Cisco *live!*

What You Make Possible

TOMORROW
starts here.

CISCO

Deploying Nexus 7000 in Data Centre Networks

BRKDCT -2951

Session Abstract

This session is targeted to Network Engineers, Network Architects and IT administrators who have deployed or are considering the deployment of the Nexus 7000.

The session begins with a Nexus 7000 hardware overview and primarily focuses on Data Centre related features and implementation best practices.

The session covers recent hardware enhancements to Nexus 7000 product line such as the new Nexus 7004 chassis, the new supervisors modules (SUP2/2E) and the new high-performance 10/40/100G linecards (M2 and F2e).

The session also discusses some of the Data Centre design examples and its best practices section will cover features such as VDC, VPC, Cisco FabricPath, Layer2, Layer3, Fabric Extenders(FEX), etc.

Attendees should have a basic knowledge of the Nexus 7000 hardware platform and software features as well as good understanding of L2 and L3 protocols.
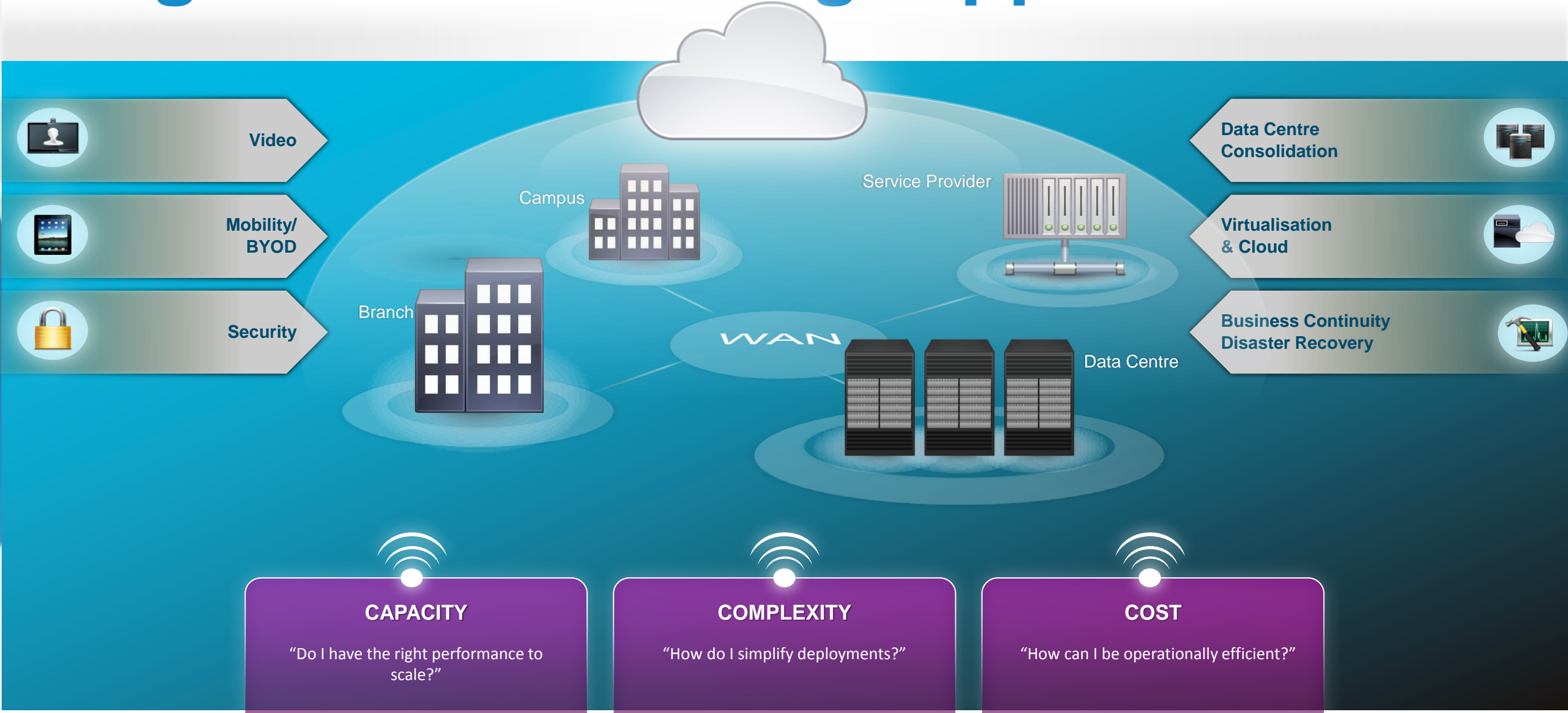
 Cisco Public

Agenda

- Evolution of Data Centre Trends & Observations

- Changes to Data Centre Fabric

- Nexus 7000 Switching Hardware Overview

- Features Overview & Best Practices

- Data Centre Design Examples

Evolution of Data Centre Trends & Observations

The IT Landscape
Mega Trends Causing Ripples

Video

Mobility/ BYOD

Security

Data Centre Consolidation

Virtualisation & Cloud

Business Continuity Disaster Recovery

Campus

Service Provider

Branch

WAN

Data Centre

CAPACITY

"Do I have the right performance to scale?"

COMPLEXITY

"How do I simplify deployments?"

COST

"How can I be operationally efficient?"

The IT Challenge: "Is My Network Ready?"

Data Centre Architecture

One design used to fit most

- Switching Design was based on the hierarchical switching - three tiers: Access, Aggregation and Core

- L2/L3 boundary at the aggregation

- Services in aggregation

- What has changed? Most everything

  - Hypervisors

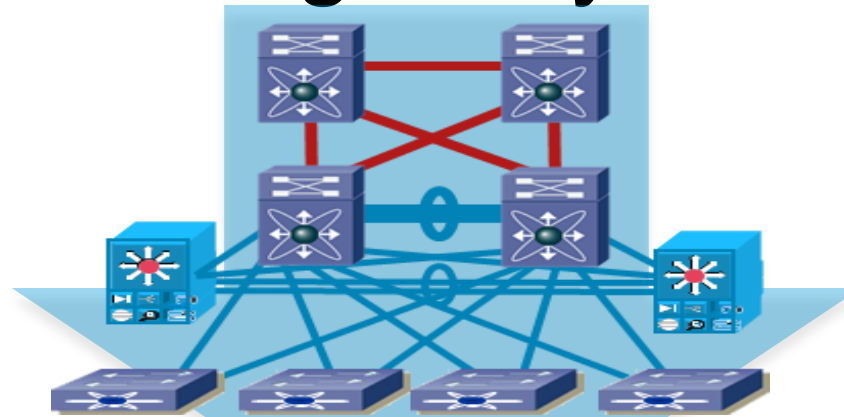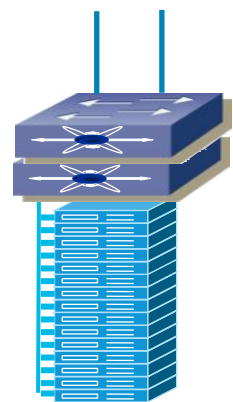 - Cloud – IaaS, PaaS, SaaS

  - MSDC

  - Ultra Low Latency

Core

Layer 3

Layer 2

Aggregation

Services

Access
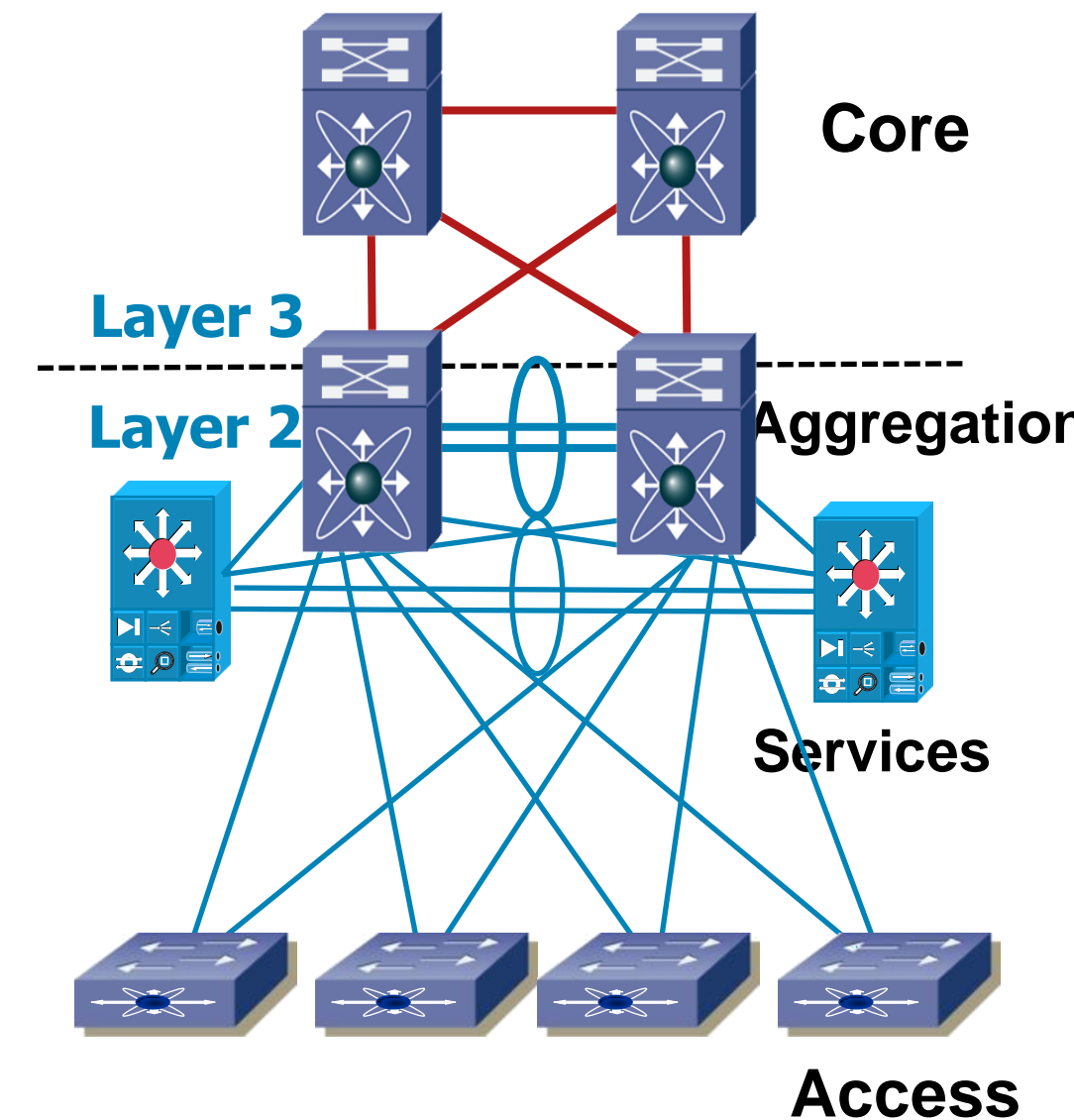
Changes to Data Centre Fabric

Data Centre Architecture
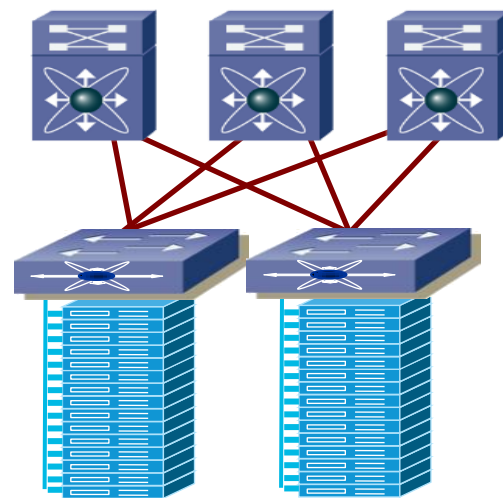
There is no 'single design' anymore

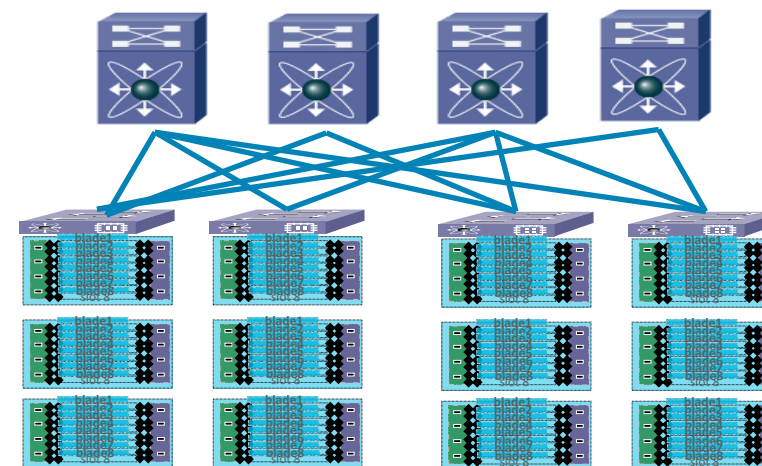### Spectrum of Design Evolution

Ultra Low Latency

- **High Frequency Trading**
- **Layer 3 & Multicast**
- **No Virtualisation**
- **Limited Physical Scale**
- **Nexus 3000 & UCS**
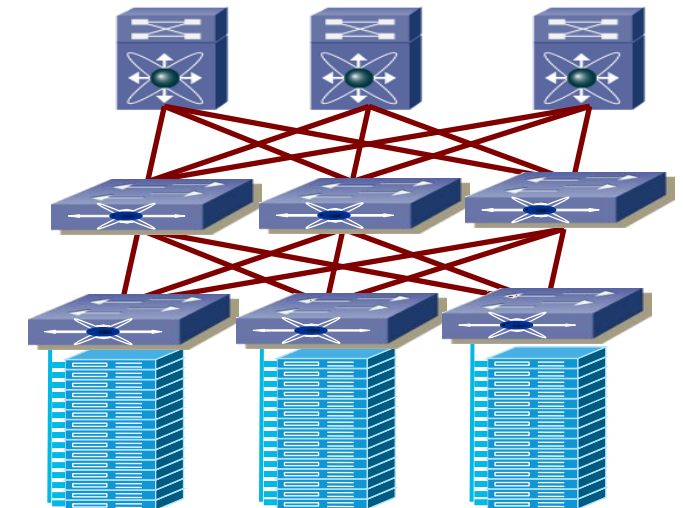- **10G edge moving to 40G**

HPC/GRID

- **Layer 3 & Layer 2**
- **No Virtualisation**
- **Nexus 2000, 3000, 5500, 7000 & UCS**
- **10G moving to 40G**

Virtualised Data Centre

- **SP and Enterprise**
- **Hypervisor Virtualisation**
- **Shared infrastructure Heterogenous**
- **1G Edge moving to 10G**
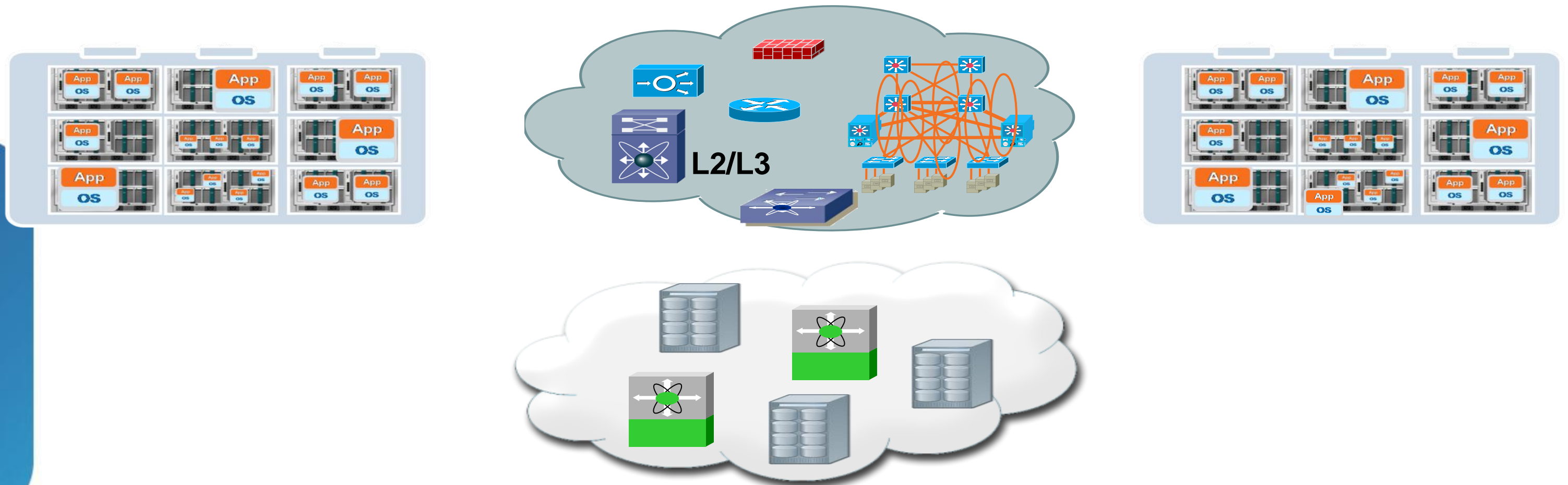- **Nexus 1000v, 2000, 5500, 7000 & UCS**

MSDC

- **Layer 3 Edge (iBGP, ISIS)**
- **1000's of racks**
- **Homogeneous Environment**
- **No Hypervisor virtualisation**
- **1G edge moving to 10G**
- **Nexus 2000, 3000, 5500, 7000 & UCS**

Building Efficient DC Fabric to Scale

Tightly Coupled Workload—Active/Active

L2/L3

- Active workload migration (e.g. vMotion) currently constrained by the latency requirements associated with storage synchronisation

- Tightly coupled workload domain has specific network, storage, virtualisation and services requirements

Building an Efficient DC Fabric to Scale

Loosely Coupled Workload—Burst and Disaster Recovery


WAN DCI and Routing

Asynchronous
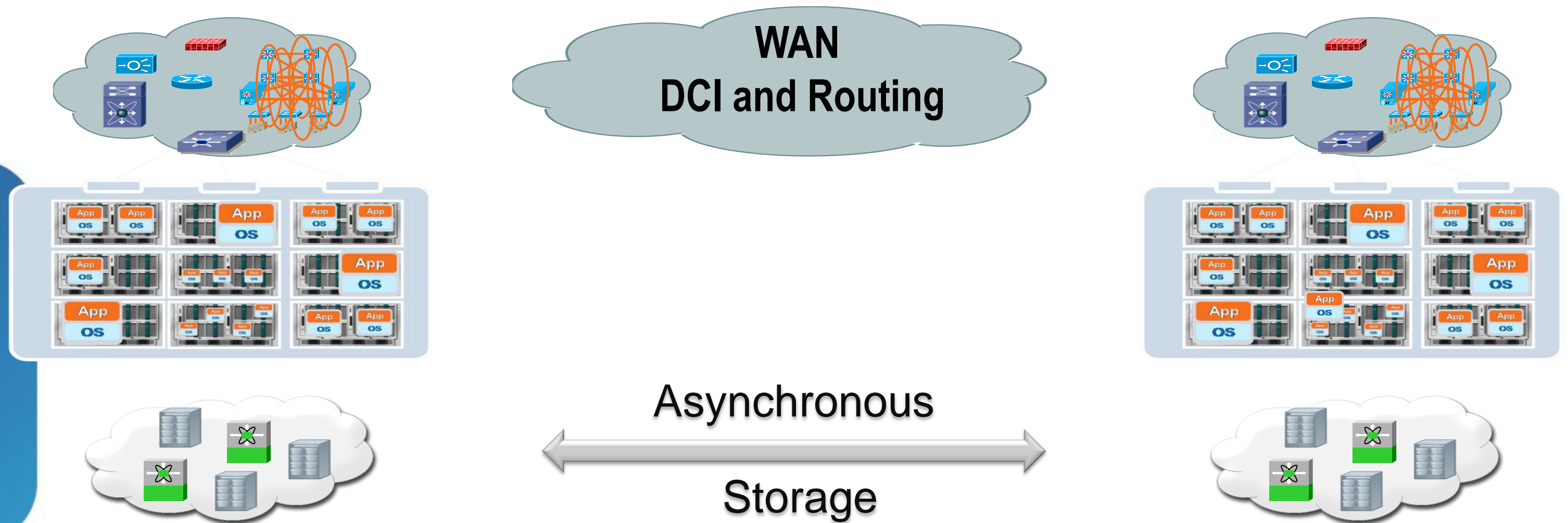
Storage

- Burst workload (adding temporary processing capacity) and Disaster Recovery leverage out of region facilities

- Loosely coupled workload domain has a different set of network, storage, virtualisation and services requirements

Building an Efficient DC Fabric to Scale

'Scaling Up' the Network Pod and Scaling 'out' the Fabric

- Scaling 'Up' of the building blocks (High Density 10G, 40 G, 100G, FEX, Adapter-FEX, vPC)

- Scaling 'Out' of the Fabric (FabricPath, OTV, SPF/EIGRP/ISIS/BGP, MPLS)

Scaling 'out' the Fabric

Scaling 'up' the aggregation block (POD)
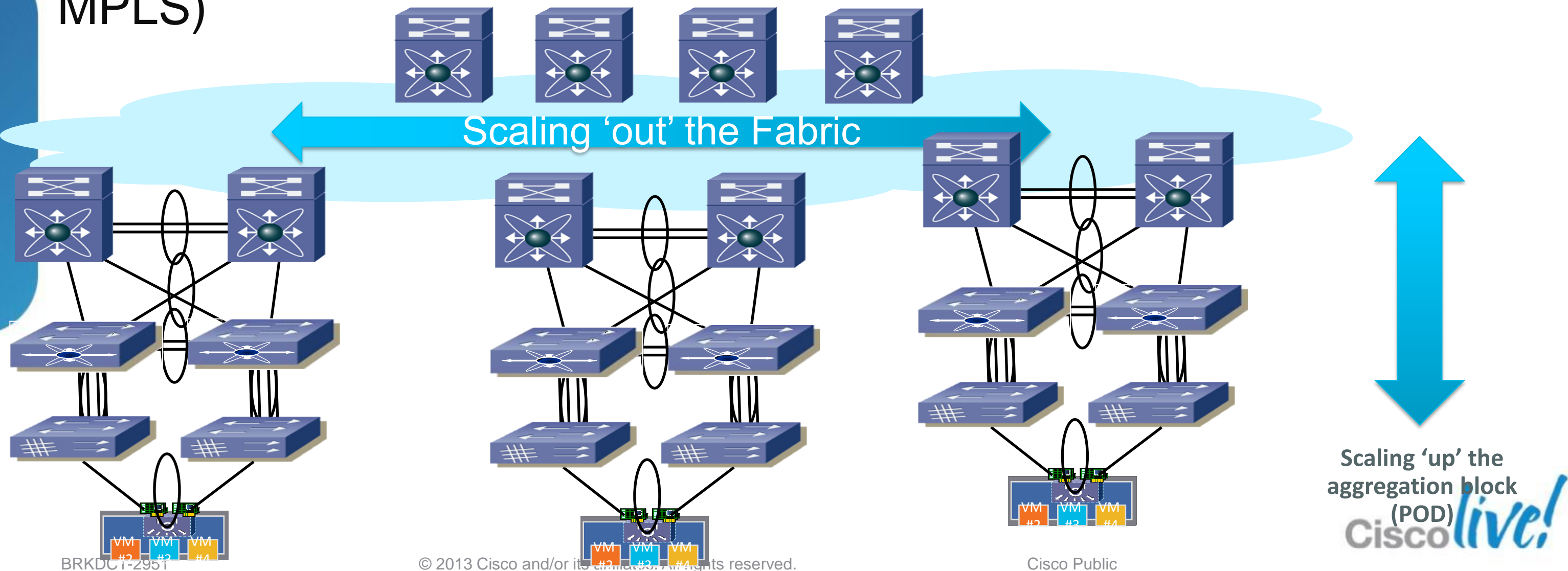
Massively Scalable Data Centres
Requires Large Fabrics

- 'N' way load sharing designs

- Topologies "flattening" and spreading wider

- Spine/Leaf (Core/Edge) design to address
  - Oversubscription & Buffering (non-blocking in the extreme case)
  - Workload Traffic Pattern Changes (East-West)
  - Availability (moving to N + 1 models)

- Nexus HW provides a solid toolset for these designs
  - Scaling Port channel Bandwidth: 8 links → 16 links, Virtual Port Channels
  - FabricPath – Nexus 7K, Nexus 5K
  - L3 ECMP – Nexus 7K,5K,3K
  - Overlay Protocols
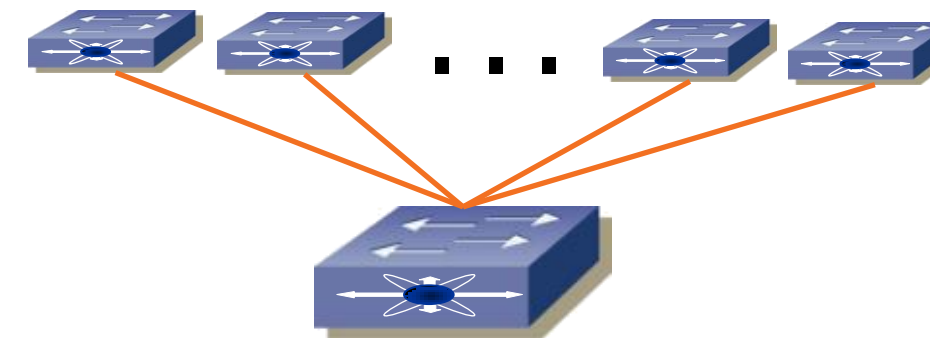
Nexus 3000 – 32 way ECMP
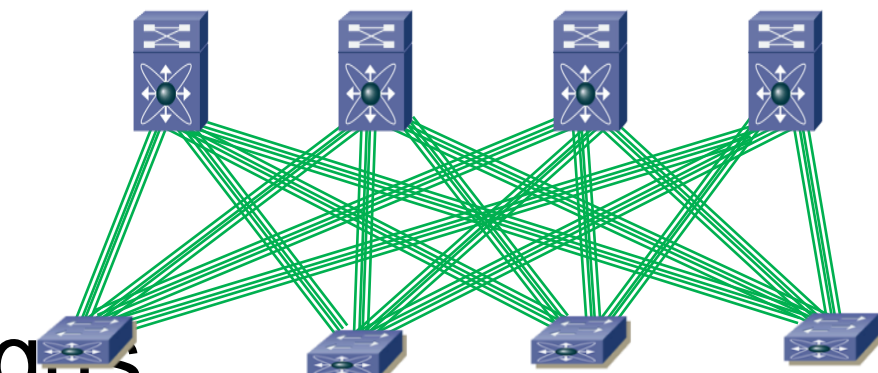
FabricPath 16 way ECMP & 16 way port channels x 10G links
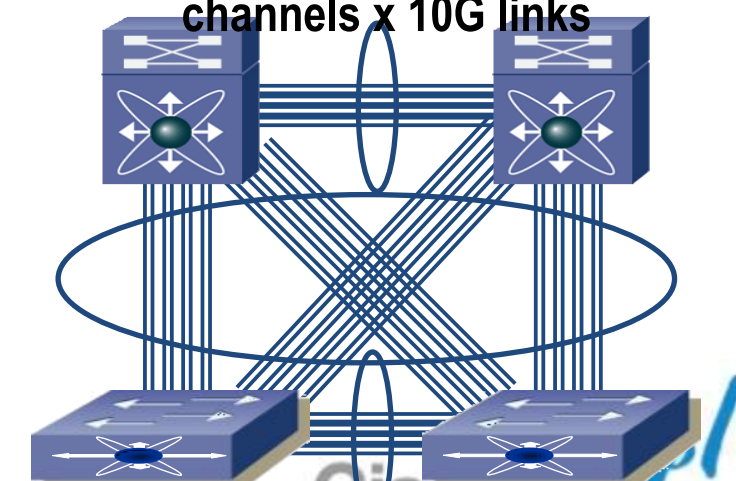
vPC Port Channel 32 x 10G links

