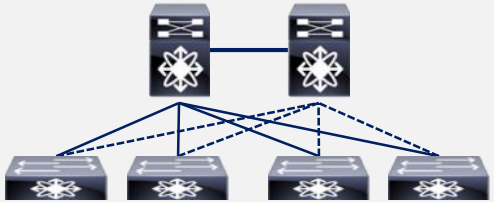
Data Center & Cloud Networking

Evolutionary Network Approach

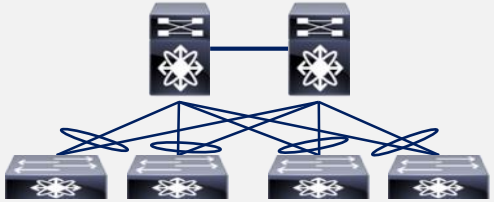
Workload Mobility
Increased App Communication
Higher Server Port Density and Bandwidth

STP based "Tiered" Design



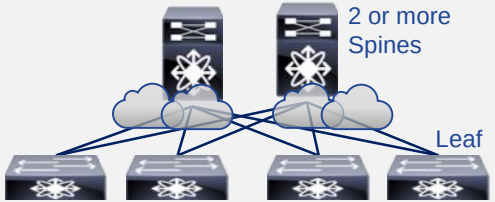
Classis STP Limitation
50% of all Links not utilized
Complex to Harden

VPC based "Tiered" Design



No STP Blocked Ports
Full Links Utilization
Faster Convergence
Macro for "best practice"

"Fabric" Design



No STP
Simple to Configure
Higher Fabric Bandwidth
Consistent Latency
FabricPath, VXLAN with MP-BGP-EVPN (Control Plane)

Spine
Scales to provide fabric bandwidth

Leaf
Scales to provide access port density

Focus Areas of Investment – Nexus Platforms

Nexus 9000 Cloud Scale

Cloud Scale ASICs

- Design Flexibility – ACI, VXLAN, Segment Routing
- Streaming Telemetry & Analytics
- Programmability

Nexus 9500 & 3600 R-Series

Broadcom Jericho

- Multicast – Media & Financial
- MPLS, VXLAN, Segment Routing
- Deep Buffers & Large Tables

Nexus 7000 Series

Cisco Custom ASICs

- Investment Protection
- Data Center Interconnect
- DC & Campus Core

Nexus 3000 Series

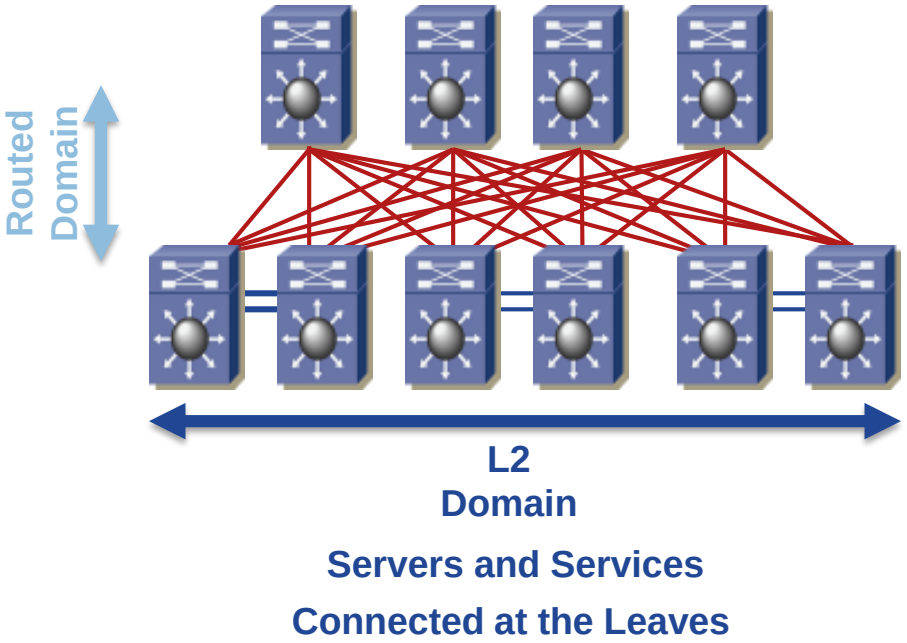
Merchant Silicon

- Customers looking for specific Merchant ASICs
- Ultra Low Latency
- Data Path Programmability

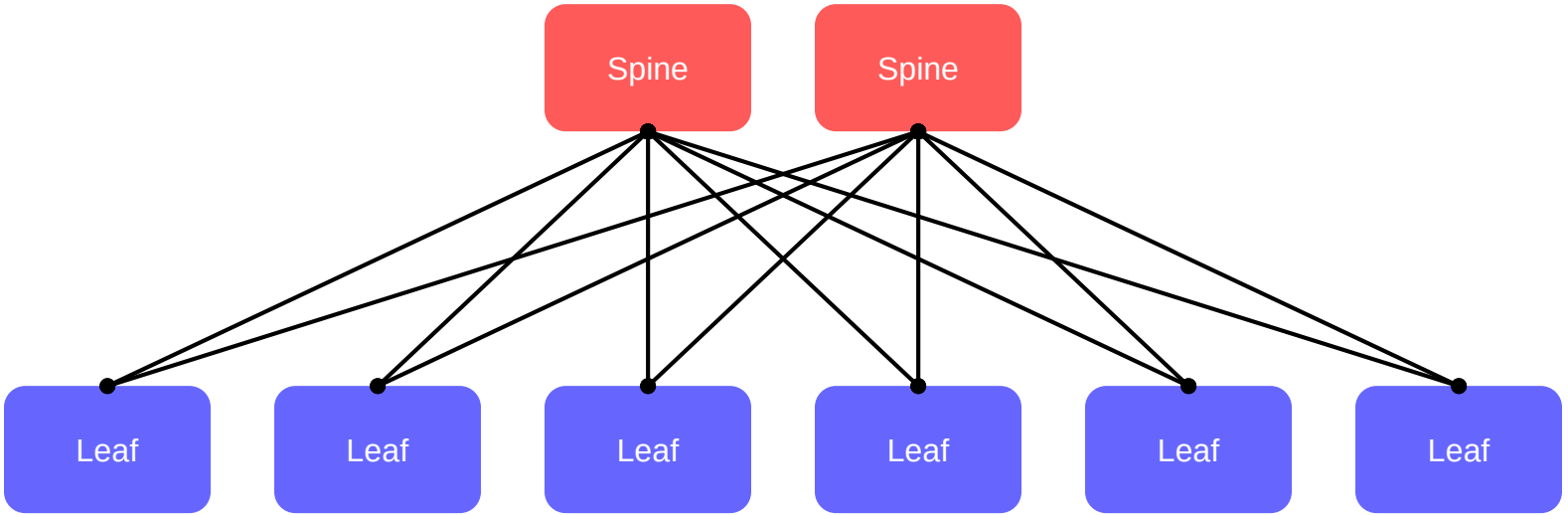
Changing Traffic Patterns in the Data Centre

Data Centre Design Evolution CLOS Fabric

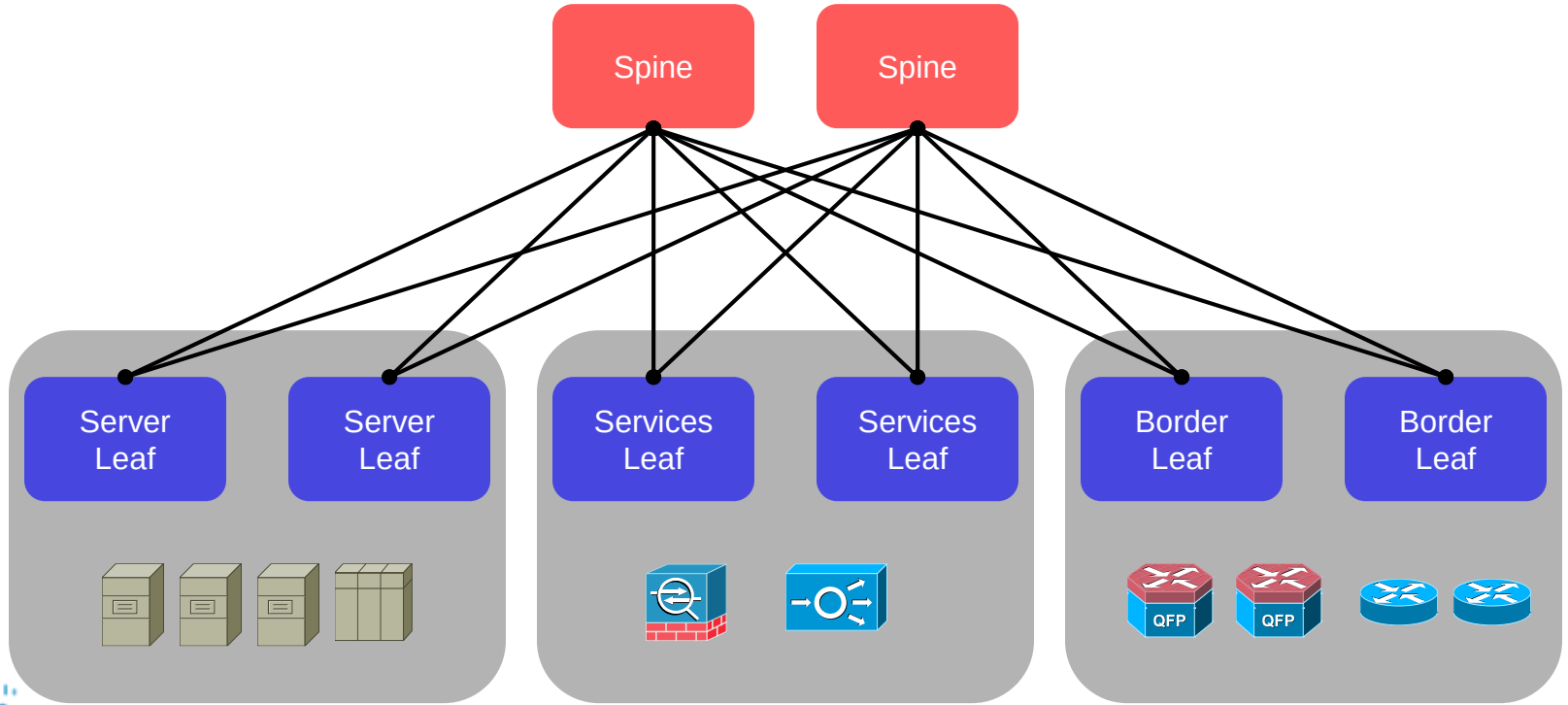
- Moving to Spine/Leaf construct
- No Longer Limited to two aggregation boxes
- Created Routed Paths between “access” and “core”
  - Routed based on MAC, IP, or VNI
- Layer 2 can be anywhere even with routing
- Automation/Orchestration, removing human error.



