TOMORROW starts here.

CISCO

Cisco live!

Introduction to Application Centric Infrastructure

BRKAPP-9000

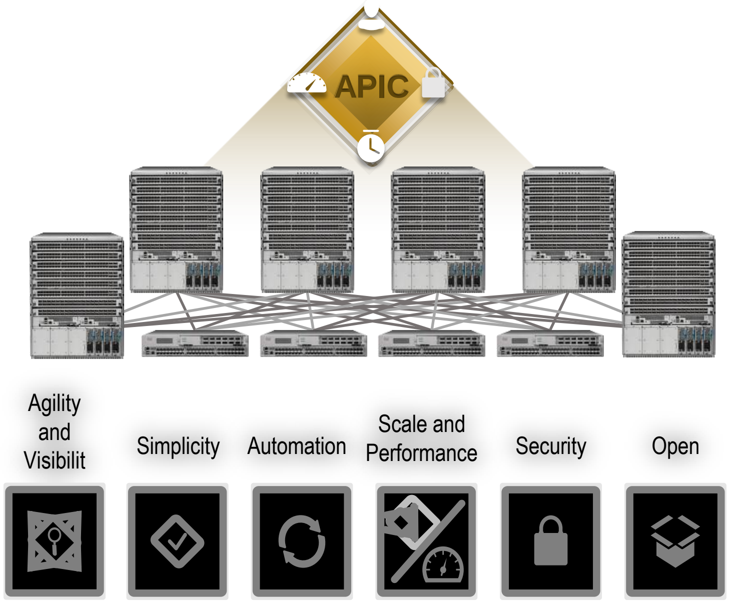
Mike Herbert
Principal Engineer

Cisco live!

What is our Goal Today?

Agenda – Application Centric Infrastructure

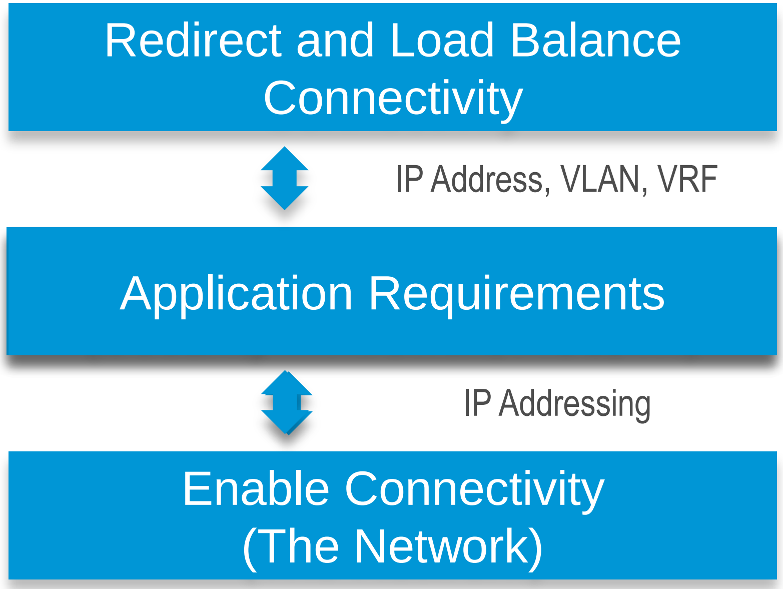
- What is ACI - Concepts and Principles
  - Why, What & How
- Foundations of ACI
  - ACI Fabric
  - Nexus 9000
  - ACI Policy Model
  - Hypervisor Integration, VMware, MSFT and KVM
  - Integration and Automation of L4-7 Services
  - APIC (The Controller)
- Integration, Migration and Co-Existence with Existing Infrastructure
- Open Standards, Open Source, Open API's



The on-going "IT pain"

- High cost, heterogeneous systems
- Redundant functionality
- Lack of agility to innovate
- Slow time to market
- Rising maintenance costs
- Rising regulatory and compliance costs, multiplied by:
  - Heterogeneous systems
  - Geographic expansion / local laws
- Falling IT Budgets

Overloaded Network Constructs



Redirect and Load Balance Connectivity

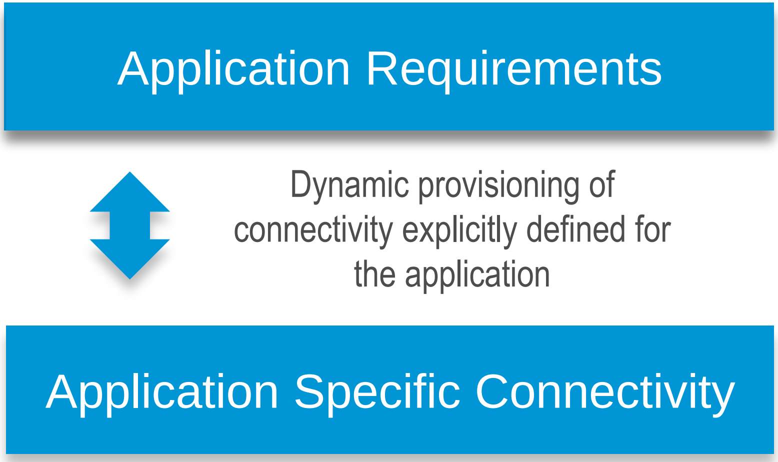
IP Address, VLAN, VRF

Application Requirements

IP Addressing

Enable Connectivity (The Network)

ACI directly maps the application connectivity requirements onto the network and services fabric



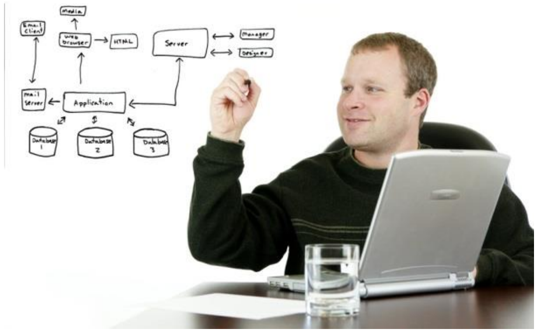
Application Requirements

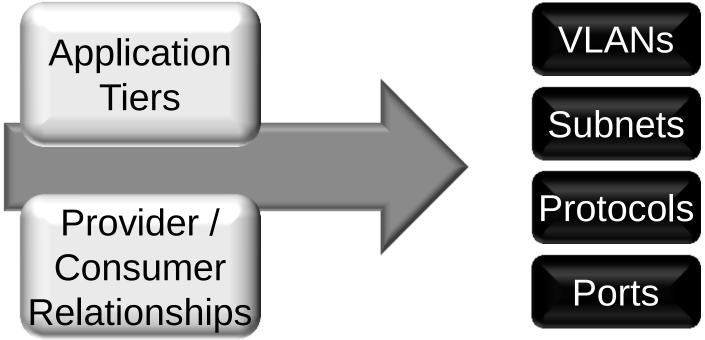
Dynamic provisioning of connectivity explicitly defined for the application

Application Specific Connectivity

Application Language Barriers

Developers





Application Tiers

Provider / Consumer Relationships

VLANs

Subnets

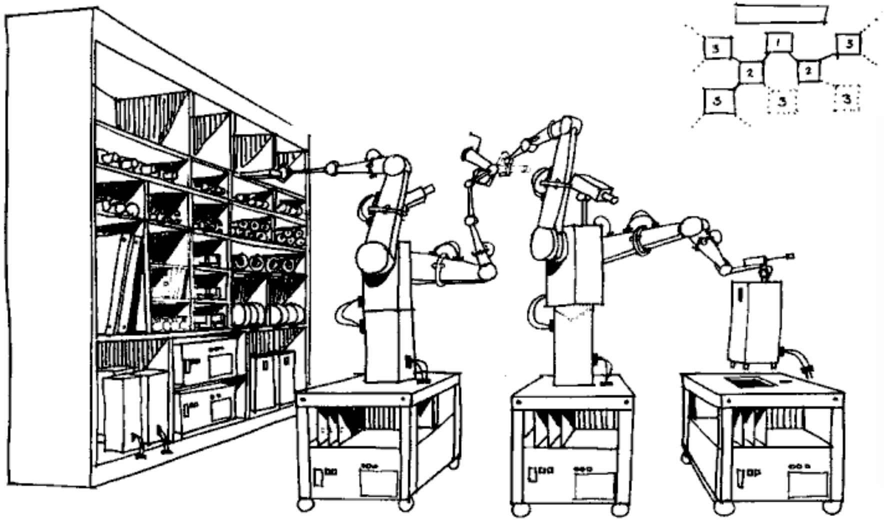
Protocols

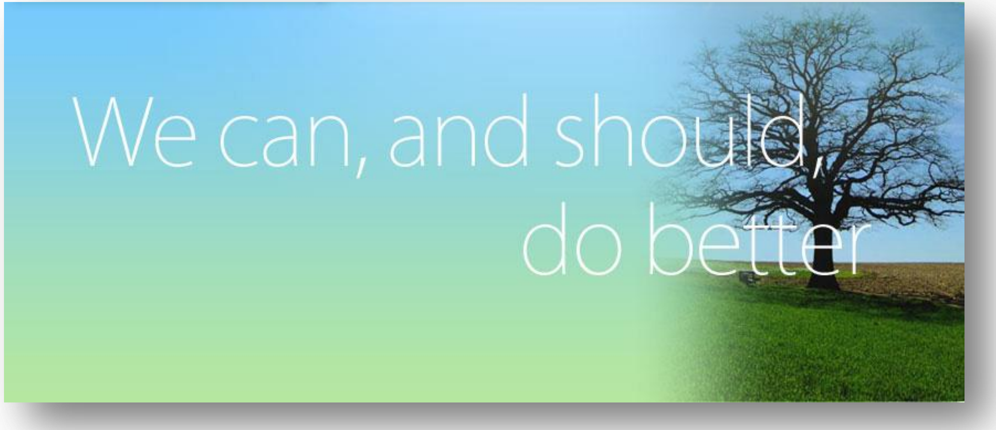
Ports

Infrastructure Teams

Developer and infrastructure teams must translate between disparate languages.

A Need for Infrastructure Automation, but…





Abstraction, the Real Objective of "SDN"

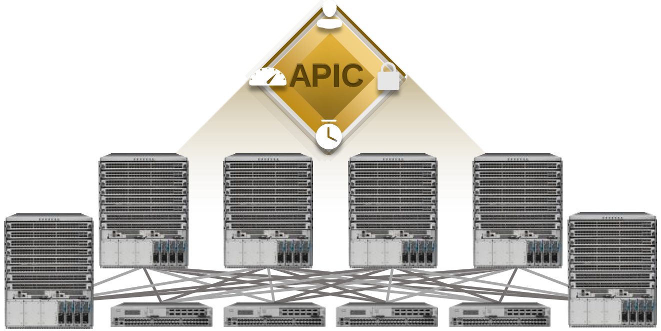
How to Avoid Death by Micromanagement

Now that you have logged on, click on the Start button and navigate to your email account. Take five minutes to read, then make a cup of tea...
© Jason Frazer
jason@twistedmusings.com.au
To increase productivity, Rob's boss decided to micro-manage his employees.