Cisco Fabric Extender Architecture

Scaling the Access Layer Fabric

FEX   FEX   FEX

Top of Rack (ToR)   End of Row (EoR)

Nexus 5000 or
Nexus 7000
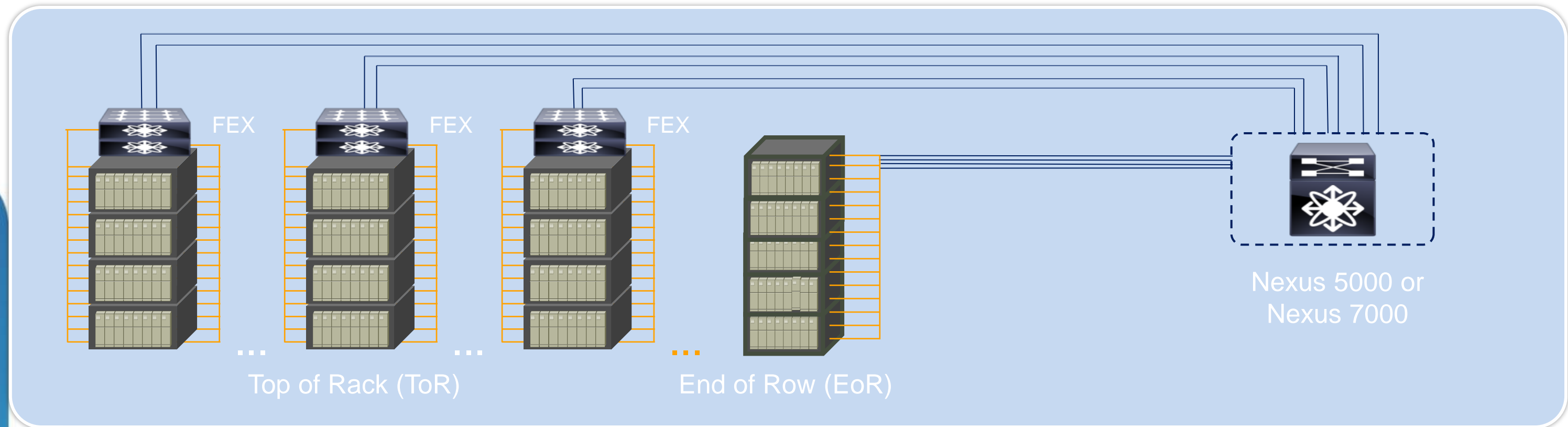
Benefits

- De-coupling and optimisation of Layer 1 and Layer 2 Topologies

- Simplified Top of Row cabling with End of Row Management paradigm

- Support for Rack and blade server connectivity

- Reducing number of management points in a ToR Model → Fewer devices to manage, monitor, upgrade

Virtual Port Channel (vPC)
Scalable L2 Network, Maximise Uplink Bandwidth and Resiliency

Traditional STP Based Network

Nexus VPC Based Network

L3

L2

L3

L2

Block

Only 50% Links Forwarding
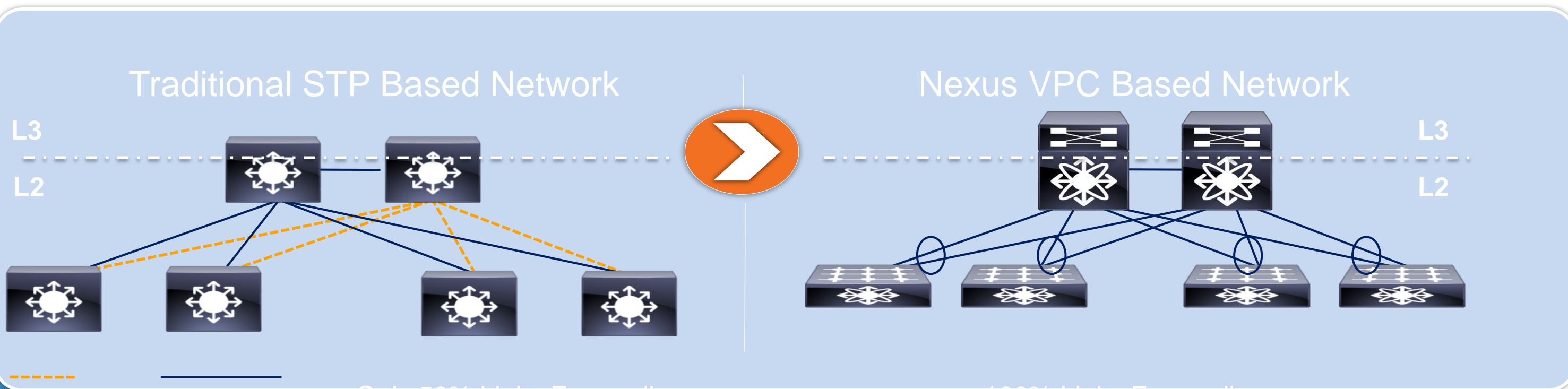
100% Links Forwarding

Features

- **Overcomes spanning tree limitations** in access and distribution layers

- **Both uplinks are active** on all VLANs

- Extends link aggregation to **two separate physical switches**

Benefits

- Double the Bandwidth— all links forwarding

- Fast convergence around failed link

- Increased data centre resiliency

- Maximise investment in infrastructure

Cisco live!

Cisco FabricPath

Extend VLANs Within the Data Centre

Nexus FabricPath Network

POD 1
VLAN1

POD 2
VLAN1

POD 3
VLAN1

Span VLANs Within the Data Centre
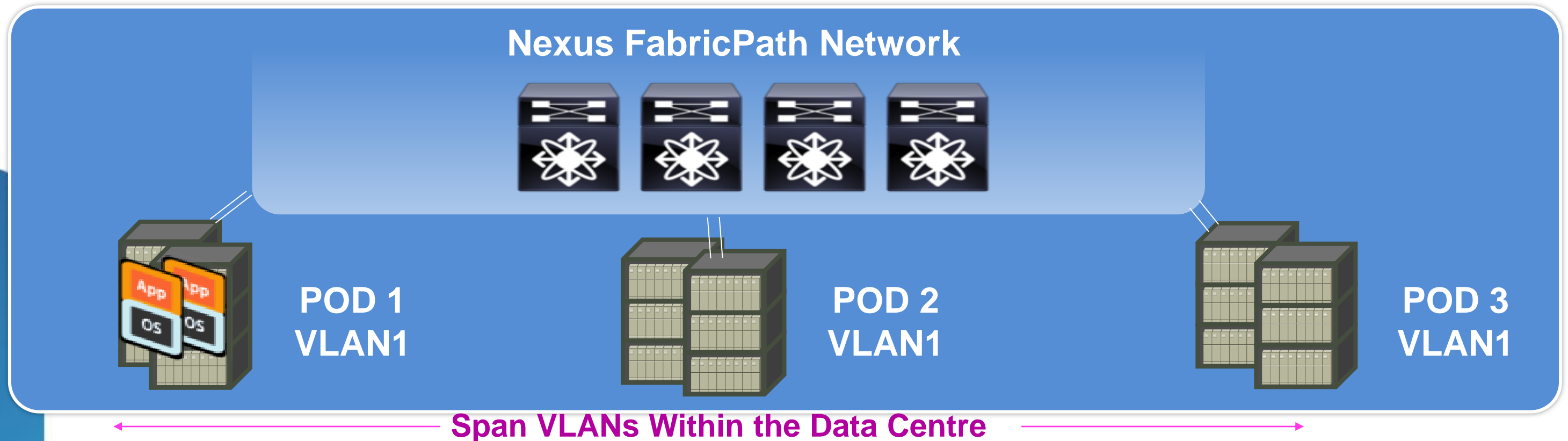
Features

- Scalable up to 12,000+ 10GE servers in single domain

- High cross-sectional bandwidth

- Extend VLANs across data centre

Benefits

- Seamless workload mobility beyond racks/pods

- Leverage compute resources across data centre for any workload

- Simplify scale out by adding compute resources for any app, anywhere in the data centre

Overlay Transport Virtualisation (OTV)

Extend VLANs Across Data Centres

IP/MPLS (Internet/Private)

OTV

DC 1
VLAN1

DC 2
VLAN1

DC 3
VLAN1

Extend VLANs Across Data Centres
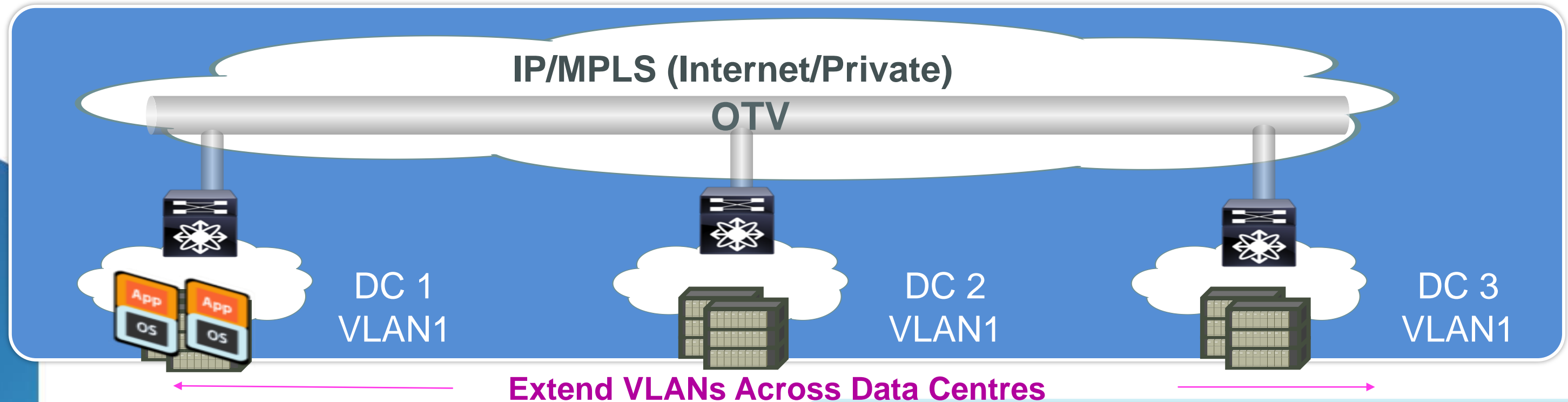
Features

- Ethernet LAN Extension over any network

- Multi datacentre scalability

- Seamless overlay—no network re-design

Benefits

- Many physical sites—one logical data centre

- Seamless workload mobility between data centres

- Leverage and optimise compute resources across data centres for any workload

- Enables disaster avoidance and simplifies recovery

What is the Nexus 7000 ?

Data-Centre class Ethernet switch designed to deliver high-availability, system scale, usability, investment protection

Supervisor Engine

I/O Modules

Chassis

Fabrics

Cisco live!
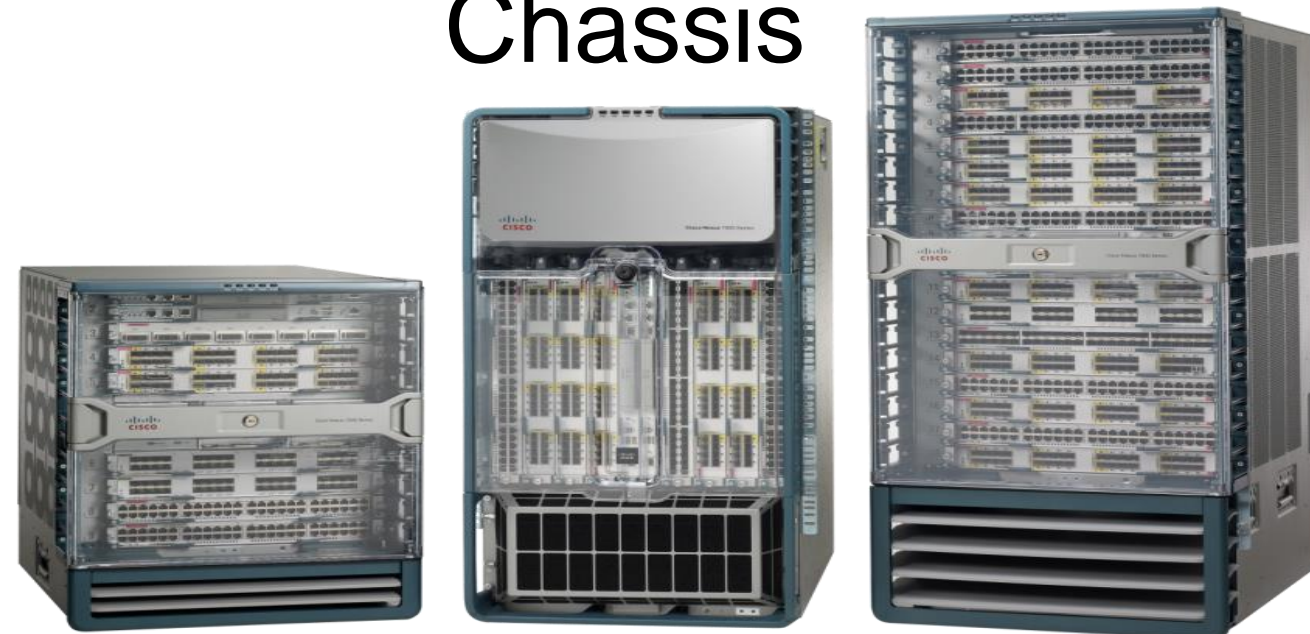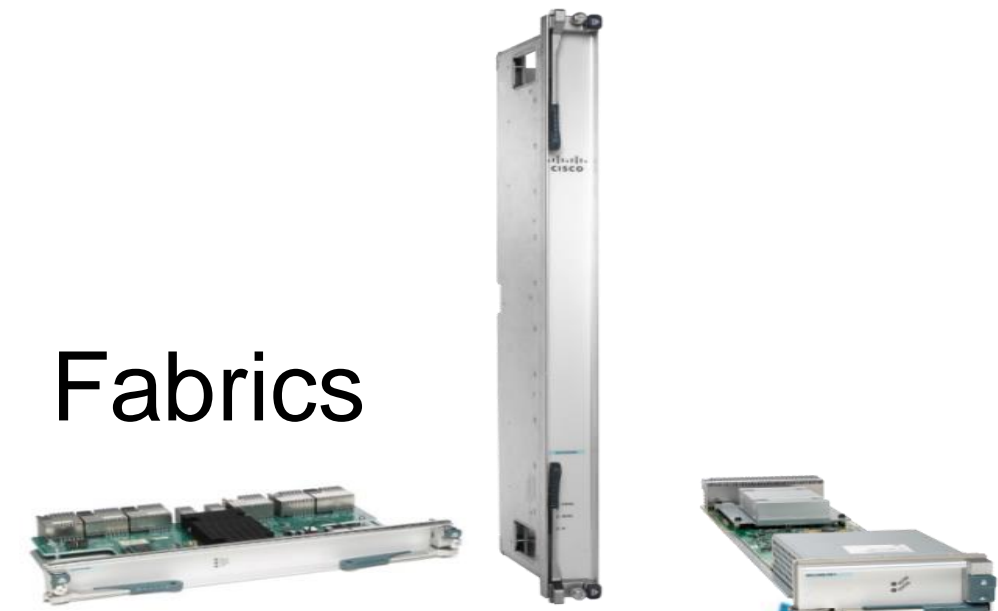
Cisco Nexus 7000 Series Chassis

Multiple Chassis Form factor

NEW

Highest 10GE Density
in Modular Switching

	Nexus 7004	Nexus 7009	Nexus 7010	Nexus 7018
Height	7 RU	14 RU	21 RU	25 RU
Max BW per Slot	440 Gig/Slot	550 Gig/Slot	550 Gig/Slot	550 Gig/slot
Max 10/40/100GE ports	96/12/4	336/42/14	384/48/16	768/96/32
Air Flow	Side-to-Rear	Side-to-Side	Front-to-Back	Side-to-Side
Power Supply Configurations	4 x 3KW AC	2 x 6KW AC/DC 2 x 7.5KW AC	3 x 6KW AC/DC 3 x 7.5KW AC	4 x 6KW AC/DC 4 x 7.5KW AC
Application	Small to Medium Core/Edge	Data Centre and Campus Core	Data Centre	Large Scale Data Centre

Cisco Public

Cisco live!

Nexus 7004

- 2 Supervisors + 2 Modules
- No Fabric Modules Required
- Up to 4 3kW Power Supply AC/DC
- Air Flow: Side to Rear
- Use cases: DC Edge, Small core/agg
- Supports FabricPath, OTV, LISP etc

Supported Modules	
M1-XL	✔
M2-XL	✔
F2/F2e	✔
SUP2/SUP2E	✔
Sup1	✖
F1	✖
M1-NonXL	✖

Side-to-back airflow

Fan tray

Supervisor slots (1-2)

I/O module slots (3-4)

Power supplies

7RU

Front

Air exhaust

6.1 maintenance

Rear

Cisco Public
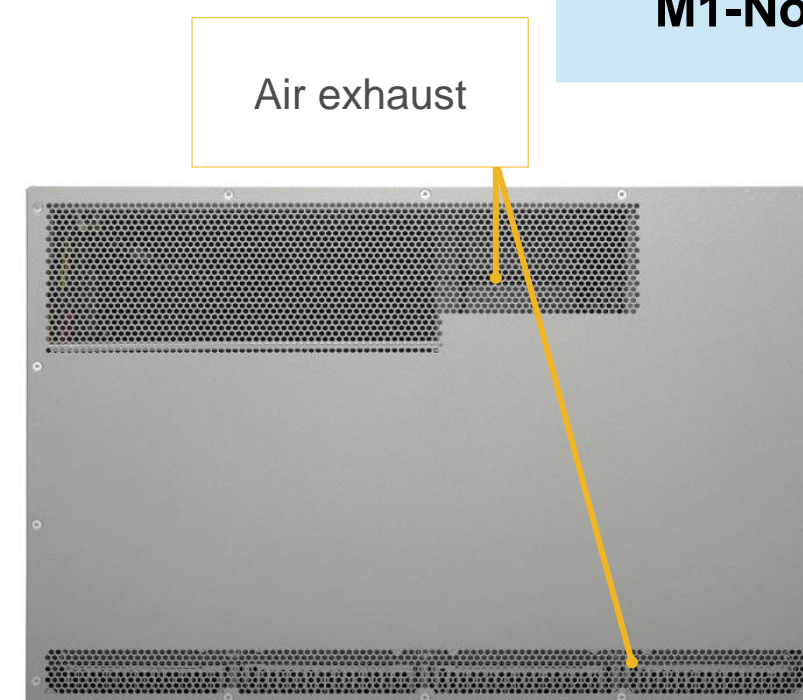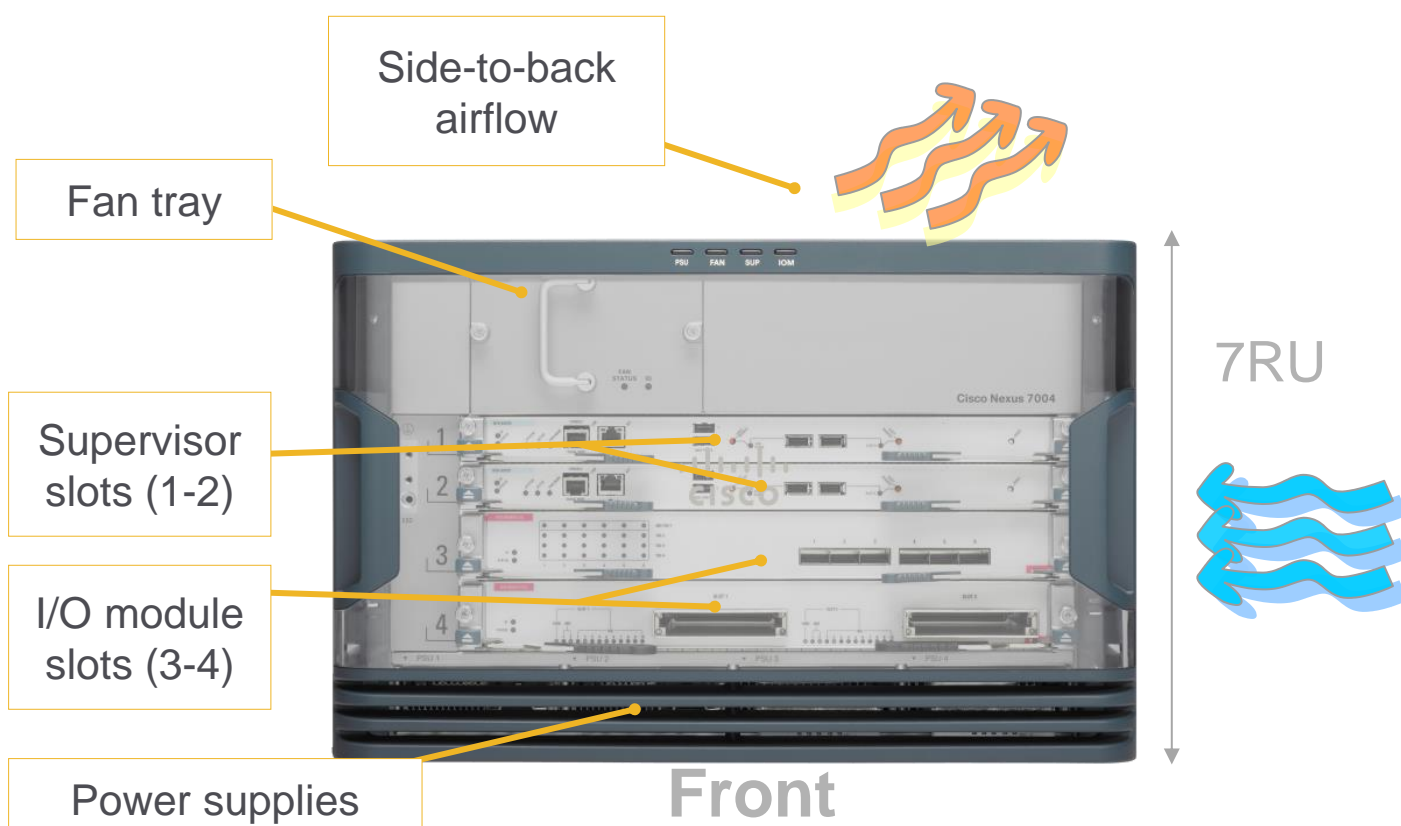
Nexus 7000 Supervisors Additions

Two New Supervisor Choices
Control Plane Scale

Sup2

Quad Core CPU
12GB RAM

Sup2E

Dual Quad Core CPU
32 GB RAM

Same Price-point as Sup1
Additional CPU and Memory
4+1 Admin VDC Support
VDC CPU Shares

8+1 VDC Support
Support for upto 48 FEX
Highest Scale
VDC CPU Shares

DESIGNED FOR: HIGH PERFORMANCE DATA CENTER ENABLING HIGHEST FEATURE SCALABILITY

Supervisor Comparison

	Sup1	Sup2	Sup2E
CPU	Dual-Core Xeon	Quad-Core Xeon	2 x Quad-Core Xeon
Speed	1.66 Ghz	2.13 GHz	2.13 GHz
Memory	8G	12 GB	32 GB
Flash Memory	Compact Flash	USB	USB
CMP	Supported	Not Supported	Not Supported
NX-OS Release	4.0 or later	6.1 or later	6.1 or later
VDCs	4	4+1	8+1
FEX	32 FEX/1536 Ports	32 FEX/1536 Ports	48 FEX/2048 Ports

Cisco Public

Fabric Modules

Fabric 1

- Each module provides **46Gbps per I/O module slot**

  - Up to 230Gbps per slot with 5 fabric modules

N7K-C7009-FAB-2

- I/O modules leverage different amount of fabric bandwidth

- Fabric access controlled using QoS-aware central arbitration with VOQ

Fabric 2

- Increases bandwidth to **110Gbps per I/O module slot!**

  - Up to 550Gbps per slot with 5 fabric modules

- Backward compatible with existing modules

N7K-C7010-FAB-1
N7K-C7010-FAB-2

- Requires NX-OS 5.2 (N7009) & NX-OS 6.0 (N7010/N7018)
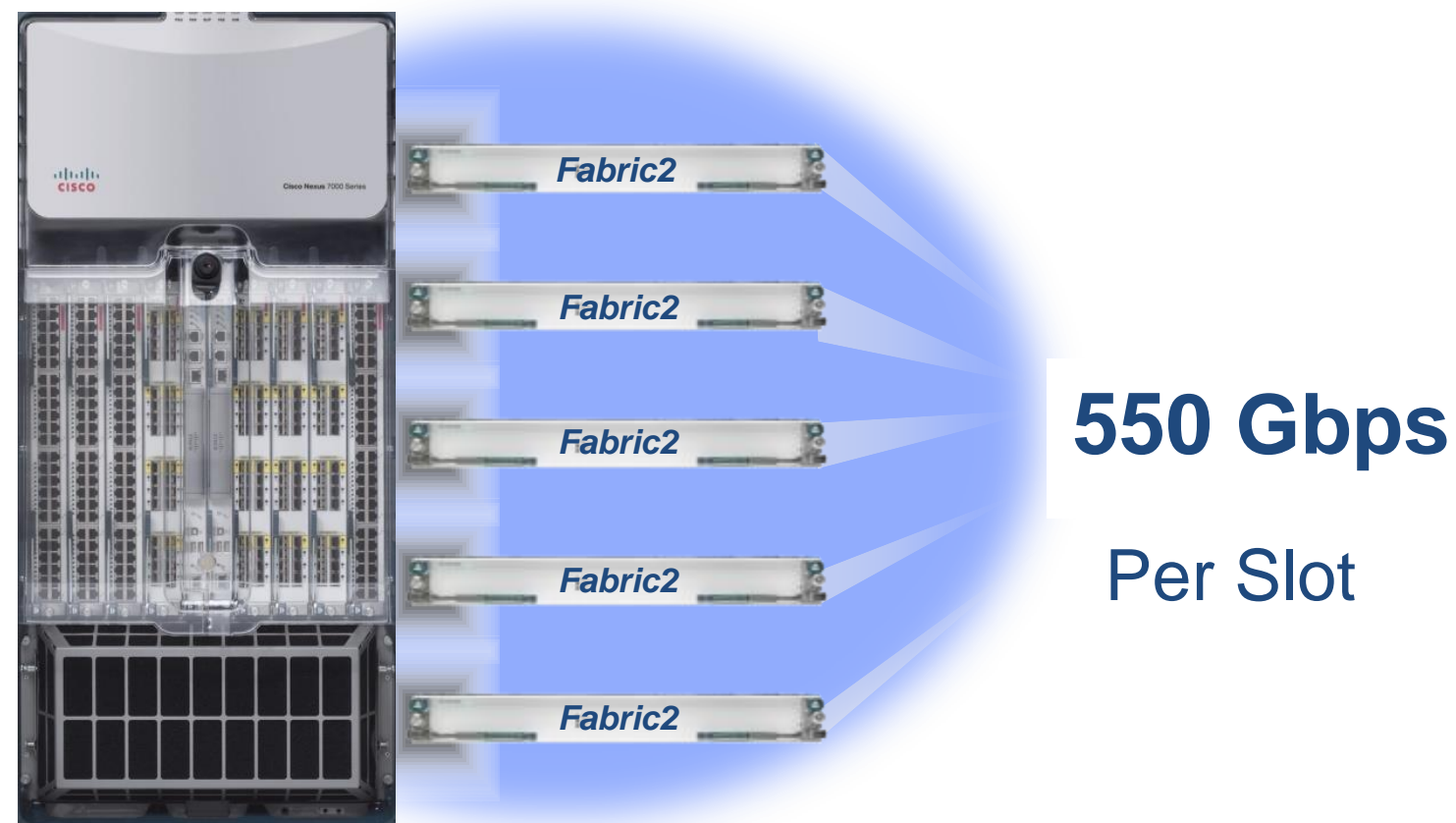
Cisco live!

