
TOMORROW
starts here.

How to Achieve True Active-Active Data Centre Infrastructures

BRKDCT-2615

Carlos Pereira

Distinguished Systems Engineer II – WW Data Centre / Cloud

(with extensive credits and thanks to my fellow Cisco collegues: Victor Moreno, Patrice Bellagamba, Yves Louis, Mike Herbert)

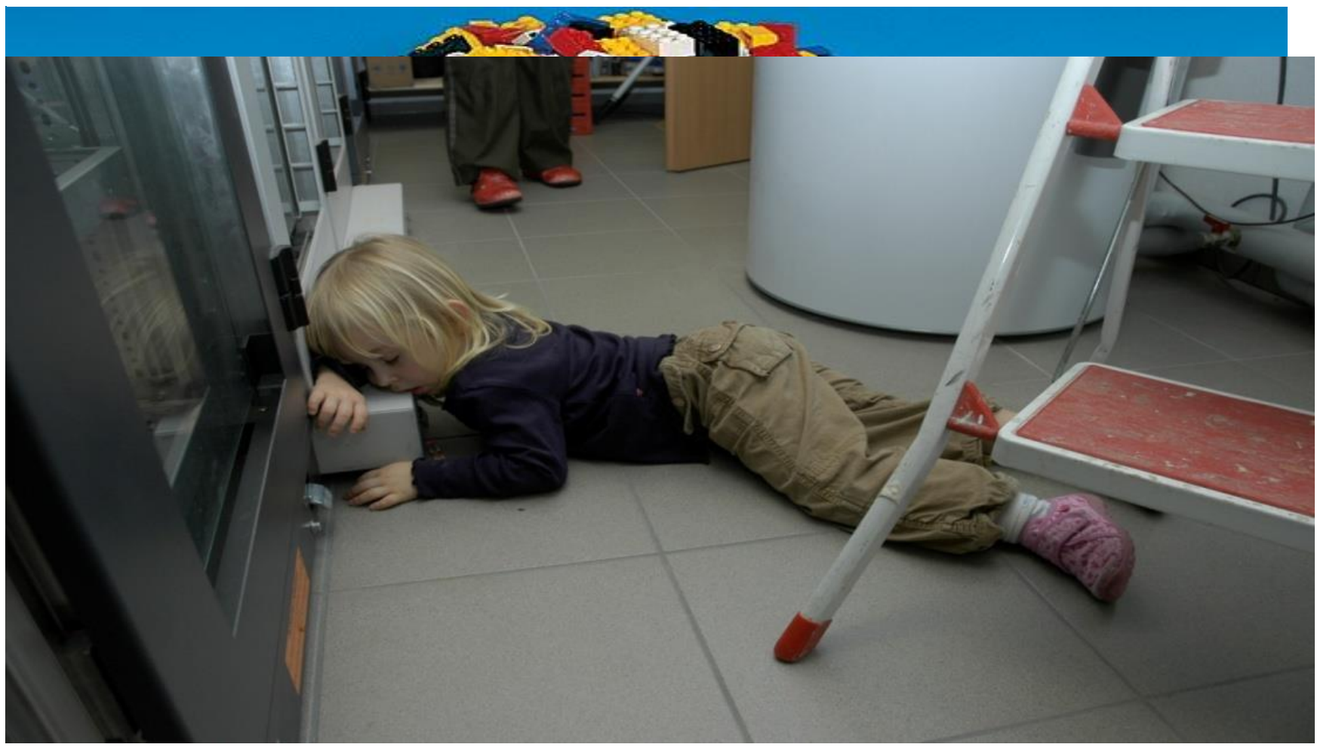
#clmel

Cisco live!

Active / Active Data Centres

Then try to figure that out

… and feel tired (or panic ☺)



Objectives

- Understand the Active/Active Data Centre requirements and considerations
- Provide considerations for Active/Active DC Design – inclusive for Metro areas - from storage, DCI, LAN extension, ACI and network services perspectives
- Brainstorm about ACI Fabric extension, stretch fabrics, application portability, VM mobility, policy synchronisation, etc.
- Share Experiences with State-full Devices placements and their impact within DCI environment

Reference slides would be
Quickly (if) covered during the session.

Legend

Load Balancer

SSL Offloader

APIC

Application
Policy
Infrastructure
Controller

SVI / HSRP Default Gw

IDS / IPS

WAN Accelerator

Firewall

Potential Collateral Effect ☺

nerd

Agenda

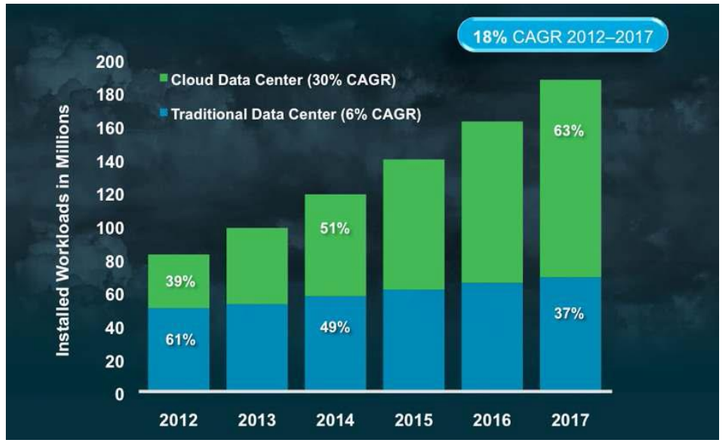
- Active-Active (A/A) Data Centre:
  - Market & Business Drivers
  - Terminology, Criticality levels and Solutions Overview
- A/A Data Centre Design Considerations:
  - Storage Extension
  - Data Centre Interconnect (DCI) – L2 & L3 scenarios
- A/A Metro Data Centres Designs

  - Network Services and Applications (Path optimisation)
- Cisco ACI and Active / Active Data Centre
- Q&A

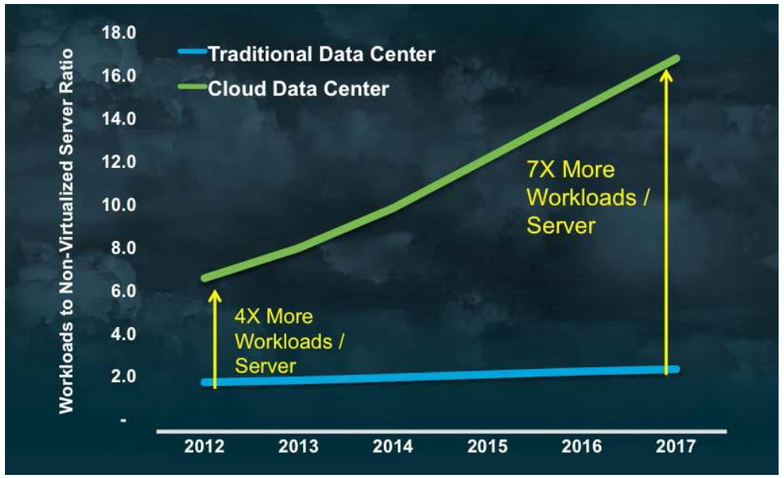
Some Important Trends Impacting the Data Centre Evolution

1 More Workloads are moving to Cloud Data Centres

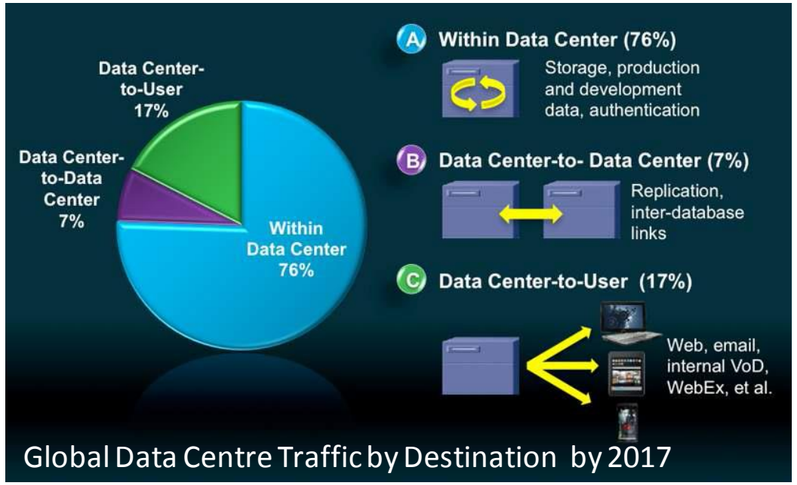
The increasing density of Business Critical workloads hosted in the Cloud is driving new Multi-site designs to handle Business Continuity, Workload Mobility, and Disaster Recovery



2 Cloud Data Centres include more virtualised workloads per server



3 Traffic in each area of the Data Centre is increasing dramatically



Source: Cisco Global Cloud Index, Forecast and Methodology, 2012–2017

4 Two Market Transitions – One DC Network

Virtual Machines

Apps Portability, Cross-Platform & Automation

LXC / Docker Containers

Applications

PaaS

Infrastructure

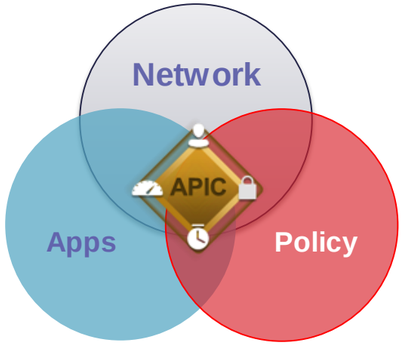
HyperScale Data Centres

Traditional Data Centre Networking

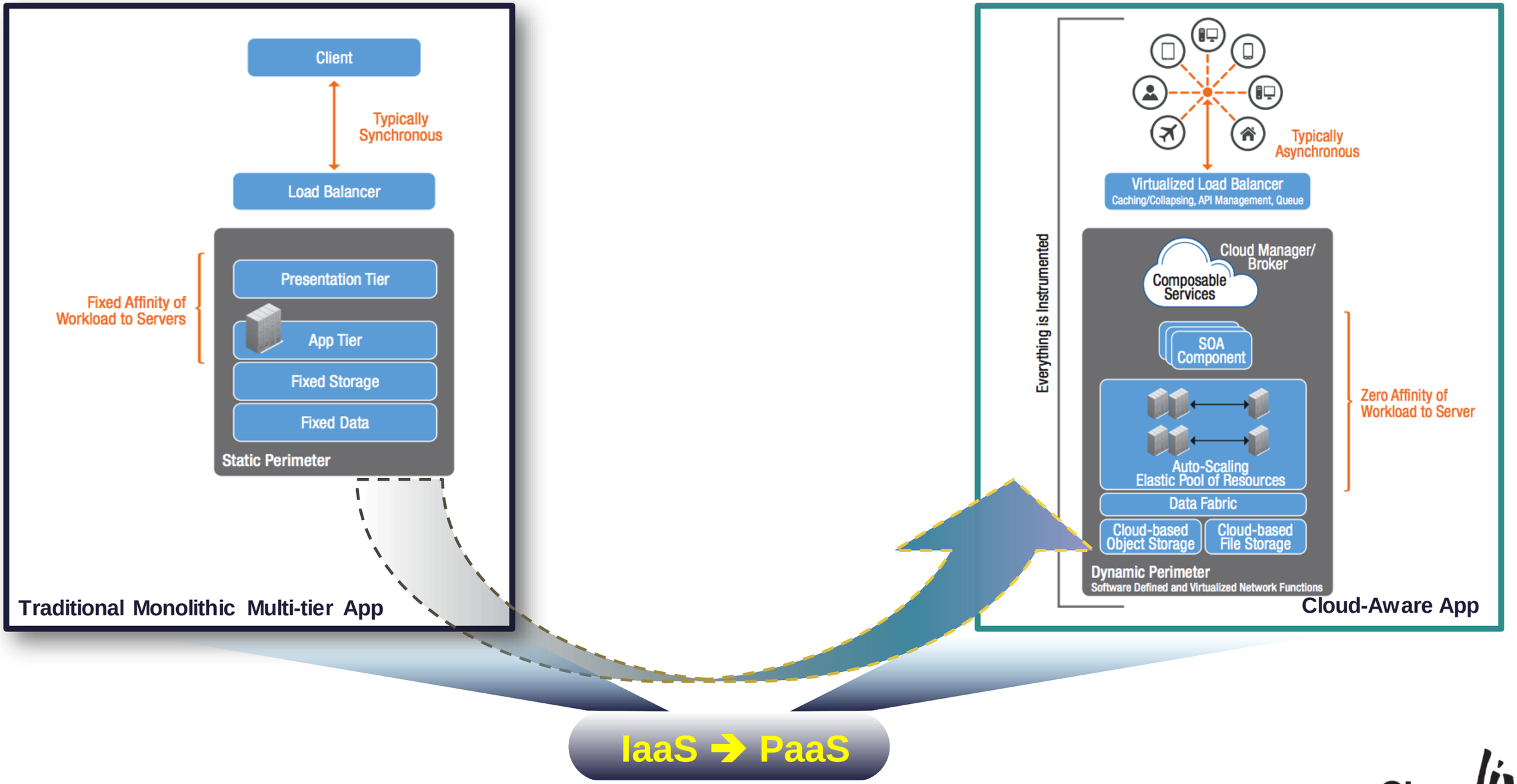
Application Centric Infrastructure (ACI)

DC Switching

Network + Services Abstraction & Automation

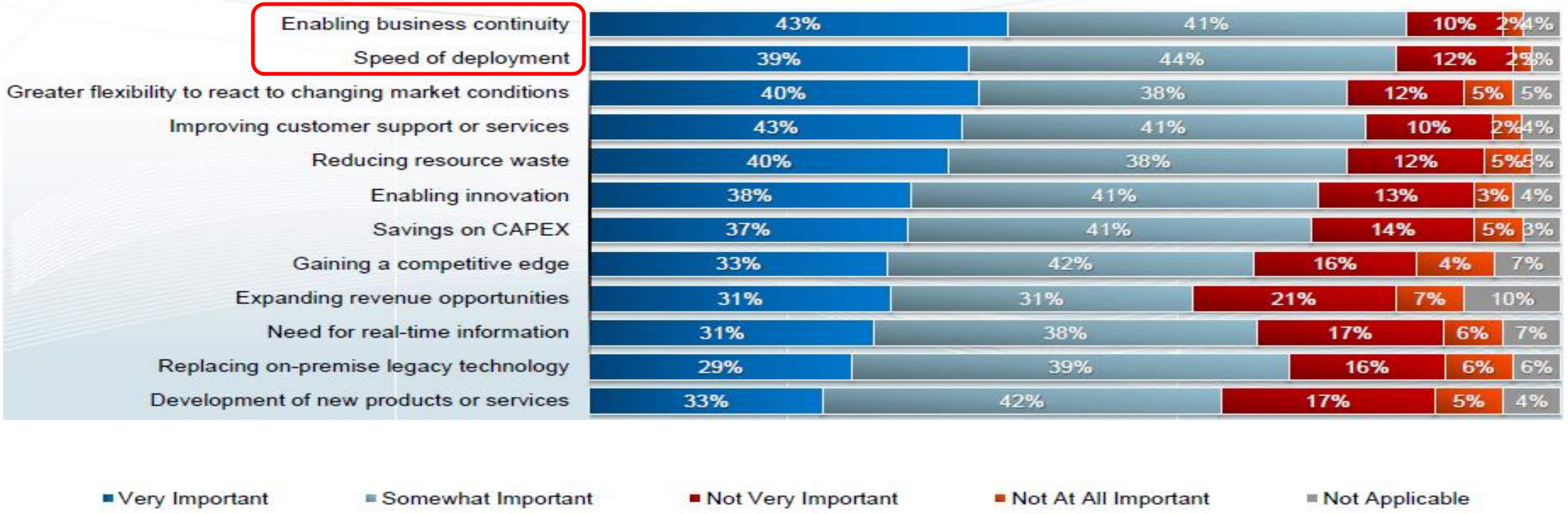


5 The App Market tTansition – From Monolithic To Cloud-aware



Business Continuity & Speed of Deployment are Top Drivers

IDG Enterprise
An IDG Communications Company



Q. How important are the following as business drivers of investment in cloud computing technology?

Source: IDG Enterprise Cloud Computing Study, January, 2012

Terminology

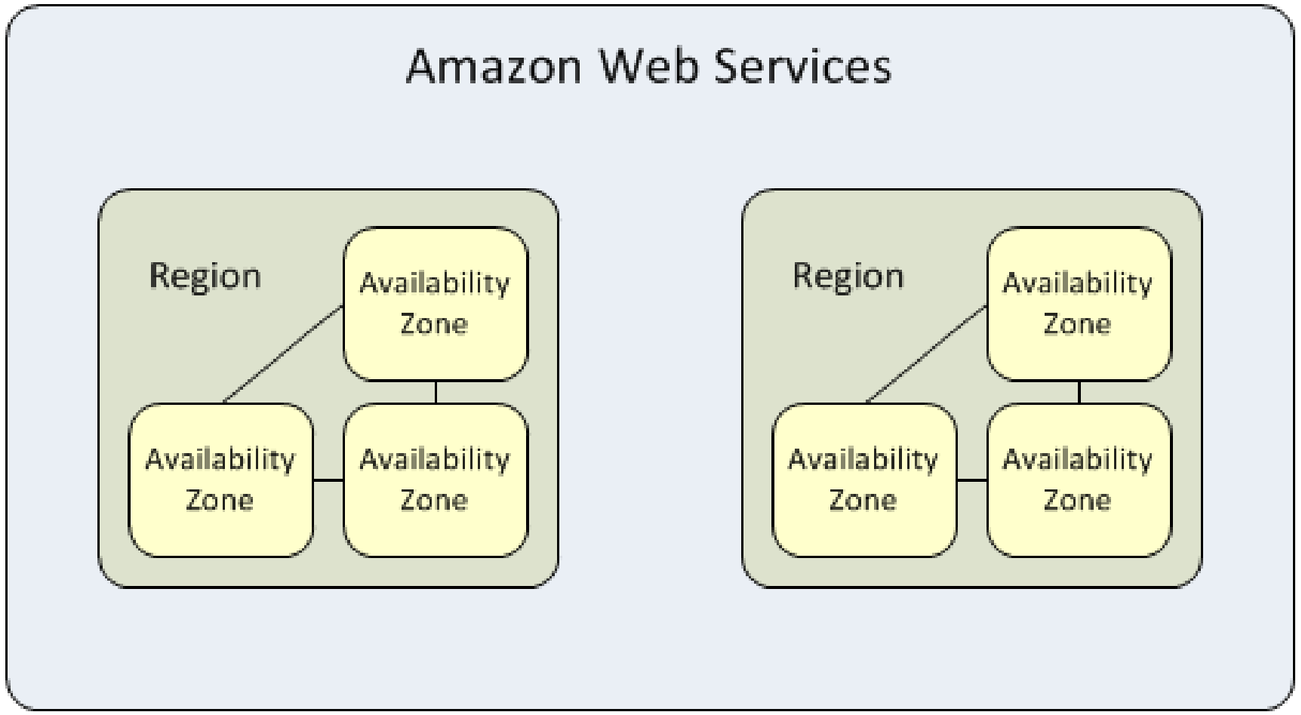
- The Terminology around Workload and Business Availability / Continuity is not always consistent
- Some examples:

“Availability Zone”

  - AWS - Availability Zones are distinct locations within a region that are engineered to be isolated from failures in other Availability Zones
  - OpenStack - An availability zone is commonly used to identify a set of servers that have a common attribute. For instance, if some of the racks in your data centre are on a separate power source, you can put servers in those racks in their own availability zone.

Availability Zone and Regions – AWS Definitions

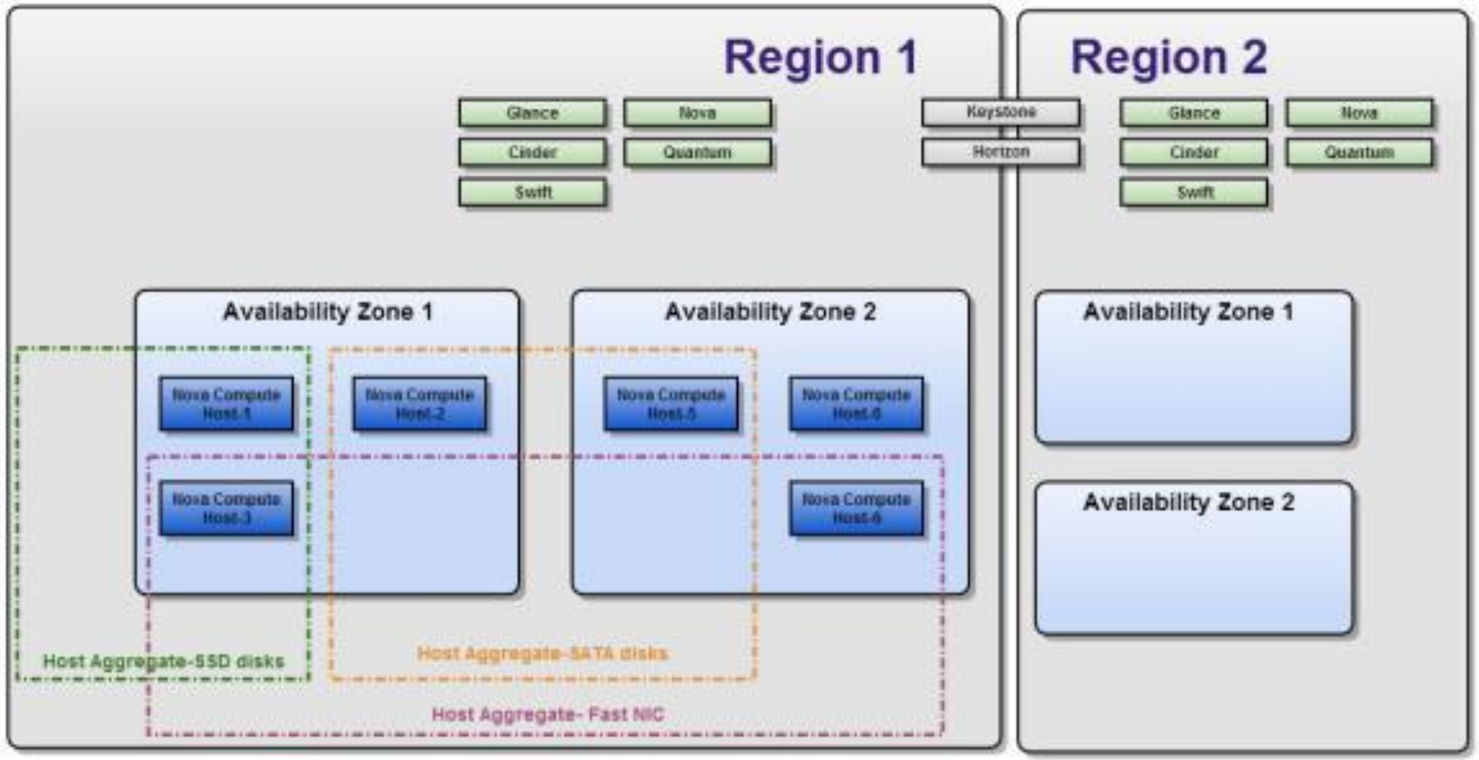
- Regions are large and widely dispersed into separate geographic locations.
- Availability Zones are distinct locations within a region that are engineered to be isolated from failures in other Availability Zones and provide inexpensive, low latency network connectivity to other Availability Zones in the same region



Availability Zone and Regions - OpenStack Definitions

- Regions - Each Region has its own full Openstack deployment, including its own API endpoints, networks and compute resources. Different Regions share one set of Keystone and Horizon to provide access control and Web portal. (Newer deployments do not share Keystone)

- Availability Zones - Inside a Region, compute nodes can be logically grouped into Availability Zones, when launching new VM instance, we can specify AZ or even a specific node in a AZ to run the VM instance.



In-Region and Out-of-Region Data Centres Business Continuity and Disaster Recovery

- **Active/active** — Traffic intended for the failed node is either passed onto an existing node or load balanced across the remaining nodes.

- **Active/passive** — Provides a fully redundant instance of each node, which is only brought online when its associated primary node fails.

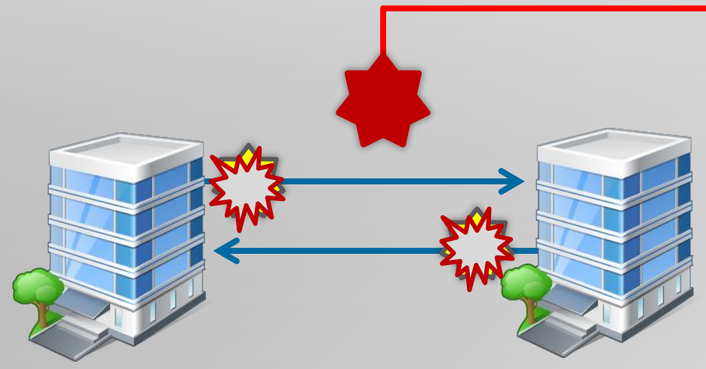
- **Out-of-Region** – Beyond the ‘Blast Radius” for any disaster

Business Continuity and Disaster Recovery

Ability to Absorb the Impact of a Disaster and Continue to Provide an Acceptable Level of Service.

"Operational Continuity"

Planned / Unplanned Service Continuity within the Metro
(including a single data centre loss)



DC-1 DC-2

Active-Active

Metro Area

"Disaster Recovery (DR)"

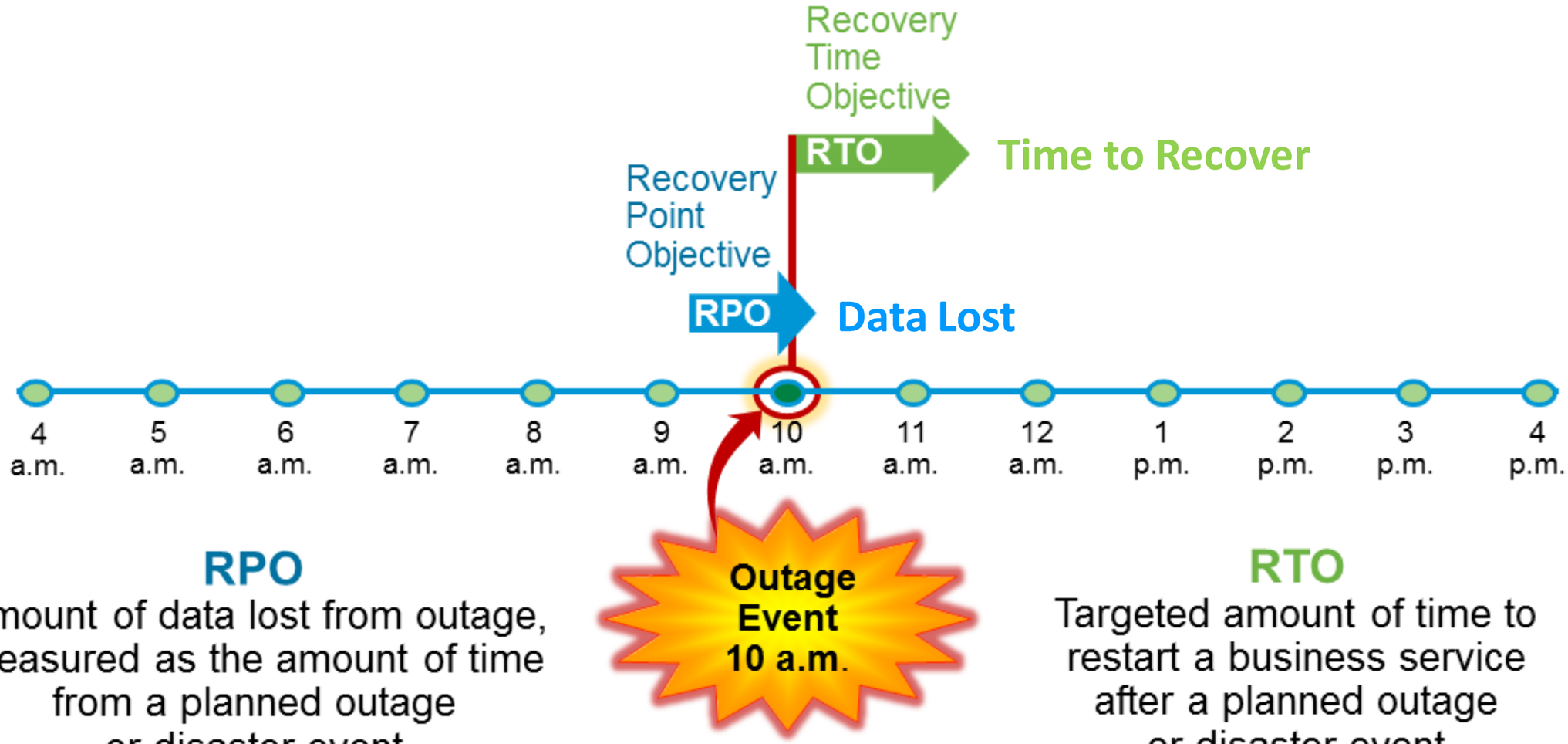
Loss of "Regional data centres" leads to recovery in a Remote data centre

DC-3

Pervasive Data Protection
+ **Infrastructure Rebalance**

"Applications and services extended across Metro and Geo distances are a natural "next step."

Industry Standard Measurements of Business Continuity



RPO

Amount of data lost from outage, measured as the amount of time from a planned outage or disaster event

RTO

Targeted amount of time to restart a business service after a planned outage or disaster event

Application Resiliency and Business Criticality Levels

Defining how a Service outage impacts Business will dictate a redundancy strategy (and cost)

Each Data Centre should accommodate all levels… Cost is important factor

Lowest RTO/RPO

Highest RTO/RPO

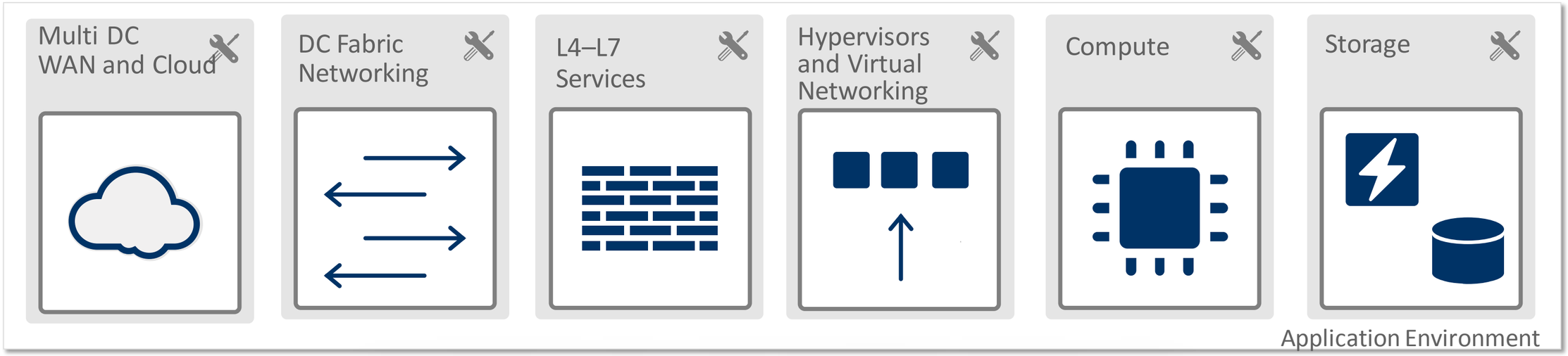
Criticality Levels	Application Level	Impact Description	Typical App Distribution	Typical Cost $
C1 RTO/RPO 0 to 15 mins	Mission Imperative	Any outage results in immediate cessation of a primary function, equivalent to immediate and critical impact to revenue generation, brand name and/or customer satisfaction; **no downtime is acceptable under any circumstances**	~ 20% of Apps	*100% Duplicate Resources, 2x Cost Multiplier (Most Costly)*
C2	Mission Critical	Any outage results in immediate cessation of a primary function, equivalent to major impact to revenue generation, brand name and/or customer satisfaction;		
C3 RTO/RPO 15+ mins	Business Critical	Any outage results in cessation over time or an immediate reduction of a primary function, equivalent to minor impact to revenue generation, brand name and/or customer satisfaction	~ 20% of Apps	*Many-to-One Resource Sharing, Lower Cost Multiplier (Less Costly)*
C4	Business Operational	A sustained outage results in cessation or reduction of a primary function	~ 40% of Apps	
C5	Business Administrative	A sustained outage has **little to no impact** on a primary function	~ 20% of Apps	

Two of the most common RTO/RPO Targets

Mapping Applications to Business Criticality Levels

Application Centric View of Data Centre Interconnect (DCI)

- Applications consume resources across the Cloud DC infrastructure
- If an Application moves between sites, each element of the Application Environment must also adjust to the new location
- **DCI extends the Application Environment between Geographic sites within Private Clouds and Public Clouds**
- Critical IT Use Cases including Business Continuity, Workload Mobility, and Disaster Recovery within Public and Private Clouds, impact each element of the Application Environment



DCI EXTENDS THE APPLICATION ENVIRONMENT ACROSS MULTIPLE SITES, SUPPORTING PHYSICAL AND VIRTUAL ELEMENTS

DCI Enables Critical Use Cases within Private Clouds and Public Clouds

Including Business Continuity and Workload Mobility between Metro/Geo Sites

Site 1

Business Continuity
Workload Mobility
Disaster Recovery and Avoidance
Site Migrations

Business Use Cases

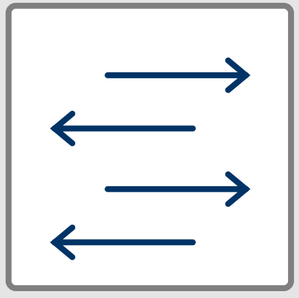
Load Balanced Workloads
Operations Maintenance
Operations Rebalancing
Application Geo-Clusters

Site 2

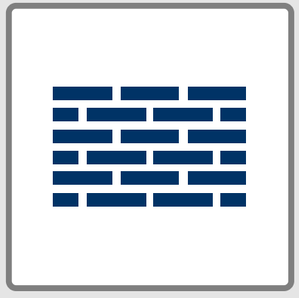
Application Environment

Multi DC
WAN and Cloud

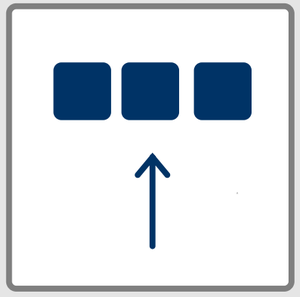
DC Fabric
Networking



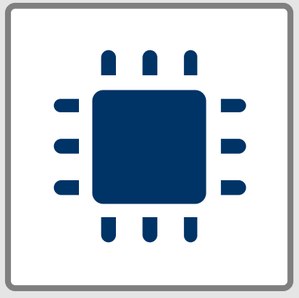
L4–L7
Services



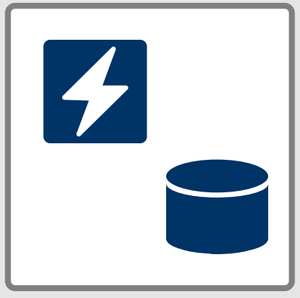
Hypervisors
and Virtual
Networking



Compute



Storage



The Application Environment Spans Many Cloud Resources

DCI Extends Cloud Resources to Support Multi-site Use Cases

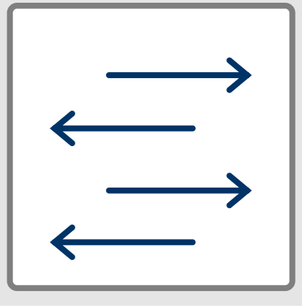
Multi DC WAN & Cloud

WAN Connectivity
- IP Internet Access
- MPLS VPN Access
- Physical or Virtual WAN router
- IP Path Optimisation (LISP, DNS, Site Selector)

L3 Routing and IGP
- OSPF
- ISIS
- BGP

Data Centre Interconnect
- Overlay Transport Virtualisation (OTV)
- EoMPLS, VPLS
- E-VPN

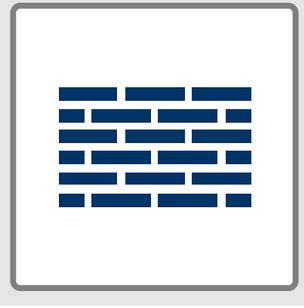
DC Fabric Networking



Data Centre Fabrics
- Virtual Port Channel (vPC)
- VxLAN based (with our without SDN Controller)
- FabricPath
- Application Centric Infrastructure (ACI)

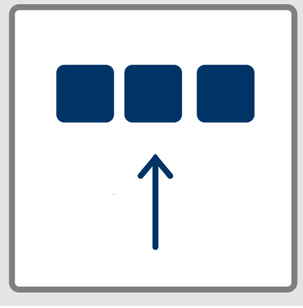
Fabric Services
- Tenancy
- Secure Segmentation (VRF, VLAN, VxLAN)
- Traffic QoS
- Bandwidth Reservation

L4–L7 Services



Physical and Virtual Services
- Firewalls
- Load Balancers
- IPSec VPN Termination
- WAN Acceleration Service
- Network Analysis
- Data Encryption

Hypervisors and Virtual Networking

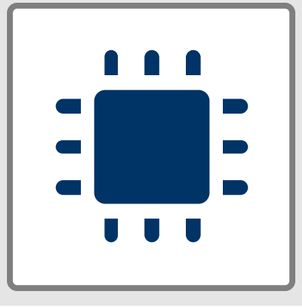


Hypervisors
- VMware vSphere
- Microsoft Hyper-V
- KVM (ex.: RedHat)

Hypervisor Services
- Live and Cold Application Migrations
- Extended Clusters
- High Availability and Recovery Services
- Site Affinity Services

Virtual Switching
- Nexus 1000v
- Virtual Interfaces

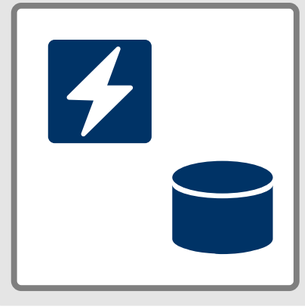
Compute



Unified Compute System (UCS)
- C-Series Rack Servers
- B-Series Blade Servers
- Physical and Virtual Interfaces
- Port and Security Profiles

Integrated PoDs
- FlexPod
- vBlock
- Low Cost Compute PoDs

Storage



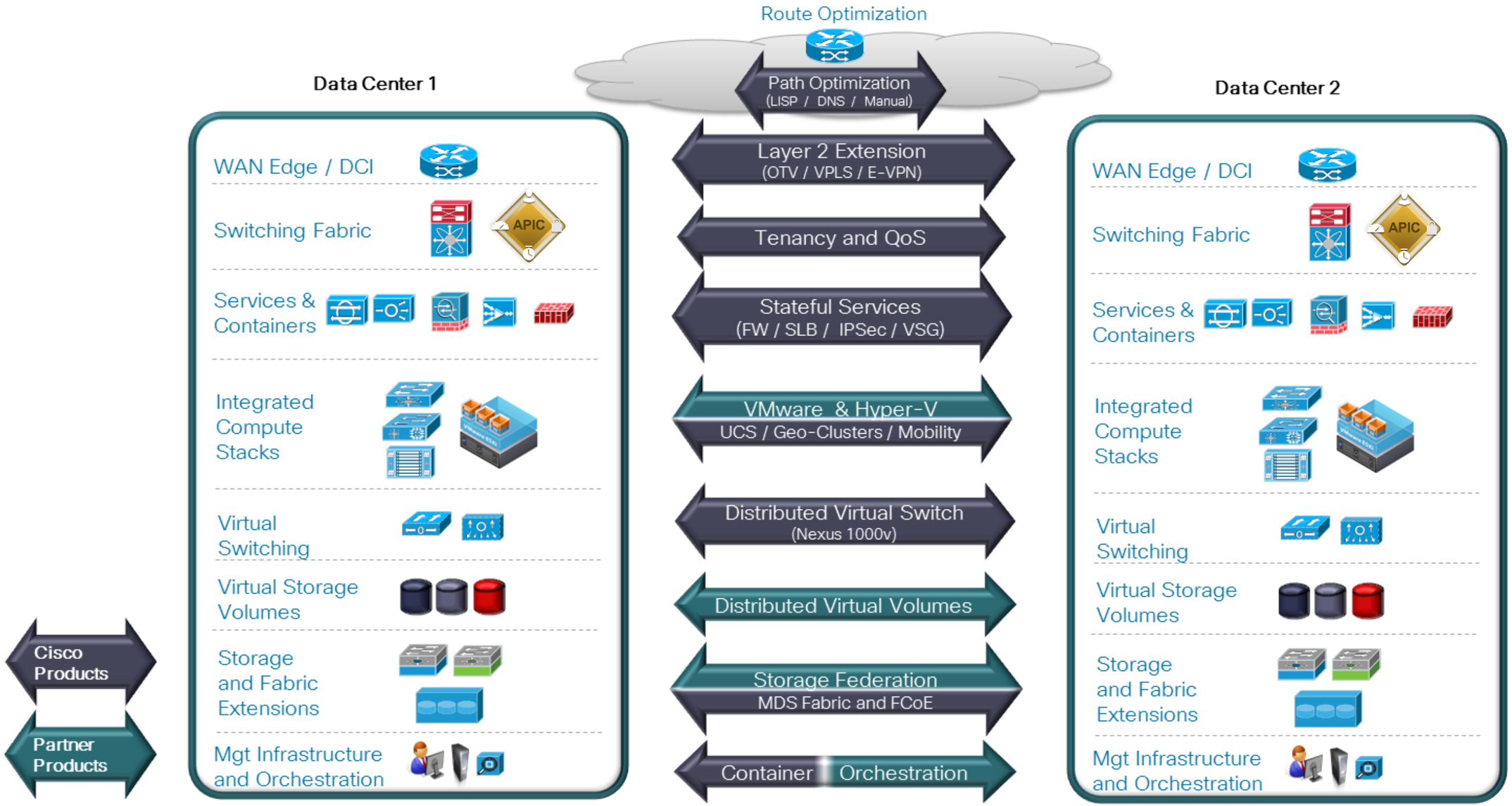
Storage
- NetApp
- EMC
- Direct Attached Storage

Storage Fabrics
- FC
- FCoE
- 10GE

Data Replication
- Synchronous
- Asynchronous
- Hypervisor Based
- DWDM / IP / FCIP

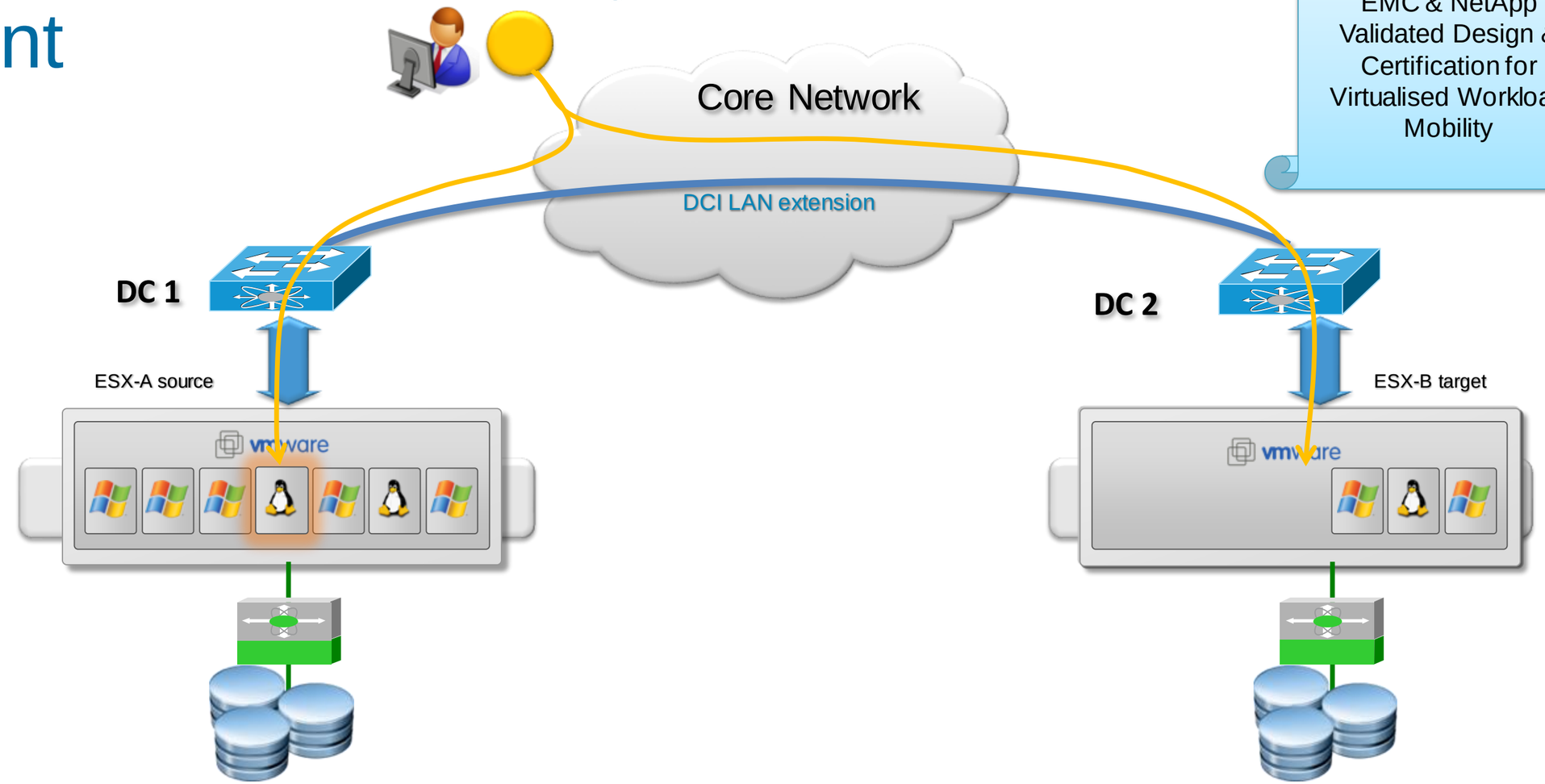
APPLICATION TEAMS CHOOSE FROM AVAILABLE DESIGN OPTIONS…
THESE FUNCTIONS ARE EXTENDED TO SUPPORT MULTI-SITE USE CASES