Networks are traditionally controlled in similar micro-managed, high touch, interactive manner

First Generation SDN is no different

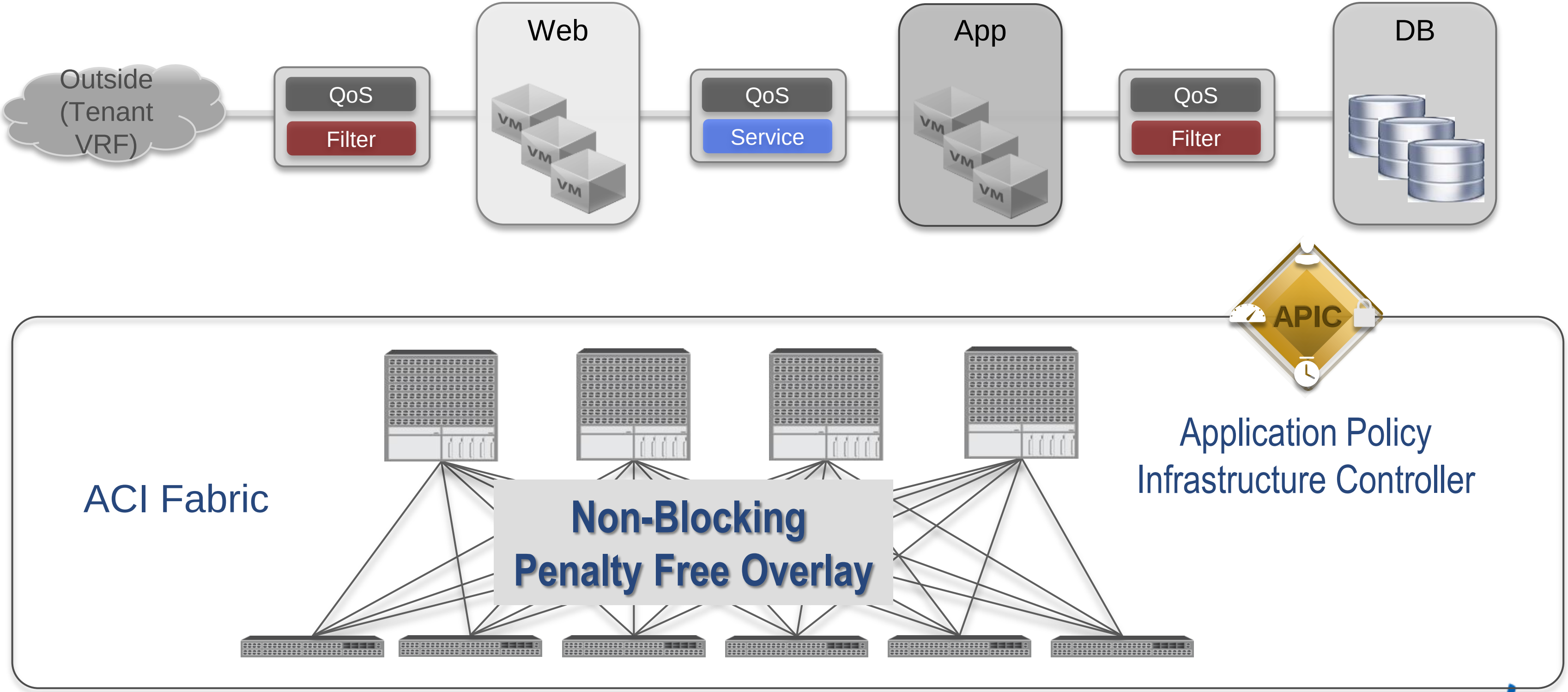
ACI Design Philosophy



- System Architecture
  - Expand Networking From Boxes To Systems
- Open Source & Multi-vendor
  - Innovations Published to Open Source
- Physical & Virtual
  - Traditional, Virtualised, & Next-Generation Non Virtualised Applications
- Velocity
  - Abstraction, Abstraction, Abstraction
- Costs
  - Best of Merchant & Custom Silicon for CAPEX Unmatched Automation for OPEX

ACI Fabric

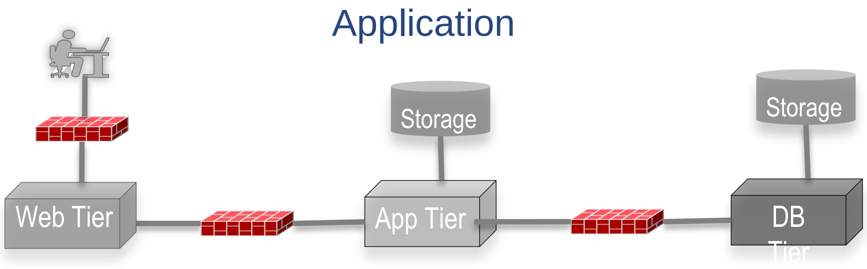
Logical network provisioning of stateless hardware



Application Network Profile

Policy Based Fabric Management

- Extend the principle of UCSM service profiles to the entire fabric
- Network Profile: Stateless Definition of Application Requirements
  - Application Tiers
  - Connectivity policies
  - L4 – L7 Services
  - XML/JSON Schema
- Fully Abstracted from the infrastructure implementation
  - Removes dependencies of the infrastructure
  - Portable across different Data centre fabrics



Network Profile fully describes the application connectivity requirements

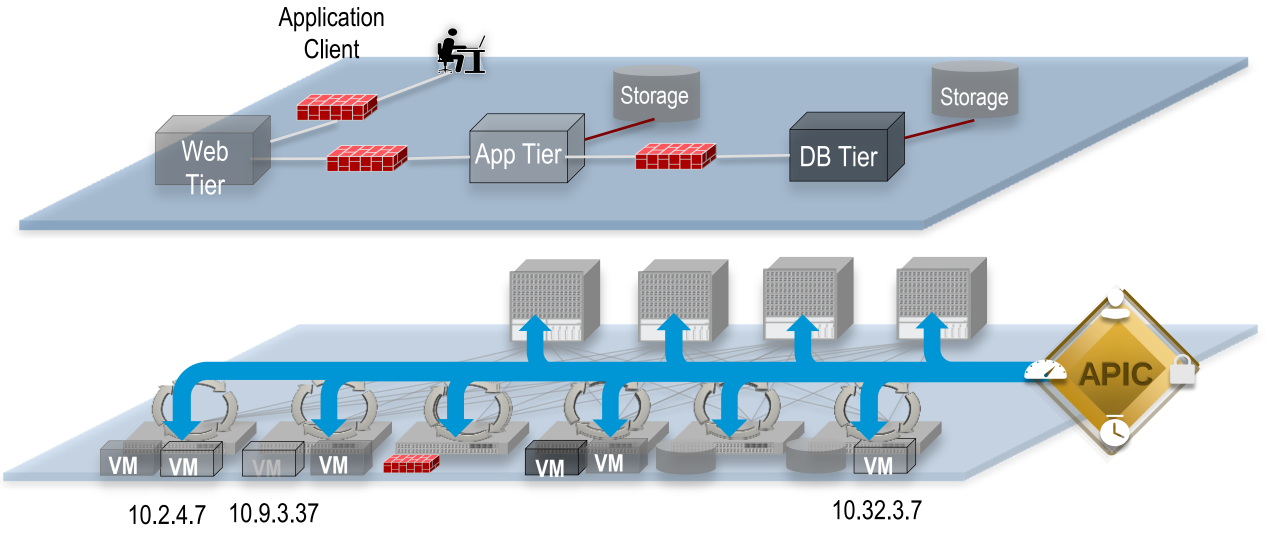
```
## Network Profile: Defines Application Level Metadata (Pseudo Code Example)

<Network-Profile = Production_Web>
 <App-Tier = Web>
   <Connected-To = Application_Client>
       <Connection-Policy = Secure_Firewall_External>
   <Connected-To = Application_Tier>
       <Connection-Policy = Secure_Firewall_Internal & High_Priority>
. . .
 <App-Tier = DataBase>
   <Connected-To = Storage>
       <Connection-Policy = NFS_TCP & High_BW_Low_Latency>
. . .
```

Application Policy Model & 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 via the Application Network Profile
  - IP addresses are fully portable *anywhere* within the fabric
  - Security & Forwarding are fully *decoupled* from any physical or virtual network attributes
  - Devices autonomously update the state of the network based on configured policy requirements

Application Awareness

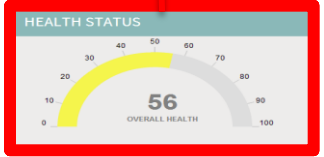
Application Level Visibility

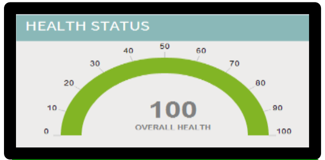
- ACI Fabric provides next generation of analytic capabilities
- Per Application, Tenants, & Infrastructure:
  - Health Scores
  - Latency
  - Atomic Counters
  - Resource Consumption
- Integrate with Workload Placement or Migration