Fabric 1 to Fabric 2 Migration

In-service upgrade

In-Service upgrade from Fabric 1 to Fabric2

Fast.
Simple.
Non-Disruptive.

Fabric2
Fabric2
Fabric2
Fabric2
Fabric2

550 Gbps

Per Slot

Fab1 to Fab2 Migration - In-Service Upgrade

Fab2 supports all existing Modules

Fab2 is required to utilise F2 / F2e Modules to full 48 port 10GE line rate capacity

Cisco live!

Nexus 7000 I/O Module Families – M and F

M family – L2/L3/L4 with large forwarding tables and rich feature set

New !

N7K-M132XP-12
N7K-M132XP-12L

N7K-M108X2-12L

N7K-M148GT-11/N7K-M148GT-11L

N7K-M148GS-11/N7K-M148GS-11L

N7K-M224XP-23L

N7K-M206FQ-23L

N7K-M202CF-22L

F family – Low-cost, high performance, low latency, low power and streamlined feature set

N7K-F132XP-15
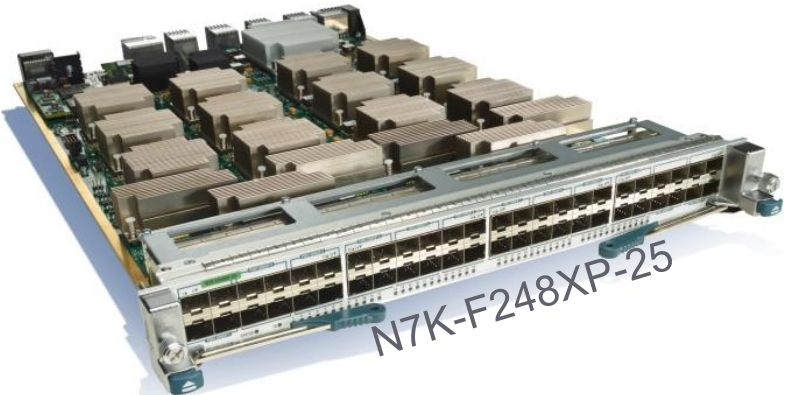
N7K-F248XP-25
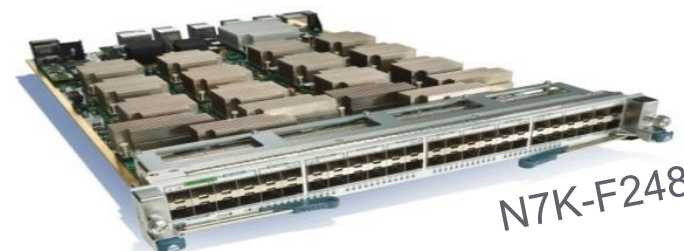
N7K-F248XP-25E
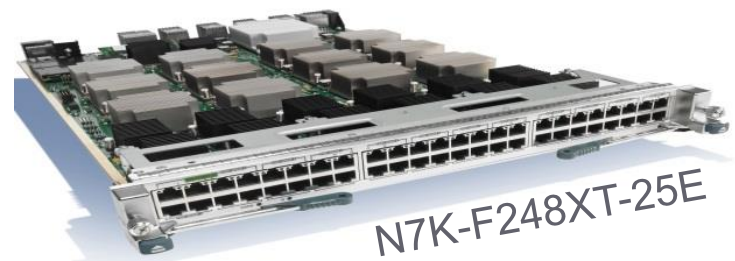
N7K-F248XT-25E

New !

Cisco live!

Which Module to Choose ?

Software Supported Features

MPLS   OTV   LISP

Multicast Features Set

Classic Layer 2 Feature Set

QoS Feature Set

M-Series

Classic Layer 3 Feature Set

Virtual Device Context

F-Series

FabricPath   FCoE

Security Features Set

Common (M or F) Baseline Features

Advanced Features

 Cisco Public
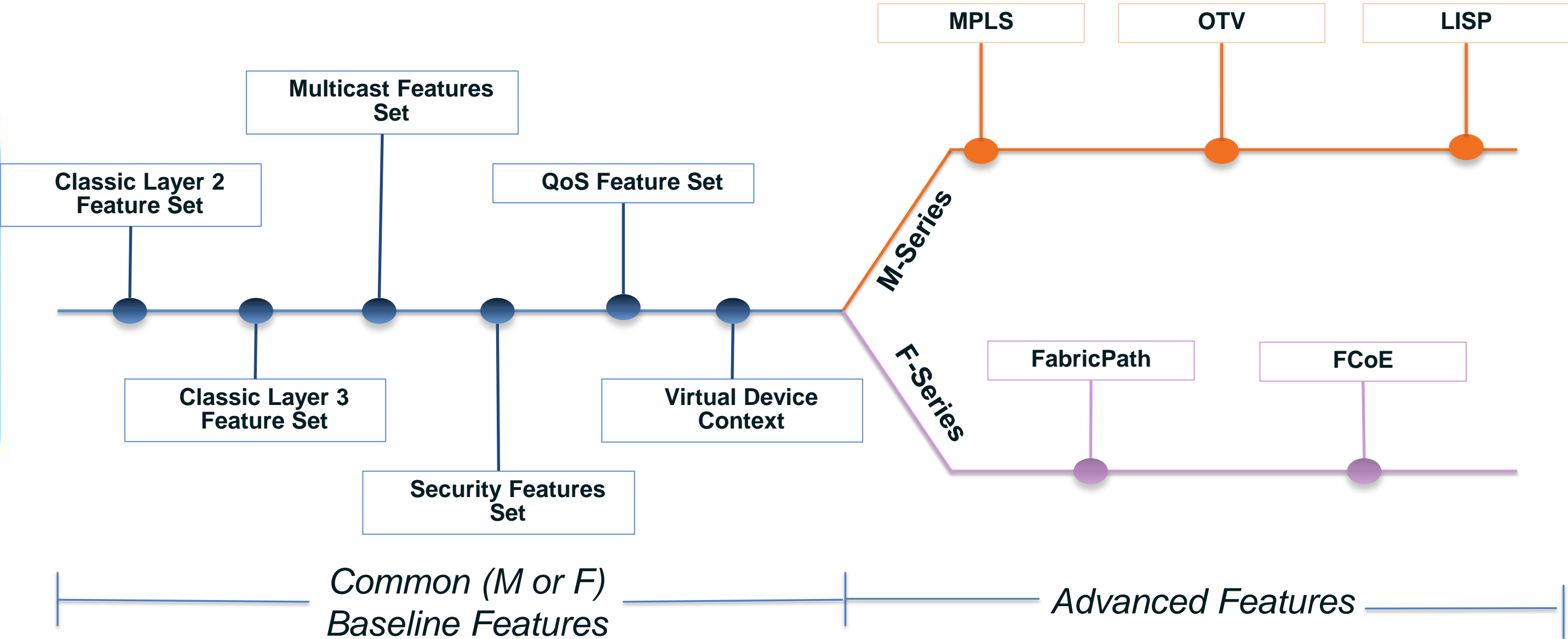
24-Port 10GE M2 I/O Module

N7K-M224XP-23L

- 24-port 10G with SFP+ transceivers (line rate)

- 240G full-duplex fabric connectivity

- Two integrated forwarding engines (120Mpps)

- Support for "XL" forwarding tables (licensed feature)

- Distributed L3 multicast replication

- 802.1AE LinkSec

N7K-M224XP-23L

M Family
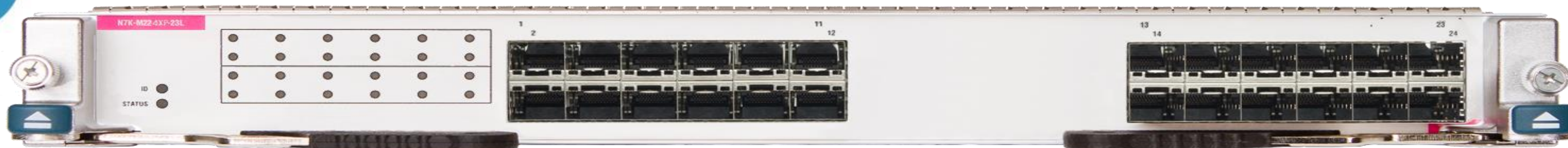
6-Port 40GE M2 I/O Module

N7K-M206FQ-23L

- 6-port 40G with QSFP transceivers

- Option to breakout to 4X10G interfaces per 40G port

- 240G full-duplex fabric connectivity

- Two integrated forwarding engines (120Mpps)

- Support for "XL" forwarding tables (licensed feature)

- Distributed L3 multicast replication

- 802.1AE LinkSec

N7K-M206FQ-23L

CORE / ISP

100G

100G
100G
100G

40 /100G
40 /100G
40 /100G
40 /100G

40 G
40 G
40 G
40 G

10 G
10 G
10 G

Cisco Public

Cisco live!

M Family

2-Port 100GE M2 I/O Module

N7K-M202CF-22L

- 2-port 100G with CFP transceivers

- Option to breakout to 2X40G or 10X10G interfaces

- 200G full-duplex fabric connectivity

- Two integrated forwarding engines (120Mpps)

- Support for "XL" forwarding tables (licensed feature)

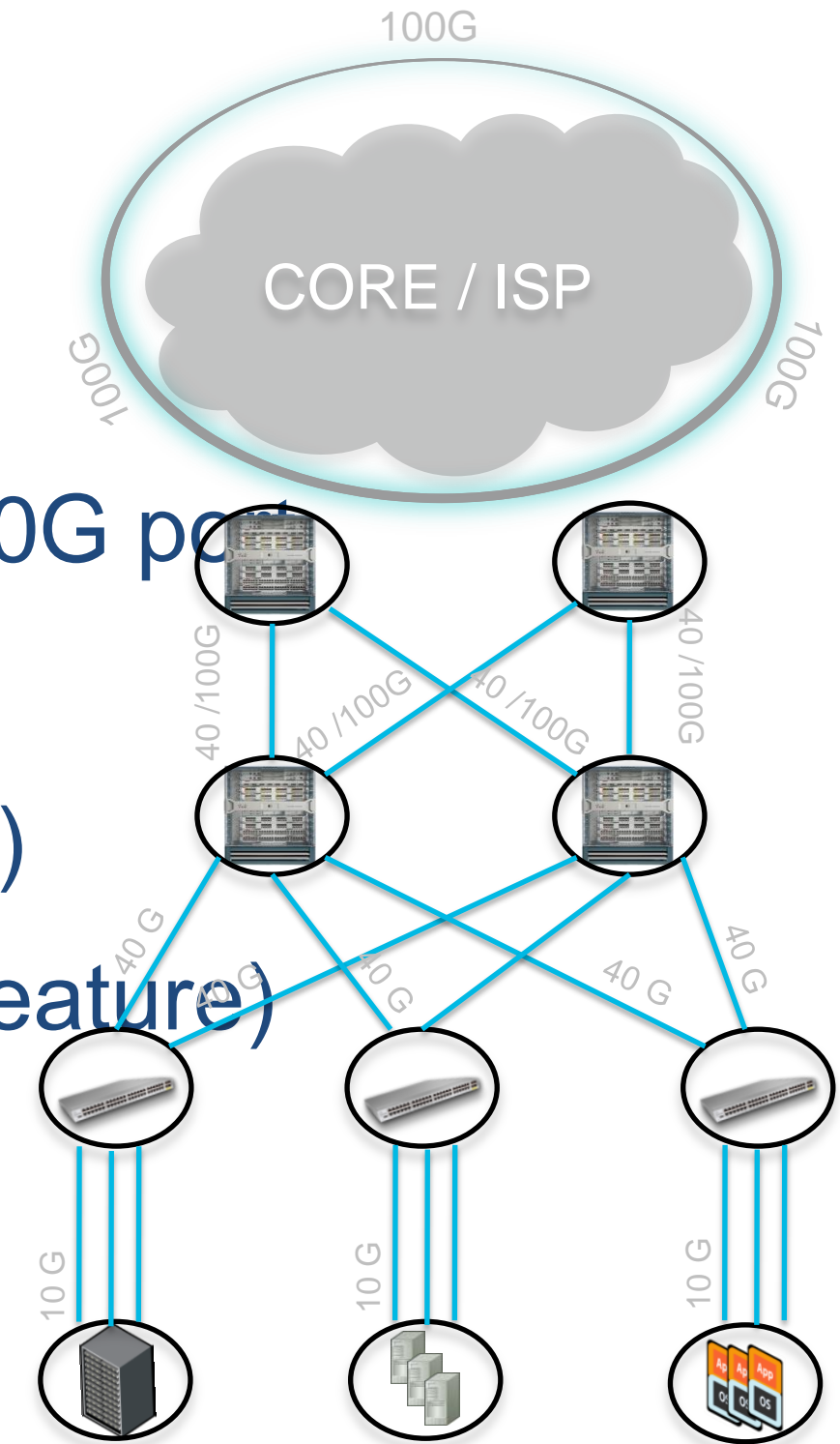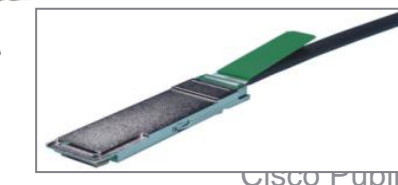- Distributed L3 multicast replication

- 802.1AE LinkSec

CORE / ISP

100G

100G

100G

100G

40 /100G

40 /100G

40 /100G

40 /100G

40 G

40 G

40 G

40 G

40 G

40 G

10 G

10 G

10 G

N7K-M202CF-22L

Cisco live!
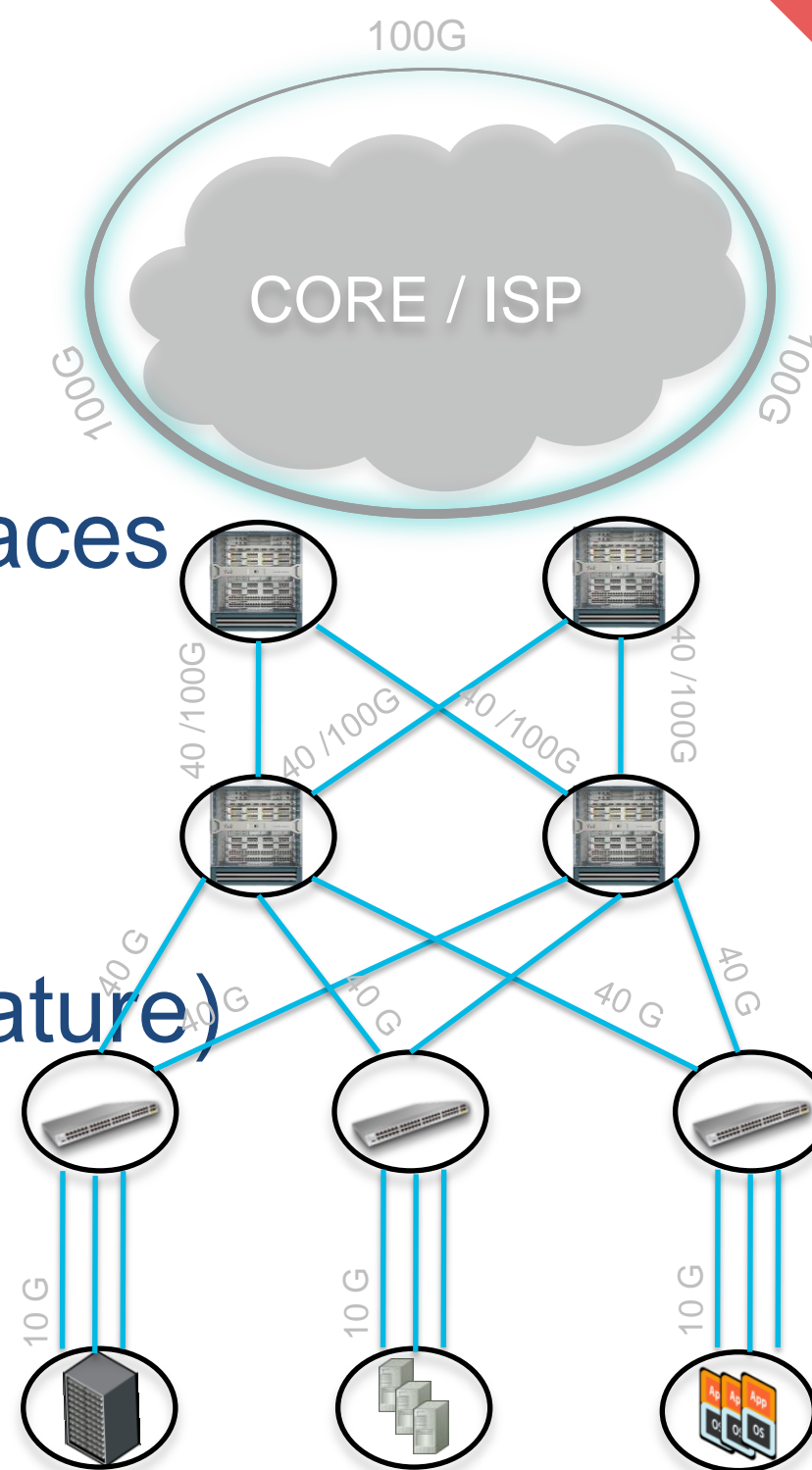
F2-Series Solution Enhanced with F2e

- High Density 48 ports 1/10GE ports (SFP+)

- F2e Series modules enables tighter integration of FabricPath with LISP & MPLS by providing M/F2e VDC Inter-operability support*

- Better scaling by utilising larger M tables

- Ports 41-48 capable of Wire rate encryption with MacSec*

N7K-F248XP-25E
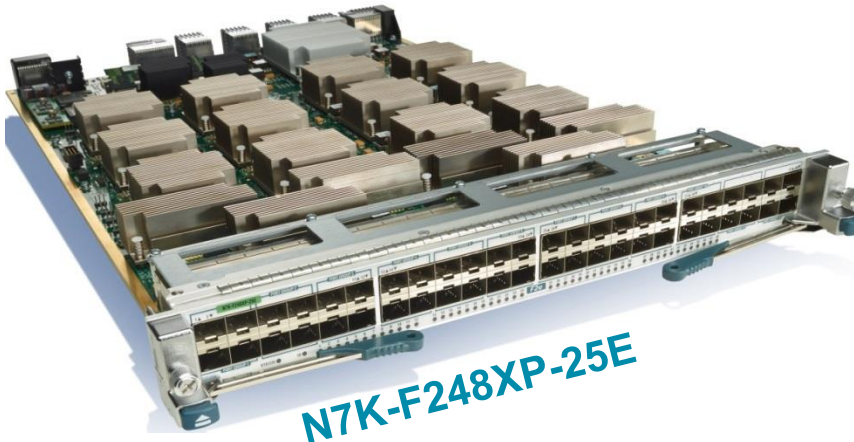
*No dedicated VDC requirement for F2e !

Interop options	Software required	F2e Behaviour
Interop with F2 Series	6.1.2 Nov 2012	In an F2 VDC, F2e behaves like F2 with full L2 and L3 feature set
Interop with M1/M2 Series	6.2 1HCY13	In an M VDC, F2e works in L2 Mode with M-Series providing L3 Proxy forwarding

F2e Copper Supports EoR/MoR

- 48 Copper -1/10 GE line rate ports
- Low Power Consumption: ~8W per Port
- Low Latency ~6usec
- Full Layer 2 and Layer 3 Support
- All ports (1-48) are MACSeC capable**
- FabricPath support for next gen Layer 2 designs

*No dedicated VDC requirement for F2e !

N7K-F248XT-25E

Flexible, Scalable & Cost effective EoR/MoR design

Reach up to 100m on Cat6A cables
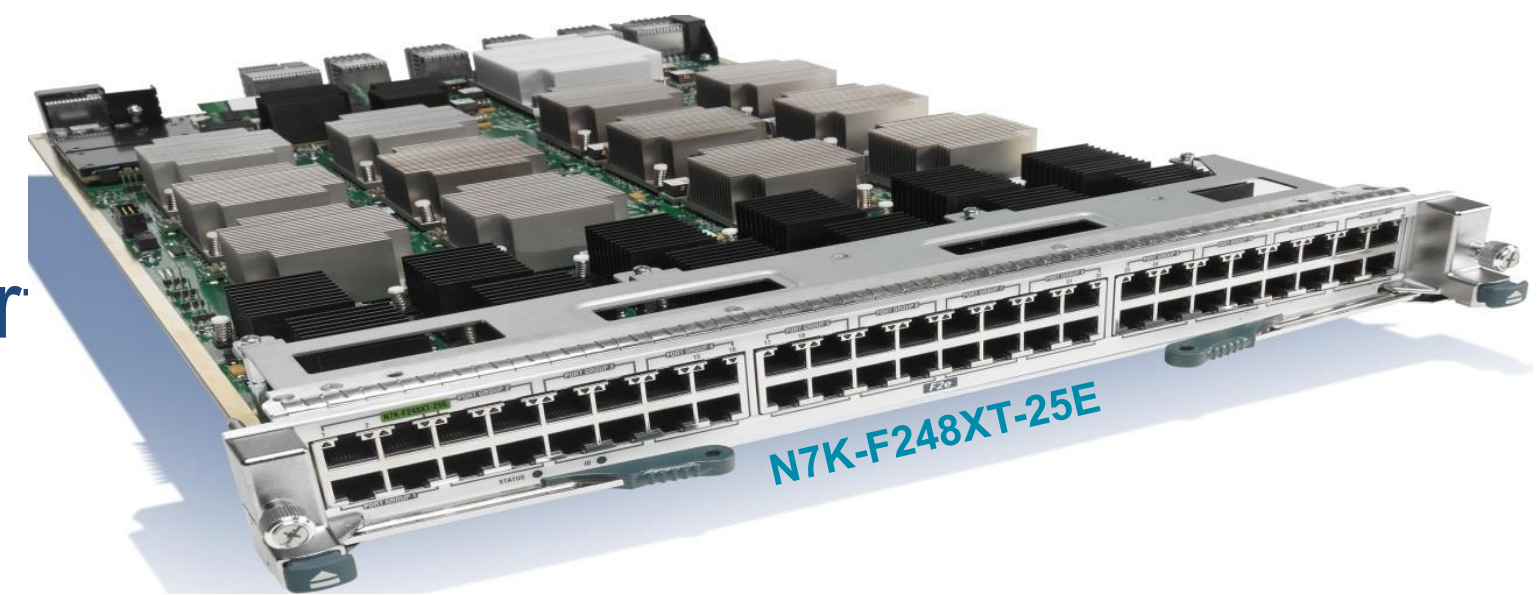
❖ **Ideal for EoR and MoR design**

❖ **Enables Cost Effective MoR/EoR Designs**

❖ **No Need for optics, standard RJ-45 connectors**

**Macsec support will be added in a future release

*Hardware Capable -Requires NX-OS 6.2, No inter-op with F1

Cisco live!

F2e Incremental Features

*No dedicated VDC requirement for F2e!

Features Common to F2 and F2e:

- 48 port 1 and 10GE (SFP+)
- Layer 2 and Layer 3 forwarding
- 1G/10G dual-speed on each interface
- Low Latency
- Nexus 2000 Support
- FabricPath
- IEEE 1588 PTP

Incremental Features in F2e:

- Interoperability with M1XL/M2* (F1 Interop not planned)
- MACSec (802.1AE) *
- Bidir PIM *
- SVI stats *
- IPv6 DSCP-to-Queue Mapping
- 48 Port 1G/10G Copper modules

* Not available at FCS

FEX Support

System scale

Nexus 7000	FEX supported (Pre NX-OS 6.1)	NEW in NX-OS 6.1
Up to 2048[1] host ports		

Cisco Nexus 7000

1 – Requires SUP2E

Cisco Nexus® 2248TP

Cisco Nexus® 2232PP

Cisco Nexus® 2224TP

Cisco Nexus® 2248TP-E
N2K-C2248TP-E

Cisco Nexus® 2232TM
N2K-C2232TM-10GE

- FEX supported with

    - SUP1, SUP2 and SUP2E support FEX

    - M132XP, M224XP & F2 series modules **(F1/M18 Series exception)**

- **Up to 48 FEX (both 1GE and 10GE FEX) modules supported with SUP2E (6.1)**

- Choice of 1G/10G interfaces with FEX

Nexus 7000 – Features & Leading Practices
Virtual Device Contexts

Nexus 7000 Virtual Device Contexts (VDC)
Partition One Physical Switch to Multiple Secure Virtual Switches

Consolidate Multiple Physical Switches into One

VDC 1
VDC 2

Layer 2 Protocols		Layer 3 Protocols	
VLAN	UDLD	OSPF	GLBP
PVLAN	CDP	BGP	HSRP
STP	802.1X	EIGRP	IGMP
LACP	CTS	PIM	SNMP
...		...	