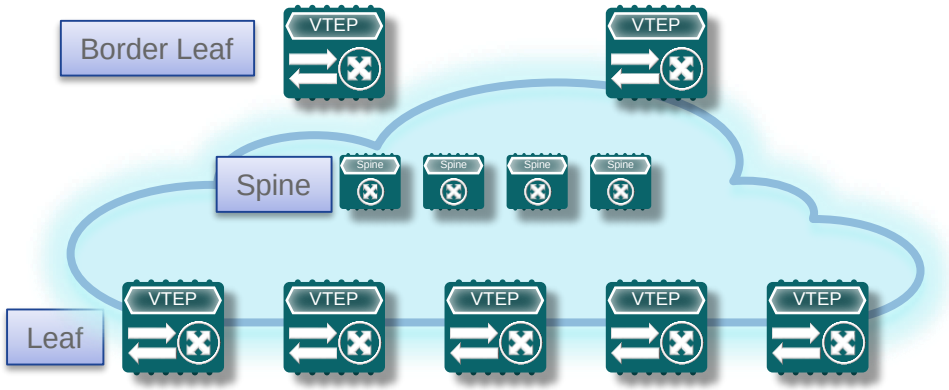
How do we design our physical fabric?



We can dedicate leaf nodes to a function.

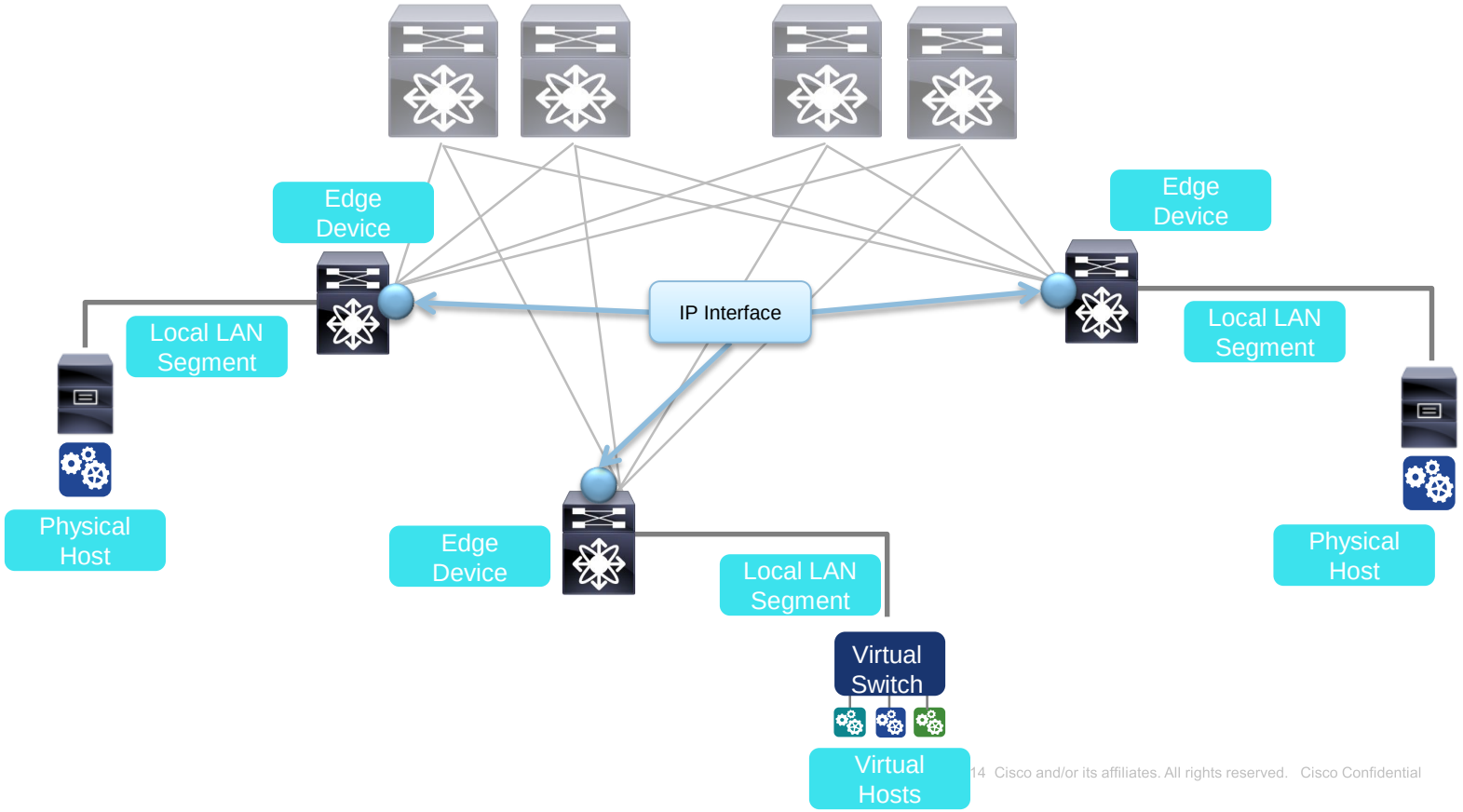


Border Leaf + Spine

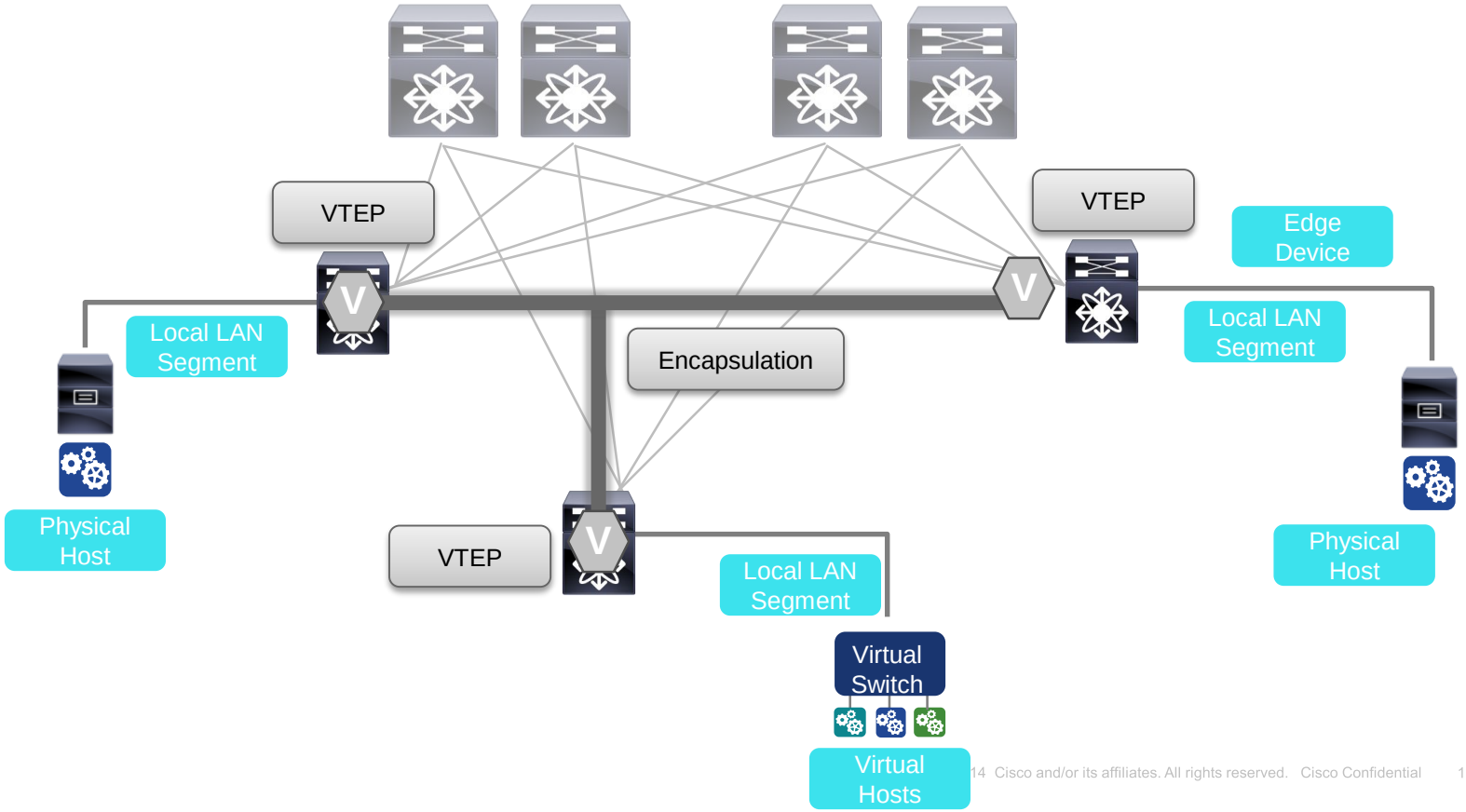


- **Leaf**
  - VXLAN Edge-Device
  - Route and Bridges Classic Ethernet frames and encapsulates them into VXLAN
  - Requires VTEP
- **Spine**
  - IP transport forwarder between Leaf (East/West)
  - Potentially hosting Rendezvous-Point (RP) for Underlay
  - Potentially hosting Route-Reflector (RR) for EVPN
  - Does not require VTEP
- **Border Leaf**
  - VXLAN Edge-Device
  - Route and Bridges Classic Ethernet frames from an outside network and encapsulates them into VXLAN (North/South)
  - Speaks IGP/EGP routing protocols with the outside network (North/South)
  - Requires VTEP