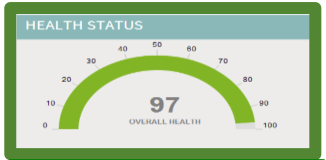
PetStore Event

Triggered Events or Queries

Actions:
No New Hosts or VM’s
Evacuate Hypervisors
Re-Balance Clusters



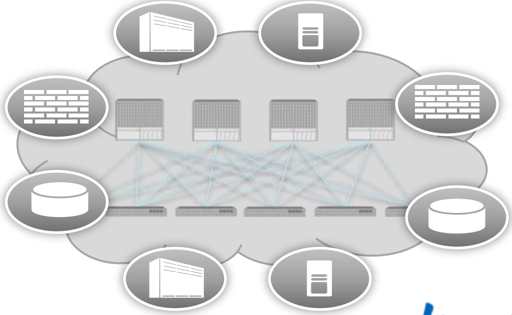




PetStore Dev
Leaf 1 & 2
Spine 1 – 3
Atomic Counters

PetStore Prod
Leaf 2 & 3
Spine 1 – 2
Atomic Counters

PetStore QA
Leaf 3 & 4
Spine 2 – 3
Atomic Counters

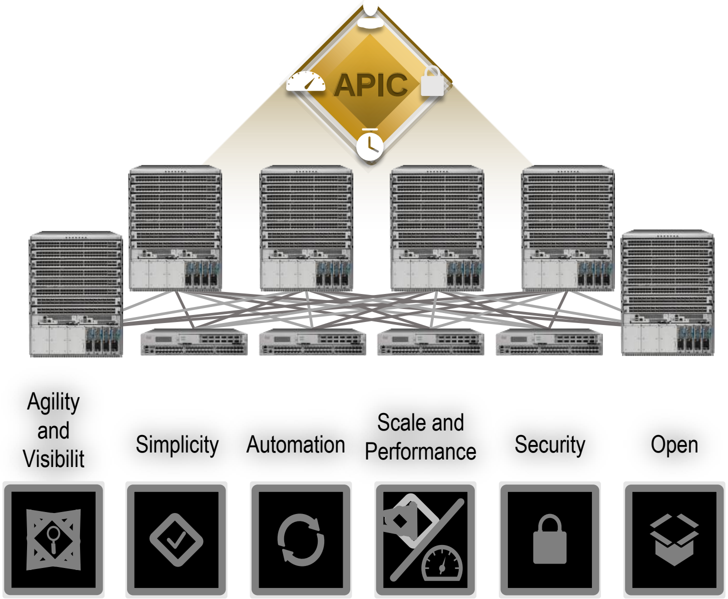


VXLAN
Per hop Visibility

Physical And
Virtual As One

Agenda – Application Centric Infrastructure

- What is ACI - Concepts and Principles
  - Why, What & How
- Foundations of ACI
  - ACI Fabric
  - Nexus 9000
  - ACI Policy Model
  - Hypervisor Integration, VMWare, MSFT and KVM
  - Integration and Automation of L4-7 Services
  - APIC (The Controller)
- Integration, Migration and Co-Existence with Existing Infrastructure
- Open Standards, Open Source, Open API's

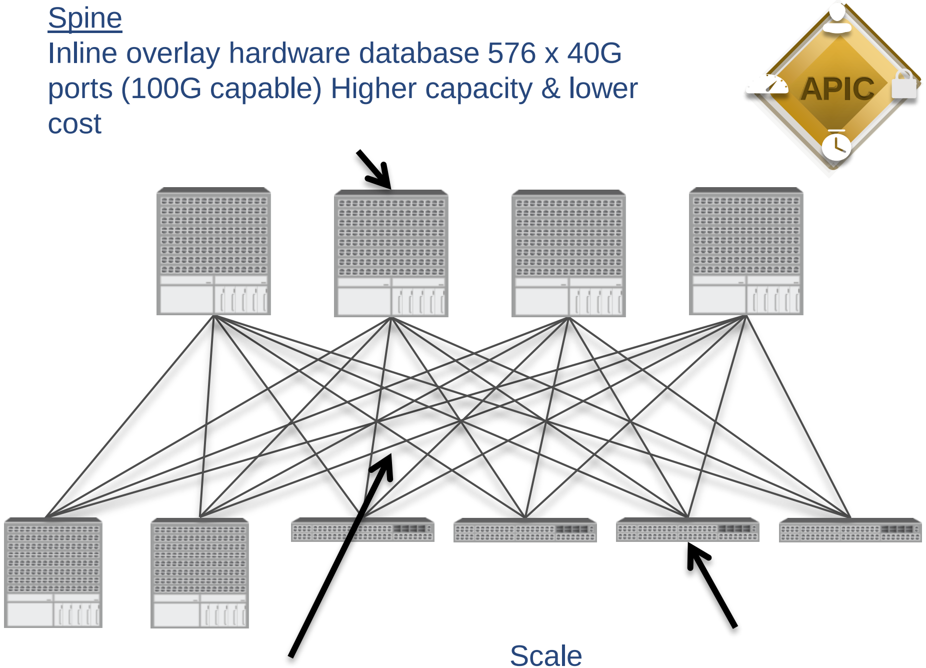


ACI - Based on a Better Network

ACI Fabric

- Industry's most efficient fabric
  - 1/10G edge - High density 40G spine (100G capable)
  - 1M+ IPv4 & IPv6 endpoints
  - 64K+ Tenants
  - 55K+ 1/10G Hosts in a single tier 3:1 oversubscribed Fabric
- Routed fabric – Optimal IP Forwarding
  - Bridging (L2) *and* Routing (L3) of VXLAN/NVGRE/VLAN at scale
  - No x86 GW's – Physical & Virtual
  - Application Agility – Place & Join without limits in Fabric
- Full visibility into virtual and physical
- Common operations from Hypervisor to Compute, To Fabric, to WAN

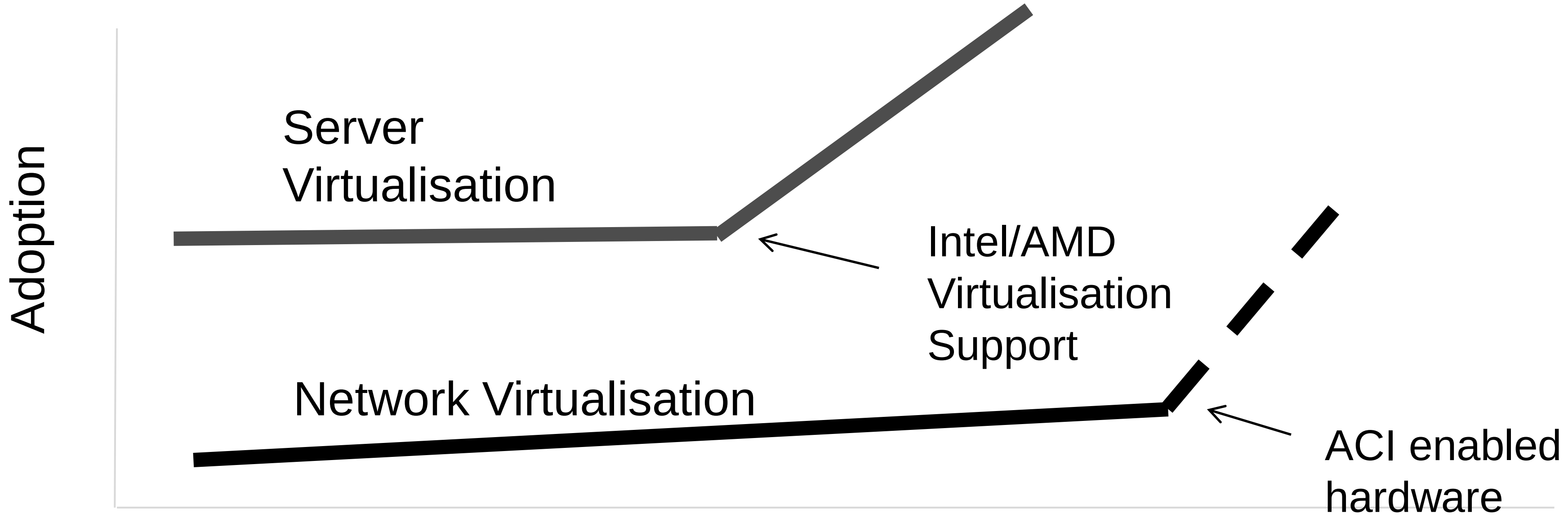
Spine
Inline overlay hardware database 576 x 40G ports (100G capable) Higher capacity & lower cost



Fabric Optimisation
Improved Utilisation
1588 Timing & Latency
ECMP based approaches

Scale
Intelligent caching
Overlay hardware offload
Improved Analytics

ACI Enabled Hardware – “Market Transition”



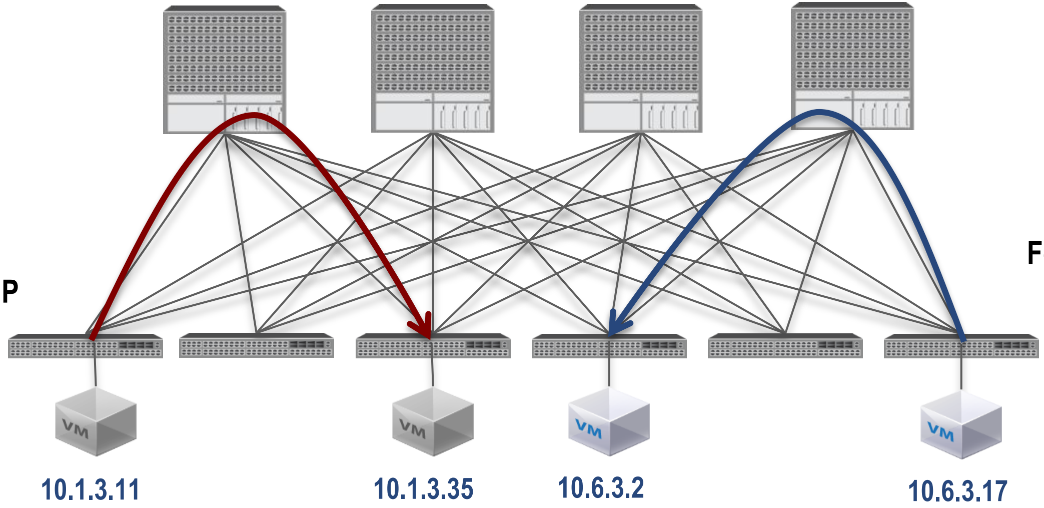
True virtualisation and abstraction requires hardware innovation

ACI - Host Routed Fabric

Layer 2 and Layer 3

IP Forwarding:

Forwarded using DIPi address, HW learning of IP address



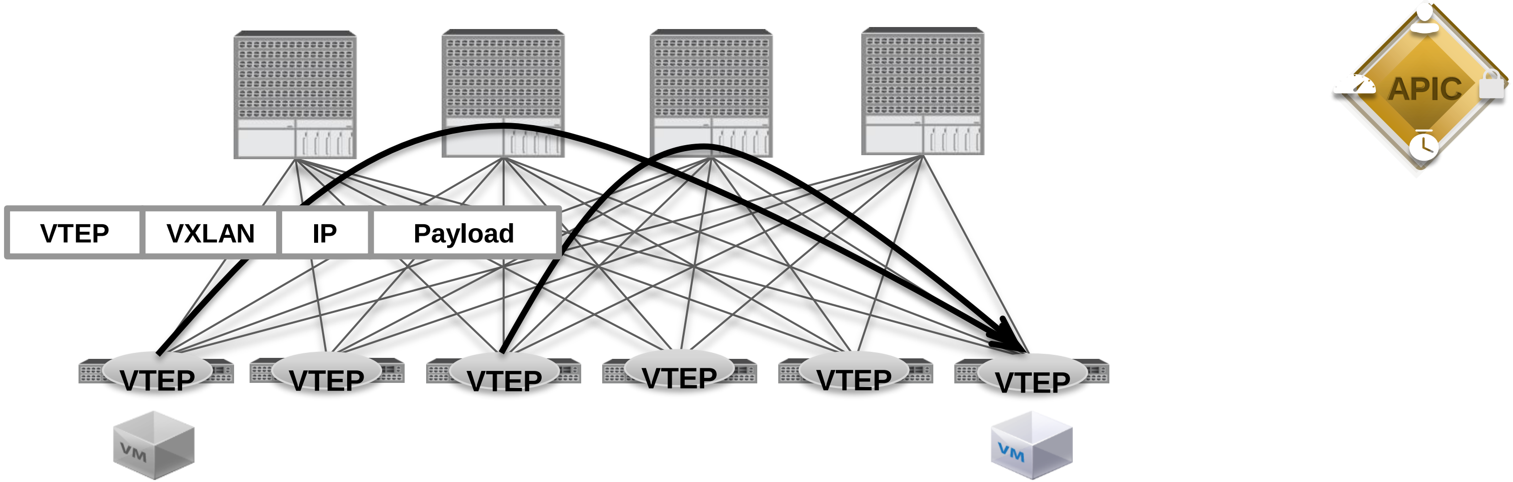
MAC Forwarding:

Forwarded using DMAC address, HW learning of MAC address

- Forward based on destination IP Address for intra and inter subnet (Default Mode)
    - Bridge semantics are preserved for intra subnet traffic (no TTL decrement, no MAC header rewrite, etc.)
    - Non-IP packets will be forwarded using MAC address. Fabric will learn MAC's for non-IP packets, IP address learning for all other packets
- Route if MAC is router-mac, otherwise bridge (standard L2/L3 behaviour)

ACI Fabric

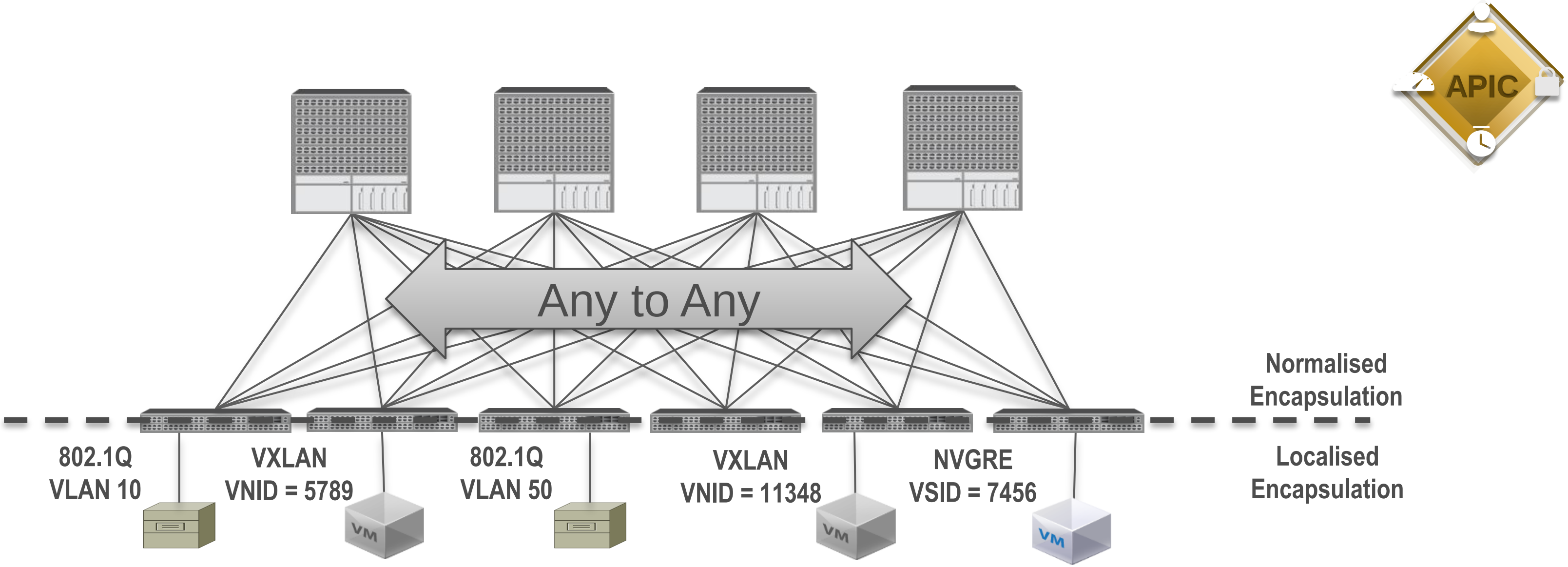
Decoupled Identity, Location & Policy



- ACI Fabric decouples the tenant end-point address, it’s “identifier”, from the location of that end-point which is defined by it’s “locator” or VTEP address
- Forwarding within the Fabric is between VTEPs (VXLAN tunnel endpoints) and leverages an extender VXLAN header format referred to as the VXLAN policy header
- The mapping of the internal tenant MAC or IP address to location is performed by the VTEP using a distributed mapping database

Physical, Any Virtual and Distributed

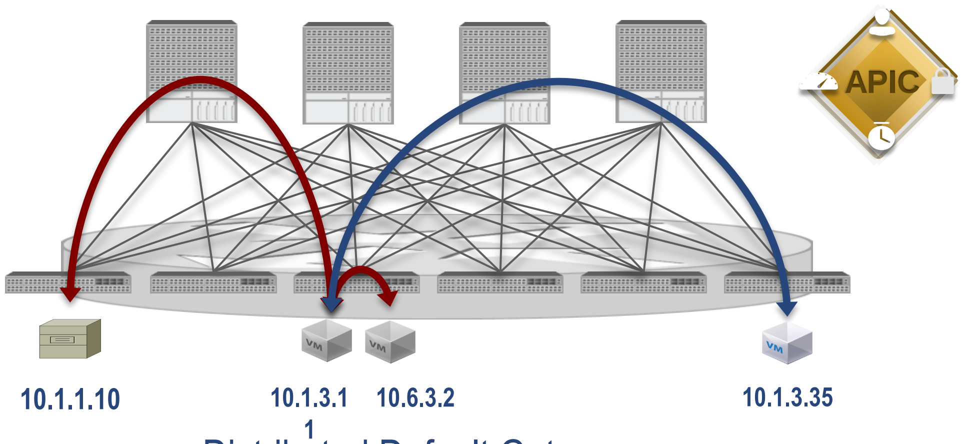
Encapsulation Normalisation



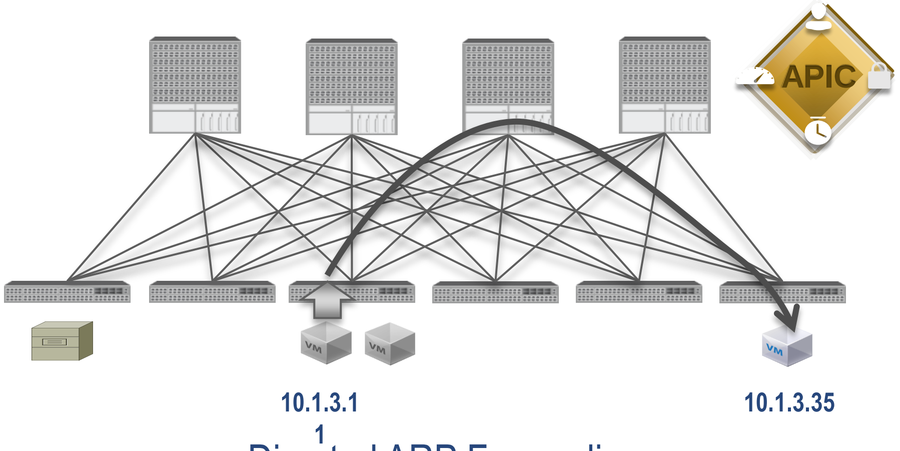
Forwarding is ‘not’ limited to nor constrained by the encapsulation type or encapsulation specific ‘overlay’ network

Location Independent Forwarding

Layer 2 and Layer 3



Distributed Default Gateway



Directed ARP Forwarding

- ACI Fabric supports full layer 2 and layer 3 forwarding semantics, no changes required to applications or end point IP stacks
- ACI Fabric provides optimal forwarding for layer 2 and layer 3
  - Fabric provides a pervasive SVI which allows for a distributed default gateway
  - Layer 2 and layer 3 traffic is directly forwarded to destination end point
- IP ARP/GARP packets are forwarded directly to target end point address contained within ARP/GARP header (elimination of flooding)

Host Routed Fabric

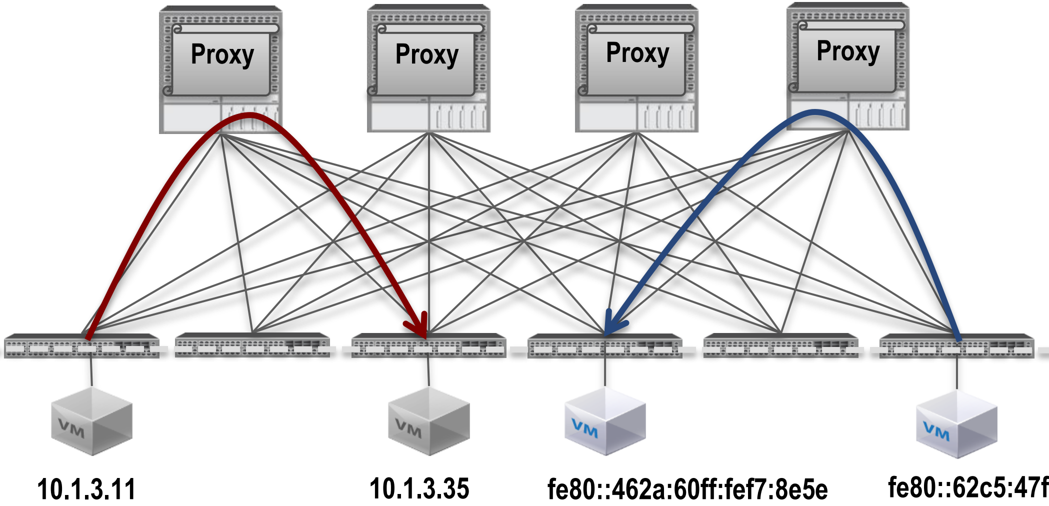
Inline Hardware Mapping DB - 1,000,000+ hosts

Global Station Table contains a local cache of the fabric endpoints

10.1.3.35	Leaf 3
*	Proxy A

10.1.3.11	Port 9

Local Station Table contains addresses of 'all' hosts attached directly to the iLeaf