Admin VDC

VDC 1
VDC 2
VDC 8

Core   Agg

DMZ   Internet
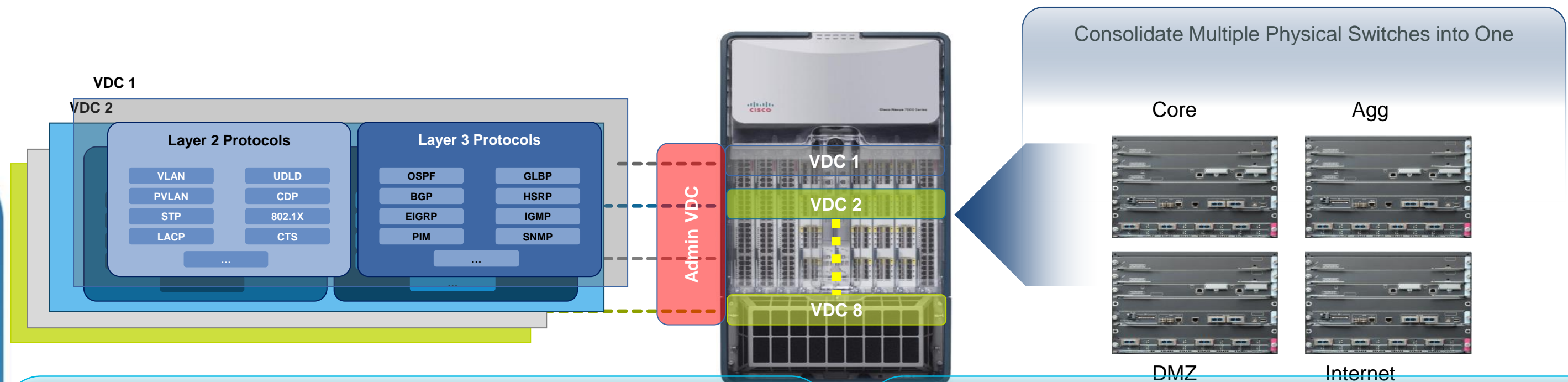
Features
- Flexible separation/distribution of hardware resources and software components
- Complete data plane and control plane separation
- Complete software fault isolation
- Securely delineated administrative contexts

Benefits
- Device consolidation, both vertical and horizontal
- Reduced number of devices—lower power usage, reduced footprint and lower CapEx/OpEx
- Fewer devices to manage
- Optimise investment

Cisco live!

Virtual Device Contexts (VDCs)
VDC Port Allocation

```
Nexus7K(config)# vdc secure-net id 2
Nexus7K(config-vdc)# allocate interface e2/1,e2/3,e2/5,e2/7
Nexus7K(config-vdc)# allocate interface …..
Nexus7K(config-vdc)# exit
```

Any individual ports on the 48 Port 1GE & 8 Port 10GE **M1**(ex : Port 1 , Port 2 etc)

Any individual ports on the 24 Port 10GE,6 port 40GE & 2 Port 100GE **M2** (ex : Port 1 ,Port 2 etc)

All ports in the same port-group on 32 port 10GE **M1** modules (Ex : 1,3,5,7 …2,4,6,8 etc)

M132
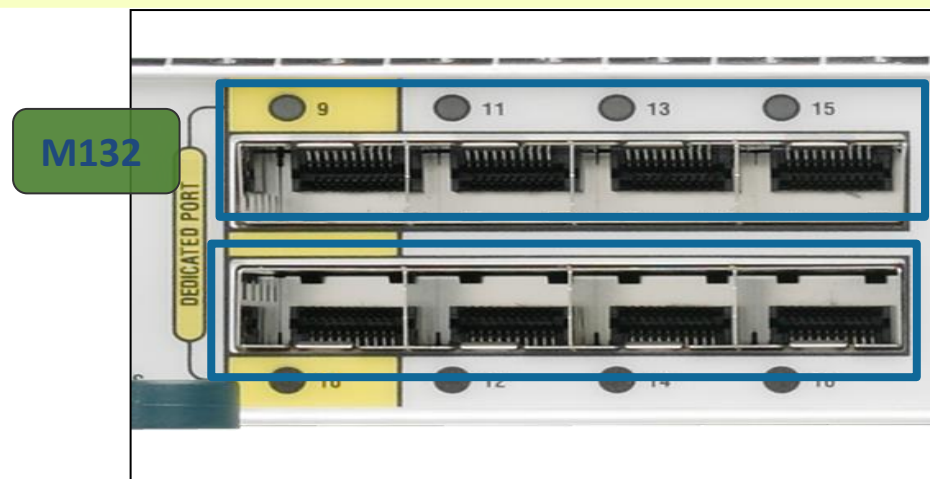
All Ports in a SoC (Port-Group) on 32/48 port 10GE **F1** / **F2/F2e** Modules (ex Ports 1,2… Ports 7,8 etc on **F1**)   & (Ports 1,2,3,4….. Ports 13,14,15,16 etc on **F2/F2e**)

F1

F2/F2e

M108

M2-10G

M148

M2-40G

M148

M2-100G

Cisco live!
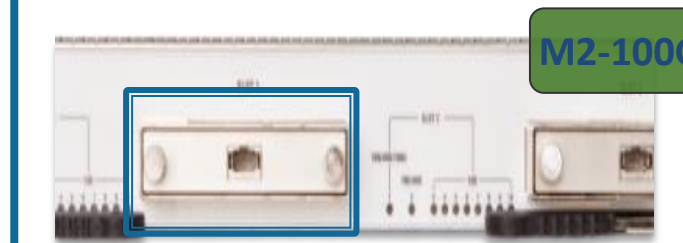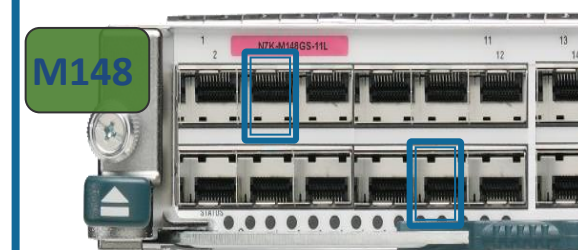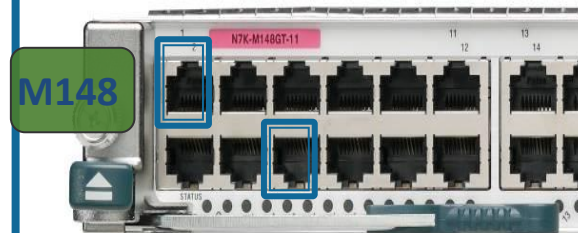
Virtual Device Contexts (VDCs)

VDC Module Allocation

- If VDC has both M1(non-XL) and M1-XL /M2-XL modules, system will operate with least common denominator mode

- VDC resource-limit module-type customise as needed

- Default VDC mode allows M1 / F1/ M1-XL / M2-XL Modules
Other dedicated modes (ex:F1,M1,M1-XL,M2-XL & F2 only) are configurable

```
Nexus7K(config)# vdc inet
Nexus7K(config-vdc)# limit-resource module-type m2-xl
```

It is recommended to allocate whole modules per VDC, Helps with better hardware resource scaling

Communication Between VDCs

- Must use front panel port to communicate between VDCs

- No soft cross-connect or backplane inter-VDC communication

VDC1 Admin

VDC2 INET

VDC3 ENET

Lab VDC

M1 Modules only

INET Facing VDC

M2-XL modules only

VDC2

VDC3
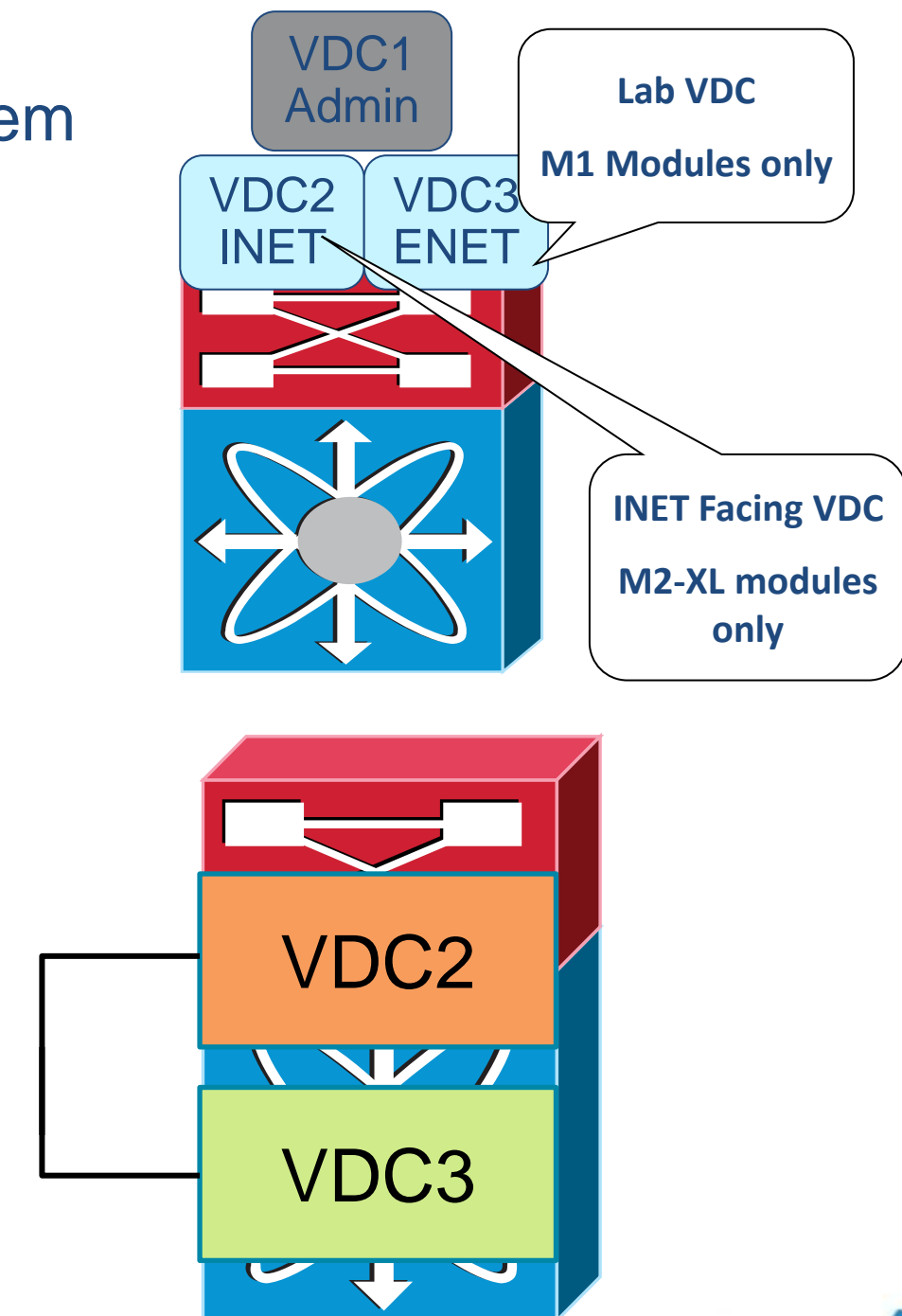
Admin VDC
Purely Administrative Context

- Available on Supervisor 2/2E
- Provides pure administrative context
    - CoPP configuration / HWRL Configuration
    - ISSU and EPLD
    - VDC creation, suspension and deletion, interface allocation
    - Show tech-support, tac-pac, debugs, GOLD Diagnostics
    - System-wide QoS, Port Channel load-balancing
    - Poweroff & out-of-service Modules
    - License Management
- Simplify configuration for data plane VDCs
    - No boot statements, CoPP policies, etc in Non-Admin VDCs
- Addresses customer security concerns about network administrator role
    - Can better leverage VDC administrator role

Admin

Management Functions

CoPP   GOLD
ISSU   Licensing
VDC   EPLD
...

Admin

Infrastructure

Kernel
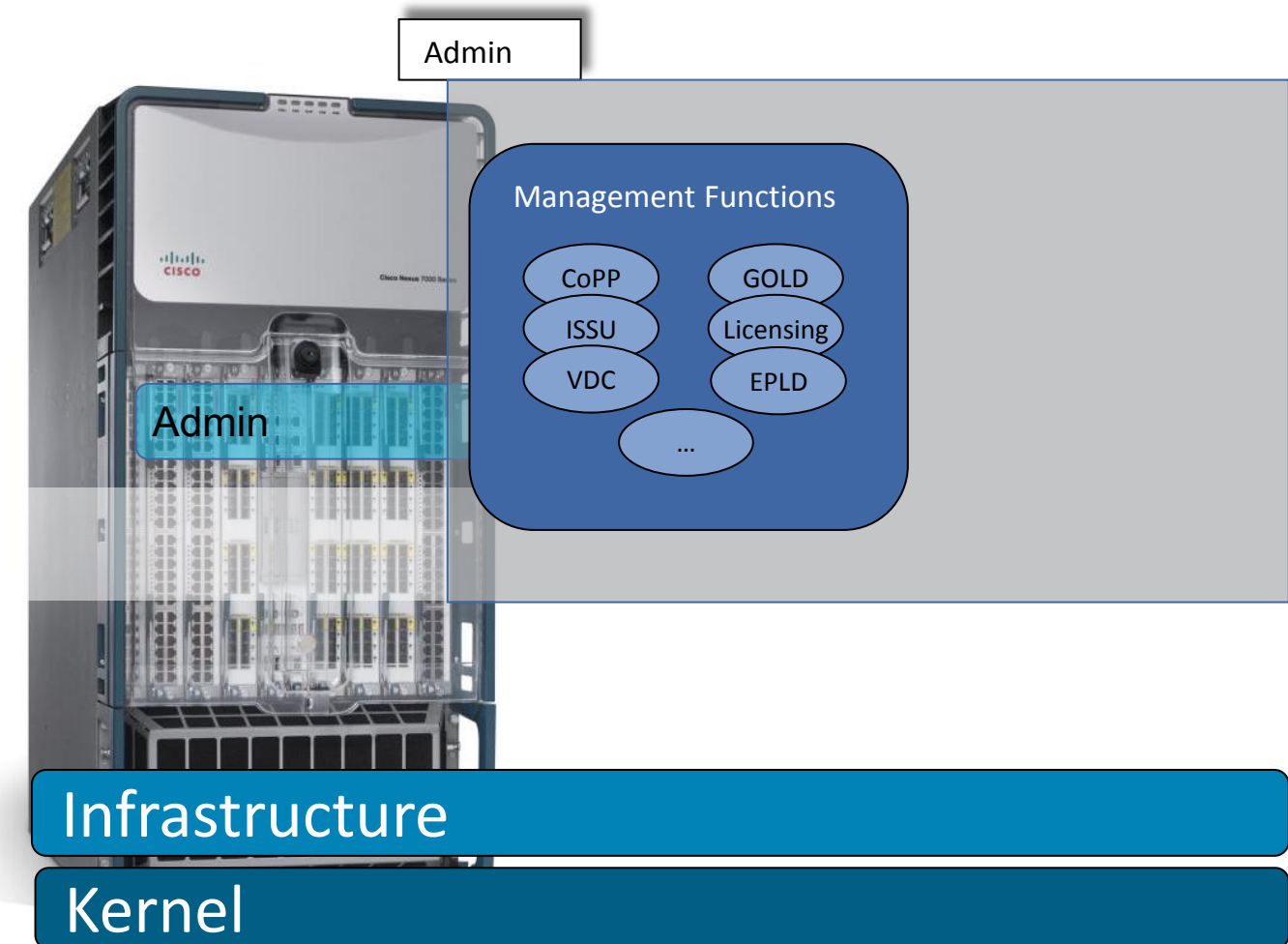
- Doesn't require Advanced or VDC License
    - Can use 1 Admin VDC + 1 Data VDC (1+1)

- Initially only available on Supervisor 2/2E
    - Planned for SUP1 in later half of CQ13

VDC CPU Shares

- Enables per-VDC CPU Access & Prioritisation

- Provides more control and protection per VDC for users

- Network administrator controls each VDC's priority

- CPU share is controlled by VDC priority & CPU is shared equally among VDCs

- User can control allocation – priorities are linear in effect

- The more VDCs configured, the lower the overall percentage per VDC

- Comes into use when CPU utilisation increases (contention)

- Available on SUP2/2E only*

- CPU shares take effect immediately (no need to restart/reload)

VDC1
Shares = 2

VDC2
Shares = 4

VDC3
Shares=1

VDC4
Shares=8

VDC5
Shares=10

VDC6
Shares=5
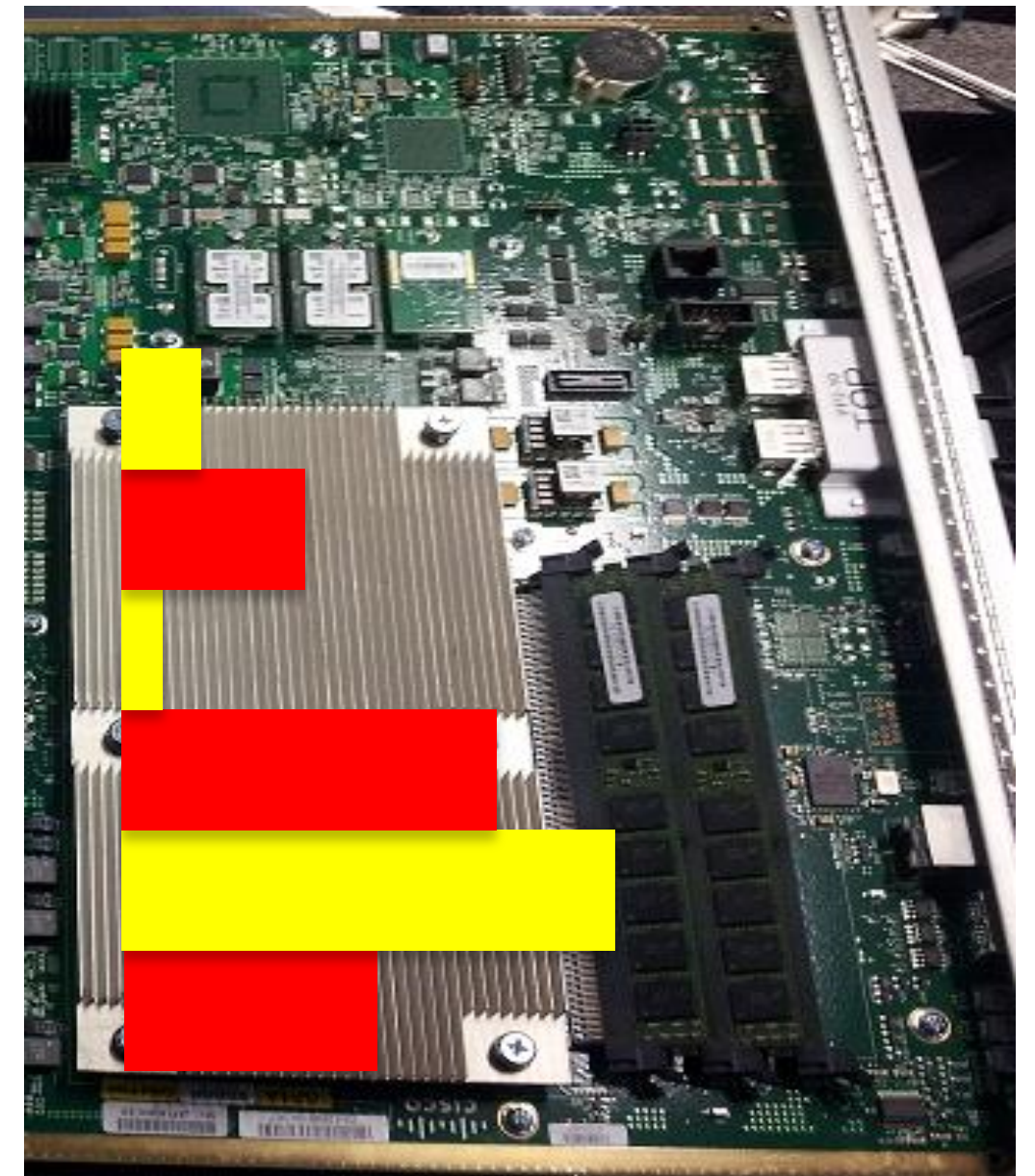
- Controlled by NX-OS scheduler in the kernel

- Processes that do not want the CPU do not affect CPU time of other processes

*SUP2 and SUP2E Require NX-OS 6.1

VDC CPU Resource Allocation

CPU Shares

- Configured under each VDC

- Shares are defined on a scale of 1-10 with default of 5

  - 10 is highest priority, 1 lowest

```
vdc Agg1 id 2
  limit-resource module-type m1 f1 m1xl m2xl
  allow feature-set ethernet
  allow feature-set fabricpath
  allow feature-set fex
  cpu-share 5
  allocate interface Ethernet4/1-8
  boot-order 1
<snp>

N7K-1# show vdc Agg1 det

vdc id: 2
vdc name: Agg1
vdc state: active
vdc mac address: 00:26:98:0f:d9:c2
vdc ha policy: RESTART
vdc dual-sup ha policy: SWITCHOVER
vdc boot Order: 1
CPU Share: 5
CPU Share Percentage: 20%
vdc create time: Mon Apr 23 15:13:39 2012
vdc reload count: 0
vdc restart count: 0
vdc type: Ethernet
vdc supported linecards: m1 f1 m1xl m2xl
```

Default CPU Share

CPU Share Percentage shows minimum % during contention

Cisco Public

Nexus 7000 –Features & Leading Practices
Virtual Port Channel

vPC – Virtual Port Channel

- vPC allows a single device to use a port channel across two neighbour switches (vPC peers) (Layer 2 port channel only)

- Eliminate STP blocked ports & reduces STP Complexity (Do not disable STP)

- Uses all available uplink bandwidth - enables dual-homed servers to operate in active-active mode

- Provides fast convergence upon link/device failure

- If HSRP enabled, both vPC devices are active/active on forwarding plane

- Available since NX-OS 4.1(3) on the Nexus 7000 & NX-OS 4.1(3)N1 on N5K

vPC Peers

MCEC

vPC Peers

MCEC
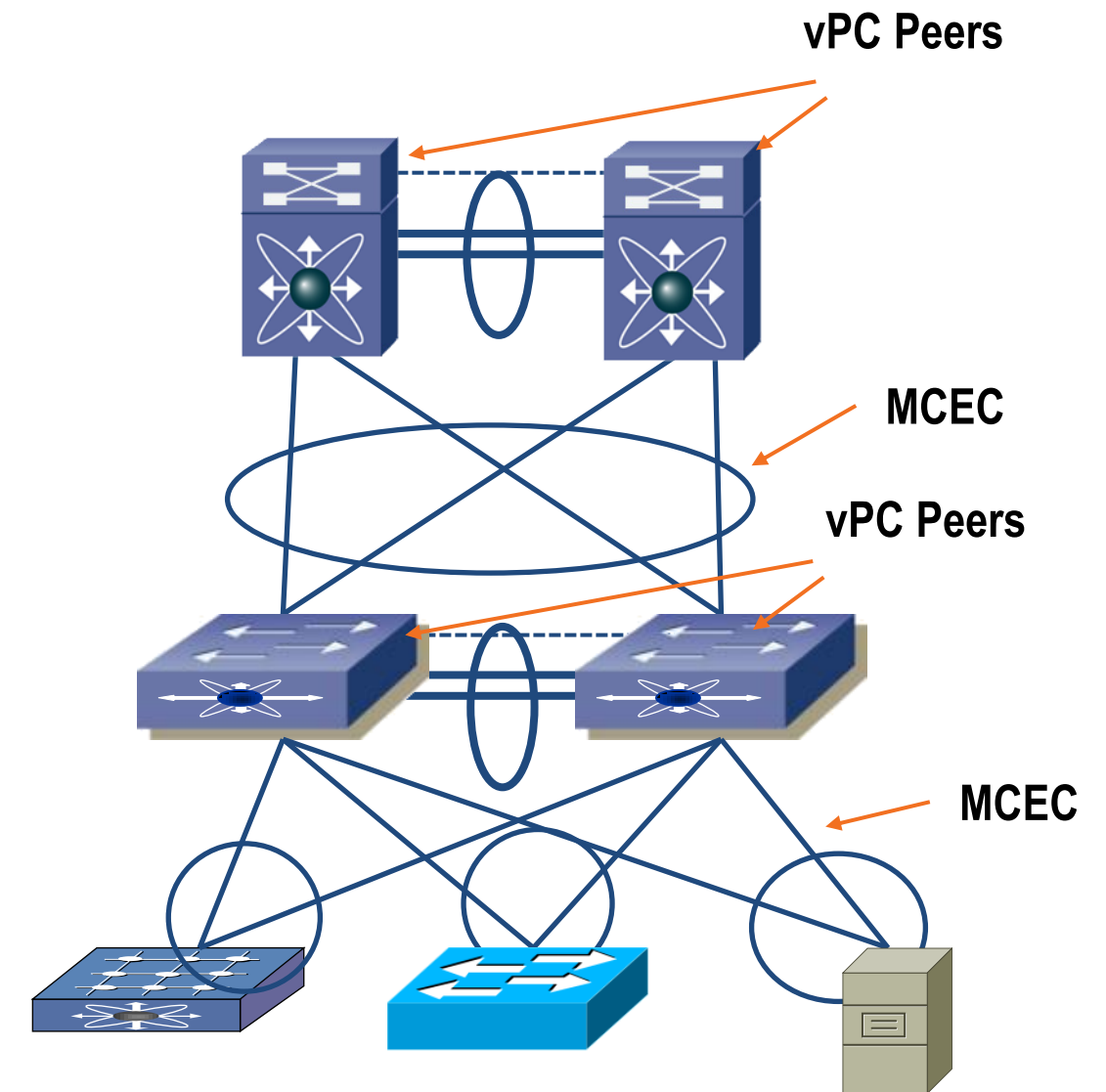
```
! Enable vpc on the switch
dc11-7010-1(config)# feature vpc

! Check the feature status
dc11-7010-1(config)# show feature | include vpc
vpc                        1          enabled
```

vPC/STP Best Practices

- **Do not disable STP !!**
- Configure vPC peers in aggregation layer as primary/secondary root
- BA (Bridge Assurance) is enabled by default on vPC peer-link
- Do not enable Loopguard and BA on vPC (disabled by default)
- Enable STP port type "edge" and port type "edge trunk" on host ports
- Enable STP BPDU-guard globally on access switches
- Selectively allow vlans on trunks

VLAN 1 - 4094
STP Pri 8192

VLAN 1- 4094
STP Pri 16384

agg1a

agg1b

BA Enabled (Default)

No BA or Loopguard (Default)

