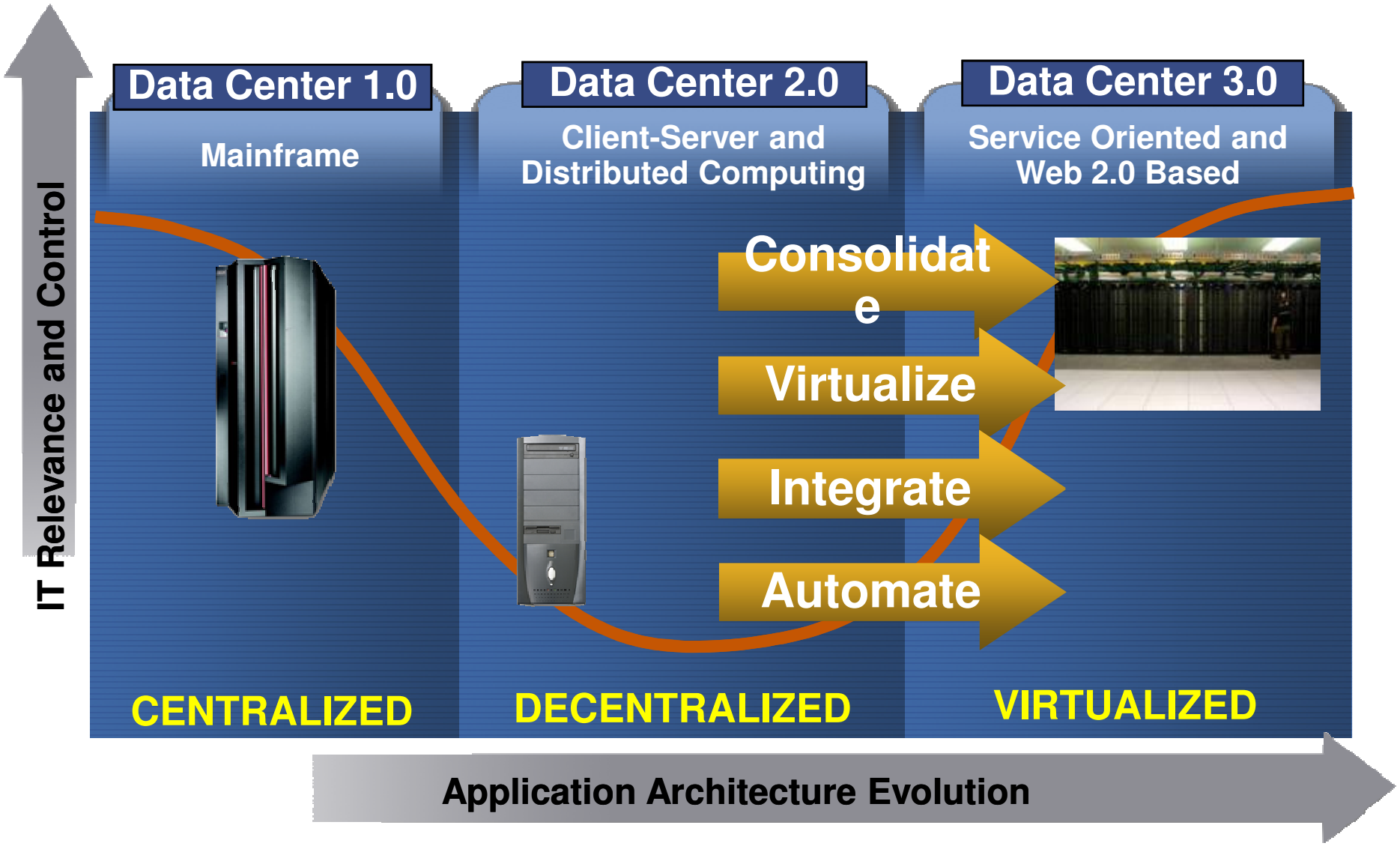
Maciej Bocian

mbocian@cisco.com

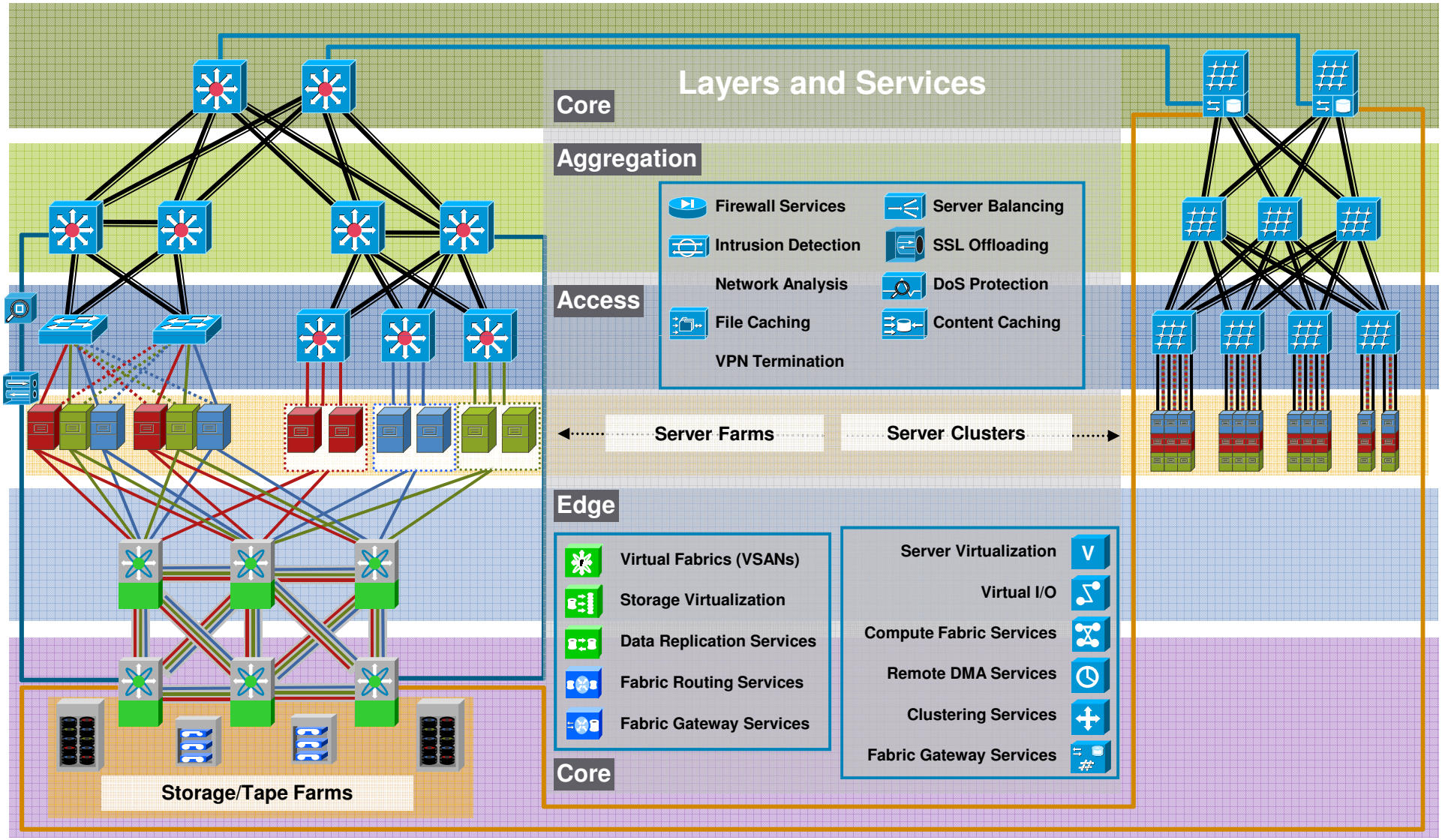
Consulting System Engineer
Data Center/Storage Networking

CCIE#7785

Data Center and Network Evolution = Growth

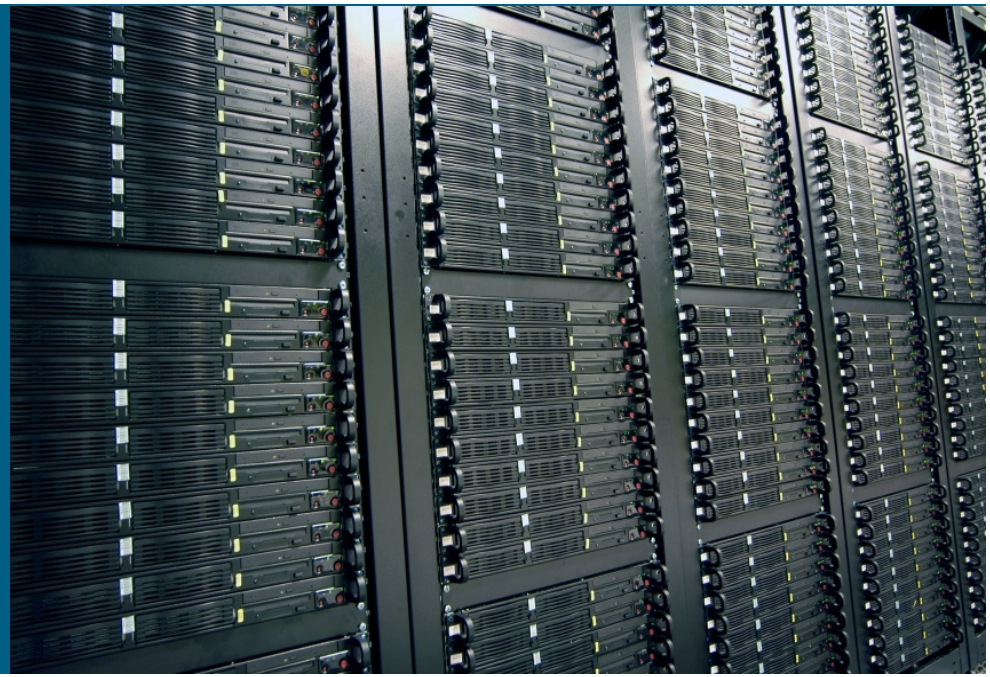


Current Data Center Architecture



Ethernet Networking

High density 10G Ethernet
Nexus 7000



Nexus 7000: First In Class

Data Center Class Platform

- Multi-Terabit system
- 550Gb/slot capable
- Optimized for 10 / 40 / 100 Gbps interfaces
- Extreme availability
- Multi-protocol (Ethernet, Storage and Unified I/O)

Data Center Class Operating System

- Self Healing Operating system
- Graceful system operation
- Virtualized Control Plane and Data Plane
- Fully Modular
- Security

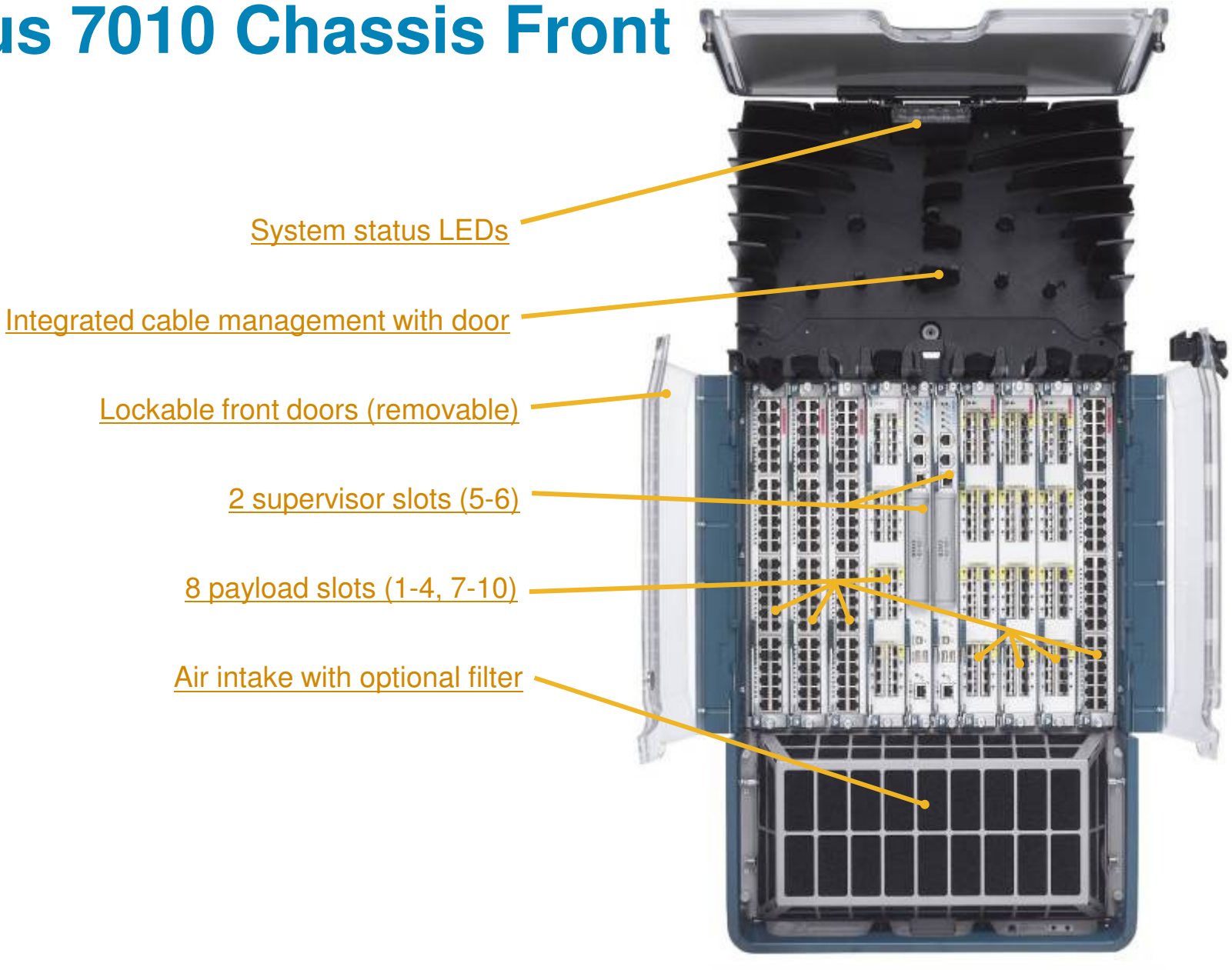
Data Center Network Manager (DCNM)

- Unified Data Center Manager
- Configuration / Provisioning / Service Enablement / Network Ops / Status / Statistics / Event Management
- Powerful feature rich web services API (XML)

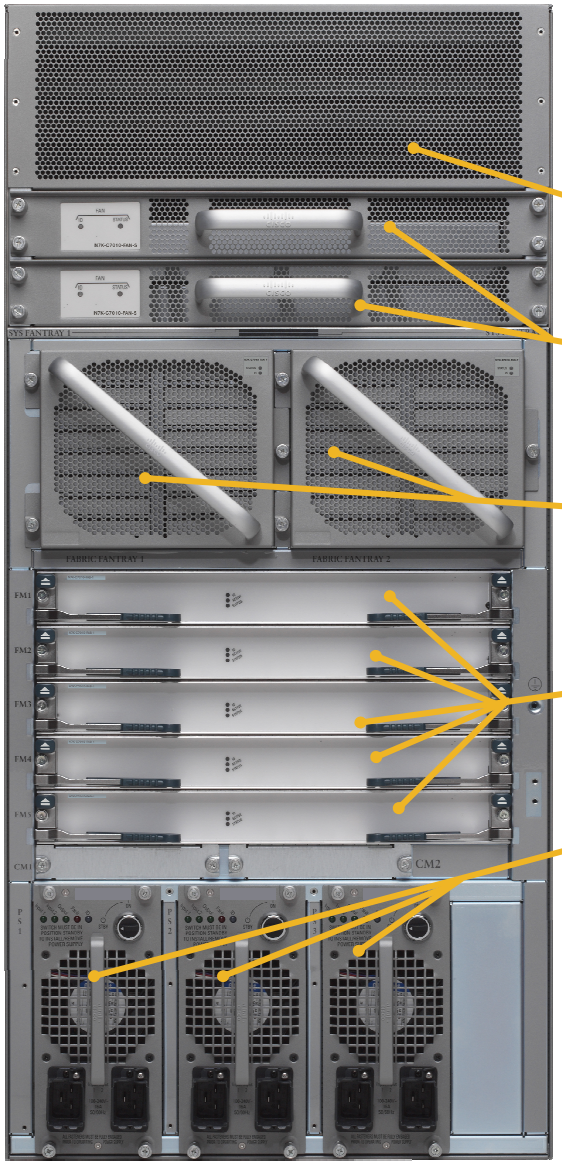
Nexus 7010 10-Slot Chassis

- First chassis in Nexus 7000 product family
- Optimized for data center environments
- High density
  - 256 10G interfaces per system
- High performance
  - 1.2Tbps system bandwidth at initial release
  - 80Gbps per slot
  - 60Mpps per slot
- Future proof
  - Initial fabric provides up to 4.1Tbps
  - Product family scaleable to 15+Tbps
  - 40/100G and Unified Fabric ready

Nexus 7010 Chassis Front



Nexus 7010 Chassis Back



Air exhaust

2 system fan trays

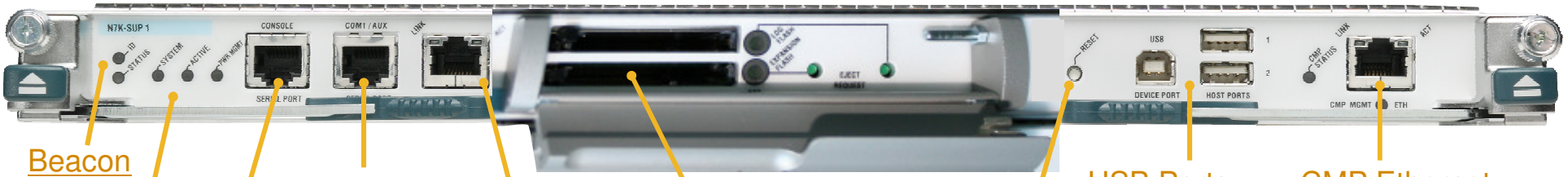
2 fabric fan trays

5 crossbar fabric modules

3 power supplies

Supervisor Engine

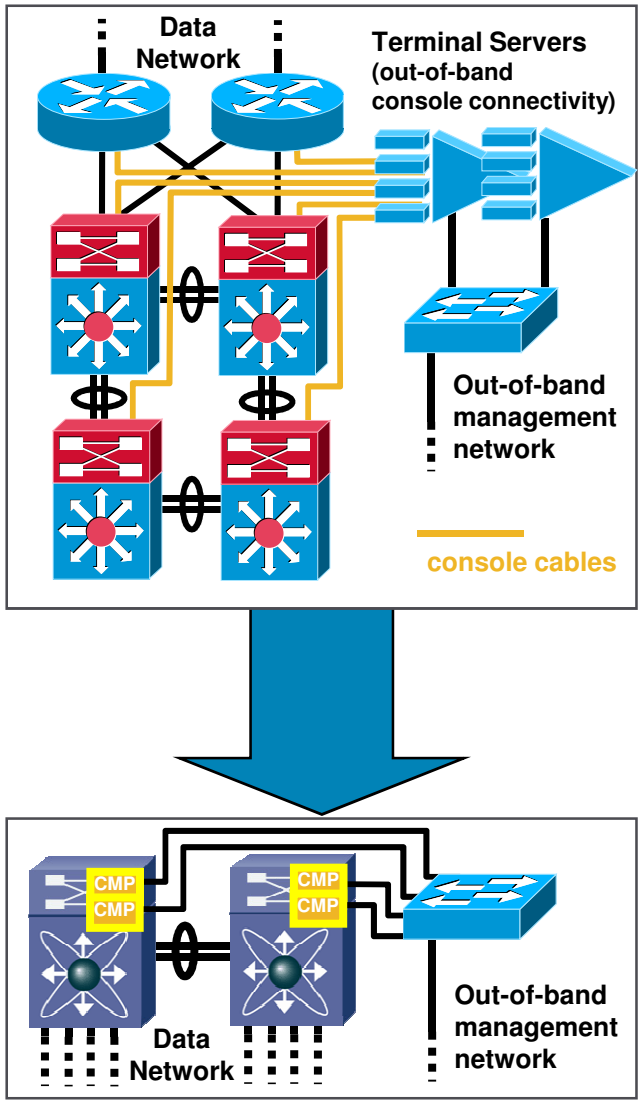
- Dual-core 1.66GHz Intel Xeon processor with 4GB DRAM
- Connectivity Management Processor (CMP) for lights-out management
- 2MB NVRAM, 2GB internal bootdisk, 2 external compact flash slots
- 10/100/1000 management port with 802.1AE LinkSec
- Console & Auxiliary serial ports
- USB ports for file transfer
- Blue beacon LED for easy identification