10.1.3.35	Leaf 3
10.1.3.11	Leaf 1
fe80::8e5e	Leaf 4
fe80::5b1a	Leaf 6

Proxy Station Table contains addresses of 'all' hosts attached to the fabric

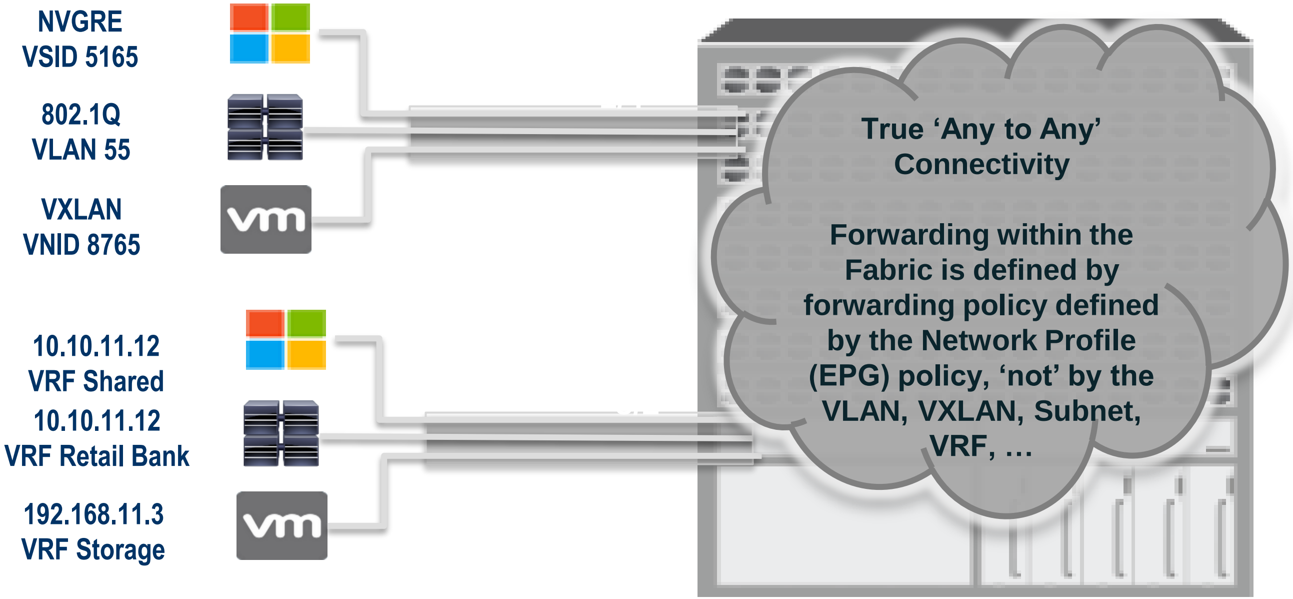
- The Forwarding Table on the Leaf Switch is divided between local (directly attached) and global entries
- The Leaf global table is a cached portion of the full global table
- If an endpoint is not found in the local cache the packet is forwarded to the 'default' forwarding table in the spine switches (1,000,000+ entries in the spine forwarding table)

Fabric Infrastructure

Endpoint based forwarding with distributed policy

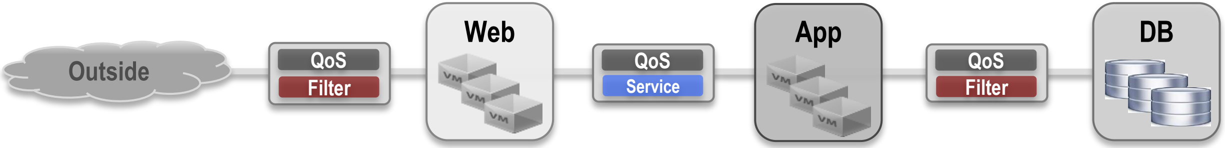
All single port can support all encapsulations simultaneously

Forwarding is defined by Policy EPG ‘Web’ can talk to EPG ‘DB’ independent of IP subnet, VLAN/VXLAN, VRF is Policy says it should



True ‘Any to Any’ Connectivity

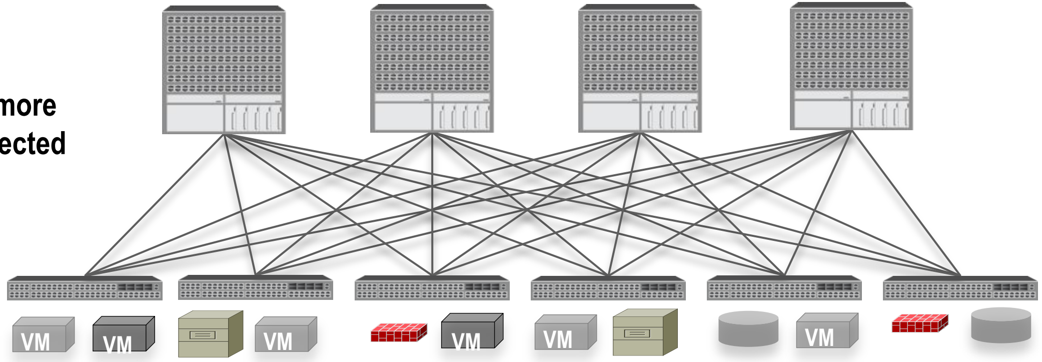
Forwarding within the Fabric is defined by forwarding policy defined by the Network Profile (EPG) policy, ‘not’ by the VLAN, VXLAN, Subnet, VRF, …



ACI Fabric

Why Focus on Next Generation Telemetry

Larger Fabrics make it more difficult to Correlate collected data to a specific Tenant/Application



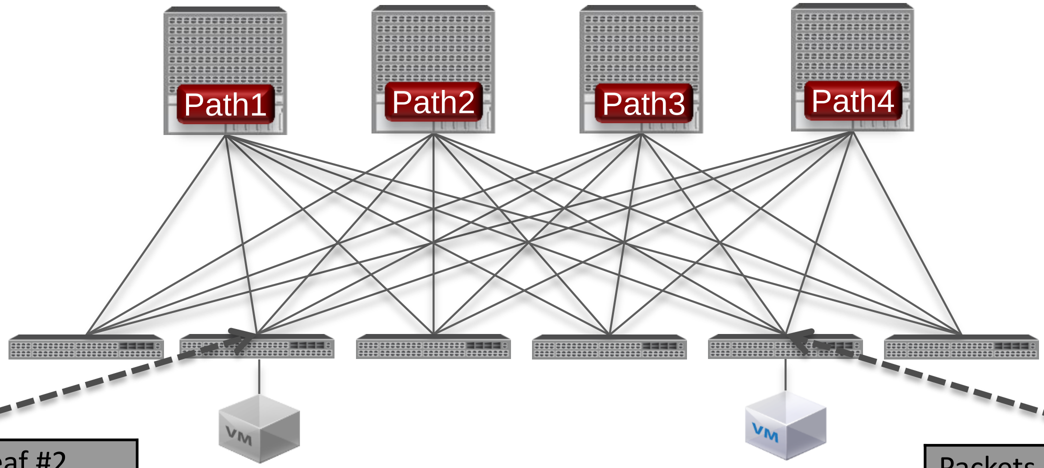
Desire for SLA Monitoring in Shared Environments

Increasing distribution of workload

- Topology and traffic pattern changes require us to re-evaluate the assumptions of Troubleshooting and Capacity Planning within the data centre
  - Higher degree of sharing combined with Distributed/Mobile Workloads require more information and more contextualised information
- ACI Fabric Capabilities
  - Atomic Counters
  - Latency Metrics

Telemetry

Atomic Counters



Packets sent from Leaf #2 to Leaf #5	
Path 1	2068
Path 2	2963
Path 3	2866
Path 4	2506

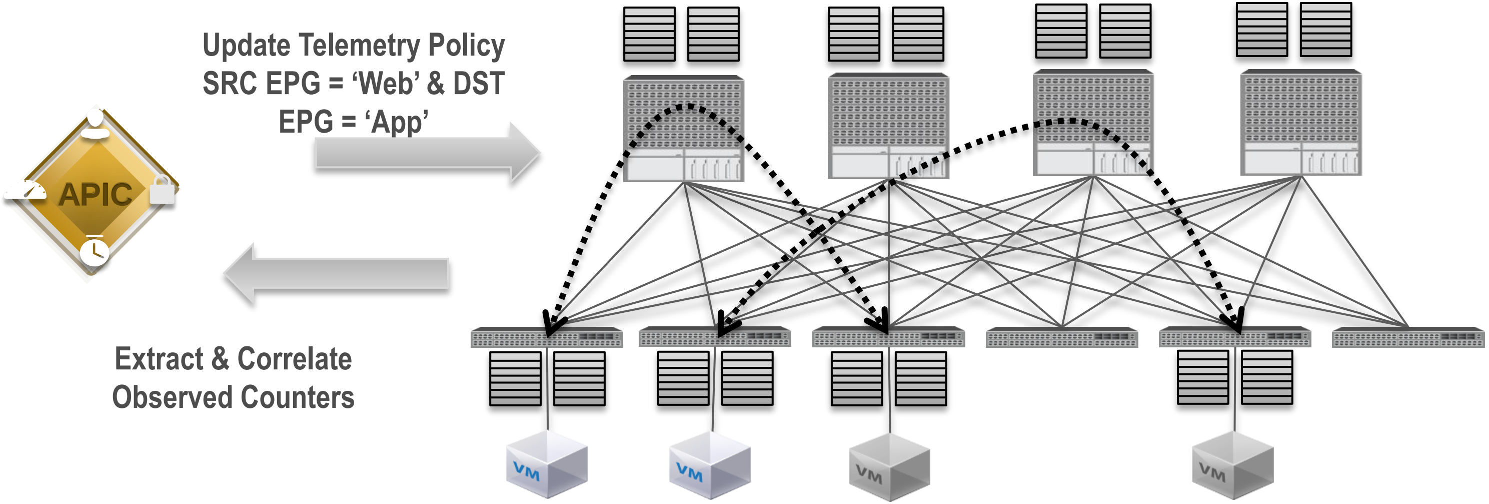
Difference	
Path 1	2
Path 2	0
Path 3	-3
Path 4	0

Packets Received on Leaf #5 sent from Leaf #2	
Path 1	2066
Path 2	2963
Path 3	2869
Path 4	2506