BPDU-guard

Port Type   Edge / Edge Trunk
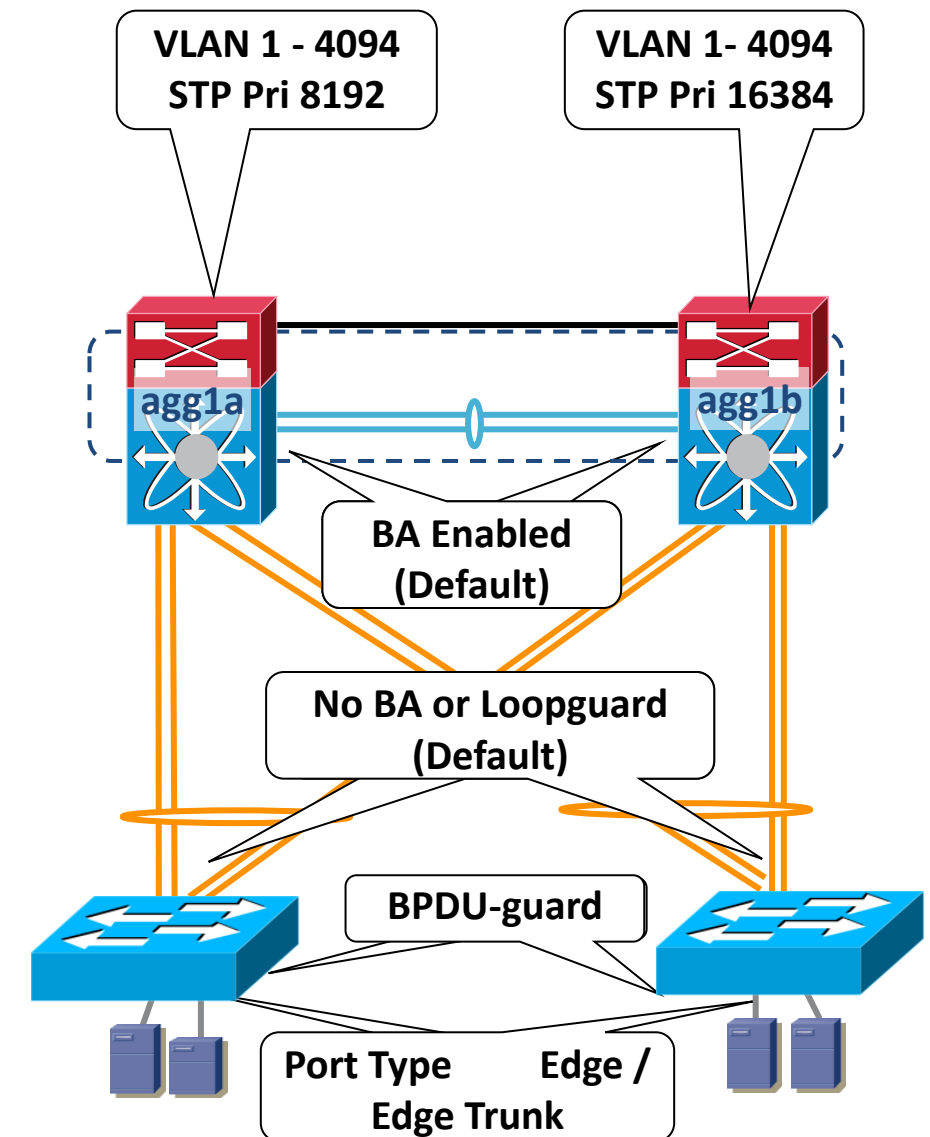
vPC Additional Recommended Config

- Strong Recommendations
    - Designate vPC primary / secondary peer
    - Implement IGP routing between vPC peers
    - Enable vPC peer-gateway
    - Enable vPC auto-recovery
    - Enable vPC ARP Sync
    - Use VPC Graceful Consistency check enhancement
    - Use Per Vlan Type-1 Check enhancement
- Optional Recommendations
    - Enable peer-switch in a pure vPC topology
    - Enable Orphan port-suspend if needed

Enable Peer-Gateway

vPC Auto-recovery

vPC Auto-recovery

Enable Peer-Gateway

vPC_PKL

routing peer

agg1a

agg1b

vPC Primary role pri 8192

vPC Secondary role pri 16384

vPC_PL

vPC ARP Sync

vPC ARP Sync
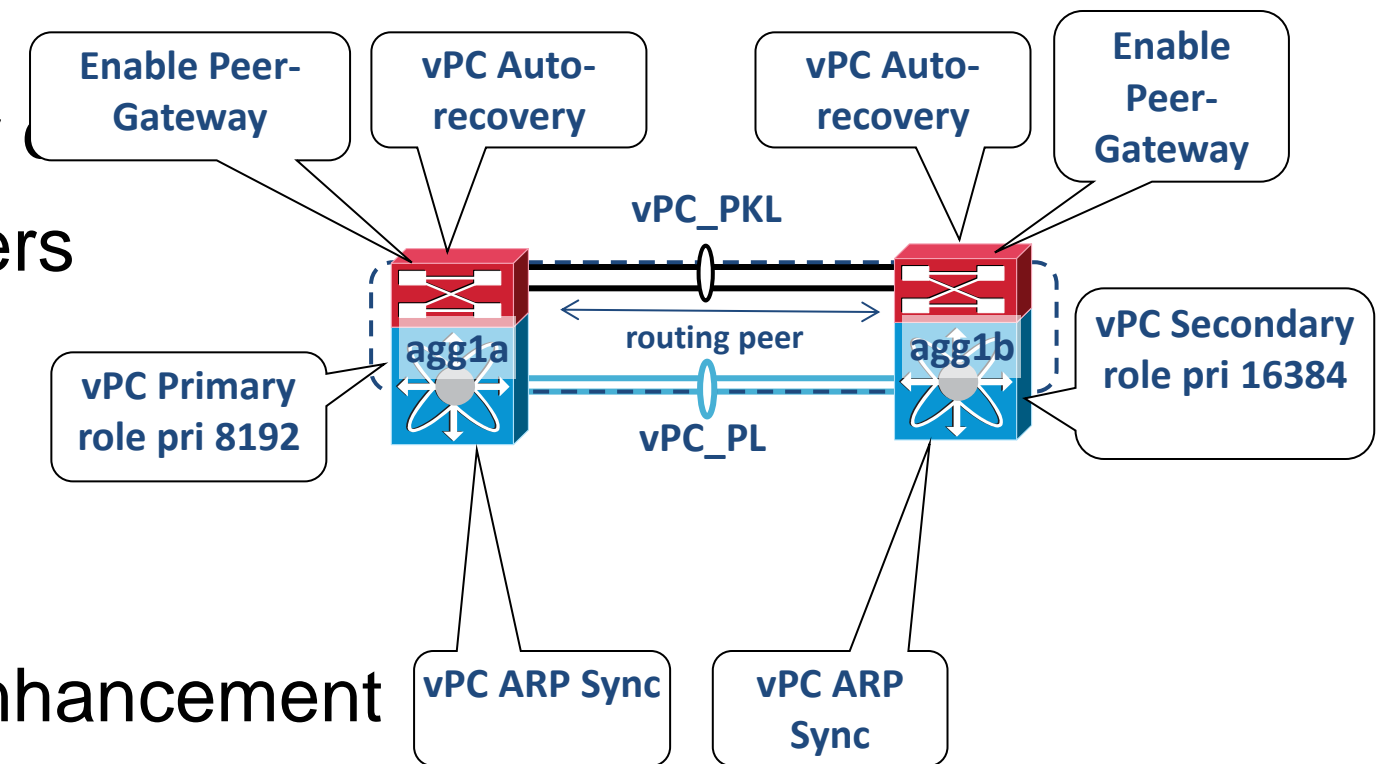
vPC Best Practices Configuration

- **Use dedicated connection for vPC peer keepalive link and assign to a separate VRF**

  - Port-channel is recommended but not required

  - Connect via an OOB mgmt network if using mgmt 0 interface

  - Do not use SVI interface between vPC peer-link as vPC keepalive link

- **Utilise diverse 10GE modules to form vPC peer-link (must be 10GE port-channel)**

  - Peer-Link port-channel requires identical modules on same & other side and can use any 10GE Module (M1,M2,F1,F2/F2e)

  - Dedicated mode (For M132) recommended , Shared mode is supported but not recommended

- **vPC peer-link must be configured as a trunk**

10.1.1.1
VRF
vPC-Keepalive

10.1.1.2
VRF
vPC-Keepalive

vPC_PKL

agg1a   agg1b

e3/47-48)

Identical modules required
on both sides

Po1

e1/1   e1/1

agg1a   agg1b

e2/1   e2/1

vPC_PL

Trunk allowed VLANs =
vPC VLANs
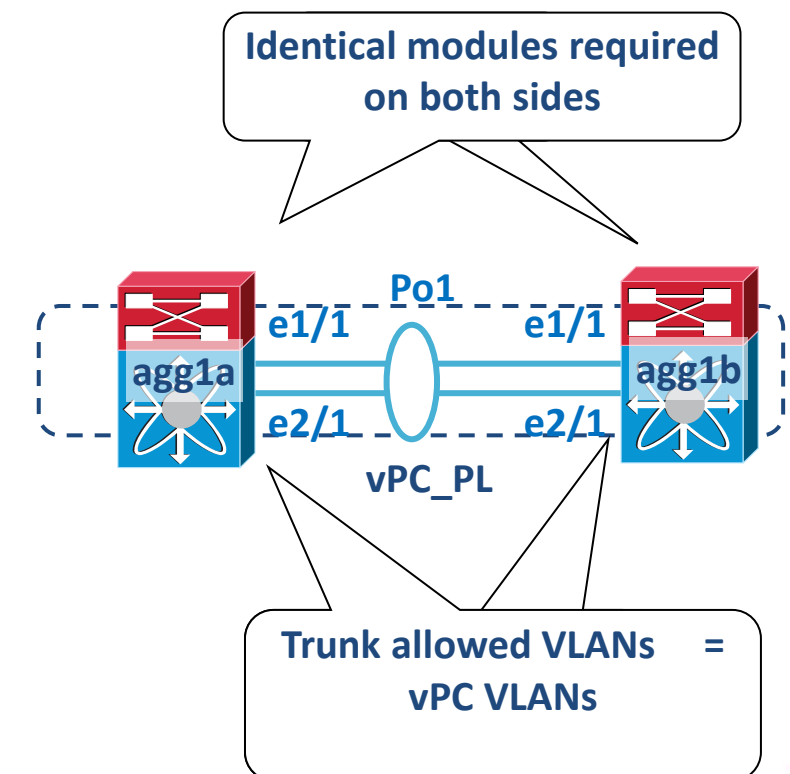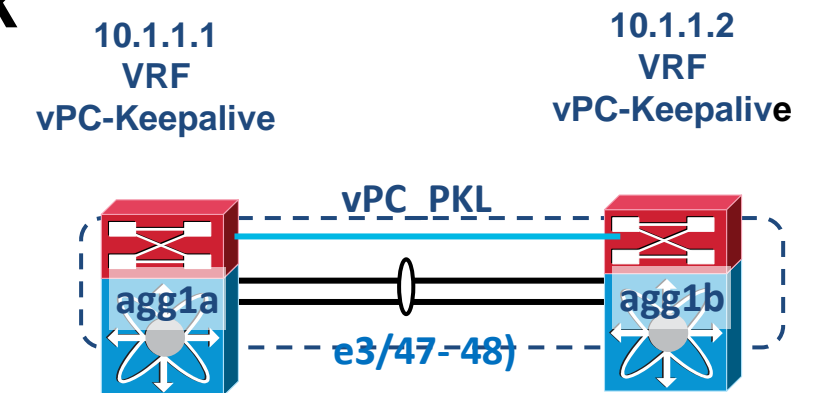
vPC Best Practices Configuration

- **Always dual home all devices to vPC domain using vPC !!**

  - Failure of peer-link can isolate single attached devices

- Match vPC with port-channel number for better administration

- Clear unnecessary VLANs on trunks

```
Nexus7K-1a(config)#
int e3/1-2
switchport
switchport mode trunk
channel-group 11 mode active
!
int port-channel 11
 switchport
 switchport mode trunk
 switchport trunk allowed vlan 100-110
 vpc 11
```


Core

vPC Primary

vPC Secondary

vPC_PKL

agg1a **vPC_PL** **agg1b**

Shut SVIs vPC VLANs

Orphan port

Isolated!!

Acc1 **Acc2** **Acc3**
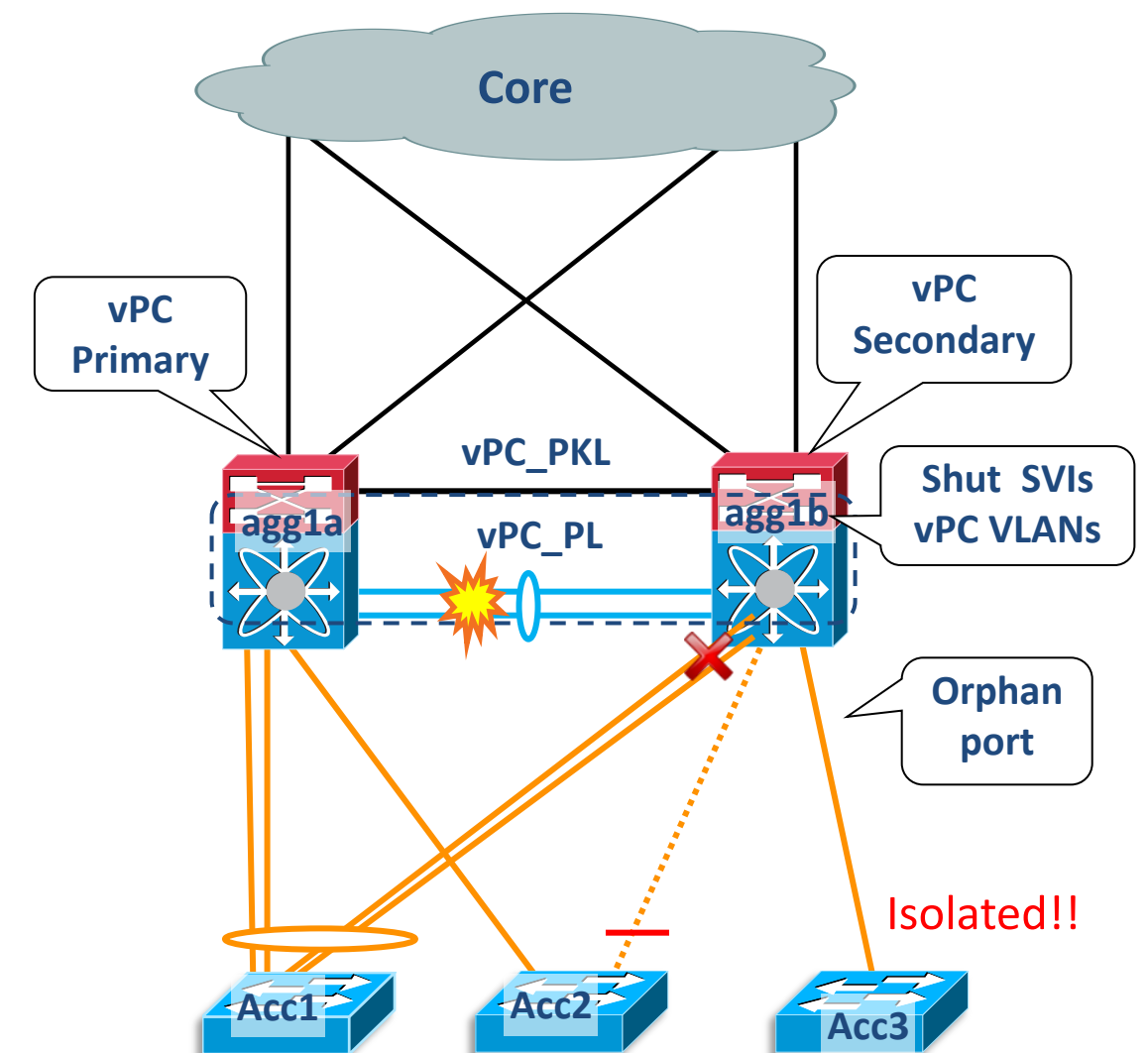
If vPC peer-link fails, the secondary vPC peer suspends local vPCs and shuts down SVIs of vPC VLANs

Orphan-Port Suspend

vPC Active / Standby NIC teaming support

- A vPC orphan port is an non-vPC interface on a switch where other ports in the same VLAN are configured as vPC interfaces

- Prior to release 5.0(3)N2 on Nexus 5000/5500 and 5.2 on Nexus 7000 an orphan port was 'not' shut down on loss of vPC peer-link

- With the supported release the orphan ports on the vPC secondary peer can (configurable) also be shut down triggering NIC teaming recovery for all teaming configurations

- Configuration is applied to the physical port*

vPC

eth 100/1/1

```
N5K-2(config)# int eth 100/1/1
N5K-2(config-if)# vpc orphan-port suspend
```

vPC Supported Server fails over correctly

Active/Standby Server does not fail over correctly since orphan port is still active

*'VPC orphan-port suspend' with FEX host interface requires 6.1.2 release due to CSCua35190
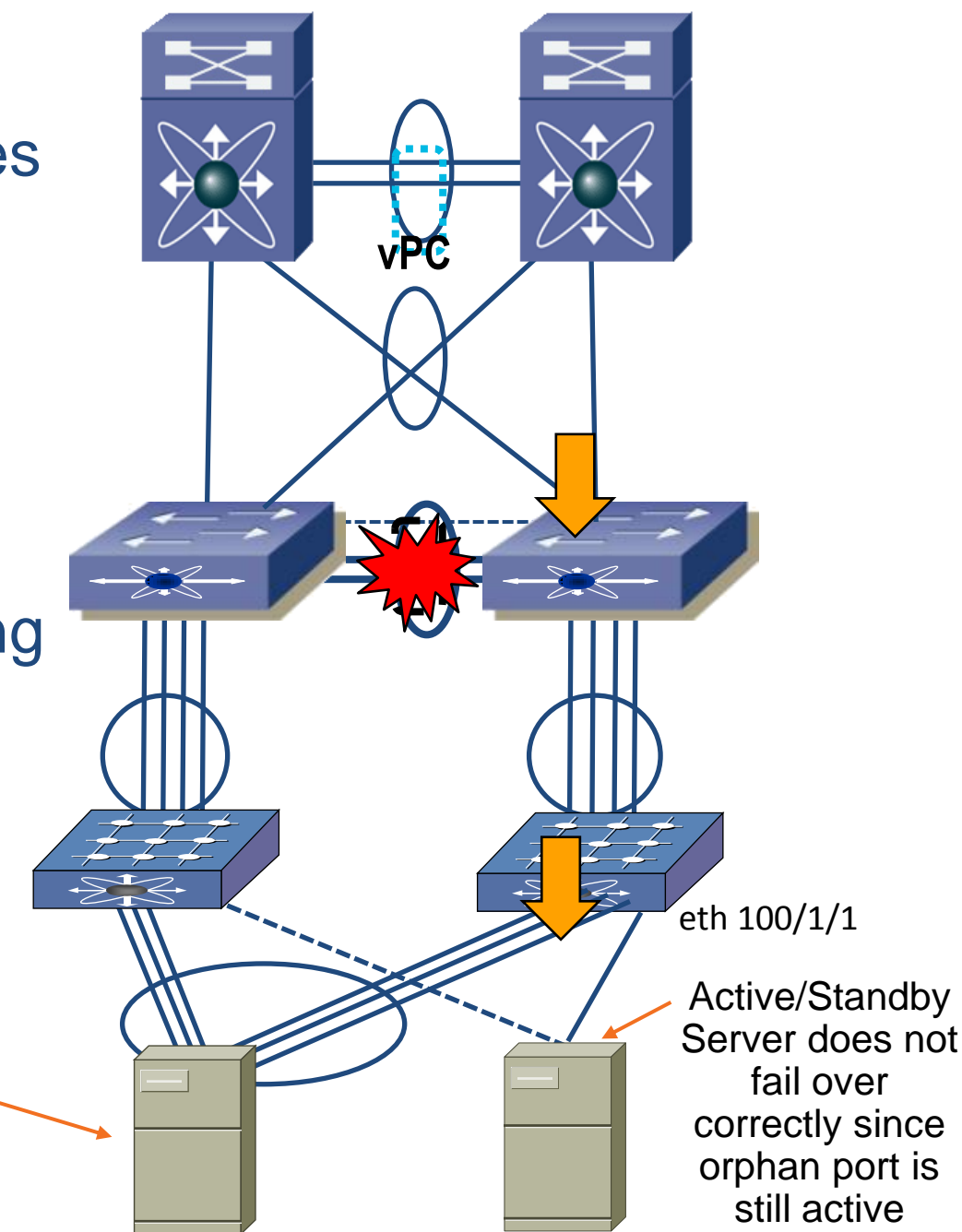
vPC Graceful Type-1 Check

- vPC member ports on S1 and S2 should have identical parameters (MTU, speed, …)

- Any inconsistency in such parameters is Type 1➔ all vlans on both vpc legs are brought down in such Inconsistency

- With graceful type-1 check, only Secondary vPC members are brought down.

- vPC member ports on primary peer device remain up

Keepalive

S1 -Primary   S2-Secondary

vPC peer-link

Type-1 Inconsistency

vPC 1

po1

CE-1
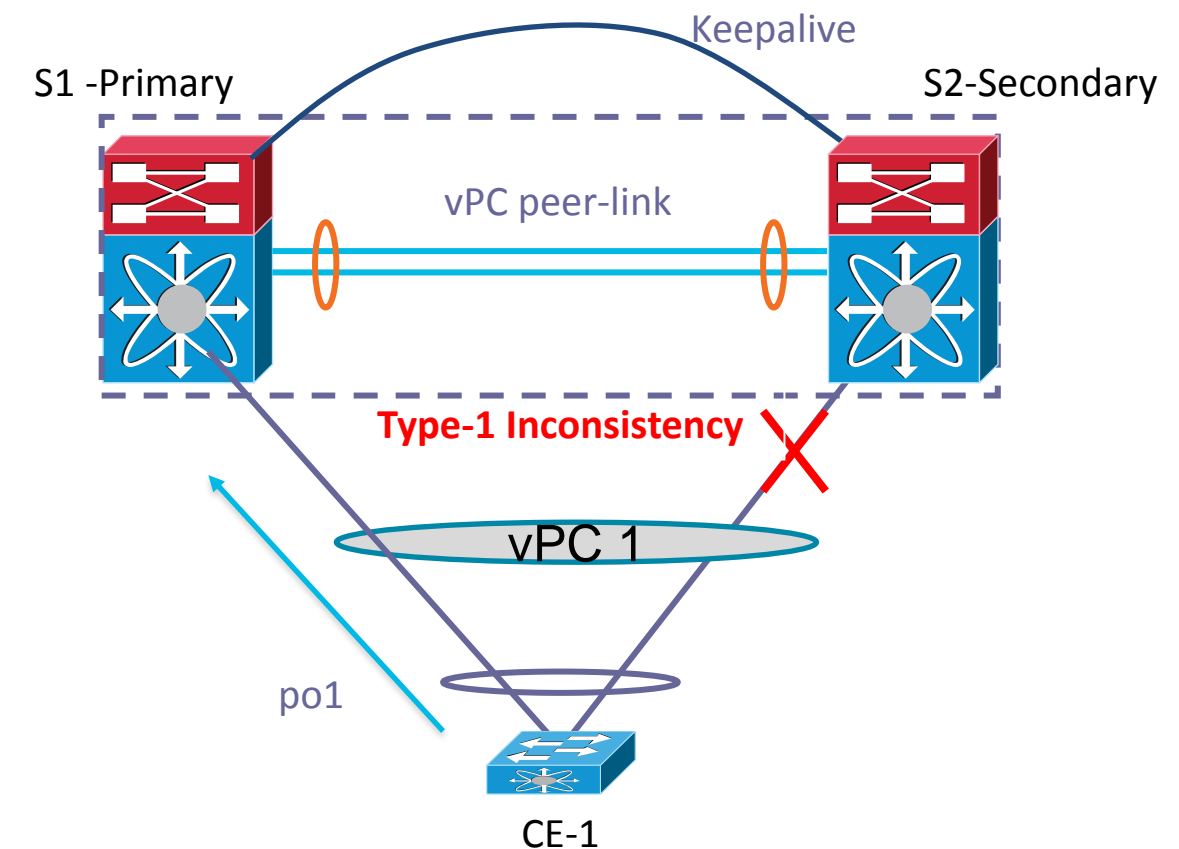
- S1(config-vpc-domain)# graceful consistency-check

- S2(config-vpc-domain)# graceful consistency-check

- Graceful Type-1 check enabled by default.

Cisco Fabric Path

- FabricPath connects a group of switches using an arbitrary topology and provides Scalability, High Bandwidth, High Resiliency, L3 integration and L2 integration

- Benefits of FabricPath

  - Eliminate Spanning tree limitations

  - Multi-pathing across all links, high cross-sectional bandwidth

  - High resiliency, faster network re-convergence

  - Any VLAN, any where in the fabric eliminate VLAN Scoping

```
Nexus7K(config)# feature fabricpath
Nexus7K(config)# feature switch-id <#>
Nexus7K(config)# interface ethernet 1/1
Nexus7K (config)# switchport mode fabricpath
```

With a simple CLI, aggregate them into a Fabric
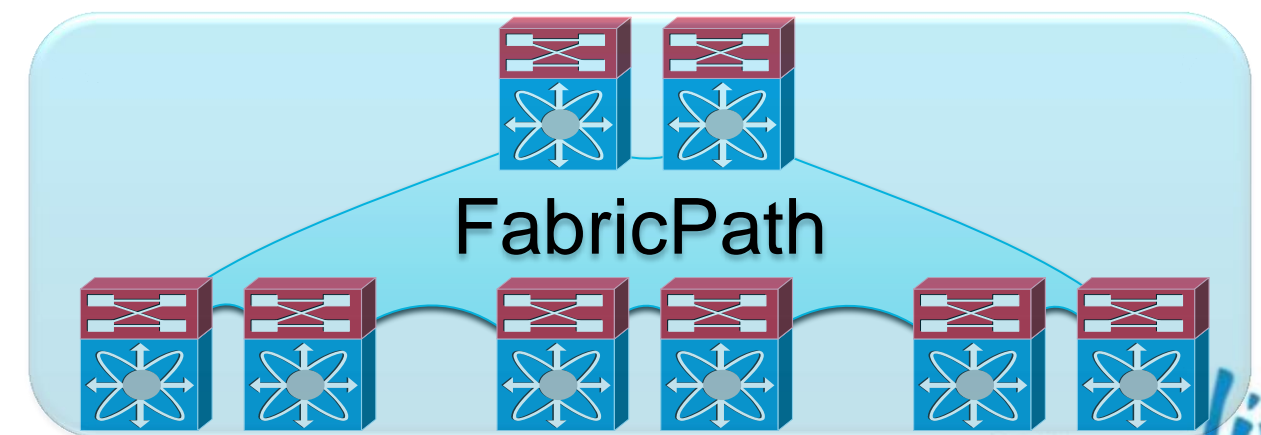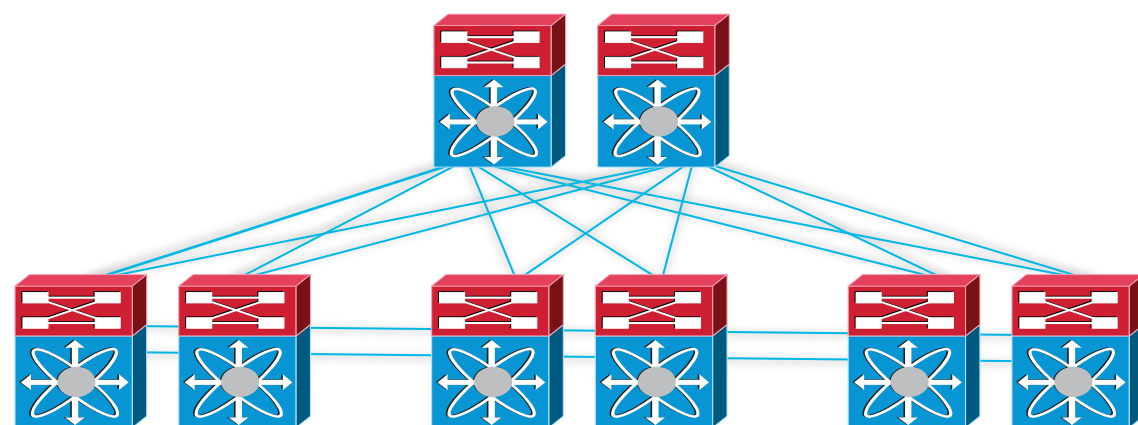
FabricPath

FabricPath Terminology

Interface connected to another FabricPath device
Sends/receives traffic with FabricPath header
Does not perform MAC learning, No STP
Exchanges topology info through L2 ISIS adjacency
Forwarding based on 'Switch ID Table'

FP Core Ports

Spine Switch

S10   S20   S30   S40

FabricPath (FP)

Ethernet frames transmitted on a Cisco FP CORE port always carry an IEEE 802.1Q tag, and as such can be conceptually considered a trunk port.