DCI Extensions Impact Each Tier of the Cloud Data Centre



Virtualised Workload Mobility Inter-DCs Deployment

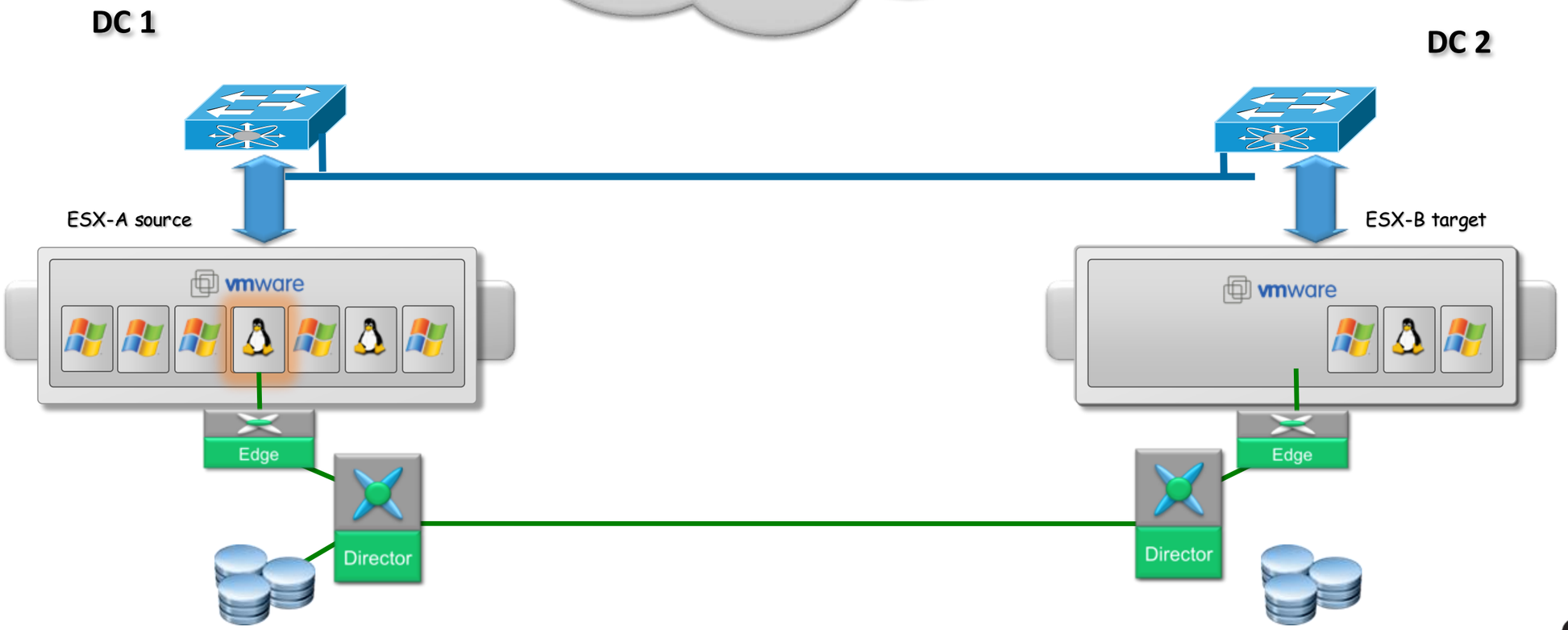
Cisco-VMware With EMC & NetApp Validated Design & Certification for Virtualised Workload Mobility



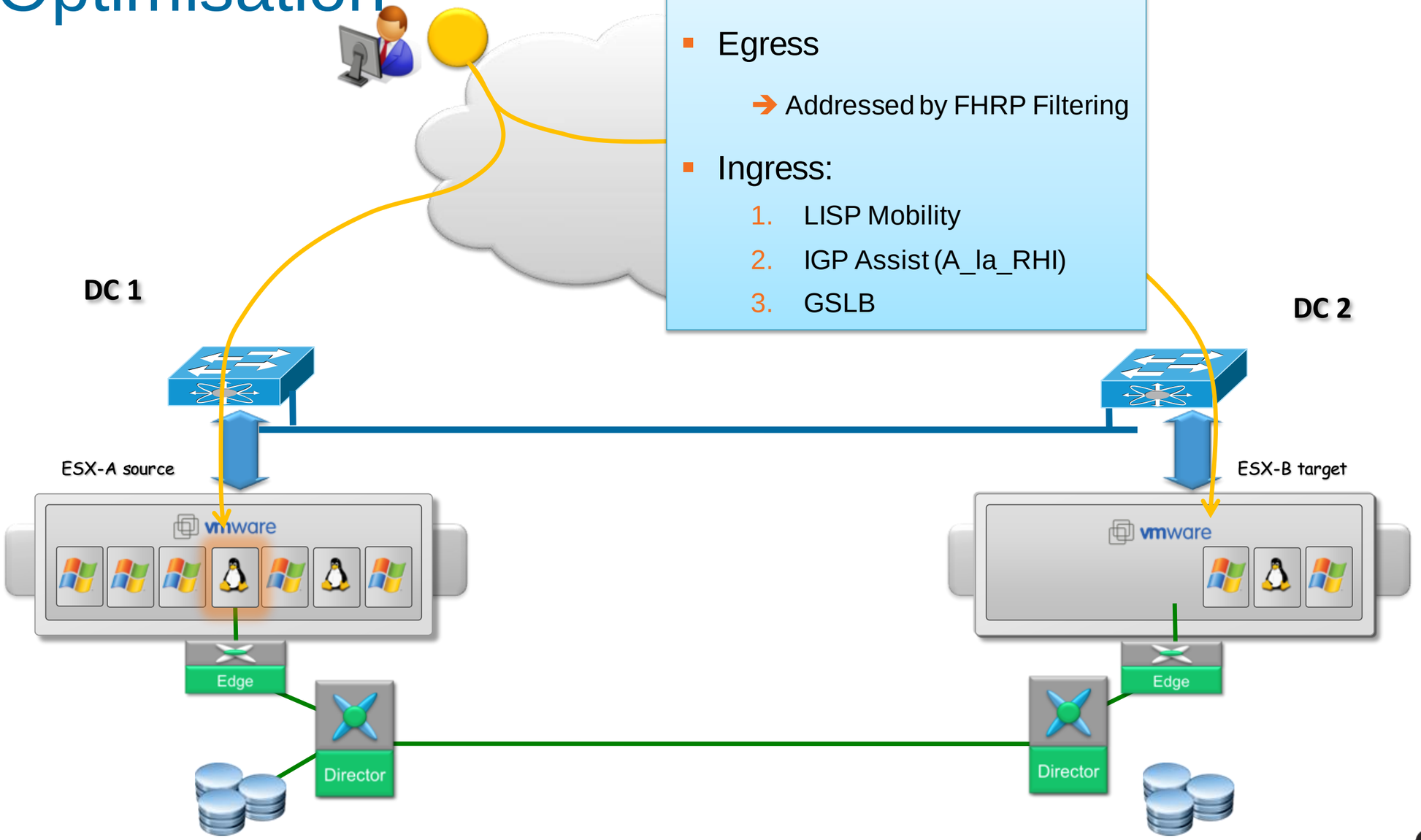
- Virtualised Workload Mobility across geographically dispersed sites
- Requires to stretch VLANs and to provide consistent LUN ID access
- Disaster Recovery Applications. Ex.: VMware Site Recovery Manager (SRM)

Data Centre Interconnect
LAN Extension

- STP Isolation is the key element
- Multipoint
- Loop avoidance + Storm-Control
  Unknown Unicast & Broadcast control
- Link sturdiness
- Scale & Convergence



Data Centre Interconnect Path Optimisation



Options

- Egress
  - ➔ Addressed by FHRP Filtering
- Ingress:
  1. LISP Mobility
  2. IGP Assist (A_la_RHI)
  3. GSLB

Ex.1: vSphere Redundancy and Mobility Options Can Extend Across Geographies

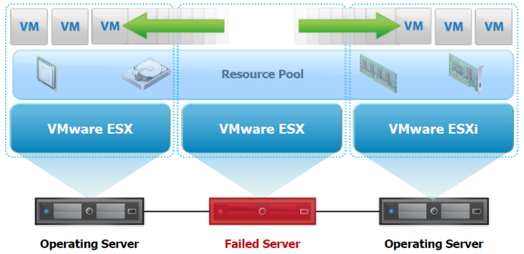
VM High Availability

VMware vSphere High Availability (HA)

Restarts Virtual Machines After Operating System or Hardware Failure

Automatically restarts VMs in the event of:
- Hardware failure
- VM failure (loss of heartbeat)

Transparent to OS and Applications
Downtime: Minutes

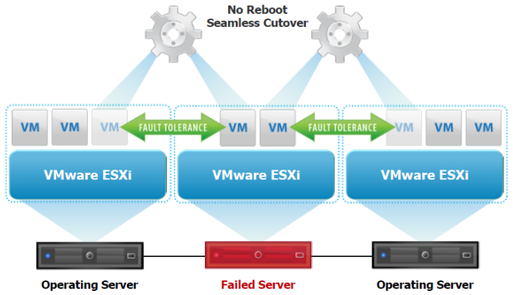


DC-1 ⟷ DC-2

VMware vSphere Fault Tolerance (FT)

Eliminates Workload Disruption Due to Hardware Failure

- A protected VM has a shadow VM in lockstep on another host
- Zero downtime and zero data loss in the event of host failure
- Downtime: Zero (Continuous Availability)

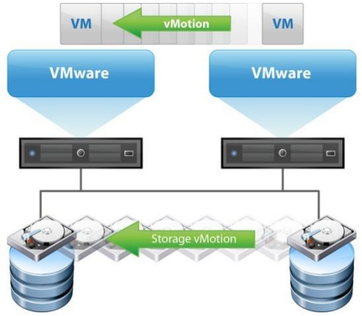


DC-1 ⟷ DC-2

VM Mobility

VMware vMotion and Storage vMotion

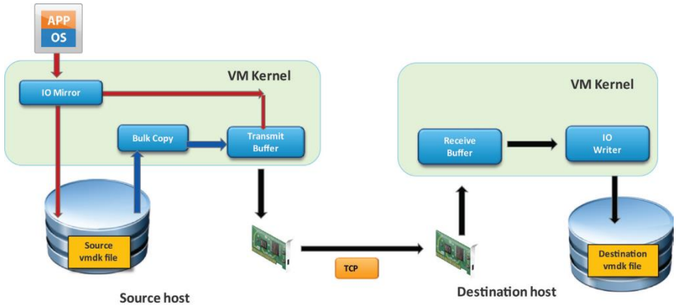
Live Migration of VMs and VM storage
- Non-disruptive migration of VMs
- Non-disruptive migration of Virtual storage



DC-1 ⟷ DC-2

"Shared Nothing" VMware vMotion

Live Migration of VMs and storage WITHOUT shared storage
- Simple management of recovery and migration plans
- Non-disruptive testing

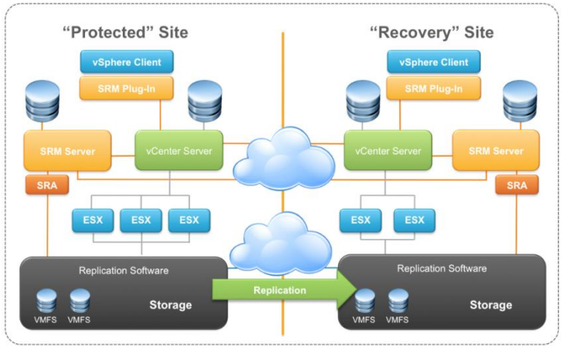


DC-1 ⟷ DC-2

Site / VM Recovery

VMware Site Recovery Manager (SRM)

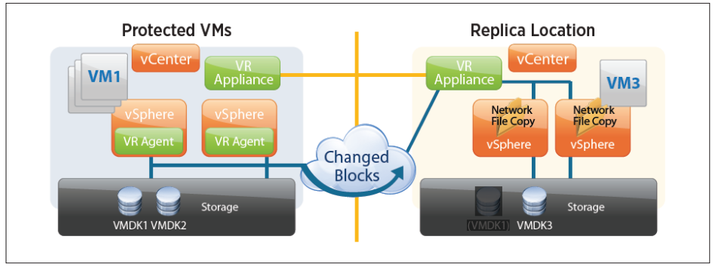
Fully automated site recovery and migration
- Simple management of recovery and migration plans
- Non-disruptive testing



DC-1 ⟷ DC-2

VMware vSphere Replication

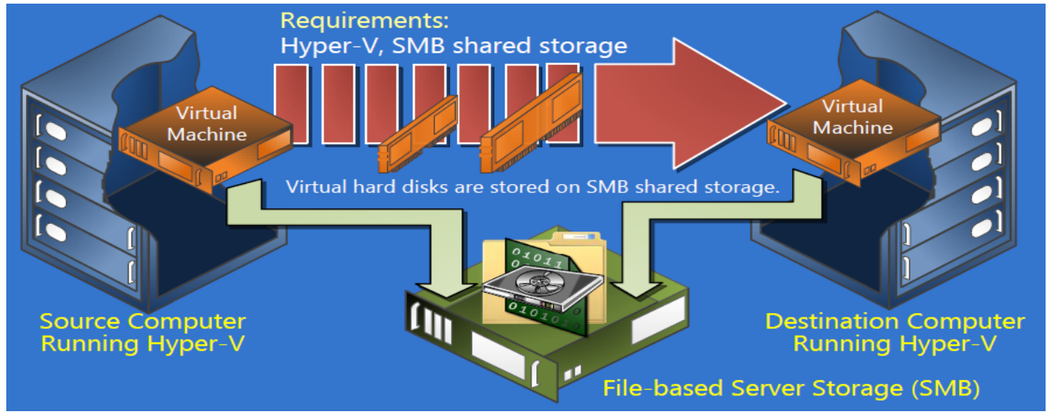
Creates VM snapshot copies available for restoration through the vCenter
- Continuous replication to another location, within or between clusters
- Hypervisor based replication, VM granularity



DC-1 ⟷ DC-2

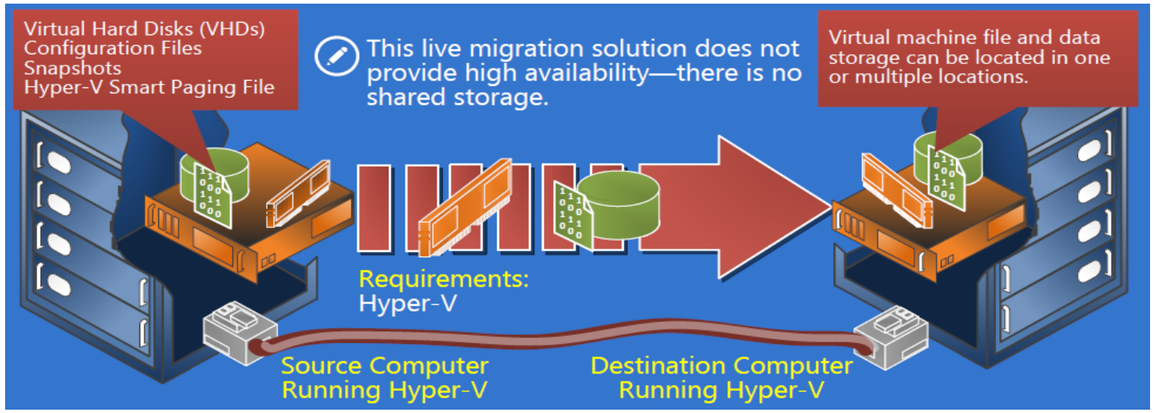
Ex.2: Hyper-V Redundancy and Workload Mobility Options Can Extend Across Geographies

Hyper-V Live Migration with Shared SMB Storage



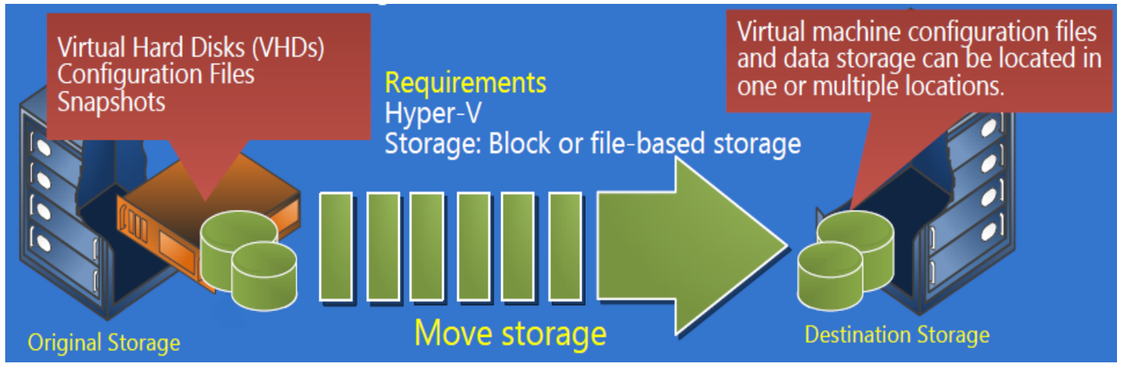
DC-1 ⟷ DC-2

Hyper-V "Shared Nothing" Live Migration



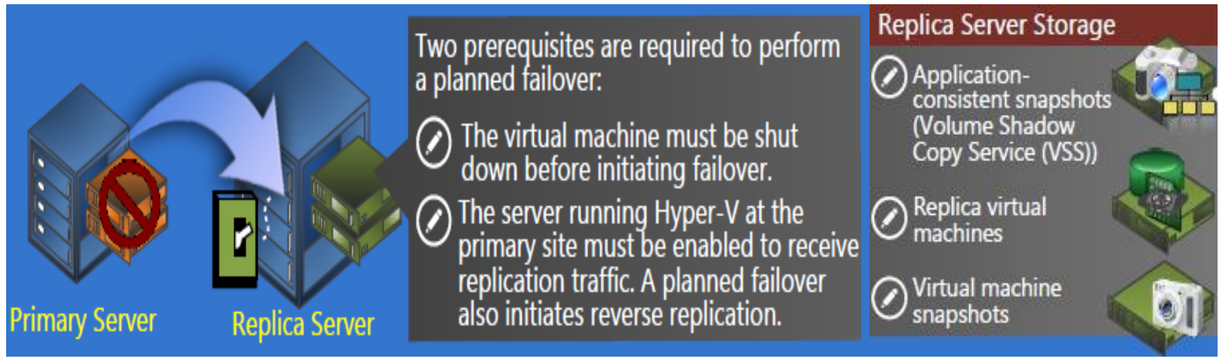
DC-1 ⟷ DC-2

Hyper-V Storage Migration



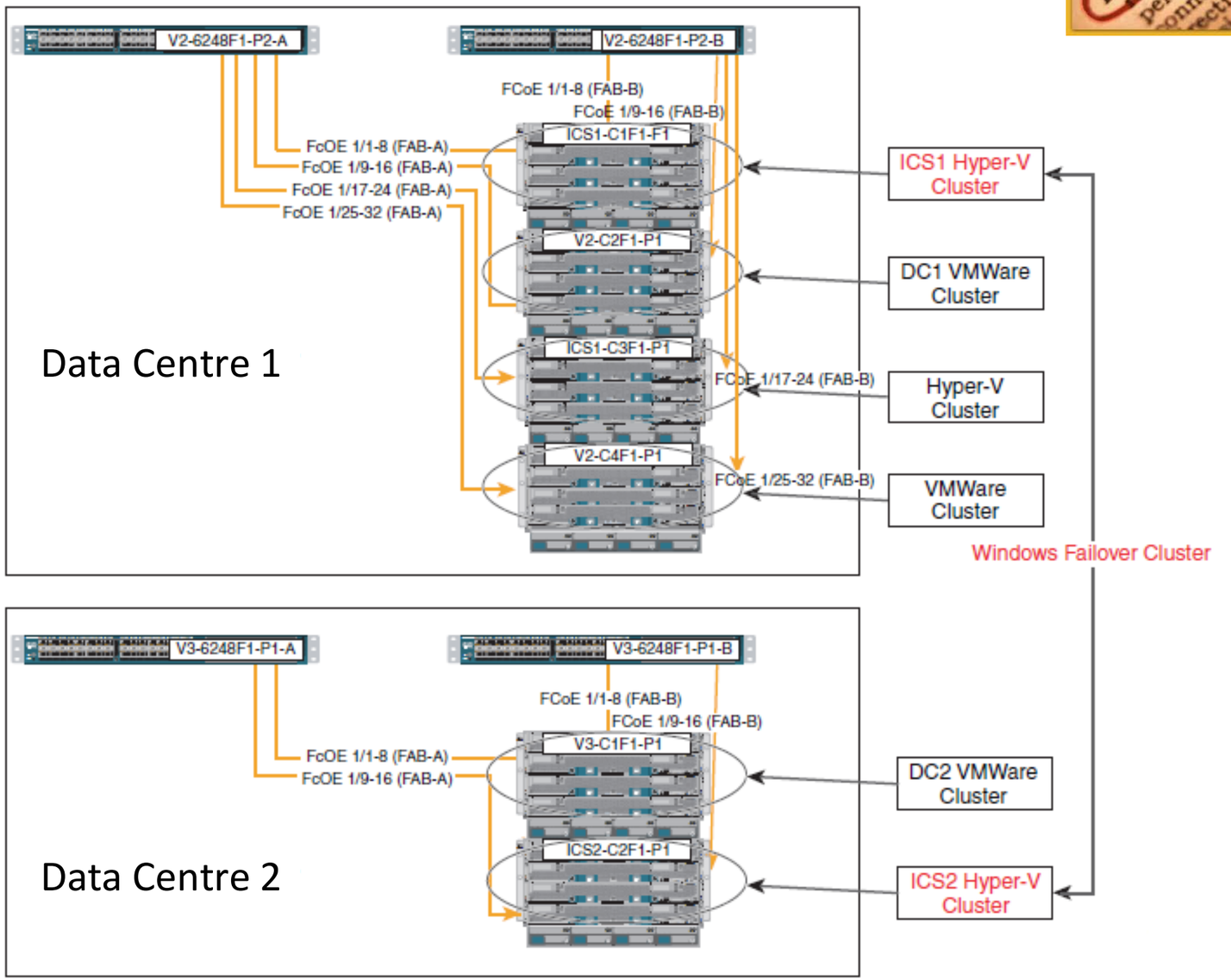
DC-1 ⟷ DC-2

Hyper-V Replica – Application Consistent Snapshots



DC-1 ⟷ DC-2

VMware vSphere
and
Microsoft Hyper-V
environments are
concurrently supported
in Cisco Cloud
solutions



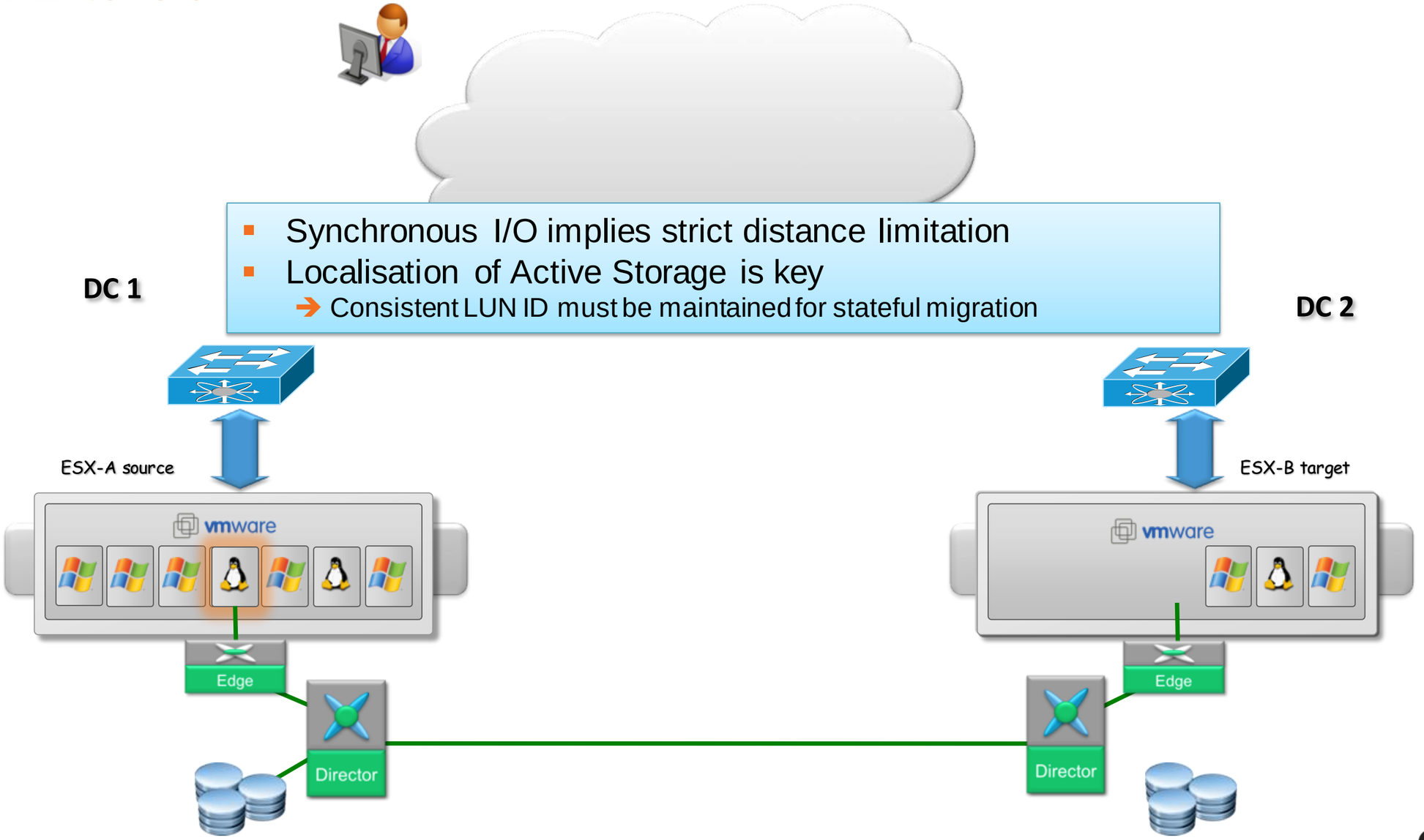
Agenda

- Active-Active (A/A) Data Centre:
  - Market & Business Drivers
  - Terminology, Criticality levels and Solutions Overview
- A/A Data Centre Design Considerations:
  - Storage Extension
  - Data Centre Interconnect (DCI) – L2 & L3 scenarios
- A/A Metro Data Centres Designs

  - Network Services and Applications (Path optimisation)
- Cisco ACI and Active / Active Data Centre
- Q&A

Data Centre Interconnect

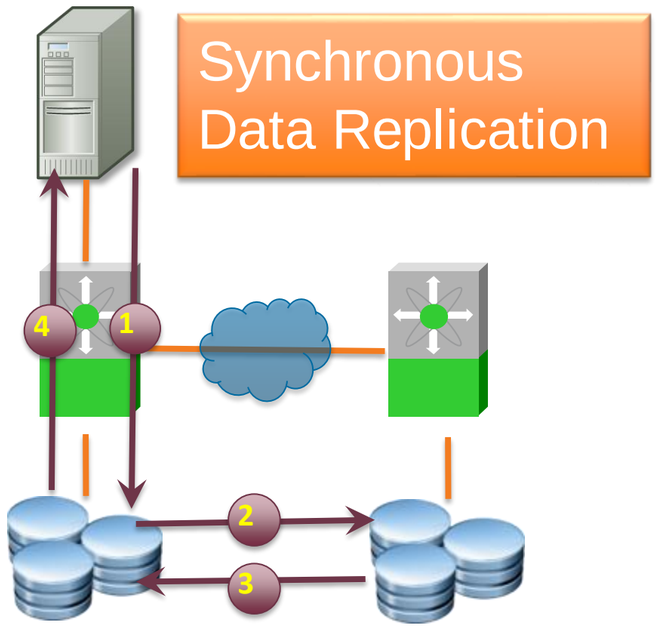
SAN Extension

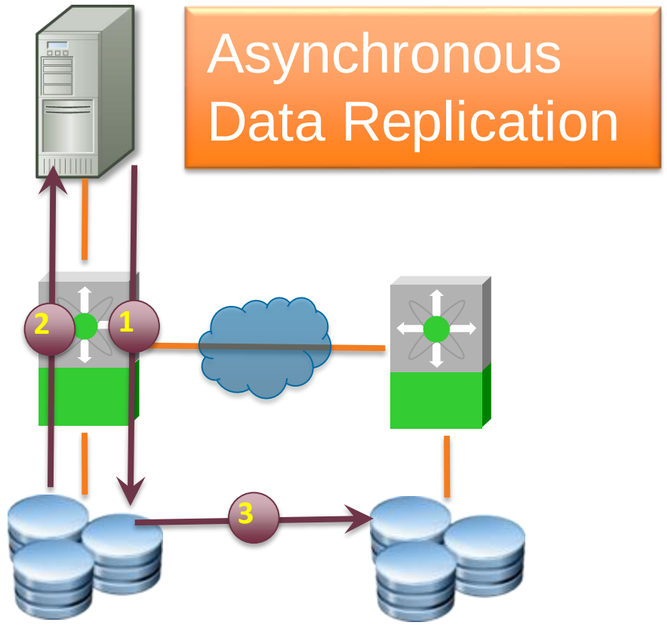


SAN Extension

Synchronous vs. Asynchronous Data Replication

- **Synchronous Data replication:** The Application receives the acknowledgement for I/O complete when both primary and remote disks are updated. This is also known as Zero data loss data replication method (or Zero RPO)
  - Metro Distances (depending on the Application can be 50-300kms max)
- **Asynchronous Data replication:** The Application receives the acknowledgement for I/O complete as soon as the primary disk is updated while the copy continues to the remote disk.
  - Unlimited distances



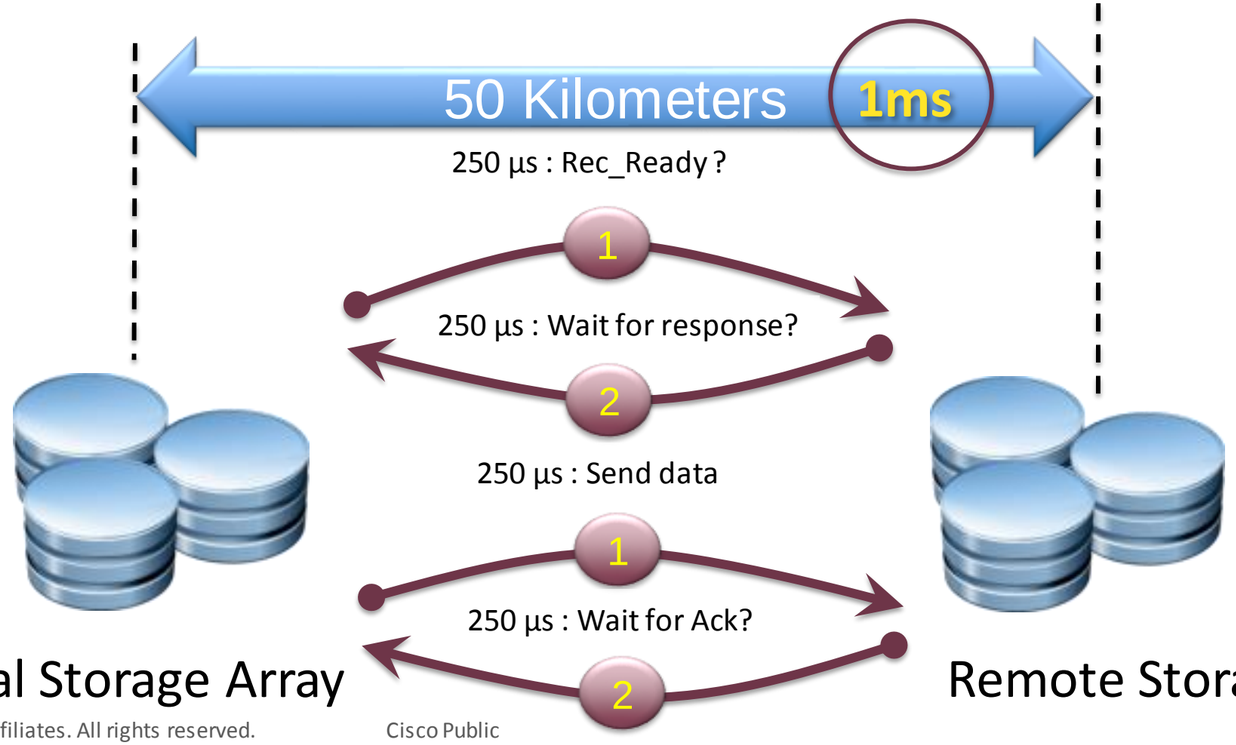


Synchronous Data Replication

Network Latency

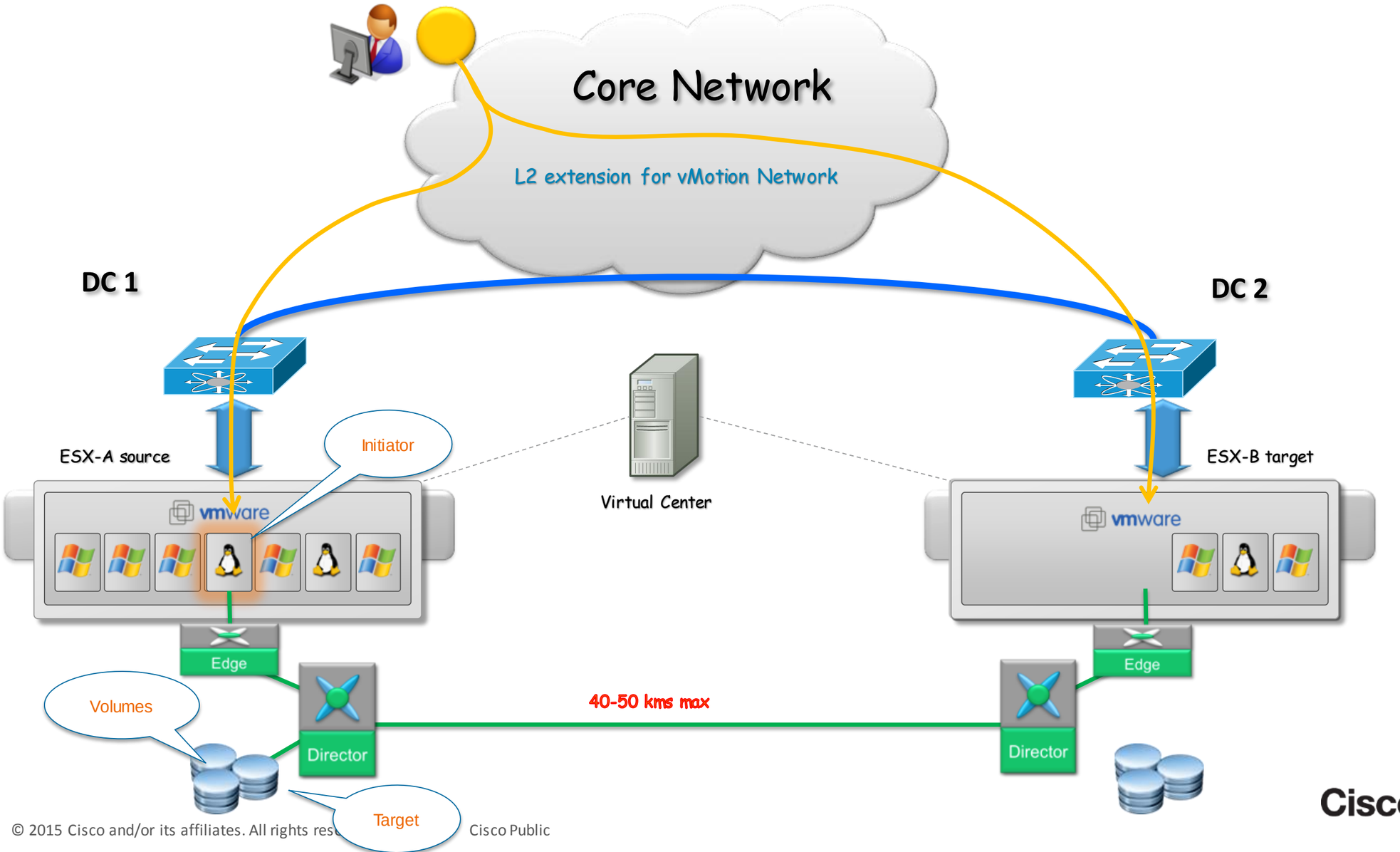
- Speed of Light is about 300000 Km/s
- Speed is reduced to 200000 Km/s ➔ 5 µs per Km (8 µs per Mile)
- That gives us an average of **1ms** for the light to cross **200 Km** of fibre

- Synchronous Replication: SCSI protocol (FC) takes a four round trips
- For each Write cmd a two round trips is about 10 µs per kilometer
  ➔ 20µs/km for 4 round trips for Synch data replication



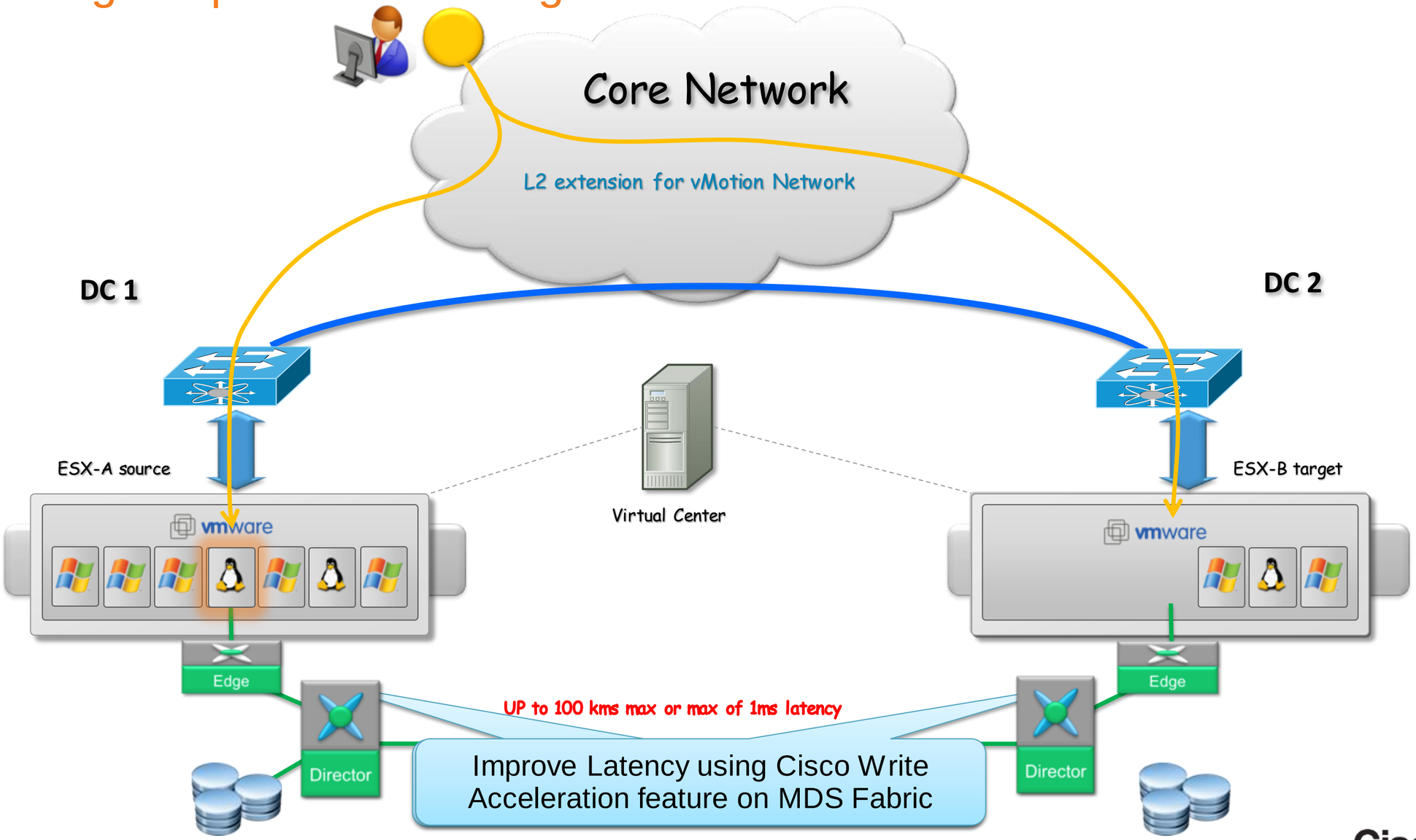
Storage Deployment in DCI

Option 1 - Shared Storage



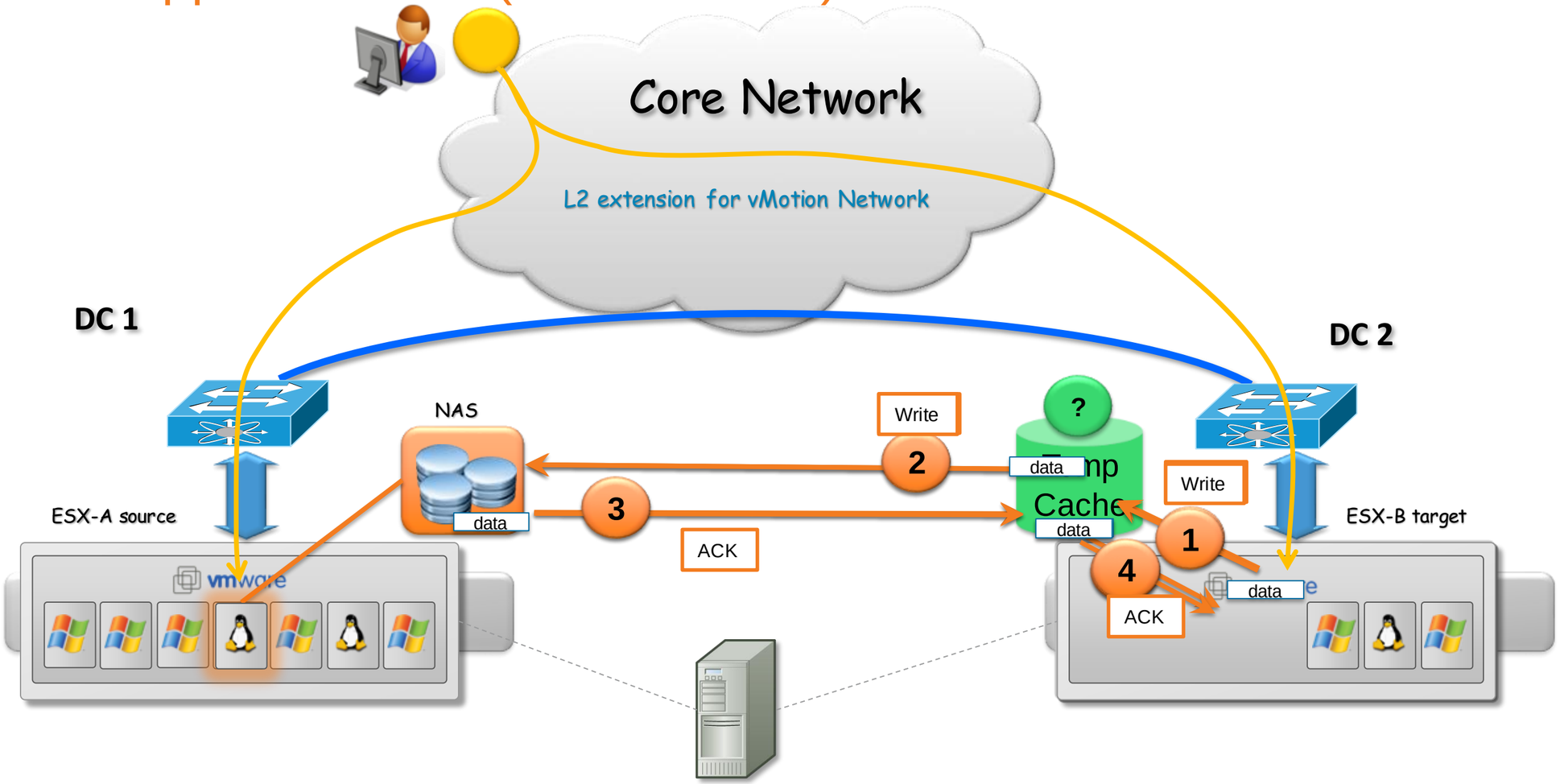
Storage Deployment in DCI

Shared Storage Improvement Using Cisco IOA



Storage Deployment in DCI

Option 2 - NetApp FlexCache (Active/Cache)

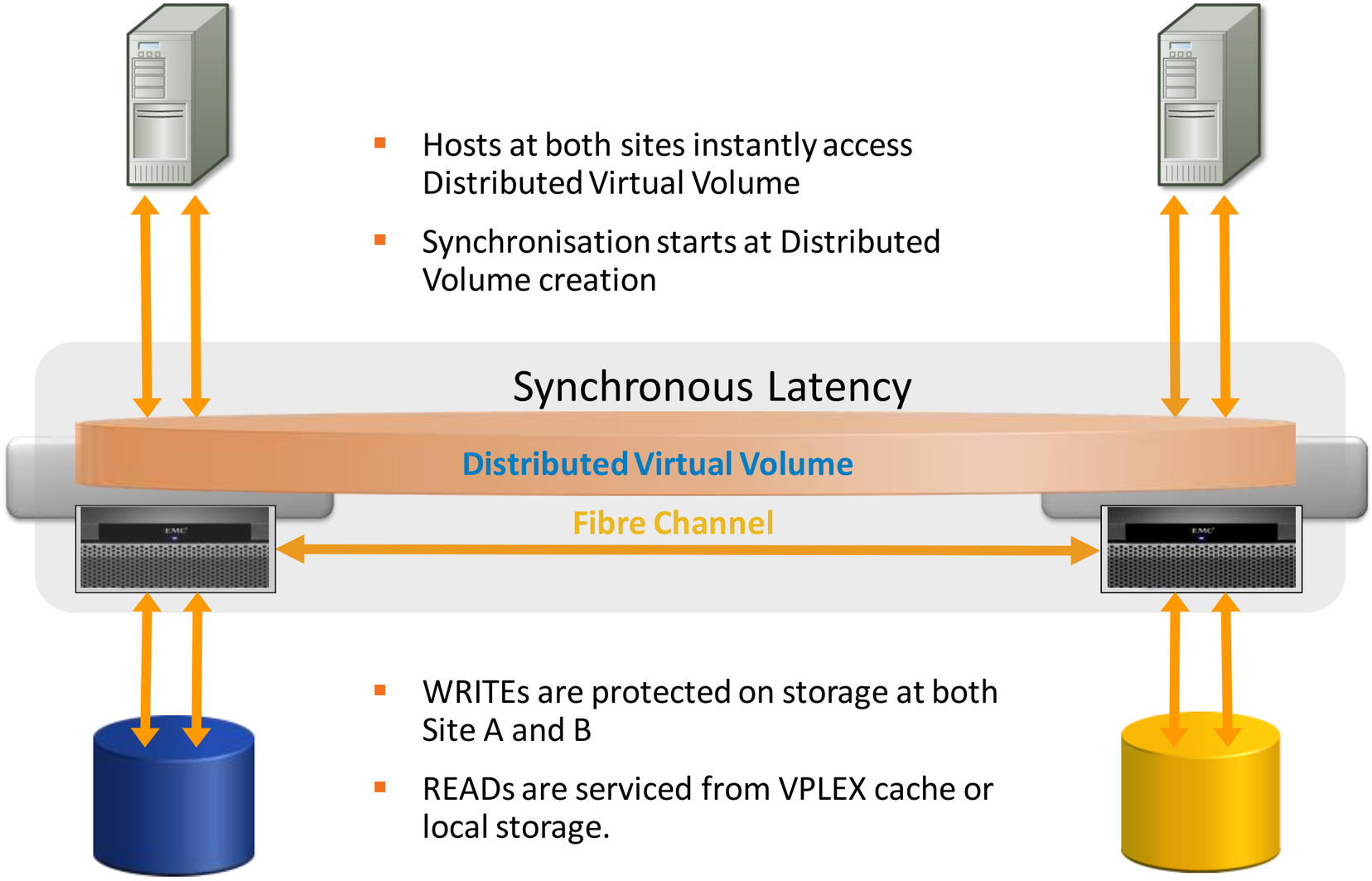


- FlexCache does NOT act as a write-back cache
- FlexCache responds to the Host only if/when the original subsystem ack'ed to it
- No imperative need to protect a Flexcache from a power Failure

http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DCI/4.0/Netapp/dciNetapp.html

Storage Deployment in DCI

Option 3 - EMC VPLEX Metro (Active/Active)

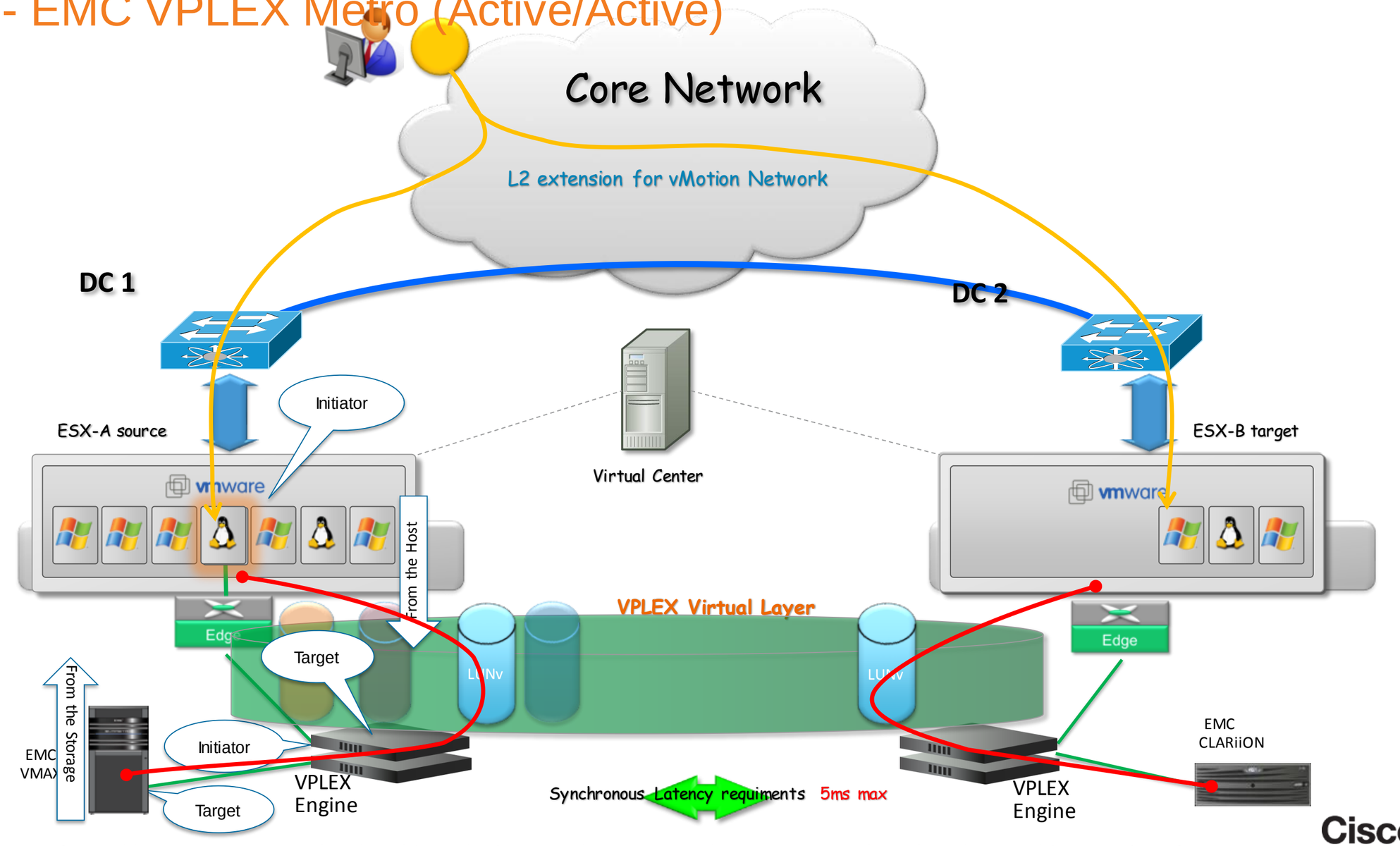


- Hosts at both sites instantly access Distributed Virtual Volume
- Synchronisation starts at Distributed Volume creation

- WRITEs are protected on storage at both Site A and B
- READs are serviced from VPLEX cache or local storage.