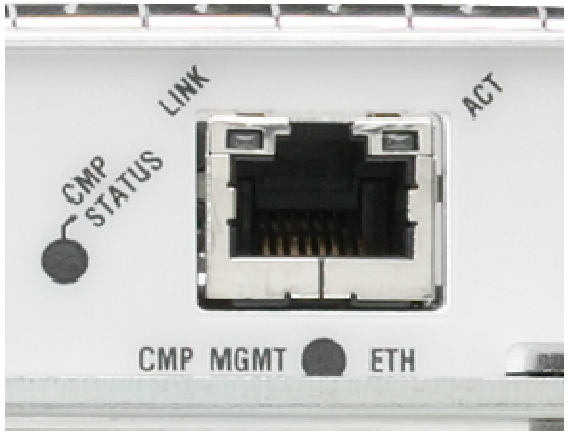
Management Ethernet Interface

- 10/100/1000 interface used exclusively for system management
- Belongs to dedicated “management” VRF
  - Prevents data plane traffic from entering/exiting from mgmt0 interface
  - Cannot move mgmt0 interface to another VRF
  - Cannot assign other system ports to management VRF
- Capable of IEEE 802.3ae LinkSec encryption (not enabled in 4.0 release)

Connectivity Management Processor (CMP)



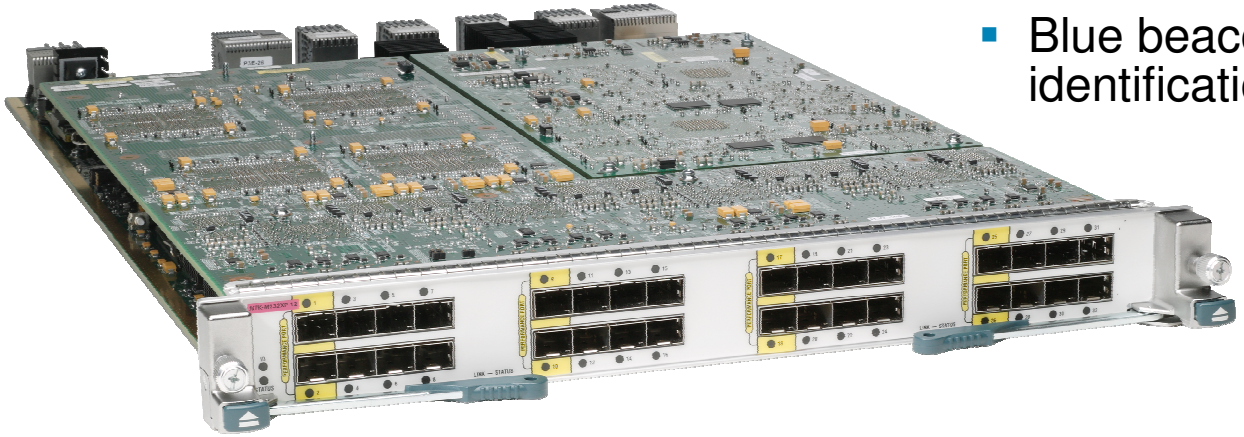
- Standalone, always-on microprocessor on supervisor engine
- Provides 'lights out' remote management and disaster recovery via 10/100/1000 interface
  - Removes need for terminal servers
- Monitor supervisor and modules, access log files, power cycle supervisor, etc.
  - Runs lightweight Linux kernel and network stack
  - Completely independent of DC-OS on main CPU



32-Port 10GE I/O Module

- 32 10GE ports with SFP+ transceivers
- 80G full duplex fabric connectivity
- Integrated 60Mpps forwarding engine for fully distributed forwarding
- 4:1 oversubscription at front panel

- Virtual output queueing (VOQ) ensuring fair access to fabric bandwidth
- 802.1AE LinkSec on every port
- Buffering:
  - Dedicated mode: 100MB ingress, 80MB egress
  - Shared mode: 1MB + 100MB ingress, 80MB egress
- Queues: 8q2t ingress, 1p7q4t egress
- Blue beacon LED for easy identification

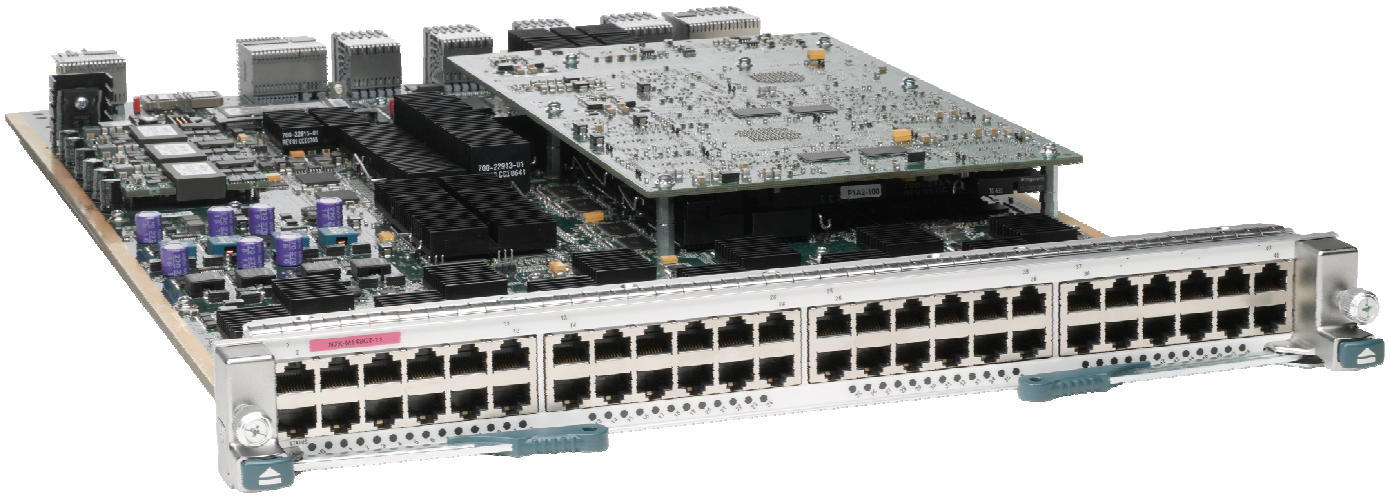


SFP+

SR – 300m over MMF
LR – 10km over SMF

48-Port 1GE I/O Module

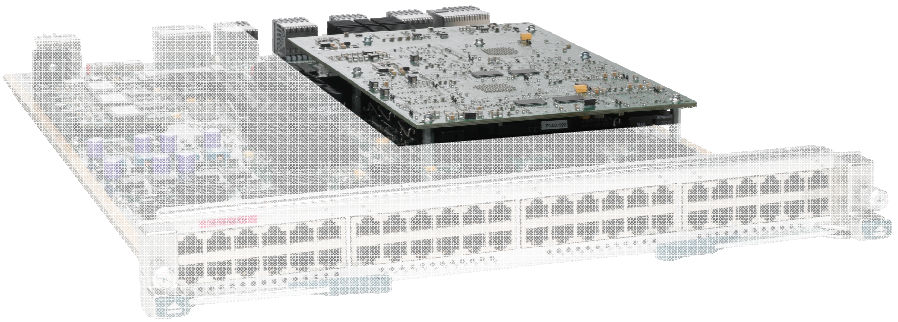
- 48 1GE 10/100/1000 RJ-45 ports
- 40G full duplex fabric connectivity
- Integrated 60Mpps forwarding engine for fully distributed forwarding
- Virtual output queueing (VOQ) ensuring fair access to fabric bandwidth
- 802.1AE LinkSec on every port
- Buffer: 7.5MB ingress, 6.2MB egress
- Queues: 2q4t ingress, 1p3q4t egress
- Blue beacon LED for easy identification



Forwarding Engine Hardware

Advanced hardware forwarding engine integrated on every I/O module

- 60Mpps Layer 2 bridging with hardware MAC learning
- 60Mpps IPv4 and 30Mpps IPv6 unicast
- IPv4 and IPv6 multicast (SM, SSM, bidir)
- IPv4 and IPv6 security ACLs
- Cisco TrustSec security group tag support
- Unicast RPF check and IP source guard
- QoS remarking and policing policies
- Ingress and egress NetFlow (full and sampled)
- GRE tunnels



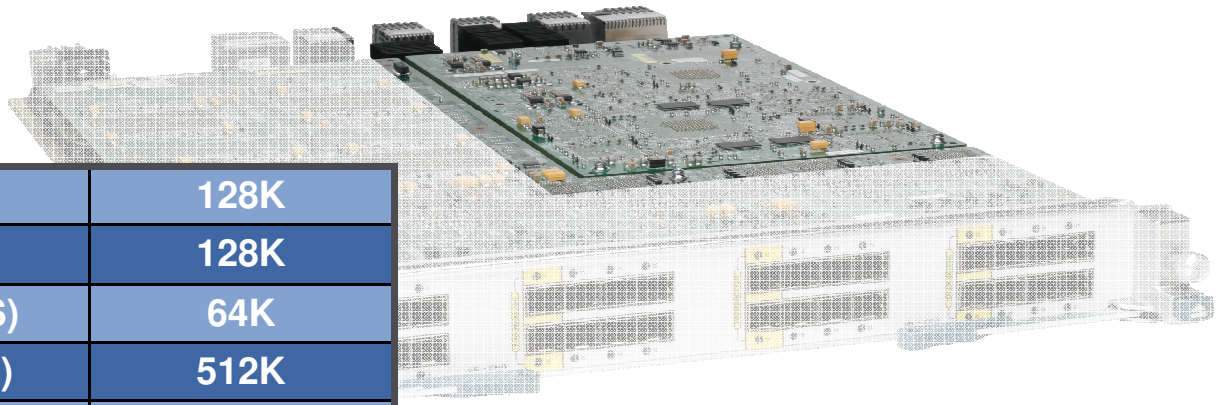
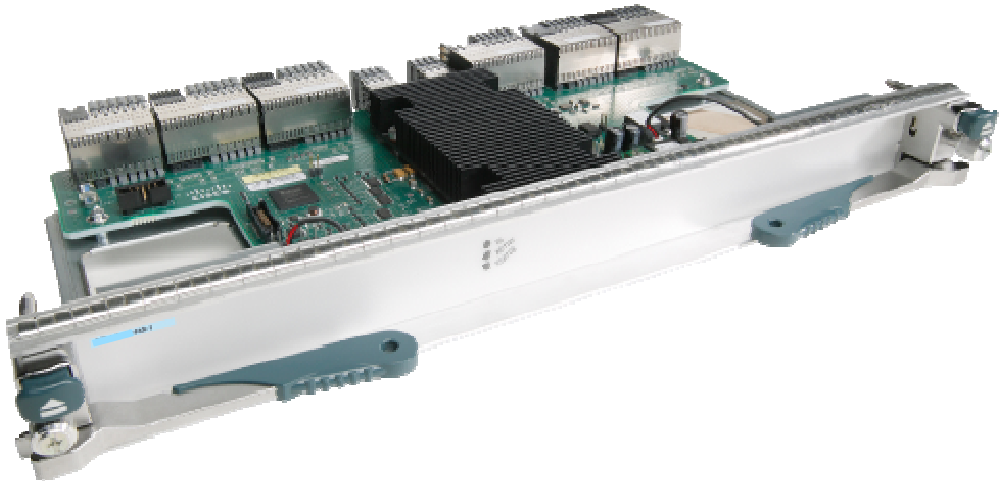


Table sizes optimized for Data Center

FIB TCAM	128K
MAC table	128K
Classification TCAM (ACL and QoS)	64K
NetFlow Table (Ingress and Egress)	512K
Policers	16K

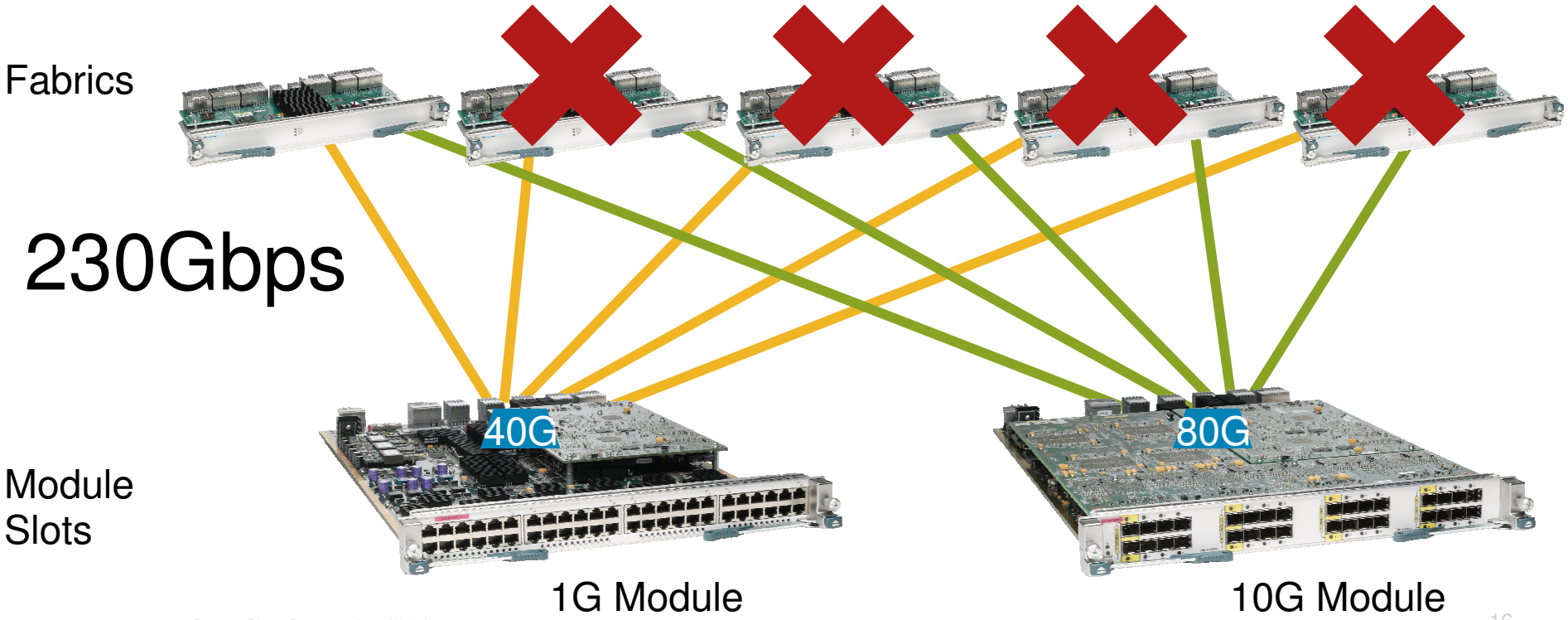
Fabric Module

- Provides 46Gbps per I/O module slot
  - Also provides 23G per supervisor slot
- Up to 230Gbps per slot with 5 fabric modules
  - Initially shipping I/O modules do not leverage full fabric bandwidth
- Load-sharing across all fabric modules in chassis
- Multilevel redundancy with graceful performance degradation
- Non-disruptive OIR
- Blue beacon LED for easy identification



Fabric Capacity and Redundancy

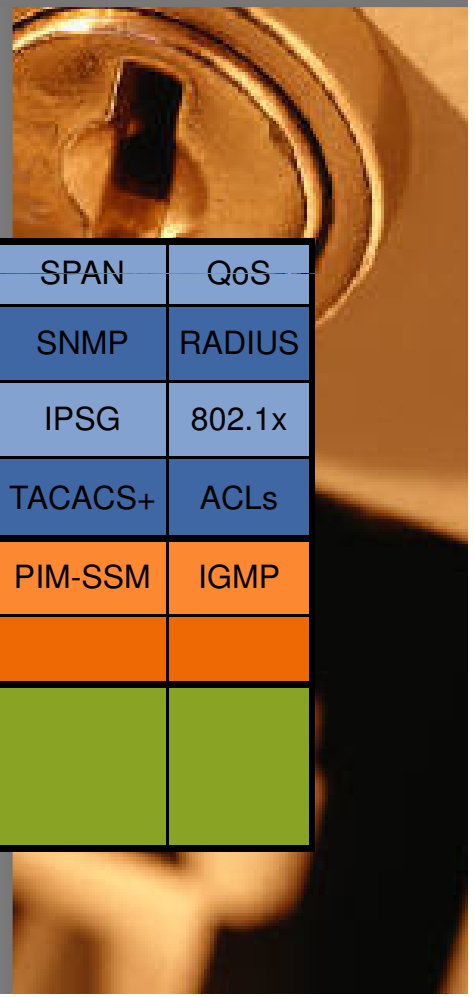
- Per-slot bandwidth capacity increases with each fabric module
- 1G module requires 2 fabrics for N+1 redundancy
- 10G module requires 3 fabrics for N+1 redundancy
- 4th and 5th fabric modules provide additional level of redundancy
- Future modules will leverage additional fabric bandwidth
- Fabric failure results in reduction of overall system bandwidth