Telemetry

Filter Based Atomic Counters

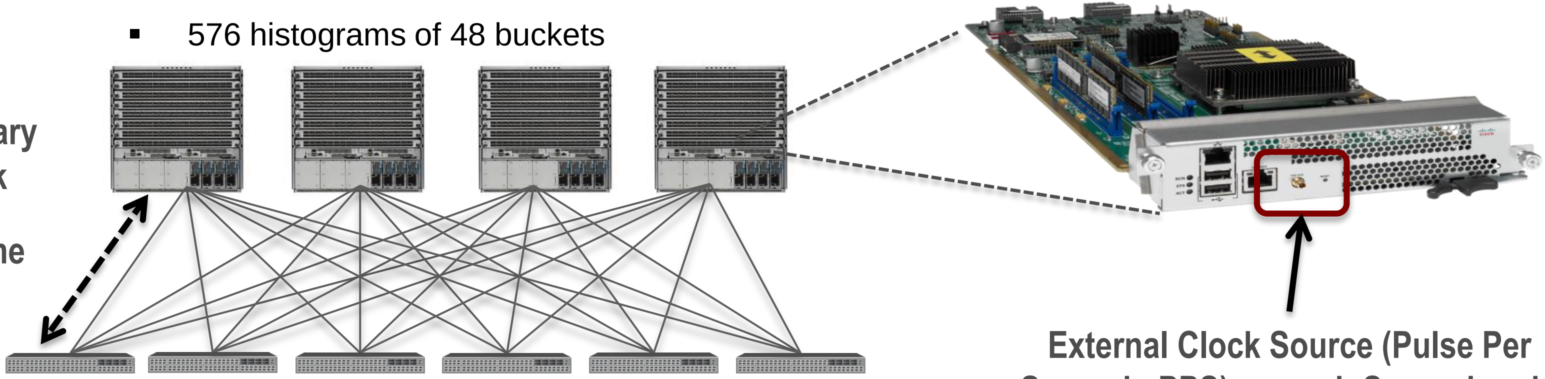
- A second Bank of counters are used for on-demand monitoring
- Counters are incremented if a programmed TCAM entry is matched & the odd/even bit is set
- TCAM match is programmed via policy on the APIC and distributed to all nodes
  - Criteria to match against: EPG, IP Address, TCP/UDP port, Tenant VRF or Bridge Domain


Update Telemetry Policy SRC EPG = 'Web' & DST EPG = 'App'
APIC
Extract & Correlate Observed Counters
VM
VM
VM
VM

Telemetry

Fabric Latency Measurements

- Matrix of Latency Measurements between all Leaves
  - Per Port Average Latency & Variance to up to 576 other iLeaves
    - Maximum, Accumulation, Sum of Square and Packet Count
  - Per Port 99% Latency (recorded to up to 576 other iLeaves)
    - 99% of all packets have recorded latency less than this value
  - 48 bucket histogram
    - 576 histograms of 48 buckets



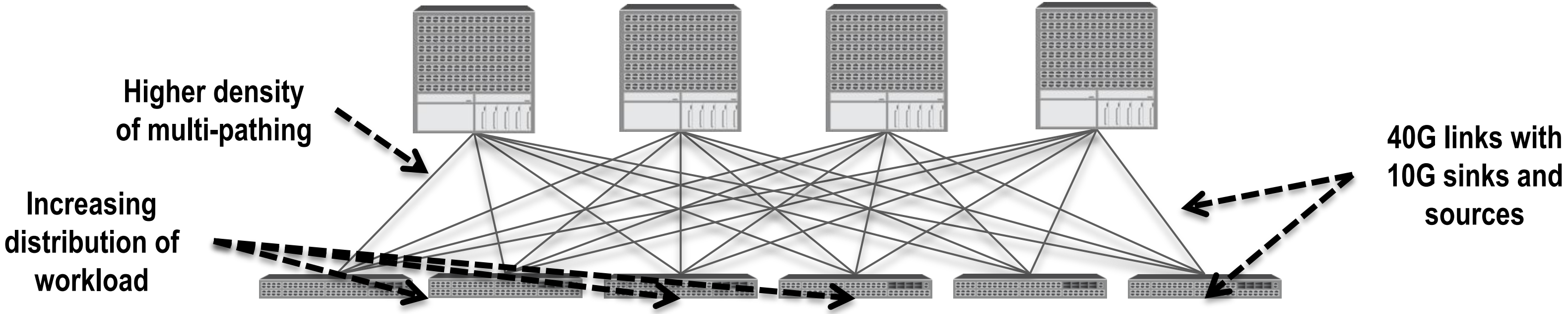
Boundary Clock

PTP Time Sync

External Clock Source (Pulse Per Second - PPS) on each Supervisor in the Spine Chassis

ACI Fabric

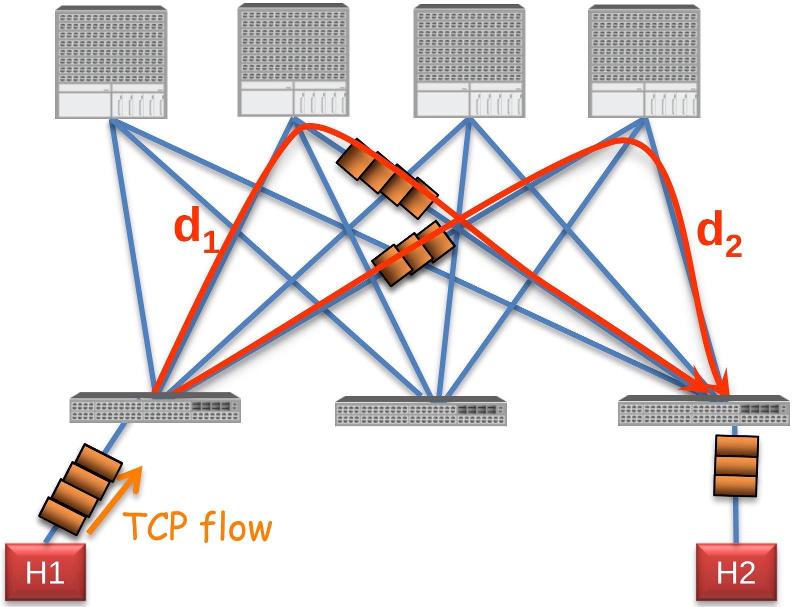
Why focus on next generation DC QoS


Higher density of multi-pathing
Increasing distribution of workload
40G links with 10G sinks and sources

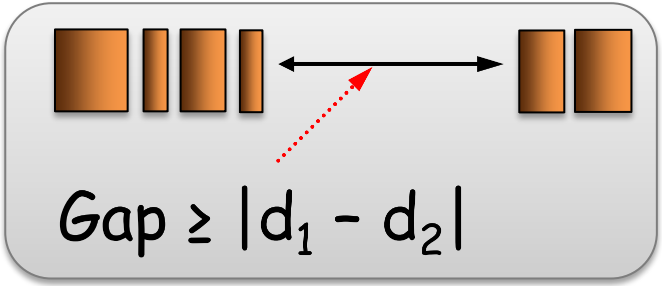
- Topology and traffic pattern changes require us to re-evaluate the assumptions of congestion management within the Data centre
  - Higher density of uplinks with greater multi-pathing ratio is resulting in more variability in congestion patterns
  - Distribution of workload is adding another dimension of traffic patterns
- Two options
  - Spend the time to statically engineering marking, queuing and traffic patterns to accommodate these new
  - Build a more systems based reactive approach to congestion management for traffic within the Data centre

ACI Fabric Load Balancing

Flowlet Switching



- State-of-the-art ECMP hashes flows (5-tuples) to path to prevent reordering TCP packets.
- ***Flowlet switching**** routes bursts of packets from the same flow independently.
- No packet re-ordering



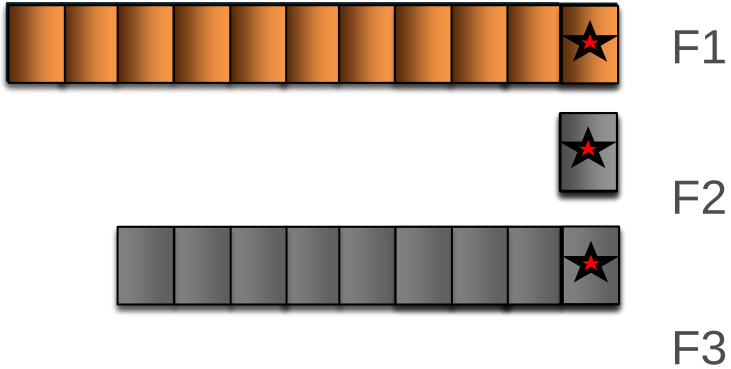
$\text{Gap} \geq |d_1 - d_2|$

*Flowlet Switching (Kandula et al ’04)

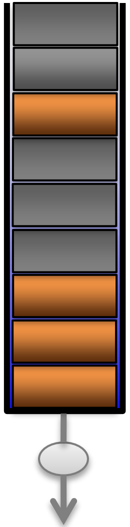
ACI Fabric Load Balancing

Dynamic Flow Prioritisation

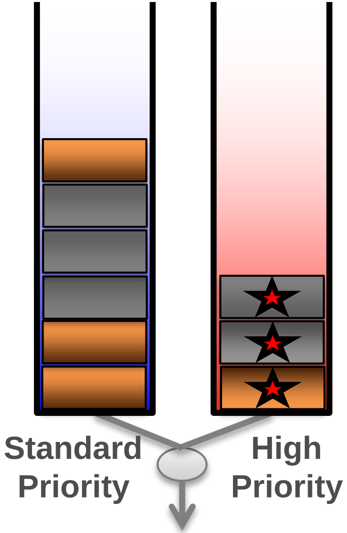
Real traffic is a mix of large (elephant) and small (mice) flows.



Key Idea:
Fabric detects initial few flowlets of each flow and assigns them to a high priority class.