Storage Deployment in DCI

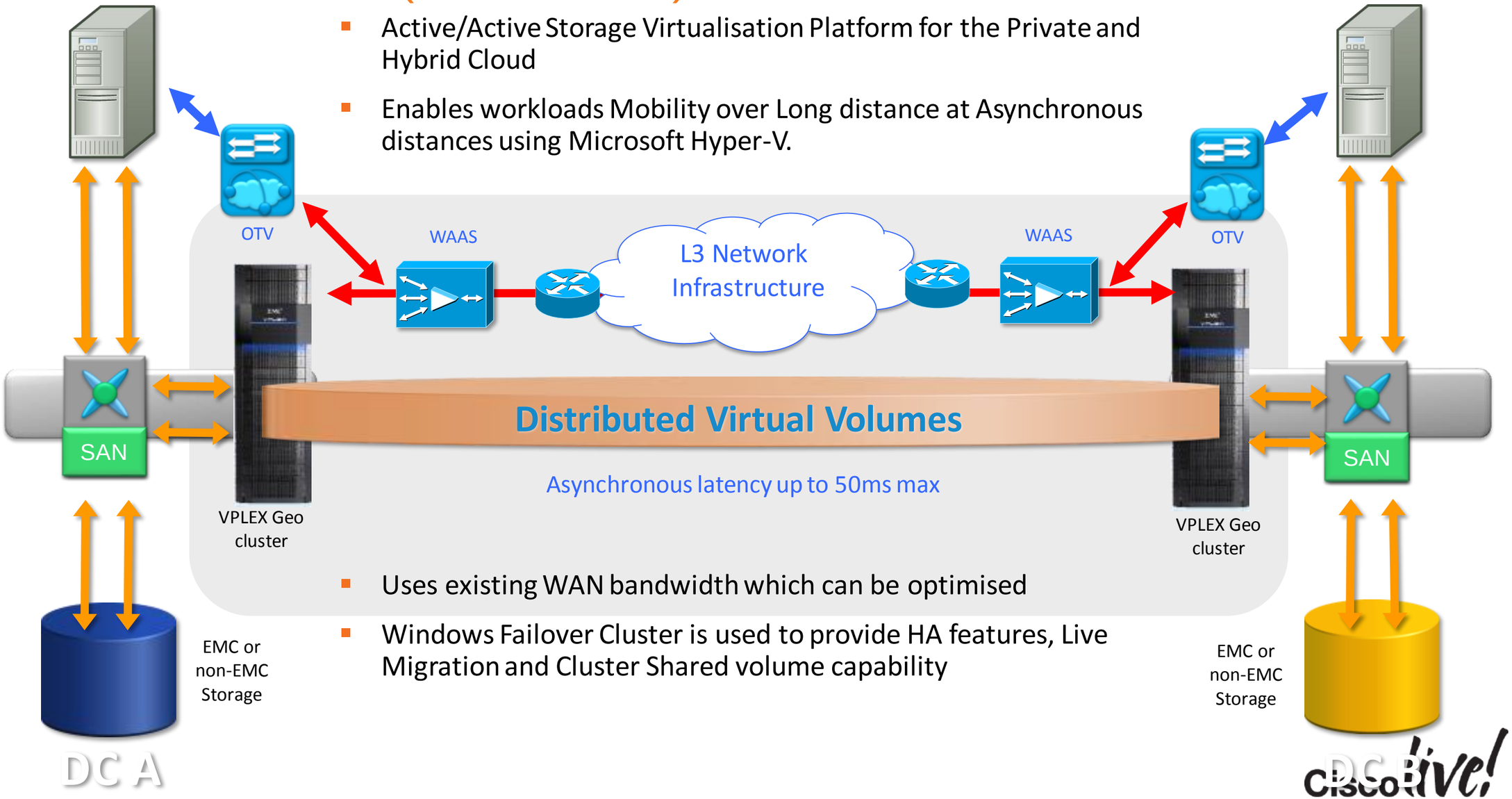
Option 3 - EMC VPLEX Metro (Active/Active)



http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DCI/4.0/_EMC/dciEmc.html

Storage Deployment in DCI

Option 4 - EMC VPLEX Geo (Active/Active)



- Active/Active Storage Virtualisation Platform for the Private and Hybrid Cloud
- Enables workloads Mobility over Long distance at Asynchronous distances using Microsoft Hyper-V.
- Uses existing WAN bandwidth which can be optimised
- Windows Failover Cluster is used to provide HA features, Live Migration and Cluster Shared volume capability

http://www.emc.com/collateral/hardware/white-papers/h8214-application-mobility-vplex-geo-wp.pdf

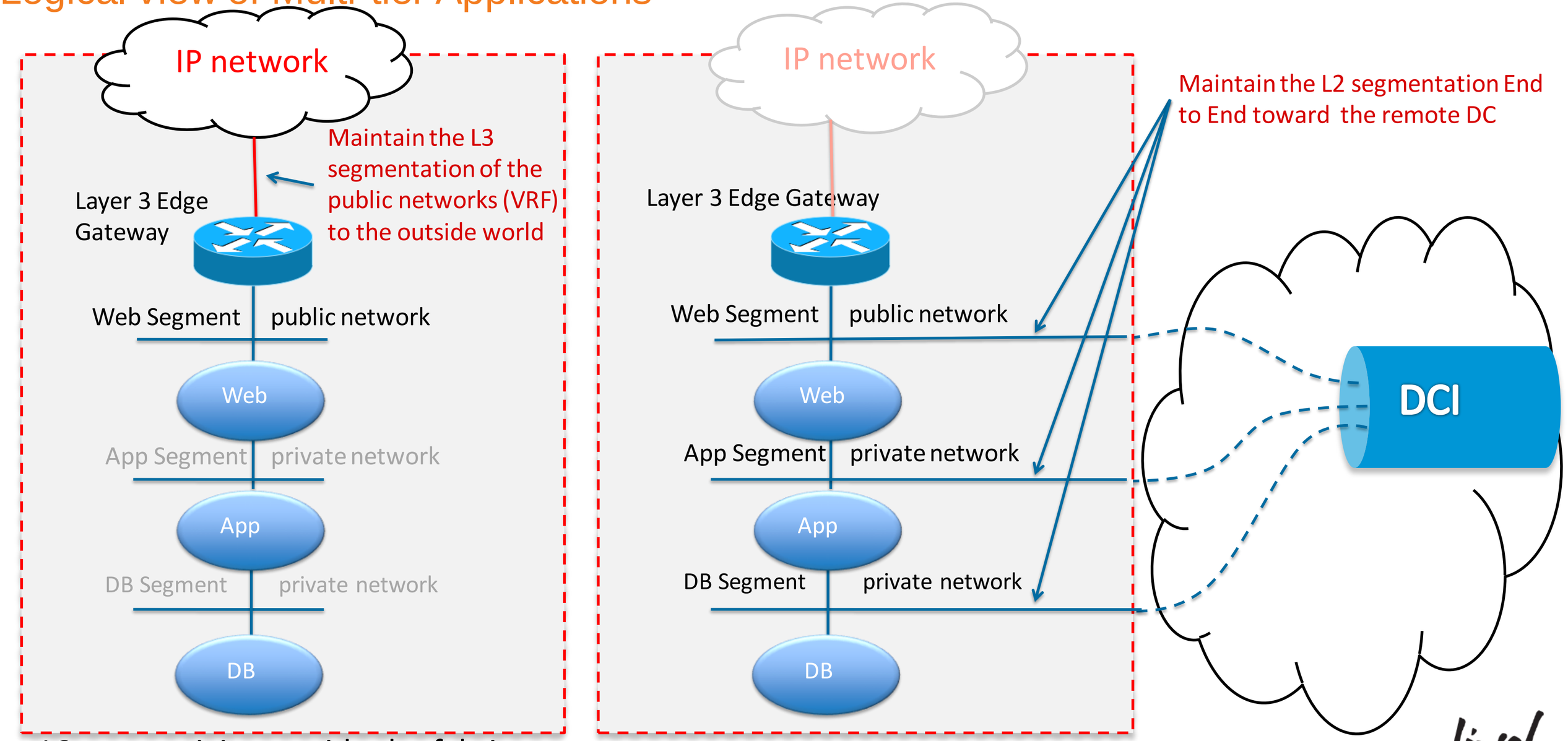
Agenda

- Active-Active (A/A) Data Centre:
  - Market & Business Drivers
  - Terminology, Criticality levels and Solutions Overview
- A/A Data Centre Design Considerations:
  - Storage Extension
  - Data Centre Interconnect (DCI) – L2 & L3 scenarios
- A/A Metro Data Centres Designs

  - Network Services and Applications (Path optimisation)
- Cisco ACI and Active / Active Data Centre
- Q&A

Extending Virtual Tenant Space Outside the Fabric

Logical view of Multi-tier Applications



LAN Extension with DCI

VLAN Types

✓ Type T0

Limited to a single access layer device

✓ Type T1

Extended inside an aggregation block (POD)

✓ Type T2

Extended between PODs part of the same DC site

✓ Type T3

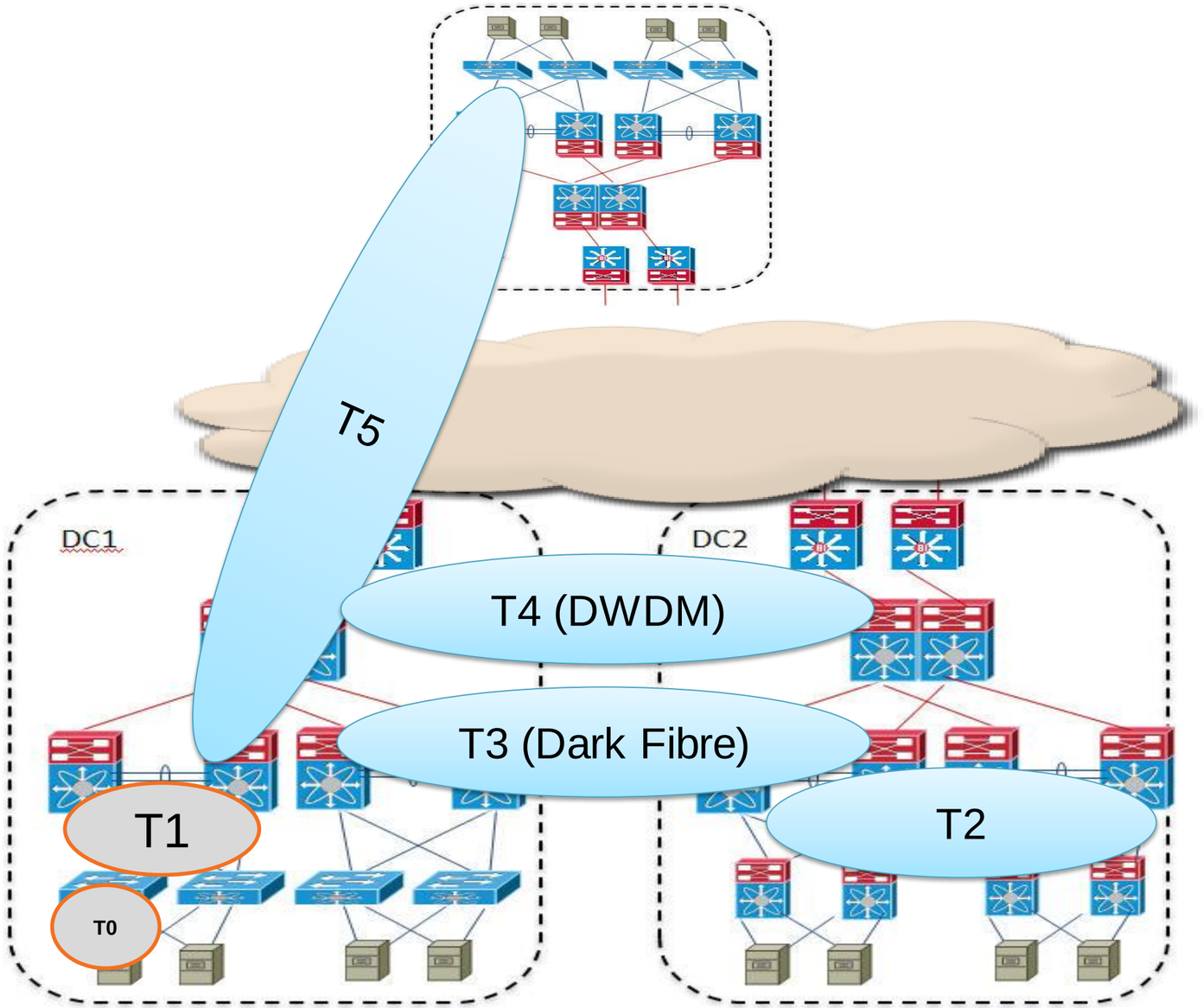
Extended between PODs part of twin DC sites connected via dedicated dark fibre links

✓ Type T4

Extended between PODs part of twin DC sites connected via xWDM links

✓ Type T5

Extended between PODs part of distant remote DC sites



LAN Extension for DCI

Technology Selection

Ethernet

Over dark fibre or protected D-WDM

- *VSS & vPC*
  - Dual site interconnection
- *FabricPath*
  - Multiple site interconnection

MPLS

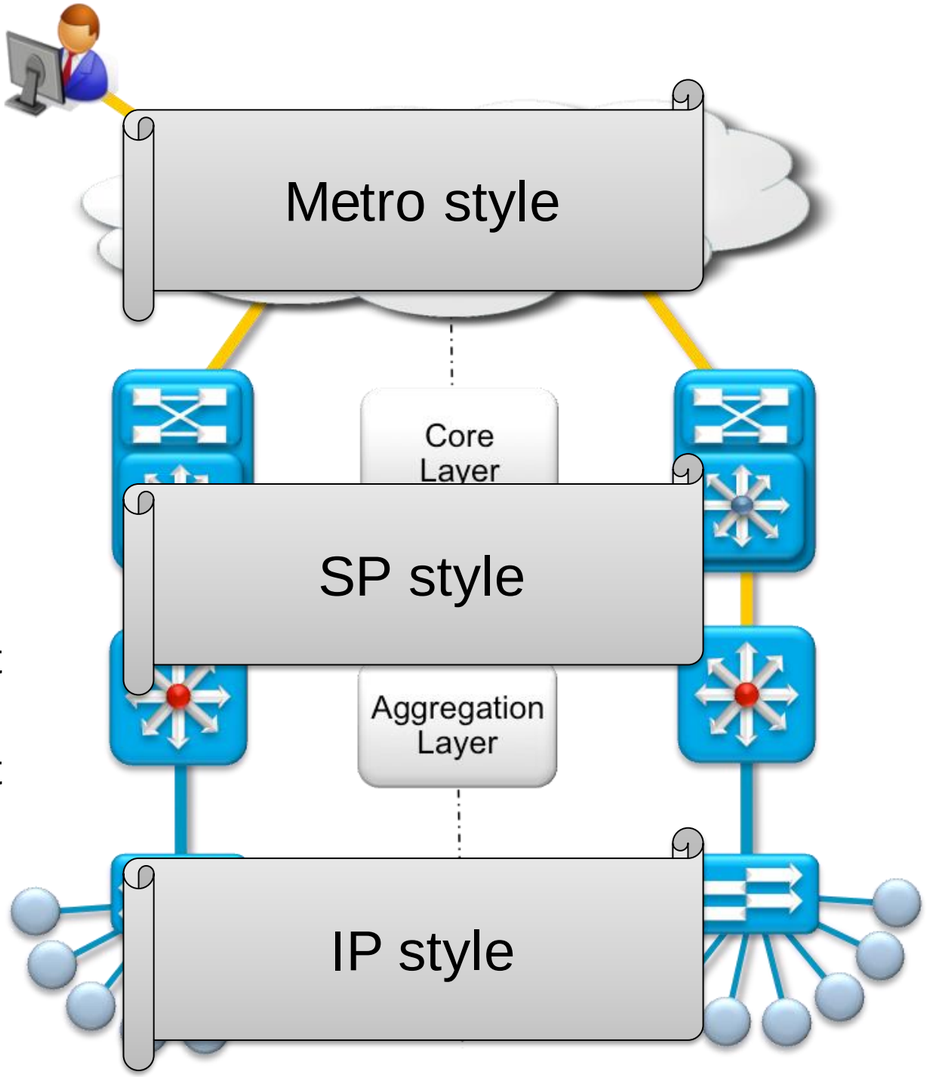
MPLS Transport

- *EoMPLS*
  - Transparent point to point
- *VPLS*
  - Large scale & Multi-tenants, Point to Multipoint
- *E-VPN*
  - Large scale & Multi-tenants, Point to Multipoint

IP

IP Transport

- *OTV*
  - Enterprise style Inter-site MAC Routing
- *LISP*
  - For Subnet extension and Path Optimisation
- *VXLAN (future for DCI)*
  - Emerging limited A/A site interconnect (requires anycast gateway)



Dual Sites Interconnection

Leveraging MECs Between Sites

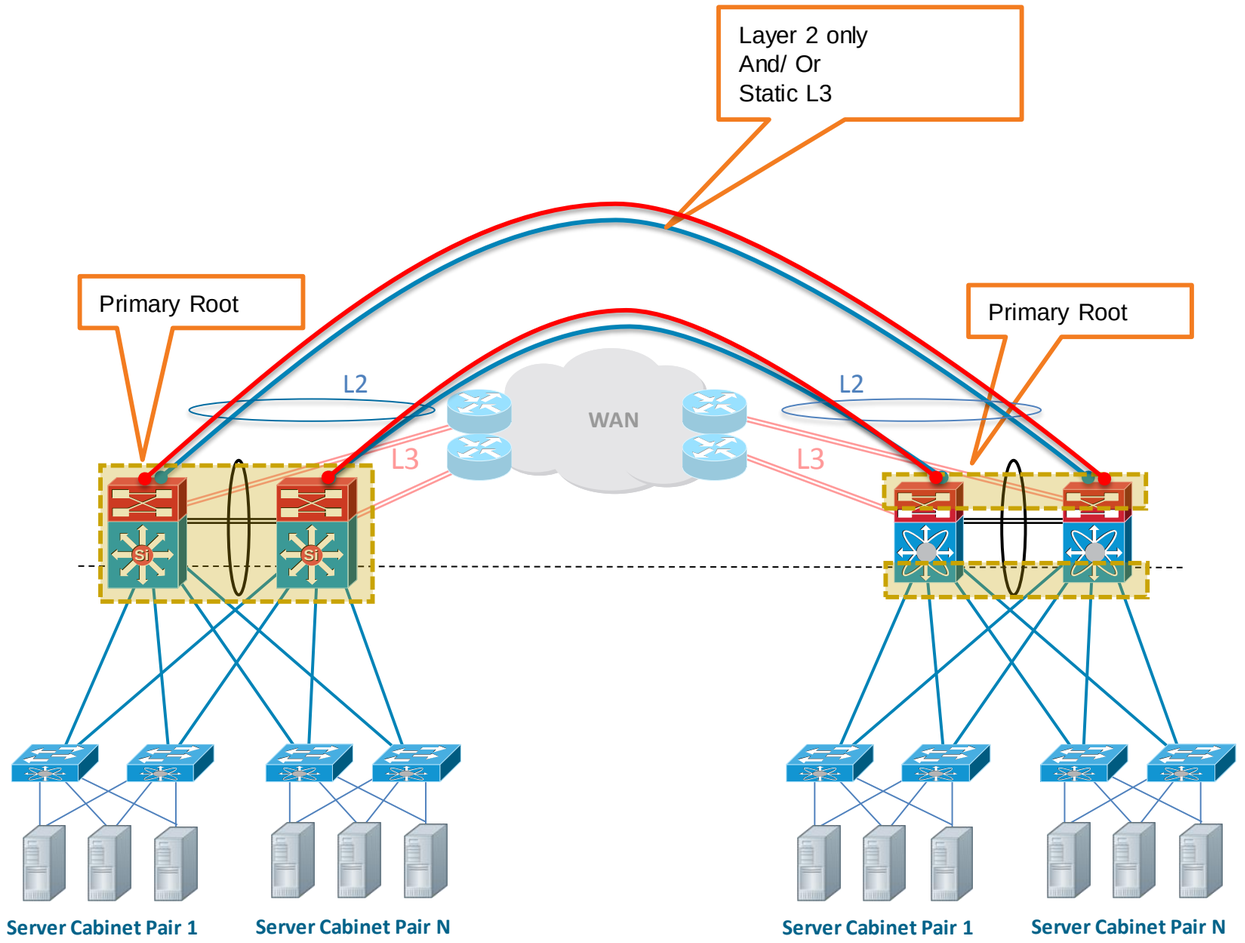
At DCI point:
- STP Isolation (BPDU Filtering)
- Broadcast Storm Control
- FHRP Isolation

- Link utilisation with Multi-Chassis EtherChannel
- Enable Fast LACP with DWDM
- DCI port-channel
  - 2 or 4 links
- Requires protected DWDM or Direct fibres

- Validated design:
  - *200 Layer 2 VLANs + 100 VLAN SVIs*
  - *1000 VLAN + 1000 SVI (static routing)*

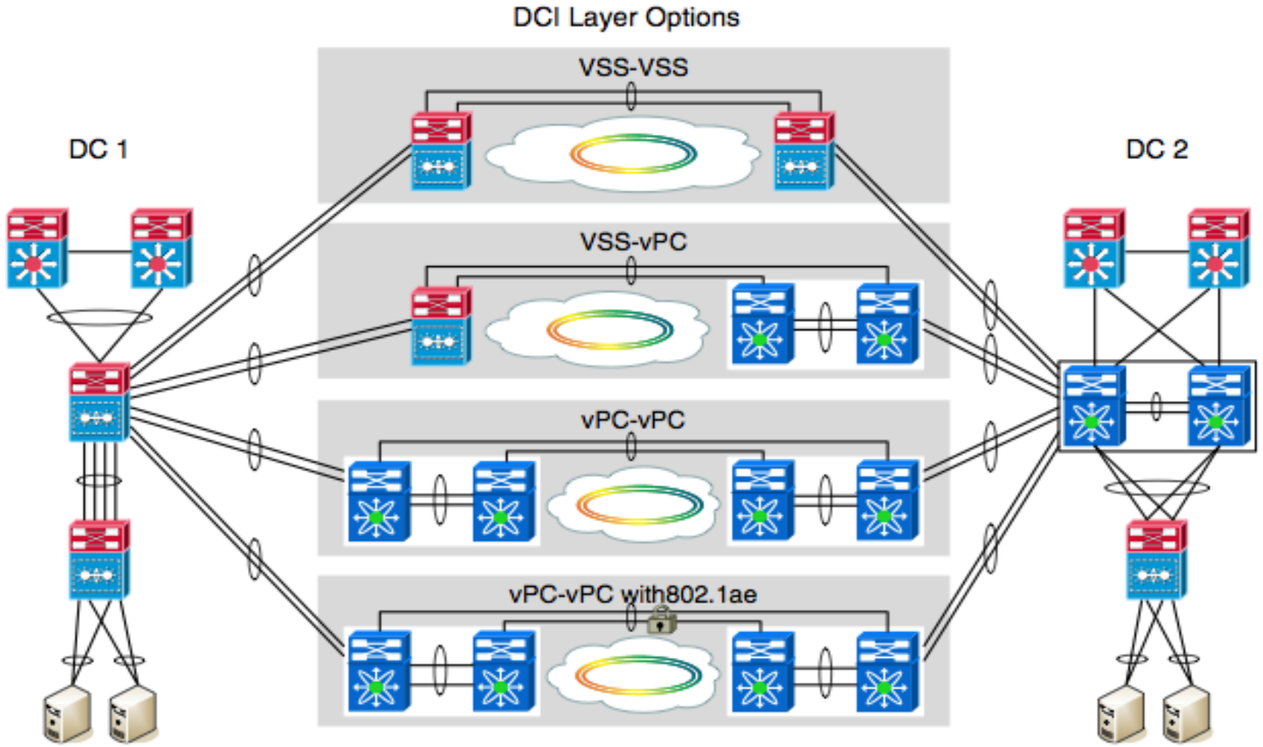
- Check if vPC supports L3 peering on your Nexus device
  - If possible, use dedicated L3 Links for Inter-DC routing!

- Support for L3 peering across all Nexus gear by CY15
  - Requires F2 Line Card (minimum)
  - NX-OS (Roadmap)



Dual Sites Use Case Summary

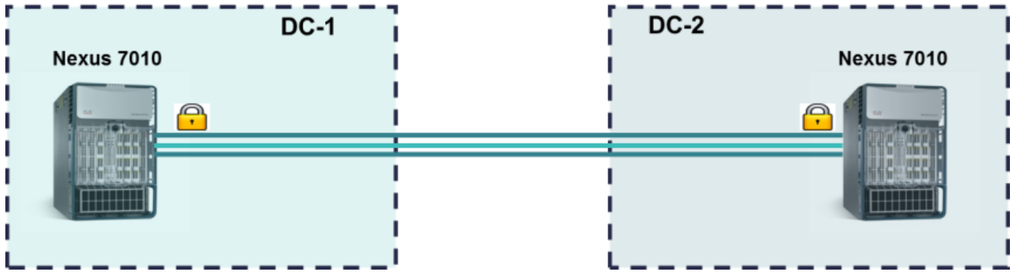
Cisco Validated Design on CCO



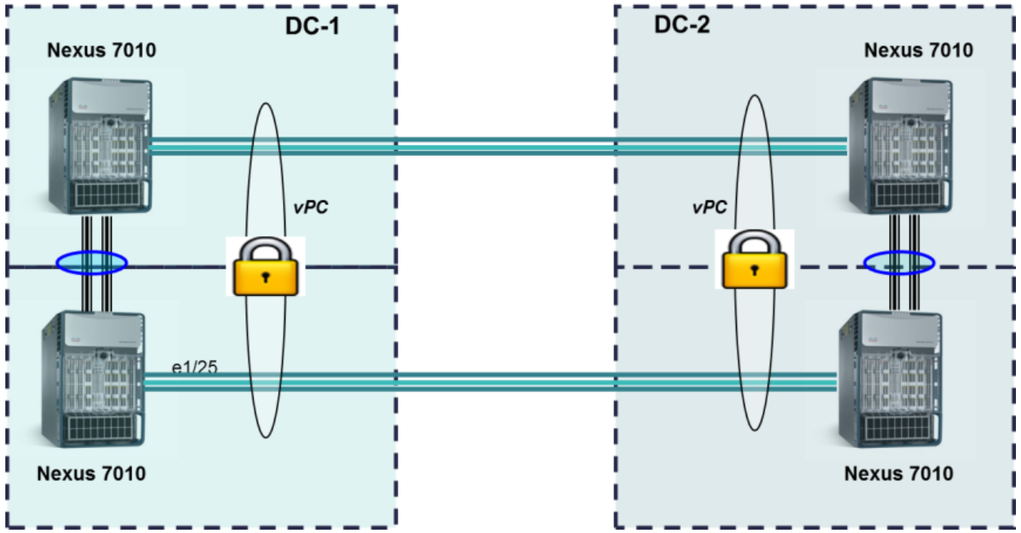
Test Case	Hardware failure Ucast	Hardware failure Mcast	Hardware restore Ucast	Hardware restore Mcast	Link Failure Ucast	Link failure Mcast	Link Restore Ucast	Link Restore Mcast
VSS-VSS	<1.7	<2.3	<1.1	<2.8	<1.3	<1.2	<1.7	<1.2
VSS-vPC	<1.3	<1.7	<2.0	<2.6	<1.2	<1.6	<1.5	<1.4
vPC-vPC	<1.5	<1.6	<2.8	<2.5	<1.2	<0.2	<0.2	<0.2

MACSec for Secure Data Centre Interconnect

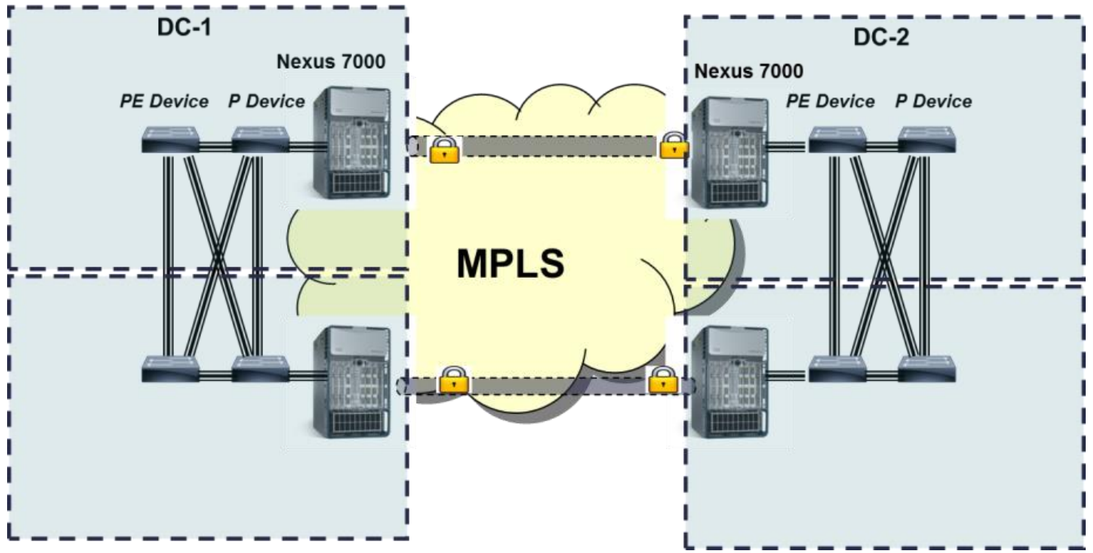
Single Access dark Fibre Connectivity



Dual Access with dark Fibre Connectivity



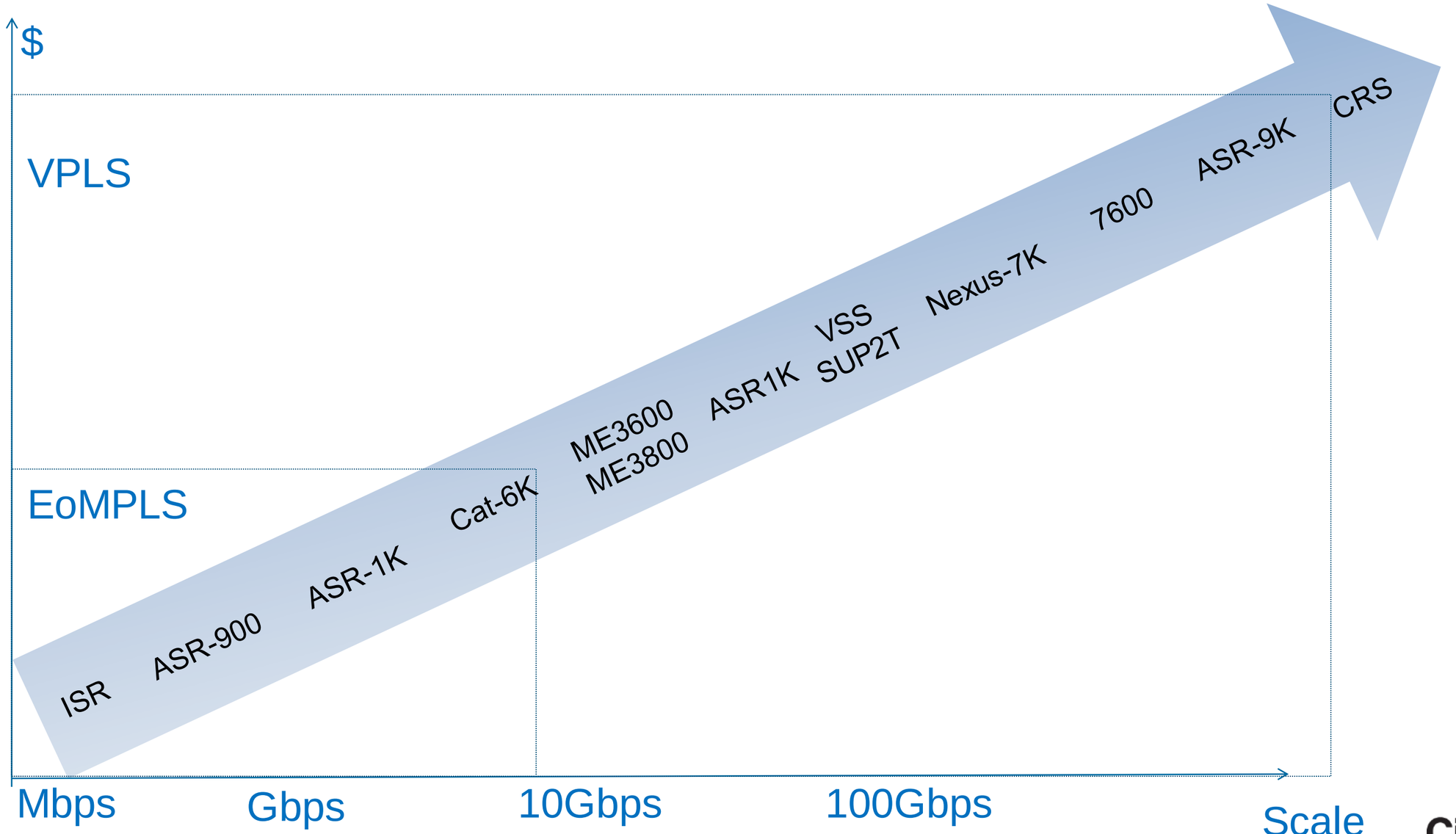
7Ks as bulk encrypters for Self managed MPLS DCI Cores – Bump in the wire



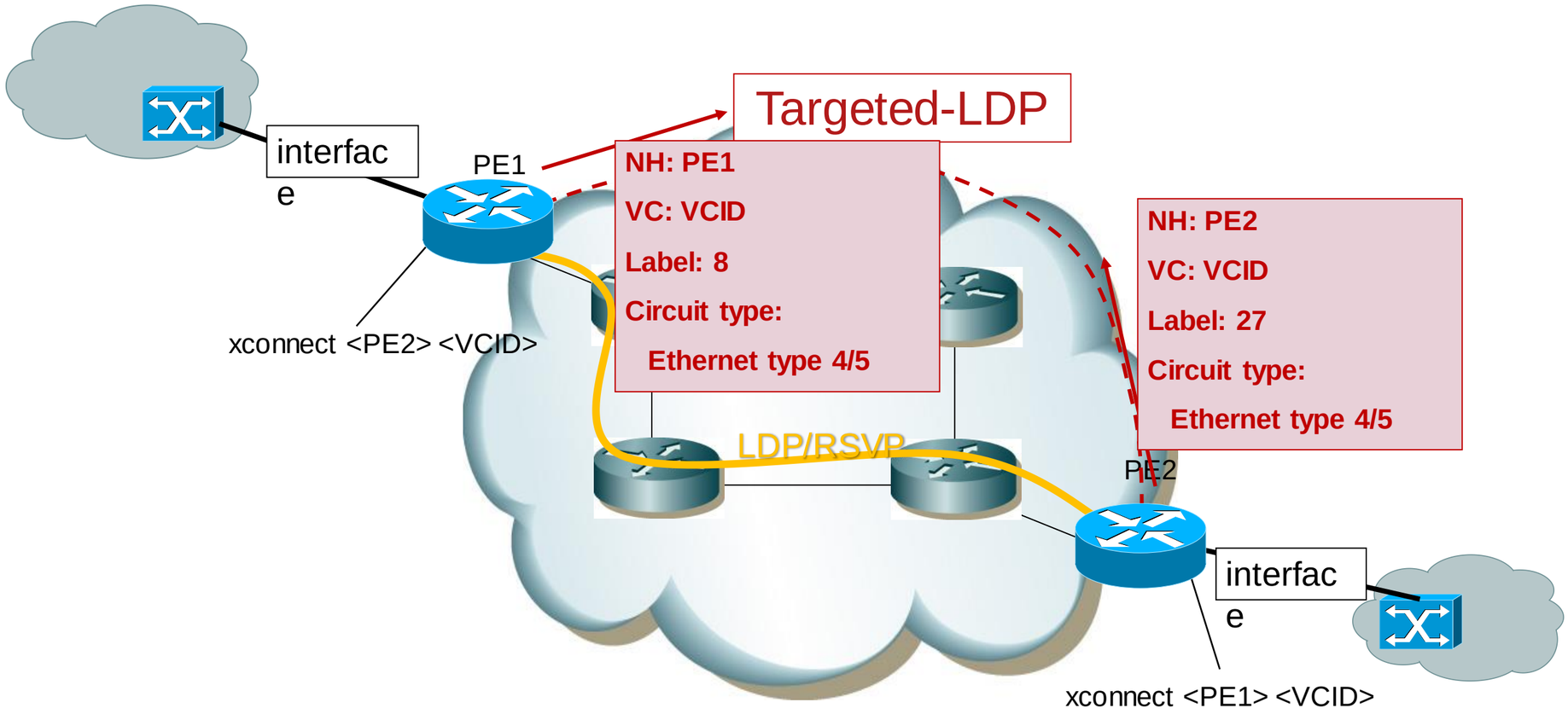
No support of MACSec on F3 Modules 40GE/100GE or Nexus 9000. Nexus 7000 M2 & Future M3 modules support 10G / 40G / 100G MACSec

MPLS for DCI

Large Choice of Devices



EoMPLS
Control Plane

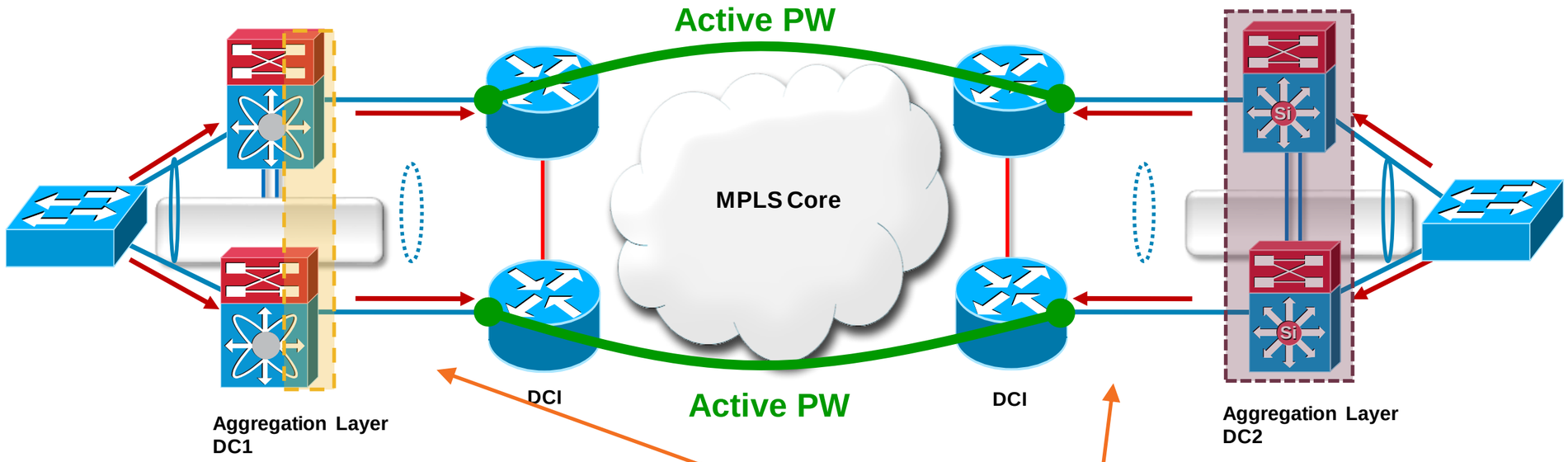


Core transport relies on
LDP + IGP (OSPF/ISIS)
or
RSVP for Traffic-Engineering

Xconnect can be applied to
- LAN Port
- LAN Sub-interface
- SVI

EoMPLS Usage with DCI

End-to-End Loop Avoidance Using Edge to Edge LACP



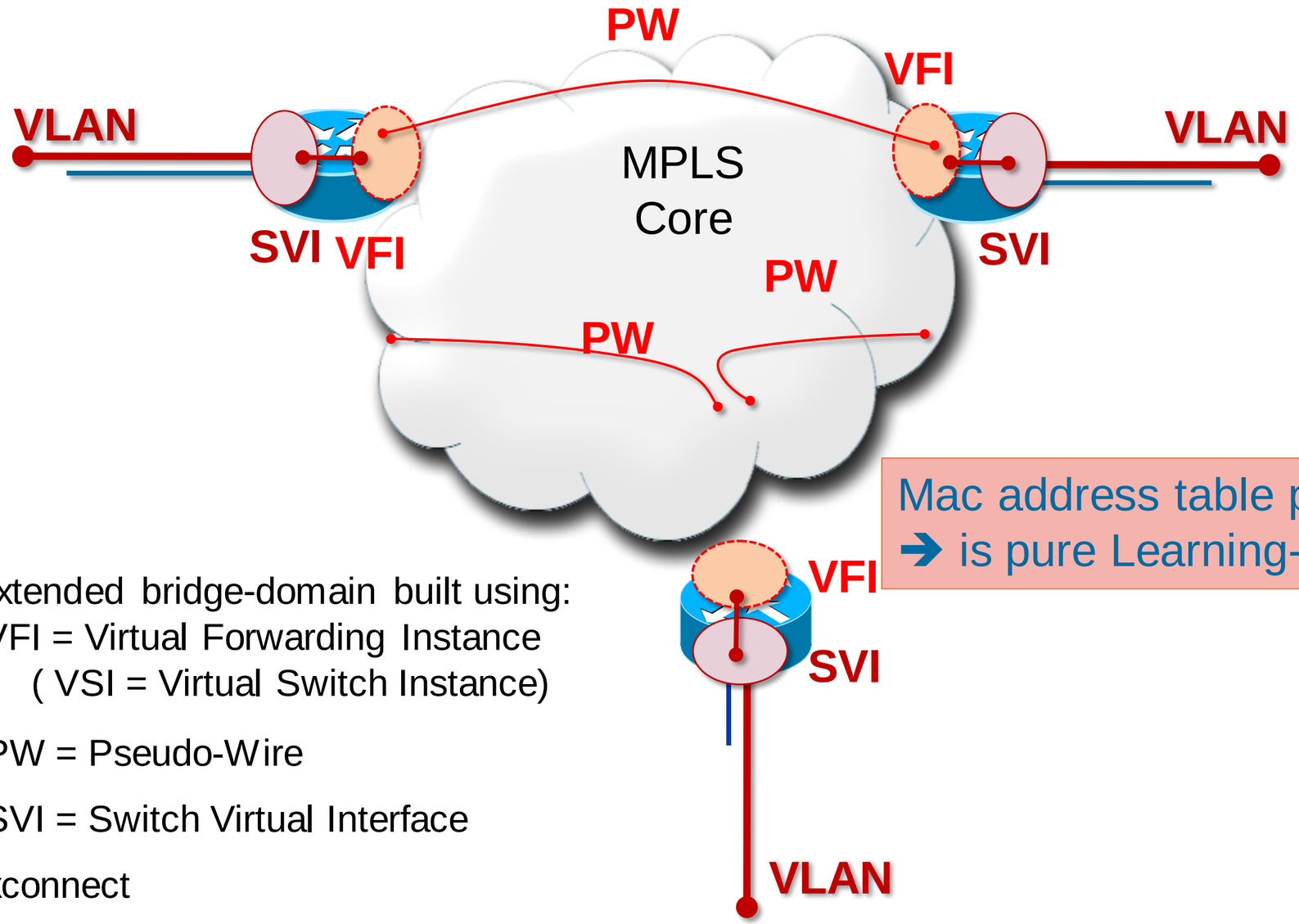
- BPDU Filtering to maintain STP domains isolation
- Storm-control for data-plane protection
- Configuration applied at aggregation layer on the logical port-channel interface

```
interface port-channel70
  description L2 PortChannel to DC 2
  spanning-tree port type edge trunk
  spanning-tree bpdufilter enable
  storm-control broadcast level 1*
  storm-control multicast level x
```

*Value to be tuned, min is 0.3

Multi-Point Topologies

What Is VPLS?