VXLAN Underlay



VXLAN Underlay



VXLAN Control/Data Plane Learning

Flood and Learn

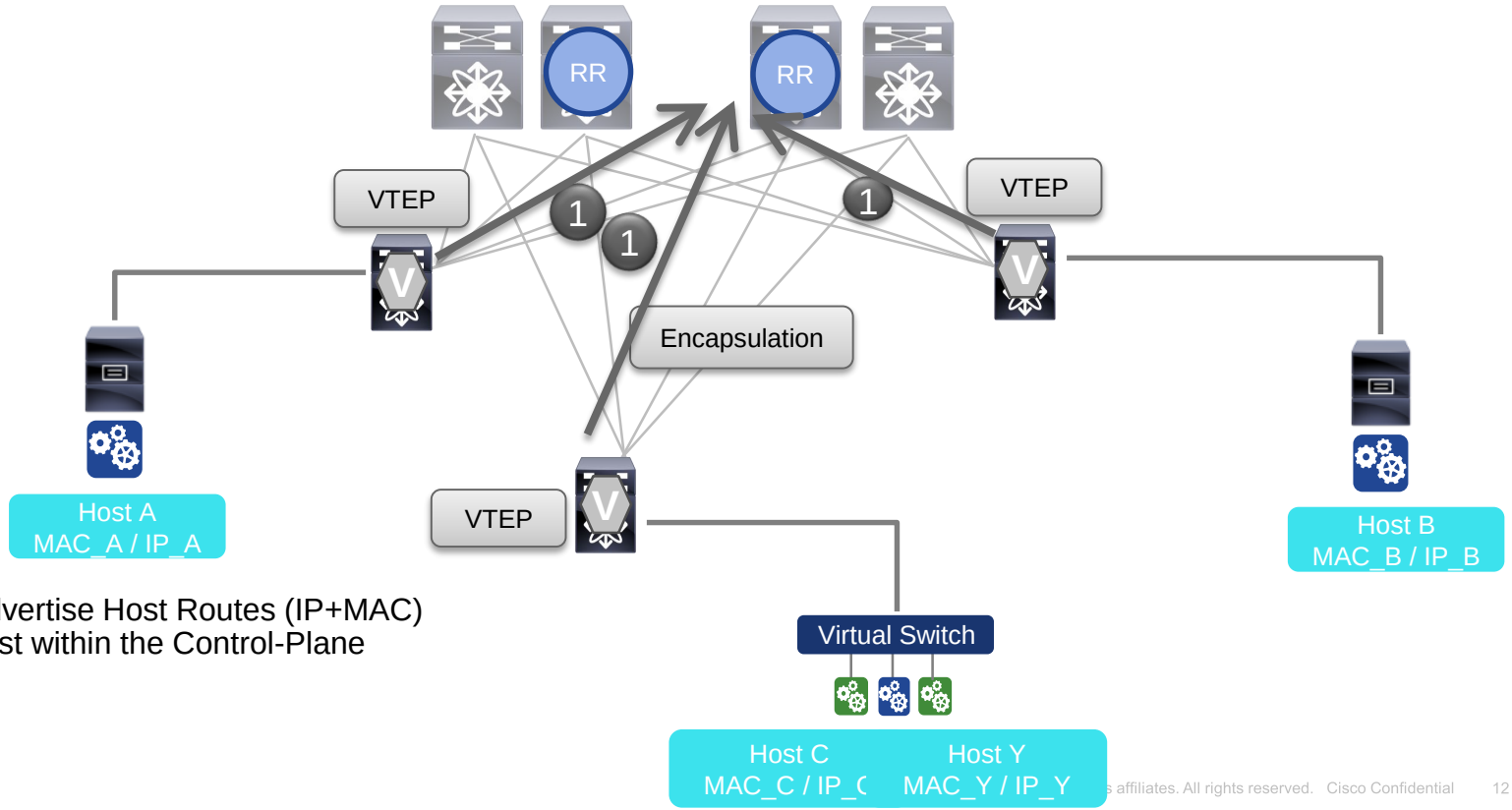
- No Control plane. Data plane learning is only option
- Data Plane Learning similar to Ethernet. Packets are flooded out all ports and over a Multicast address to find destination device.

BGP Based Control Plane

- Control plane uses standards-based BGP
- Layer 2 MAC and Layer 3 IP info distribution by BGP
- Forwarding decision based on control plane to minimise flooding
- IETF Draft L2VPN-EVPN evolved to RFC 7432

Protocol Learning & Distribution

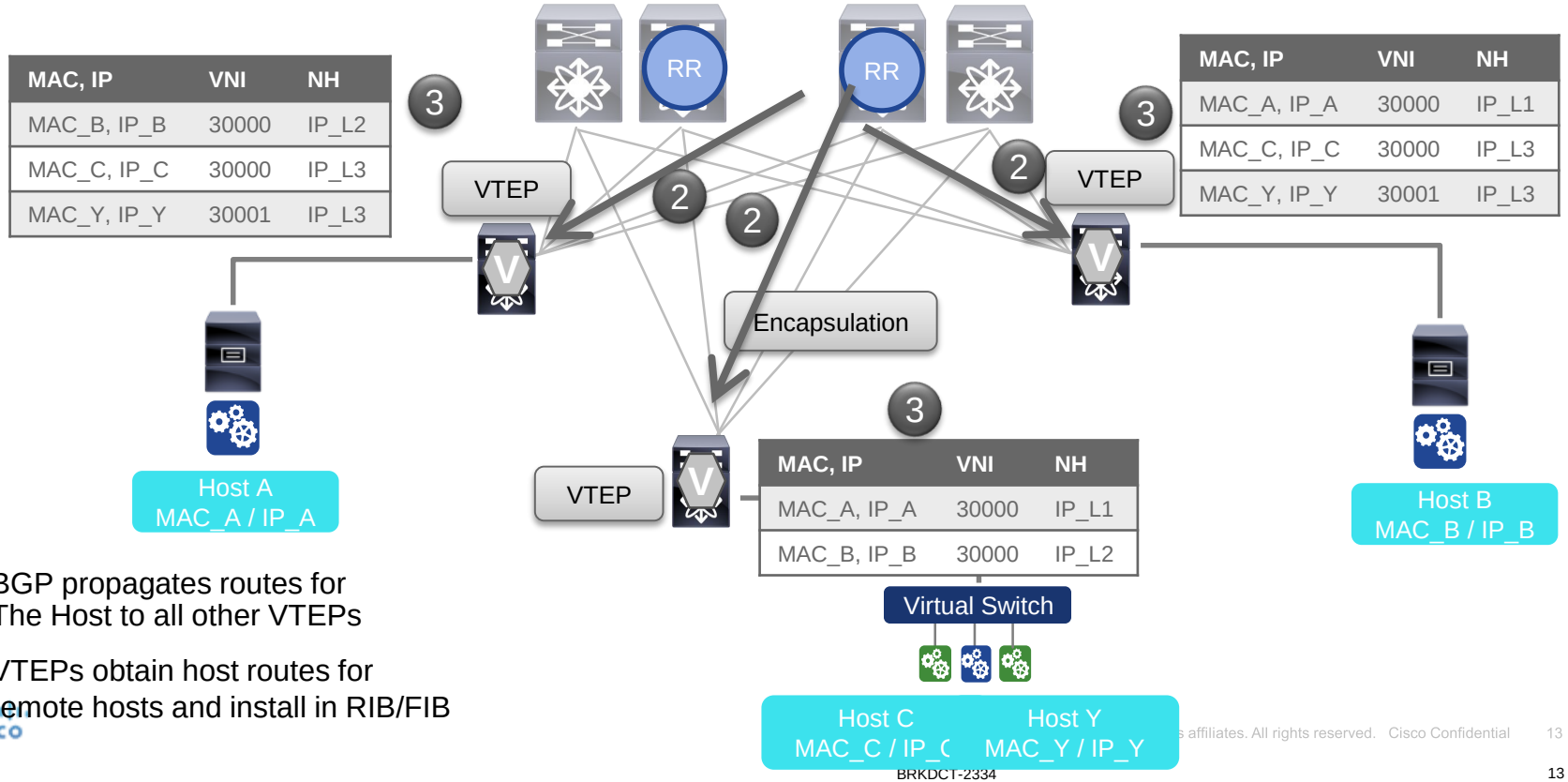
VXLAN/EVPN



1 VTEPs advertise Host Routes (IP+MAC) for the Host within the Control-Plane

Protocol Learning & Distribution

VXLAN/EVPN



2 BGP propagates routes for The Host to all other VTEPs

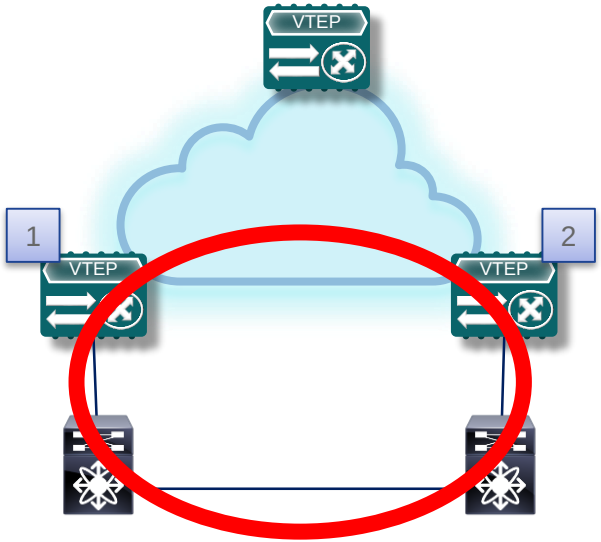
3 VTEPs obtain host routes for remote hosts and install in RIB/FIB

VNI Scalability Per Platform

5600, 7000, 9300, and 9500 Have Different Scalability Numbers

- Reference the VXLAN Verified Scalability Limits (Unidimensional) at a high level
- Focus on the Validated Deployment Case studies
- Can you support 750, 900, 1000, 1500, or 1600 VNIs?
- How Many TORs can communicate? Can I use Ingress replication or does my design require Multicast?
- Routes
  - Underlay Routes
  - Overlay Routes
  - Host Routes
  - MAC addresses

Southbound Loop Protection (today)



- EVPN detects excessive MAC moves
- Once detected, MAC is blackholed
- Loop persists but no active impact
  - Wave behavior
  - Until MAC is cleared (timer)

- Note:
  - Topology Loop persist!
  - No Loop detection
  - No Loop mitigation

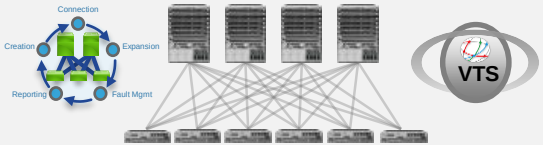
Fabric Management Options

Programmable Network

Modern NX-OS with enhanced NX-APIs

DevOps toolset used for Network Management
(Puppet, Chef, Ansible etc.)