NX-OS Licensing

Simple, Flexible Licensing Model

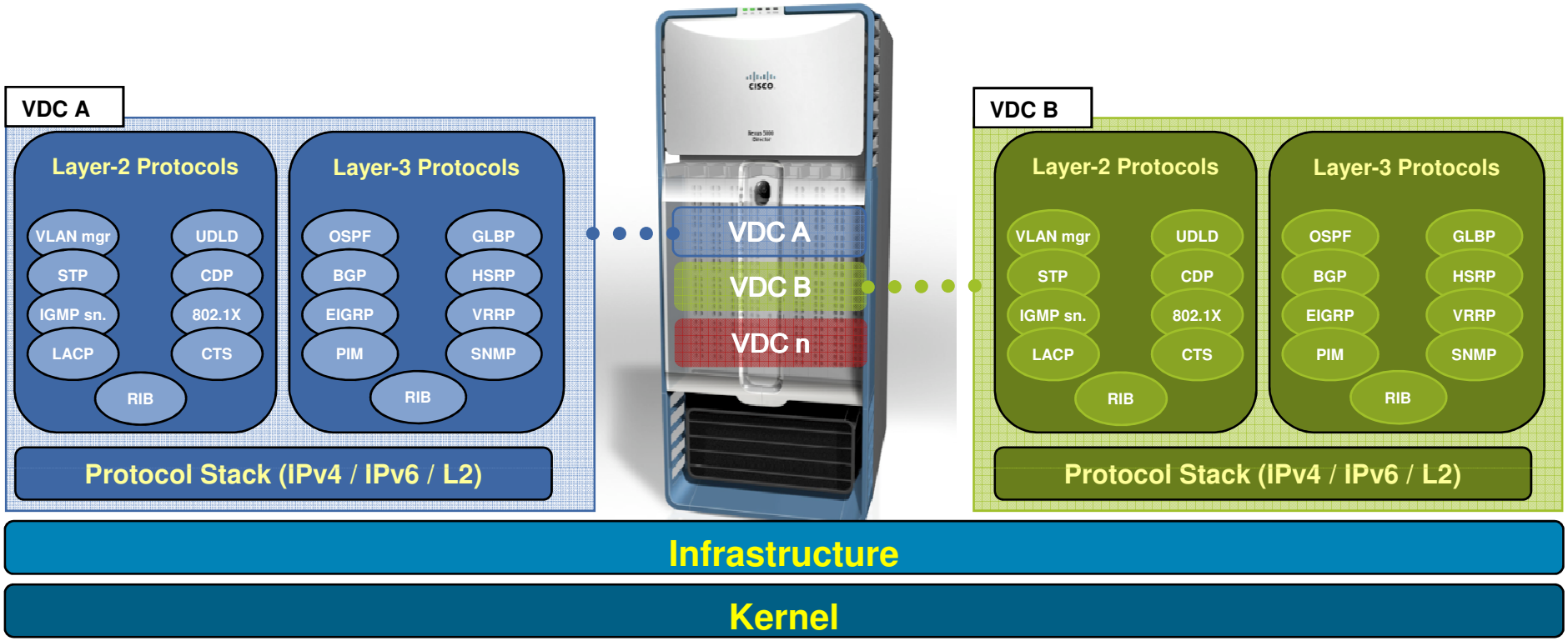
- There are three levels of enforced licensing: Base, Enterprise Services, and Advanced Services
- Grace periods facilitate feature testing and trials without buying a license (for example, 120 days), with some restrictions. The Cisco Trusted Security does not have a grace period because of export restrictions on strong cryptography



Base	ISSU	PVRST+	MSTP+	802.1Q	LACP	PVLANs	NetFlow	SPAN	QoS
	RIP/RIPng	IGMP snooping	DHCP helper	uRPF check	Port Security	SSHv2	RBAC	SNMP	RADIUS
	HSRP	GLBP	VRRP	VRF lite	CoPP	DHCP snooping	DAI	IPSG	802.1x
	Jumbo Frames	UDLD	Storm control	EEM	Cisco GOLD	Call Home	NAC	TACACS+	ACLs
Enterprise Services	OSPF	EIGRP	IS-IS	BGP	Graceful Restart	PIM-SM	Bidirectional PIM	PIM-SSM	IGMP
	MSDP	PBR	GRE						
Advanced Services	VDCs	Cisco Trusted Security							

Note: Enterprise Services is NOT included with Advanced Services license

Virtual Device Contexts (VDCs)

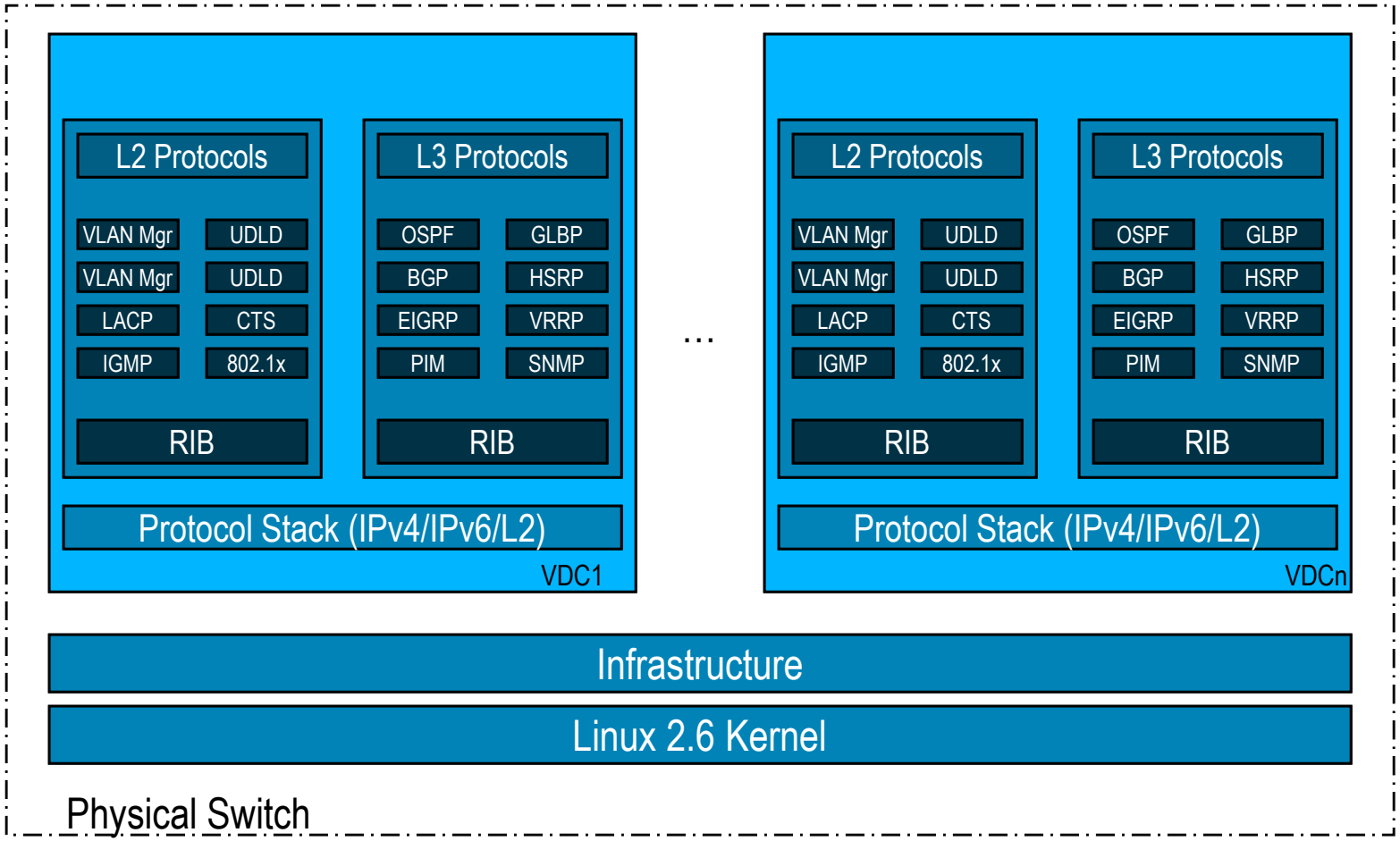


- **VDC – Virtual Device Context**
  - Flexible separation/distribution of **Software Components**
  - Flexible separation/distribution of **Hardware Resources**
  - Securely delineated **Administrative Contexts**

- **VDCs are not…**
  - The ability to run different OS levels on the same box at the same time
  - based on a ***hypervisor*** model; there is a single 'infrastructure' layer that handles h/w programming…

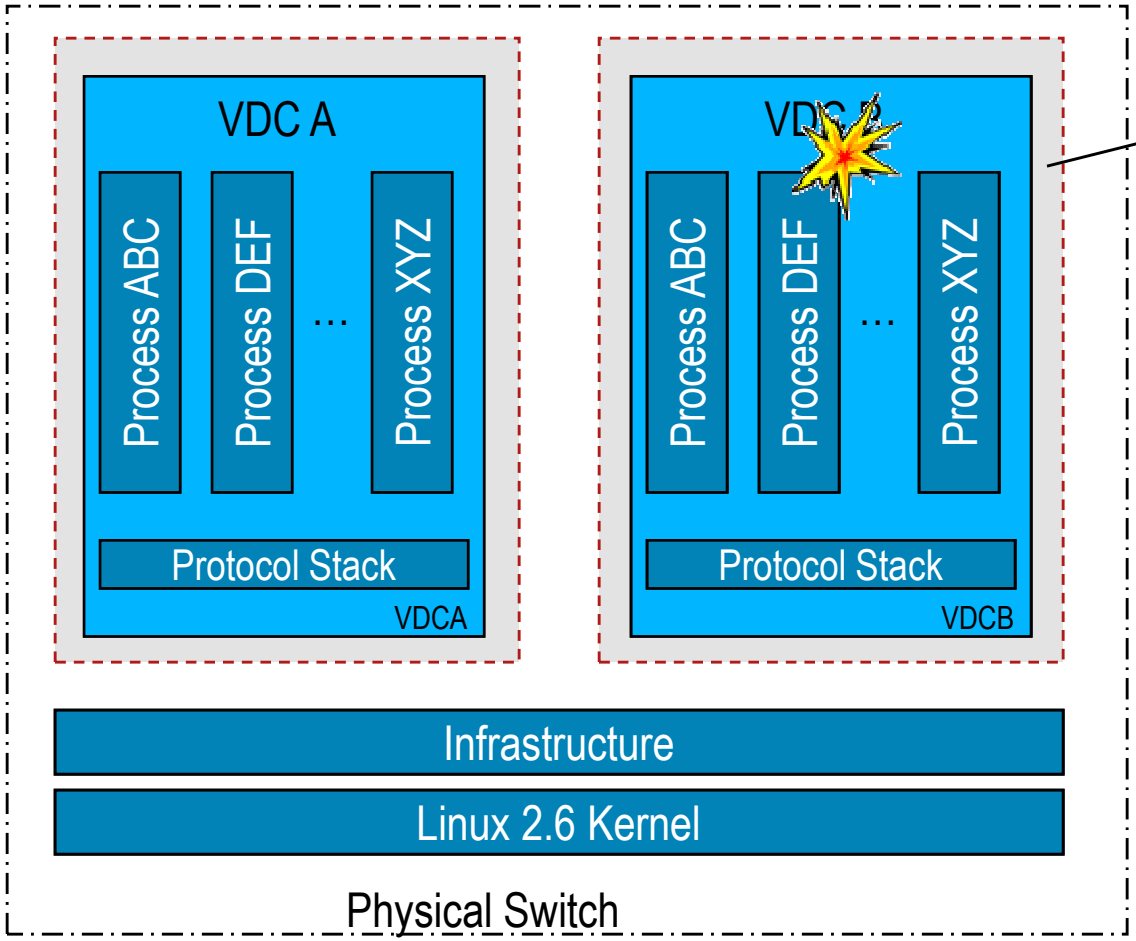
Virtual Device Contexts
An Introduction to the VDC Architecture

Virtual Device Contexts provides virtualization at the device level allowing multiple instances of the device to operate on the same physical switch at the same time…



Virtual Device Contexts
VDC Fault Domain

A VDC builds a fault domain around all running processes within that VDC - should a fault occur in a running process, it is truly isolated from other running processes and they will not be impacted…



Fault Domain

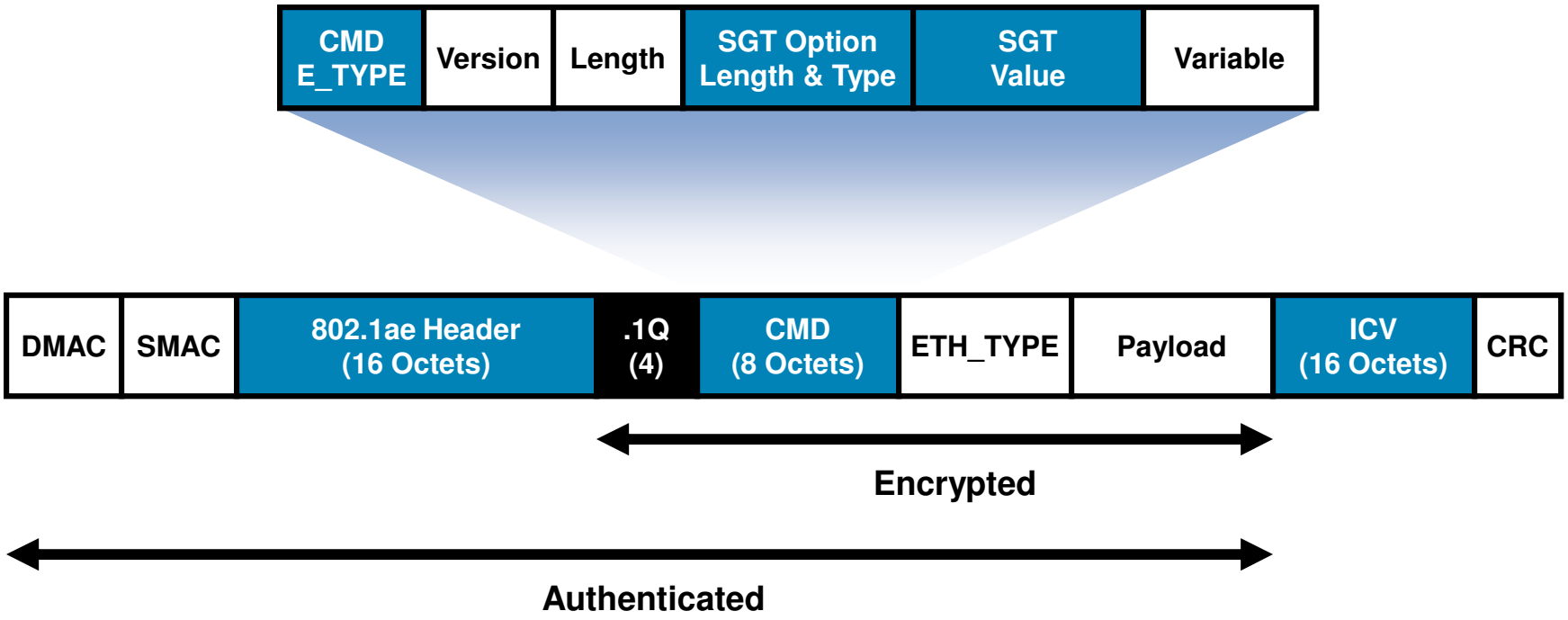
Process “DEF” in VDC B crashes

Processes in VDC A are not affected and will continue to run unimpeded

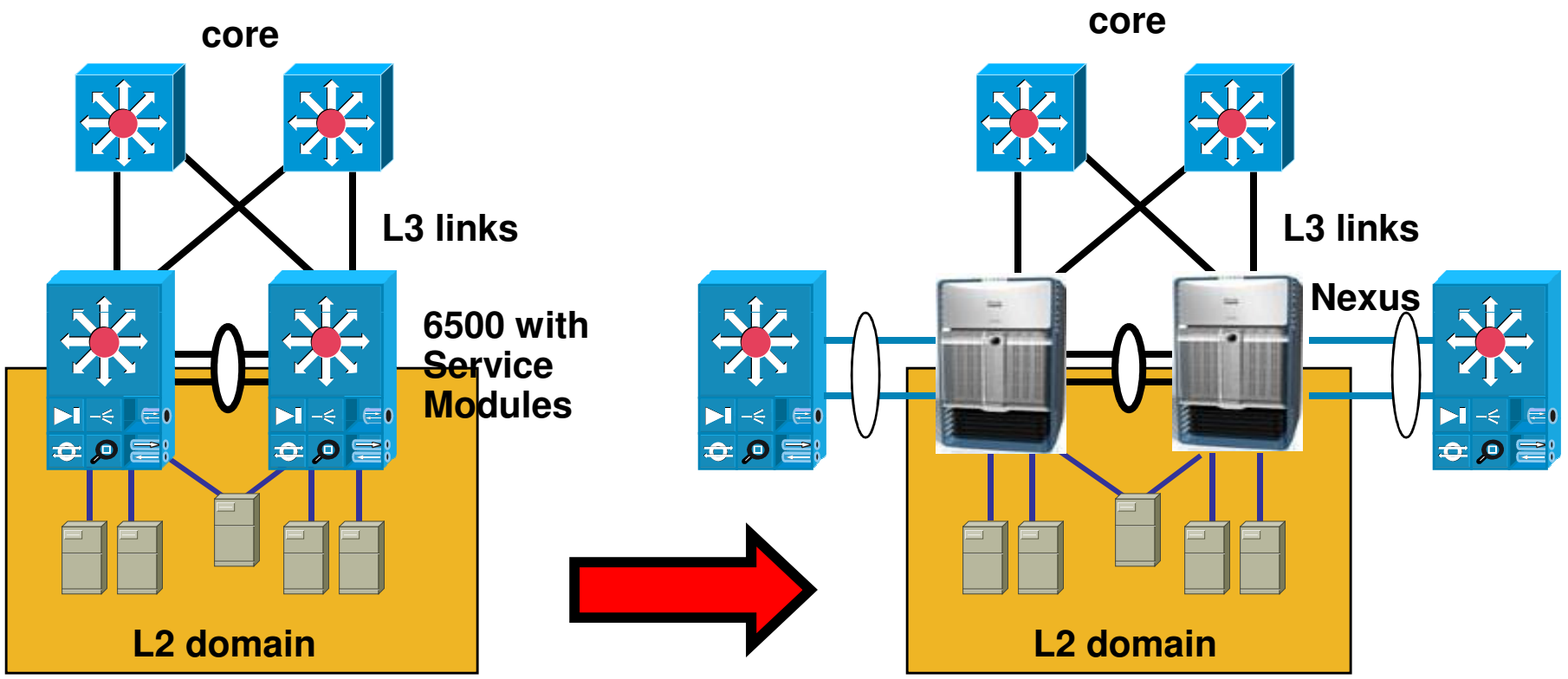
This is a function of the process modularity of the OS and a VDC specific IPC context

Cisco TrustSec on Nexus
TrustSec Linksec (802.1ae) Frame Format

The encryption used by TrustSec follows IEEE Standards-based LinkSec (802.1ae) encryption, where the upper layers are unaware of the L2 header/encryption. The key difference here is that TrustSec-capable devices also add an extra 8-byte field called Cisco MetaData (CMD), used to carry the SGT throughout the network…