Mac address table population
➔ is pure Learning-Bridge

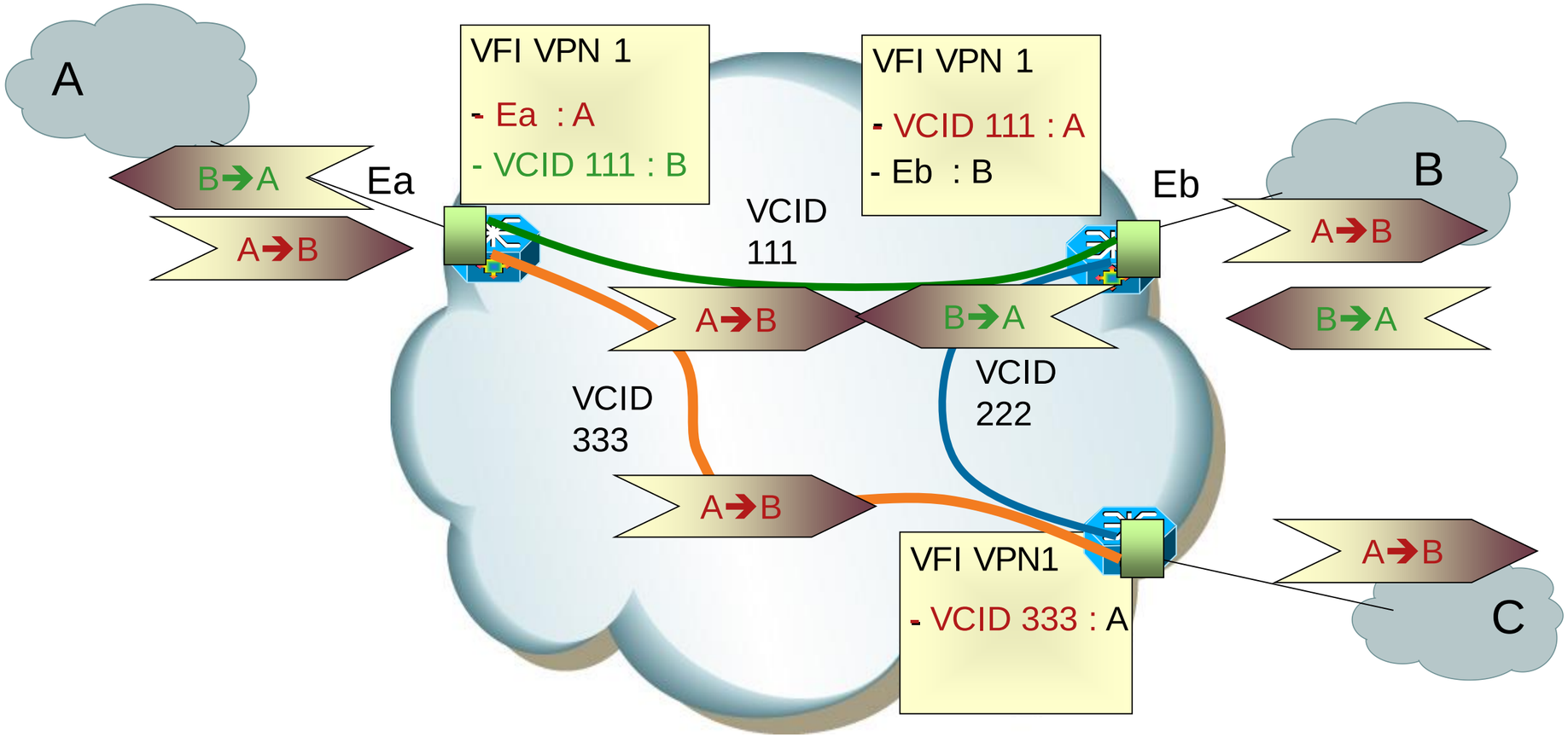
One extended bridge-domain built using:

- VFI = Virtual Forwarding Instance
  (VSI = Virtual Switch Instance)
- PW = Pseudo-Wire
- SVI = Switch Virtual Interface
- xconnect

VPLS

L2 Signalling and Forwarding (aka Transparent-Bridging)

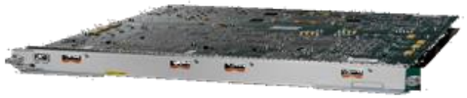
a VSI/VFI operates like a conventional L2 switch!



VPLS Cluster Solutions

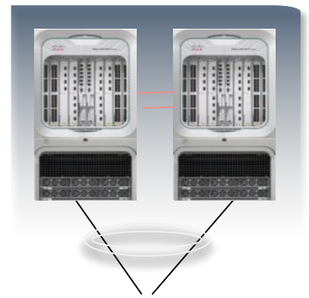
- Using clustering mechanism
  - –Two devices in fusion as one
    - VSS Sup720
    - VSS Sup2T
    - ASR9K nV virtual cluster
    - ➔ One control-plane / two data-planes
  - Dual node is acting as one only device
    - Native redundancy (SSO cross chassis)
    - Native load balancing
    - Capability to use port-channel as attachment circuit

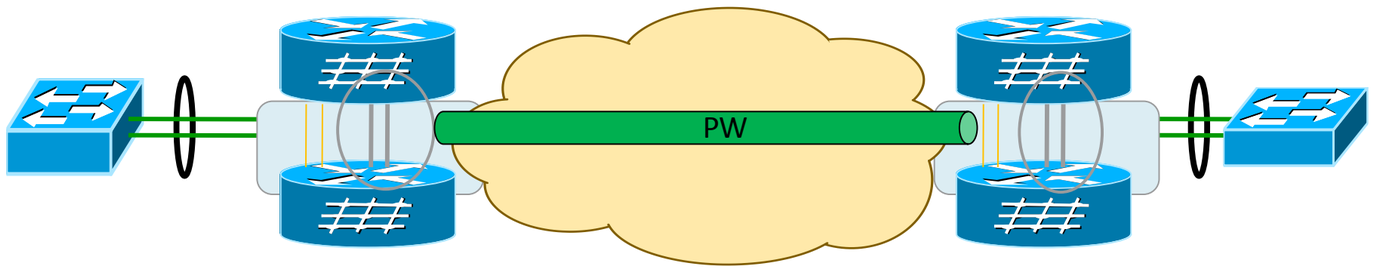
SUP720+ES



SUP2T

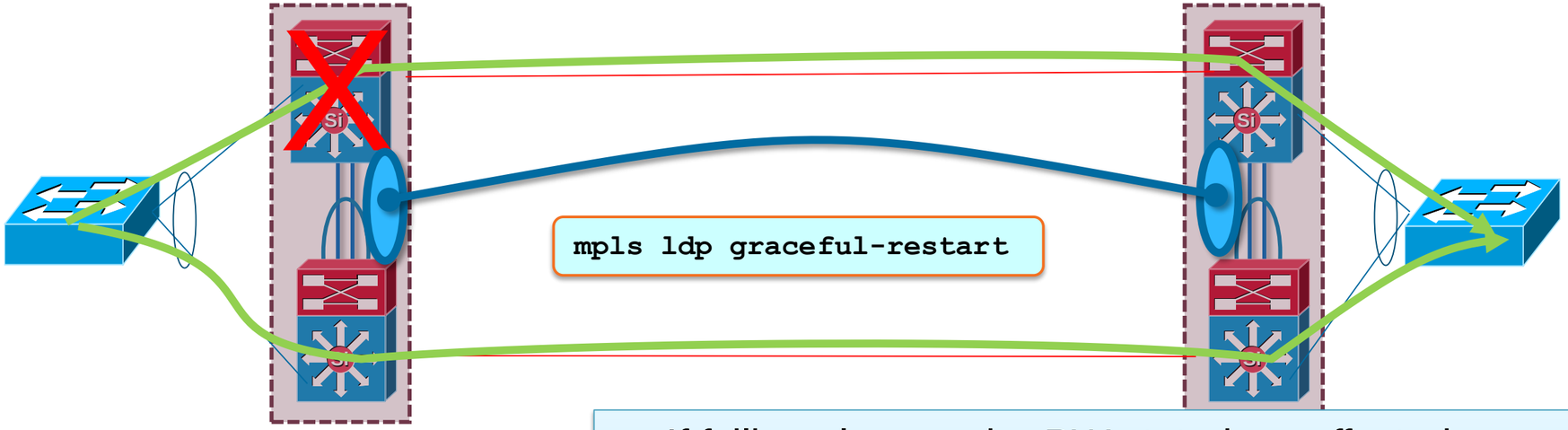
ASR9K nV





Cluster VPLS – Redundancy

Making Usage of Clustering

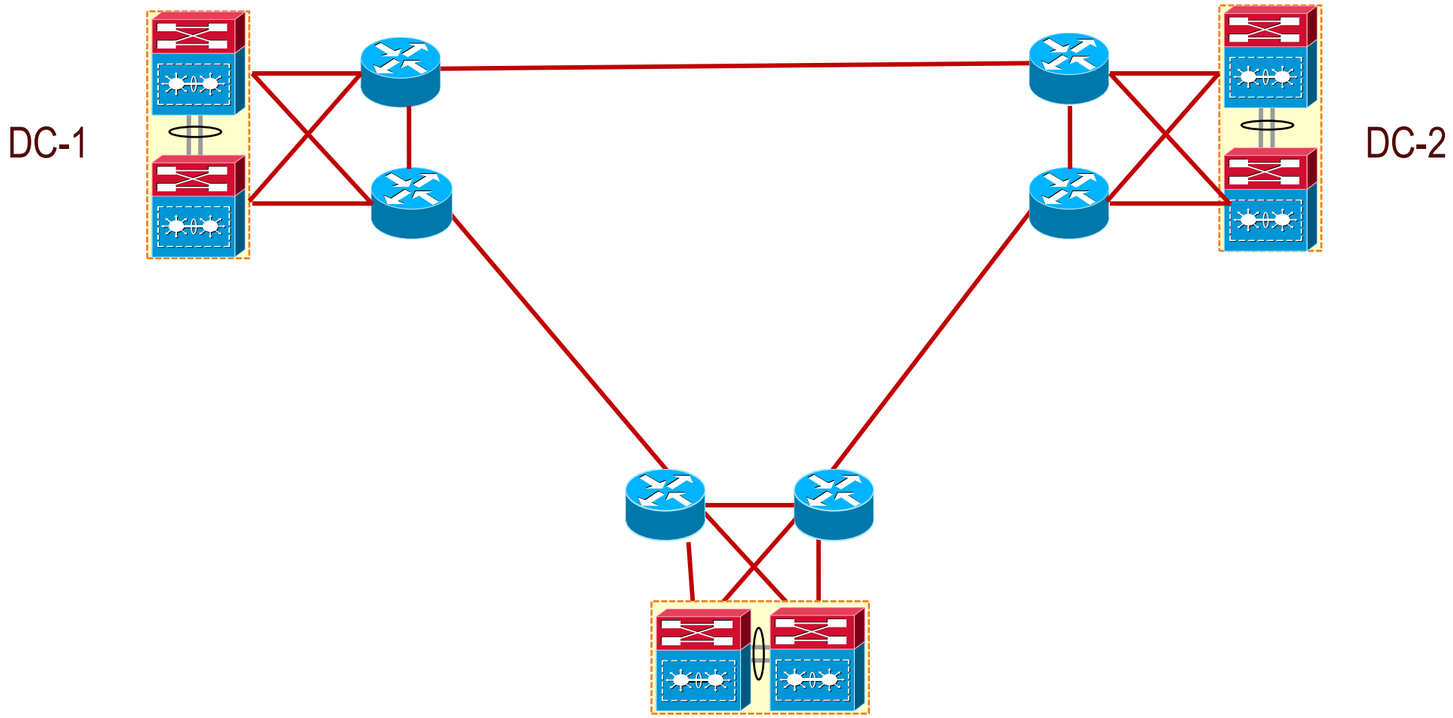


VSS		Failover (msec)	Fallback (msec)
Bridged traffic	➔	224	412
	➔	326	316

- If failing slave node: PW state is unaffected
- If failing master node:
  - PW forwarding is ensured via SSO
  - PW state is maintained on the other side using Graceful restart
- Edge Ether-channel convergence in sub-second
- Traffic is directly going to working VSS node
- Traffic exits directly from egress VSS node
- Quad sup SSO for SUP2T since 1QCY13

VPLS Cluster - Deployment Consideration

ECMP Core Requirements

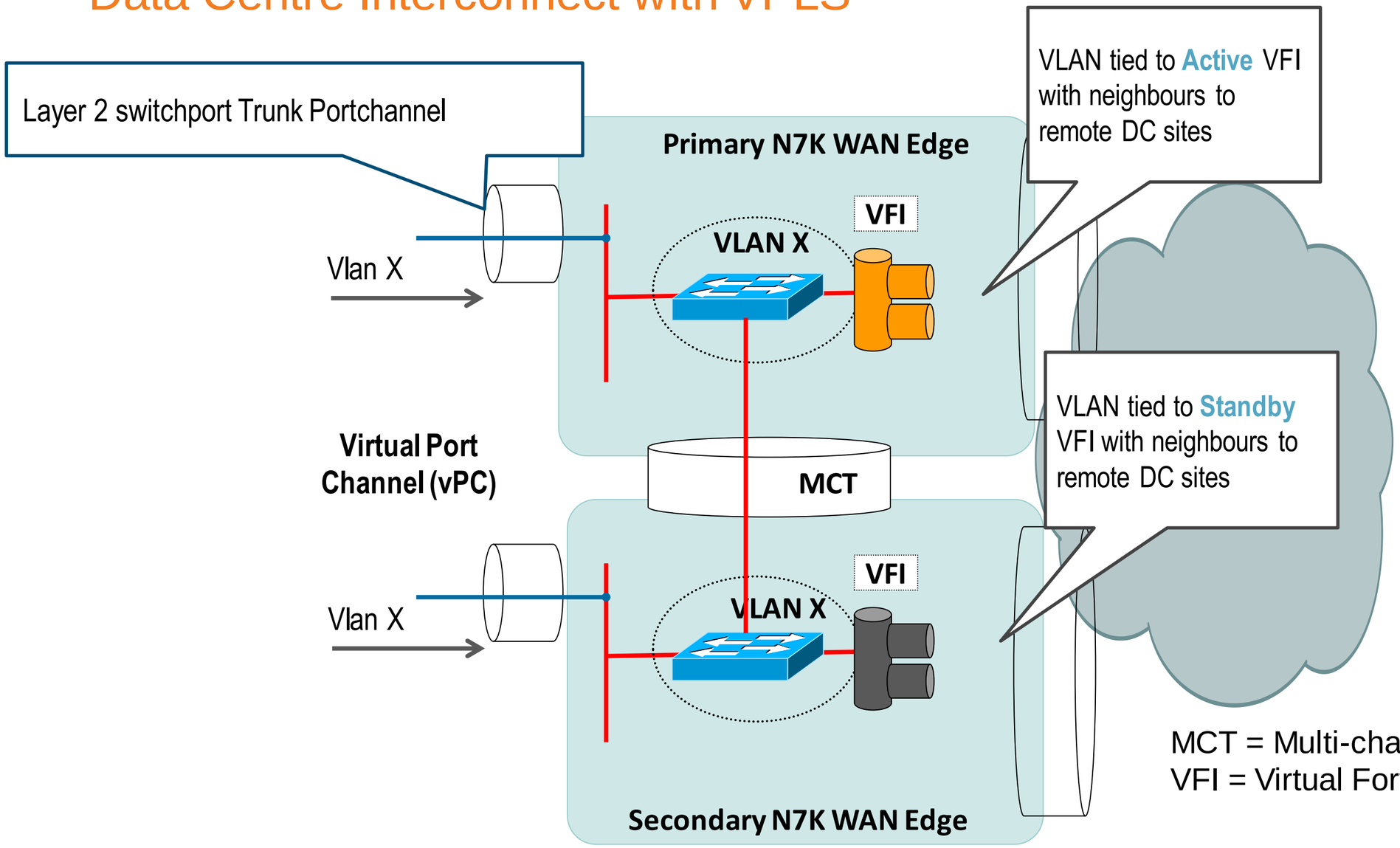


Build a symmetric core with two ECMP paths between each VSS

Nexus 7000

Data Centre Interconnect with VPLS

Since 6.2(2)



MCT = Multi-chassis Trunk Interface
VFI = Virtual Forwarding Instance

Nexus 7000 - Data Centre Interconnect with VPLS

Sample Configuration – Nexus 7000

PE 1

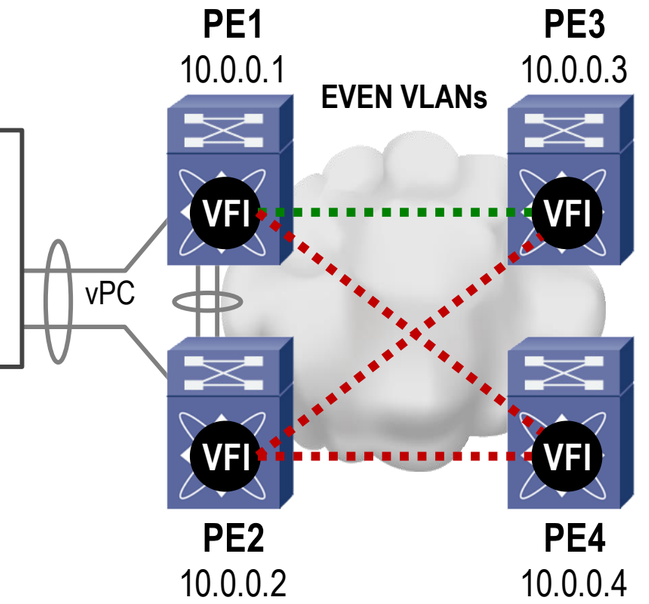
- Primary VFI owner for EVEN vlans
- Secondary owner for ODD vlans

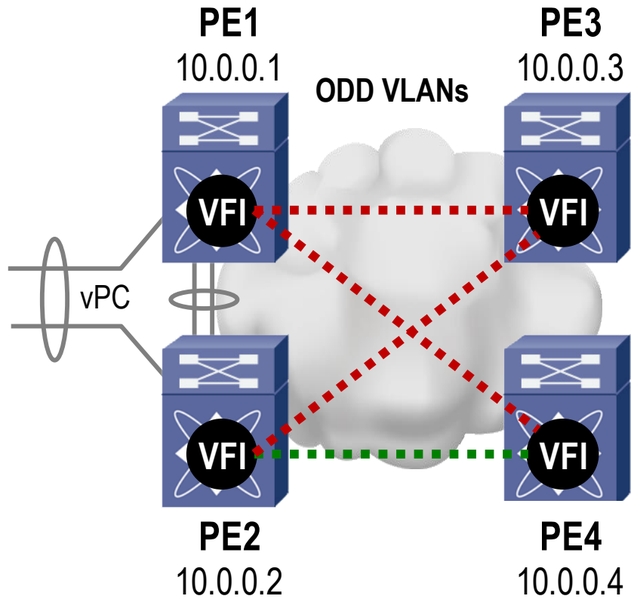
```
vlan 80-81
!
vlan configuration
 member vfi vpls-80
!
vlan configuration 81
 member vfi vpls-81
!
l2vpn vfi context vpls-80
  vpn id 80
  redundancy primary
  member 10.0.0.3 encapsulation mpls
  member 10.0.0.4 encapsulation mpls
!
l2vpn vfi context vpls-81
  vpn id 81
  redundancy secondary
  member 10.0.0.3 encapsulation mpls
  member 10.0.0.4 encapsulation mpls
!
interface port-channel50
 switchport mode trunk
 switchport trunk allowed vlan 80,81
```

PE 2

- Primary VFI owner for ODD vlans
- Secondary owner for EVEN vlans

```
vlan 80-81
!
vlan configuration 80
 member vfi vpls-80
!
vlan configuration 81
 member vfi vpls-81
!
l2vpn vfi context vpls-80
  vpn id 80
  redundancy secondary
  member 10.0.0.3 encapsulation mpls
  member 10.0.0.4 encapsulation mpls
!
l2vpn vfi context vpls-81
  vpn id 81
  redundancy primary
  member 10.0.0.3 encapsulation mpls
  member 10.0.0.4 encapsulation mpls
!
interface port-channel50
 switchport mode trunk
 switchport trunk allowed vlan 80,81
```

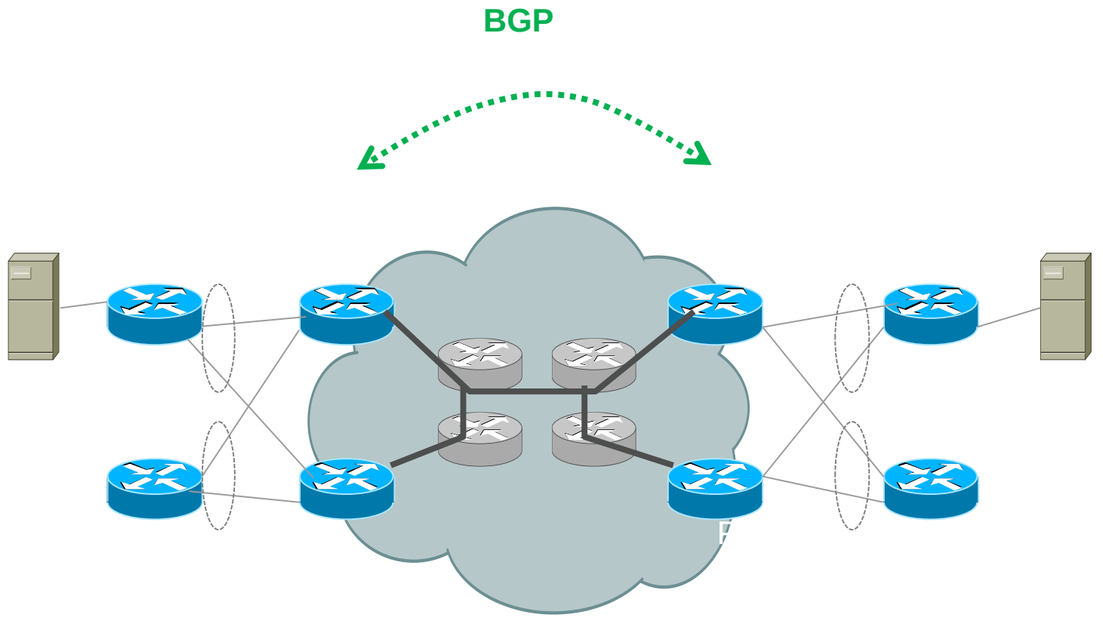




Note: Virtual Port Channel (vPC) configuration not shown

E-VPN
Main Principles

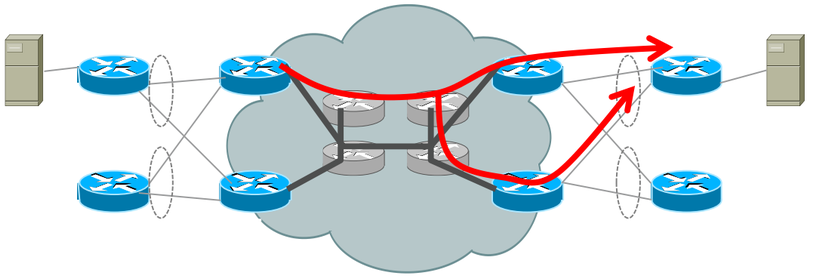
- Control-Plane Distribution of Customer MAC-Addresses using BGP
  - PE continues to learn C-MAC over AC
  - When multiple PEs announce the same C-MAC, hash to pick one PE
- MP2MP/P2MP LSPs for Multicast Traffic Distribution
- MP2P (like L3VPN) LSPs for Unicast Distribution
- Full-Mesh of PW no longer required !!



Solution Overview
(draft-ietf-l2vpn-pbb-evpn-)

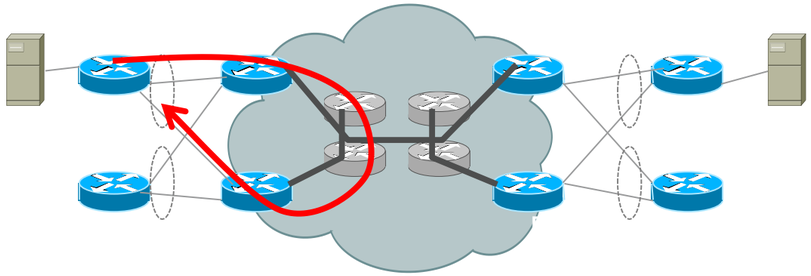
DF Election with VLAN Carving

- Prevent duplicate delivery of flooded frames.
- Uses BGP **Ethernet Segment Route**.
- Performed **per Segment** rather than per (VLAN, Segment).
- Non-DF ports are blocked for flooded traffic (multicast, broadcast, unknown unicast).



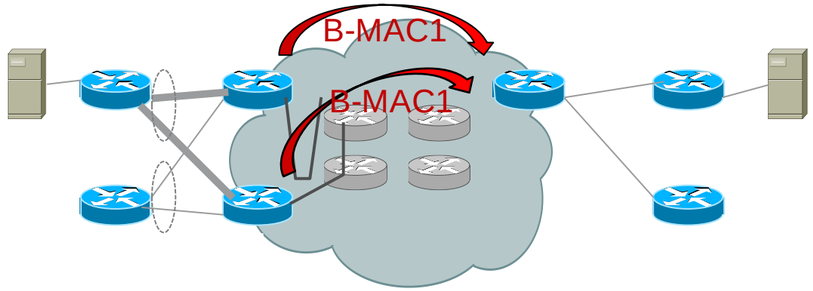
Split Horizon for Ethernet Segment

- Prevent looping of traffic originated from a multi-homed segment.
- Performed **based on B-MAC** source **address** rather than ESI MPLS Label.



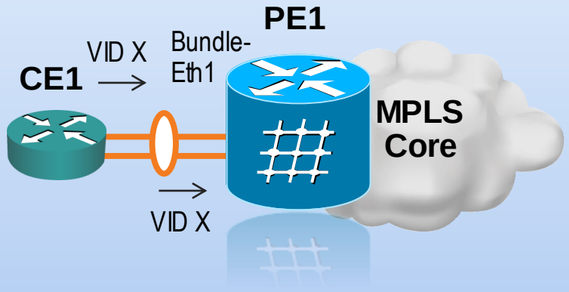
Aliasing

- PEs connected to the same multi-homed Ethernet Segment advertise the **same** B-MAC address.
- Remote PEs use these **MAC Route advertisements for aliasing** load-balancing traffic destined to C-MACs reachable via a given B-MAC.



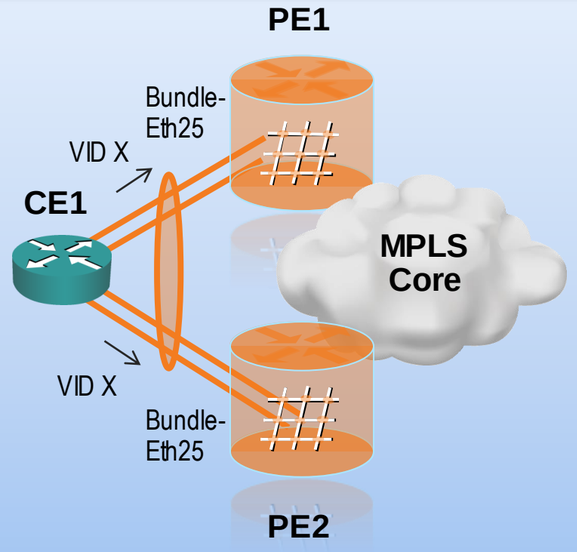
Supported Access Topologies

Single Home Device (SHD) Single Home Network (SHN)



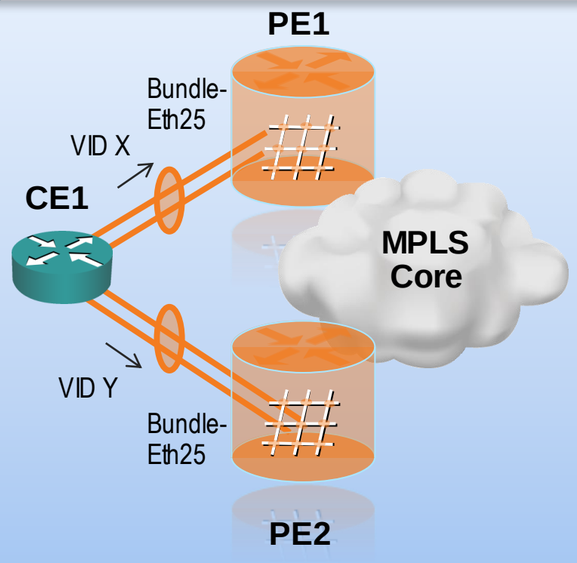
- Physical Interfaces
- Bundle Interfaces (shown)

Dual Home Device (DHD) Active / Active Per-Flow LB



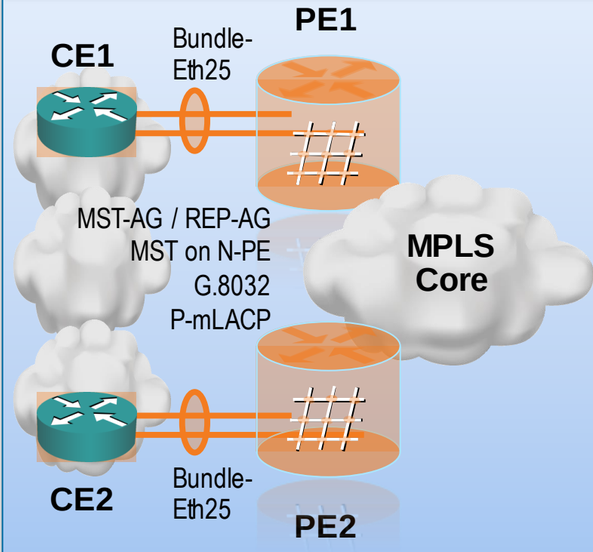
- Bundle Interface

Dual Home Device (DHD) Active / Active Per-Service LB



- Physical Interfaces
- Bundle Interfaces (shown)

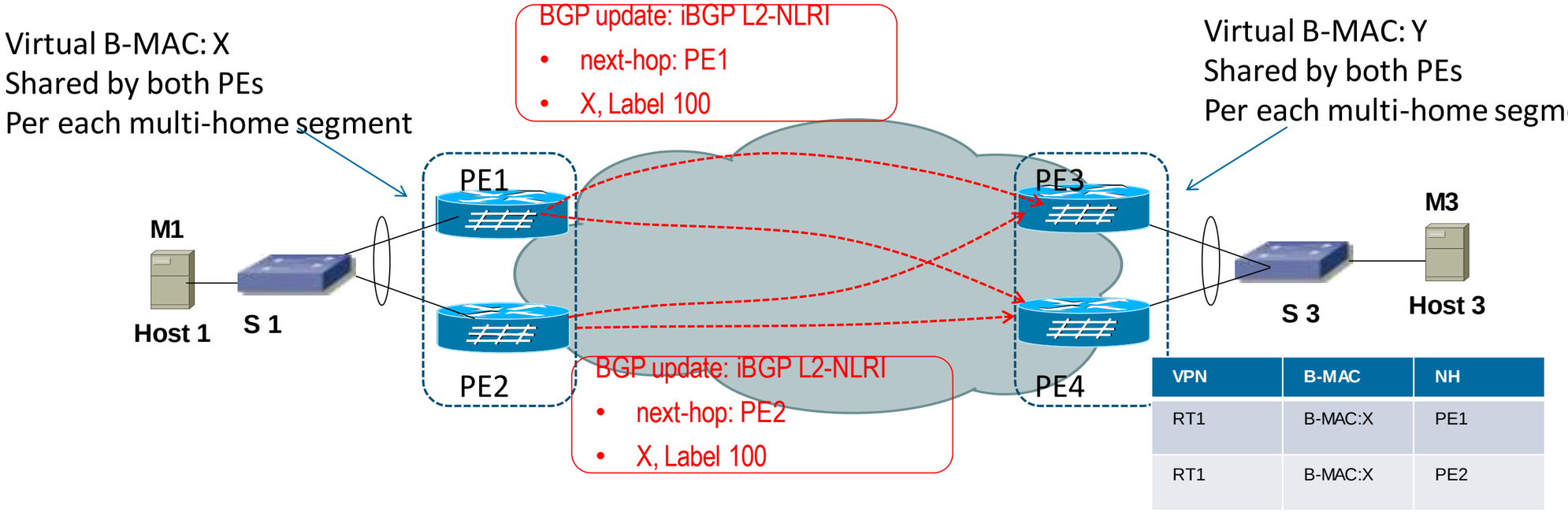
Dual Home Network (DHN) Active / Active Per-Service LB



- Physical Interfaces
- Bundle Interfaces (shown)

PBB-EVPN Principle (Provider Backbone Bridge)

BGP MAC Routing

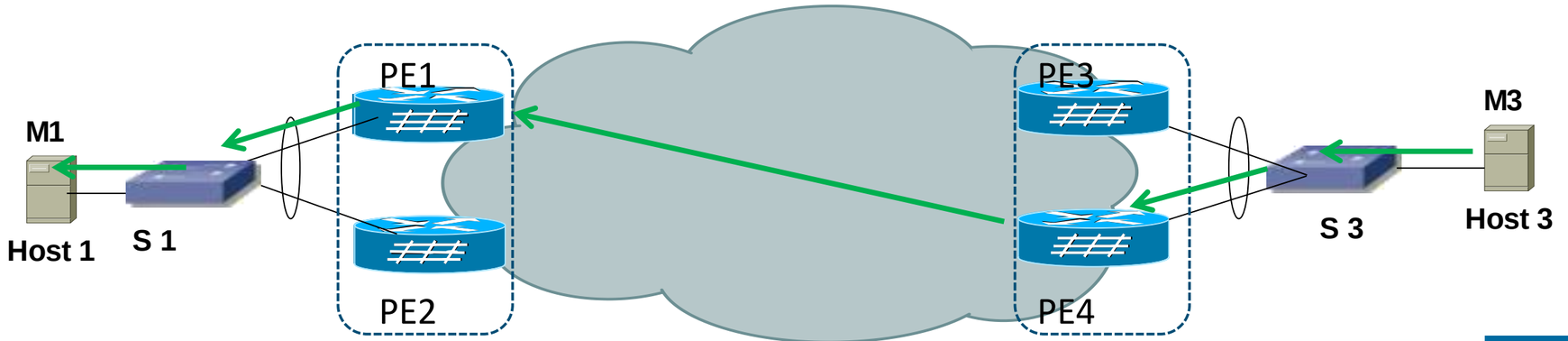


VPN	B-MAC	NH
RT1	B-MAC:X	PE1
RT1	B-MAC:X	PE2

- PE1 and PE2 share the same virtual B-MAC (: X), for the same multi-homing segment. In case of mc-lag, PE can learn the virtual B-MAC according to Switch 1 LACP system MAC automatically
- Both PE1 and PE2 advertise virtual B-MAC to the remote PEs via BGP
- Remote PE3 and PE4 receive the BGP route, and install the virtual B-MAC in its L2FIB table. BGP policy could be used for active/active path or primary/standby path
- PE3 and PE4 does the same thing, PE1 and PE2 learn the virtual B-MAC: Y

PBB-EVPN Principle (Provider Backbone Bridge)

Packet Forwarding



VPN	C-MAC	NH
BD1	M1	
BD1	M2	

B-MAC	NH
B-MAC:X	PE1, PE2
B-MAC:	…

- Host 3 send packet to Host 1, D-MAC=M1
- S3 does per-flow load balancing, and choose PE4
- PE4 look up the C-MAC table, find M1 and its associated B-MAC=X, and the two ECMP next-hop: PE1 and PE2
- PE4 does per-flow load balancing, choose PE1
- PE1 receive the packet, de-cap the PBB, learn the C-MAC: M3 and the B-MAC association
- PE1 look up its C-MAC table, and forward packet to S1

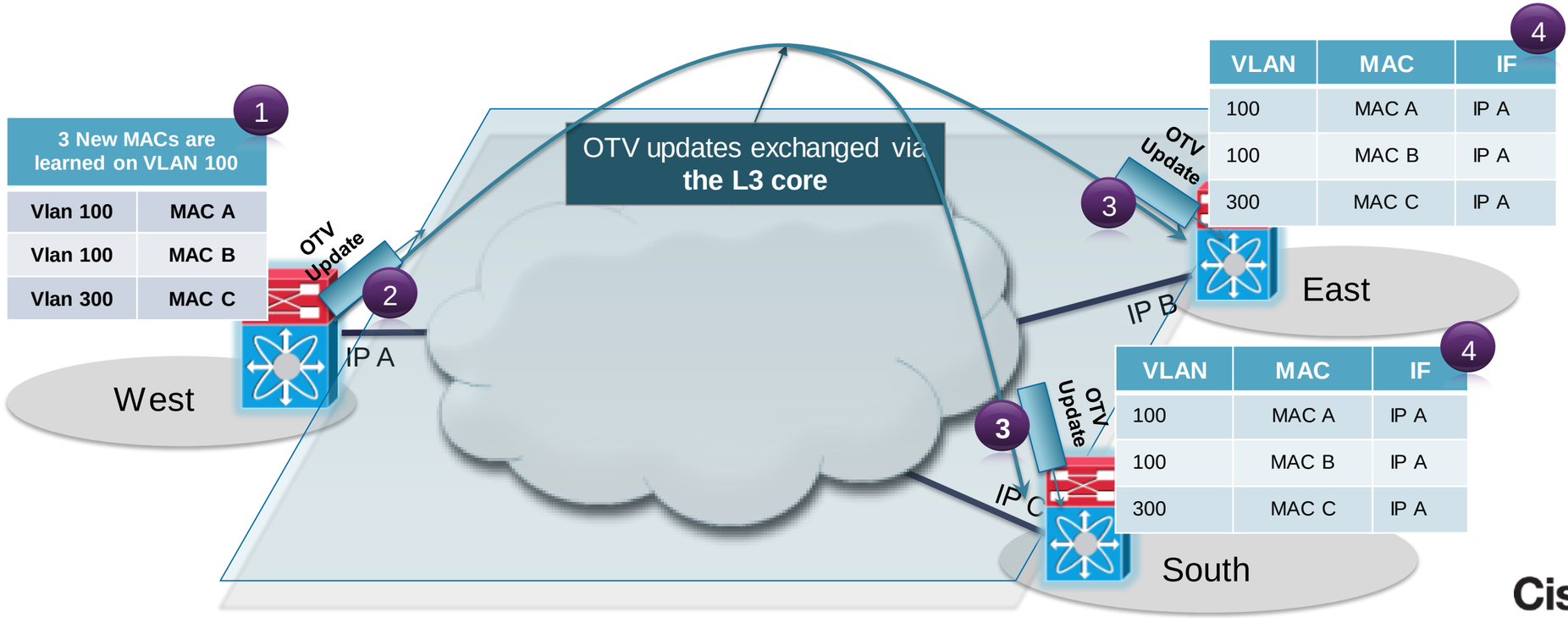
Overlay Transport Virtualisation (OTV) in a Nutshell

- OTV is a MAC-in-IP method that extends Layer 2 connectivity across a transport network infrastructure
- OTV supports both multicast and unicast-only transport networks
- OTV uses ISIS as the control protocol
- OTV offers a Build-in Loop Prevention
- OTV is Site Independence
- OTV doesn’t rely on Circuit State, it’s a Point-to-Cloud model.
- OTV doesn’t rely on Flood-&-Learn, it’s a push model to distribute host MAC reachability, preserving the failure Boundary.

Overlay Transport Virtualisation

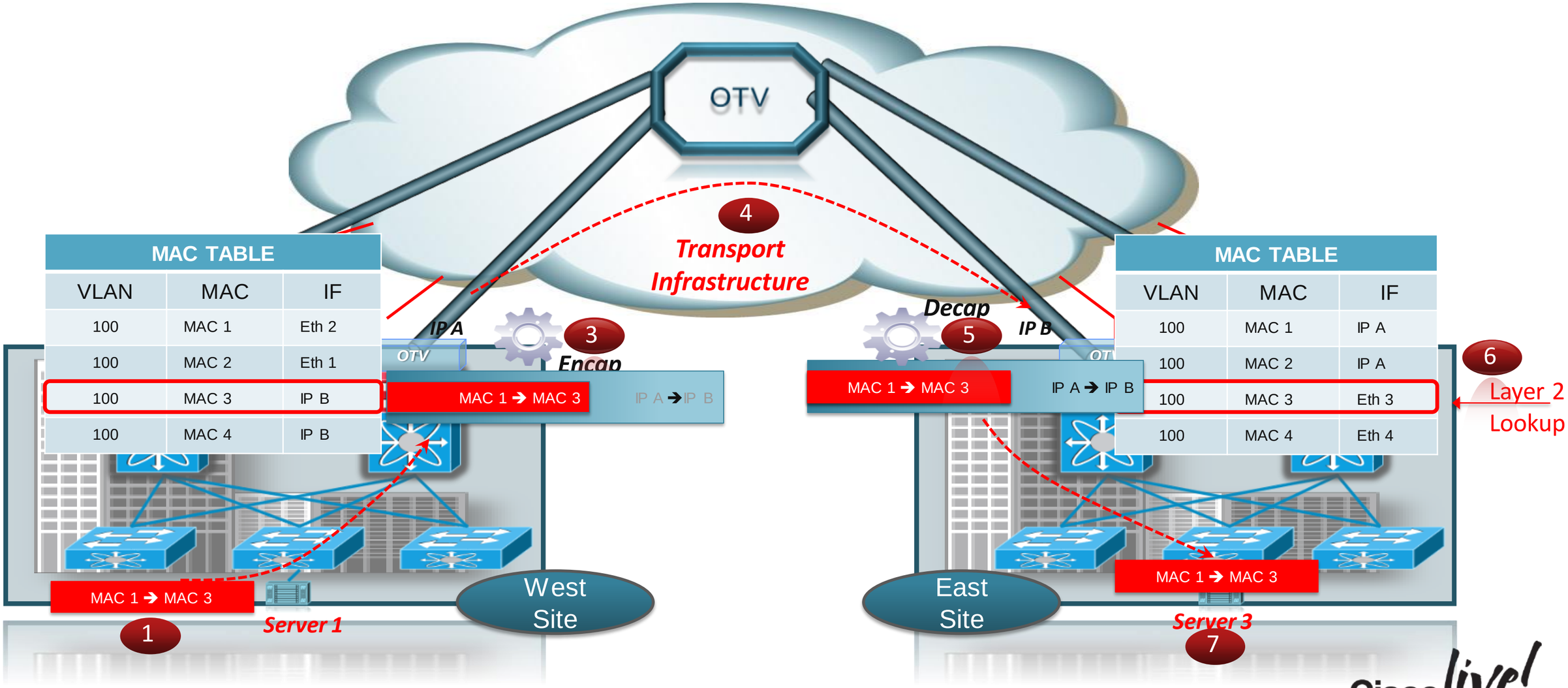
OTV Control Plane

- Neighbour discovery and adjacency over
  - Multicast (Nexus 7000 and ASR 1000)
  - Unicast (Adjacency Server Mode: Nexus 7000 (5.2) an ASR 1000 (3.9S))
- OTV proactively **advertises/withdraws** MAC reachability (control-plane learning)
- IS-IS is the OTV Control Protocol - No specific configuration required



OTV Overview

Inter-Sites Packet Flow



New Features

- F1/F2E used as Internal Interfaces
- Selective Unicast Flooding
- Dedicated Data Broadcast Group
- OTV VLAN Translation
- OTV Fast Convergence

Starting 6.2(2)

- Tunnel Depolarisation with Secondary IP

Starting 6.2(6)*

- Loopback Join-Interface

Targeted Future 7.1

* F3: recommended 6.2.(8)

New Features

6.2(2) and above

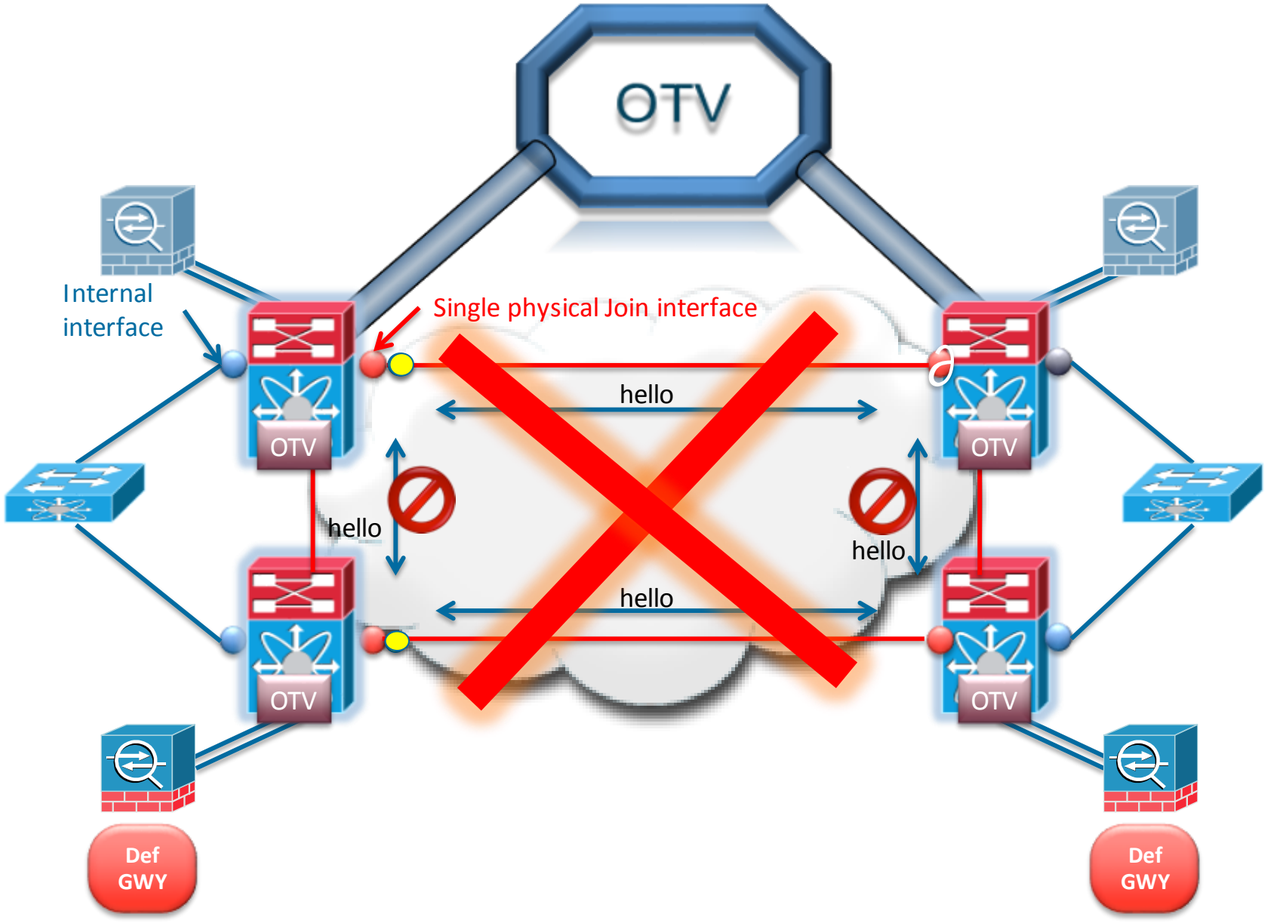
- F1 or F2E can be mixed with M-series VDC and used as OTV internal interface
- F3 (6.2(6) and above) can be used as OTV Internal and Join interface
- Unicast MAC address can be statically configured to flood across OTV
- Dedicated Broadcast Group
  - Allows separation and prioritisation of control traffic vs. data plane broadcast traffic