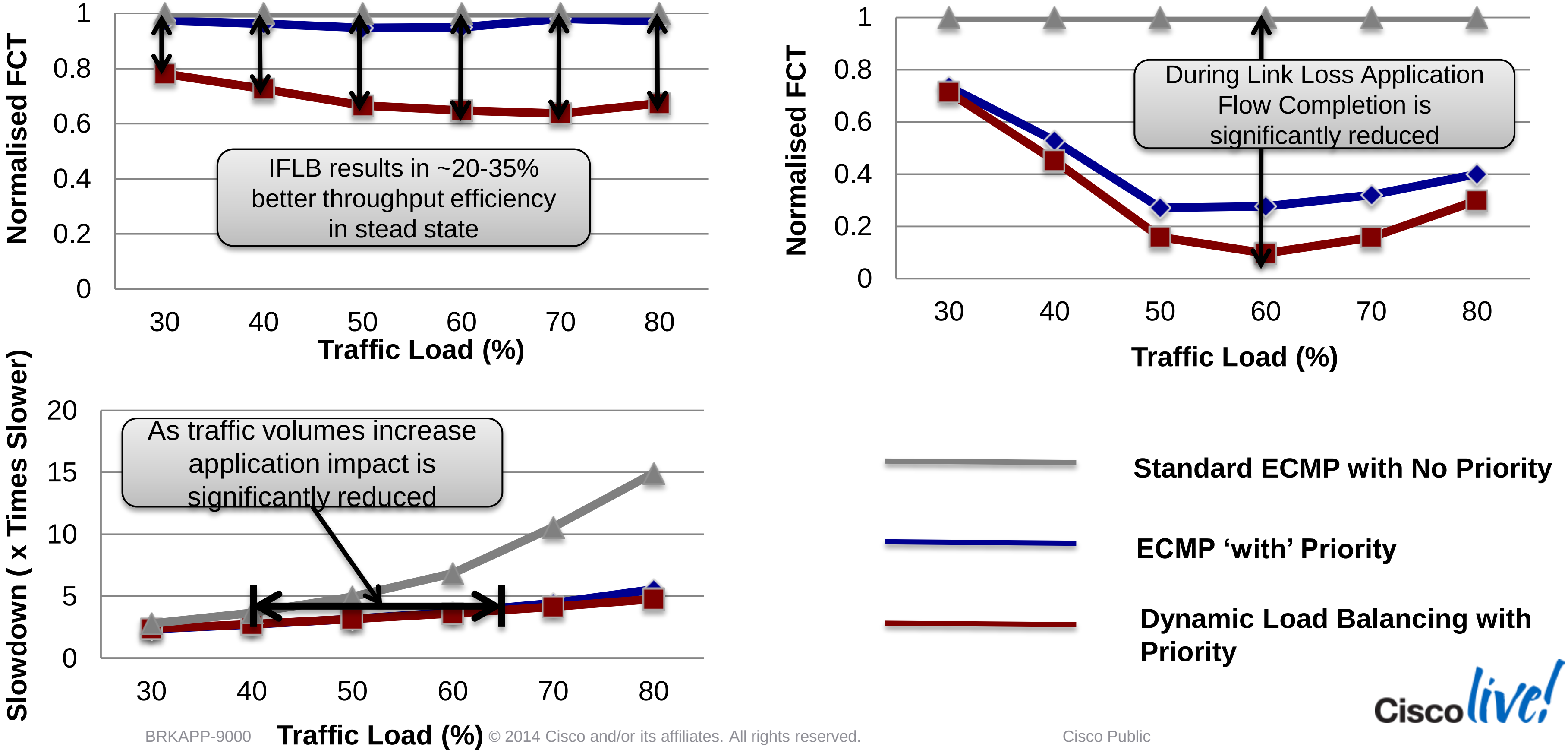
Standard (single priority): Large flows severely impact performance (latency & loss). for small flows



Dynamic Flow Prioritisation: Fabric automatically gives a higher priority to small flows.

Application Performance Improvements

ACI Fabric Load Balancing



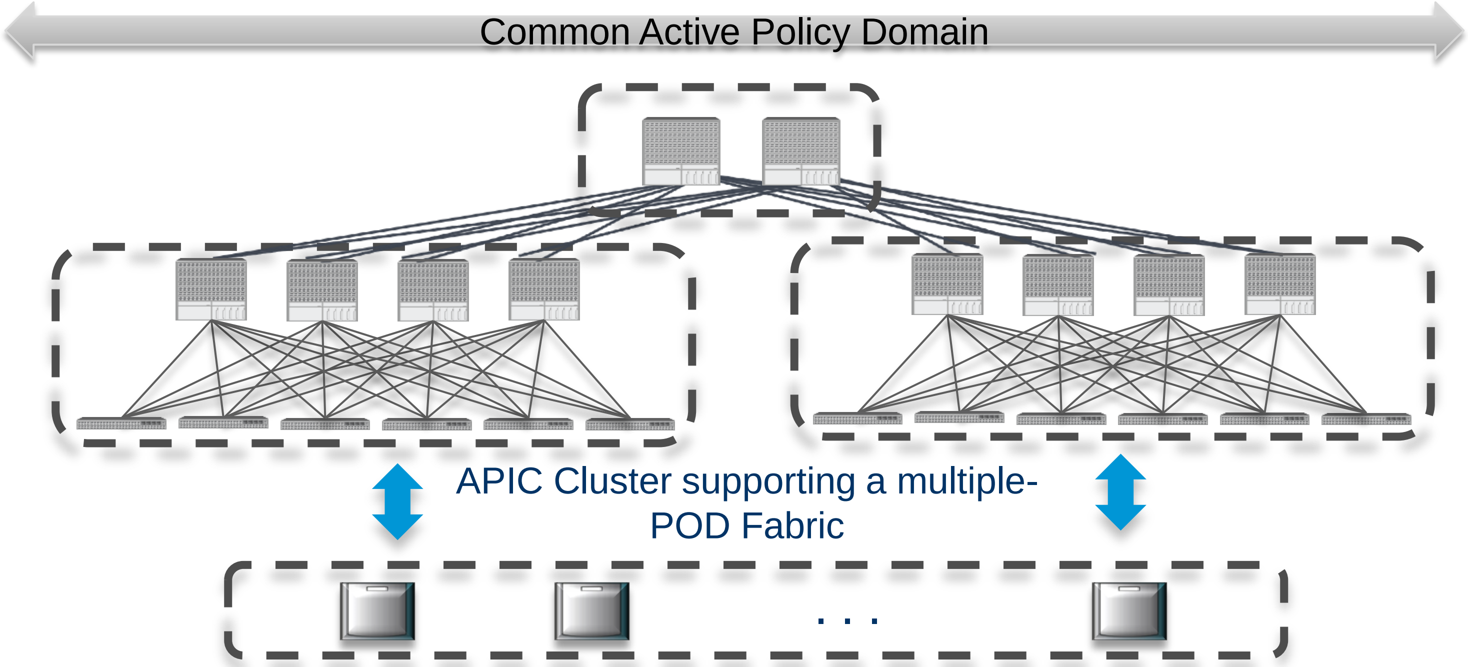
IFLB results in ~20-35% better throughput efficiency in stead state