Collapsed Aggregation/Access

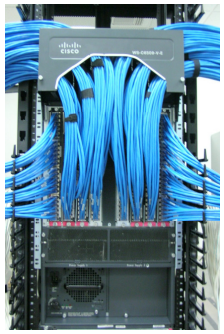


Cisco High End Switching Portfolio
Catalyst 6500 and Nexus 7000

Nexus 7000

230 GbE / Slot
10Gb optimized

15T Switching; 500G+ / slot
10G/40G/100G optimized
Unified Fabric (DCE)



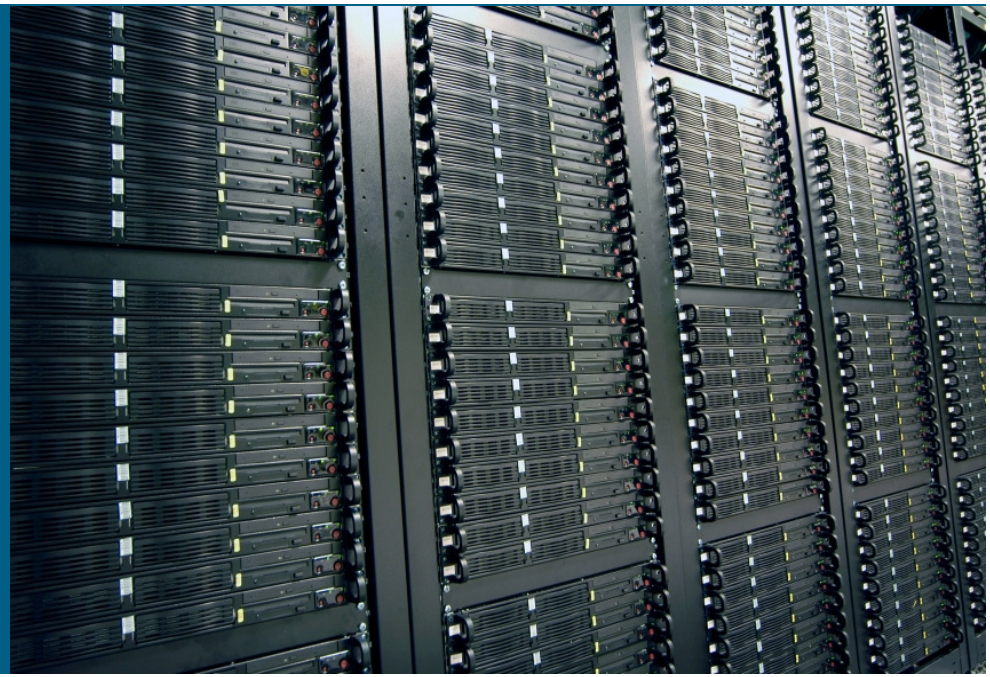
Catalyst 6500

40Gb / Slot
1Gb/10Gb optimized
Service modules

2T Switching; 80Gb+ / slot
1Gb/10Gb/40Gb optimized
Service Modules

Fibre Channel over Ethernet

Fibre Channel over Ethernet
High density & low cost 10GE
Nexus 5000



Introduction to FCoE

- IP networks are often built with Ethernet
- Usually Ethernet switches discard frames, while Fibre Channel switches do not
  - Thanks to the Buffer to Buffer credit mechanism
- Some optional Ethernet extensions allow Ethernet switches to not discard frames
  - The Pause mechanism defined in IEEE 802.3 Annex 31B
  - A mechanism different than the Buffer to Buffer credit, but able to achieve the same result
- If these extensions are implemented, Fibre Channel frames can be mapped directly on top of Ethernet
  - FCoE: Fibre Channel over Ethernet

What is Fibre Channel over Ethernet?

- From a Fibre Channel standpoint it's

  FC connectivity over a new type of cable called… an Ethernet cloud

- From an Ethernet standpoints it's

  Yet another ULP (Upper Layer Protocol) to be transported, but… a challenging one!

- And technically…

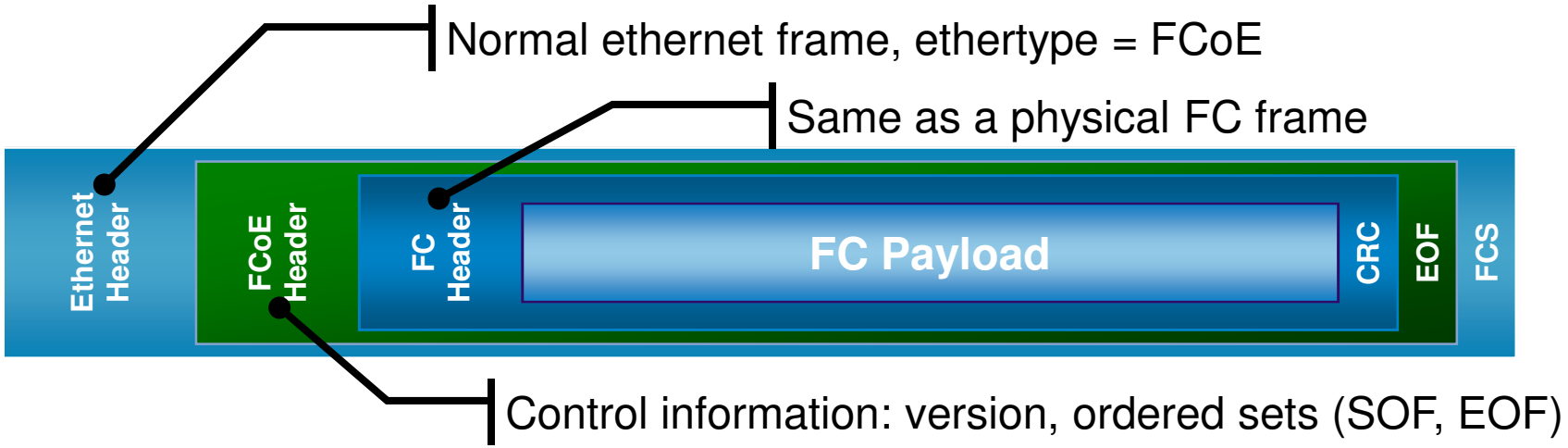
FCoE is an extension of Fibre Channel onto a Lossless Ethernet fabric

FCoE Enablers

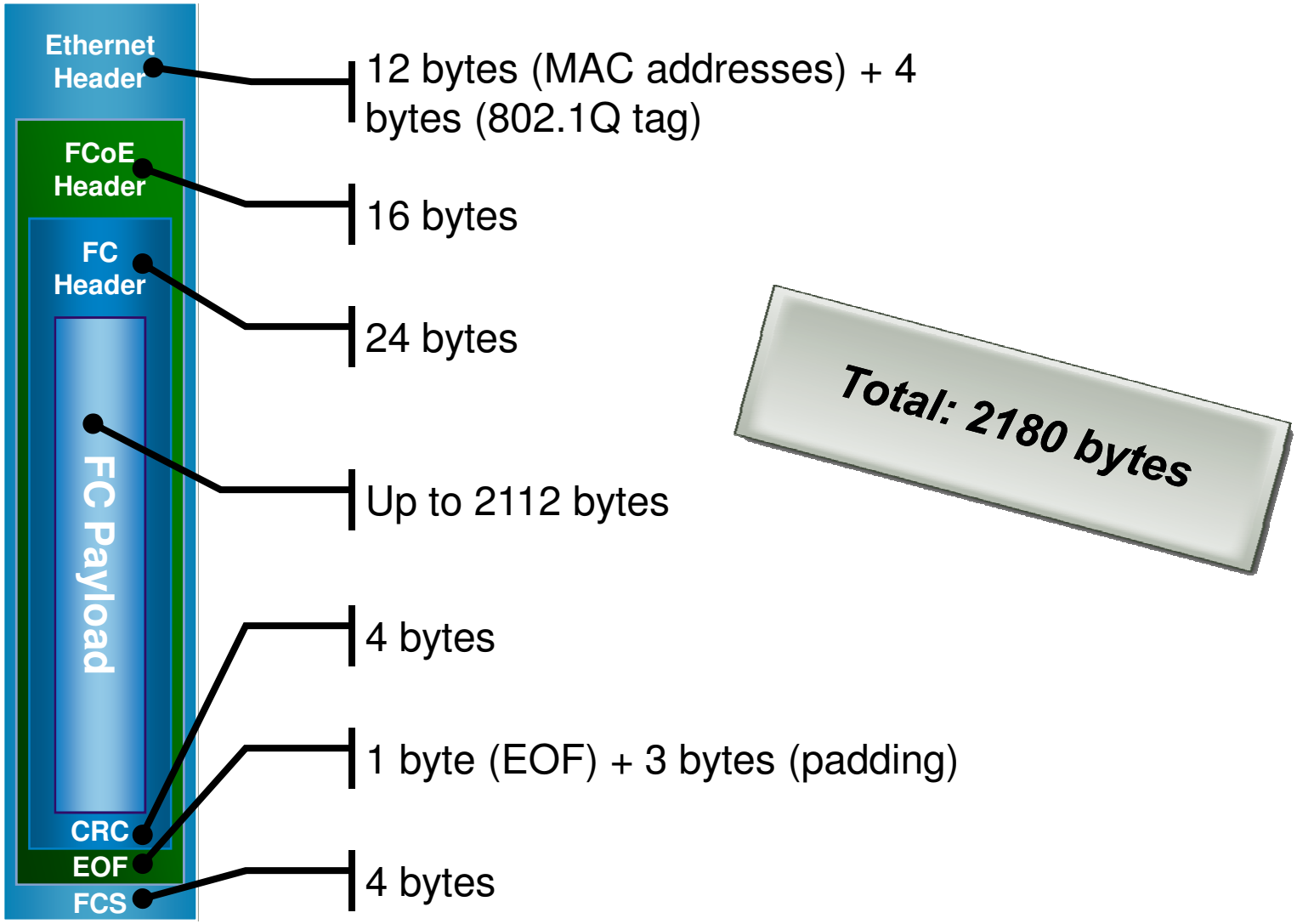
- 10Gbps Ethernet
- Lossless Ethernet

  Matches the lossless behavior guaranteed in FC by B2B credits
- Ethernet jumbo frames

  Max FC frame payload = 2112 bytes

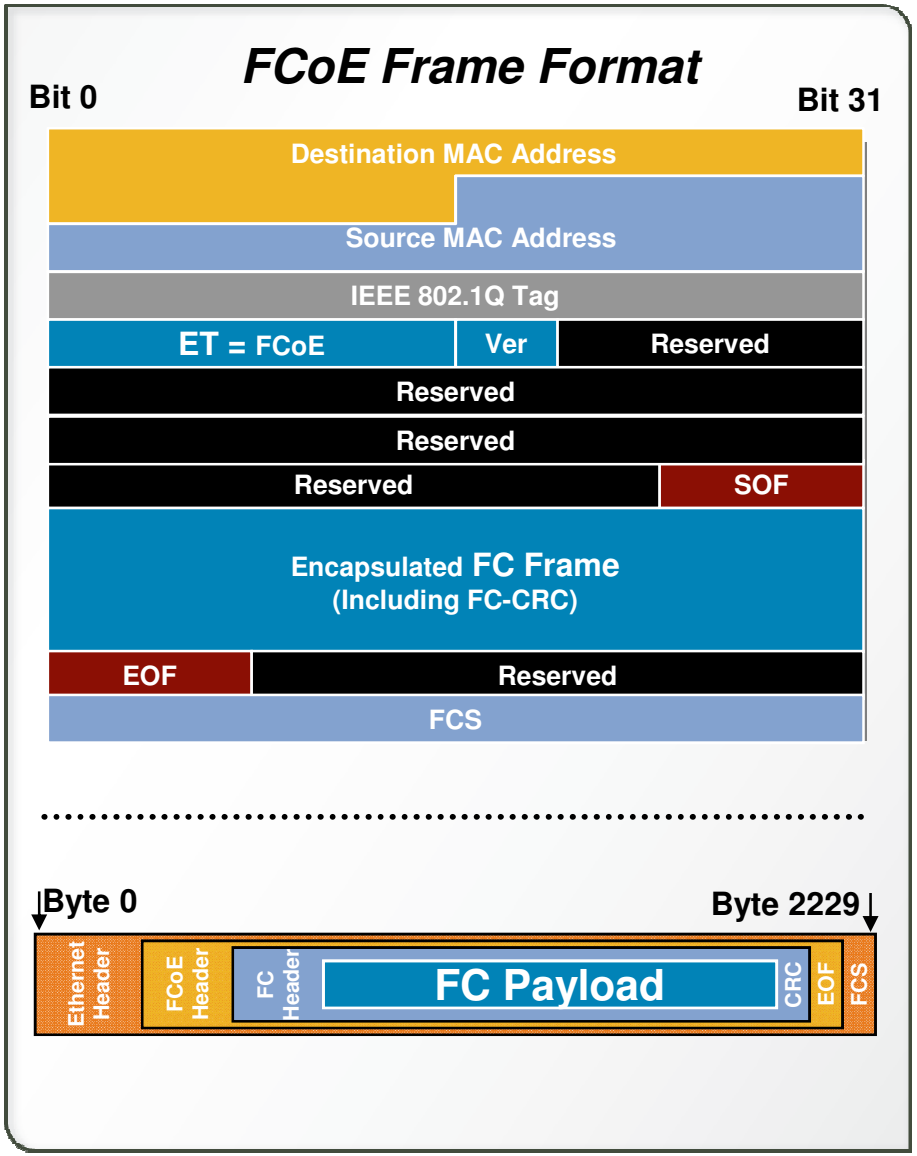


FCoE frame size



FCoE Frame Format Specification

- An extension of FC over lossless Ethernet
- FCoE Specification in ANSI INCITS FC T11.3
  - Frame Format agreement Aug. 2007
  - Target completion Summer 08
- Optional features being defined in IEEE
  - PAUSE enhancements 802.3x mechanisms
  - Priority Flow Control (PFC)
  - Jumbo Ethernet frames: FC encapsulation of 2112Bytes



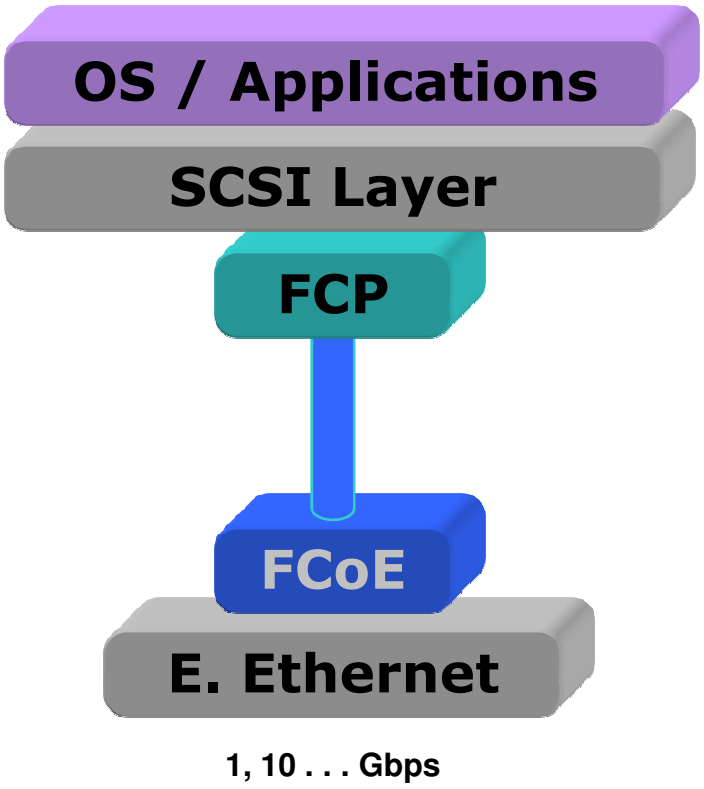
Introduction to FCoE (cont)

- FCoE appears as Fibre Channel to a host or a switch (and therefore to a user)
  - Easy to understand
    - Completely based on the FC model
  - Same operational model
    - Same host-to-switch and switch-to-switch behavior of FC
  - Same techniques of traffic management
    - E.g., in order delivery or FSPF load balancing, Multipathing
  - Same management and security models
    - WWNs, FC-IDs, hard/soft zoning, DNS, RSCN
  - Simplified troubleshooting
    - All the FC tools, plus all the Ethernet tools
  - Aligned with the FC-BB-2 model

Can Ethernet be lossless?