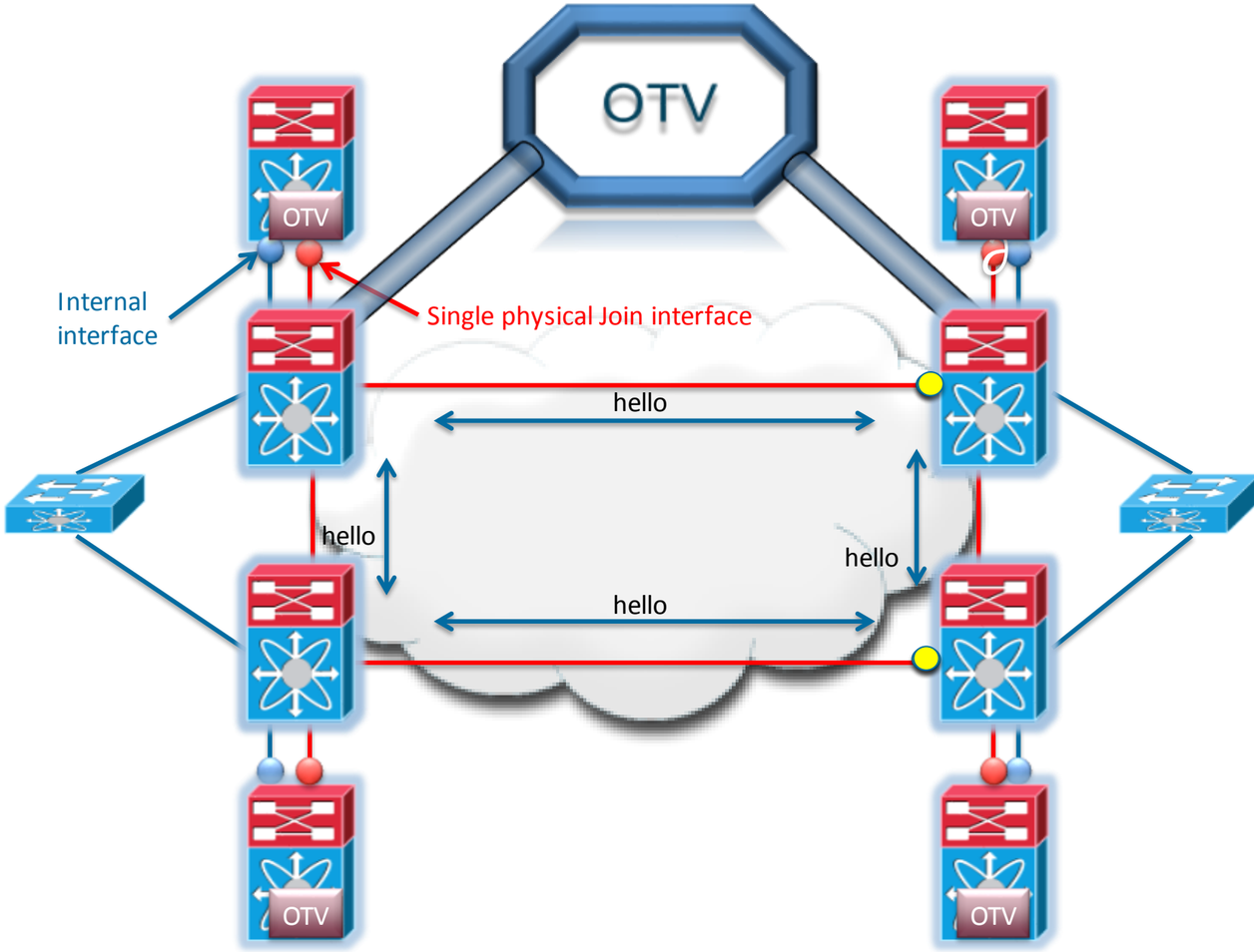
```
otv flood mac 0200.0AC9.00A2 vlan 202

interface Overlay1
  otv join-interface port-channel100
  otv broadcast-group 239.1.1.5
  otv control-group 239.1.1.1
  otv data-group 232.1.1.0/24
  otv extend-vlan 200-209
```

Improving OTV Design

Physical Join interface

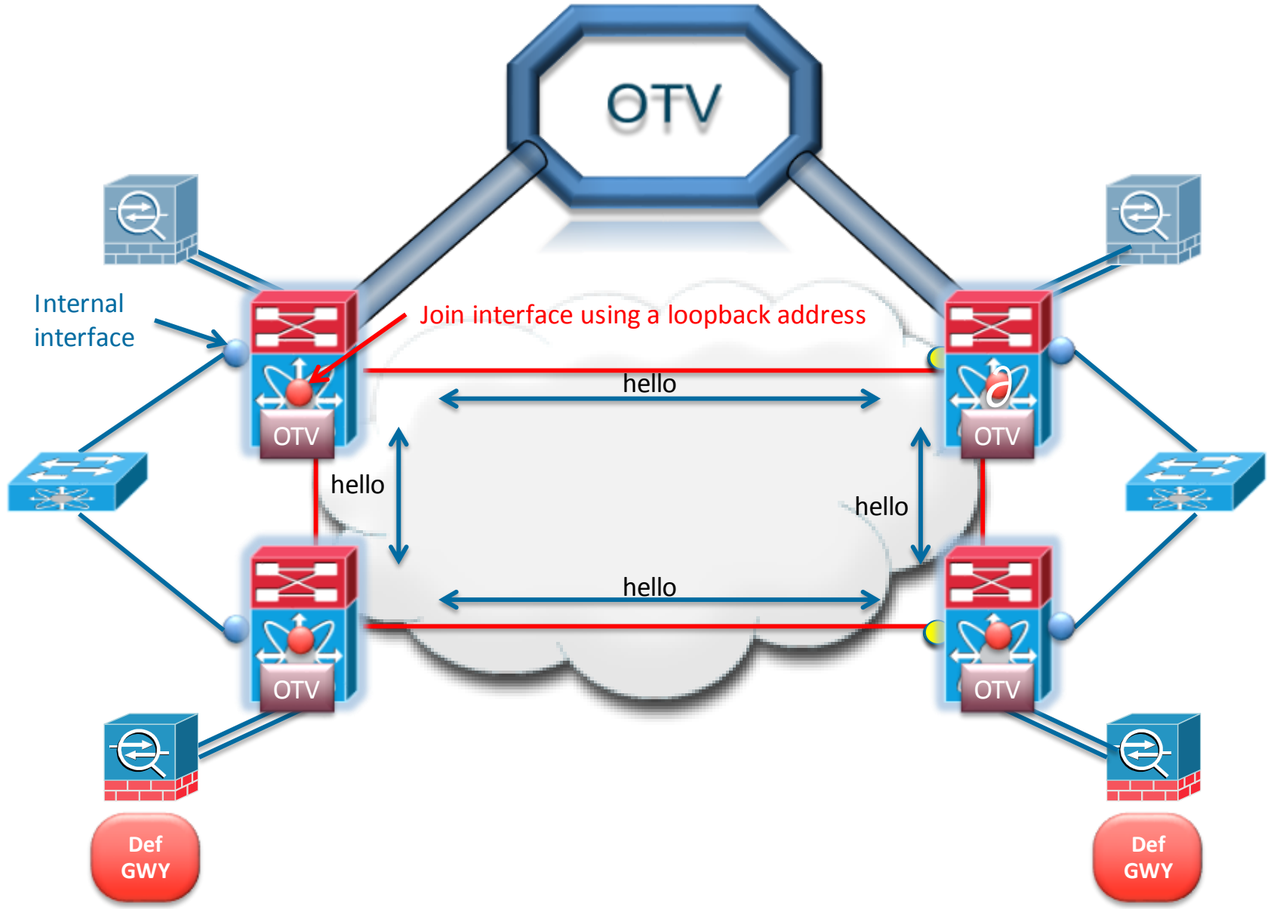




Traditional deployment of OTV in the stick with SVI in the aggregation layer

Improving OTV Design

Loopback Join interface

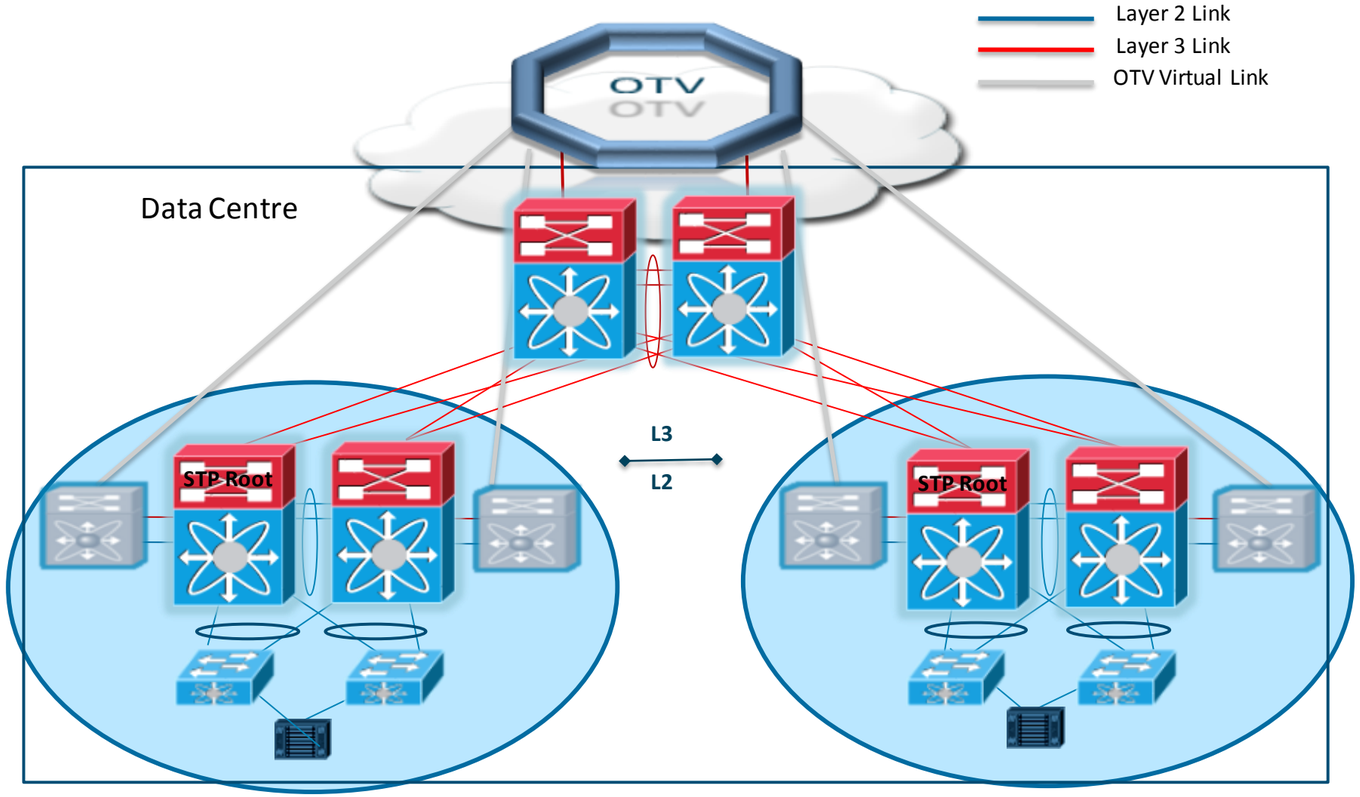


- The Source interface is now a logical loopback address and takes over the original physical Join interface
- All available layer 3 uplinks Join'ing the OTV overlay can be used to reach destination DC
- Hash will be calculated with the source-interface as the source address
- Secondary interface can be leveraged to better load distribute the traffic over multiple layer 3 paths

Placement of the OTV Edge Device

Option 1 - OTV in the DC Aggregation

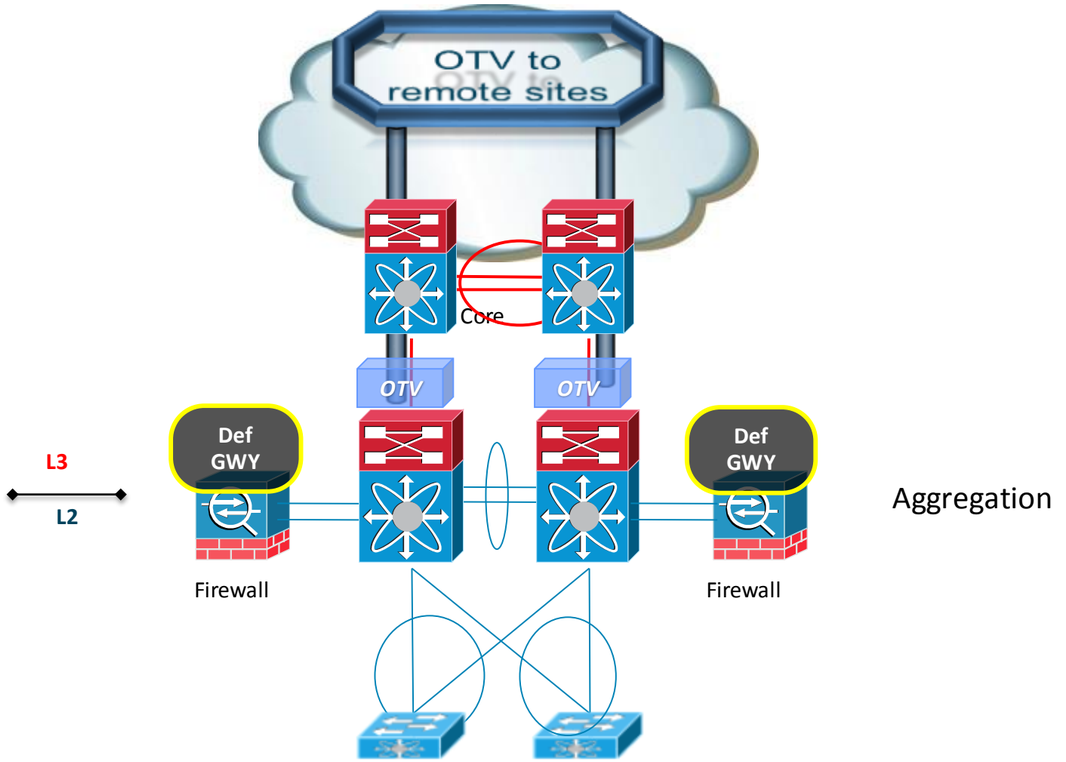
- OTV VDC deployment replicated in each POD
- Inter-PoD & inter-DC LAN extension with a pure L3 core
- Isolated STP domain in each POD
  - STP filtered across the OTV overlay by default
  - Independent STP root bridge per POD
- vPC facing the access layer devices
  - Loop free topology inside each POD
  - Loop free topology per POD can be also be through FabricPath.



Placement of the OTV Edge Device

Option 2 - OTV at the Aggregation with L2-L3 Boundary on External Firewalls

- The Firewalls host the Default Gateway
- No SVIs at the Aggregation Layer
- No Need for the OTV VDC



DCI Convergence Summary

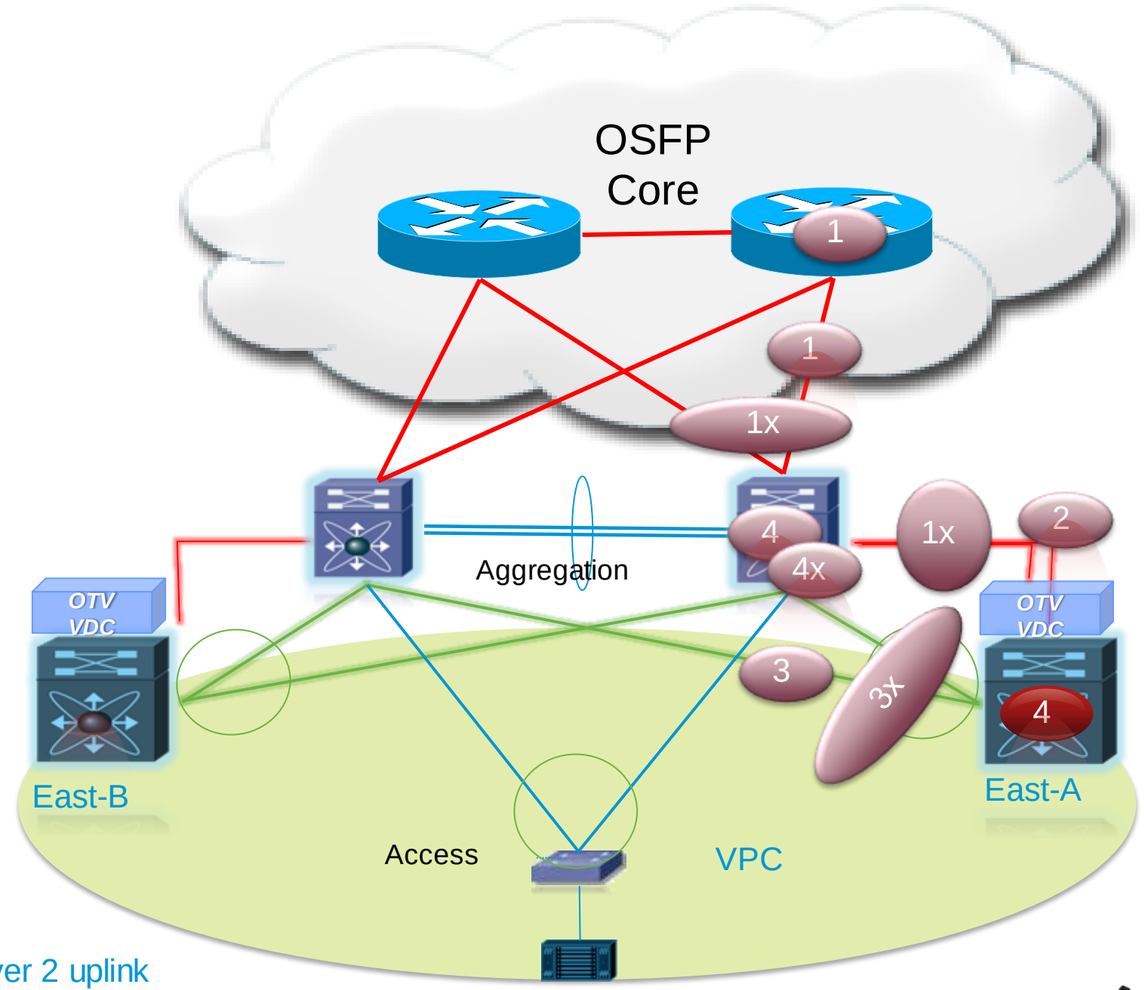
Robust HA is the guiding principle

Common Failures:

1. Core failures
   Multipath routing (or TE FRR) ➔ **sub-sec** ✓
2. Join interface failures
   Link Aggregates across line-cards ➔ **sub-sec** ✓
3. Internal Interfaces failures
   Multipath topology (vPC) & LAGs ➔ **sub-sec** ✓
4. ED component failures
   HW/SW resiliency ➔ **sub-sec** ✓

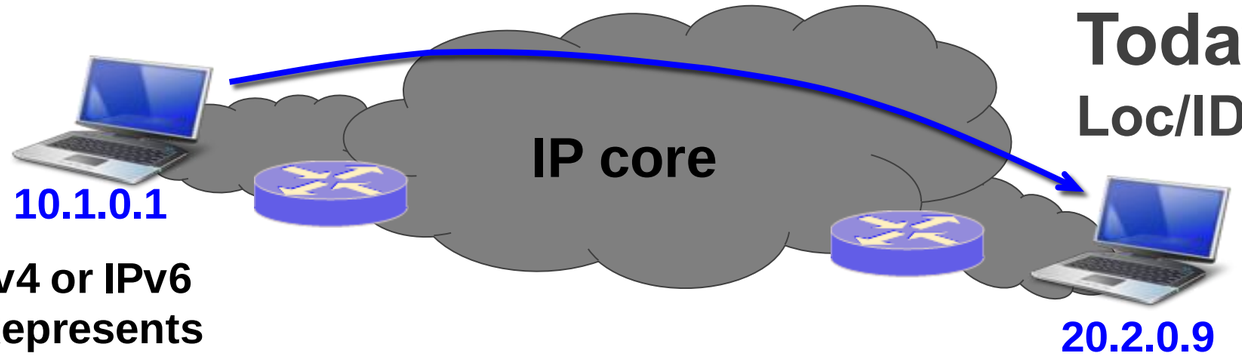
Extreme failures (unlikely):

1x. Core partition
3x. Site partition
4x. Device down
Require OTV reconvergence
➔ **< 5s** ✓



Location Identity Separation Protocol

What do we mean by “Location” and “Identity”

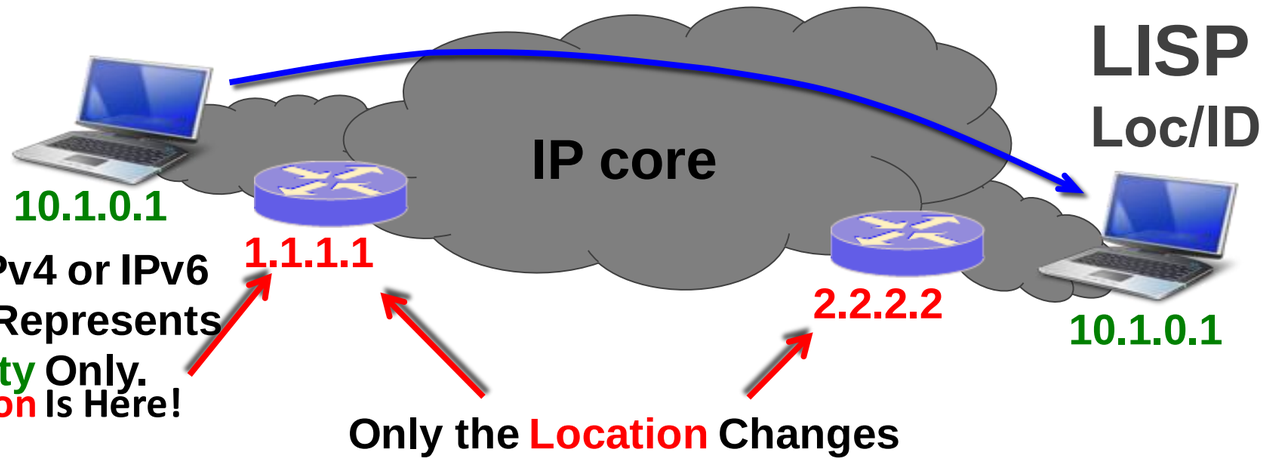


Today’s IP Behaviour

Loc/ID “Overloaded” Semantic

Device IPv4 or IPv6 Address Represents Identity and Location

When the Device Moves, It Gets a New IPv4 or IPv6 Address for Its New Identity and Location



LISP Behaviour

Loc/ID “Split”

Device IPv4 or IPv6 Address Represents Identity Only.
Its Location Is Here!

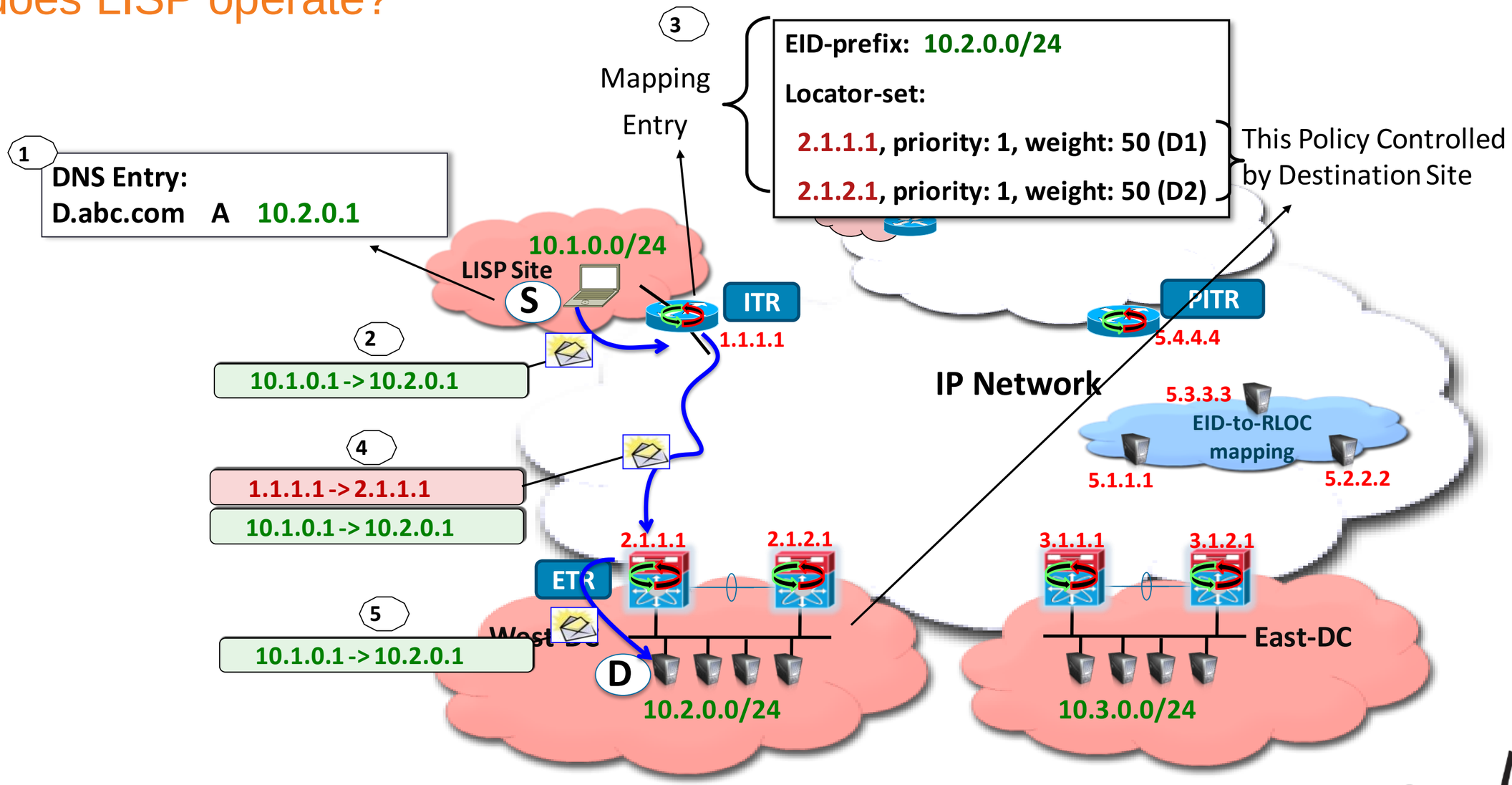
When the Device Moves, Keeps Its IPv4 or IPv6 Address. It Has the Same Identity

Only the Location Changes

Cisco live!

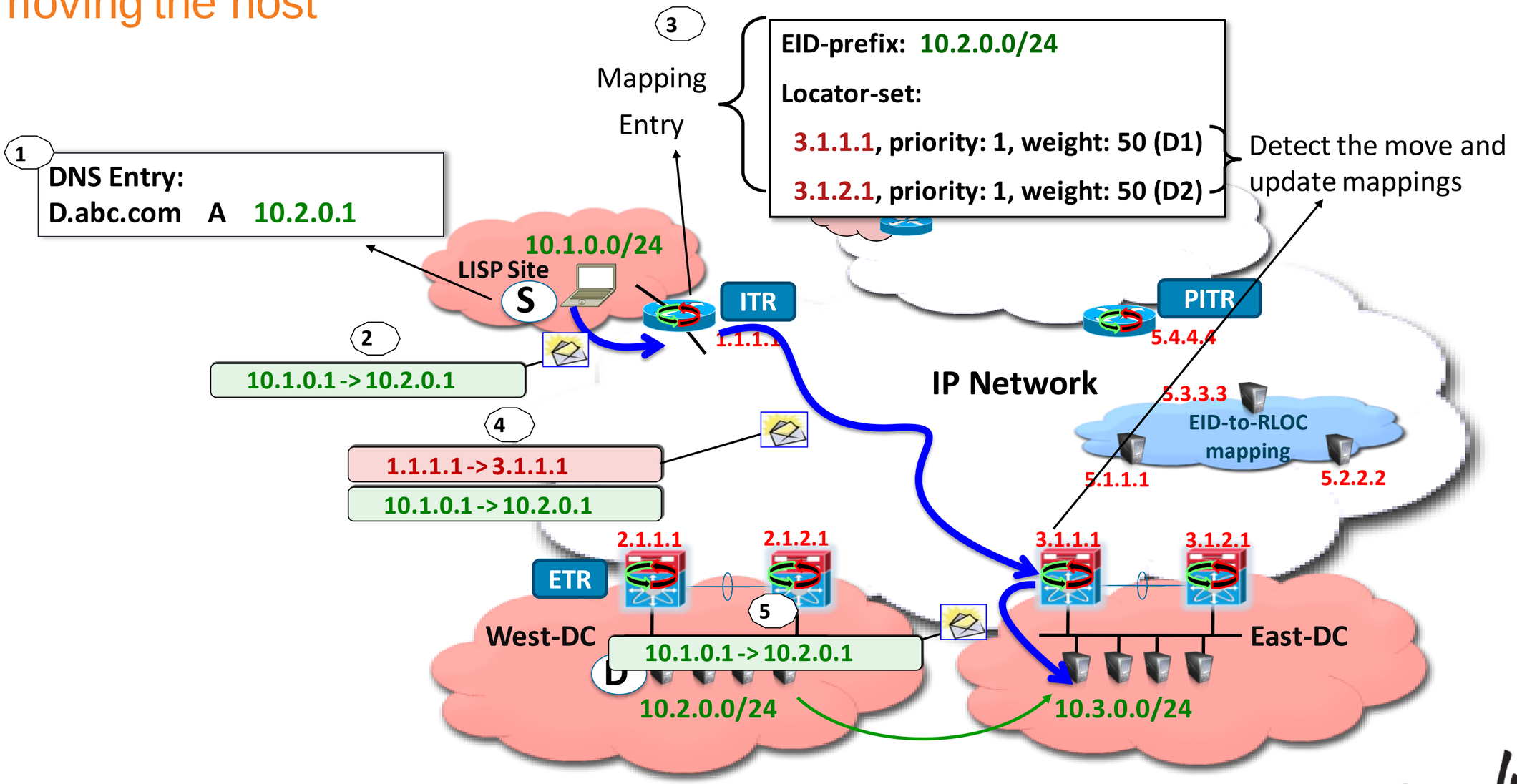
A LISP Packet Walk

How does LISP operate?



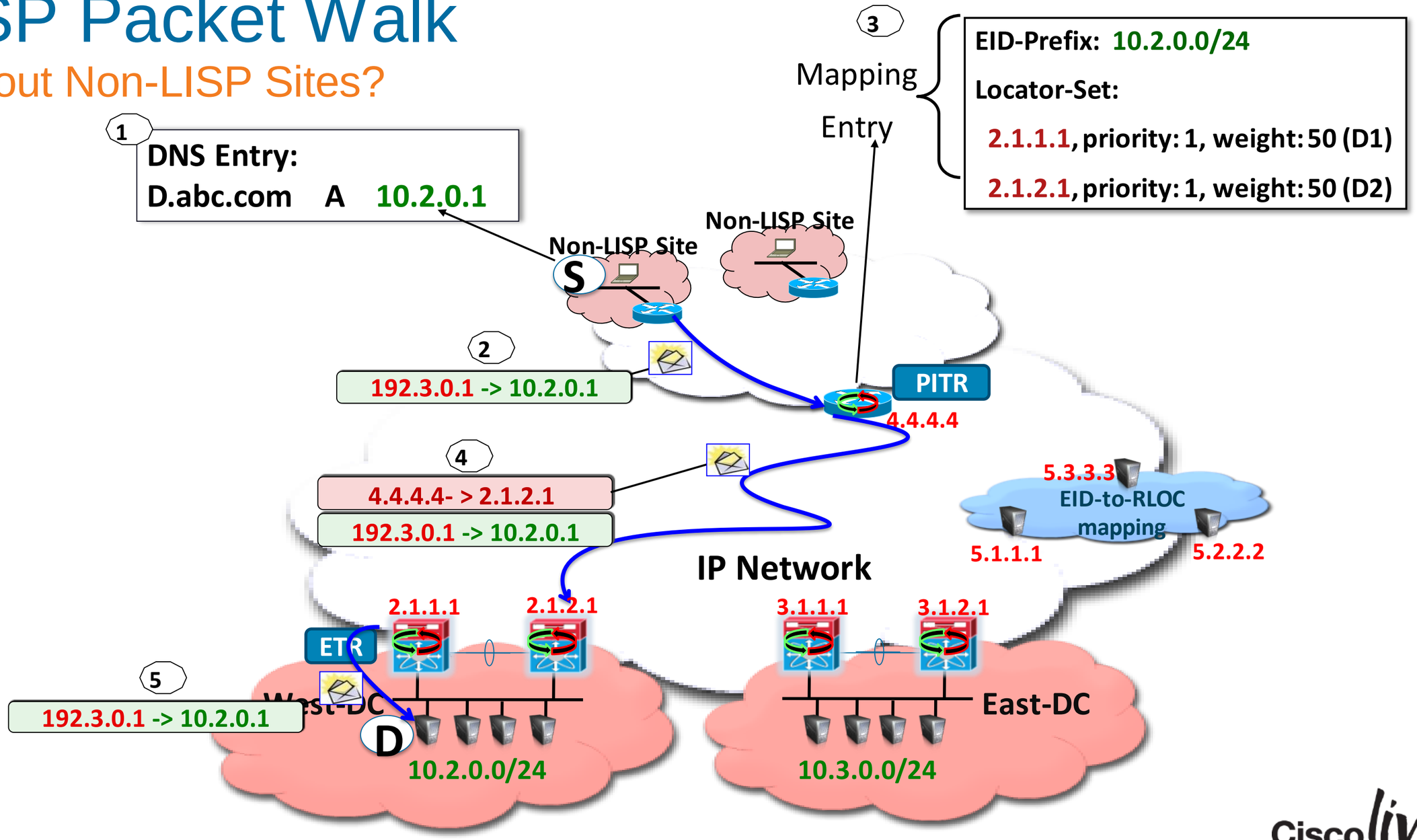
A LISP Packet Walk

After moving the host



A LISP Packet Walk

How about Non-LISP Sites?

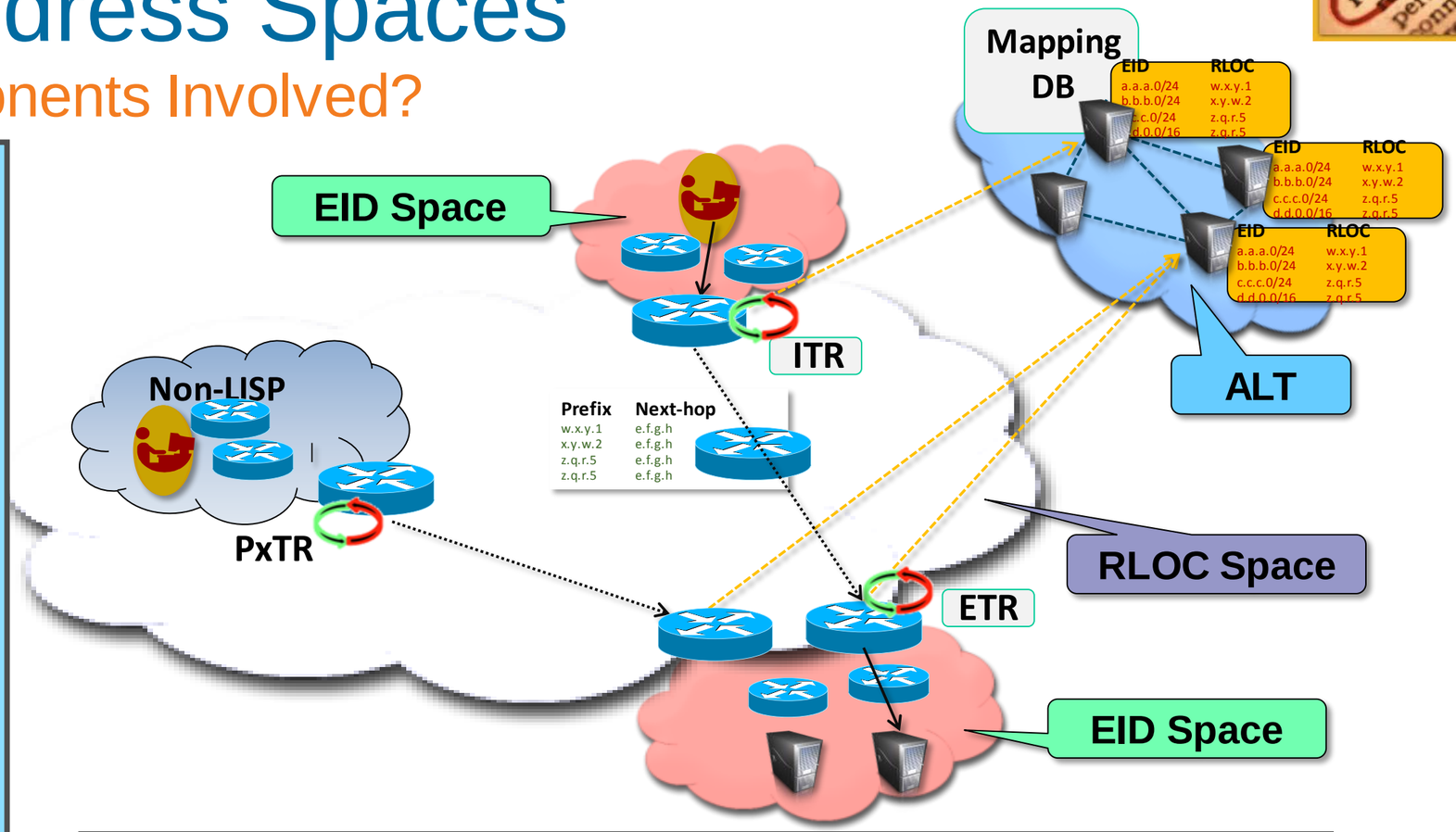


LISP Roles and Address Spaces

What are the Different Components Involved?

LISP Roles

- **Tunnel Routers - xTRs**
  - Edge devices encap/decap
  - Ingress/Egress Tunnel Routers (ITR/ETR)
- **Proxy Tunnel Routers - PxTR**
  - Coexistence between LISP and non-LISP sites
  - Ingress/Egress: PITR, PETR
- **EID to RLOC Mapping DB**
  - RLOC to EID mappings
  - Distributed across multiple Map Servers (MS)



Address Spaces

- **EID = End-point Identifier**
  - Host IP or prefix
- **RLOC = Routing Locator**
  - IP address of routers in the backbone

LISP Host-Mobility

Needs:

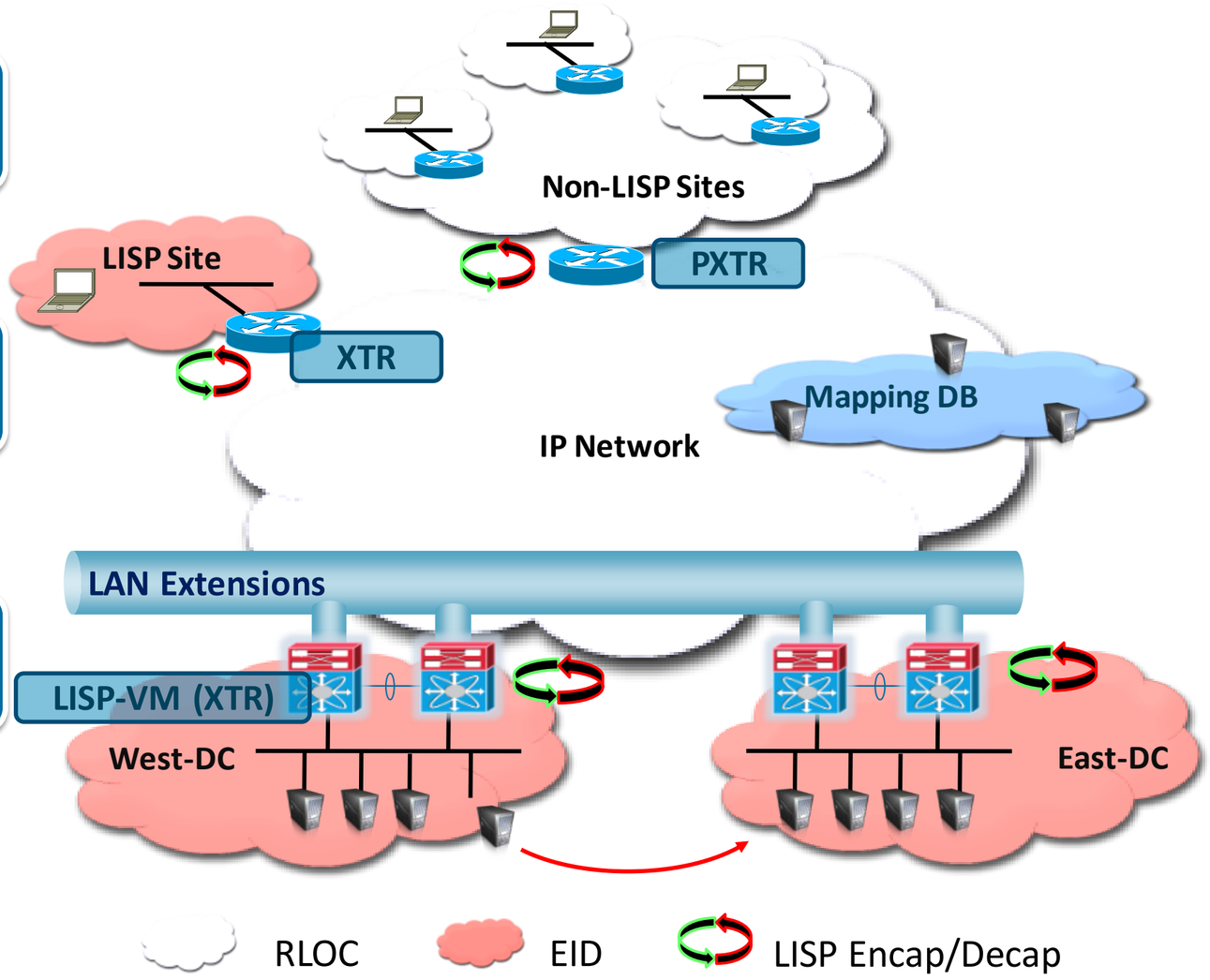
- Global IP-Mobility **across subnets**
- Optimised routing **across extended subnet sites**

LISP Solution:

- Automated **move detection** on XTRs
- Dynamically update EID-to-RLOC mappings
- **Traffic Redirection** on ITRs or PITRs

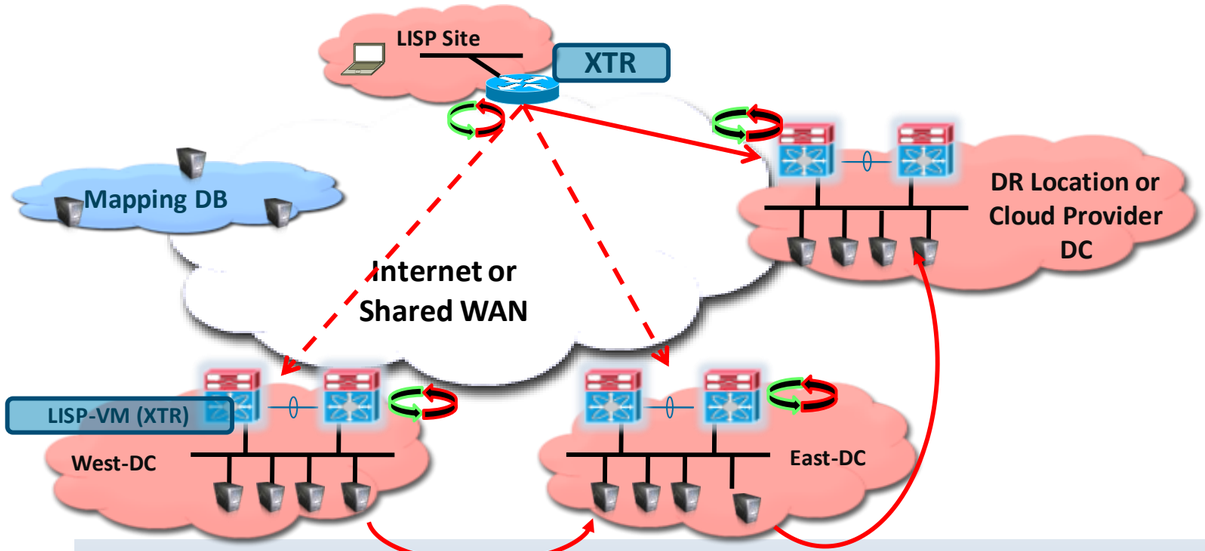
Benefits:

- Direct Path (no triangulation)
- Connections maintained across move
- No routing re-convergence
- No DNS updates required
- Transparent to the hosts
- Global Scalability (cloud bursting)
- IPv4/IPv6 Support



Host-Mobility Scenarios

Moves Without LAN Extension



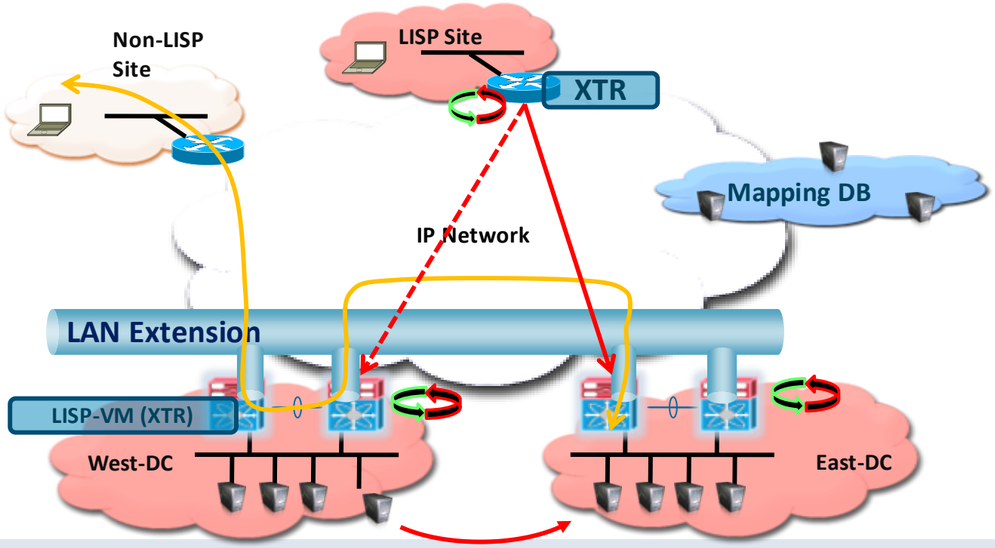
IP Mobility Across Subnets

Disaster Recovery

Cloud Bursting

Application Members in One Location

Moves With LAN Extension



Routing for Extended Subnets

Active-Active Data Centres

Distributed Clusters

Application Members Distributed (Broadcasts across sites)