S100   S200   S300

Leaf Switch

1/1   1/2

Classical Ethernet (CE)

A   B

- The association between Mac Address and Switch-id is maintained at the leaf
- Traffic is encapsulated with in fabric

CE Edge Ports

Interface connected to traditional network device
Sends/receives traffic in standard 802.3 Ethernet frame format
Participates in STP domain
Forwarding based on MAC table
CE Edge port can be switchport access or trunk

Cisco live!

FabricPath
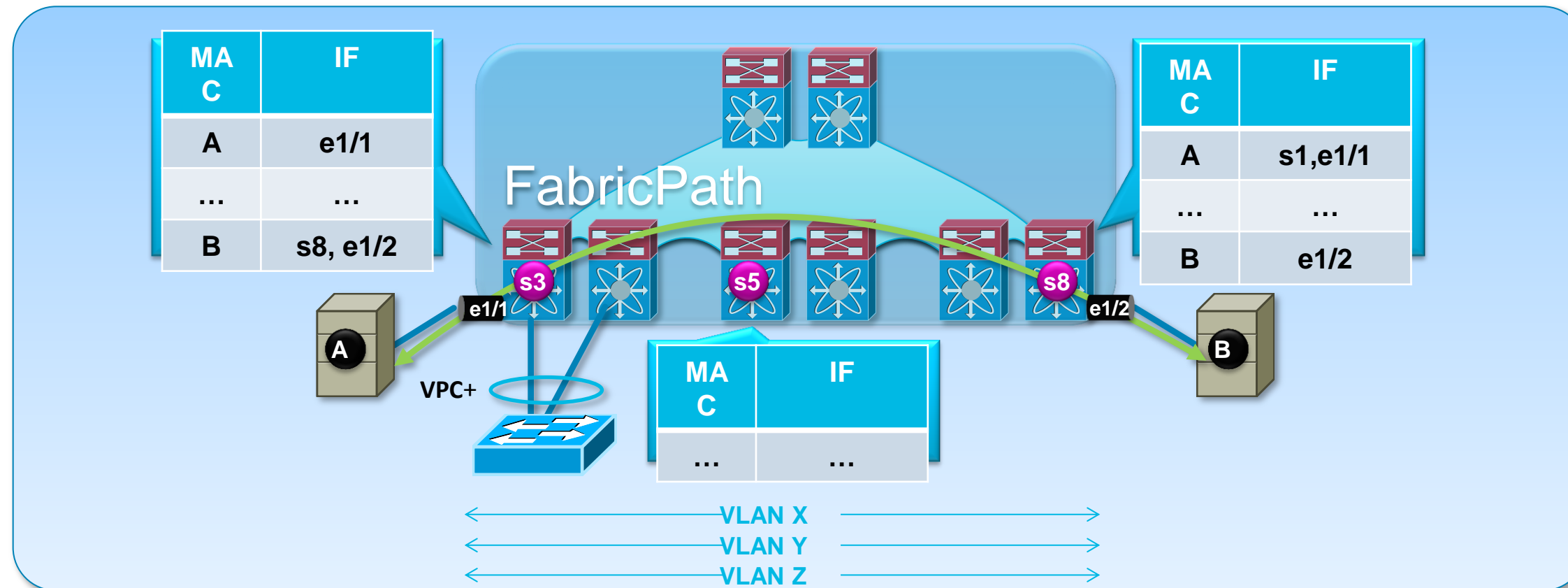## Conversational Learning & VPC+

- Per-port MAC address table only needs to learn the peers that are reached across the fabric

- **A virtually unlimited number of hosts can be attached to the fabric**

- Allows extending VLANs with no limitation (no risks of loop)

- Devices can be attached active/active (VPC+) to the fabric using IEEE standard port channels and without resorting to STP

FabricPath switch-id

- Configure switch-ID manually for all switches in the network

  fabricpath switch-id 1

- Make sure switch-ID (as well as vPC+ emulated switch-ID) is unique in the whole FP fabric

- Suggested numbering scheme for switch-ID

  SPINE switch : 2 digit switch-ID

  LEAF switch : 3 digit switch-ID

  vPC+ system : 4 digit switch-ID

- Enable FP mode VLAN in all FP fabric switches

```
Nexus7K(config)# vlan 10-400
Nexus7K(config-vpc-domain)# mode fabricpath
```

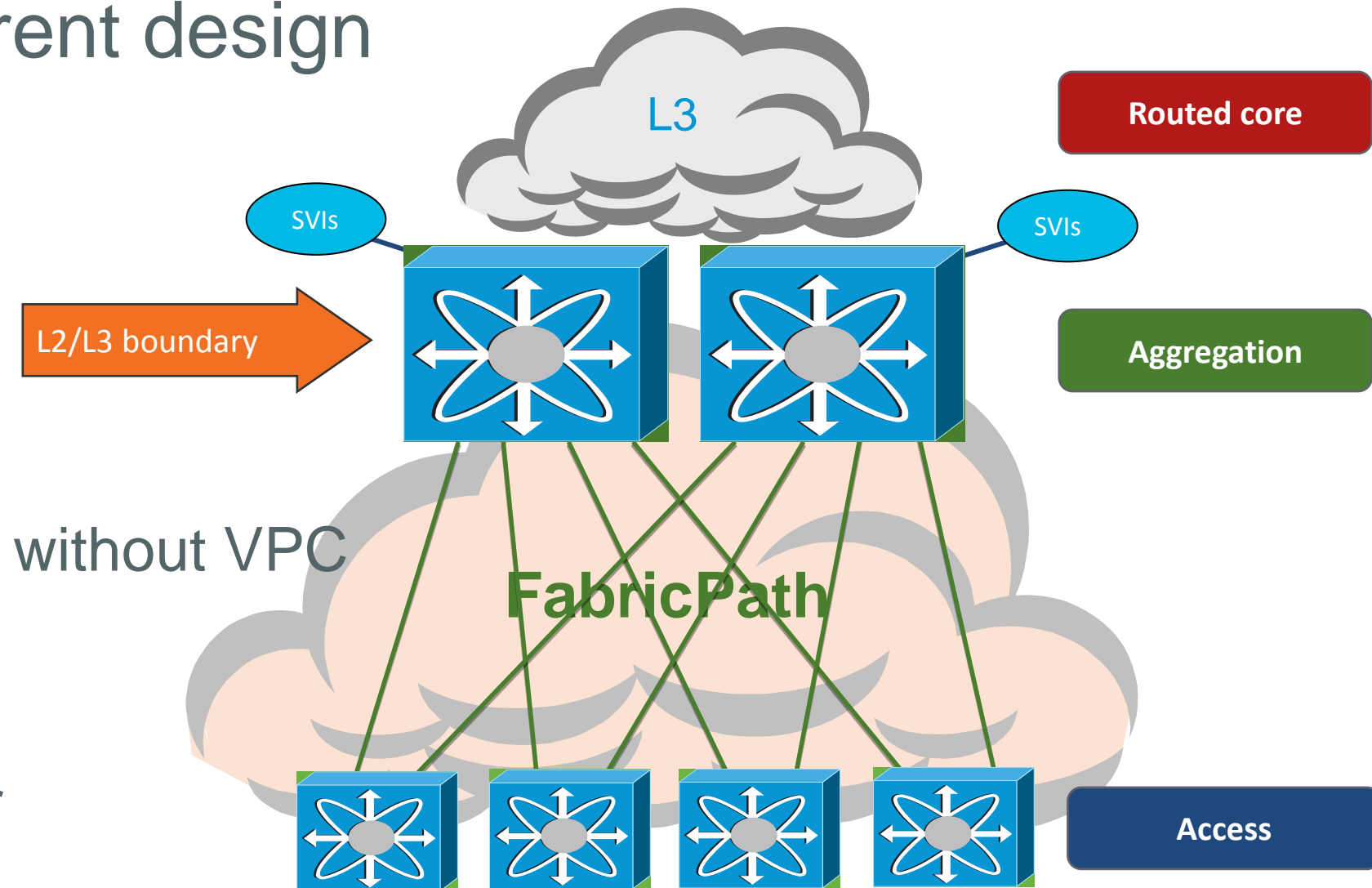- VLAN pruning must be performed manually on CE Edge port

```
Nexus7K(config)# int eth1/10
Nexus7K(config-if)# switchport mode trunk
Nexus7K(config-if) # switchport trunk allowed vlan 100-1000
```

Routing At Aggregation

- Evolutionary extension of current design practices

- Design benefits:

  Simplified configuration

  Removal of STP

  Traffic distribution over all uplinks without VPC port-channels

  Active/active gateways

  "VLAN anywhere" at access layer

  Topological flexibility

- Scalability considerations

Today: 16K unique host MACs across all routed VLANs

L3

SVIs

SVIs

L2/L3 boundary

FabricPath

Routed core

Aggregation

Access

Routing At Aggregation
Option to Scale-out the Spine Layer

VLAN
100-200

VLAN
300-400

VLAN 100-400

VLAN 100-400

FabricPath

FabricPath
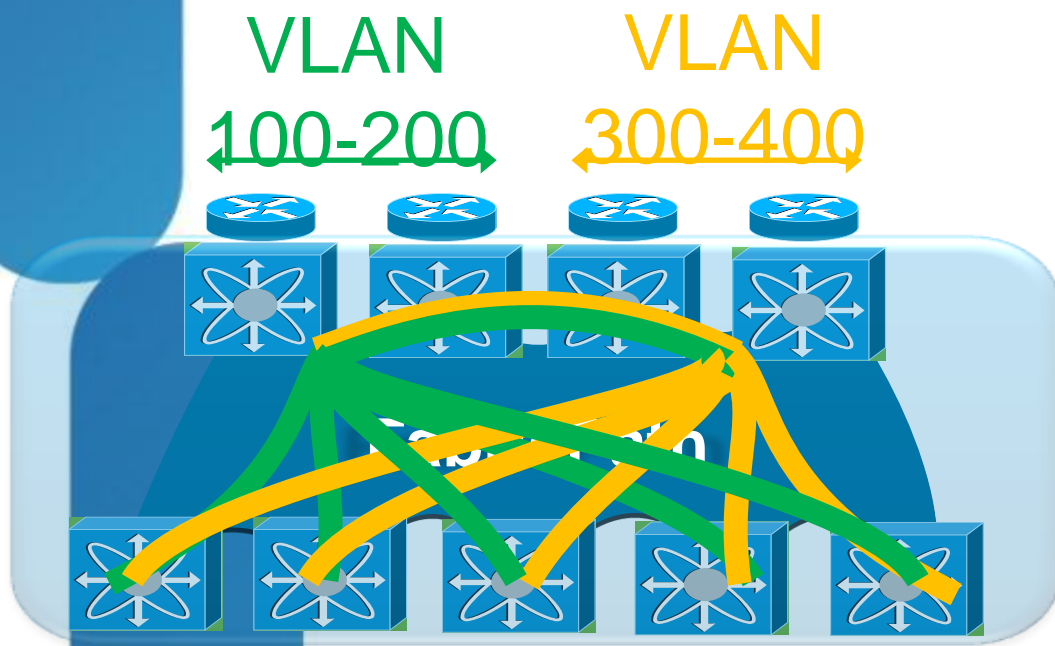
Split VLANs

GLBP

Anycast HSRP

- Some polarisation
- Inter-VLAN traffic can be suboptimal
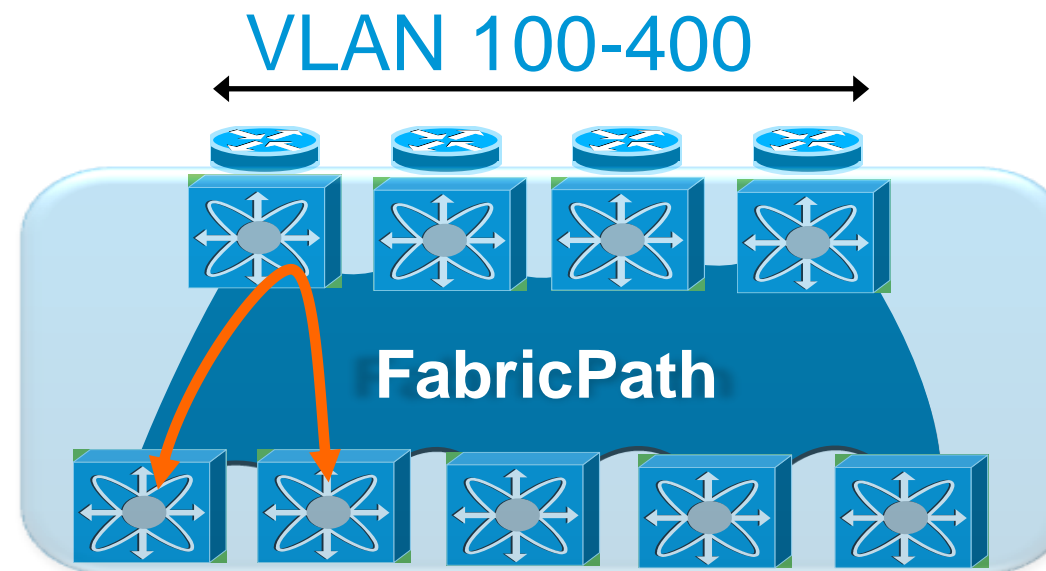
- Host is pinned to a single gateway
- Less granular load balancing
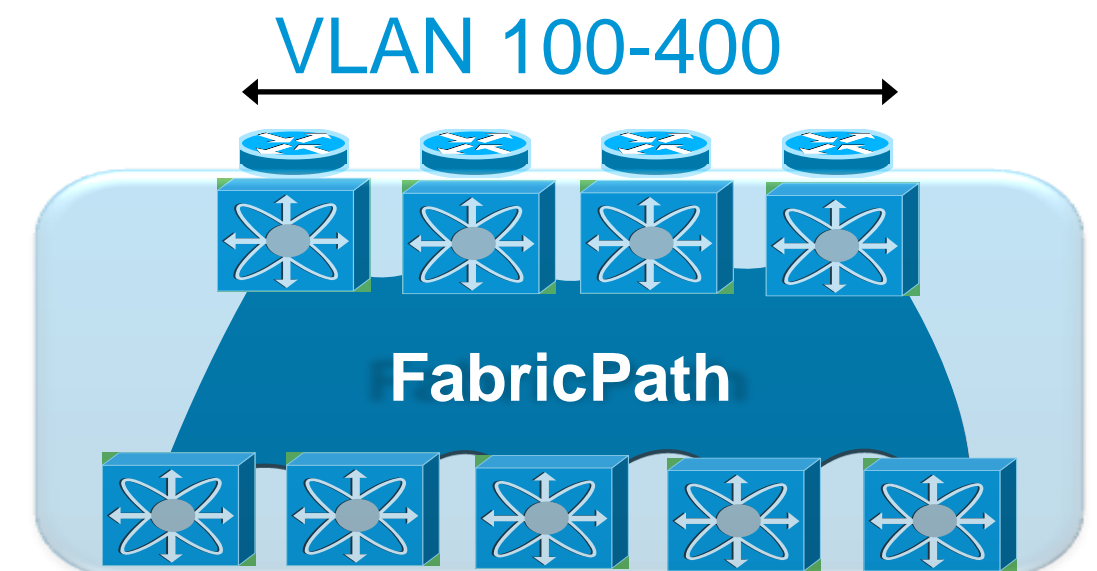
- All active
- Available in NX-OS 6.2 release

Cisco live!

Centralised Routing

Removing Routing from the FP Spine Layer

Centralised Routing Design Alternate View

L3

L2/L3 boundary

FabricPath

Layer 3 services leaf switches

FabricPath spine

Server access leaf switches

=

FabricPath spine

FabricPath

Layer 3 services border leaf switches

L2/L3 boundary

L3

Run VPC+ for active/active HSRP

All VLANs available at all leaf switches

FHRP between L3 services switches for FHRP

Cisco live!
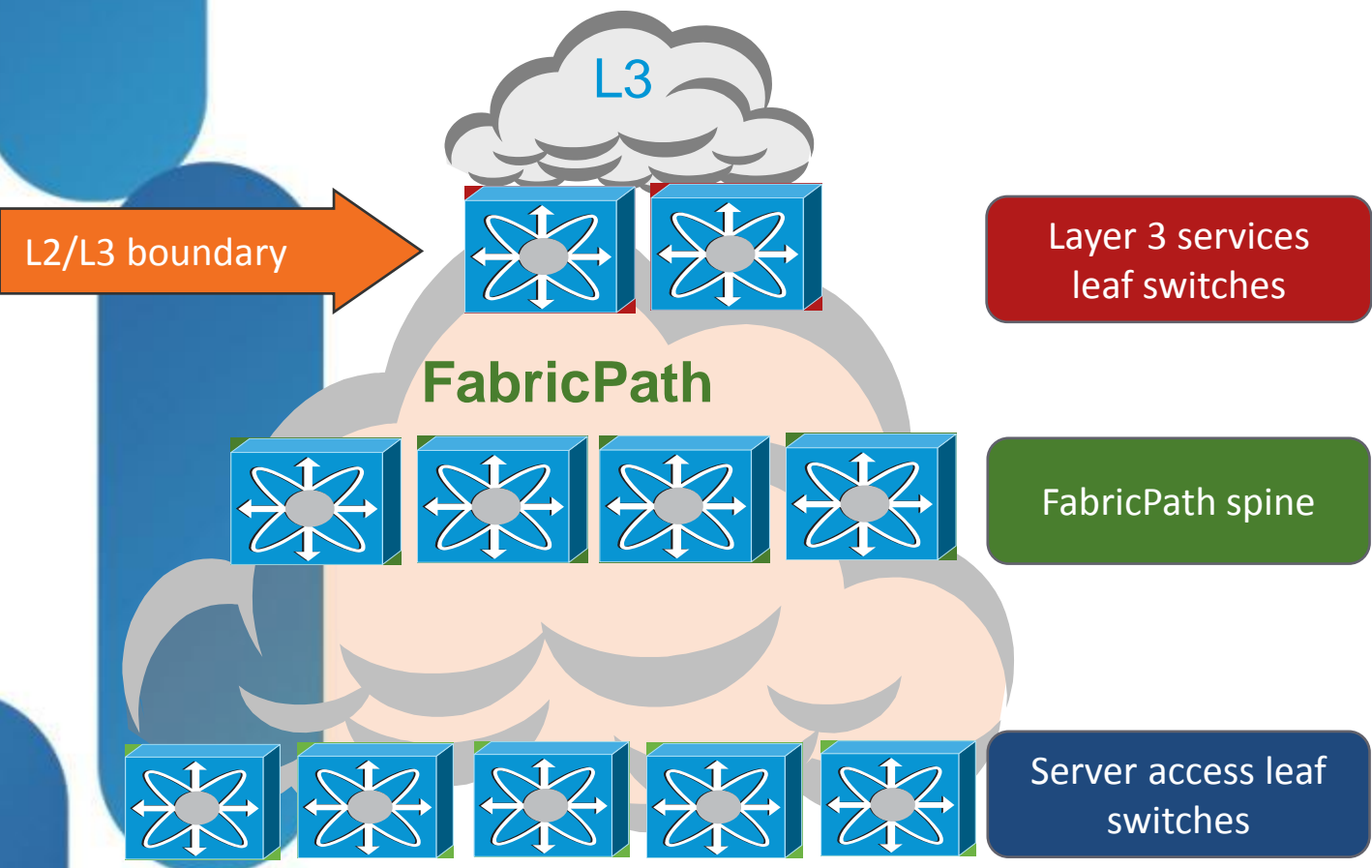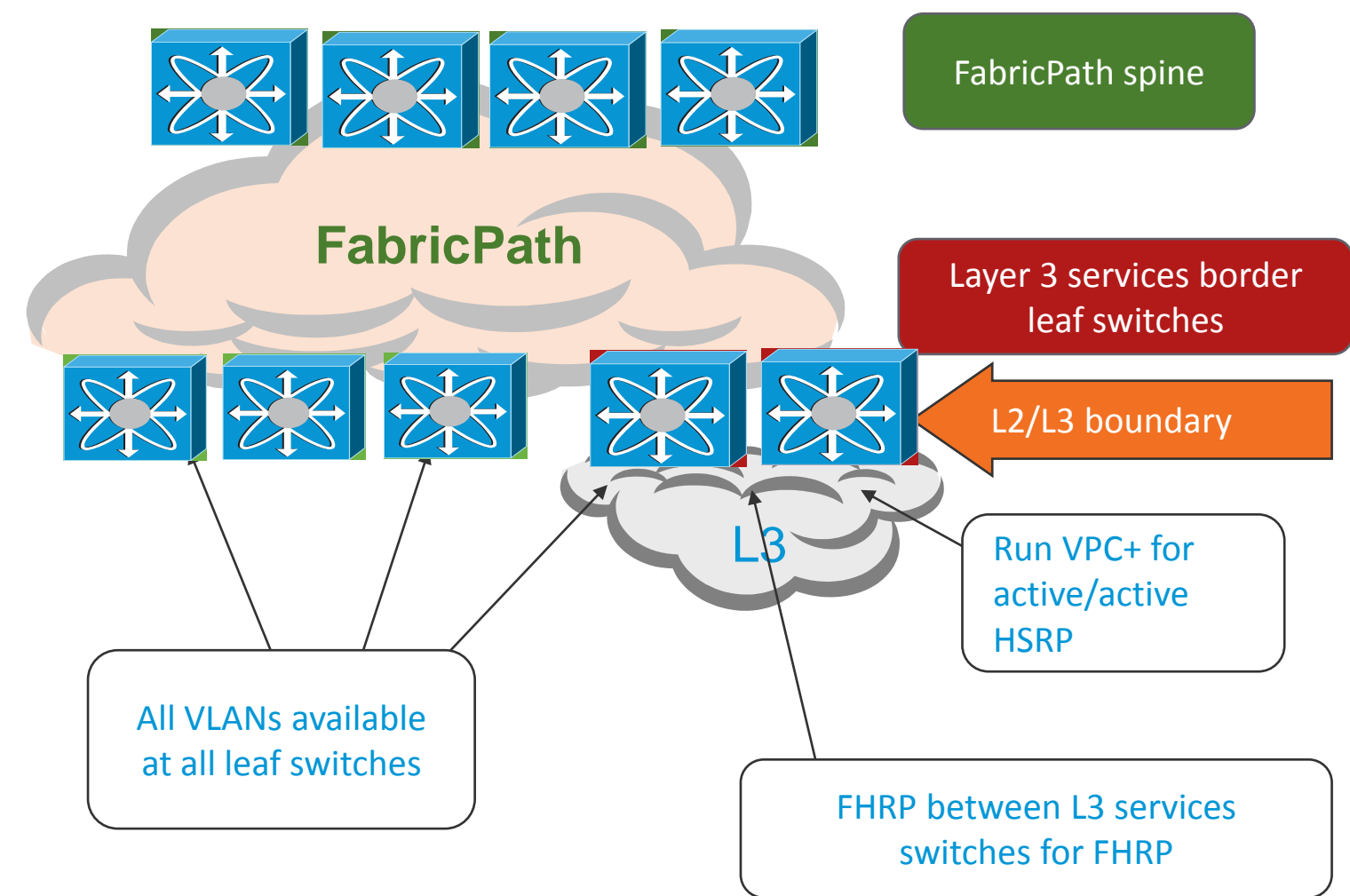
Centralised Routing

Key Design Highlights

- Traditional "aggregation" layer becomes pure FabricPath spine

  - Provides uniform any-to-any connectivity between leaf switches

  - Only FabricPath bridging occurs in spine

- FabricPath leaf switches, connecting to spine, have specific "personality"

  - Most leaf switches provide server connectivity, like traditional access switches in "Routing at Aggregation" designs

  - Two or more leaf switches provide L2/L3 boundary, inter-VLAN routing and
  - North ↔ South routing (Border Leaves)

  - Other (or same) leaf switches provide access to L4-7 services or have L4-7 services personality (future)

- Same MAC scalability considerations of routing at aggregation design

  - 16K unique host MACs today, 128K MACs with 6.2 release and Nexus 6K (at FCS)

Overlay Transport Virtualisation
"MAC in IP" supporting Layer 2 VPNs Over any Transport

Virtual/
Private Cloud
VLAN 1

Physical

HFT/HPC

Cisco
Unified Fabric

Cisco® OTV (L2 Extension)

DC2

DC3

VLAN 1

NAS

SAN

Storage
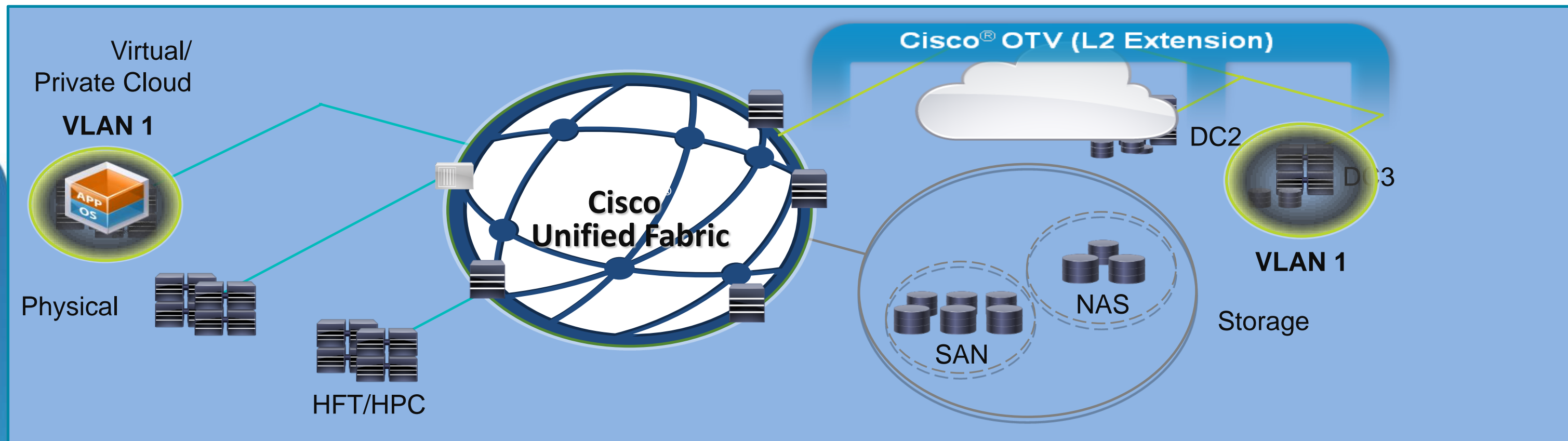
USE CASES
- Inter and intra DC connectivity— across L3
- Use all data centre capacity
- Back up data centre, rapid recovery
- Reduced data centre maintenance outage

RESULTING IN
- Scalability across multiple data centres
- Seamless overlay—no network redesign required
- Single-touch site configuration
- High resiliency
- Maximised bandwidth

Cisco Public

Cisco live!