Programmable Fabric

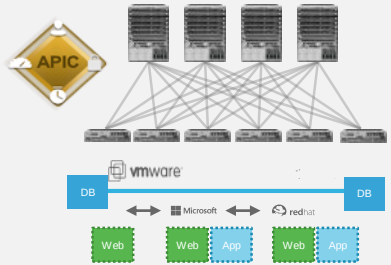


VxLAN-BGP EVPN standard-based

3rd party controller support

Cisco Controller for software overlay provisioning and management across N2K-N9K

Application Centric Infrastructure



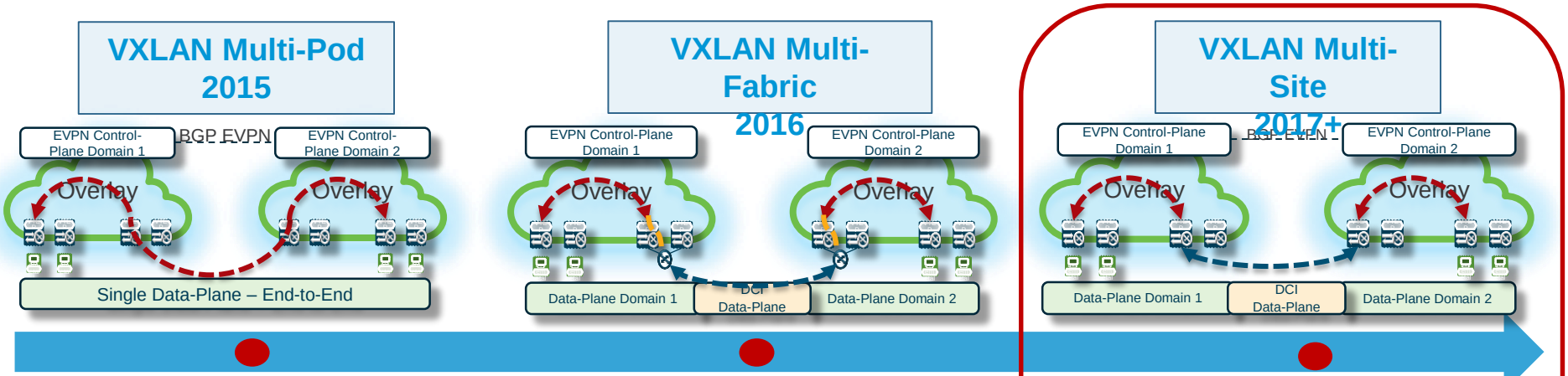
Turnkey integrated solution with security, centralised management, compliance and scale

Automated application centric-policy model with embedded security

Broad and deep ecosystem

Automation, API's, Controllers and Tool-chain's

VXLAN Multi-X Connectivity



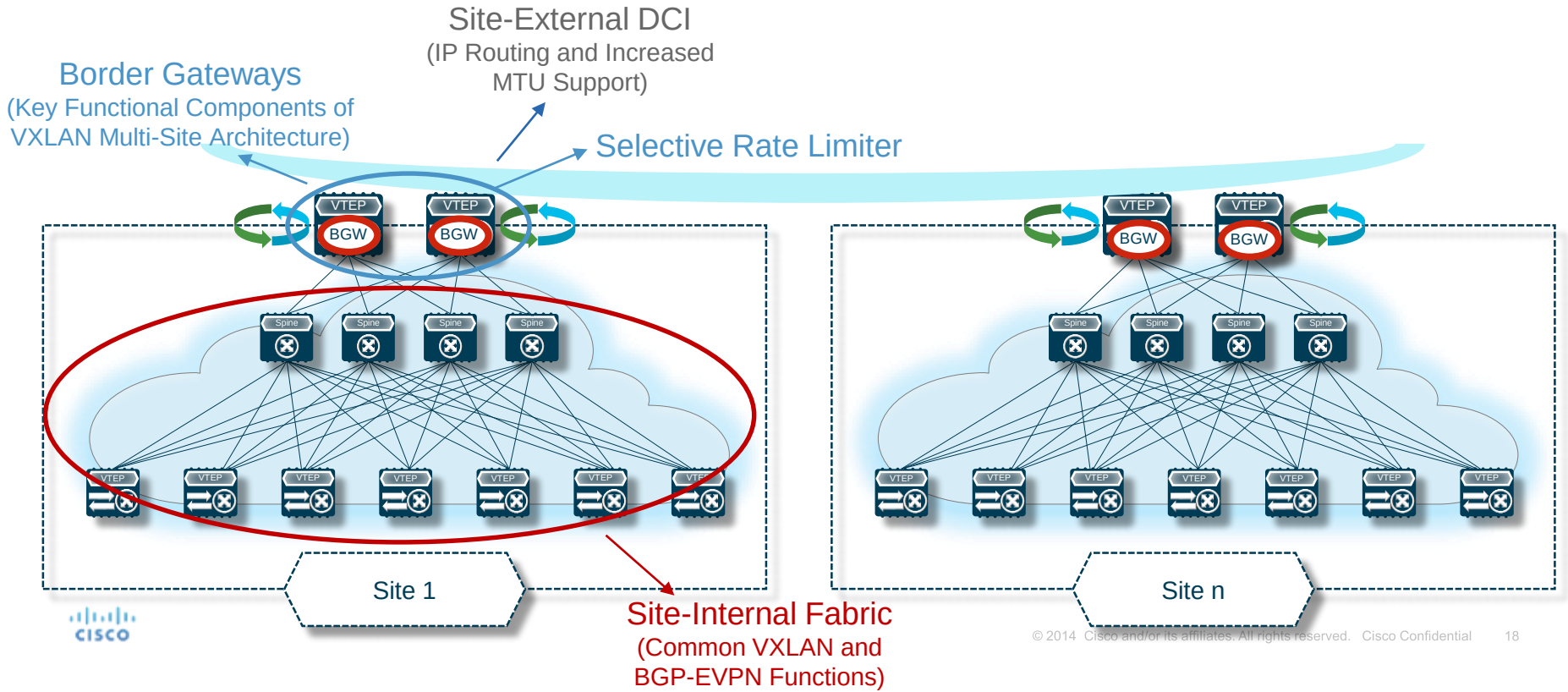
- Single Fabric with End-to-End Encapsulation
- Build Hierarchy in the Underlay – Flatten it in the Overlay

- Multiple Fabrics – Normalized through Ethernet
- Multiple Fabrics Interconnect using DCI (Layer 2 and Layer 3)

- Multiple Fabrics with Integrated DCI (DCI^2)
- Better Scale and Failure Domain Isolation between Fabrics

VXLAN Multi-Site - Functional Components

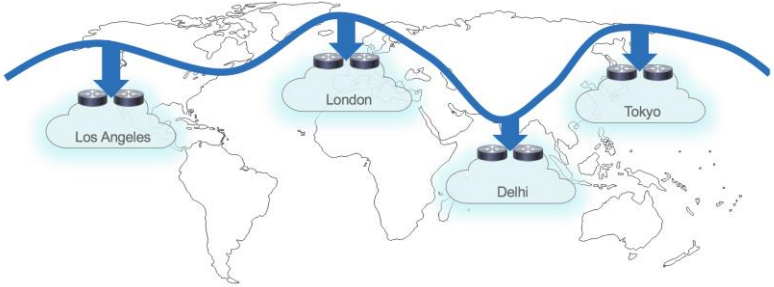
https://tools.ietf.org/html/draft-sharma-multi-site-evpn



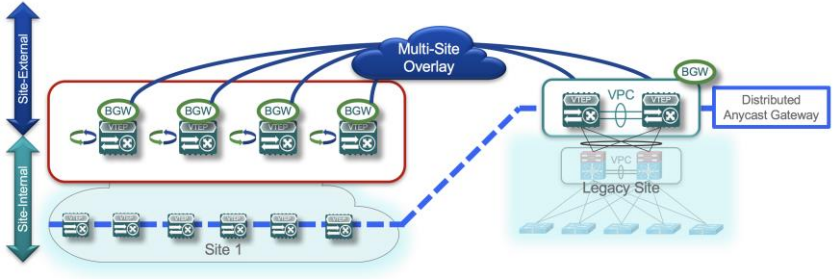
VXLAN Multi-Site

Main Use Cases

Scale-Up Model to Build a Large Intra-DC Network



Network Extension across Multiple Sites

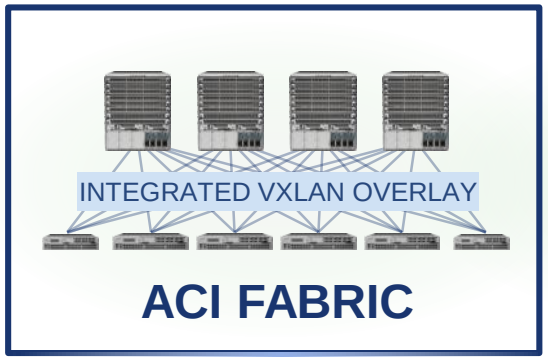


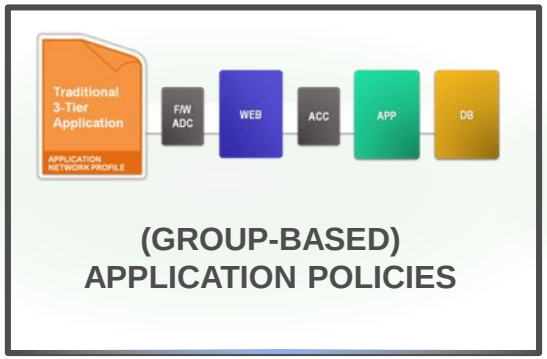
Integration with Legacy Networks (Coexistence and/or Migration)

19

Application Centric Infrastructure

Declarative Intent-based Automation
Logical Network Provisioning of Stateless Hardware
Rapid Deployment of Applications onto Networks with Scale, Security and Full Visibility



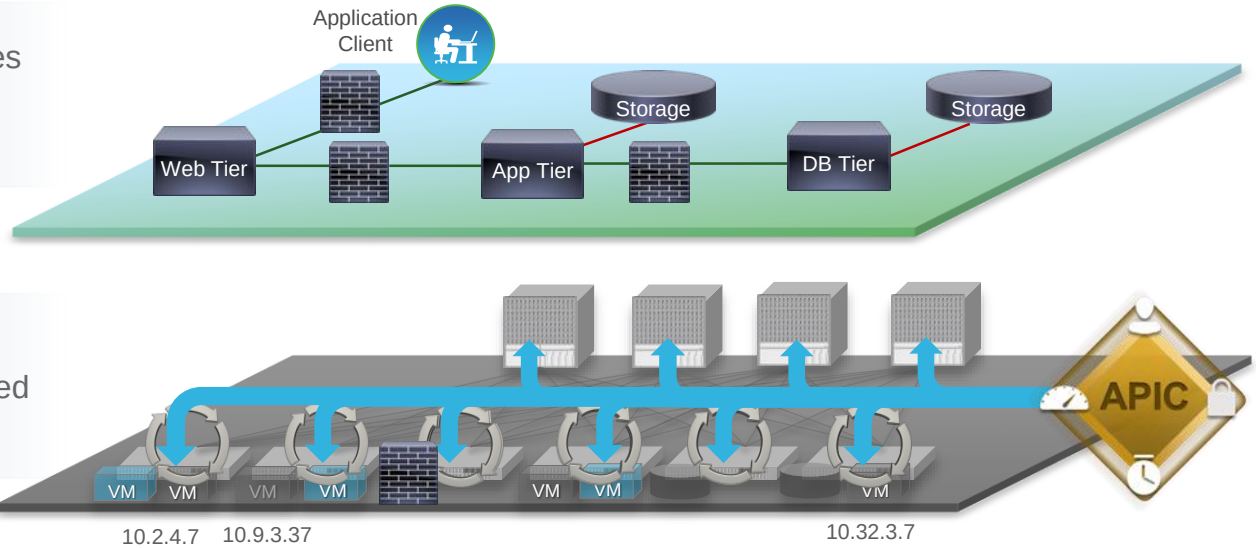


ACI

Application Policy Model and Instantiation

Application policy model: Defines the application requirements (application network profile)

Policy instantiation: Each device dynamically instantiates the required changes based on the policies