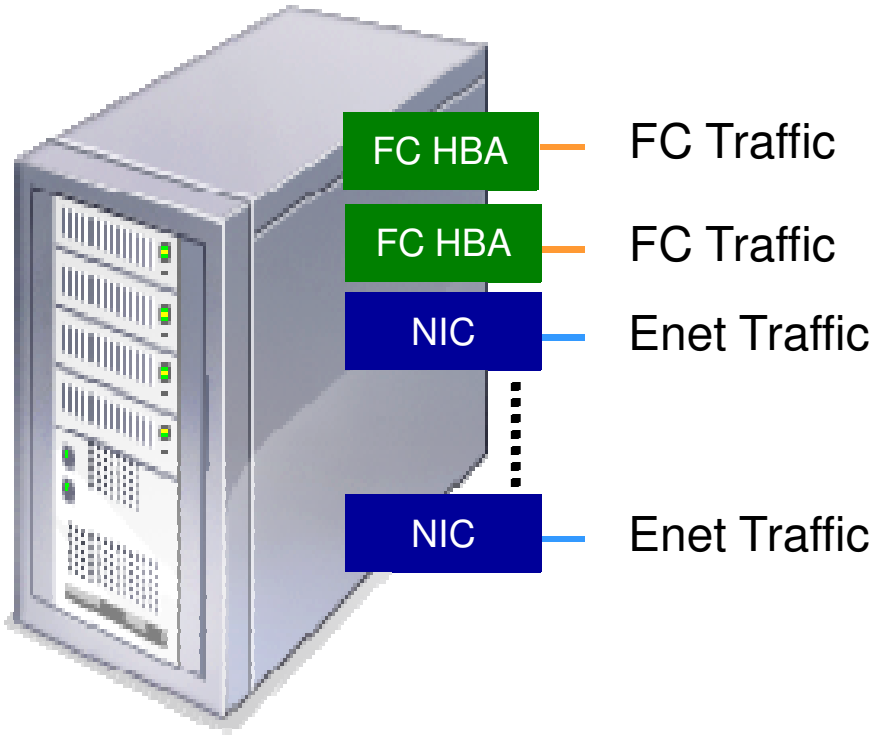
- Yes, with 802.3x Pause
  - For short distances equivalent to FC credits
- But… when customers turn it on, the results are confusing
  - Standard allows for asymmetric implementations
  - Vendor implementations are inconsistent
- Anyhow Pause is not enough
  - It applies to the whole link
- Cisco proposes Priority Flow Control
  - One pause per IEEE 802.1p priority code point
  - Public domain (no standard activity yet)
  - Embraced by many network gear vendors

Encapsulation technologies



- FCP layer is untouched
- Allows same management tools for Fibre Channel
- Allows same Fibre Channel drivers
- Allows same Multipathing software
- Simplifies certifications with OSMs
- Evolution rather than Revolution

Unified I/O Use Case

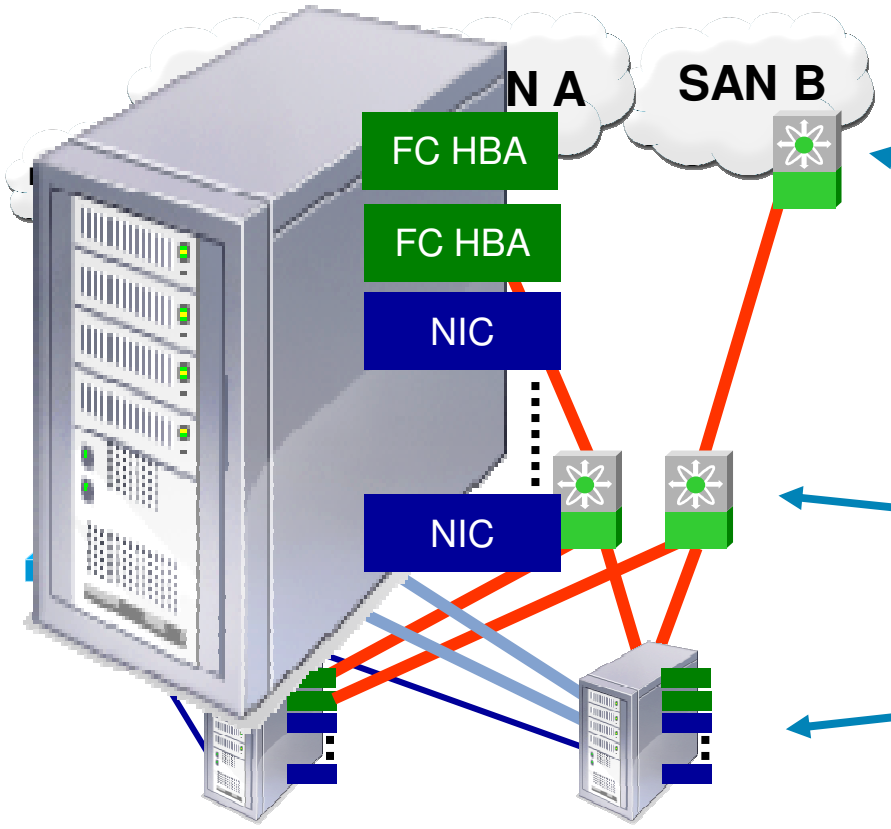


Today:

- Parallel LAN/SAN Infrastructure
- Inefficient use of Network Infrastructure
- 5+ connections per server – higher adapter and cabling costs
  - Adds downstream port costs; cap-ex and op-ex
  - Each connection adds additional points of failure in the fabric
  - Power and cooling
- Longer lead time for server provisioning
- Multiple fault domains – complex diagnostics
- Management complexity – firmware, driver-patching, versioning

Unified I/O Use Case

Today:

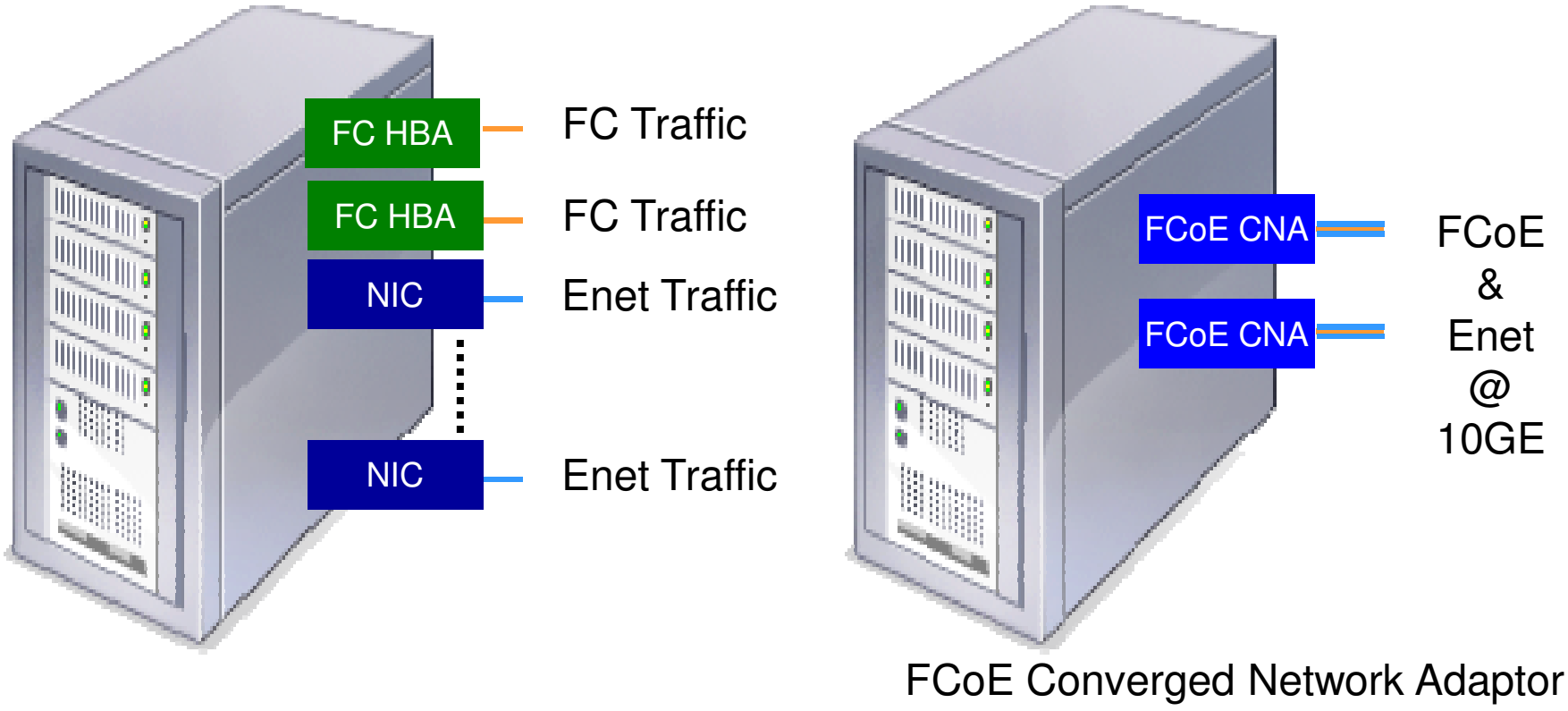


- Aggregation/Core switches
- Access – Top of the Rack switches
- Servers

Ethernet
FC

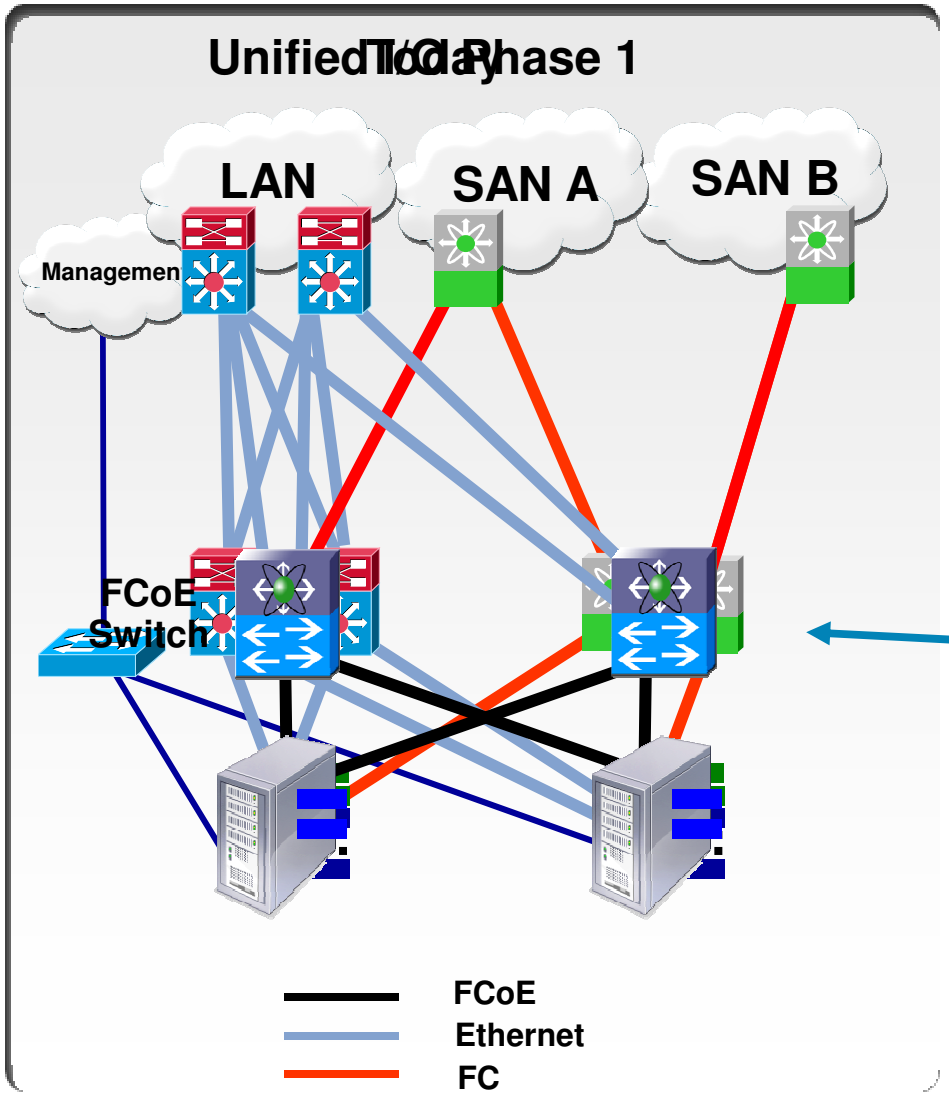
FCoE I/O Consolidation Benefit

Fewer HBA/NIC's per Server



Customers purchase fewer NIC's and HBA's

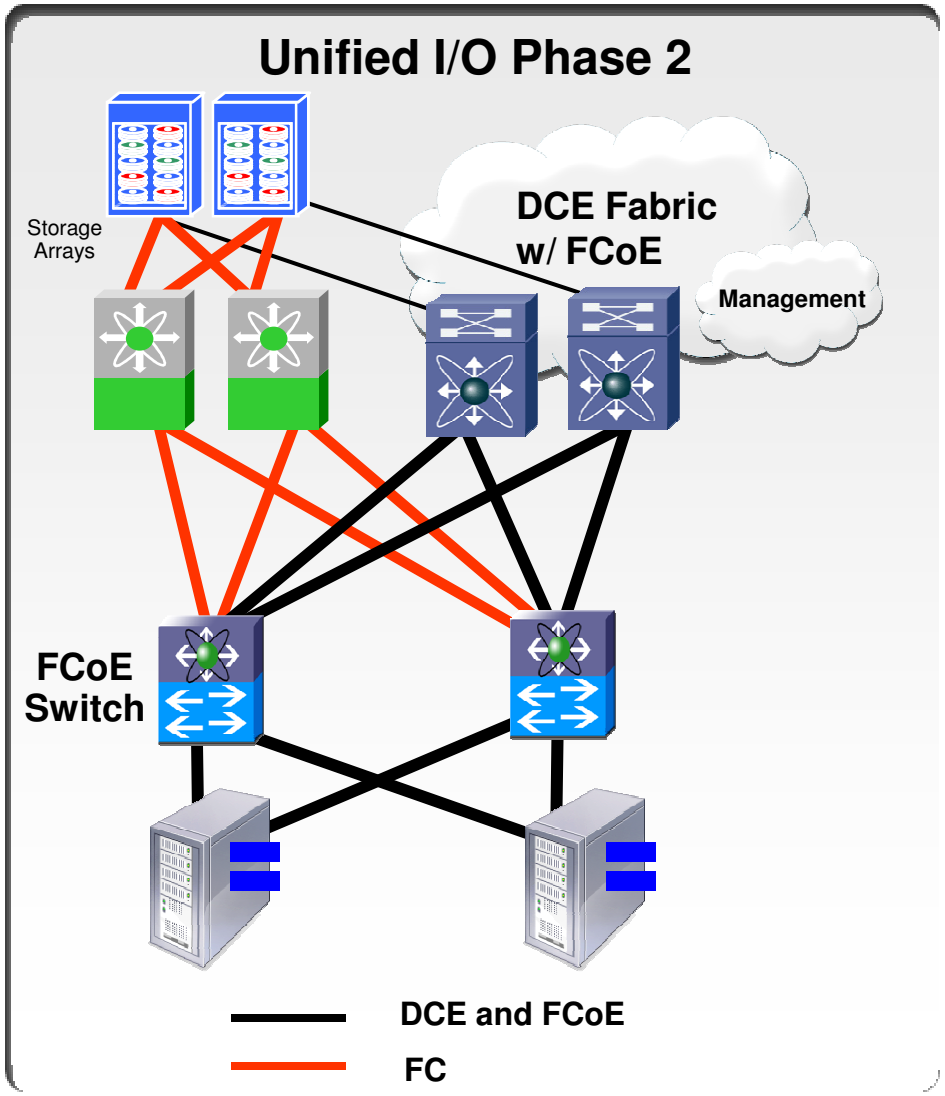
Unified I/O Use Case



Unified I/O Phase 1

- Reduction of server adapters
- Fewer Cables
- Simplification of access layer & cabling
- Gateway free implementation - fits in installed base of existing LAN and SAN
- L2 Multipathing Access – Distribution
- Lower TCO
- Investment Protection (LANs and SANs)
- Consistent Operational Model
- **One set of ToR Switches**

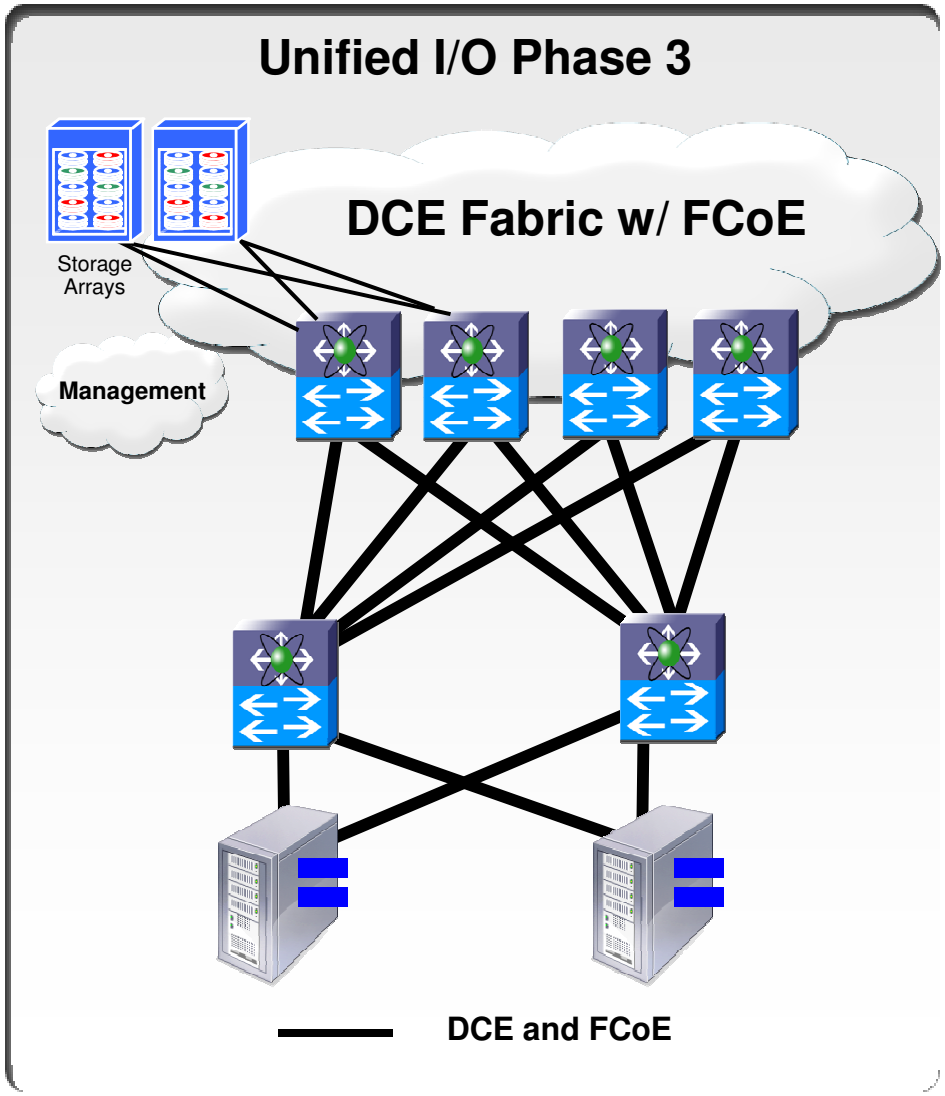
Unified I/O Use Case



Unified I/O Phase 2

- Elimination of parallel network infrastructure
- L2/L3 Multipathing end to end
- Faster infrastructure provisioning
- Lower TCO
- Disk array access via DCE or Native FC