Summary - Where to Deploy LISP and OTV

Roles and Places in the Network

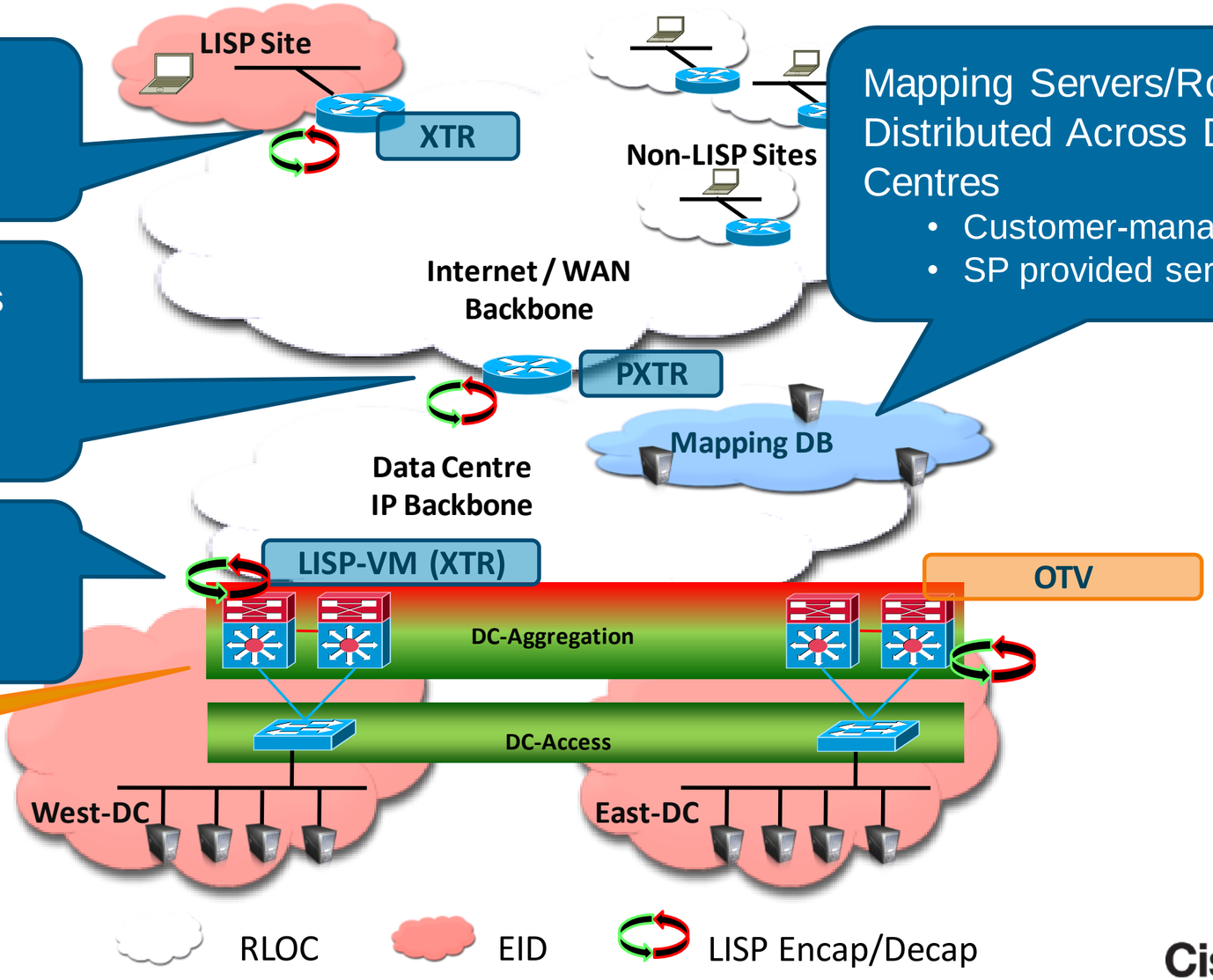
XTR: Branch Routers @ LISP Sites
- Customer-managed/owned
- SP-Managed CE service

PXTR: Border Routers @ Transit Points
- Customer backbone routers
- Customer router @ co-location
- SP provided router/service

LISP-VM XTR: Aggregation Routers @ Data Centre
- Customer-managed/owned

OTV: Aggregation Routers @ Data Centre
- Customer-managed/owned

Mapping Servers/Routers: Distributed Across Data Centres
- Customer-managed/owned
- SP provided service



LISP References

- LISP Information
  - Cisco LISP Site …… http://lisp.cisco.com (IPv4 and IPv6)
  - Cisco LISP Marketing Site ……... http://www.cisco.com/go/lisp/
  - LISP Beta Network Site …… http://www.lisp4.net or http://www.lisp6.net
  - LISP DDT Root ……... http://www.ddt-root.org
  - IETF LISP Working Group ……..… http://tools.ietf.org/wg/lisp/
- LISP Mailing Lists
  - Cisco LISP Questions …… lisp-support@cisco.com
  - IETF LISP Working Group …… lisp@ietf.org
  - LISP Interest (public) …… lisp-interest@puck.nether.net
  - LISPmob Questions ……... users@lispmob.org

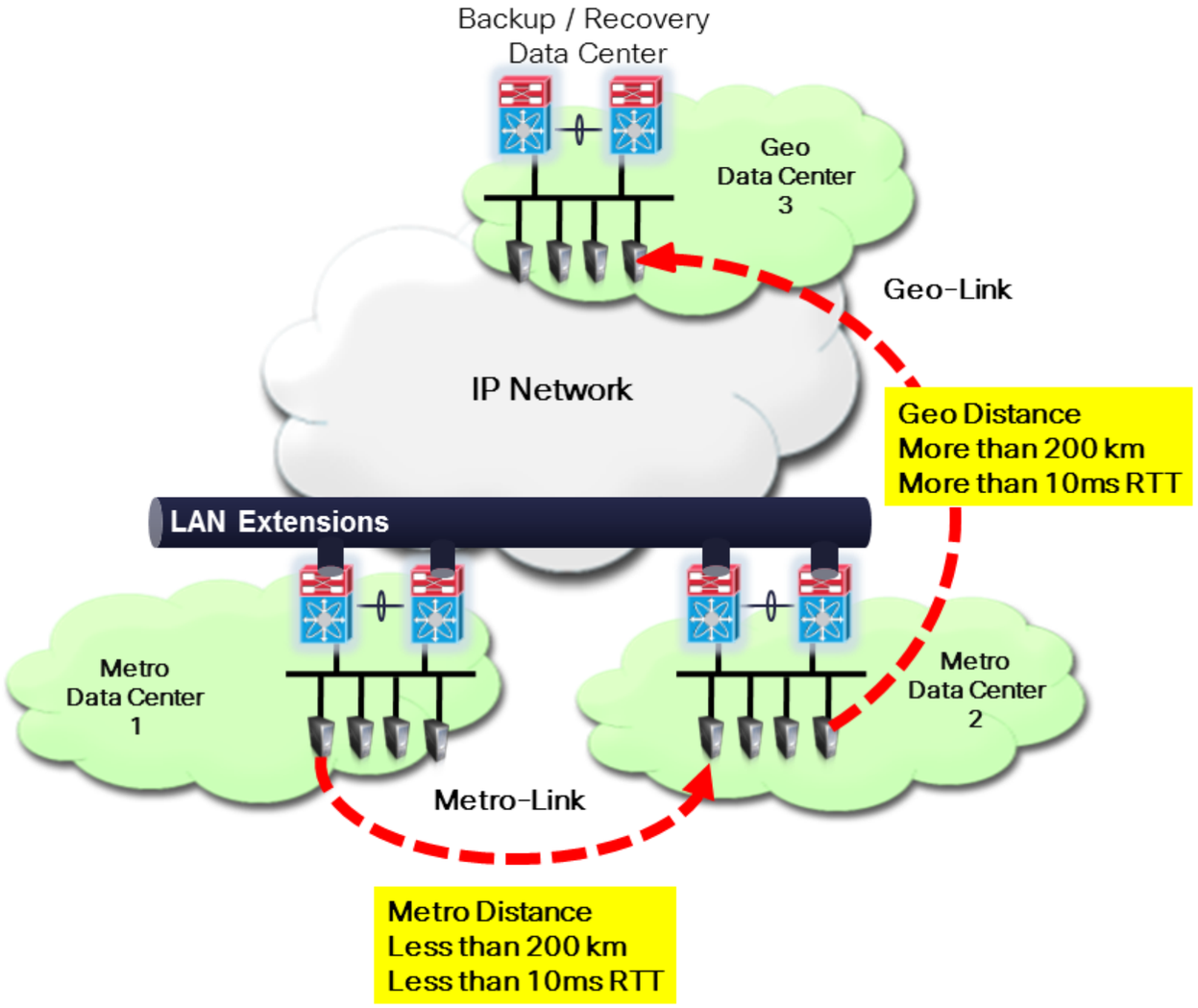
Cisco live!

Agenda

- Active-Active (A/A) Data Centre:
  - Market & Business Drivers
  - Terminology, Criticality levels and Solutions Overview
- A/A Data Centre Design Considerations:
  - Storage Extension
  - Data Centre Interconnect (DCI) – L2 & L3 scenarios
- A/A Metro Data Centres Designs

  - Network Services and Applications (Path optimisation)
- Cisco ACI and Active / Active Data Centre
- Q&A

Example: 3-Site Data Centre Interconnect Model



Geo Data Center with Optional LAN Extension

Workload Mobility Across Subnets or within Extended Subnets

Geo Data Center (DC-3)

Cold Workload Migration

National Disaster Recovery

Application Members contained to Single Site

Metro Data Centers With LAN Extension

Workload Mobility with Extended Subnets

Metro Data Centers (DC-1 and DC-2)

Live + Cold Workload Mobility

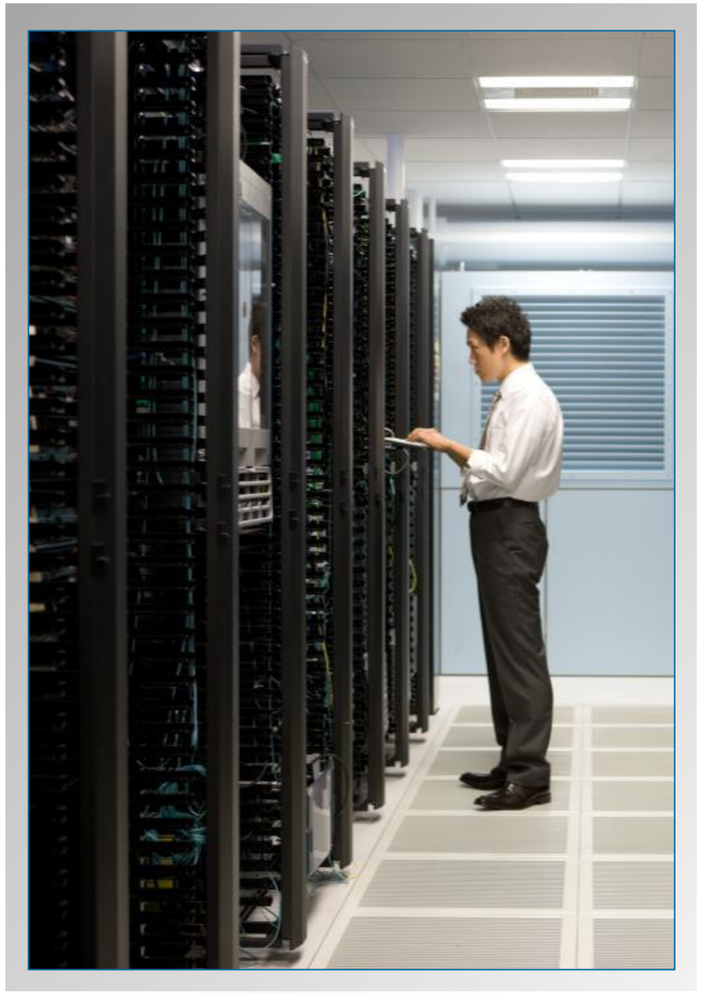
Regional Disaster Recovery

Distributed Applications Clusters

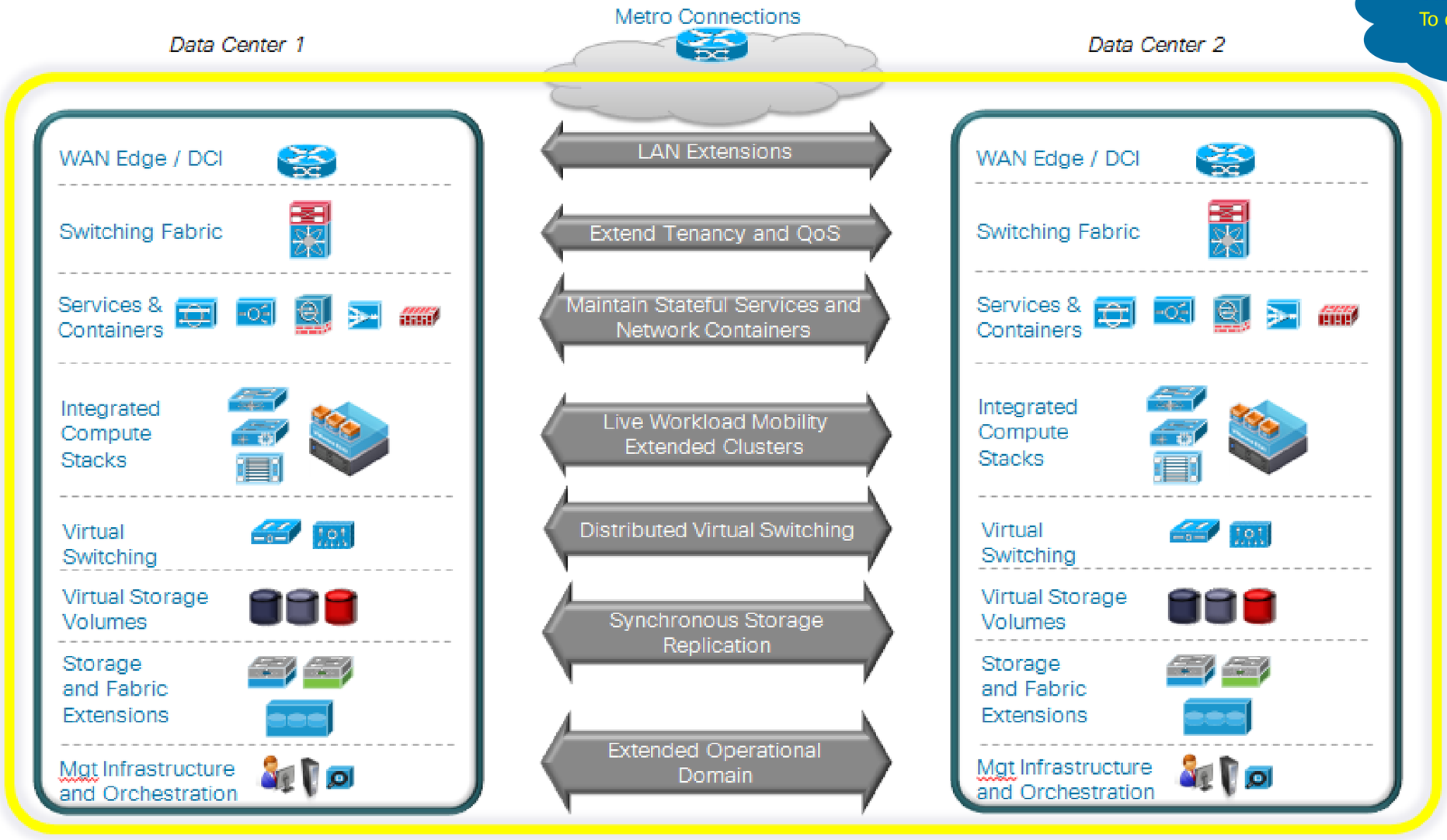
Application Members may be Distributed between Sites

Metro Virtual Data Centre

- High-availability application and data solution architecture which leverages a dual data centre physical infrastructure
- Addresses all levels of the data centre stack
  - ✓ physical layer, network layer, server platform resources, storage resources, application networking services, data tier structure
  - ✓ No physical single points of failure
  - ✓ Optimised use of capacity through virtualisation
- Management and interaction of applications in a paired data centre environment
- Disaster Avoidance and Prevention by Pro-actively migrates seamlessly Virtual Machines with NO interruption
- Support Disaster Recovery capability beyond dual data centre resiliency
- Active-active capability and workload rotation to accelerate incident response time and increase confidence
- Capable of "no data loss" within Metro (synchronous replication, RPO=0)



Active / Active Metro Design



Cool ! What if they are both ACI on the Metro sites ? To come on the 2nd half of this session ☺

These Sites are Operating as ONE Logical Data Centre

Live Workload Mobility Requirements for the Active-Active Metro Design

Move an **"Active"** Virtual Workload across Metro Data Centres while maintaining Stateful Services

Business Continuity Use Cases for Live Mobility

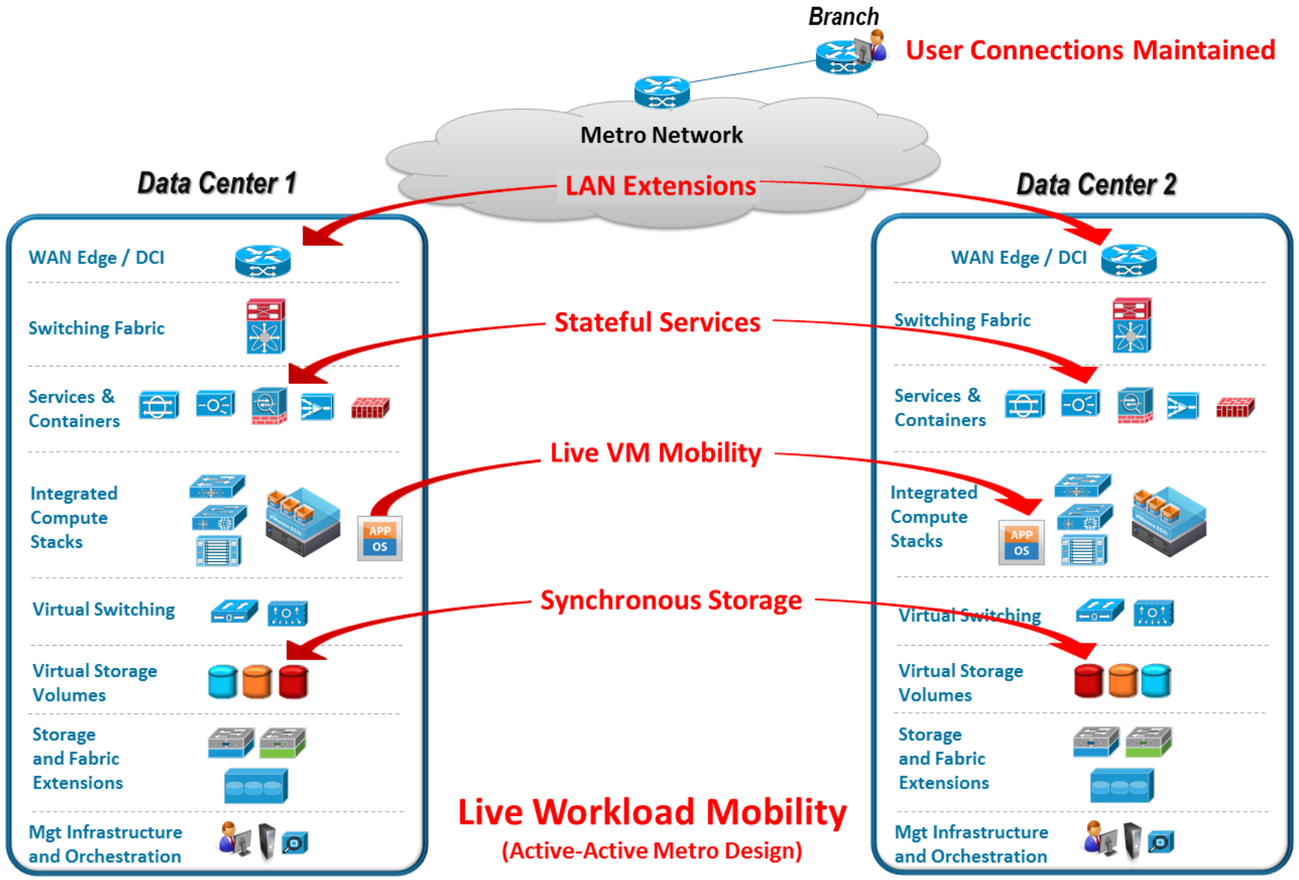
Most Business Critical Applications (Lowest RPO/RTO)
Live Workload Migrations
Operations Rebalancing / Maintenance / Consolidation of Live Workloads
Disaster Avoidance of Live Workloads
Application Geo-Clusters spanning Metro DCs

Hypervisor Tools for Live Mobility

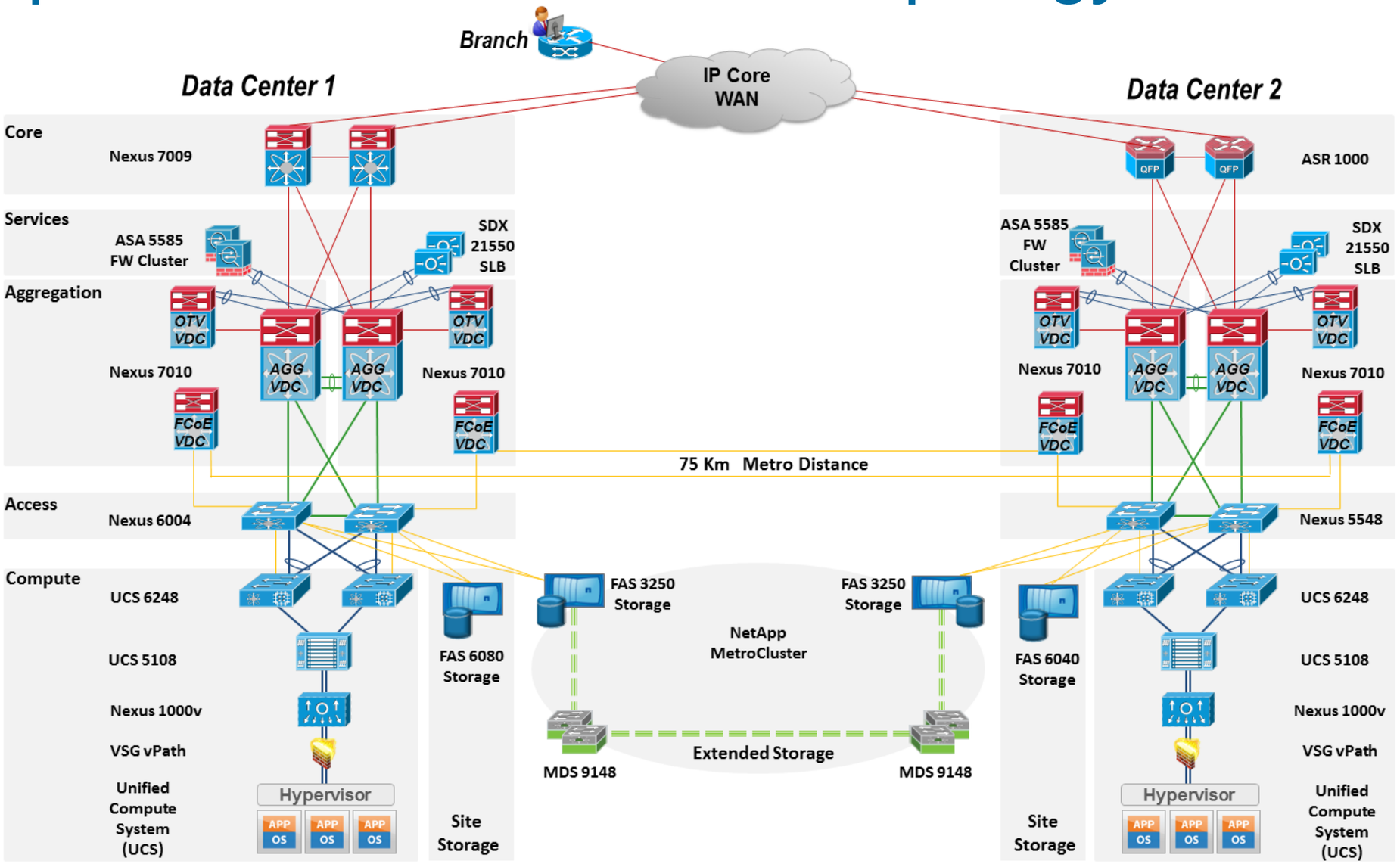
VMware vMotion or Hyper-V Live Migration
Stretched Clusters across Metro DCs
Host Affinity rules to manage resource allocation
Distributed vCenter or System Centre across Metro DCs

Metro DC Infrastructure to support Live Workload Mobility

Network: Data Centre Interconnect and IP Path Optimisations
Virtual Switches Distributed across Metro
Maintain Multi-Tenant Containers
Services: Maintain Stateful Services for active connections
Minimise traffic tromboning between Metro DCs
Compute: Support Single-Tier and Multi-Tier Applications
Storage: Storage extended across Metro, Synchronous Data Replication
Distributed Virtual Volumes
Hyper-V Shared Nothing Live Migration (Storage agnostic)

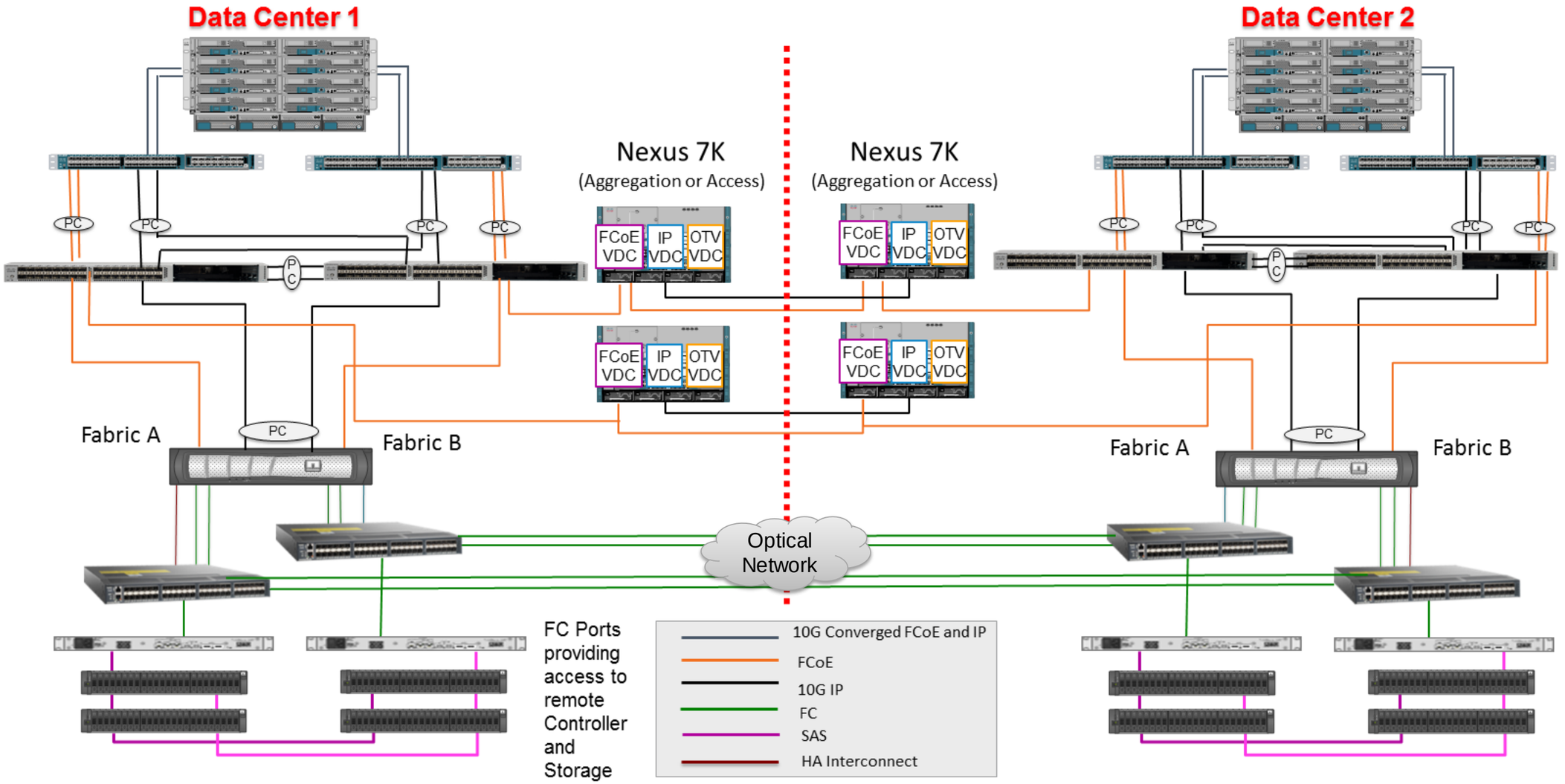


Example Active-Active Metro Topology



Cisco Public

Extended Storage Example: Multi-Hop FCoE using NetApp Fabric MetroCluster

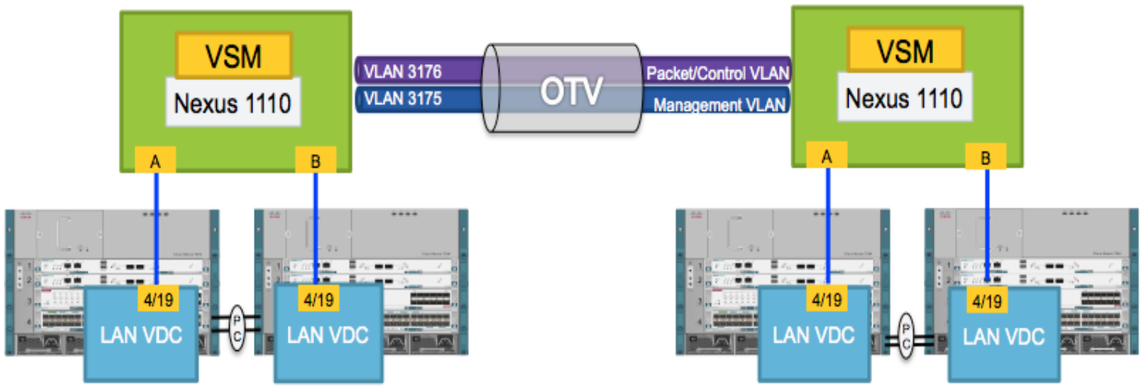


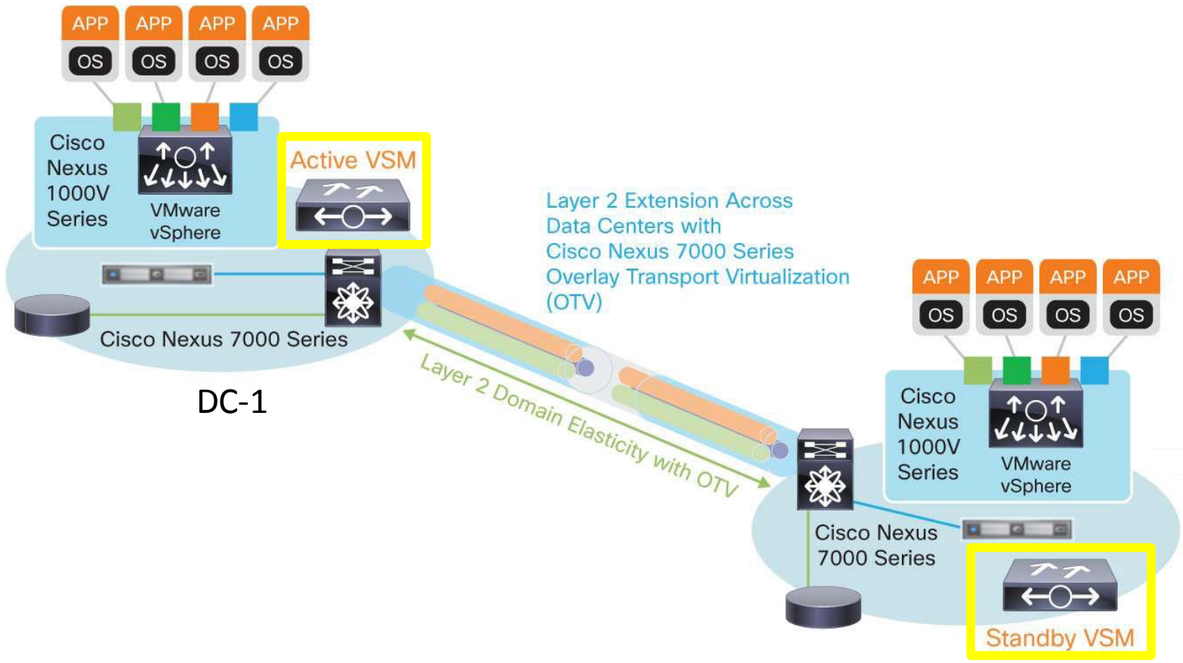
Nexus 1000v Extensions Across Geographies

VSMs and VEMs can span Metro distances for enhanced availability

VSM Extended Across Data Centres:
Supports splitting Active and Standby Nexus 1000V Virtual Supervisor Modules (VSMs) across two data centres to implement cross-DC clusters and VM mobility while ensuring high availability.

VEM support across Metro distance:
VSM's in the data centre can support VEM's at remote branch offices.

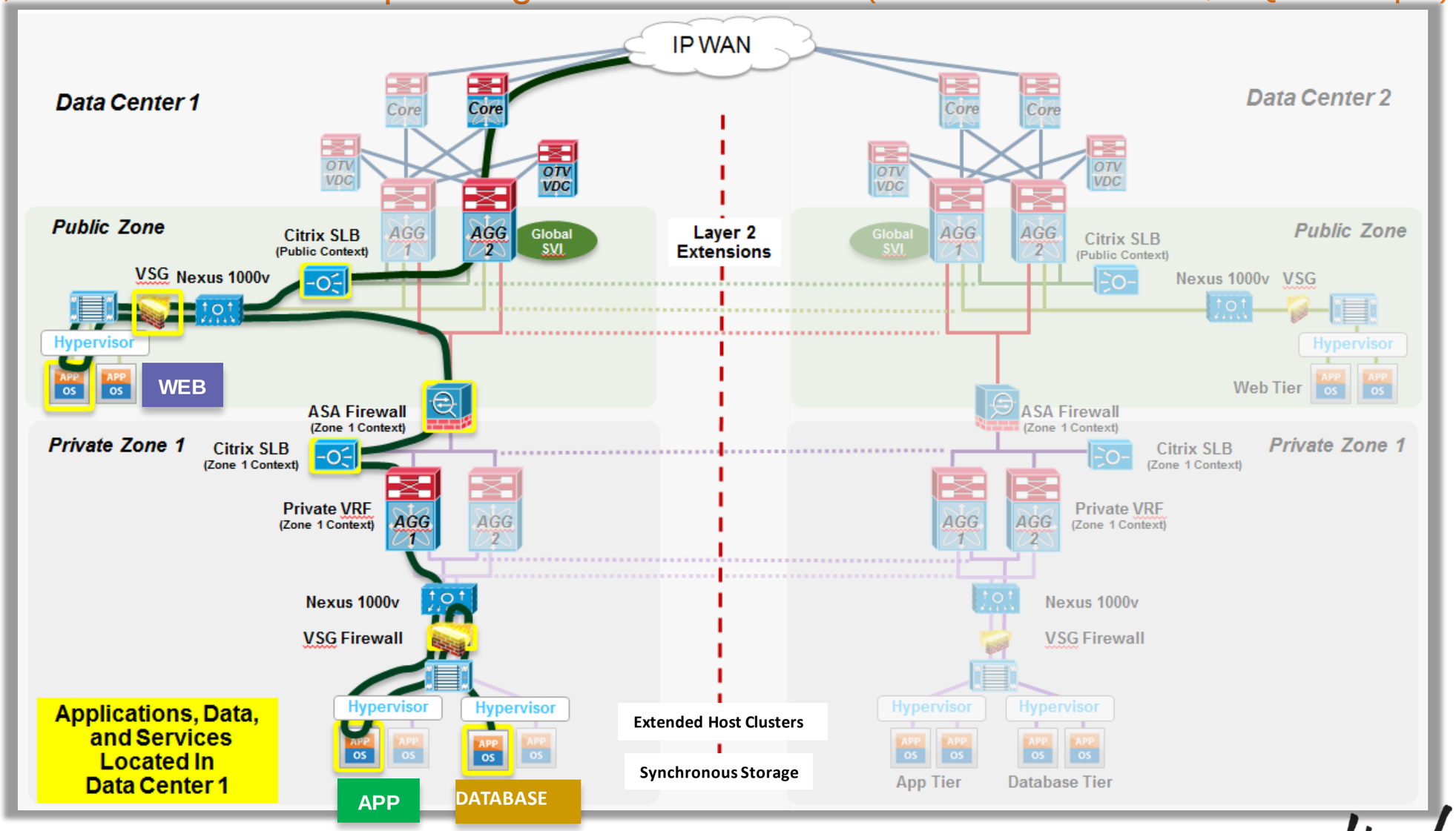




Live Workload Migration Baseline Configuration for a 3-Tier Application

The Application, Data, and Services are Operating in Data Centre 1 (Microsoft SharePoint, SQL example)

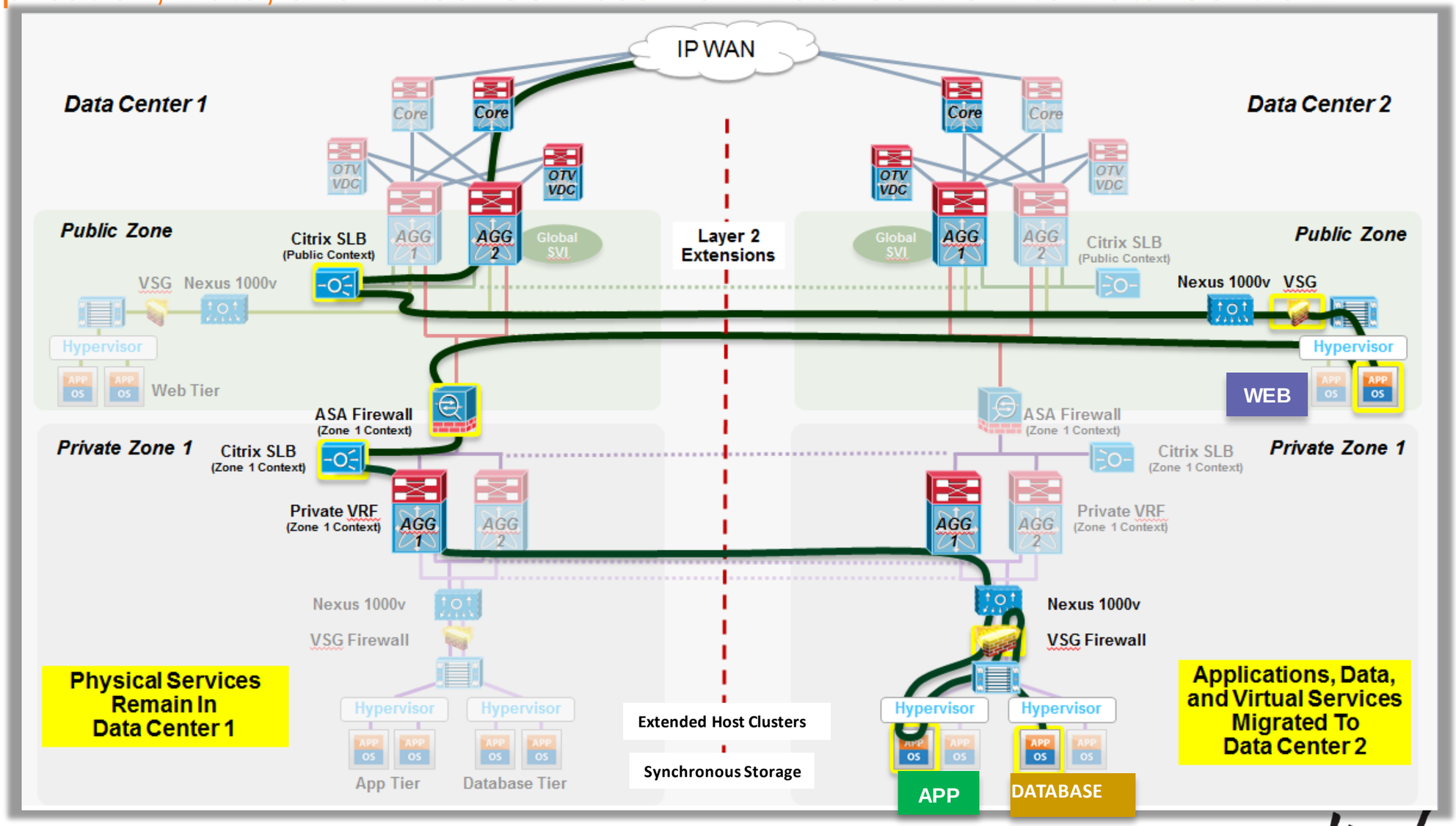
- ✓ 3-Tier Application (Web, App, Database) in a Palladium Container in DC-1
- ✓ Public Zone for Web Tier
- ✓ Protected Private Zone for App Tier and Database Tier
- ✓ Mix of Physical and Virtual Services (FW, SLB, VSG)
- ✓ Application, Data, and all Services operating in DC-1
- ✓ LAN extensions, extended ESXi clusters, and Synchronous Storage Replication between Metro sites



Step 1 of a Live Workload Migration:

Live Migrate the Application, Data, and Virtual Services from Data Centre 1 to Data Centre 2

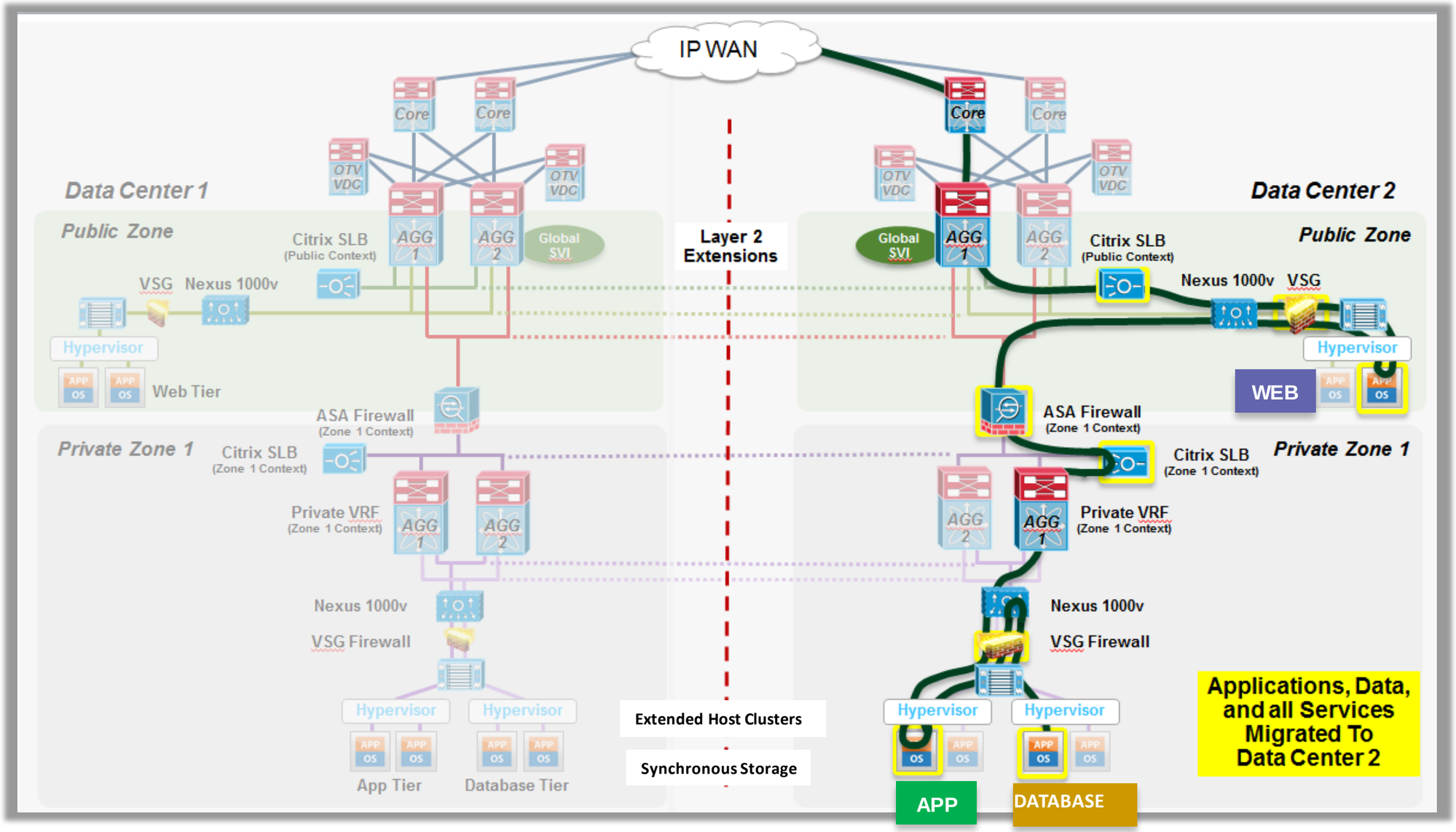
- ✓ Live vMotion all 3-Tiers of the Application to new hosts in DC-2 (Web, App, Database)
- ✓ Virtual Services follow VMs to DC-2, using Port Profile and Security Profile (N1Kv, VSG)
- ✓ Layer 2 Extensions permit tromboning back to DC-1 to maintain Stateful Services for physical appliances (FW, SLB)
- ✓ Palladium Container is extended between Metro sites, maintaining Security, Tenancy, QoS, IP addressing, and User connections
- ✓ Live Migration with minimal disruption (2 seconds or less) for IP/MAC learning of new hosts over network extensions + vMotion quiesce



Step 2 of a Live Workload Migration:

Cutover to a new Network Container in DC-2… the Application, Data, and all Services are Moved to Data Centre 2

- ✓ Migrate Physical Services (FW, SLB) to a new Palladium container in DC-2
- ✓ Redirect external users to DC-2, using a routing update (or LISP future)
- ✓ The Application, Data, and all Services are now operating in DC-2
- ✓ Layer 2 Extensions allowed the preservation of IP Addressing for Apps and Services during migration
- ✓ Network Container move with minimal disruption (Sub-second)



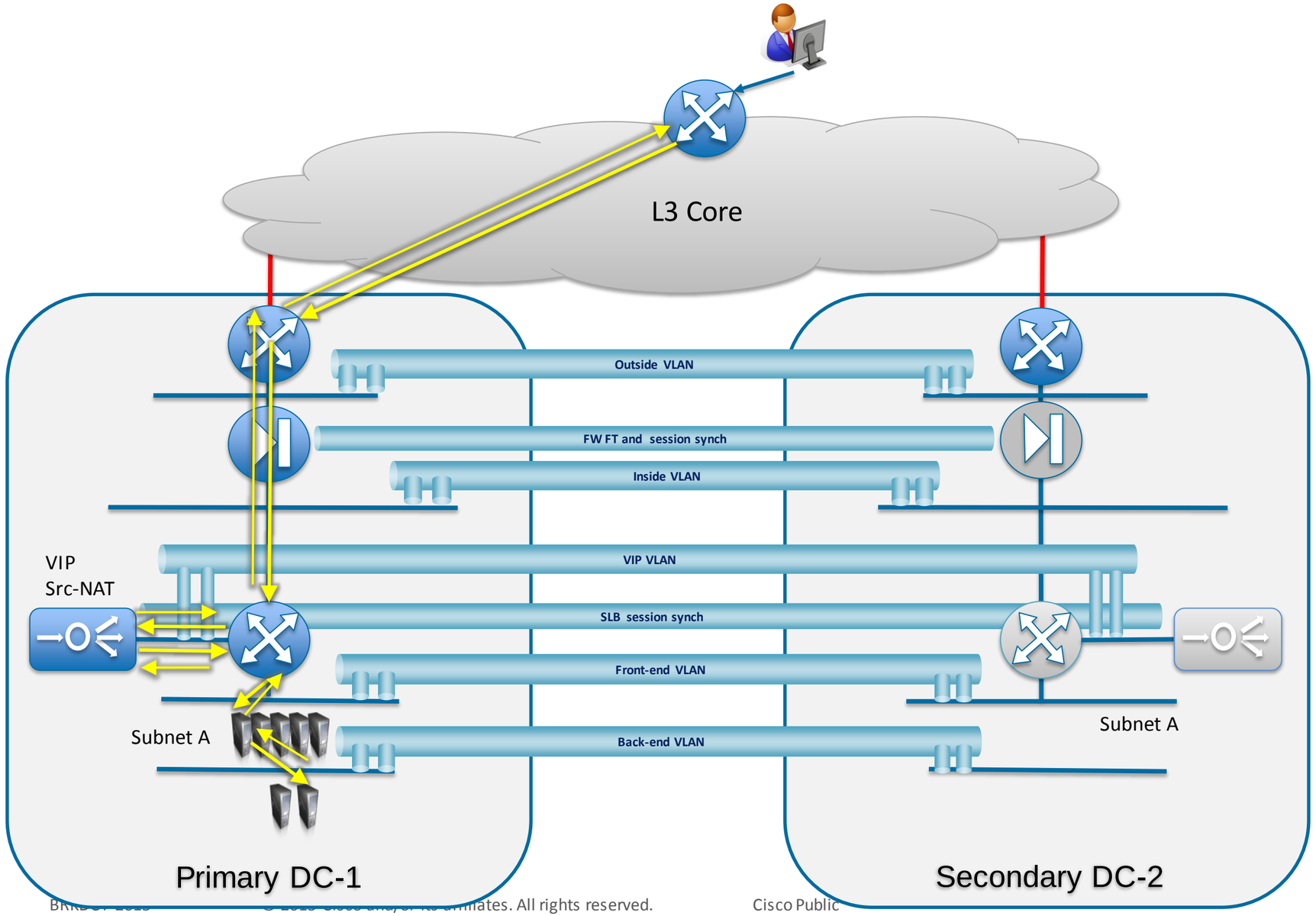
Agenda

- Active-Active (A/A) Data Centre:
  - Market & Business Drivers
  - Terminology, Criticality levels and Solutions Overview
- A/A Data Centre Design Considerations:
  - Storage Extension
  - Data Centre Interconnect (DCI) – L2 & L3 scenarios
- A/A Metro Data Centres Designs