All forwarding in the fabric is managed through the application network profile

- IP addresses are fully portable **anywhere** within the fabric
- Security and forwarding are fully **decoupled** from any physical or virtual network attributes
- Devices autonomously update the state of the network based on configured policy requirements

Access Methodology

CLI (Command-line interface)

- Means of interacting with a computer program where user issues commands to the program in the form of successive lines of text

GUI (Graphical user interface)

- Interface that allows users to interact with devices through graphical icons and visuals

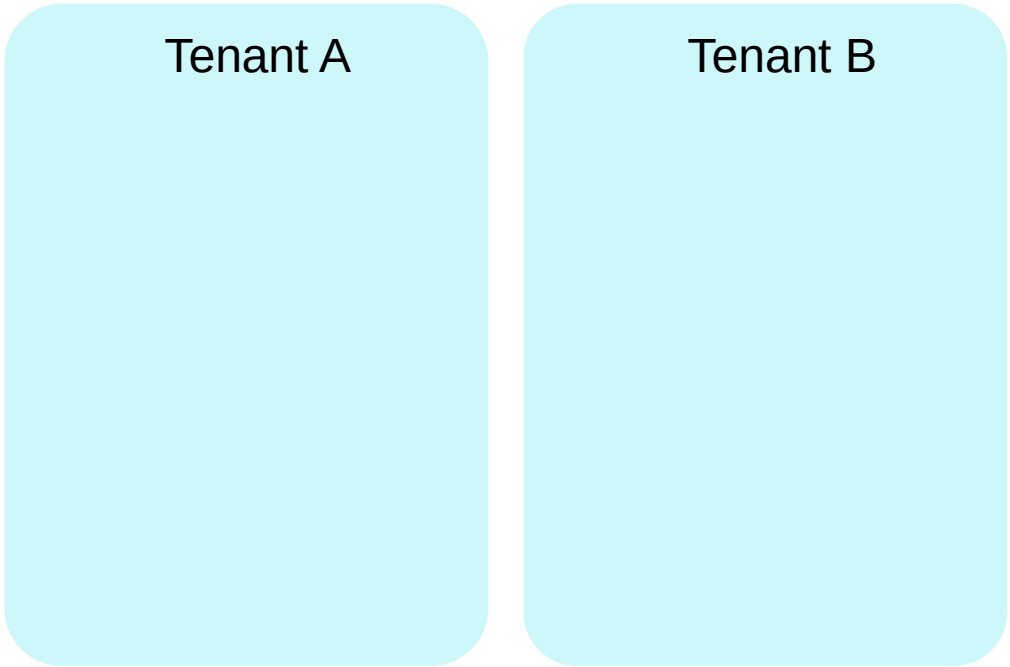
Programmable interface

- Software components / objects exposed to be called directly by other programs

- Open Source Tool
  - ACI Toolkit – Configuration Roll Back, Endpoint Tracker and other applications

Some new (or not so new) terms: Tenants, VRF (Private Network), Bridge Domains, Application Network Profiles, Endpoint Groups, Contracts/Filters

Application Policy Logical Construct

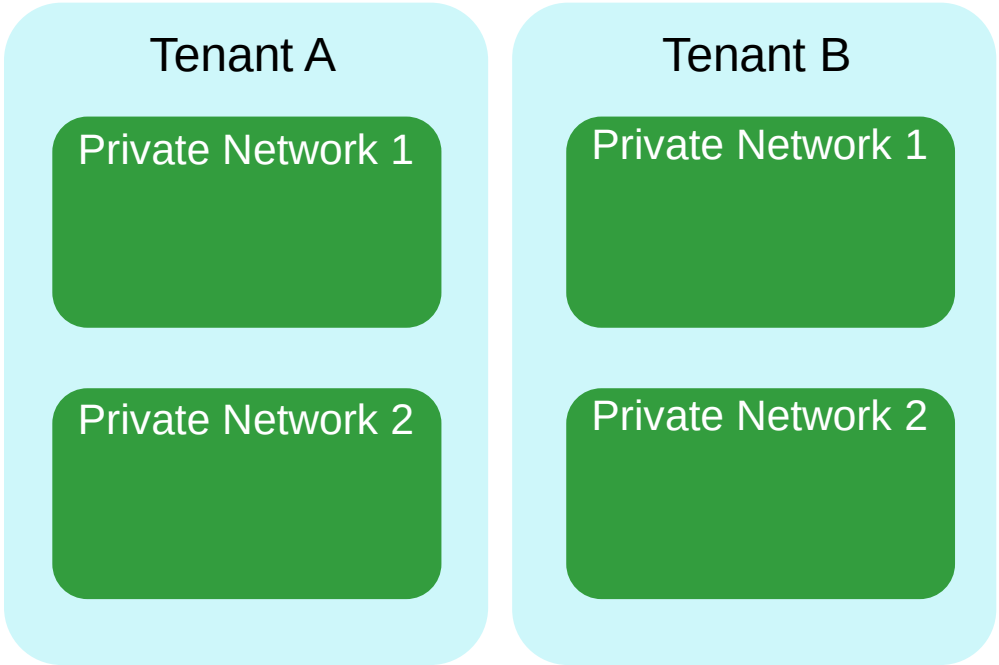


A **Tenant** is a container for all network, security, troubleshooting and L4 – 7 service policies.

Tenant resources are isolated from each other, allowing management by different administrators.

Application Policy Logical Construct

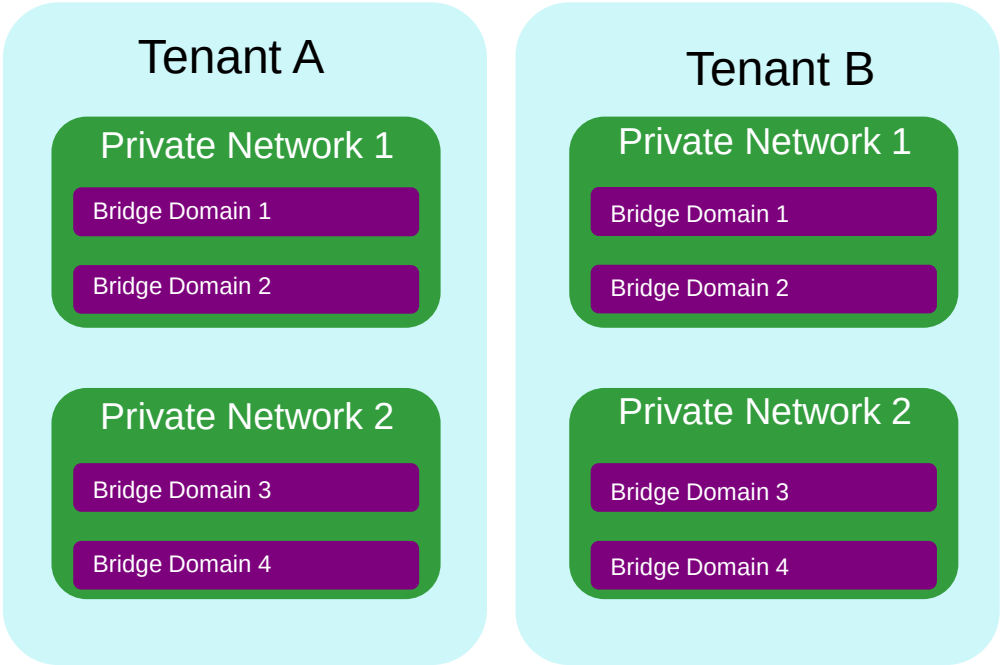
Private Networks (VRFs)



Private networks (also called VRFs or contexts) are defined within a tenant to allow isolated and potentially overlapping IP address space.

Application Policy Logical Construct

Bridge Domains

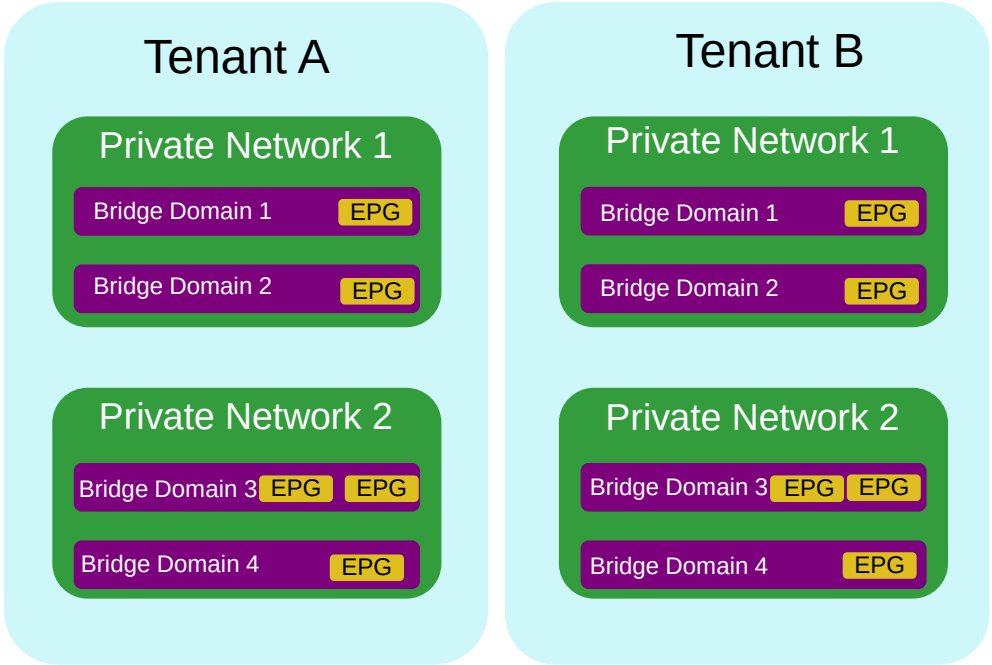


Within a private network (VRF), one or more bridge domains must be defined.

A bridge domain is a L2 forwarding entity within the fabric, used to define L2 forwarding domain and to constrain broadcast and multicast traffic.

Application Policy Logical Construct

End Point Groups (EPGs)



EPGs exist within a single bridge domain only – they do not span bridge domains.