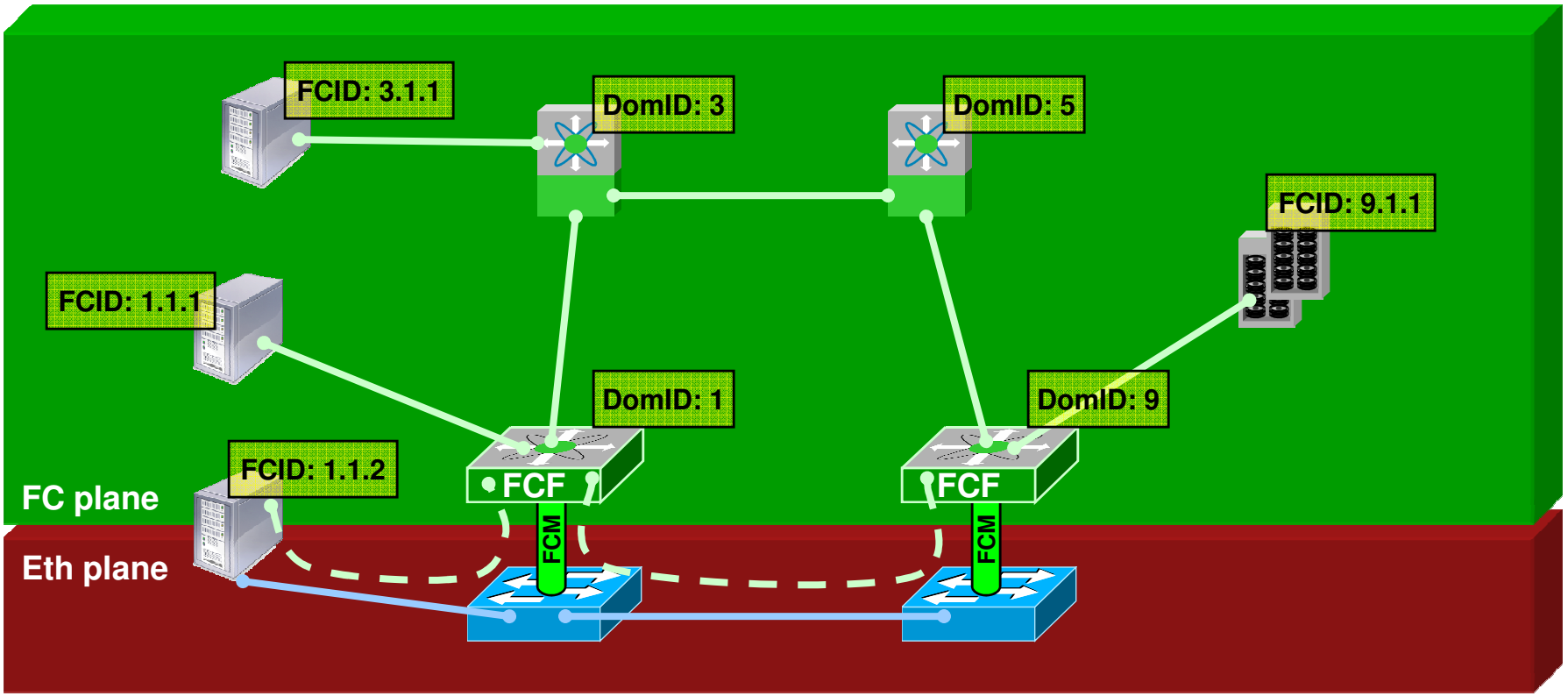
Unified I/O Use Case



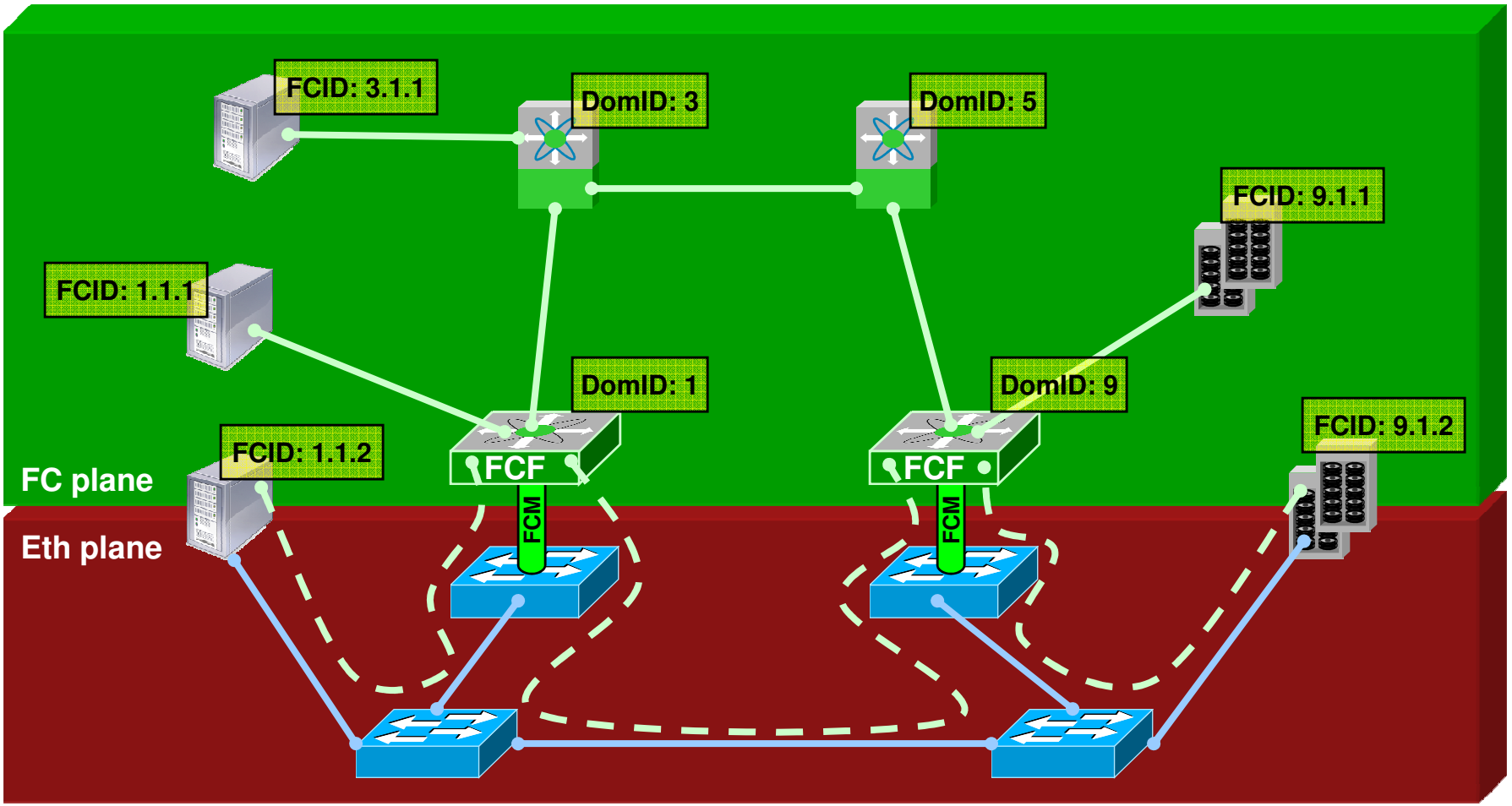
Unified I/O Phase 3

- Datacenter wide Unified Fabric for LAN and SAN
- L2/L3 Multipathing end to end
- Consistent network policies across datacenter
- Lower TCO

FCoE Architecture, day one

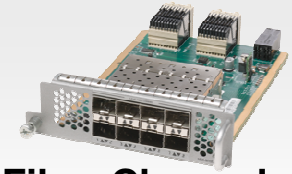


FCoE Architecture, the big picture

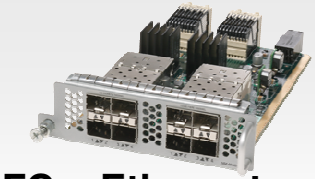


Cisco Nexus 5000 Series

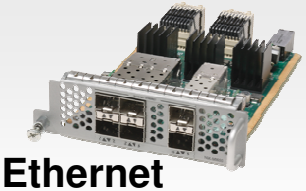
56-Port L2 Switch
- 40 Ports 10GE/FCoE/DCE, fixed
- 2 Expansion module slots



Fibre Channel
- 8 Ports 1/2/4G FC



FC + Ethernet
- 4 Ports 10GE/FCoE/DCE
- 4 Ports 1/2/4G FC



Ethernet
- 6 Ports 10GE/FCoE/DCE

NX-OS

DC-NM and Fabric Manager

Nexus 5020 Switch Fabric Attributes

Optimized for I/O Virtualization

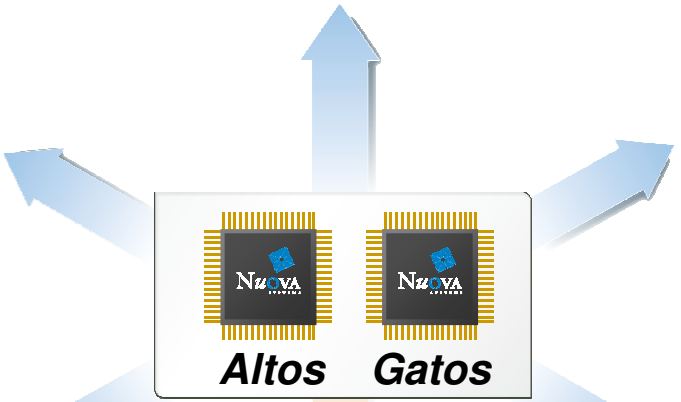
Multiprotocol
VNTag

Scale

52 10GE ports
Lossless, no packet drop fabric

Performance

1.2 Tbps Bandwidth
Low latency ~ 2.5us

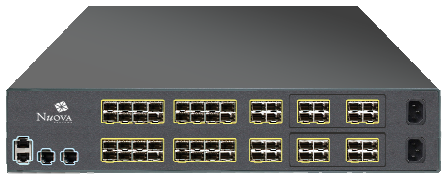


Services

L2+ switch service in H/W
Security
Per Priority Pause
Congestion Management
QoS

Compatibility

Cisco DCOS and FM
Cisco Trusted Security



Multiprotocol Top-of-Rack Switch

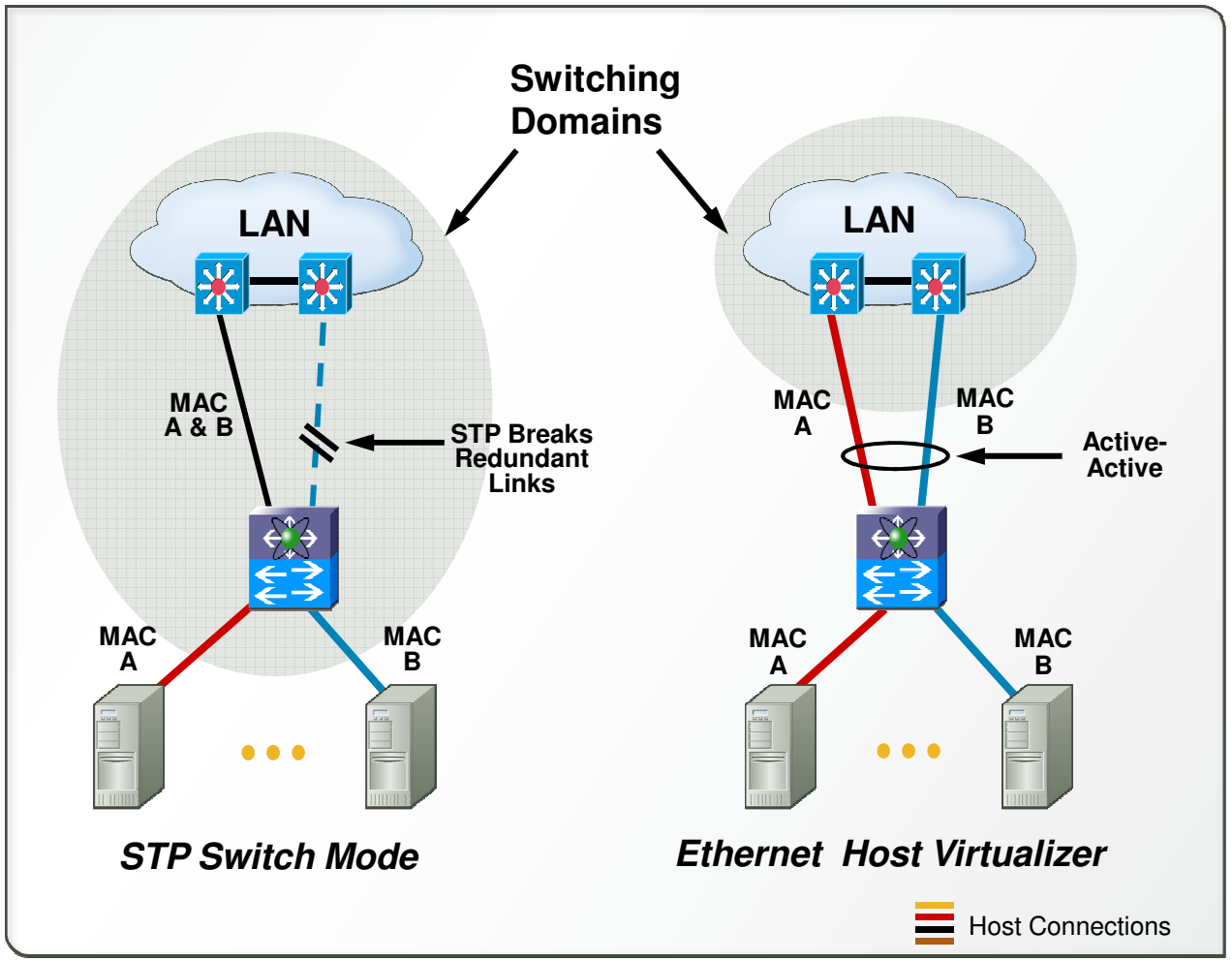
Nexus Data Center Server Switch

A switch optimized for server/data-center deployments

Feature	*Benefit*
Wirespeed 10Gbps	High I/O per server, low latency
Low Latency (~2.5us across switch)	Increase application transaction rate
Priority Flow Control (PFC)	Match native FC fabric characteristics (a loss-less fabric)
Consolidated I/O: Single Switch for LAN, SAN, IPC	Lower costs – fewer switches & adapters to buy Less complexity – fewer items to manage Power savings – fewer switches, fewer adapters
Ethernet Host Virtualizer (future)	No spanning tree to manage for uplinks Double the uplink bandwidth
N-Port Virtualizer	No Fibre Channel domain required on switch
Virtual NIC Tag (future)	Consistent network policies to virtual NICs

Ethernet Switch Mode & Ethernet Host Virtualizer

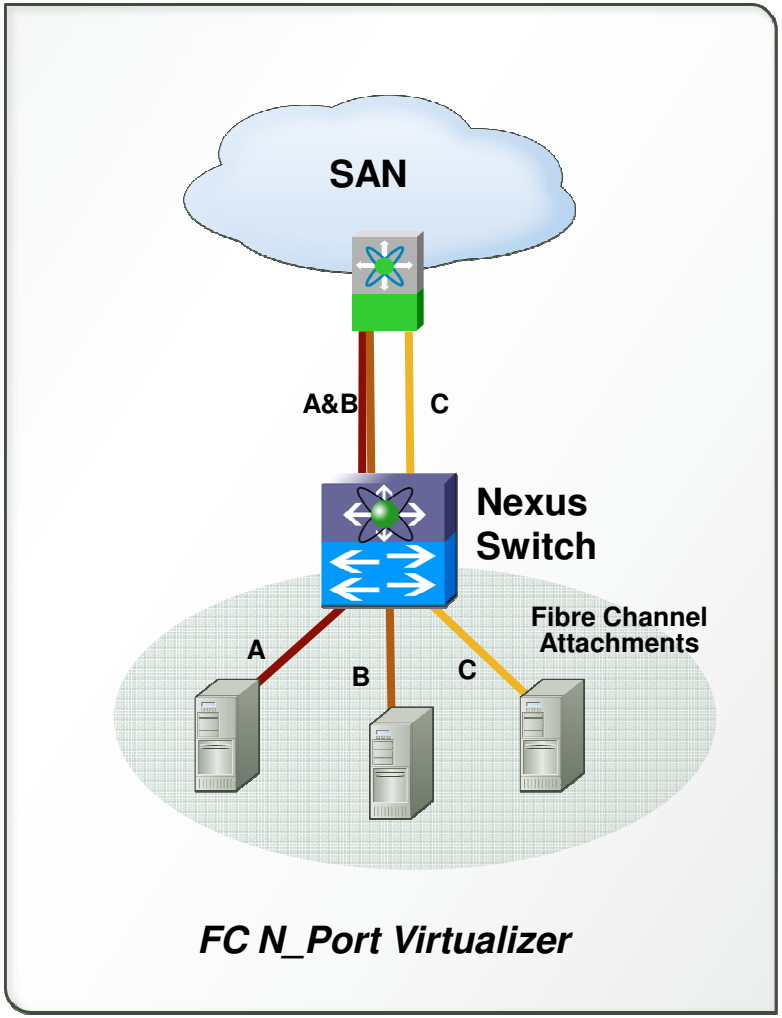
- Spanning Tree Breaks Redundant Links
- Ethernet Host Virtualizer
  - Pins Mac A to one uplink port
  - Pins Mac B to one uplink port
  - Doubles bandwidth



Both Modes of Operation Supported for Integration into Existing LAN Fabrics

Fibre Channel N_Port Virtualizer

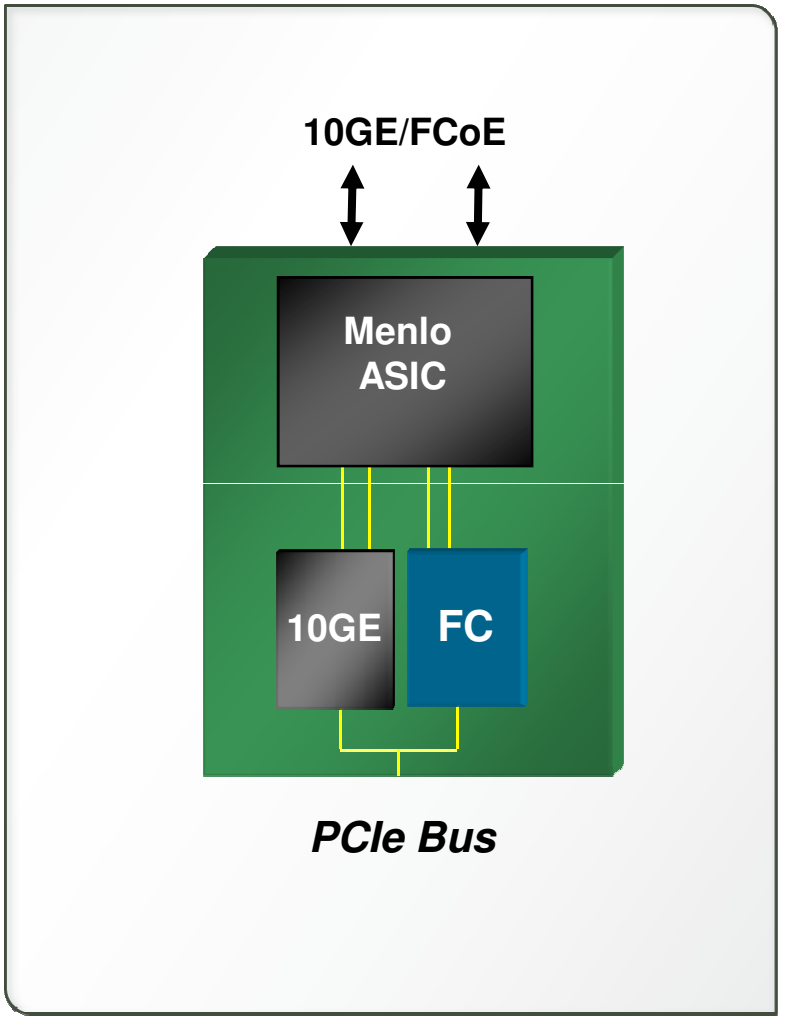
- Provides physical port level virtualization of multiple FC end nodes to one F_Port off a FC Switch
  - Nexus Switch operates in N_Port Proxy Mode (not in FC Switch mode)
  - Simplifies multi-vendor interoperation
- Eliminates the FC domain on Nexus switch
- Simplifies management
- Used in conjunction with NPIV