  - Network Services and Applications (Path optimisation)
- Cisco ACI and Active / Active Data Centre
- Q&A

Network Service Placement for Metro Distances

A/S stateful devices stretched across 2 locations – nominal workflow



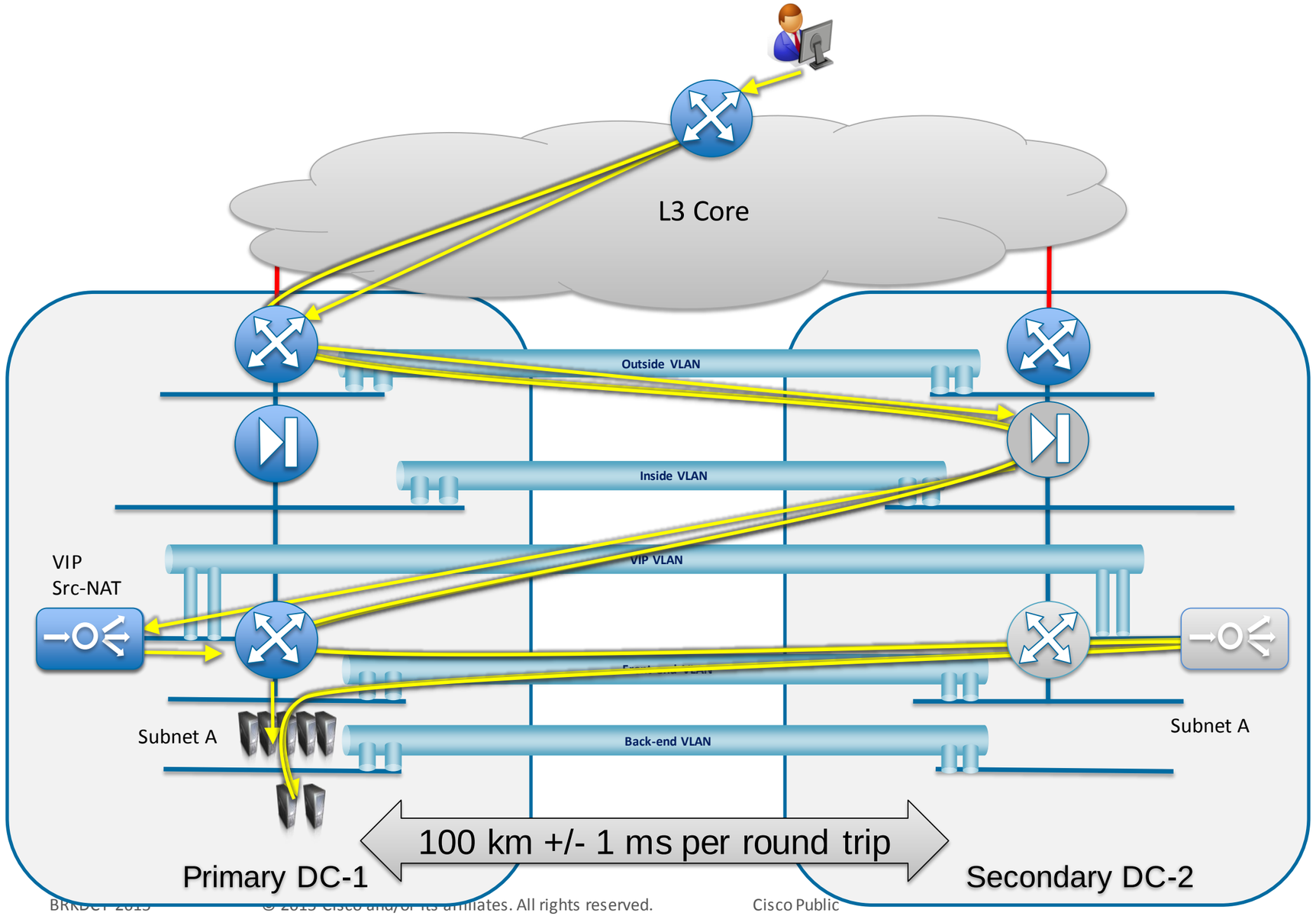
- Historically this has been well accepted for most of Metro Virtual DC (Twin-DC)

- Network Services are usually active on primary DC
- Distributed pair of Act/Sby FW & SLB on each location
- Additional VLAN Extended for state synchronisation between peers
- Source NAT for SLB VIP

Note: With traditional pair cluster this scenario is limited to 2 sites

Network Service Placement for Metro Distances

Ping-Pong impact with A/S stateful devices stretched across 2 locations



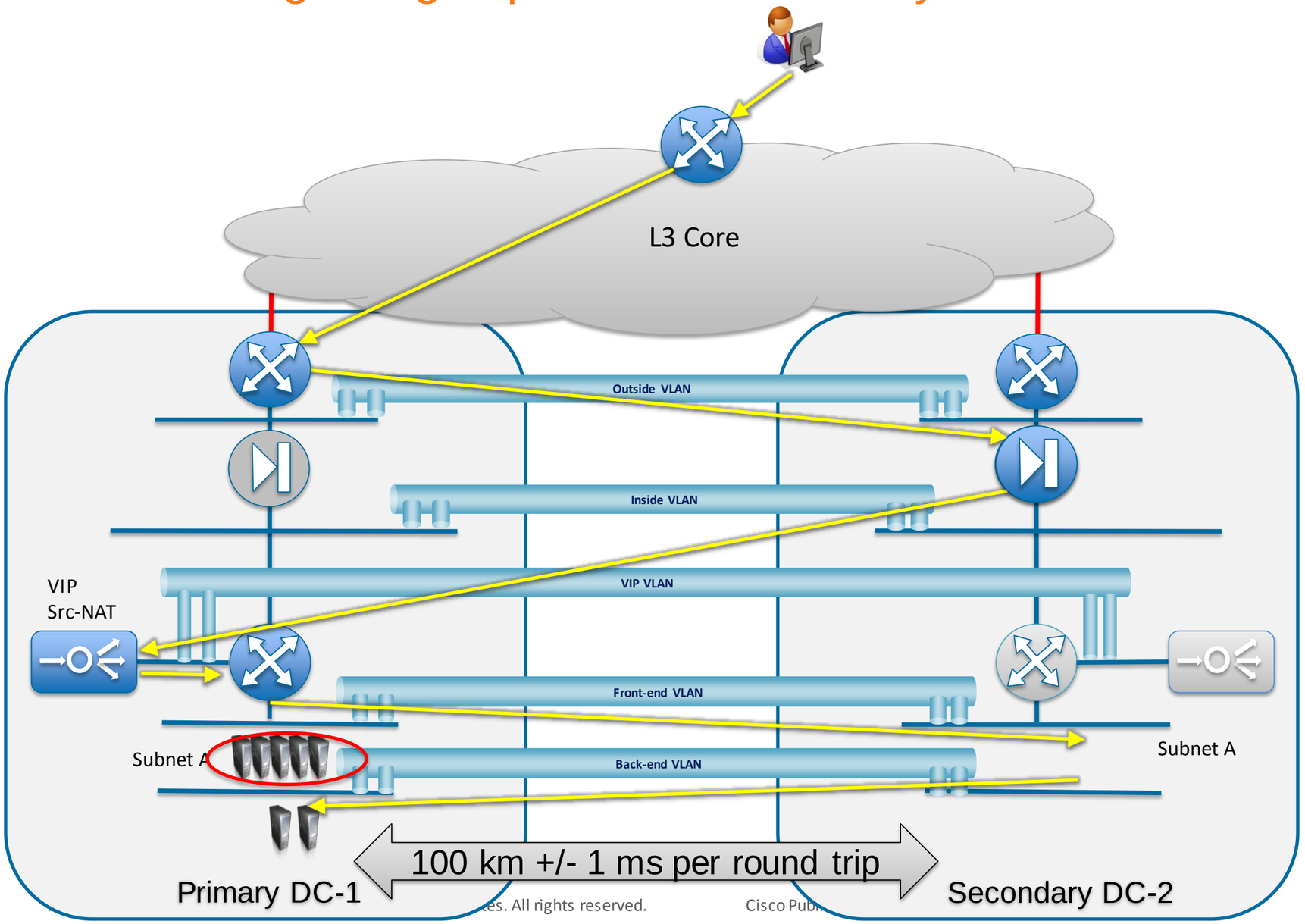
- Historically limited to Network services and HA clusters offering stateful failover & fast convergences
- Not optimum, but has been usually accepted to work in “degraded mode” with predictable mobility of Network Services under short distance

- FW failover to remote site
- Source NAT for SLB VIP
- Consider +/- 1 ms for each round trip for 100 km
- For Secured multi-tier software architecture, it is possible to measure + 10 round-trips from the initial client request up to the result.
- Interface tracking optionally enabled to maintain active security and network services on the same site

Network Service Placement for Metro Distances

Additional Ping-Pong impact with IP mobility between 2 locations

- Network team is not necessarily aware of the Application/VM mobility
- Uncontrolled degraded mode with unpredictable mobility of Network Services and Applications

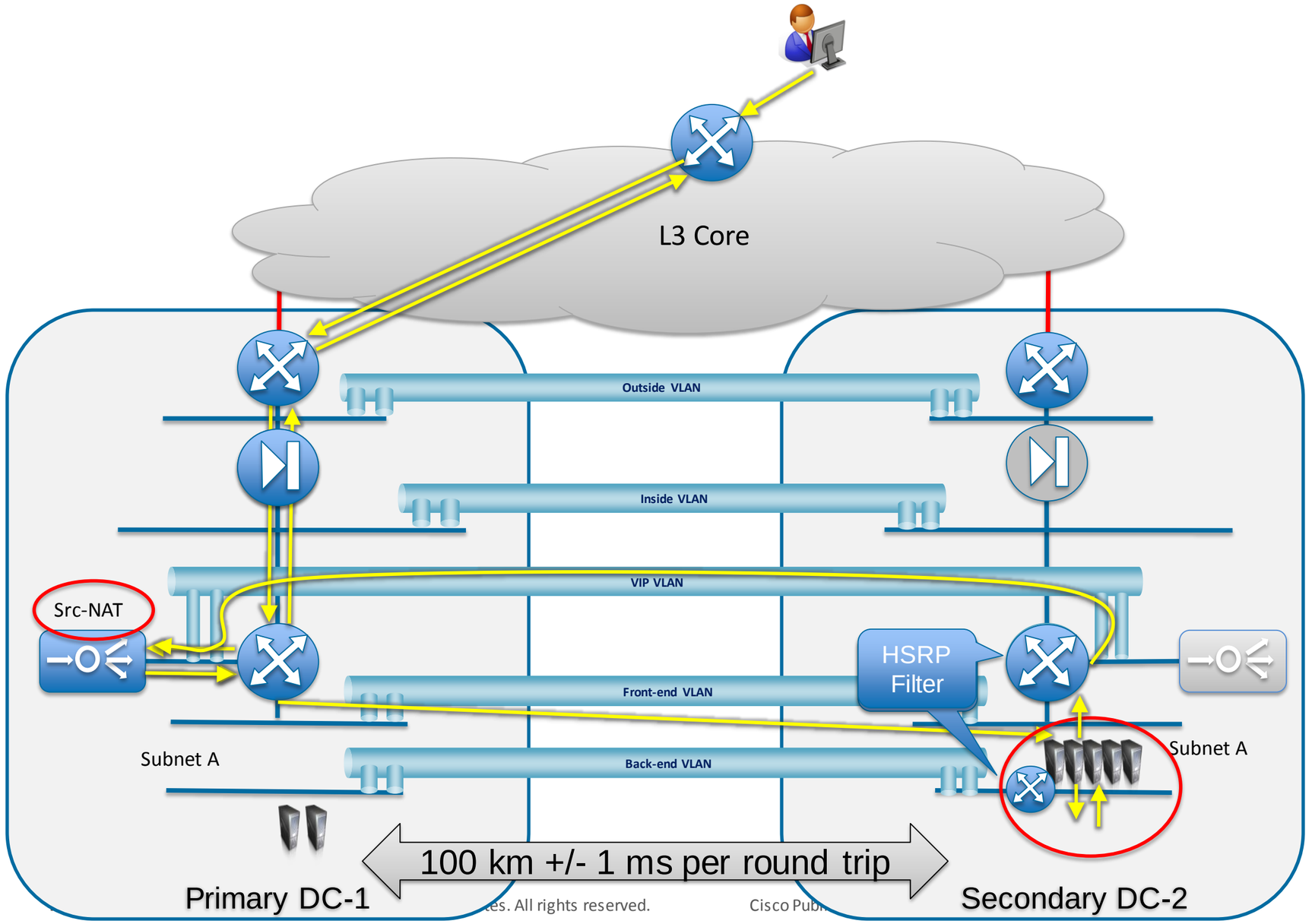


- FW failover to remote site
- Front-end server farm moves to remote site
- Source NAT for SLB VIP maintains the return path thru the Active SLB
- Partial move of a server farm is not optimised
- Understand and identify the multi-tier frameworks

Cisco live!

Network Service Placement for Metro Distances

Stateful Devices and Trombone effect for IP Mobility between 2 locations



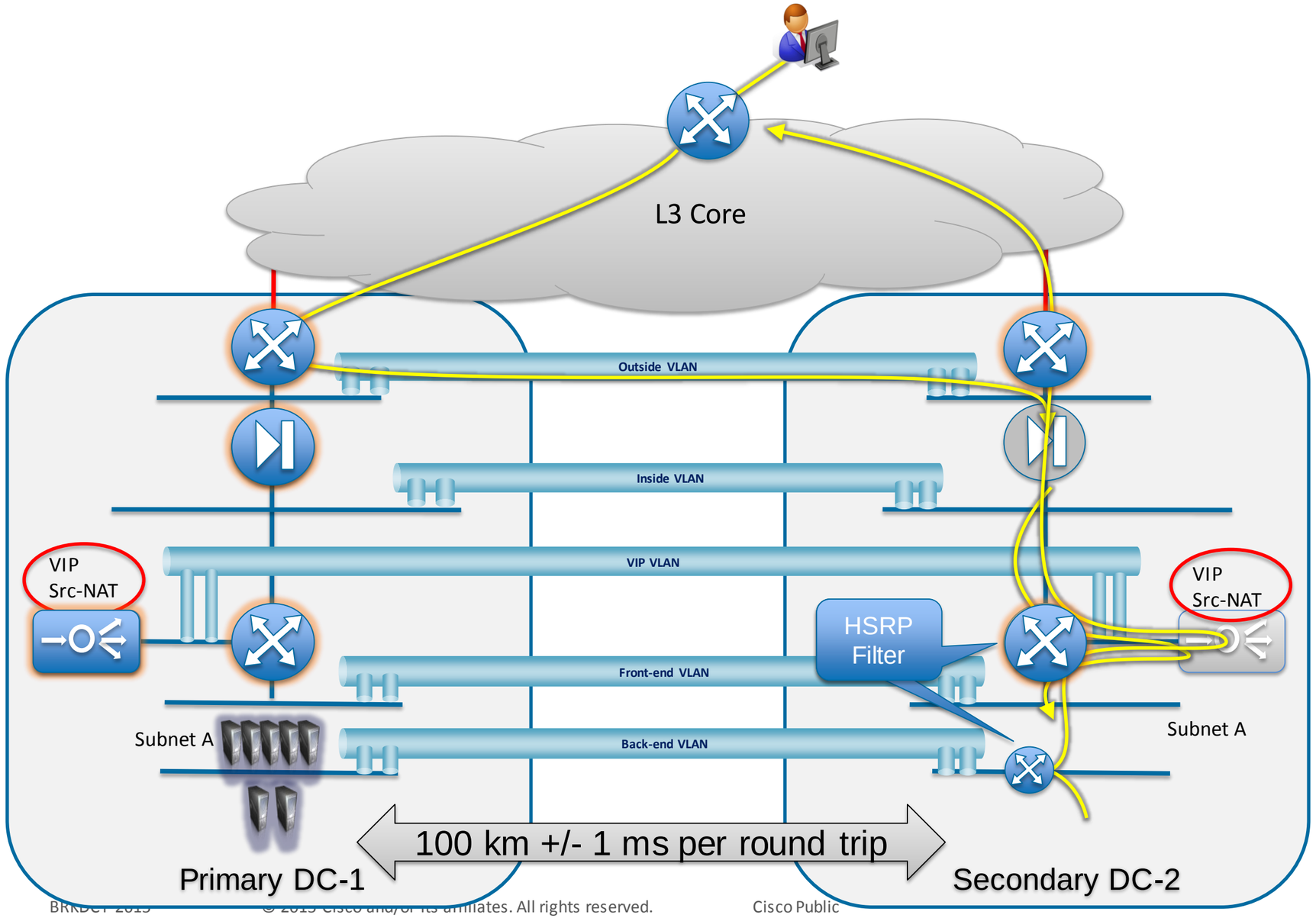
- Limited relation between server team (VM mobility) and Network Team (HSRP Filtering) and Service Team (FW, SLB, IPS..)
- Ping-Pong effect with active services placement may impact the performances

- It is preferred to migrate the whole multi-tier framework and enable FHRP filtering to reduce the trombone effect
  - FHRP filtering is ON on the Front-end & Back-end side gateways
- Source NAT for SLB VIP maintains the return path thru the Active SLB
- Understand and identify the multi-tier frameworks

Cisco live!

Network Service Placement for Metro Distances

Intelligent placement of Network Services based on IP Mobility localisation



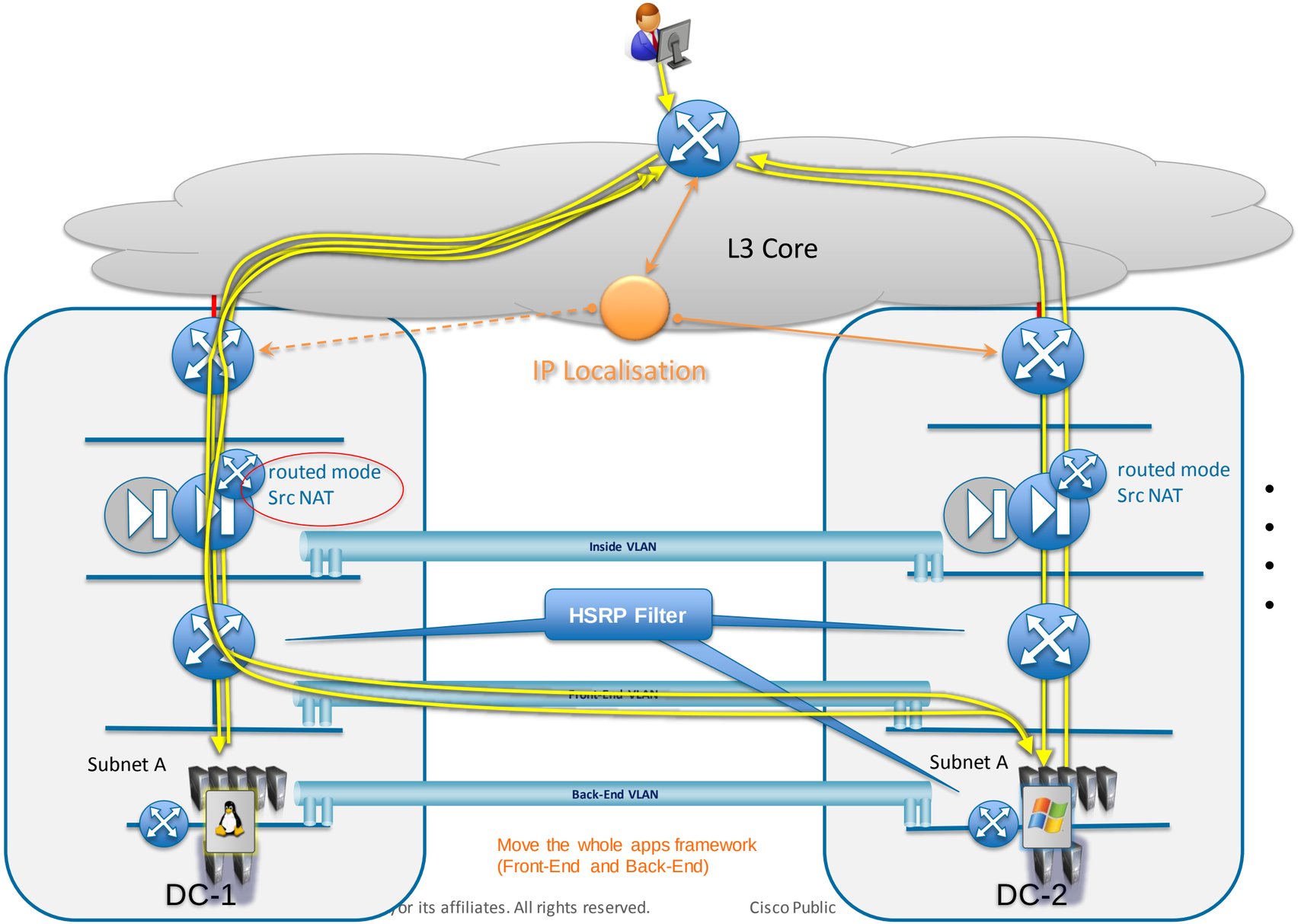
- Improving relations between sillo'ed organisations increases workflow efficiency
- Reduce trombon'ing with active services placement

- Move the FW Context associated to the application of interests
- Interface Tracking to maintain the stateful devices in same location when possible
- Return traffic keeps symmetrical via the stateful devices and source-NAT
- Intra-DC Path Optimisation almost achieved , however Ingress Path Optimisation may be required

- Sillo'ed organisations
  - Server/app
  - Network/HSRP filter
  - service & security
  - Storage

Network Service Placement for Long Distances

Active/Standby Network Services per Site with Extended LAN (Hot Live migration)



- Subnet is extended with the VLAN
- Src NAT on each FW is mandatory
- No Config nor State Synch
- Per-box per-site configuration
- Ingress Path Optimisation can be initiated to reduce trombone effect due to active services placement
- With Ingress Path optimisation the current Active TCP sessions are interrupted and re-established on the remote site

- Extend the VLAN of interests
- FW and SLB maintain stateful session per DC.
- No real limit in term of number of DC
- Granular migration is possible only using LISP or IGP Assist or RHI (if the Enterprise owns the L3 core)

Can Cisco ASA Firewall help here ?

Current Statement:

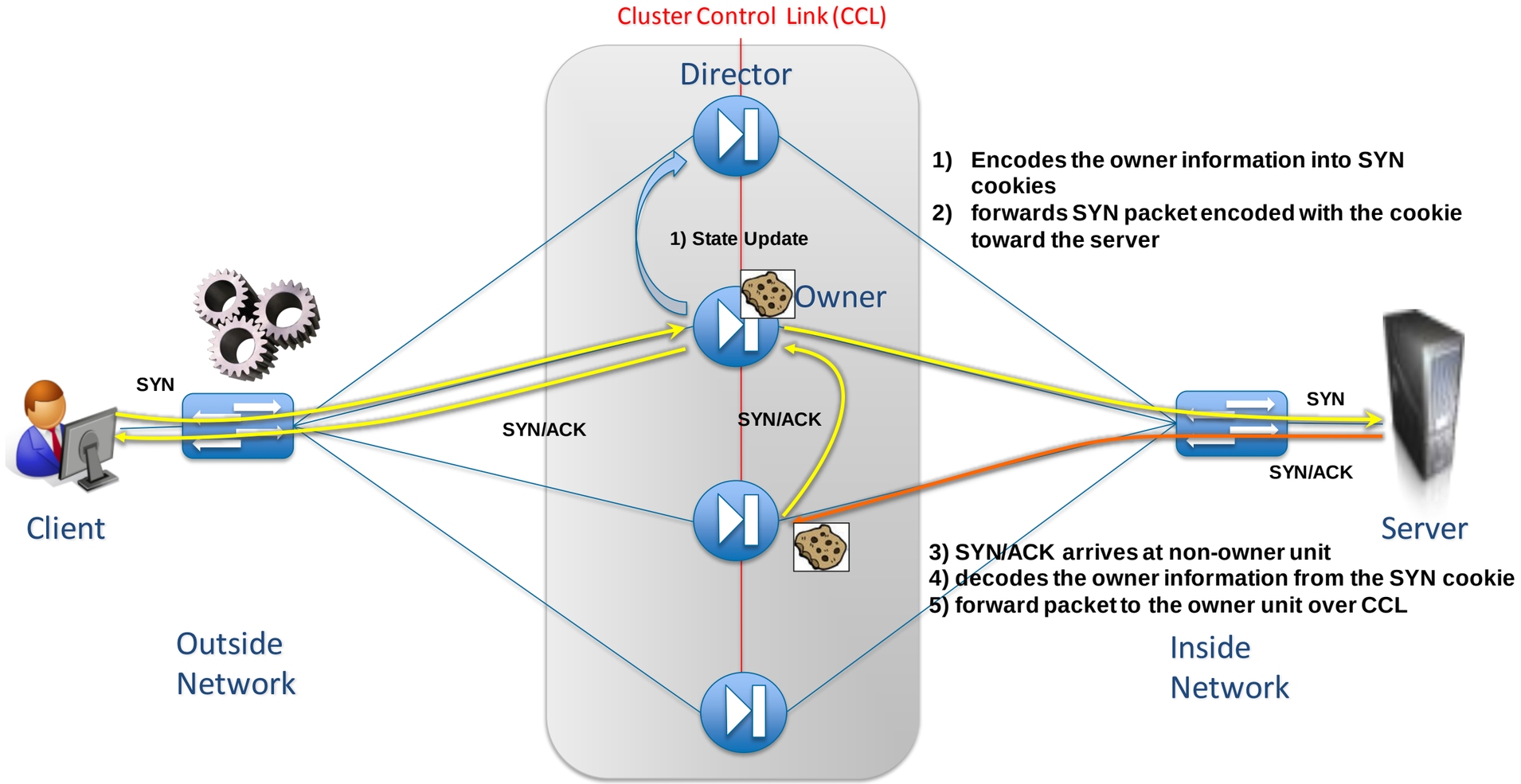
ASA clustering spanned across multiple locations is now supported in all modes:

- 9.1(4) Individual Interface mode (L3) – ASA in Routed mode (Firewall Routed Mode)
- 9.2(1) Spanned Interface mode (L2) – North-South Insertion (Firewall Transparent Mode)
  - *NO extended VLAN used by the Cluster LACP*
- 9.3(2) Spanned Interface mode (L2) – East-West Insertion (Firewall Transparent Mode)

- Non currently supported: Spanned Interface mode with ASA in Firewall Routed mode

ASA Clustering (9.0)

TCP SYN cookies with Asymmetrical Traffic workflows

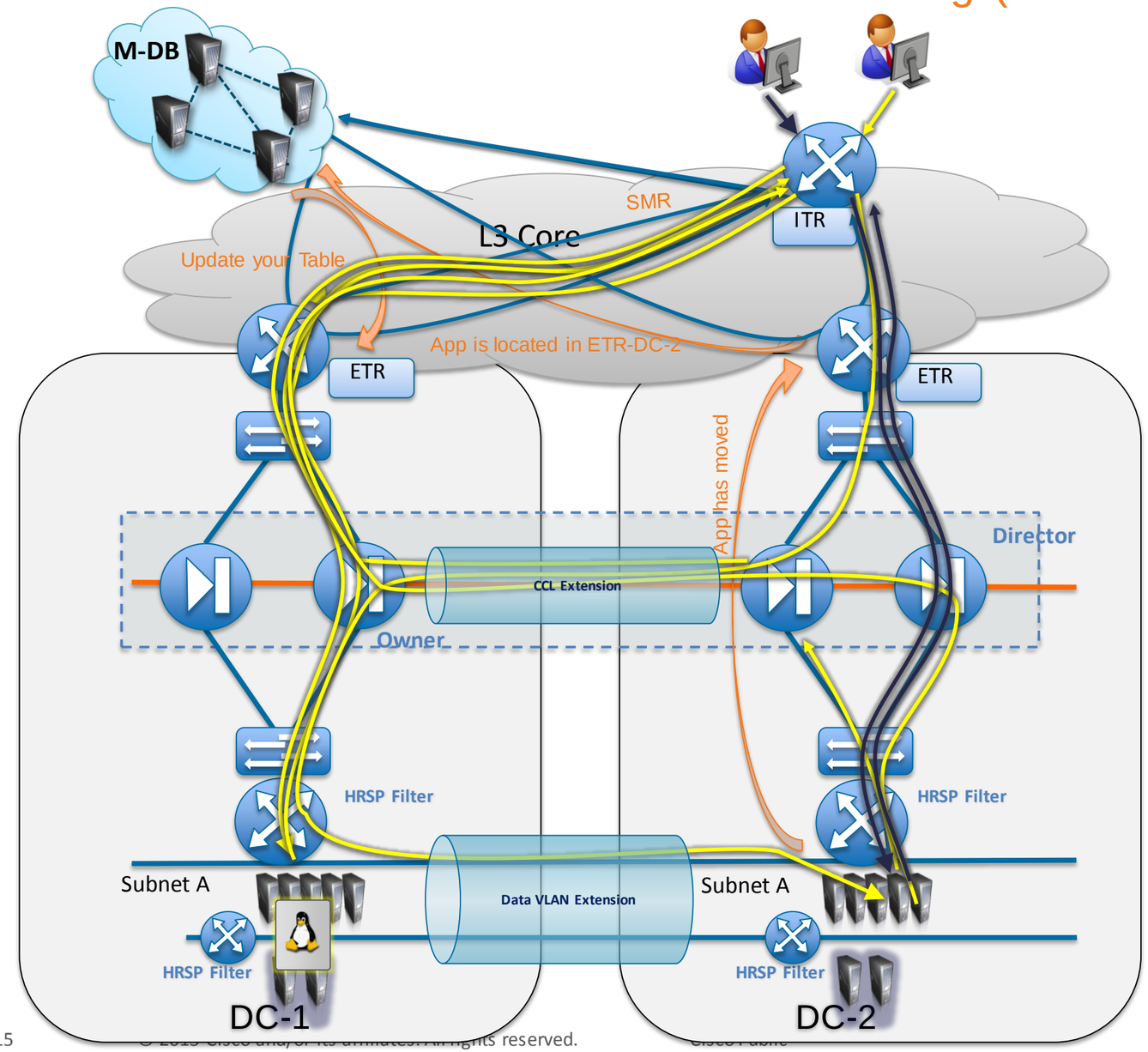


It is possible that the SYN/ACK from the server arrives at a non-owner unit before the connection is built at the director.

- As the owner unit processes the TCP SYN, it encodes within the Sequence # which unit in the cluster is the owner
- Other units can decode that information and forward the SYN/ACK directly to the owner without having to query the director

Single ASA Clustering Stretched Across Multiple DC

Case 1: LISP Extended Subnet Mode with ASA Clustering (Stateful Live migration with LAN extension)



- One Way Symmetric Establishment is achieved via the CCL
- Current active sessions are maintained stateful
- New Sessions are optimised dynamically
- Up to 10ms max one-trip latency, extend the ASA cluster for config and state synch

1 - End-user sends Request to App
2 - ITR intercepts the Req and check the localisation
3 - MS relays the Req to the ETR of interests which replies location for Subnet A being ETR DC-1
3” - ITR encaps the packet and sends it to RLOC ETR-DC-1

4 – LISP Multi-hop notifies ETR on DC-2 about the move of the App
5 – Meanwhile ETR DC-2 informs MS about new location of App
6 – MR updates ETR DC-1
7 – ETR DC-1 updates its table (App:Null0)
8 – ITR sends traffic to ETR DC-1
9 – ETR DC-1 replies with a Solicit Map Req
8 – ITR sends a Map Req and redirects the Req to ETR DC-2

Recommendations

1. Layer 2 extensions represent a challenge for optimal routing
2. Consider the implications of stretching the network and security services over multiple DCs
3. For migration over long distances, when possible enable network path optimisation for traffic :
   - Client to server communication (Ingress Optimisation)
   - Server to Client communication for symmetrical return traffic (Egress optimisation)
   - Server to Server communication (bandwidth and Latency optimisation)
4. Otherwise provision enough bandwidth (2 times the needs) and compute the total latency due to ping-pong workflows
5. When moving a VM /Tier, move all the framework
6. Network and security policies must be maintained
7. Consider FW clustering stretched across DC's to reduce the hair-pining workflow

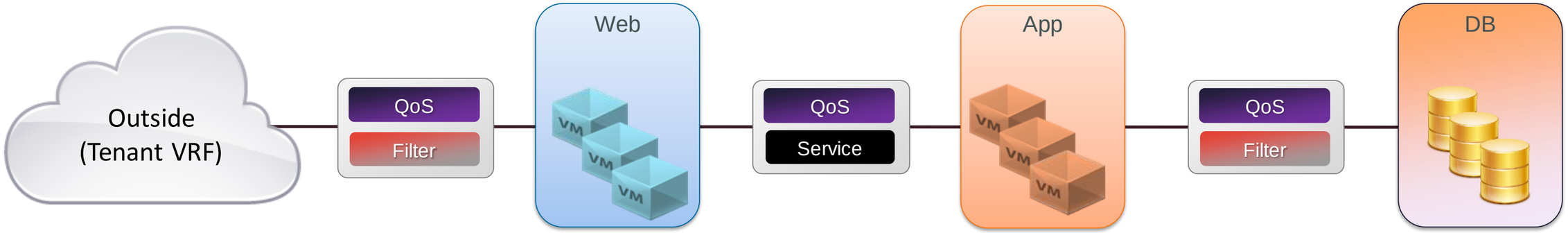
http://yves-louis.com/DCI/?p=785 (Post 27.n)

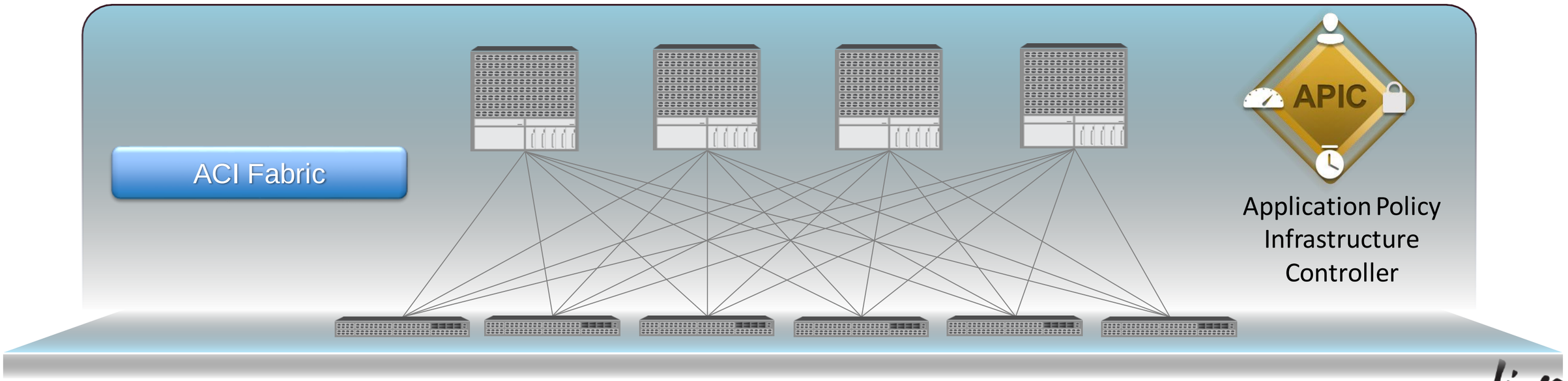
Agenda

- Active-Active (A/A) Data Centre:
  - Market & Business Drivers
  - Terminology, Criticality levels and Solutions Overview
- A/A Data Centre Design Considerations:
  - Storage Extension
  - Data Centre Interconnect (DCI) – L2 & L3 scenarios
- A/A Metro Data Centres Designs

  - Network Services and Applications (Path optimisation)
- Cisco ACI and Active / Active Data Centre
- Q&A

ACI Introduces Logical Network Provisioning of Stateless Hardware with Application Network Profile (ANP)



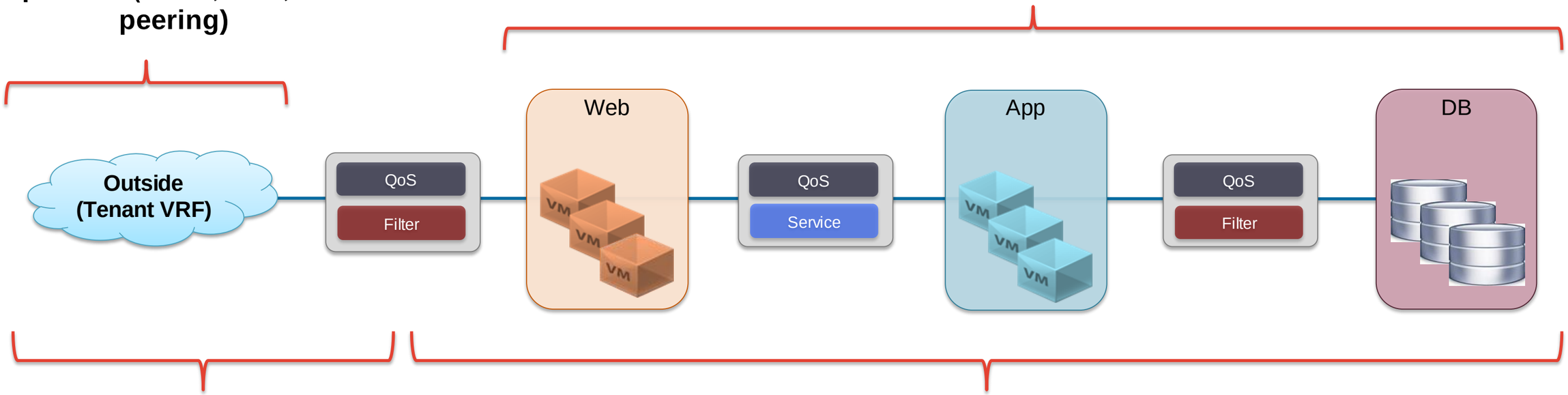


Fabric Infrastructure

Important Concepts – Inside and Outside

‘Outside’ EPG associated with external network policies (OSPF, BGP, … peering)

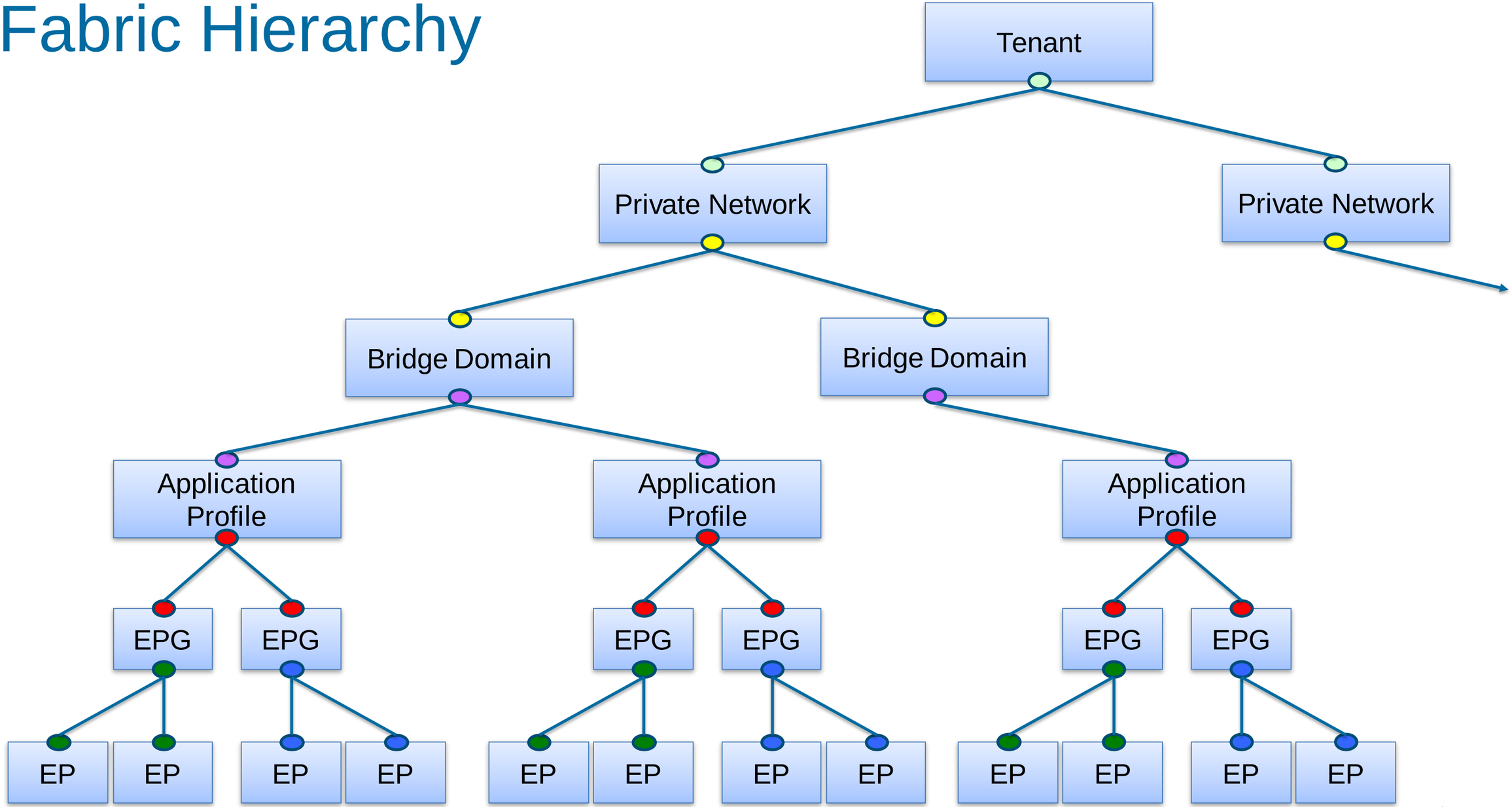
Forwarding Policy for ‘inside’ EPG’s defined by associated Bridge Domain network policies



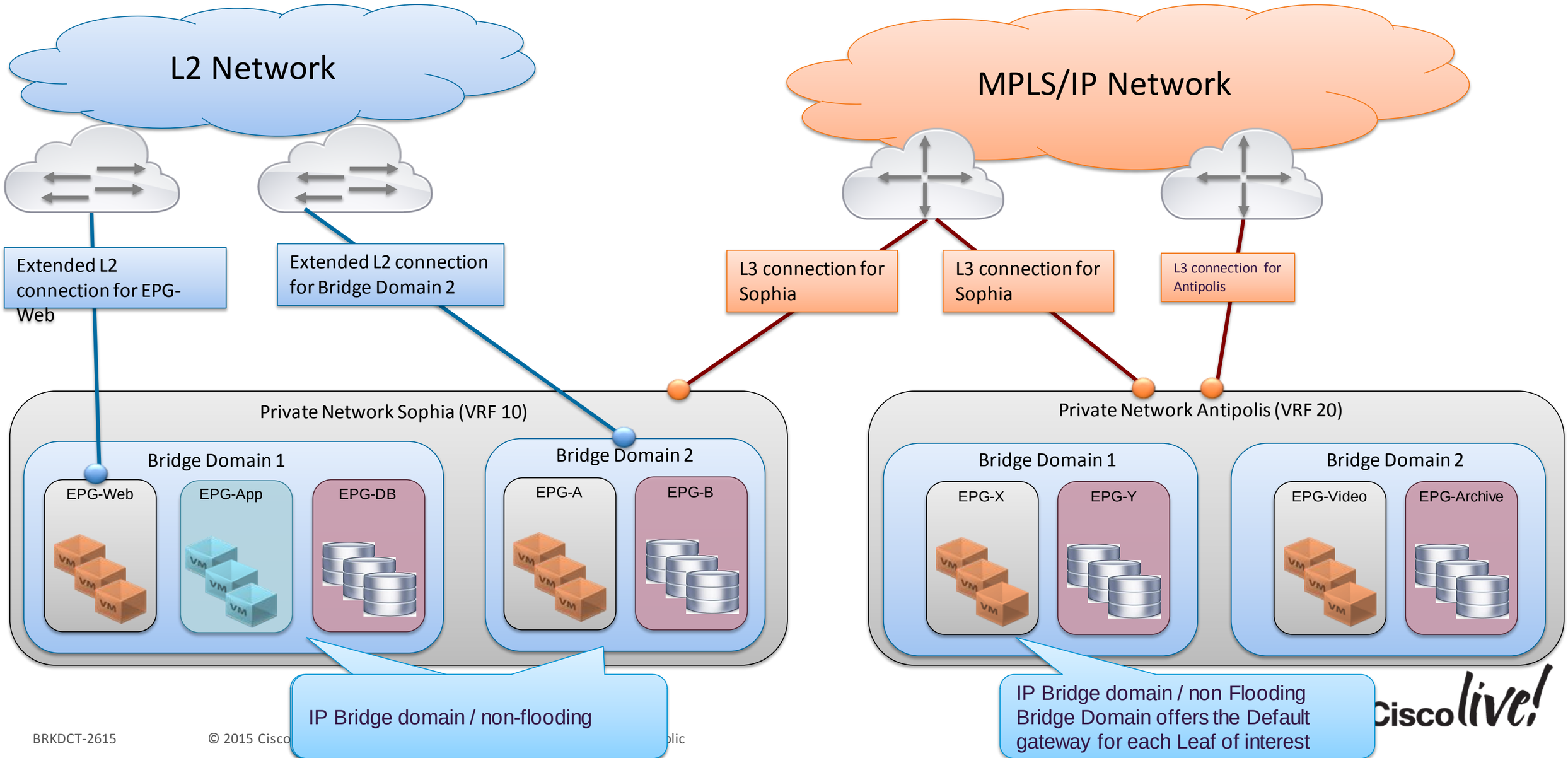
Location for Endpoints that are ‘Outside’ the Fabric are found via redistributed routes sourced from the externally peered routers (Network Level Granularity)

Location for Endpoints that are ‘Inside’ the Fabric are found via the Proxy Mapping DB (Host Level Granularity)