OTV at a Glance

- Ethernet traffic between sites is encapsulated in IP: "MAC in IP"

- Dynamic encapsulation based on MAC routing table

- No Pseudo-Wire or Tunnel state maintained

MAC1 → MAC2

IP A → IP B   MAC1 → MAC2

MAC1 → MAC2

Encap

Decap

MAC	IF
MAC1	Eth1
MAC2	IP B
MAC3	IP B

OTV

OTV

IP A

OTV

IP B

Communication between
MAC1 (site 1) and MAC2 (site 2)

Server 1
MAC 1

Server 2
MAC 2
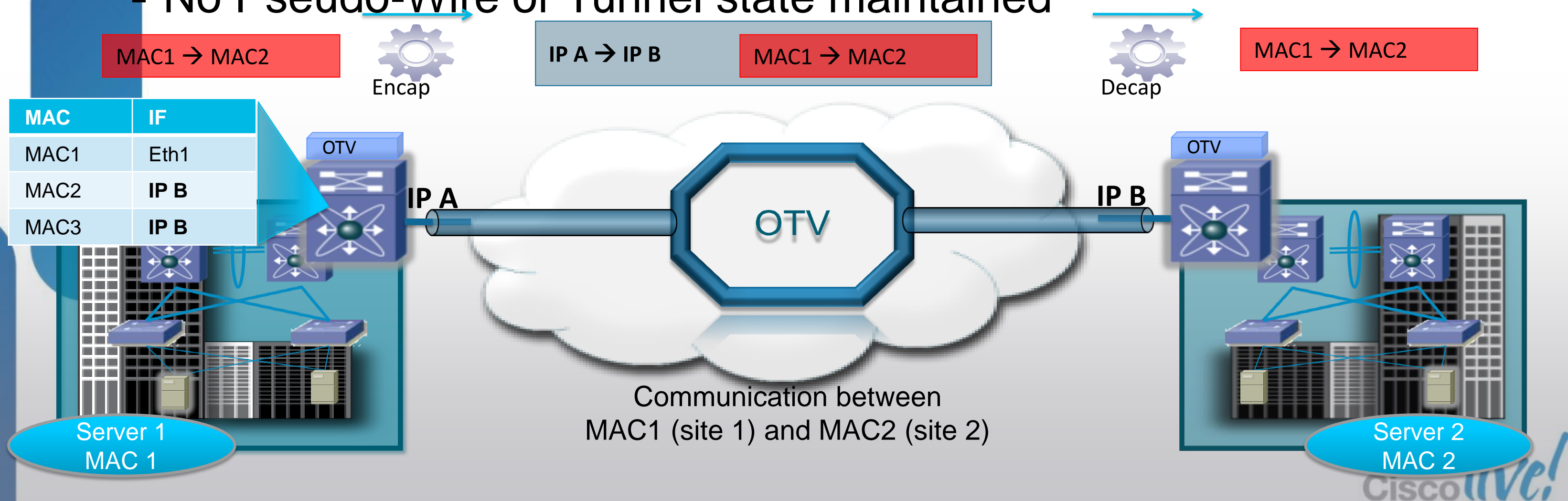
OTV Terminology

- Edge Device (ED): connects the site to the (WAN/MAN) core; responsible for performing all the OTV functions

- Authoritative Edge Device (AED): Elected ED that performs traffic forwarding for a set of VLAN

- Internal Interfaces: interfaces of the ED that face the site

- Join interface: interface of the ED that faces the core

- Overlay Interface: logical multi-access multicast-capable interface. It encapsulates Layer 2 frames in IP unicast or multicast headers

OTV

Overlay Interface

OTV

Core

L2 L3

Internal Interfaces

Join Interface

Cisco live!
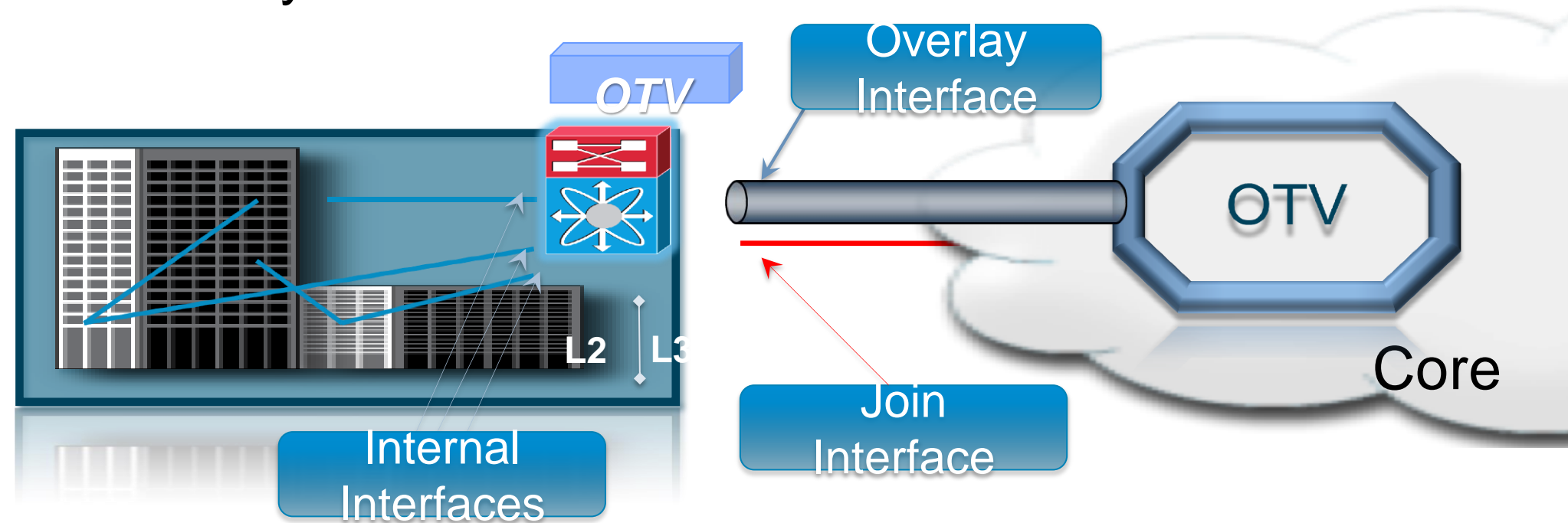
OTV & SVI

- On Nexus 7000 a given VLAN can *either* be associated with an SVI *or* extended using OTV

  - This would theoretically require a *dual-system solution*

  - The VDC feature allows to deploy a dual-vdc solution

- OTV VDC as an appliance

- Single L2 internal interface and single Layer 3 Join Interface

OTV VDC   **Default VDC**   **Default VDC**   **OTV VDC**

N7K-1   **N7K-2**

—— L3 Link
—— L2 Link
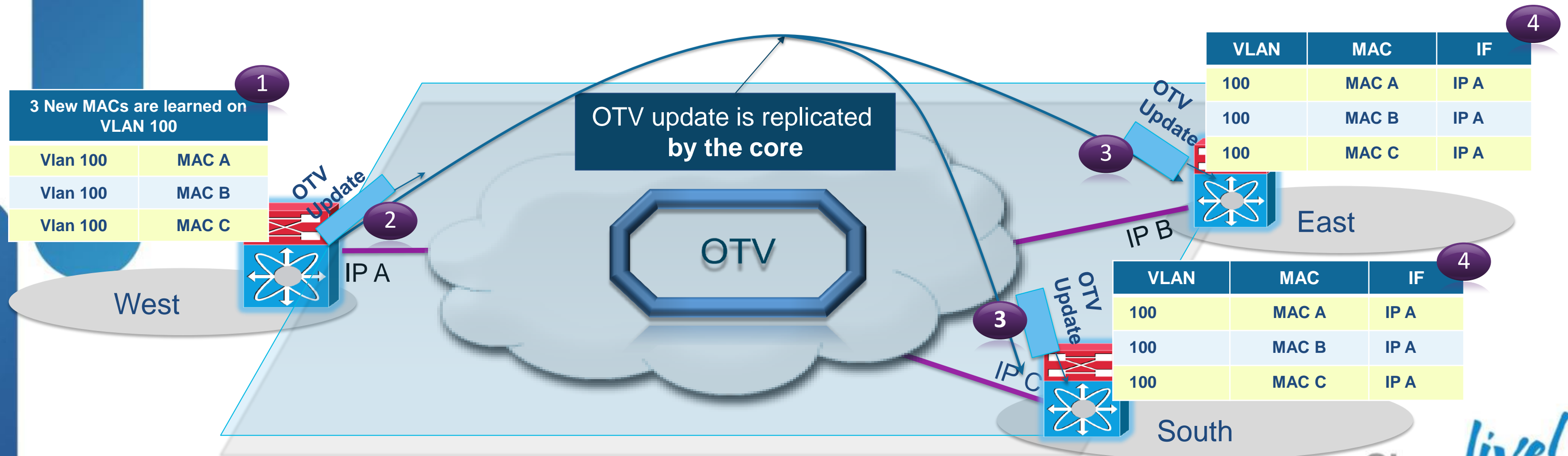
Physical View

   Cisco Public
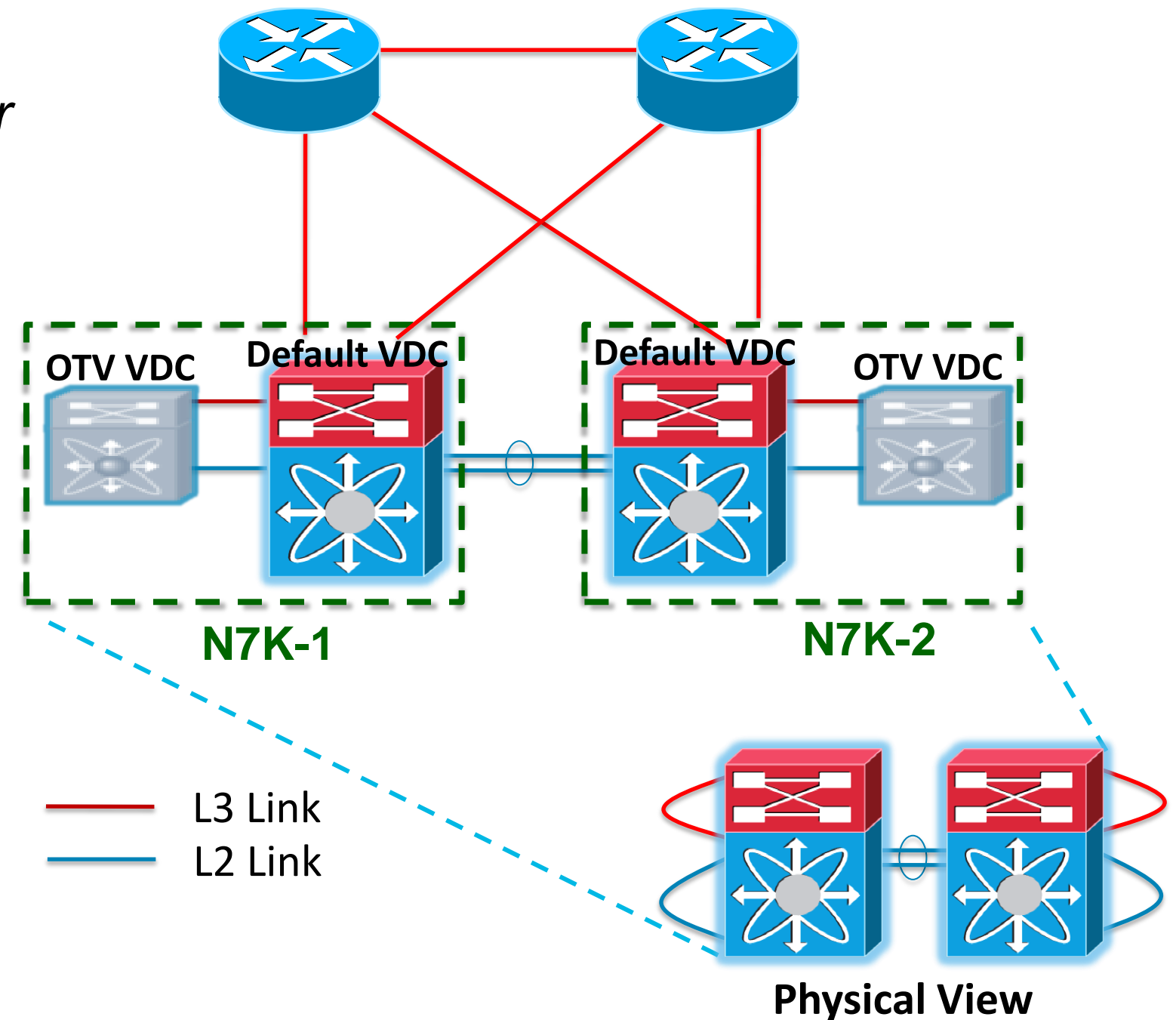
OTV Control Plane

MAC Address Advertisements (Multicast-Enabled Transport)

- When an Edge Device learns a new MAC address it advertises it together with its associated VLAN IDs and the IP address of the join-interface

- A single OTV update can contain multiple MACs from different VLANs

- With a multicast-enabled transport a **single update reaches all neighbours.**

1

3 New MACs are learned on VLAN 100	
Vlan 100	MAC A
Vlan 100	MAC B
Vlan 100	MAC C

OTV Update

2

IP A

West

OTV update is replicated **by the core**

OTV

3

OTV Update

IP B

East

4

VLAN	MAC	IF
100	MAC A	IP A
100	MAC B	IP A
100	MAC C	IP A

3

OTV Update

IP C

South

4

VLAN	MAC	IF
100	MAC A	IP A
100	MAC B	IP A
100	MAC C	IP A

Cisco live!

OTV Configuration

OTV over a Multicast Transport
Minimal configuration required to get OTV up and running

```
feature otv
otv site-identifier 0x1*
otv site-vlan 99
interface Overlay100
  otv join-interface e1/1
  otv control-group 239.1.1.1
  otv data-group 232.192.1.0/24
  otv extend-vlan 100-150
```

```
feature otv
otv site-identifier 0x3*
otv site-vlan 99
interface Overlay100
  otv join-interface e1/1.10
  otv control-group 239.1.1.1
  otv data-group 232.192.1.0/24
  otv extend-vlan 100-150
```

```
feature otv
otv site-identifier 0x2*
otv site-vlan 99
interface Overlay100
  otv join-interface Po16
  otv control-group 239.1.1.1
  otv data-group 232.192.1.0/24
  otv extend-vlan 100-150
```
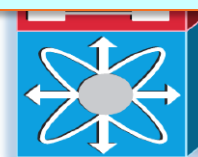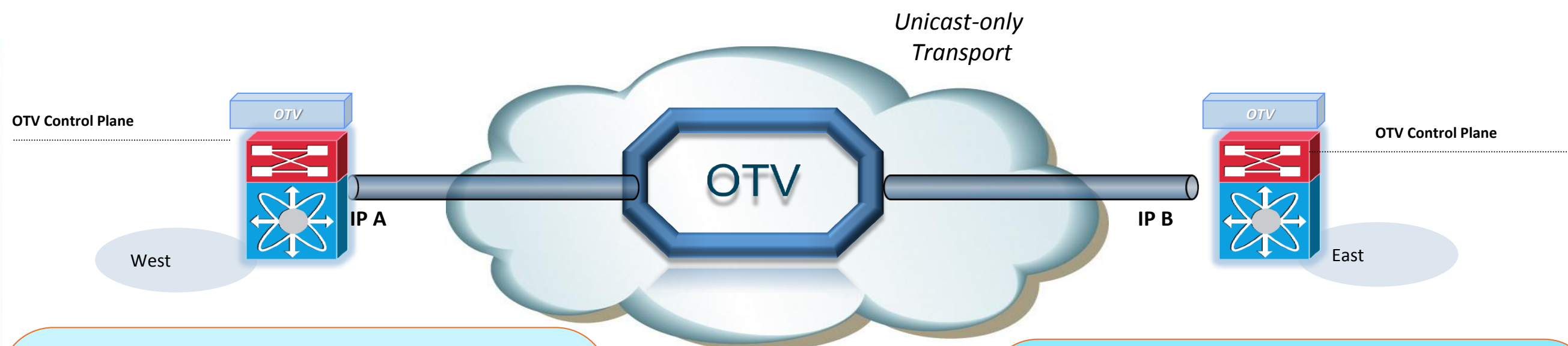
West

IP A

South

IP B

East

*Introduced from release 5.2

OTV Control Plane

Neighbor Discovery (Unicast-Only Transport)

- Ideal for connecting a small number of sites

- With a higher number of sites a multicast transport is the best choice

Unicast-only Transport

OTV Control Plane — OTV — OTV — IP A — West

OTV — OTV Control Plane — IP B — East

Mechanism

- Edge Devices (EDs) register with an *"Adjacency Server"* ED
- EDs receive a full list of Neighbors (oNL) from the AS
- OTV hellos and updates are encapsulated in IP and **_unicast_** to each neighbor

End Result

Neighbor Discovery is automated by the *"Adjacency Server"*

All signaling must be replicated for each neighbor

Data traffic must also be replicated at the head-end

Cisco live!

OTV Configuration
OTV over Unicast Transport

Primary Adjacency Server

```
feature otv
otv site-identifier 0x1
otv site-vlan 99
interface Overlay100
  otv join-interface e1/1
  otv adjacency-server unicast-only
  otv extend-vlan 100-150
```

Secondary Adjacency Server

```
feature otv
otv site-identifier 0x2
otv site-vlan 99
interface Overlay100
  otv join-interface e1/1.10
  otv adjacency-server unicast-only
  otv use-adjacency-server 10.1.1.1 unicast-only
  otv extend-vlan 100-150
```
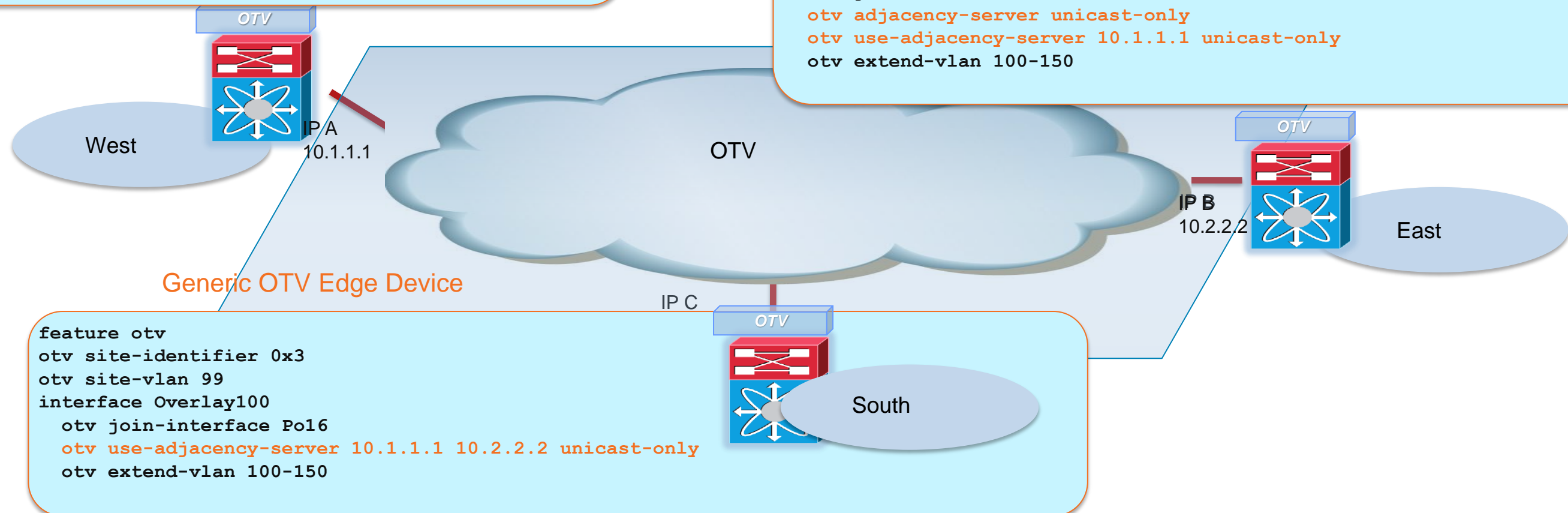
OTV

West

IP A
10.1.1.1

OTV

IP B
10.2.2.2

East

Generic OTV Edge Device

IP C

```
feature otv
otv site-identifier 0x3
otv site-vlan 99
interface Overlay100
  otv join-interface Po16
  otv use-adjacency-server 10.1.1.1 10.2.2.2 unicast-only
  otv extend-vlan 100-150
```
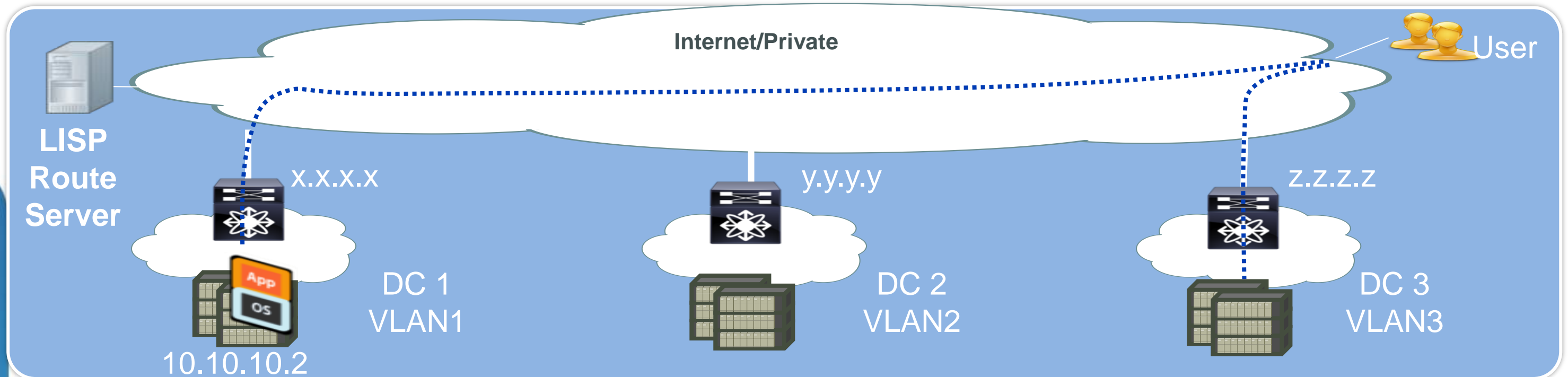
South

*Introduced from release 5.2

Nexus 7000 LISP

Location ID/Separation Protocol – Global IP Address Portability

Internet/Private

User

LISP Route Server

x.x.x.x

y.y.y.y

z.z.z.z

DC 1
VLAN1

DC 2
VLAN2

DC 3
VLAN3

10.10.10.2

Features

- IP address portability across subnets
- Auto detection and re-route of traffic/session
- Highly scalable technology

Benefits

- Seamless workload mobility between data centres and cloud
- Direct Path (no triangulation), connections maintained during move
- No routing re-convergence, no DNS updates required
- Transparent to the hosts and users

Nexus 7000 – Software Features
FCoE

Nexus 7000 FCoE

SAN Core

FC

FCoE

Hosts

Access
Nexus 7000

Common Core
Nexus 7000

FCoE

FCoE

FCoE

FCoE

Storage
Targets
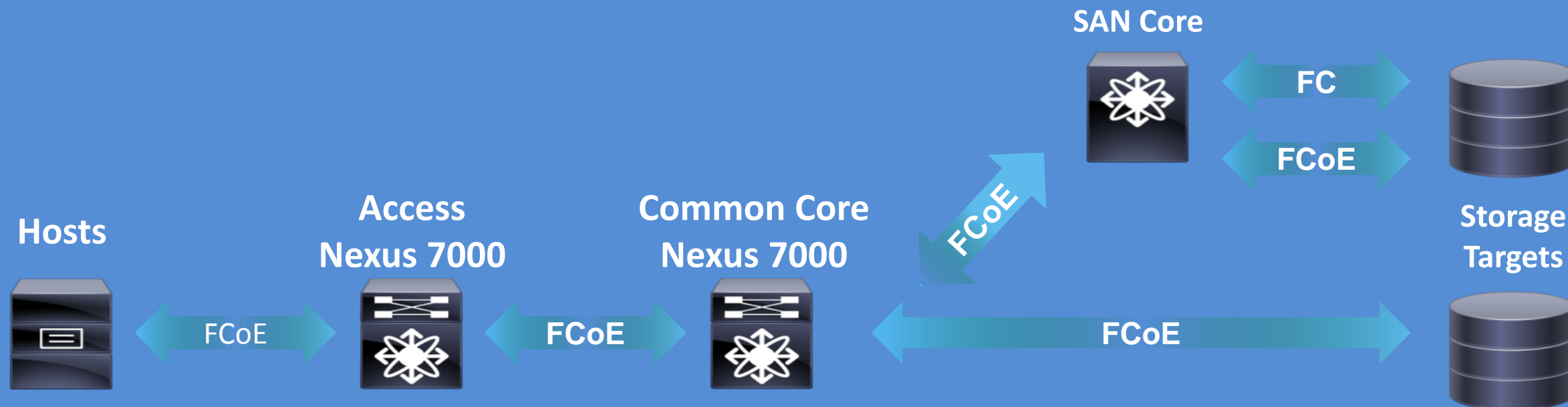
Features

- Industry's highest performance **Director-Class SAN platform**

- **Lossless Ethernet** (DCB)

- **Multi-hop FCoE support:** Spans Nexus 7000, Nexus 5000, and MDS 9500

Benefits

- Wire once flexibility over single Ethernet Fabric

- Reduce network Sprawl— switches, cables, adapters, etc.