EPGs defines the policy enforcement entities/classes. Class-based policies are applied between EPGs

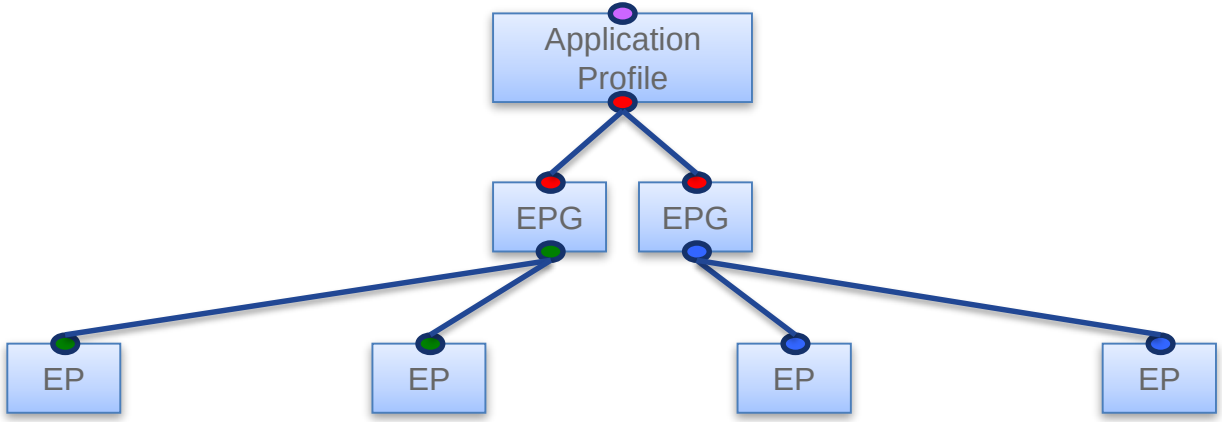
End Point (EP) Definition

EPs are devices which attach to the network either virtually or physically, e.g:

- Virtual Machine
- Physical Server (running Bare Metal or Hypervisor)
- External Layer 2 device
- External Layer 3 device
- VLAN
- Subnet
- Firewall
- Load balancer

End Point Group (EPG) Definition

An Endpoint Group (EPG) is a set of devices (end points) that share the same policy requirements.



Virtual Port, Physical Ports, External L2 VLAN, External L3 subnet

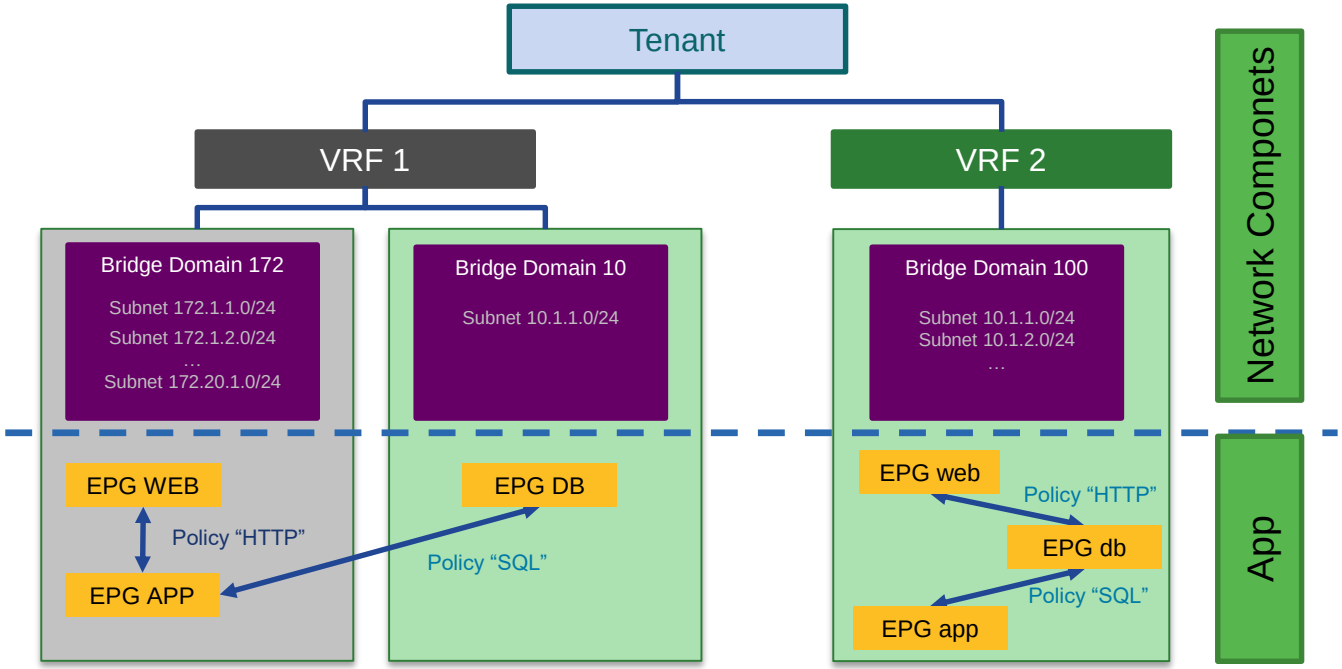
By default …

endpoints in different EPGs can’t communicate at all.

By default …

endpoints inside an EPG can communicate freely.

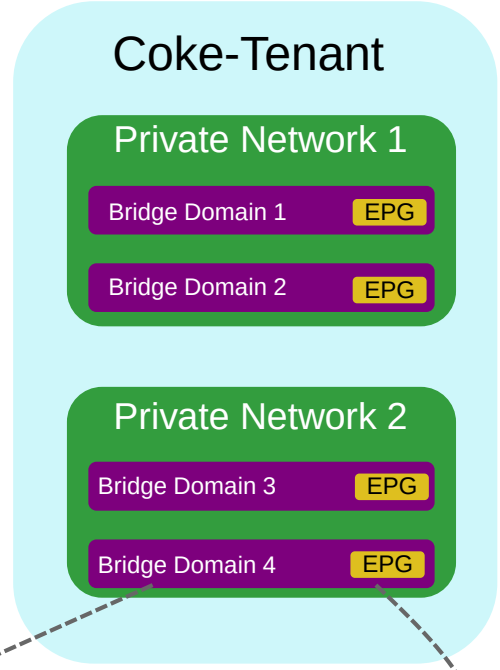
Application Policy Logical Construct



Application Policy Logical Construct

Mapping the Configuration to the Packet

- ACI Fabric leverages VXLAN Encapsulation to build network overlay
- VXLAN Source Group is used as a tag/label to identify the specific end point for each application function (EPG)
- Policy is enforced between an ingress or source application tier (EPG) and an egress or destination application tier (EPG)
- Policy can be enforced at source or destination

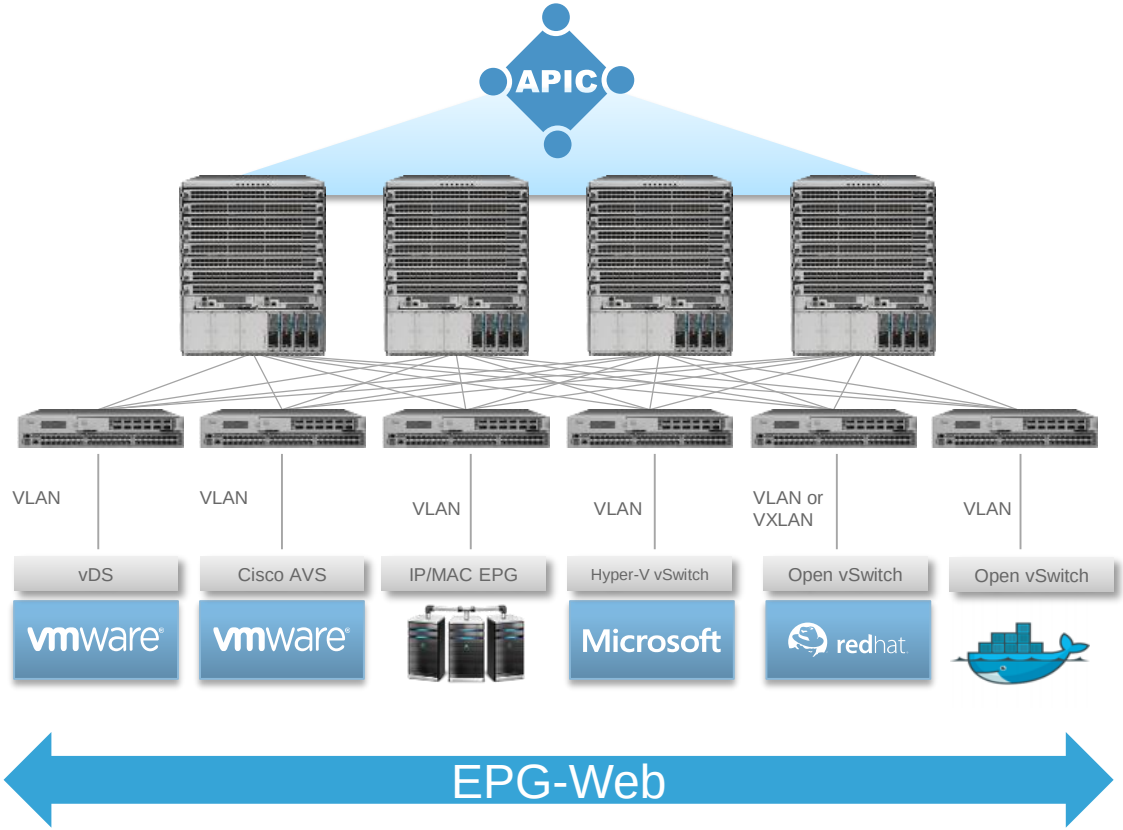


VXLAN Header:

Attribute Based Identity

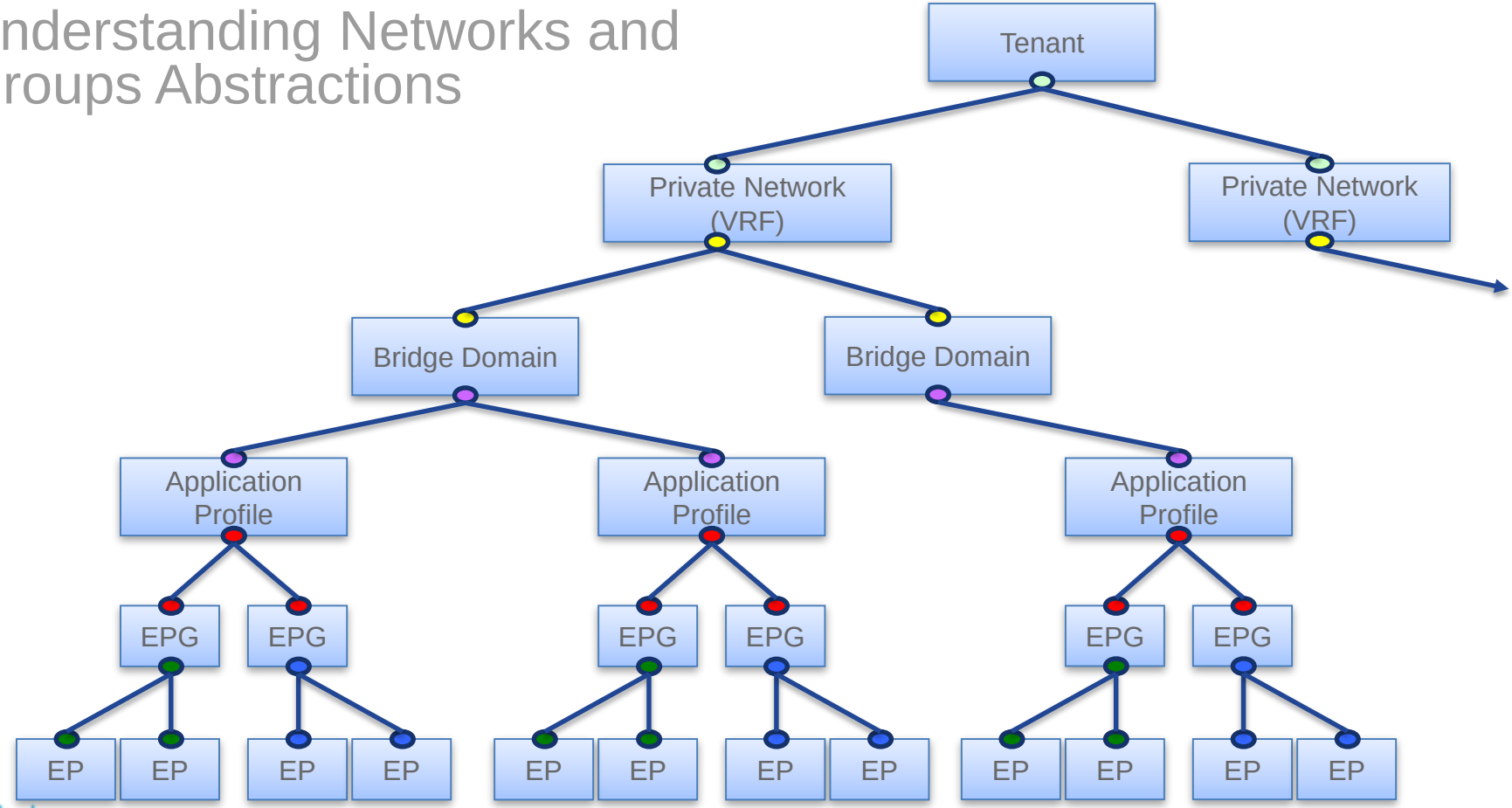
FYI

Attributes	Type
MAC Address Filter	Network
IP Address Filter	Network
VNic Dn (vNIC domain name)	VM
VM Identifier	VM
VM Name	VM
Hypervisor Identifier	VM
VMM Domain	VM
Datacenter	VM
Custom Attribute (VMWare AVS/vDS only)	VM
Operating System	VM

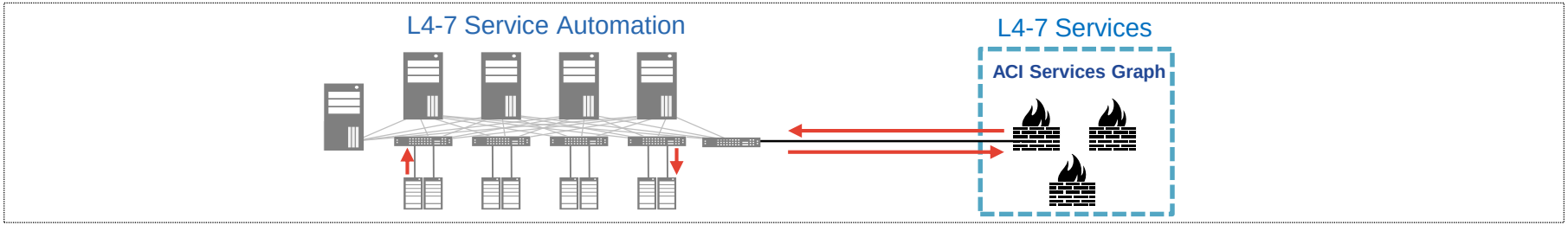


Micro-Segmentation Across any Workload

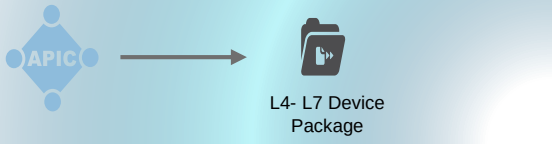
Understanding Networks and Groups Abstractions



L4-L7 Service Automation – Support for All Devices



Service Policy Mode



Full L4-L7 Centralized Service Automation
(With Device Package)

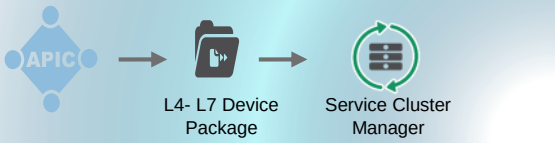
Large Ecosystem and Investment Protection

Network Policy Mode

Centralized Network Automation
(With NO Device Package)

Support for L4-L7 Cluster Managers

Service Manager Mode



Full L4-L7 Automation with Operational Flexibility
(With Device Package)

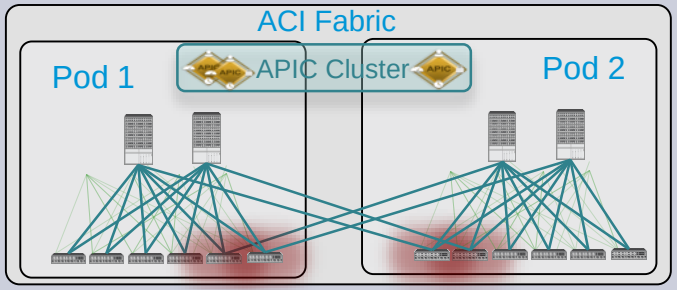
Large Ecosystem and Investment Protection

Interconnecting Multiple Sites ACI

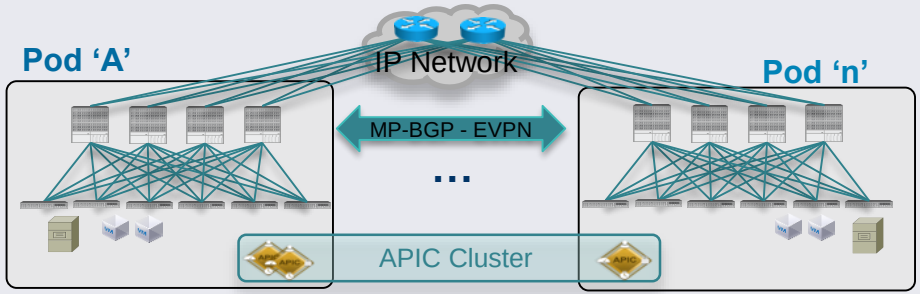
Interconnecting ACI Fabrics Design Options

Single APIC Cluster/Single Fabric

Stretched Fabric

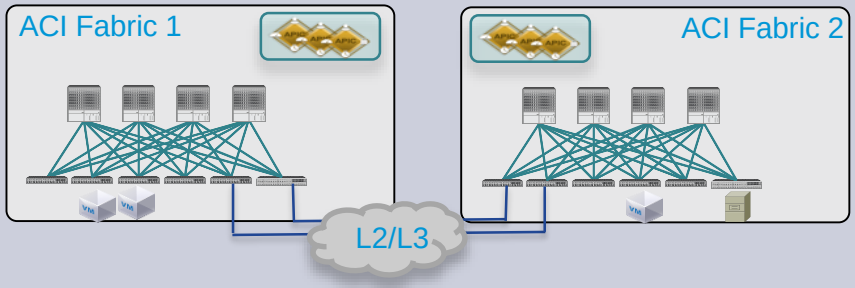


Multi-Pod (from 2.0 release)

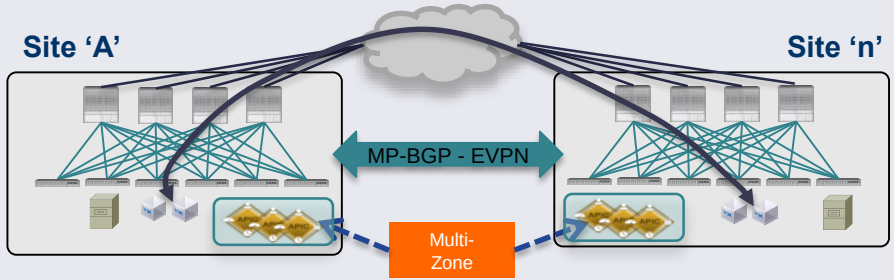


Multiple APIC Clusters/Multiple Fabrics

Multi-Fabric (with L2 and L3 DCI)



Multi-Site (Q3CY17)

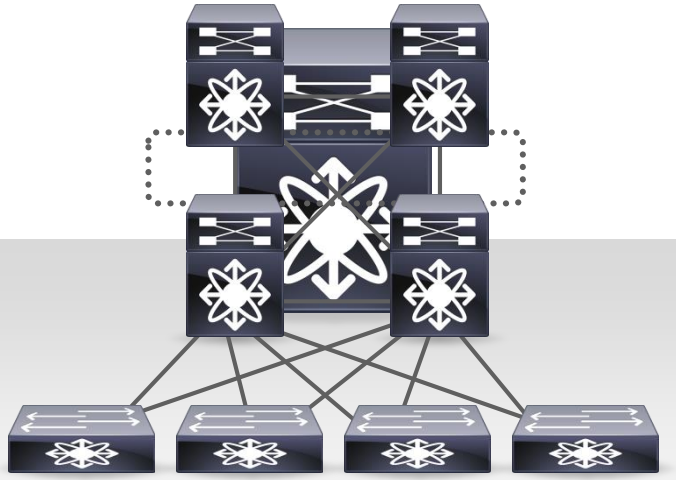


Why Leverage 25GB Ethernet?