FC N_Port Virtualizer

Menlo: I/O Consolidation Adapter

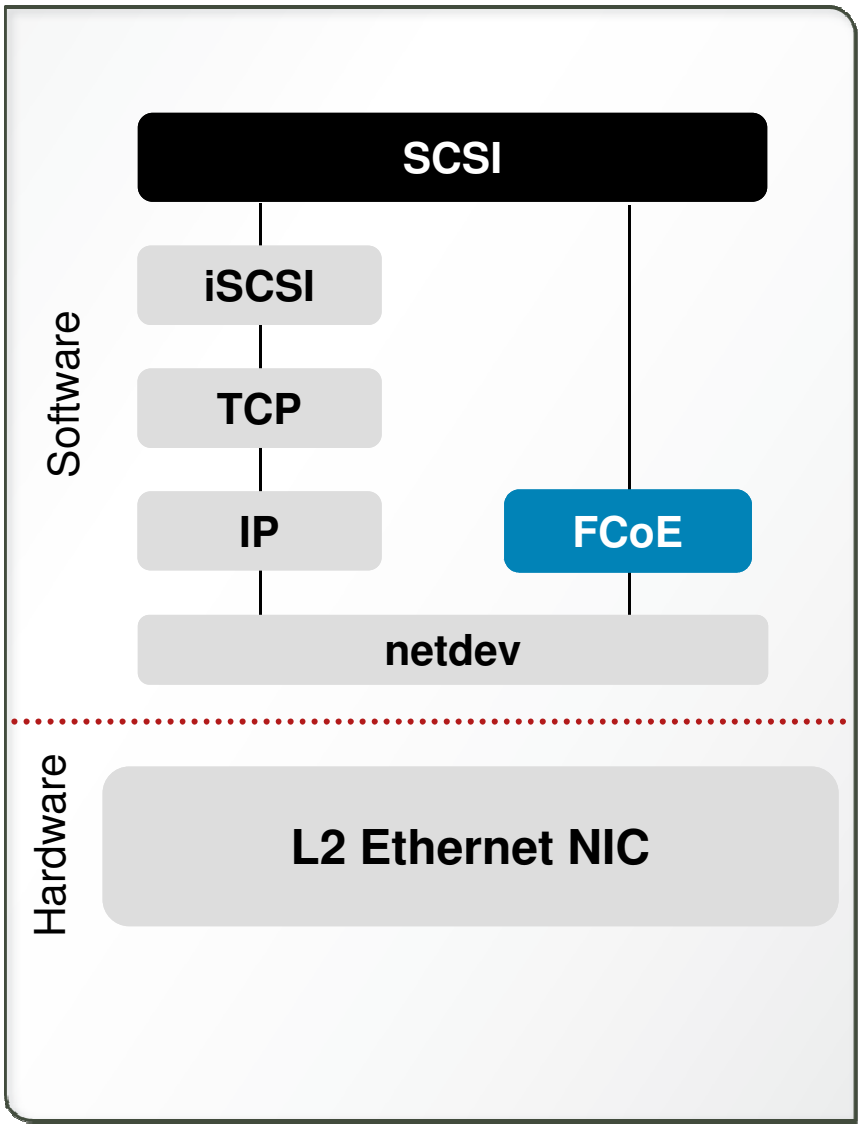
- Off the shelf NIC and HBA ASICs from: Qlogic, Emulex
  - Dual 10GE/FCoE ports
- Support for native drivers and utilities
  - Customer certified stacks
- Replaces multiple adapters per server
- Consolidates 10GE and FC on a single interface
- Minimum disruption in existing customer environments
- Available same time as Nexus Switch



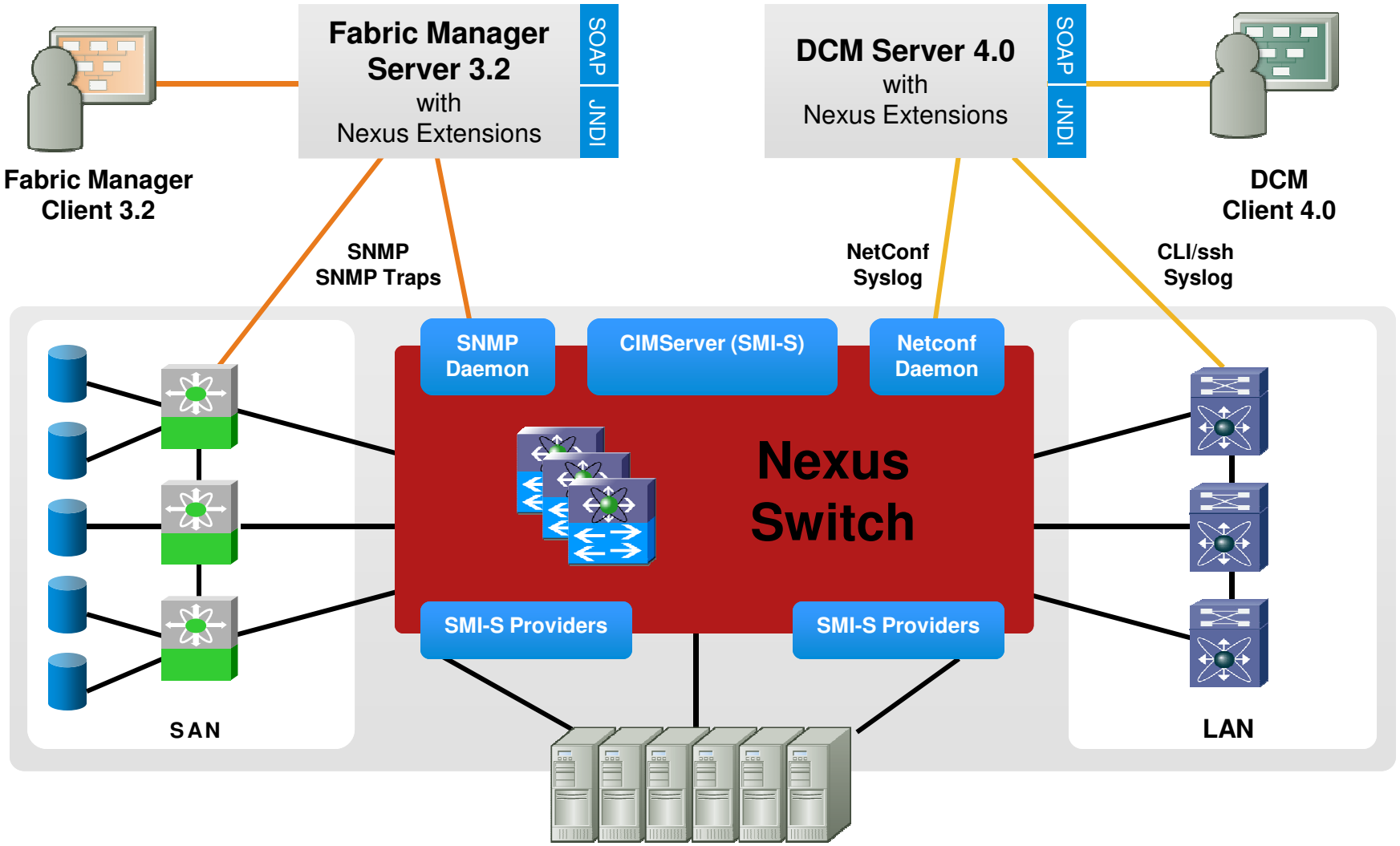
Menlo ASIC – Nexus designed multiplexer and FCoE offload protocol engine

FCoE SW Stack

- Supported on Intel Oplin 10GE adapters
  - SW upgraded turns 10GE adapter into FCoE adapter
- Software implementation
  - Initiator and Target mode
  - FCP, FC class 3
  - Fully supports Ethernet pause frames (per priority pause)
- Supported OS
  - Linux: Redhat & SLES
  - Windows
- “Free” access to the SAN
- Publicly available 2H’07

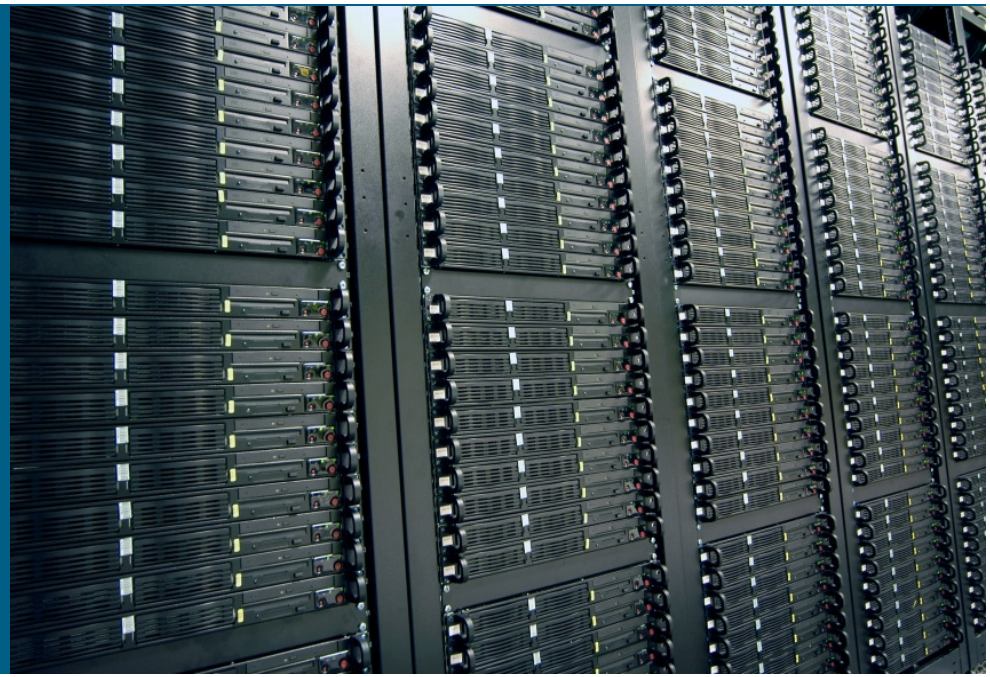


Fibre Channel and Ethernet Management



VMware virtual switch

Full-featured Cisco switch for virtual machines: Nexus 1000V



Introducing Cisco Virtual Network Link

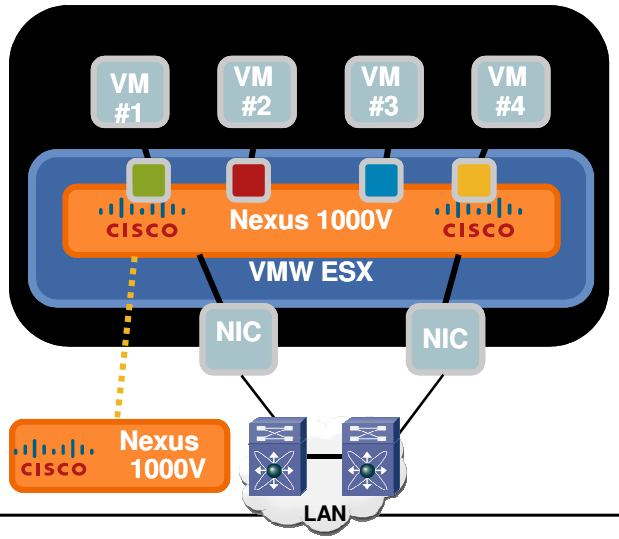
Virtualizing the Network Domain

Policy Based VM Connectivity

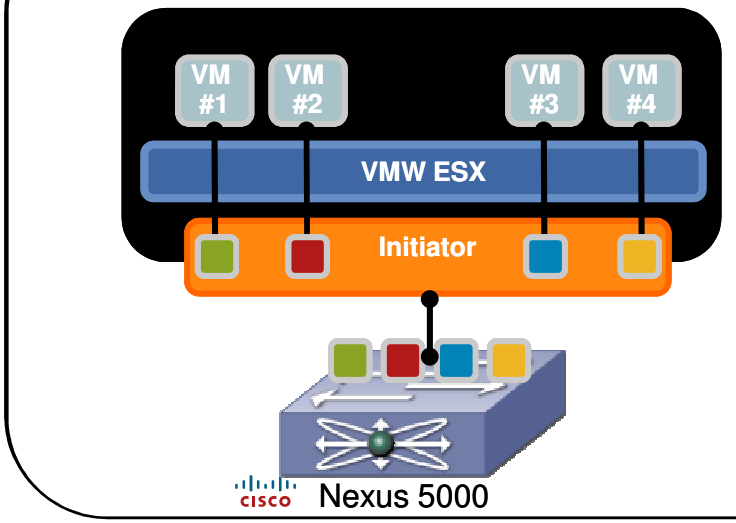
Mobility of Network & Security Properties

Non-Disruptive Operational Model

Cisco Nexus 1000V
(Software Based)



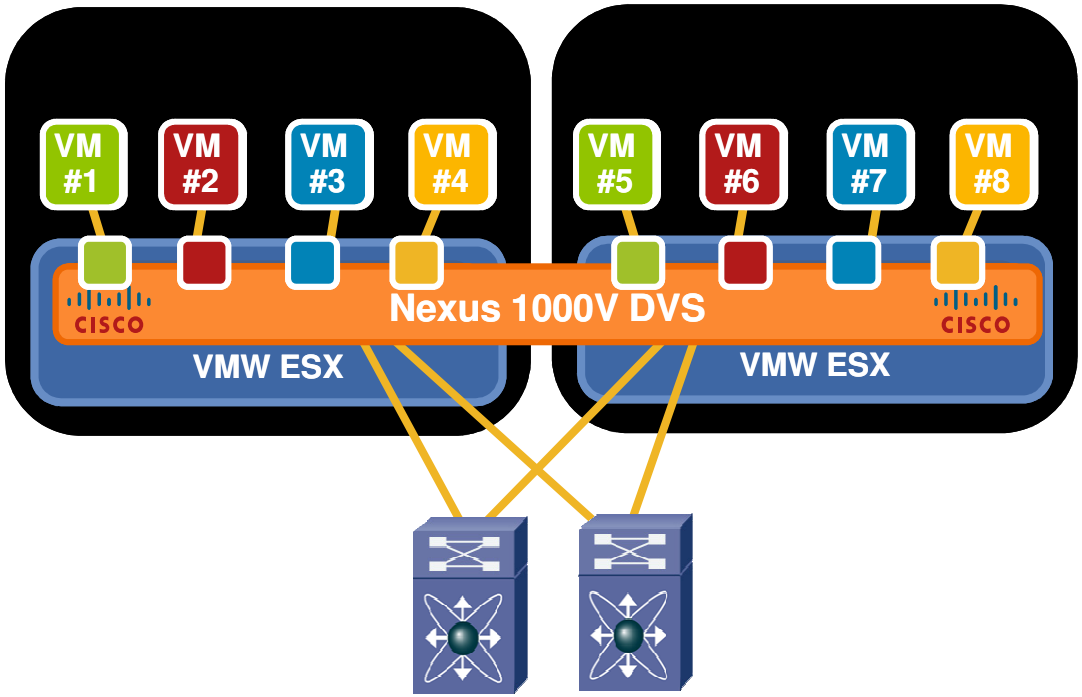
Nexus 5000 with VN-Link
(Hardware Based)



Two Complimentary Models to Address Evolving Customer Requirements

Cisco Nexus 1000V

Industry First 3rd Party Distributed Virtual Switch



- **Nexus 1000V provides enhanced VM switching for VMware ESX**
- **Features Cisco VN-Link:**
  - Policy Based VM Connectivity
  - Mobility of Network & Security Properties
  - Non-Disruptive Operational Model
- **Ensures proper visibility & connectivity during VMotion**

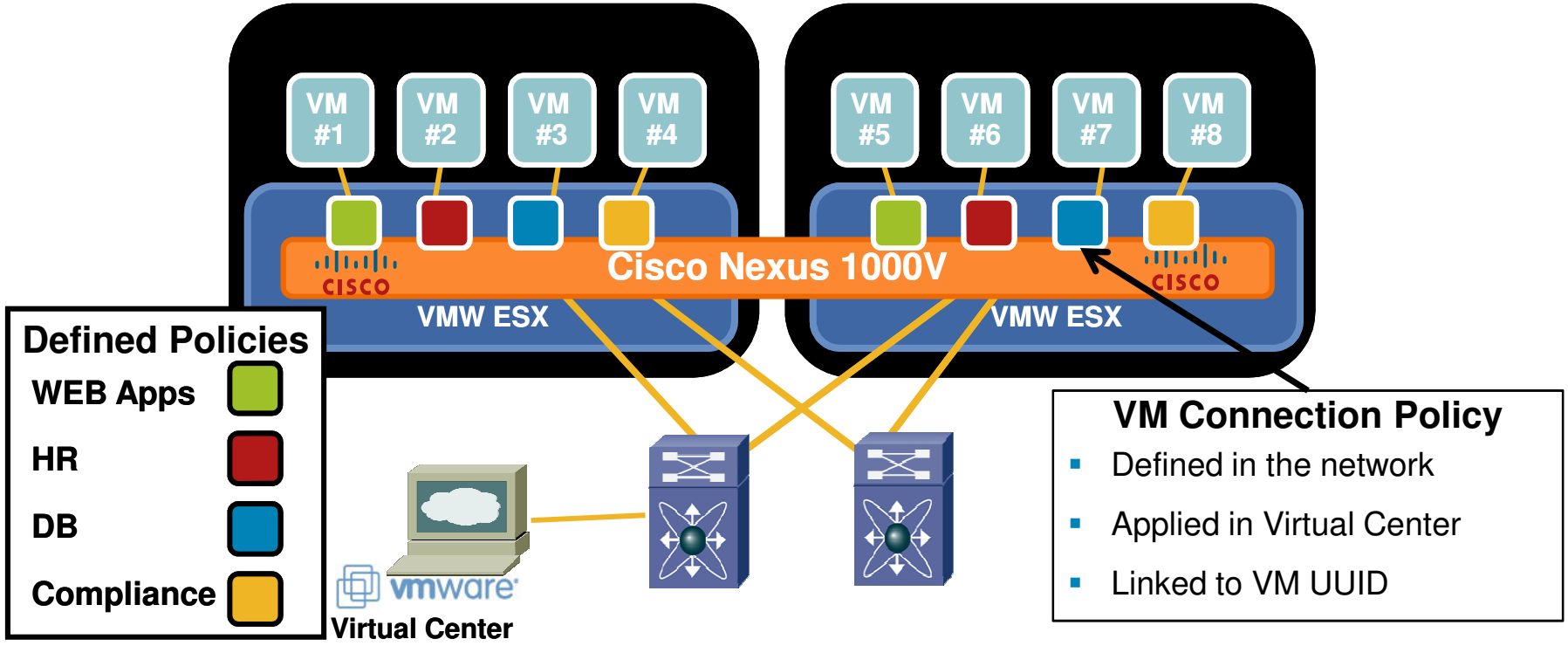
Enabling Acceleration of Server Virtualization Benefits