Fabric Hierarchy



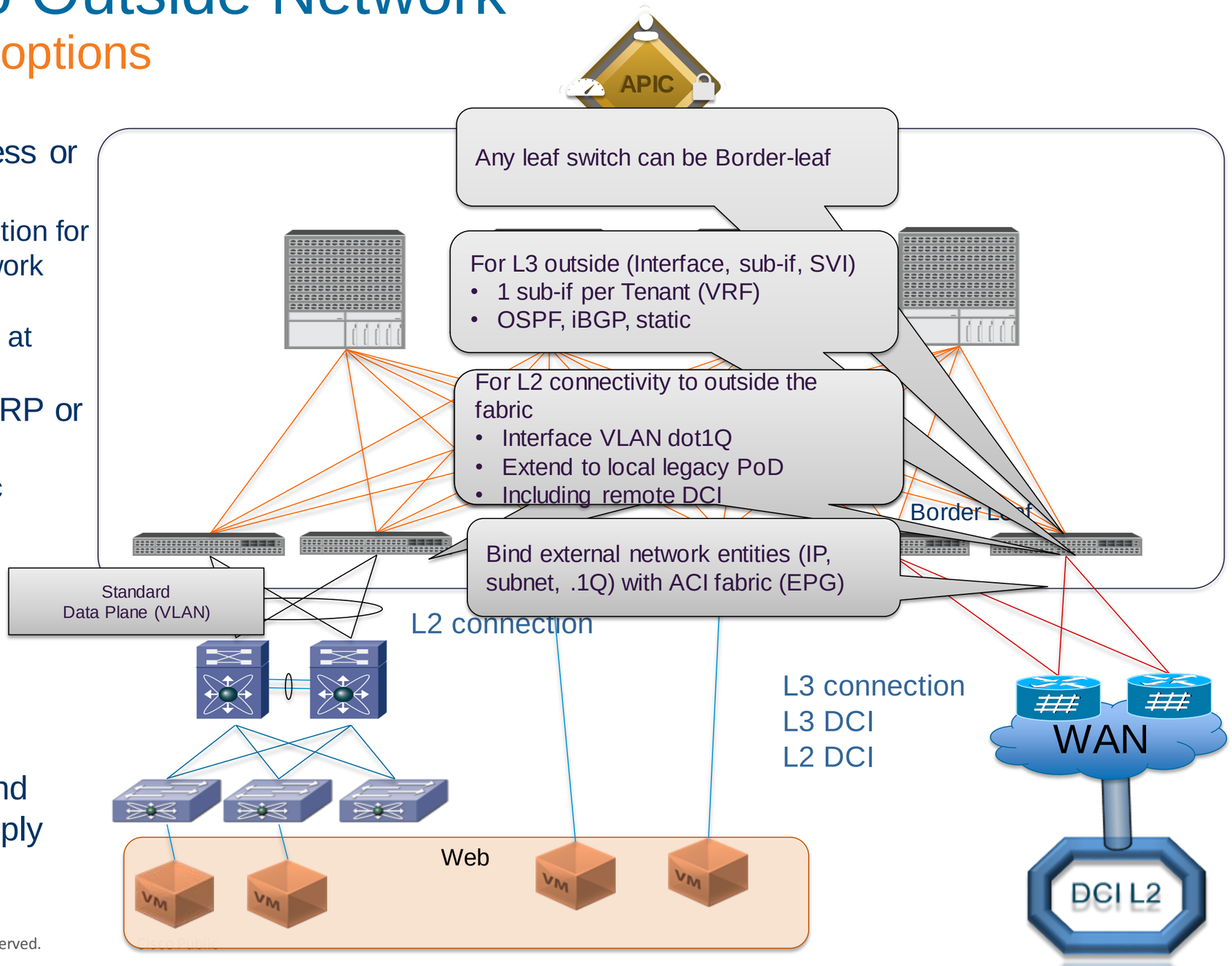
ACI Connection to Outside Network
Relationship to rest of components (Connectivity view)



ACI Connection to Outside Network

Reference topology and options

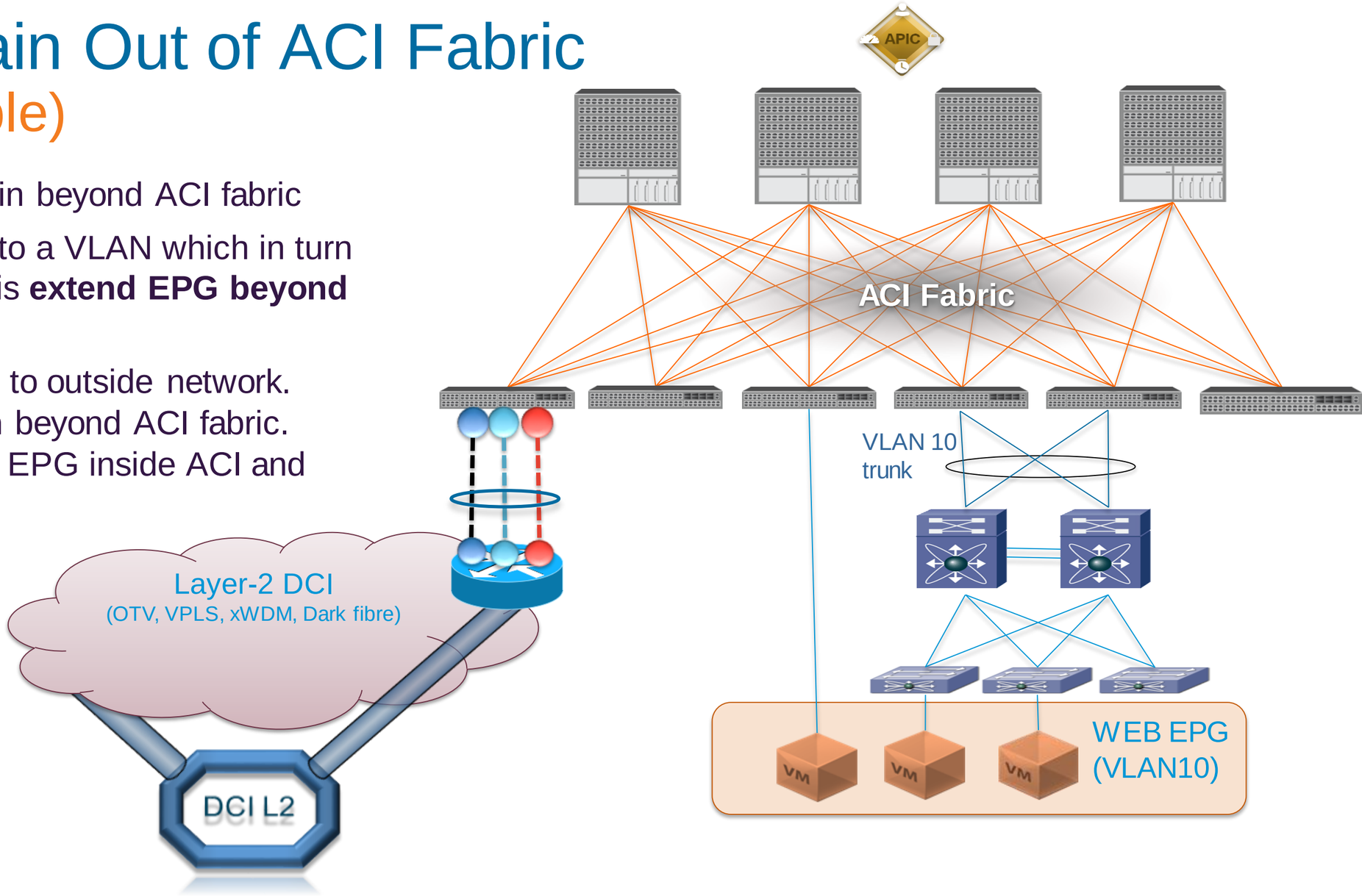
- L3 connection to outside (Internet access or DCI/L3 for DRP)
  - L3 (L3 port, sub-interface, SVI) connection for tenants connecting to existing DC network
  - VRF-lite for tenant isolation
  - OSPFv2 (NSSA),iBGP and static route at FCS
- L2 connection to outside (DCI/L2 for DRP or DAP)
  - Extend L2 domain outside of ACI fabric
  - brownfield migration
  - L2 extend across POD/site
  - vPC and STP connection
- WAN Edge focus:
  - ASR 9000, Nexus 7000, ASR 1000
- Existing principles of Inbound, Outbound traffic flows, security, DNS/GSS still apply



Extend L2 Domain Out of ACI Fabric

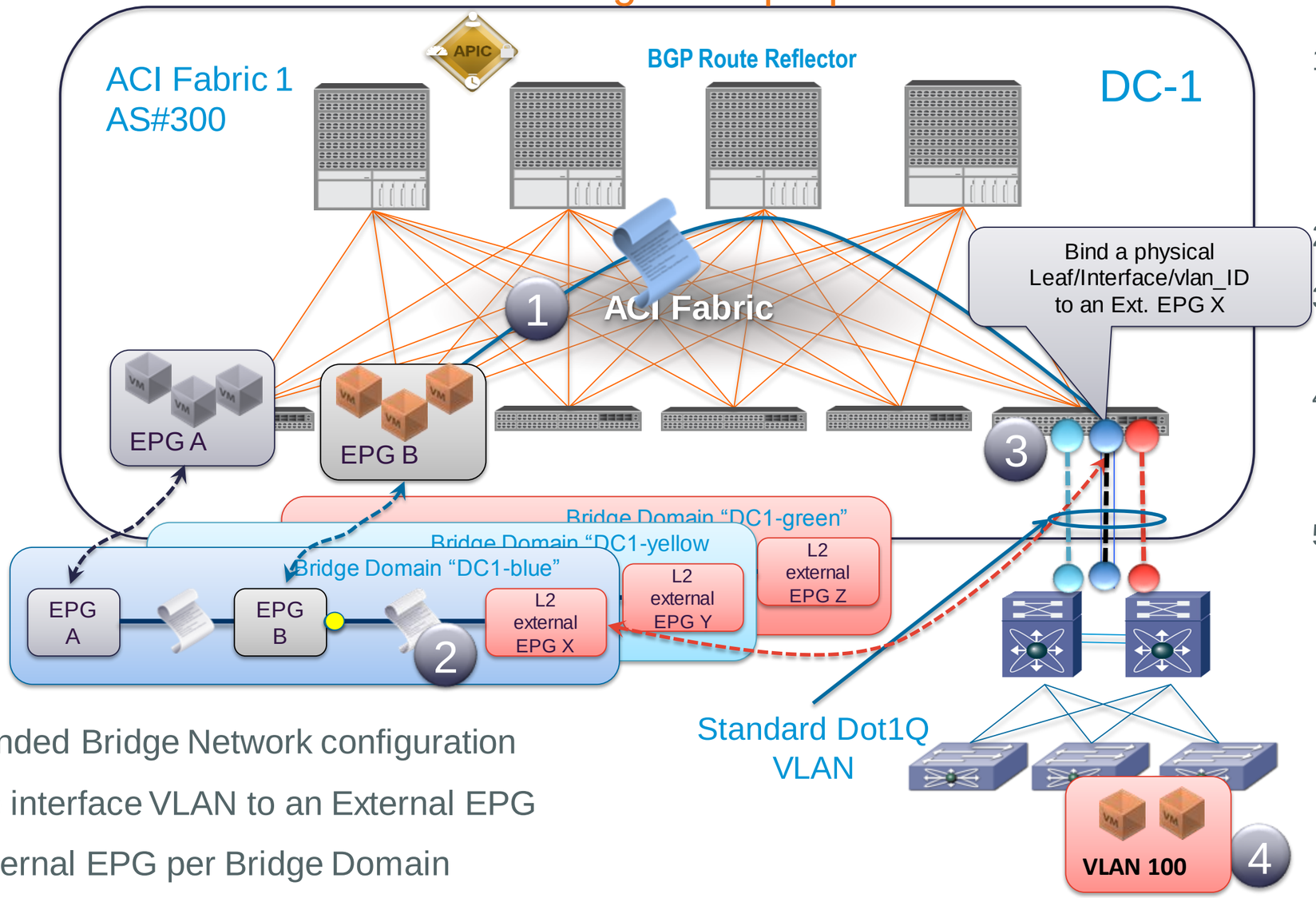
3 options (2 available)

- Three ways of extend L2 domain beyond ACI fabric
  1. Manually assign a port to a VLAN which in turn mapped to an EPG. This **extend EPG beyond ACI fabric**
  2. Create a L2 connection to outside network. **Extend bridge domain** beyond ACI fabric. Allow contract between EPG inside ACI and EPG outside of ACI
  3. Remote VTEP (future)



ACI L2 Connection to Outside Network

Integration with brownfield DC and Migration purpose



1. Extends the Bridge Domain beyond ACI fabric mapping the Leaf/Interface/VLAN to the External EPG
2. Apply contract (filter, QoS…)
3. Interface-VLAN trunk to outside
4. L2 adjacency establishment between the BD and outside the fabric
5. Traffic is policed in one or both ways according to the contract

Via the Extended Bridge Network configuration

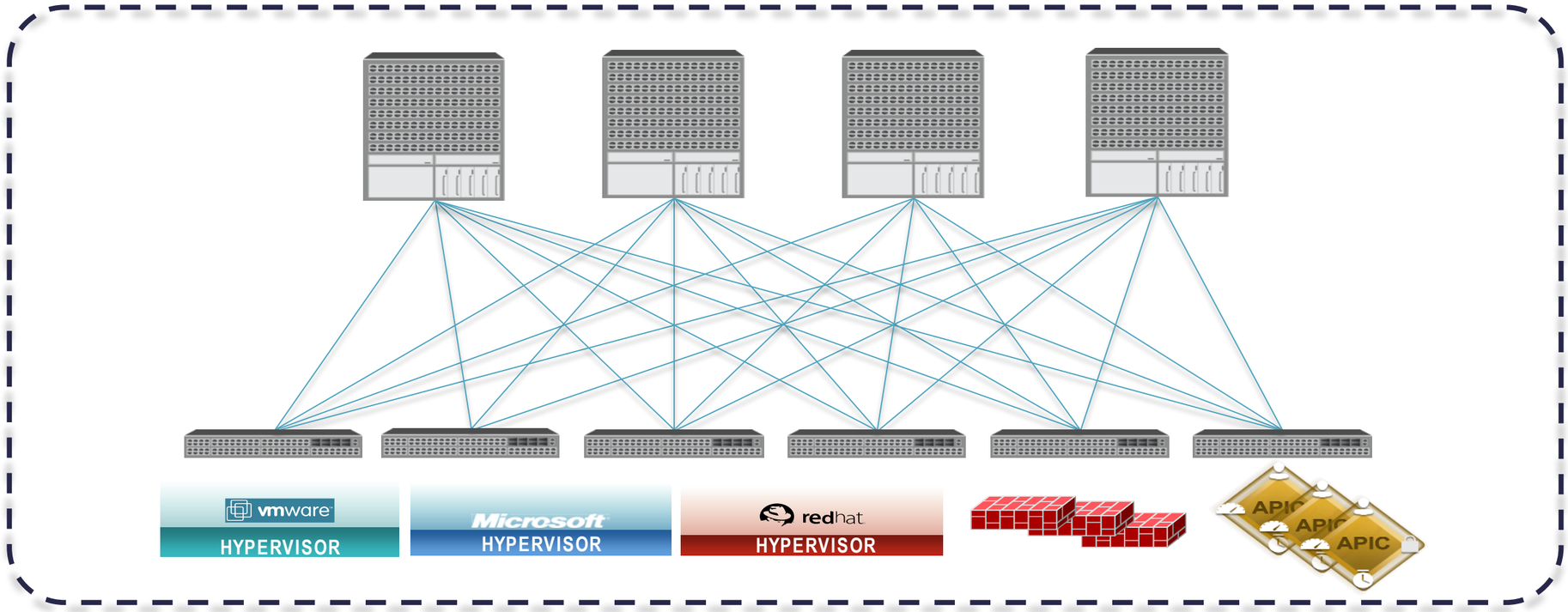
- Map the interface VLAN to an External EPG
- One External EPG per Bridge Domain
- Ethernet: Interface, VLAN, VxLAN (future)

Use cases

- Extend to legacy Access PoD
- Migration purposes
- UCS/F.I. attachment

Single Fabric Scenarios

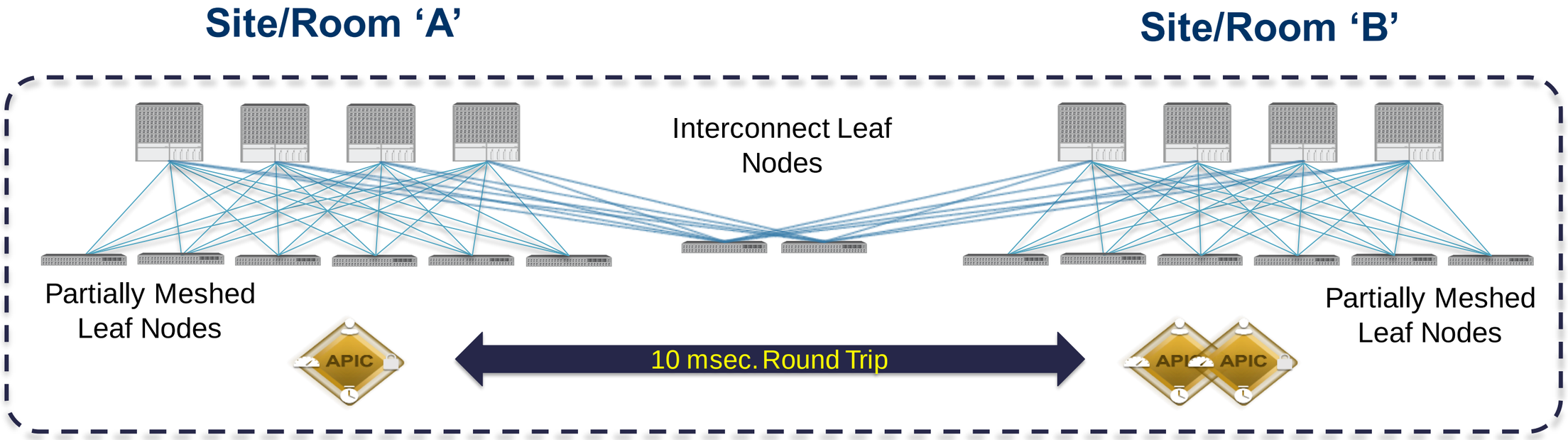
Single Site (Single Pod)



- Single Fabric + Single Site
  - Single Operational Zone (Network, VMM Cluster, Storage, FW/LB are all treated as if it is ‘one’ zone)
  - Can support multiple Openstack Availability Zones (Multiple VMM, Compute Zones)

Single Fabric Scenarios

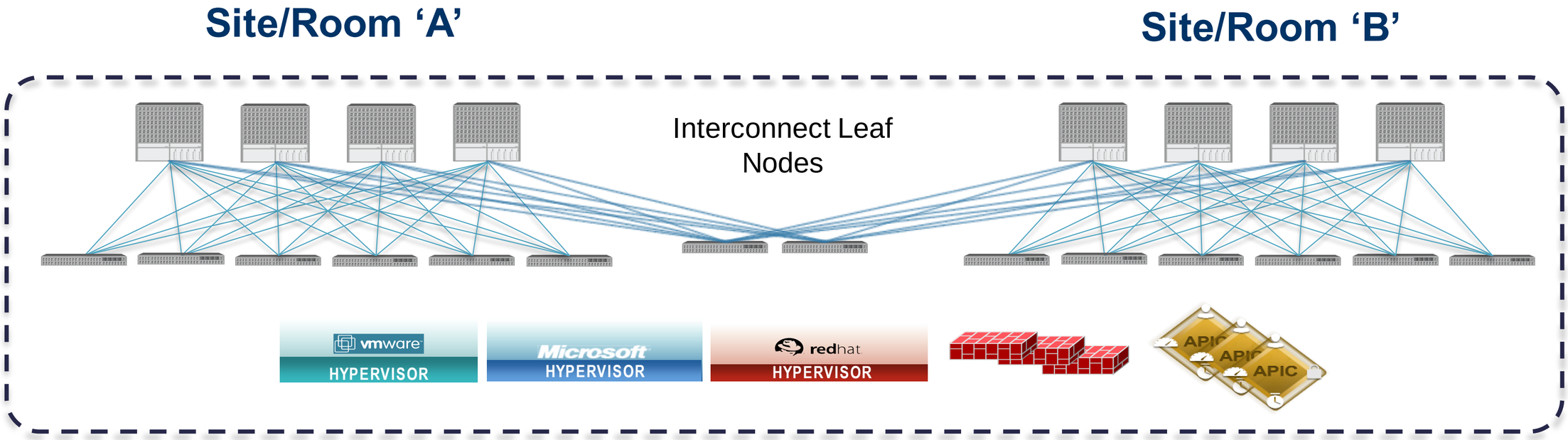
Multi-Site (Stretched) Fabric



- Single Fabric + Multi-Site
- Use Cases
  - Multi-Building cross campus and metro distances (Dual site cross-metro design is a very common topology)
  - Multi-Floor, Multi-Room Data Centres (cabling restrictions prevent full mesh)

Single Fabric Scenarios

Multi-Site (Stretched) Fabric

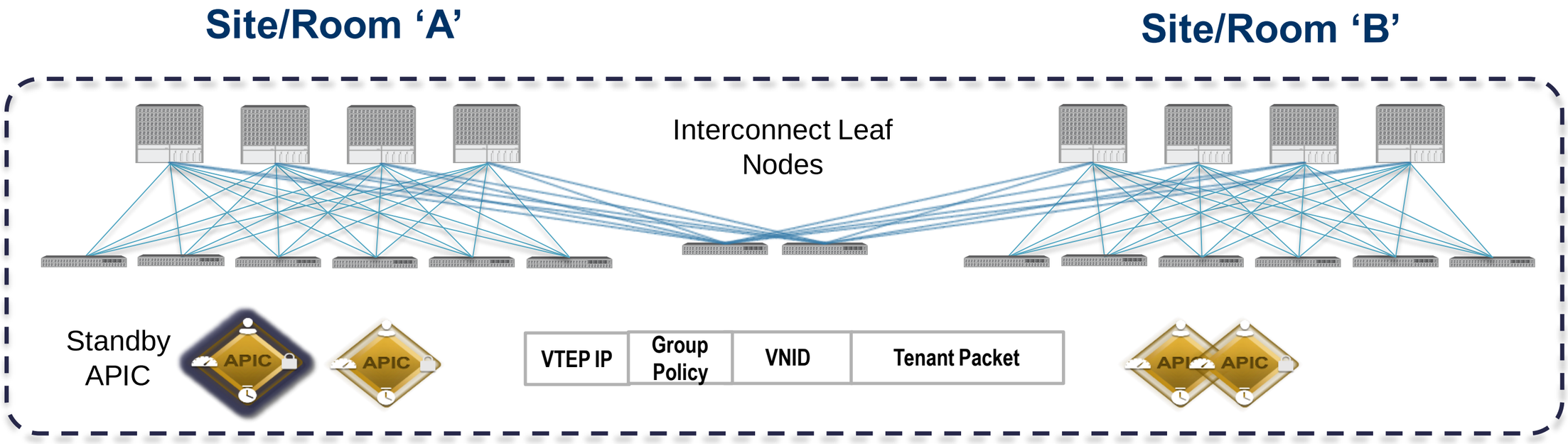


- Single Fabric + Multi-Site
  - Single Operational Zone (VMM, Storage, FW/LB are all treated as if it is ‘one’ zone)
    - e.g. Single vCenter with Synchronized Storage
- Interconnect between sites
  - Direct Fibre (40G), DWDM (40G or multiple 10G), Pseudo Wire (10G or 40G)

http://www.cisco.com/c/en/us/td/docs/switches/datacenter/aci/apic/sw/kb/b_kb-aci-stretched-fabric.html

Single Fabric Scenarios

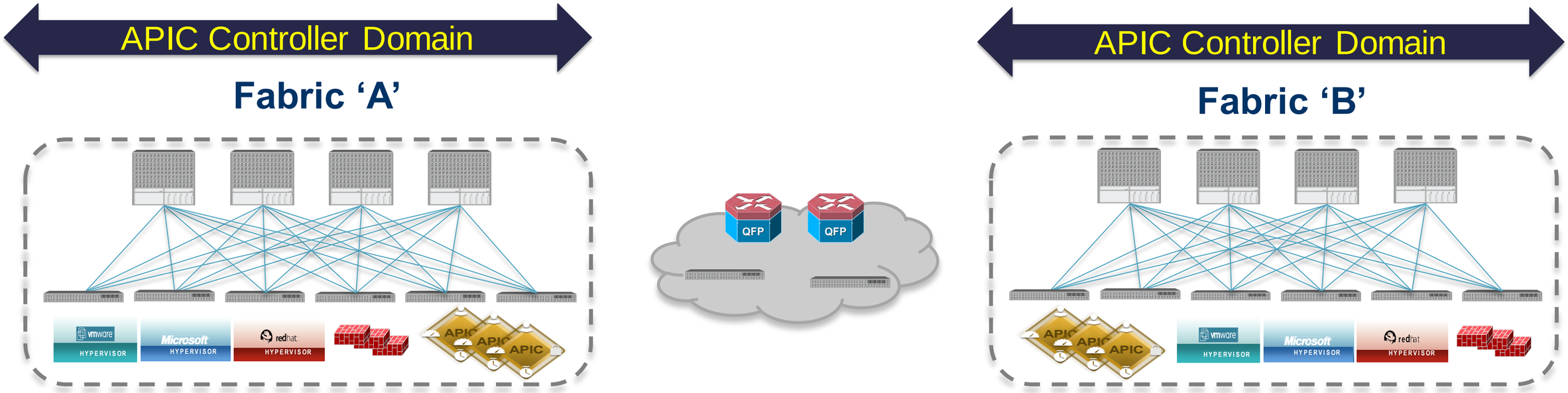
Multi-Site (Stretched) Fabric



- Single APIC Cluster (Leverages 3 Active + 1 Standby APIC for Site Failure Scenarios)
- Single Operational Domain (Changes are immediately propagated to all nodes in the fabric)

- Single Infrastructure
- Single VRF, Bridge Domain and EPG name space (one network)
- WAN links shared

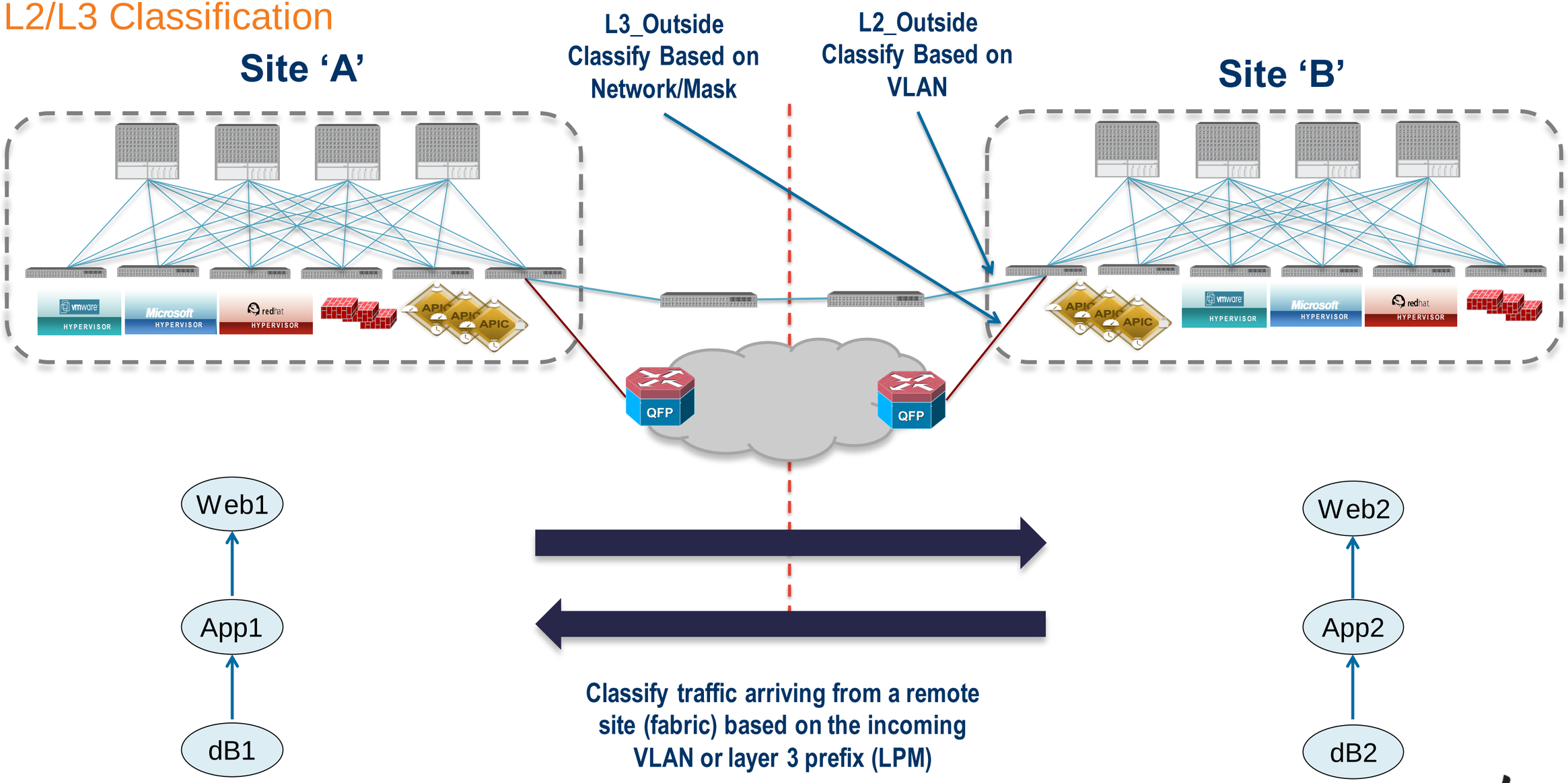
Multi-Fabric Scenarios



- Multi-Fabric Use Cases
  - Multiple Fabrics within a single site (includes Multi-Pod, Multi-Floor, Multi-Room Data Centres)
  - Multi-Building cross campus and metro distances (Majority of larger Enterprise and SP/Cloud customers require a dual site active/active ‘in region’ design)
  - Multiple Fabrics across ‘out of region’ Data Centres (almost all customers leverage some form of out of region Data Centre)

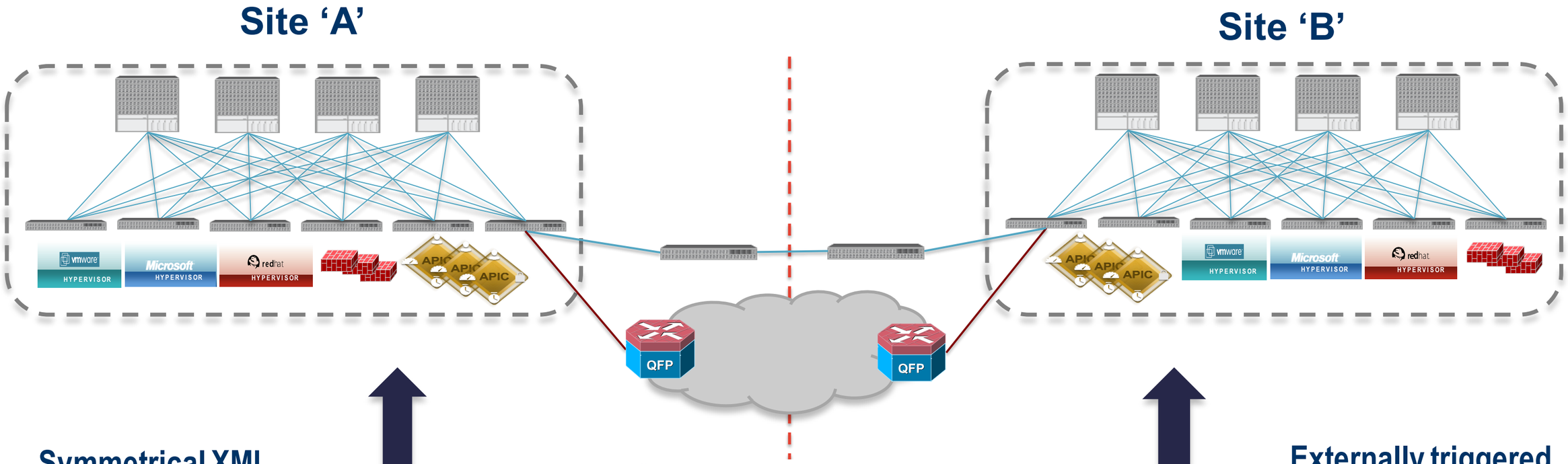
Multi-Fabric – Current Options

L2/L3 Classification



Multi-Fabrics – Current Options

External Synchronisation of Fabric Policy



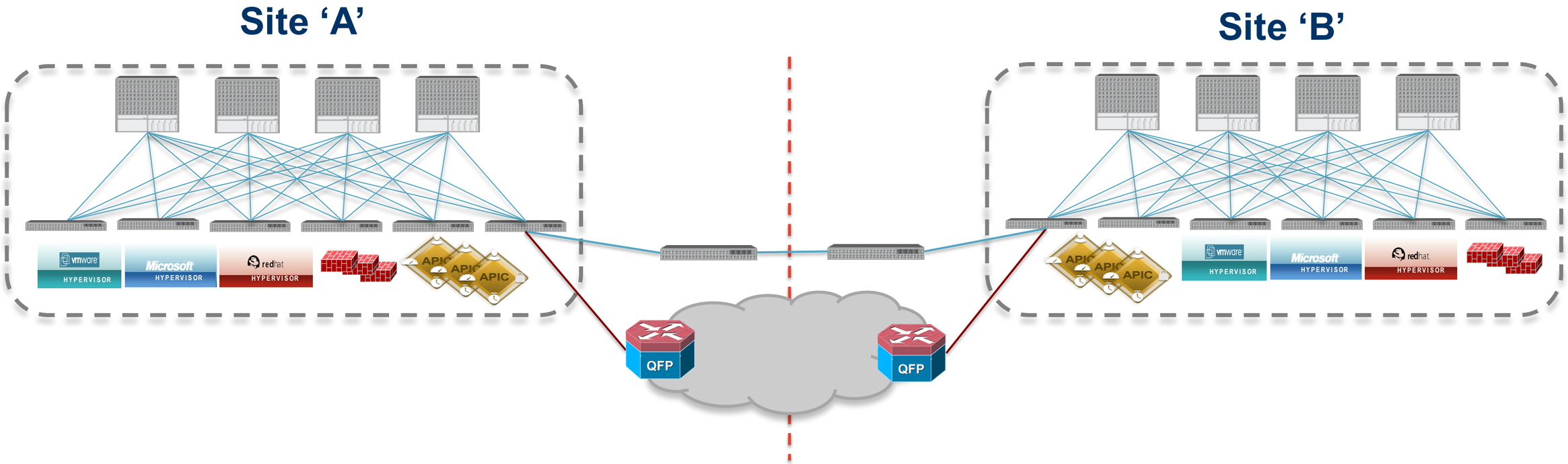
Symmetrical XML Configuration will maintain consistent operation between fabrics

```
<polUni>
   <fvTenant dn="uni/tn-Customer1" name="Customer1">
   <fvCtx name="customer1-router"/>
   <fvBD name="BD1">
      <fvRsCtx tnFvCtxName="customer1-router" />
      <fvSubnet ip="10.0.0.1/24" scope="public"/>
   </fvBD>
     <vzBrCP name="ALL">
          <vzSubj name="any">
          <vzRsSubjFiltAtt tnVzFilterName="default"/>
       </vzSubj>
    </vzBrCP>
   <fvAp name="web-and-ordering">
    <fvAEPg name="VLAN10">
      <fvRsBd tnFvBDName="BD1" />
      <fvRsCons tnVzBrCPName="ALL"/>
      <fvRsProv tnVzBrCPName="ALL" />
    </fvAEPg>
  </fvTenant>
</polUni>
```

Externally triggered Export and Import between Fabrics is another option to maintain consistency

Multi-Fabrics – Current Options

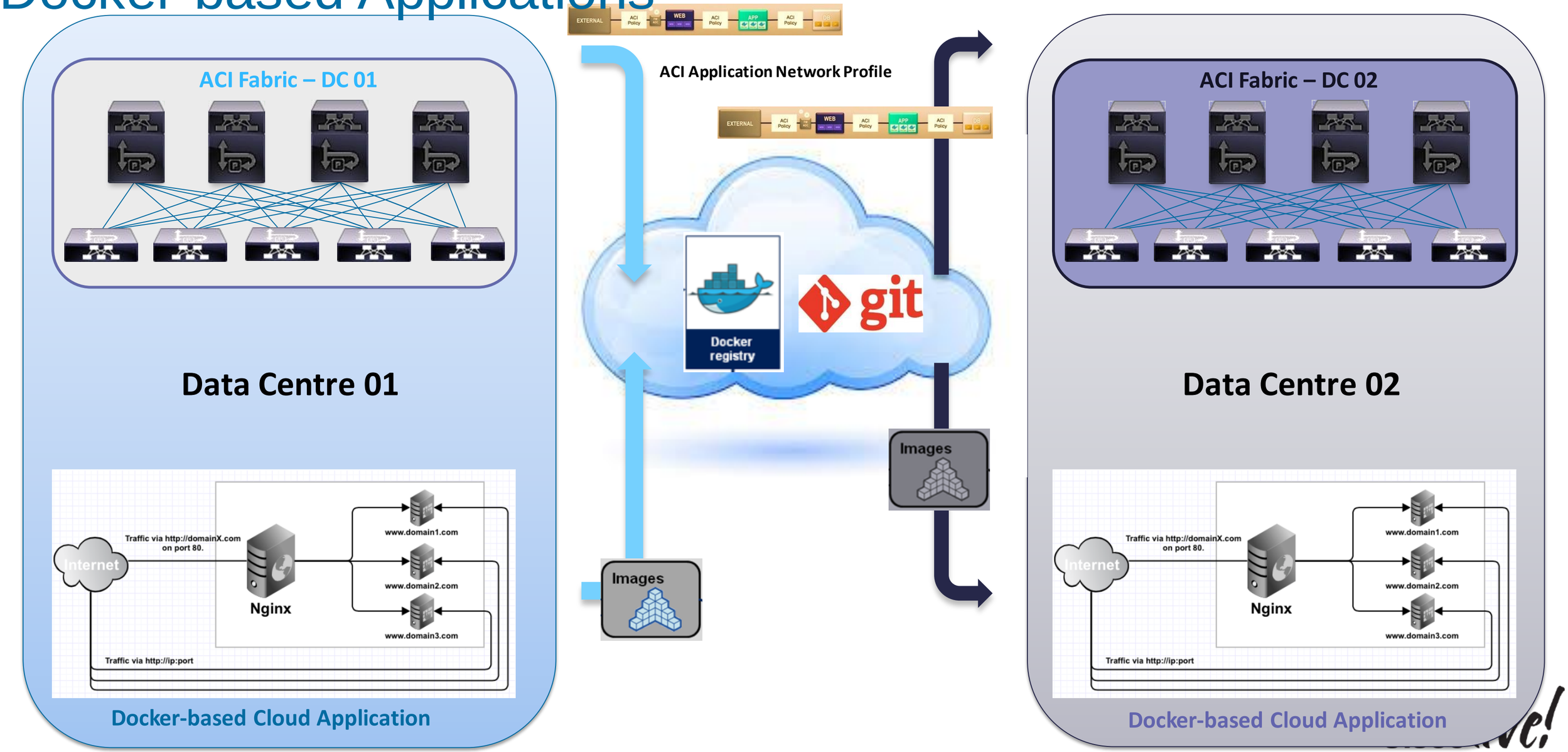
Multiple Domains



- Multiple APIC Clusters (N+1 Redundancy for each Fabric)
- Multiple Operational Domains (Changes are immediately propagated to all nodes ‘within’ each fabric, external tools correlate across fabrics)

- Replication of VRF, Bridge Domain and EPG name spaces (multiple networks)
- Layer 2 and/or Layer 3 interconnect between fabrics
- VMM Clusters can be extended if ‘not’ integrated with APIC

Multi-site Abstraction and Portability of Network Metadata and Docker-based Applications



Agenda

- Active-Active (A/A) Data Centre:
  - Market & Business Drivers
  - Terminology, Criticality levels and Solutions Overview
- A/A Data Centre Design Considerations:
  - Storage Extension
  - Data Centre Interconnect (DCI) – L2 & L3 scenarios
- A/A Metro Data Centres Designs

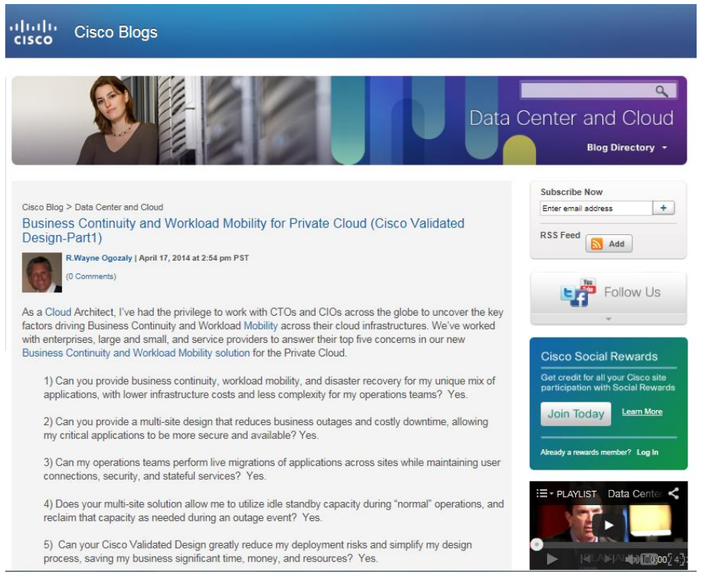
  - Network Services and Applications (Path optimisation)
- Cisco ACI and Active / Active Data Centre
- Q&A

Social Media Collateral for Cisco DCI CVD

Cloud Business Continuity and Workload Mobility content

15,000+ views

Blogs on Cisco.com



http://blogs.cisco.com/datacenter/business-continuity-and-workload-mobility-for-the-private-cloud-cisco-validated-design-part-1/

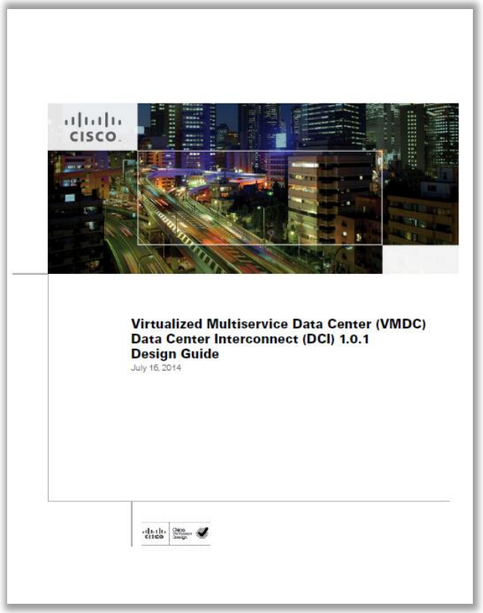
TechWise TV Show

TechWise TV
Business Continuity and Workload Mobility
http://cs.co/9000my6i

5,000+ downloads

Design Guide

Business Continuity and Workload Mobility for Cloud



http://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Data_Center/VMDC/DCI/1-0-1/DG/DCI.html


Q & A

Complete Your Online Session Evaluation

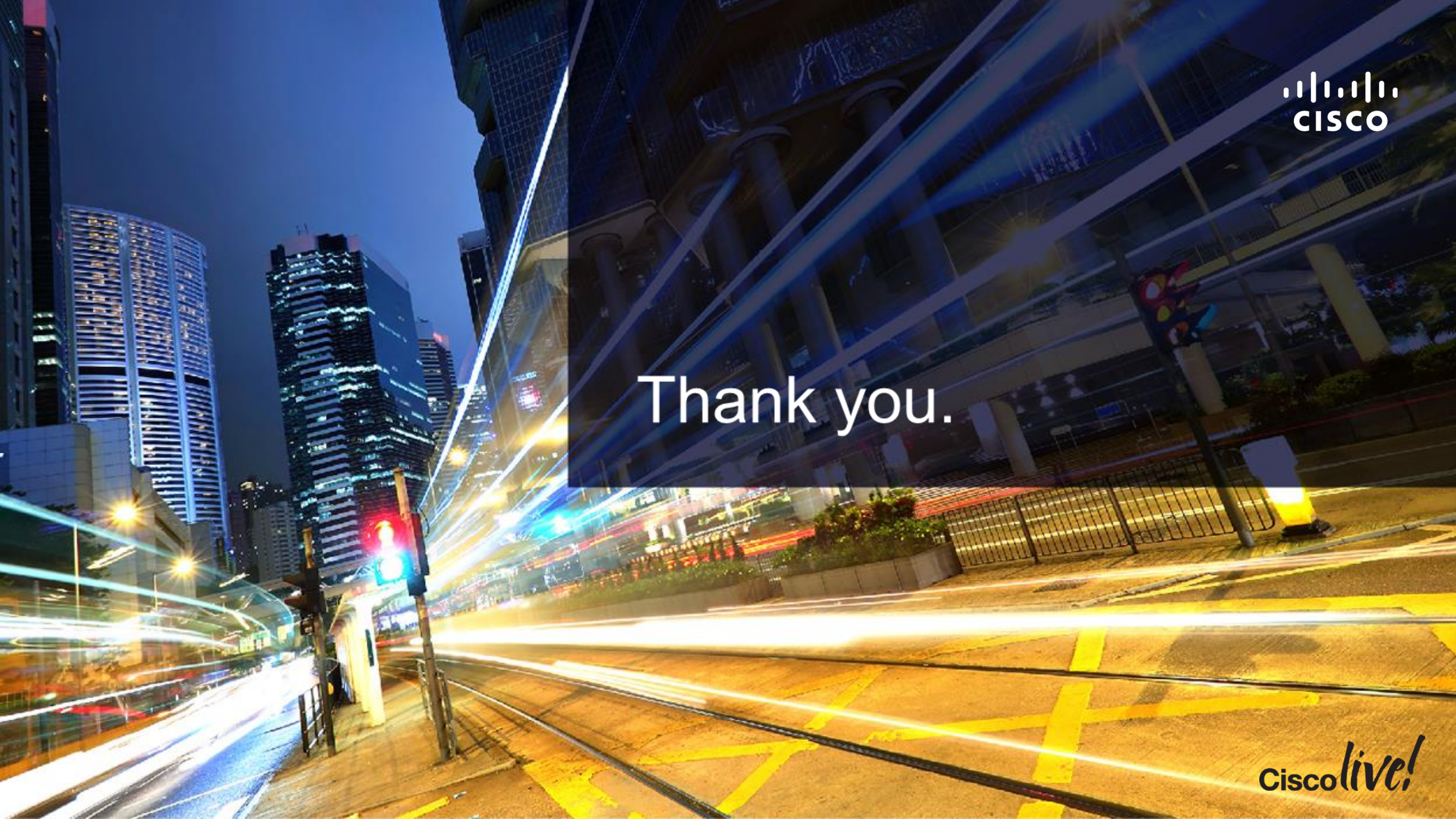
Give us your feedback and receive a Cisco Live 2015 T-Shirt!

Complete your Overall Event Survey and 5 Session Evaluations.

- Directly from your mobile device on the Cisco Live Mobile App
- By visiting the Cisco Live Mobile Site http://showcase.genie-connect.com/clmelbourne2015
- Visit any Cisco Live Internet Station located throughout the venue

T-Shirts can be collected in the World of Solutions on Friday 20 March 12:00pm - 2:00pm

Learn online with Cisco Live!
Visit us online after the conference for full access to session videos and presentations. www.CiscoLiveAPAC.com


Thank you.


CISCO