- Up to 45% access layer CapEx savings

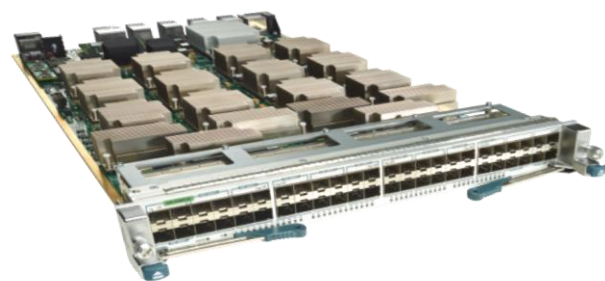- Seamlessly integrate converged networks with existing MDS FC SANs

FCoE ON F2 MODULE

High Performance Director Class Convergence

- Highest performance storage platform in the industry 15-Tbps

- Highest density of line-rate 10G FCoE ports - Access, Aggregation, or Core

- Multi-protocol: FCoE, iSCSI, and NAS

F2 module: N7K-F248XP-25

REQUIRES
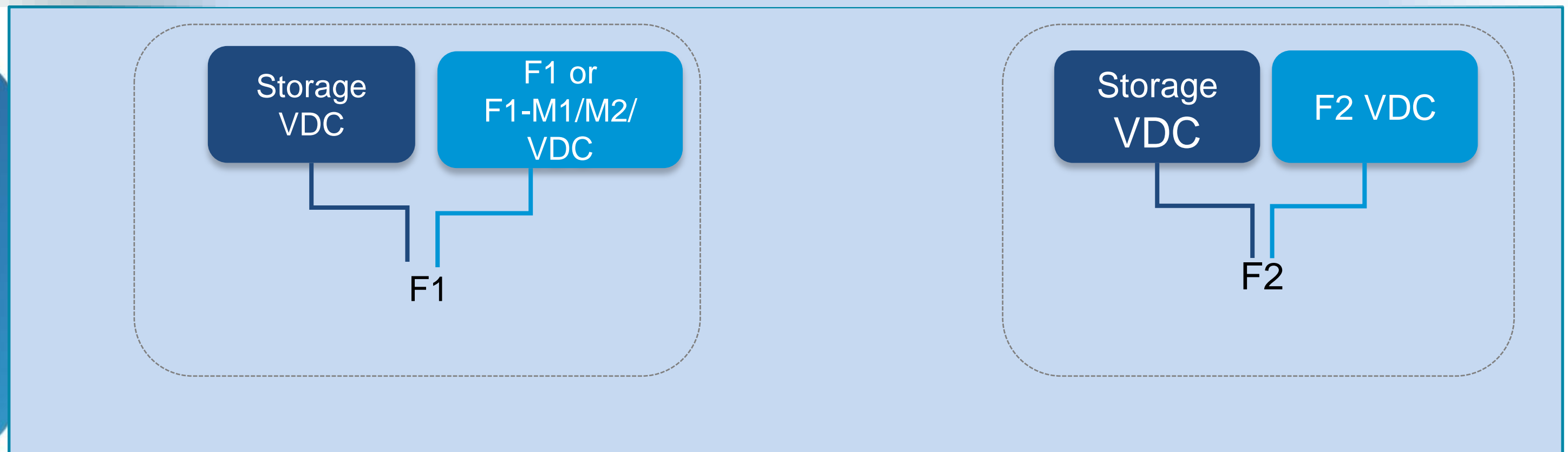
- SUP2/2E
- Fabric 2 Modules for full bandwidth

Cisco*live!*

Storage VDC Supported Options

Logical representation

Dedicated VDC – Shared interfaces

Notes

- F1 and F2 cannot co-exist in the same VDC
- Only one storage VDC per chassis

Software Licensing

Features installed by individual licenses or enabling the license grace period (120 days)

- Grace period not recommended

- Installation is non-disruptive to features already running under the grace period

- Backup the license after license is installed

- System generates periodic Syslog, SNMP or Call home messages

Feature License	Features
Enterprise LAN	OSPF, EIGRP, BGP, ...
Advanced LAN	CTS, VDC
MPLS	MPLS, L3VPN
VDC	4 VDC & 1 Admin VDC
Scalable Feature	M1-XL TCAM
Transport Services	OTV, LISP
Enhanced L2 Package	FabricPath
FCoE	FCoE
Storage	VSAN routing and access list

```
Nexu7K# show license usage
Feature                          Ins  Lic    Status Expiry Date Comments
                                      Count

---------------------------------------------------------------------------
LAN_ADVANCED_SERVICES_PKG        Yes   -     In use Never         -
LAN_ENTERPRISE_SERVICES_PKG      No    -     In use              Grace 119D 22H
```

Software Upgrade

- Synchronise the kickstart image with the system image

- Utilise cold start upgrade procedure for non-production devices

```
Nexus7K(config)# boot system bootflash:<system-image>
Nexus7K(config)# boot kickstart bootflash:<kickstart-image>
Nexus7K# copy run startup-config
Nexus7K# reload
```

- Utilise"install all" to perform ISSU with zero service interruption

   Issue "show install all impact" to determine upgrade impact

```
Nexus7K# install all kickstart bootflash:<kickstart-image> system bootflash:<system-image>
```

- Refer to release notes and installation guide

- Avoid disruption to the system during ISSU upgrade (STP topology change, module removal, power interruption, etc)

EPLD Upgrade

- EPLD upgrade is used to enhance HW functionality or to resolve known issues

- EPLD upgrade is an independent process from software upgrade and not dependent on NX-OS

- EPLD upgrade is typically not required

- Performed on all Field Replaceable Modules

- In redundant configuration, requires reload of IO modules

```
Nexus7K# sh ver <type> <#> epld
Nexus7K# sh ver mod 3 epld

EPLD Device                Version
----------------------------------
Power Manager                4.008
IO                           1.016
Forwarding Engine            1.006
FE Bridge(1)               186.006
FE Bridge(2)               186.006
Linksec Engine(1)            2.006
---deleted---
Linksec Engine(8)            2.006
```

Cisco live!

EPLD Upgrade – Best Practices

- Upgrade to the latest EPLD image prior to bringing hardware into production environment (staging HW, replacement HW, etc)

  Only use "Install all EPLD" on non-production systems

  ```
  Nexus7K# install all epld bootflash:<EPLD_image_name>
  ```

- When performing supervisor EPLD upgrade for a system with dual-sup, first upgrade the standby supervisor, then switchover and upgrade previous active supervisor

  Make sure EPLD image is on both supervisor's flash

  ```
  Nexus7K# install module <module> epld bootflash:<EPLD_Image_name>
  ```

- In a redundant system, only EPLD upgrade for I/O modules can disrupt traffic since the module needs to be power-cycled

Hardware Installation Considerations

- Two supervisors for high availability and ISSU

- Two M1/M2 modules in mixed mode chassis (Mx/Fx)

- A minimum of three fabric modules to provide N+1 redundancy

- Use five 2nd generation fabric modules for full performance for newer I/O modules

- Perform chassis / system grounding

- Perform additional diagnostics on staged devices before production

  - Configure complete boot-up diagnostic level (default)

  - Administratively shutdown all ports to run Portloopback test over night

 - Power-cycle after burn-in period to perform boot-up diagnostic

Nexus 7000 Design Examples

Data Centre Design Example 1

3-Tier Data Centre Design Leveraging VDCs

- Large Data Centre utilising 3-Tier DC design
- Nexus 7000s in Core and Aggregation
- 10GE/GE ToR and GE MoR access layer switches
- 40GE/100GE high density Aggregation to Core connections
- Implement vPC / double-sided vPC for redundant active/active server connections
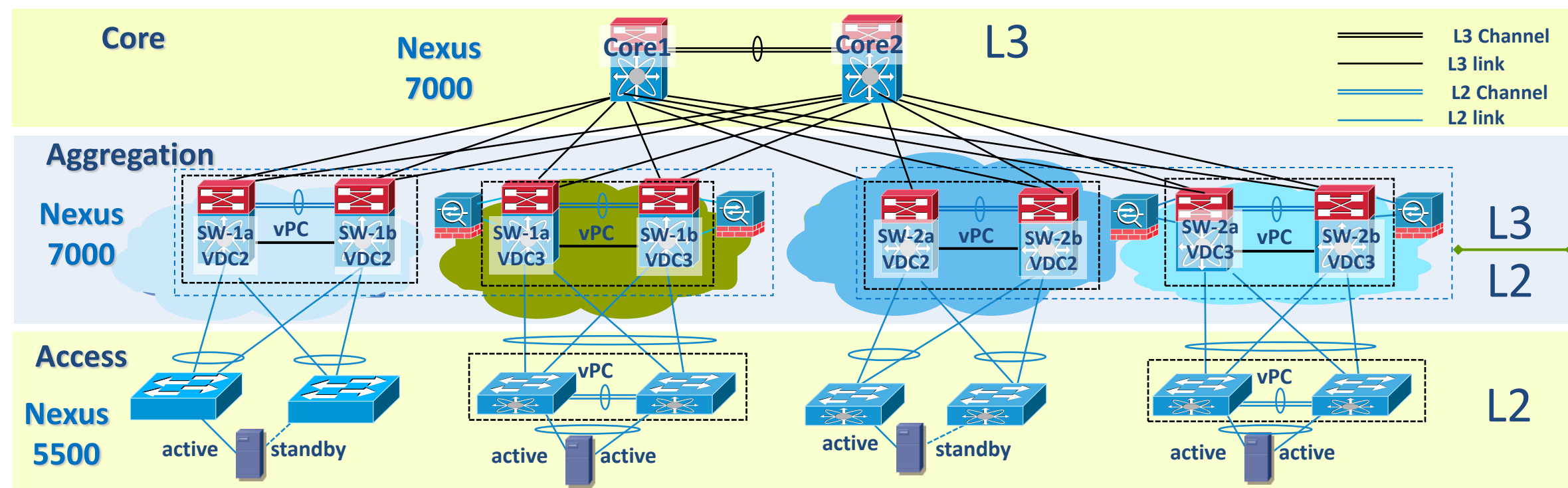
Data Centre Design Example 2

## 3 Tiered DC Design with Nexus 5000 & Nexus 2000

- Nexus 7000 in core and aggregation, N5K / N2K in access layer
- Utilise Nexus 2000 for scaling with Nexus 7000
- Implement vPC / double-sided vPC / Host vPC for redundancy
- Different vPC redundancy models can be utilised to support active/active or active/standby server connections
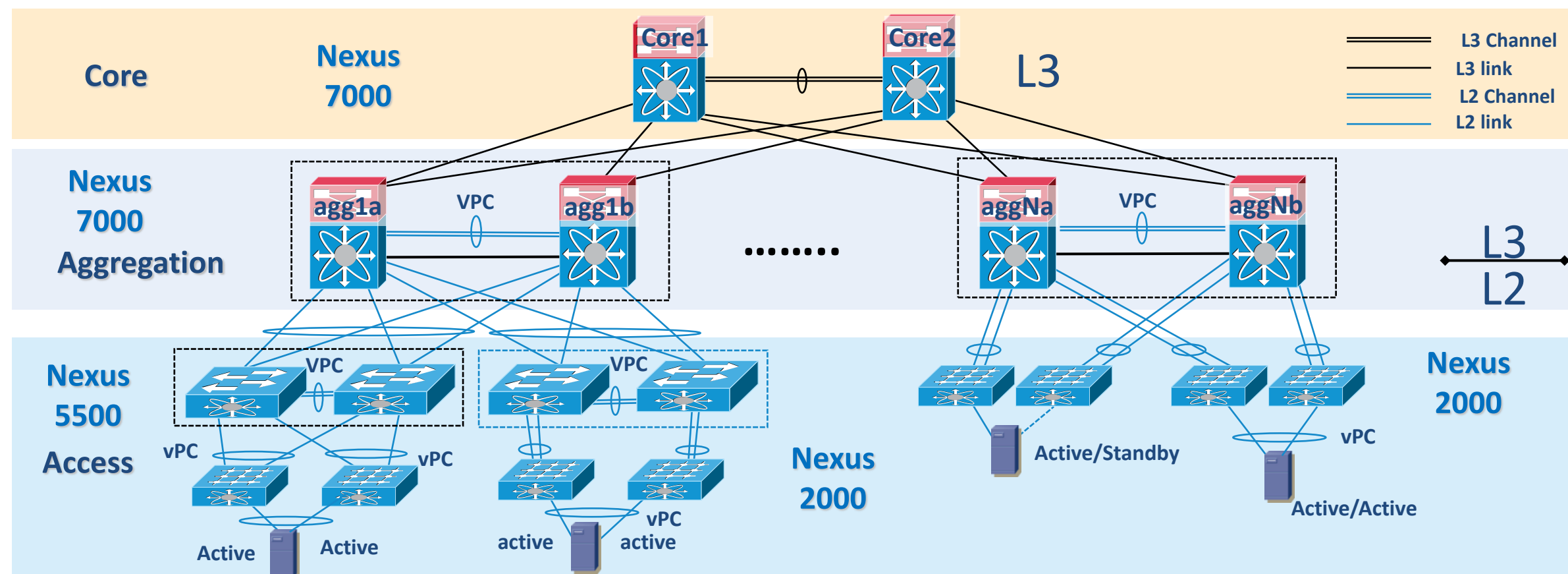
Data Centre Design Example 3

Small DC with Virtualised 3-Tier Design using N7004

- Small Data Centre with a "virtualised" 3-Tier DC design using Nexus 7004
- Utilise VDCs on two devices to create a core and aggregation layer
- GE and 10GE ToR access layer switches
- Implement vPC / double-sided vPC for redundant active/active server connections
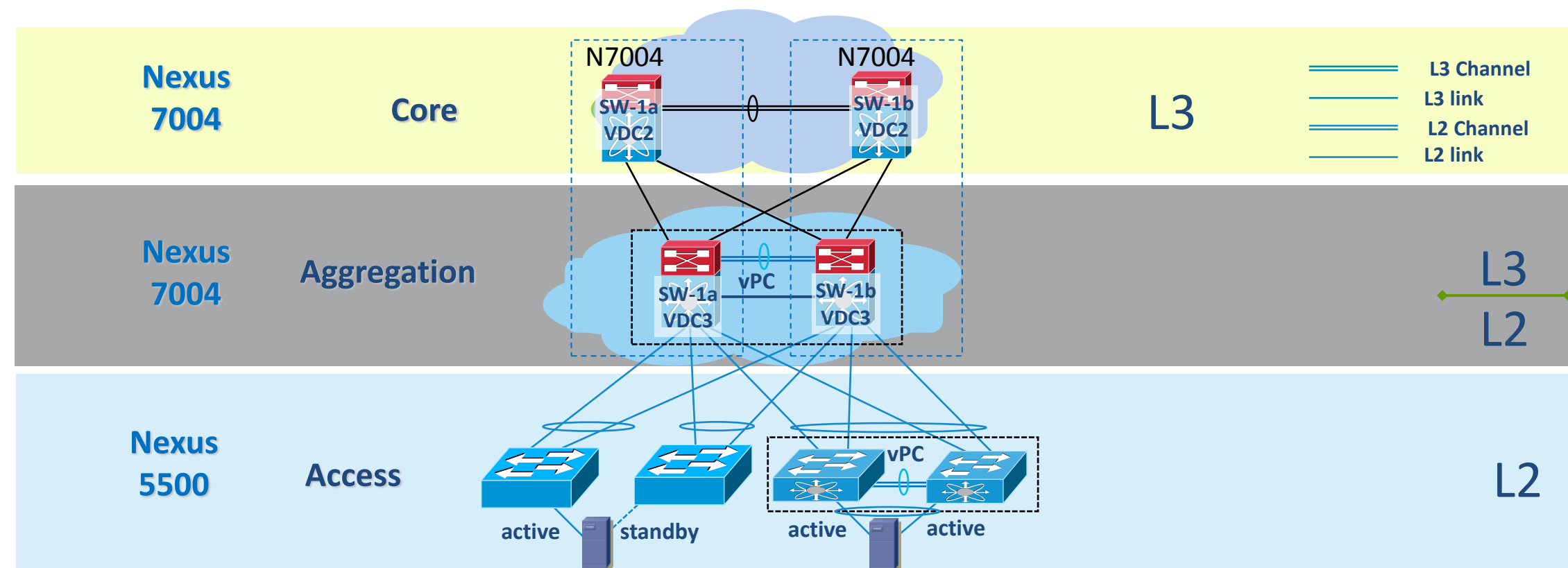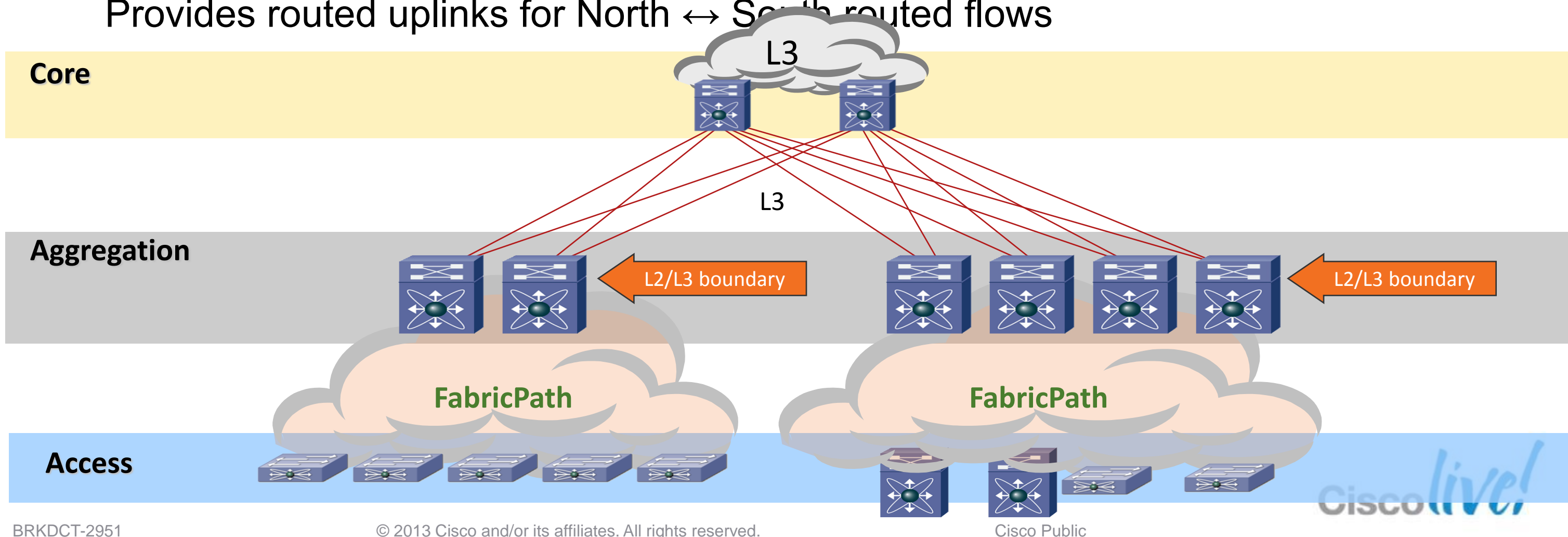
Data Centre Design Example 4

FabricPath Design with Routing at Aggregation

- Fabric Path High-level design options with Routing at Aggregation
- Aggregation serves as FabricPath spine as well as L2/L3 boundary

  Provides FabricPath bridging for East ↔ West intra-VLAN traffic

  Provides SVIs for East ↔ West inter-VLAN routing

 Provides routed uplinks for North ↔ South routed flows

——	Layer 3 Link
——	Layer 2 CE
——	Layer 2 FabricPath

Core

L3

Aggregation

L3

L2/L3 boundary

L2/L3 boundary

FabricPath

FabricPath

Access

Cisco Public

Data Centre Design Example 5

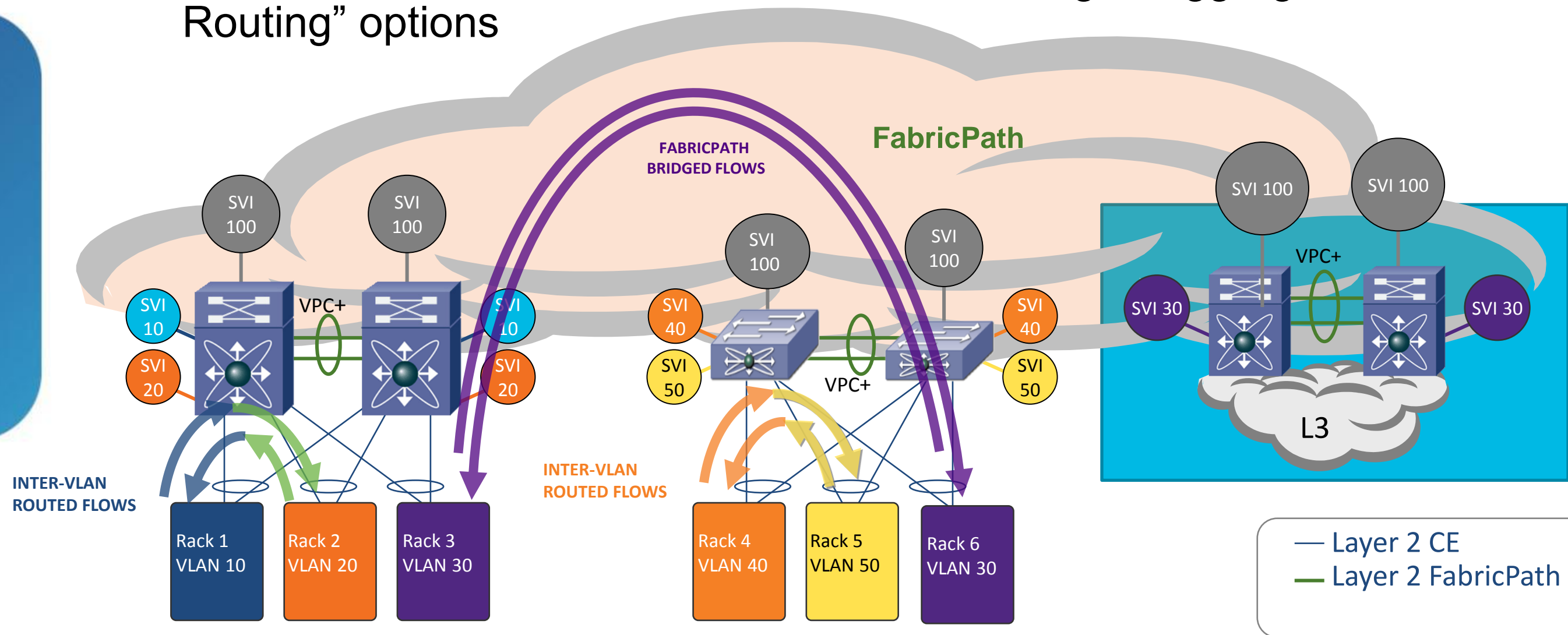## Cisco FabricPath Centralised Routing Design

- Centralised routing at FabricPath Leafs, Aggregation layer no longer does the routing

- Transparent spine Interconnecting the leafs together

- Leaf switches provide server access and some leaf can provide L3 Services

- Very flexible in terms of scaling and enables vlan anywhere

L3

L2/L3 boundary

Layer 3 services
leaf switches

FabricPath spine

FabricPath

Server access leaf
switches
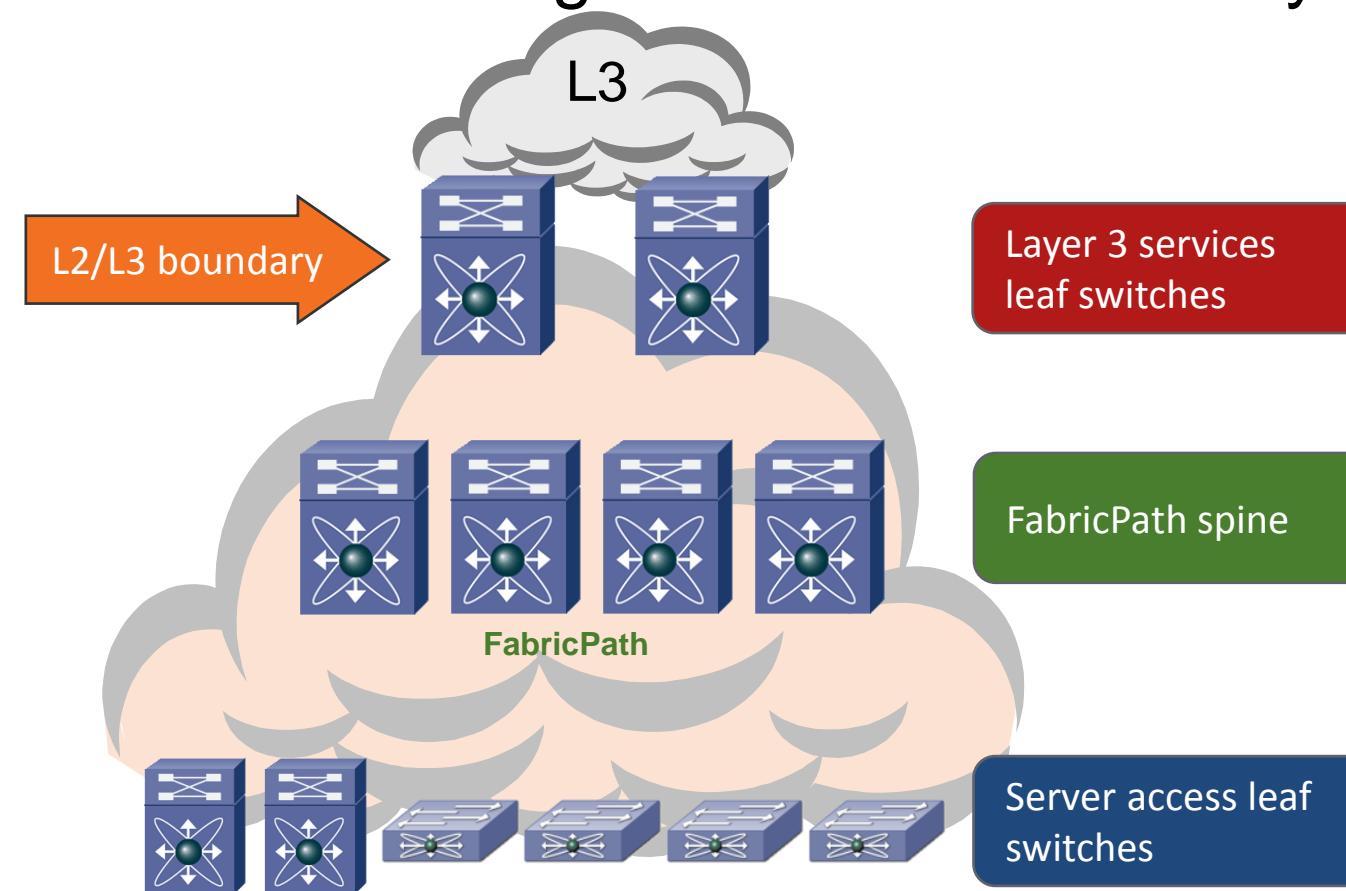
Data Centre Design Example 6

FabricPath Design with Distributed Routing

- Fabric Path High-level design options with Distributed routing

  - For POD-local VLANs, same as "Routing at Aggregation" options

  - For DC-wide and multi-POD VLANs, "Routing at Aggregation" or "Centralised Routing" options

FabricPath

FABRICPATH BRIDGED FLOWS

INTER-VLAN ROUTED FLOWS

INTER-VLAN ROUTED FLOWS

Rack 1 VLAN 10
Rack 2 VLAN 20
Rack 3 VLAN 30
Rack 4 VLAN 40
Rack 5 VLAN 50
Rack 6 VLAN 30

L3

— Layer 2 CE
— Layer 2 FabricPath

Conclusion

- Understand requirements and features available from the products

- Build a topology choosing the leading practices to ensure solution is resilient and scalable

- Pilot and test the solution

- Put it to production

- Fine tune and optimise solution as required

Cisco live!

Q & A

Complete Your Online Session Evaluation

Complete your session evaluation:

- Directly from your mobile device by visiting www.ciscoliveaustralia.com/mobile and login by entering your username and password

- Visit one of the Cisco Live internet stations located throughout the venue

- Open a browser on your own computer to access the Cisco Live onsite portal

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. Activate your account by visiting www.ciscolive365.com

Cisco Public