Cisco Nexus 1000V

Faster VM Deployment

VN-Link: Virtualizing the Network Domain

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

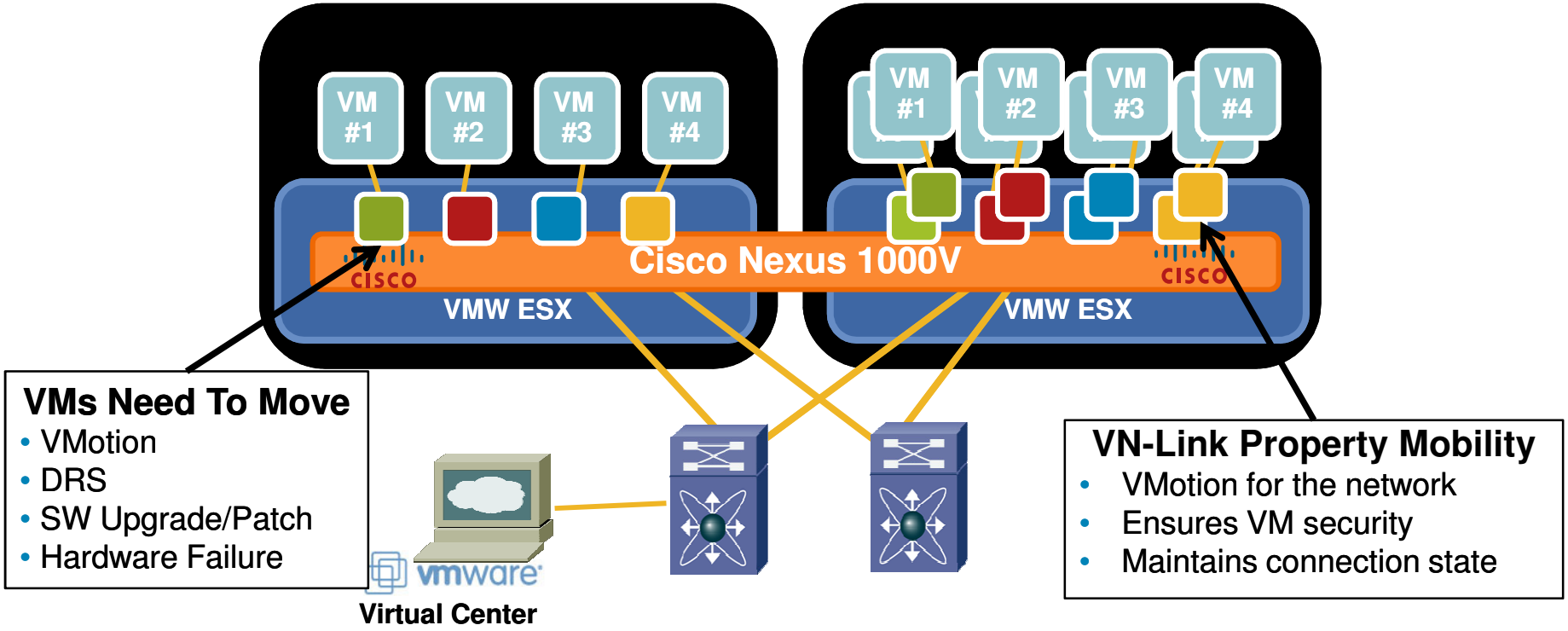


Cisco Nexus 1000V

Richer Network Services

VN-Link: Virtualizing the Network Domain

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

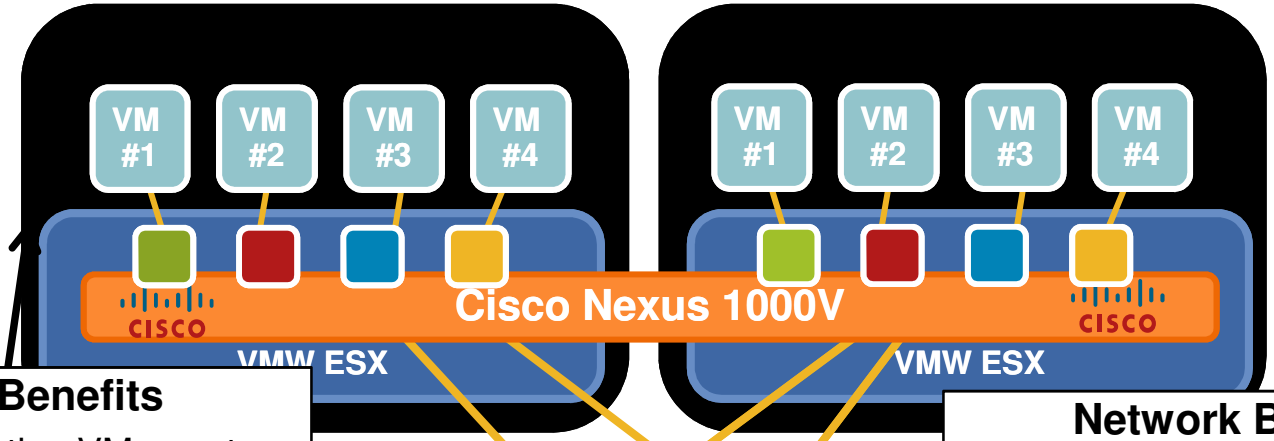


Cisco Nexus 1000V

Increase Operational Efficiency

VN-Link: Virtualizing the Network Domain

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



Server Benefits

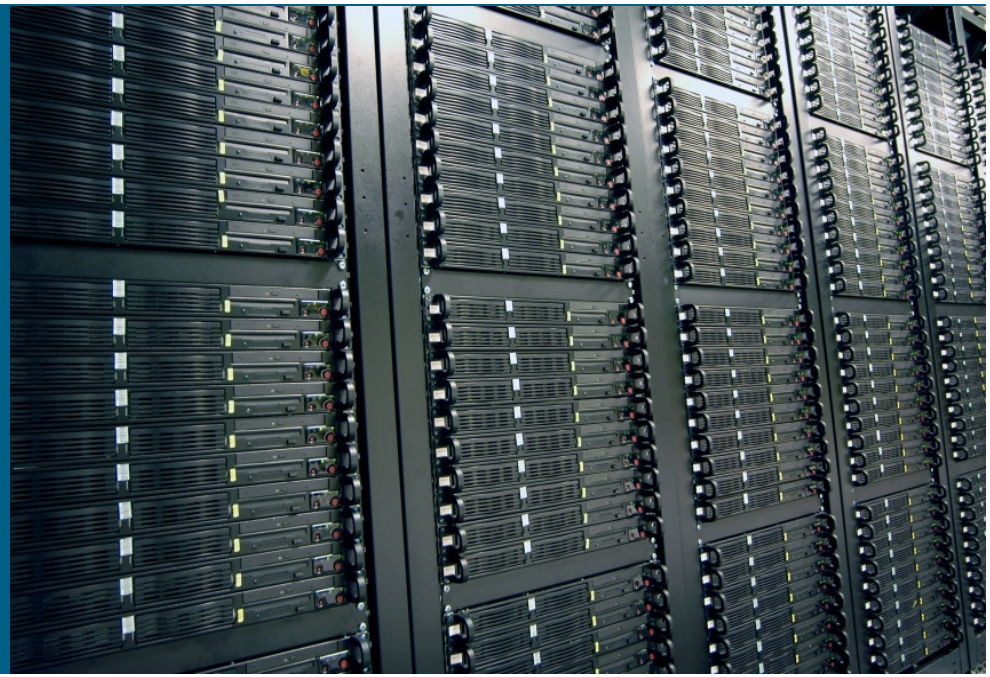
- Maintains existing VM mgmt
- Reduces deployment time
- Improves scalability
- Reduces operational workload
- Enables VM-level visibility

Network Benefits

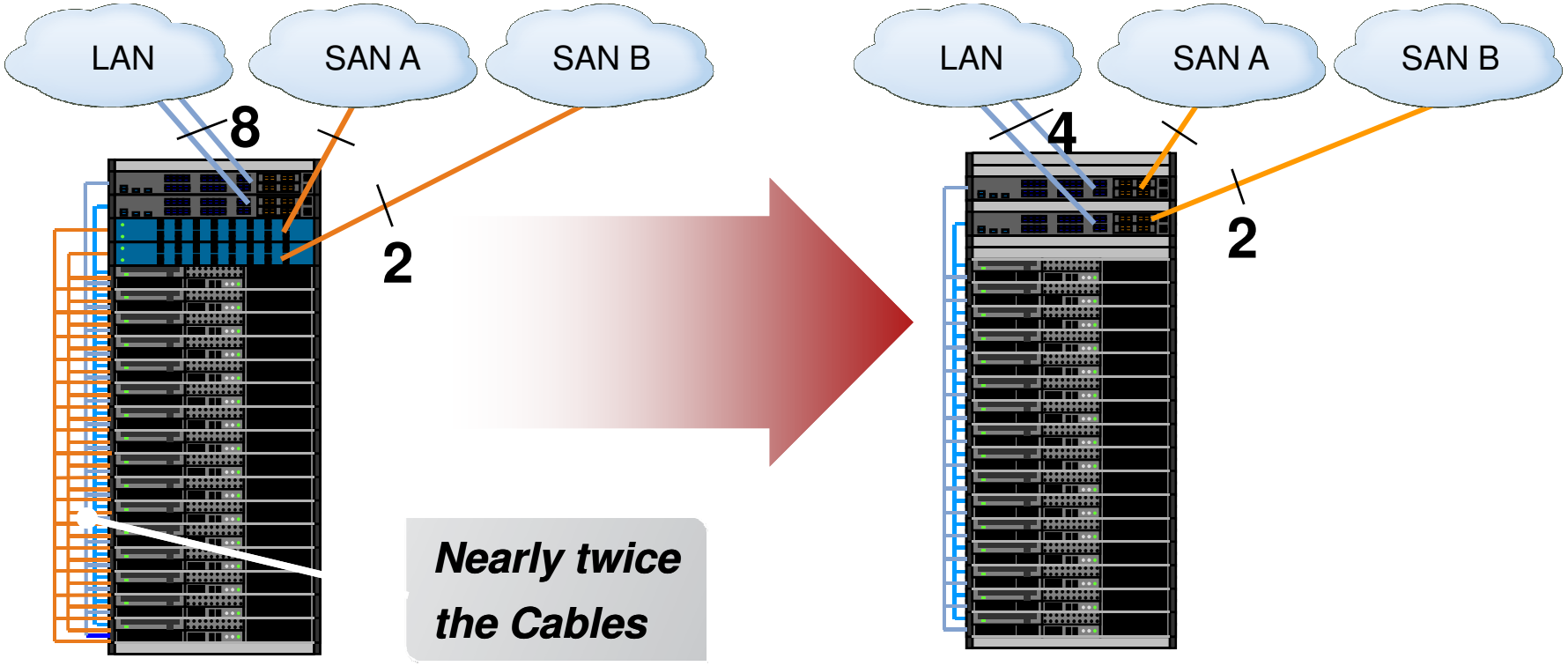
- Unifies network mgmt & ops
- Improves operational security
- Enhances VM network features
- Ensures policy persistence
- Enables VM-level visibility

Unified Fabric

Benefits of Unified Fabric and future of Data Center Architecture



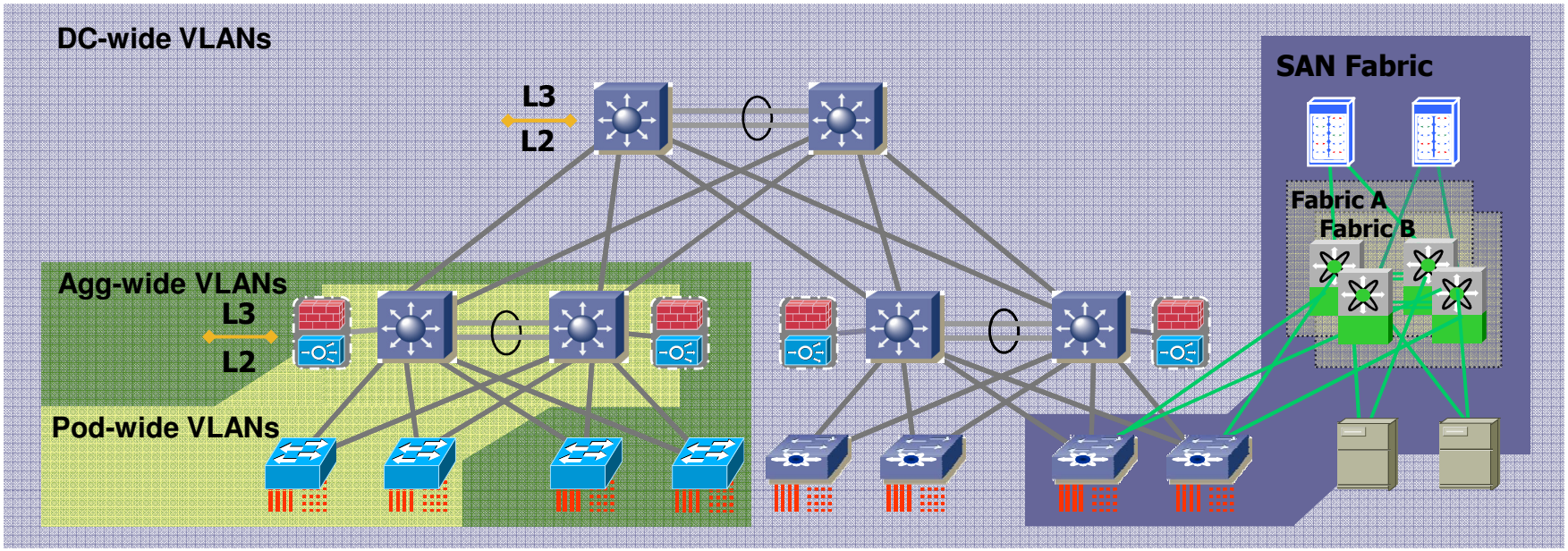
Use Case



16 Servers	Enet	FC	Total
Adapters	20	20	40
Switches	2	2	4
Cables	40	40	80
Mgmt Pts	2	2	4

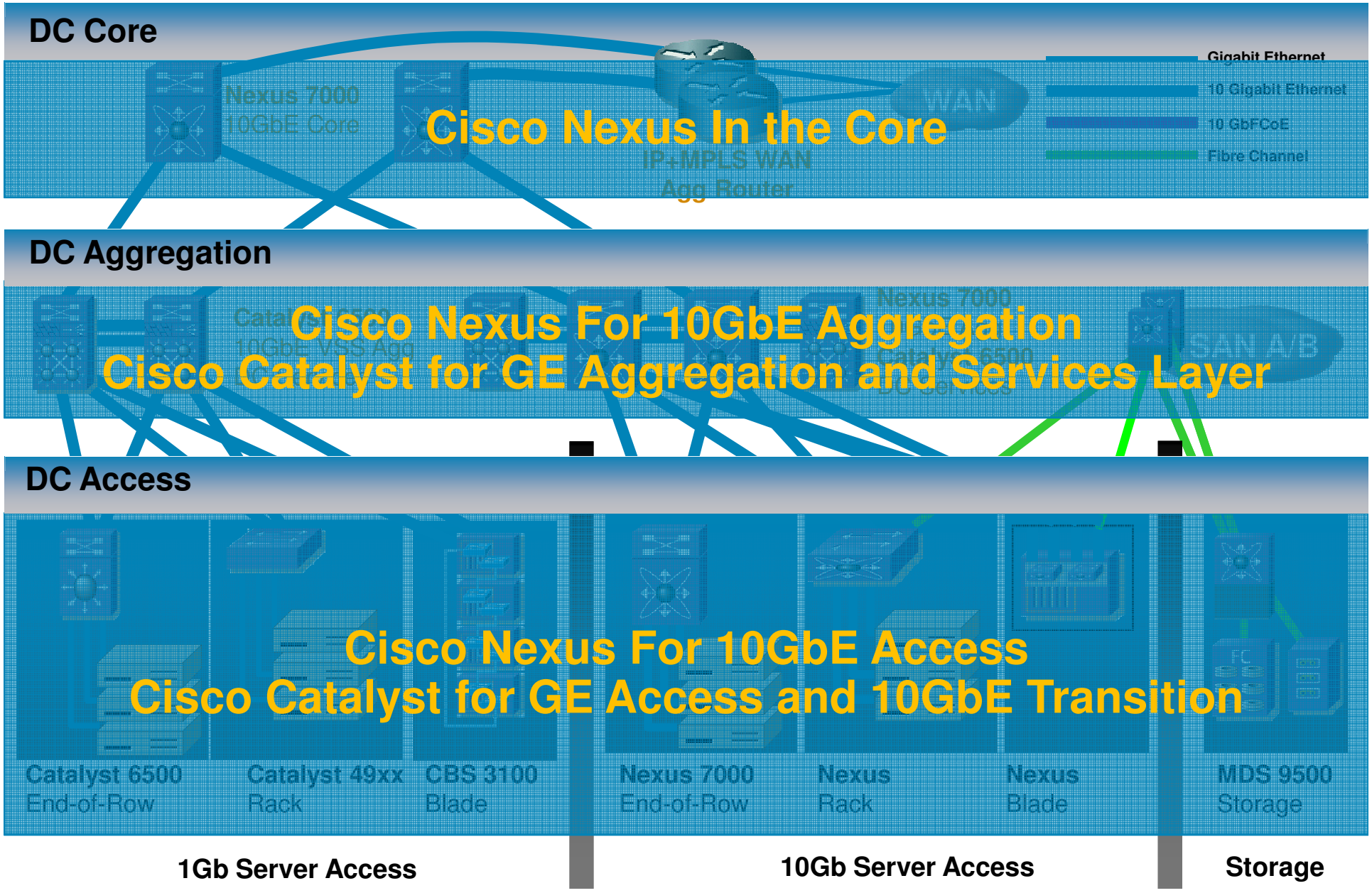
16 Servers	Enet	FC	Total
Adapters	20	0	20
Switches	2	0	2
Cables	40	0	40
Mgmt Pts	2	0	2

Future of Data Center Architecture



- **Topology Layers:**
  - Core Layer: Support high density L3 10GE aggregation
  - Aggregation Layer: Support high density L2/L3 10GE aggregation
  - Access Layer: Support EoR/MoR, ToR, & Blade for 1GE, 10GE, DCE & FCoE attached servers
- **Topology Service**:
  - Services through service switches attached at L2/L3 boundary
- **Topology Flexibility:**
  - Pod-wide VLANs, Aggregation-wide VLANs or DC-wide VLANs
  - Trade off between flexibility and fault domain

Summary


DC Core
Nexus 7000
10GbE Core
IP+MPLS WAN
Agg Router
WAN
Gigabit Ethernet
10 Gigabit Ethernet
10 GbFCoE
Fibre Channel
Cisco Nexus In the Core
DC Aggregation
Catalyst 6500
10GbE VSS Agg
Nexus 7000
Catalyst 6500
SAN A/B
Cisco Nexus For 10GbE Aggregation
Cisco Catalyst for GE Aggregation and Services Layer
DC Access
Catalyst 6500
End-of-Row
Catalyst 49xx
Rack
CBS 3100
Blade
Nexus 7000
End-of-Row
Nexus
Rack
Nexus
Blade
MDS 9500
Storage
Cisco Nexus For 10GbE Access
Cisco Catalyst for GE Access and 10GbE Transition
1Gb Server Access
10Gb Server Access
Storage

mbocian@cisco.com

Cisco Networkers Barcelona
26 – 29.01.2009

Cisco Networkers
2009
January 26-29 Barcelona, Spain

http://www.cisco.com/web/europe/cisco-networkers/2009/index.html