- Server IO Doubling every 24 Months
- Core Networking Doubling every 18 Month
- Clients starting to use multiple interfaces per Server again
- Maximise Switch Throughput
- Minimise # of Cables and TOR switches
- SFP-25G Transceivers same form factor at SFP-10G
  - 1, 2, 3, 5 meter Twinax
  - SR Optics 100m OM4

Ethernet Alliance Introduction-to-25GbE-Webinar_D2p1.pdf

CapEx Optimization with VDC System Level Consolidation



Collapsed Architectures

Resource optimization over common Infra

Isolating domains: DMZ, Internal, Extranet

Business Challenges

Reduced physical device footprint, while meeting business needs

Reduced OpEx and driving new architectures

Resource optimization and On-Demand allocation

Compliance with Industry and Regulatory standards

VDC Benefits

Lower CapEx – Reduced number of physical switches

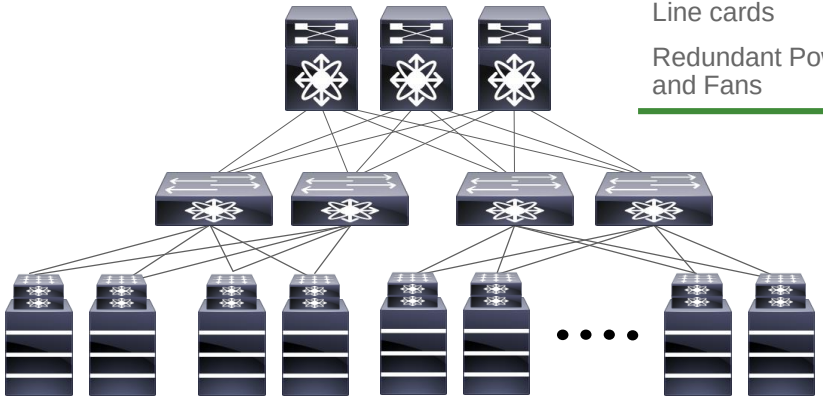
Lower OpEx – Reduced power and management requirements

Flexible separation/distribution of resources

Hardware and software fault isolation

Non-Stop Operations
Providing Ultra High Availability

- Redundant Supervisors
- Separate Data & Control Plane
- OIR PS, Fan modules
- Redundant Fabric
- OIR SUP, Fabric & Line cards
- Redundant Power Supply and Fans



Hardware | System | Network

Process Restartal

Port Channels

ast

(16 way

MP

Benefits

5-9s availability

Business Continuity

Seamless upgrades

Enabling new services

Software Upgrade

NX-OS HA

- Industry Leading Data Center HA Solution
- Mandatory for Mission Critical Data Centers
- Focus on Operational Excellence

ISSU

- Hitless – Non-Stop Forwarding
- Layer 2 and Layer 3
- Upgrade & Downgrade
- Only Platform in the Industry to Support Hitless ISSU for L2/L3

Direction:
- No support for ISSD
- More structured recommendations for software upgrades

Patching

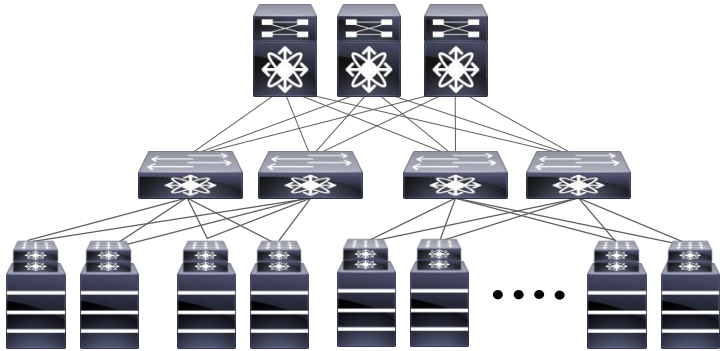
- Non-Disruptive Bug Fix for re-startable/ stateful processes
- Works with or without ISSU
- Chef and Puppet Agent Support
- Patch Management Tool

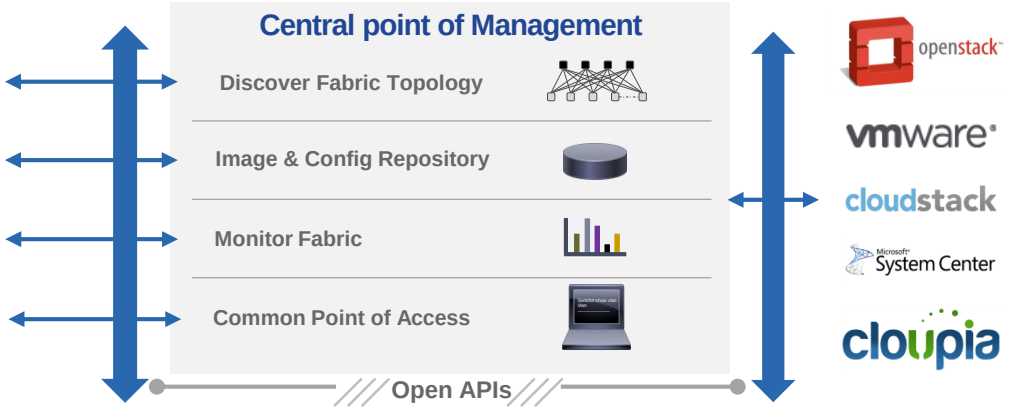
Direction:
- Limited number of Patches supported
- May be disruptive for certain processes

Maintenance Mode

- Graceful Insertion Removal
- Per VDC or entire switch
- Support per protocol used
  - vPC/FabricPath/vxlan
  - BGP/OSPF/..
  - OTV/LISP/MPLS
- Faster Reboot Improves Availability

Simplified Management
Consolidated, Automated, Aware





Fabric Topology Discovery

Detect Topology, mis-cabling

Image and Config Management

Power-On Auto Provisioning

Auto deploy nodes

Monitoring Fabric

Stats collection, VM location determination

Common Point of Access

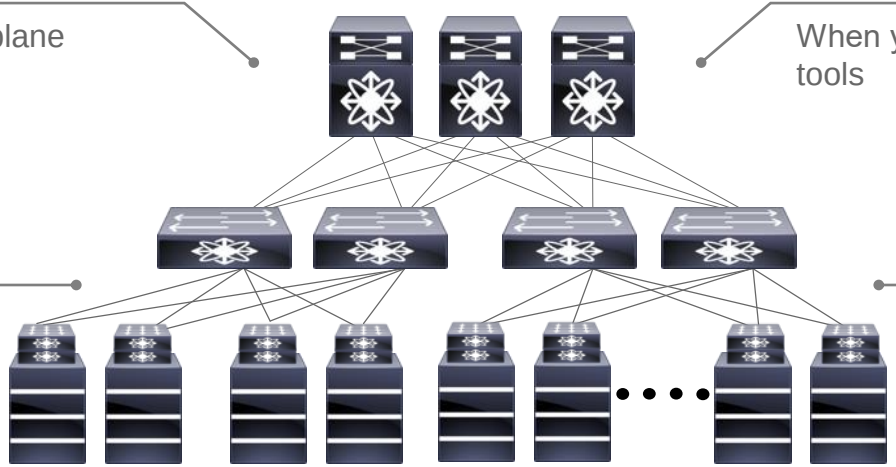
Access and run commands on multiple devices

Simplified Management for Ease of Operations

Unified Fabric Programmability
It's All About Options

OpenFlow
Standards based data plane programmability

Python
When you don't need power tools



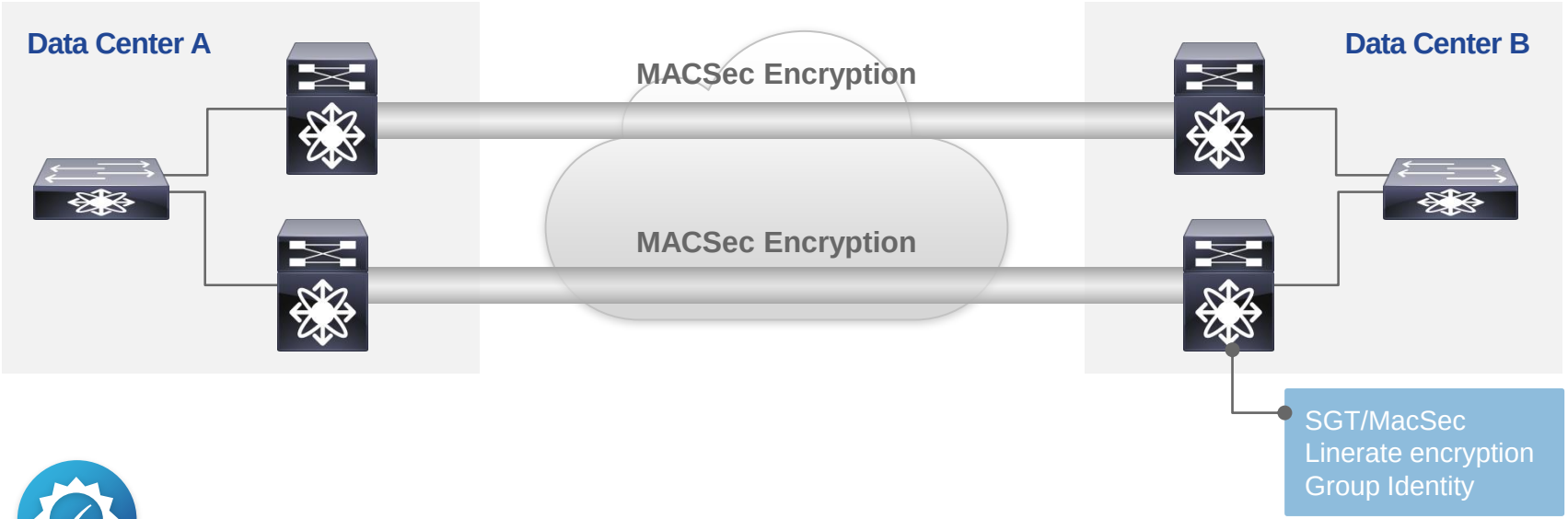
PoAP
Simplified provisioning & configuration control

JSON/REST
With low-level access provided by NxAPI, there are no limits

Puppet/Chef/ OpenStack
Orchestration and provisioning

SNMP & XML/NetConf
Traditional management tools CLI

Data Plane Security
Securing Connectivity at Scale



BENEFITS

- Prevent Intellectual Property Leaks
- Lower CapEx: Reduced number of physical devices for Encryption
- Flexible separation/ distribution of resources
- Compliance for Regulations

Security: Protecting IP and Infrastructure

Securing the Control Plane

- Control Plane Policing (CoPP)
- Control and Data Plane separation
- Authentication Protocol

Securing the Management Plane

- SSH
- SCP
- Role-based Access Control

Securing the Data Plane

- Line-rate MACSec
- Access Control Lists (ACLs)
- Security Group ACLs
- uRPF Check
- IP Source Guard, Port Security
- Dynamic ARP Inspection
- DHCP Snooping
- PVLAN

Extensive Security Portfolio

- Lower CapEx – Reduced number of physical devices for Encryption
- Flexible separation/distribution of resources
- Regulatory Compliance

Visibility and Monitoring

- Flexible Netflow
- NAM
- IEEE1588v2 Timestamping
- SPAN (ERSPAN, ACL SPAN, SPAN on drop, Exception SPAN)

Advanced Analytics on Nexus

Security Cannot Be Achieved Without Visibility

Microburst Monitoring

- Find out how many Microbursts were received

Buffer Monitoring

- See Buffer usage at real time

Latency Monitoring

- Find out precise port to port latency

Advanced SPAN

- **SPAN-on-Drop:** Correlate packet drop to applications
- **SPAN-on-Latency:** Span when latency exceeds a threashold
- **Exception SPAN:** Find out which malicious source was hogging the CPU
- **Selective SPAN:** SPAN selective traffic with Rule-Based SPAN/ACL-VLAN filters
- **ERSPAN with PTP timestamp:** Find out latency from point A to point B in your network

400G Drivers

Data Center traffic growth driving speed transitions in the access and aggregation layers

- ASIC readiness from Cisco and merchant vendors
- First generation ASICs optimized and targeted at Early Adopters (e.g. MSDC)
- First generation ASICs based on 12/16 nm technology
- Second generation ASICs likely based on 7 nm technology