During Link Loss Application Flow Completion is significantly reduced

As traffic volumes increase application impact is significantly reduced

Distribution of Workloads

Aggregation of Fabric PODs

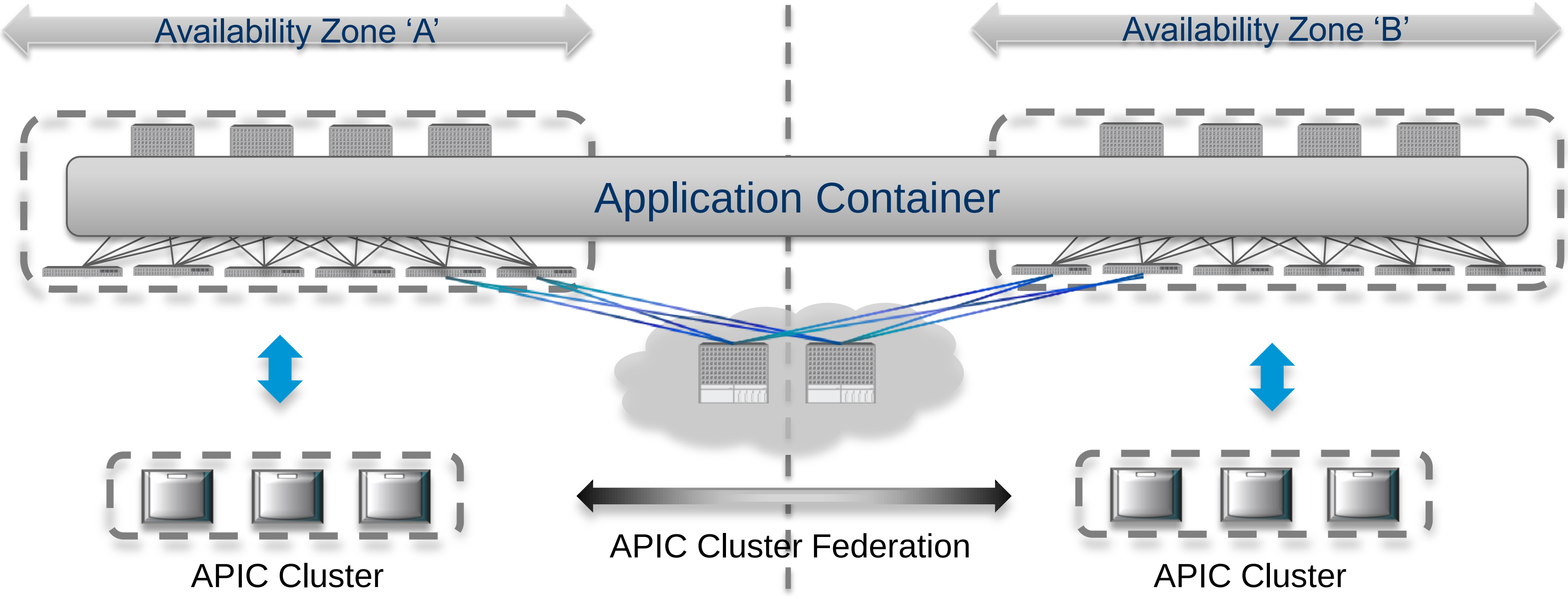


Common Active Policy Domain

APIC Cluster supporting a multiple-POD Fabric

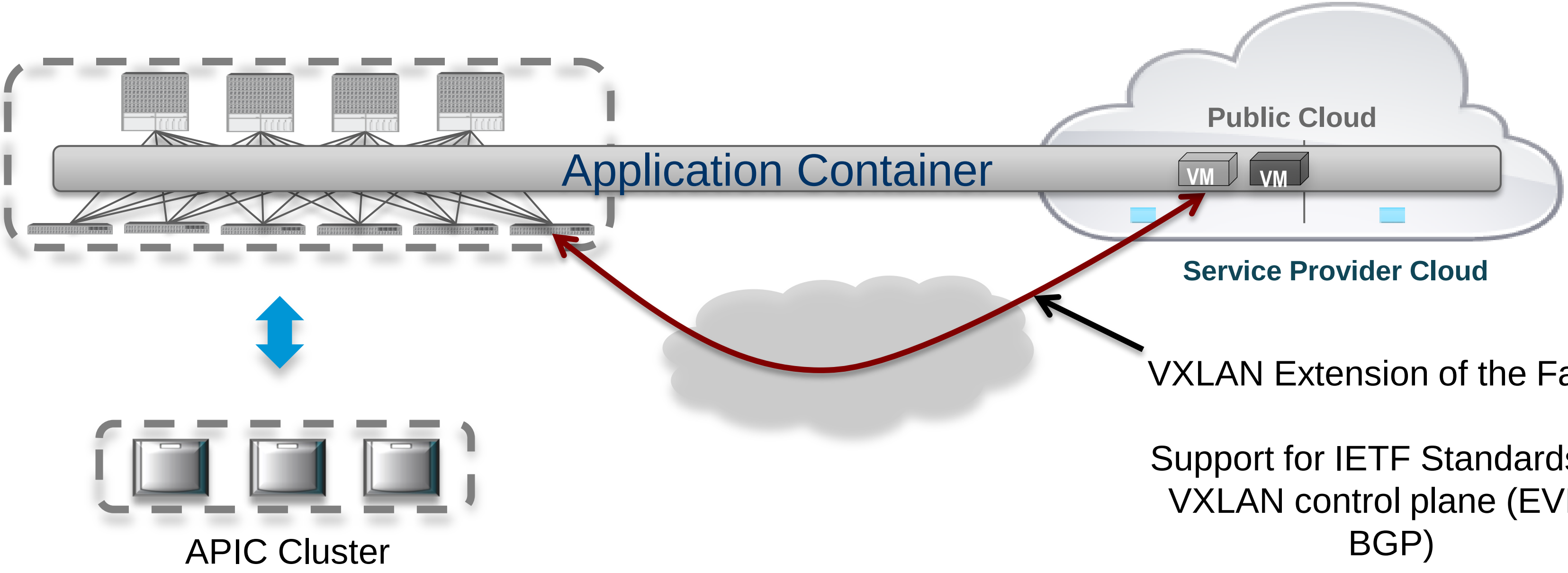
Availability Zones

Distributed Application Containers



Workload Extension

Extended Containers

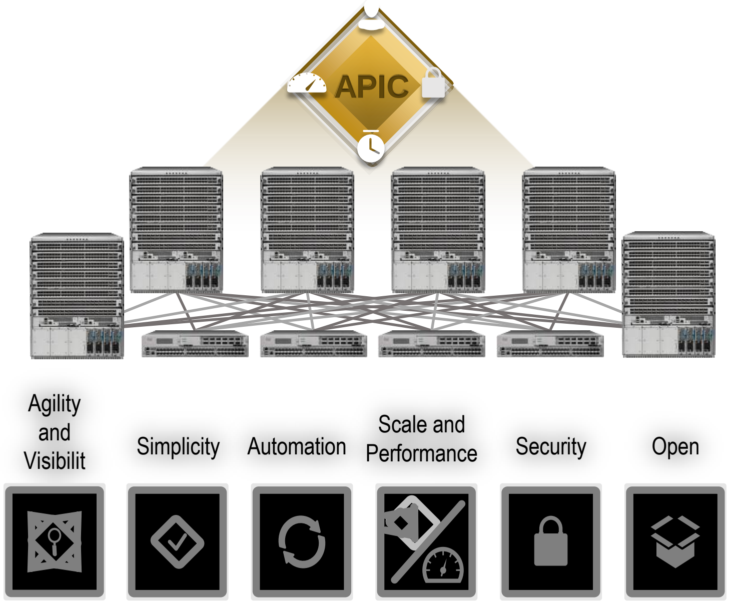


VXLAN Extension of the Fabric

Support for IETF Standards for VXLAN control plane (EVPN BGP)

Agenda – Application Centric Infrastructure

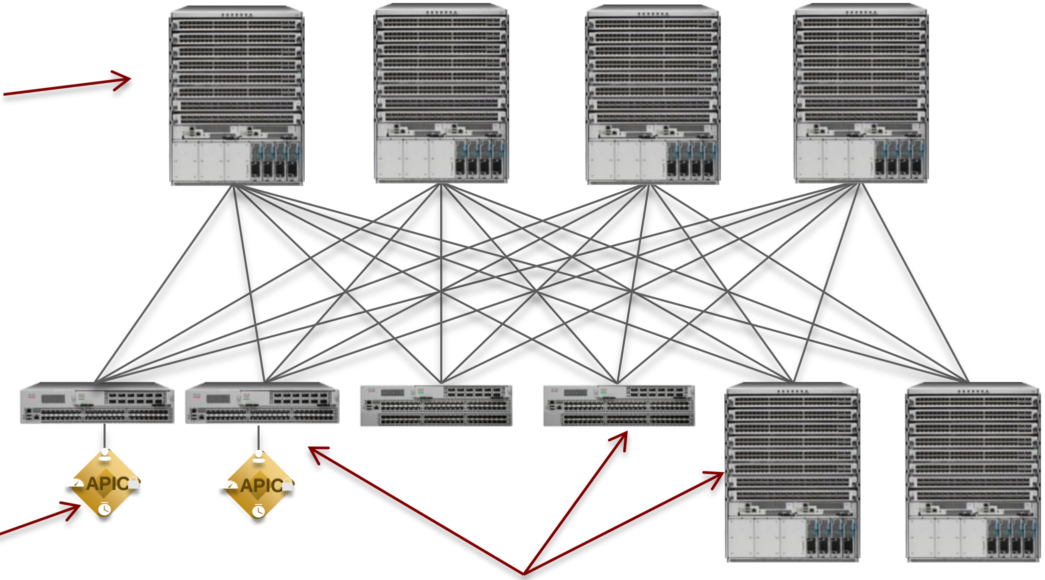
- What is ACI - Concepts and Principles
  - Why, What & How
- Foundations of ACI
  - ACI Fabric
  - Nexus 9000
  - ACI Policy Model
  - Hypervisor Integration, VMWare, MSFT and KVM
  - Integration and Automation of L4-7 Services
  - APIC (The Controller)
- Integration, Migration and Co-Existence with Existing Infrastructure
- Open Standards, Open Source, Open API’s



ACI Fabric

Fabric Spine Nodes

- 16 Slot Modular
- 8 Slot Modular
- 4 Slot Modular
- Mini-Spine (36 ports)



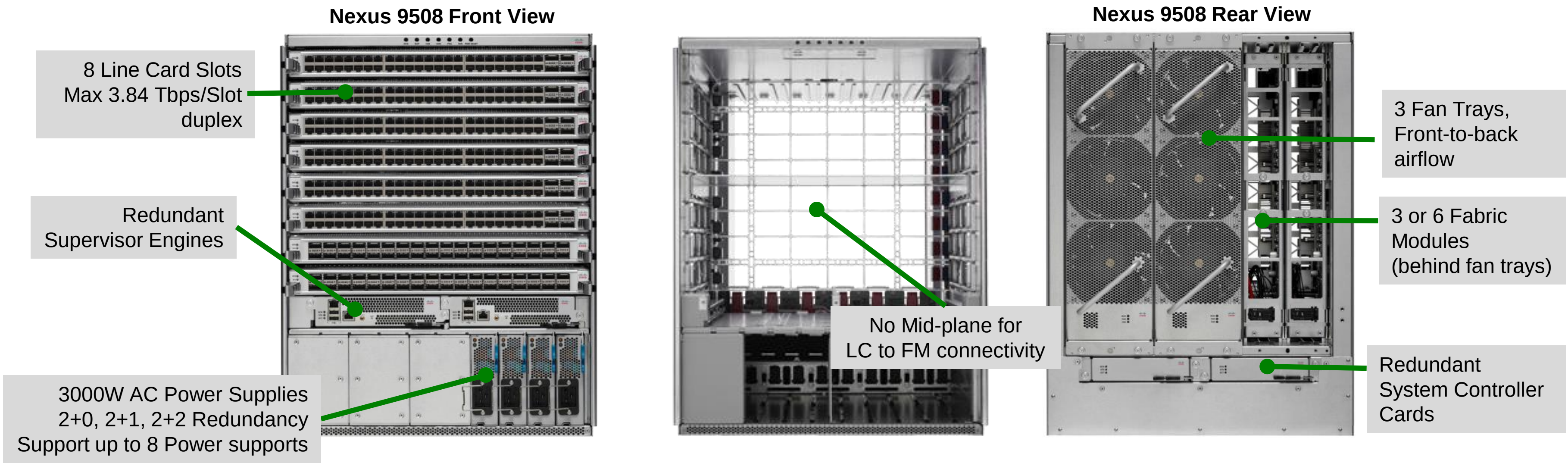
APIC Servers

UCS ‘C’ Series (Intel)

Fabric Leaf Nodes
4, 8 & 16 slot Modular (post FCS)
48 x 1/10 + 12 x 40G
96 x 1/10 + 8 x 40G
Variety of 1 & 2 RU form factors (post FCS)

ACI Optimised Hardware

Nexus 9500 Modular Chassis

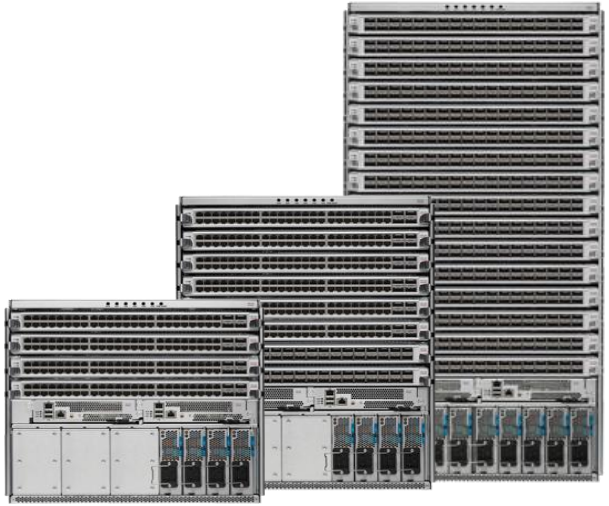


Nexus 9508 Front View
Nexus 9508 Rear View
8 Line Card Slots
Max 3.84 Tbps/Slot duplex
Redundant Supervisor Engines
3000W AC Power Supplies
2+0, 2+1, 2+2 Redundancy
Support up to 8 Power supports
No Mid-plane for LC to FM connectivity
3 Fan Trays, Front-to-back airflow
3 or 6 Fabric Modules (behind fan trays)
Redundant System Controller Cards

Mechanical Advancements

- No Mid-Plane (Better airflow, Better MTBF, Longevity)
- Both a Supervisor ‘and’ a System Controller (Better Control Plane Scale)
- Power footprint future proofed for 100/400G
- Common Components across 4, 8 & 16 slot chassis

Modular Switch Platform – Nexus 9500



- Nexus 9500 Modular Chassis
  - 4, 8 & 16 payload slots
  - Common Supervisor, Power Supply, Line Cards

40G Aggregation

36 ports 40G QSFP+ (Non Blocking)

9600 Series

1/10G Access and 10/40G Aggregation

48 ports 10G SFP+ & 4 ports 40G QSFP+
48 ports 1/10G-T & 4 ports 40G QSFP+
(non blocking)

36 ports 40G QSFP+ ((1.5:1 oversubscribed)

9500 Series

40G Fabric Spine

36 ports 40G QSFP+ (Non Blocking)

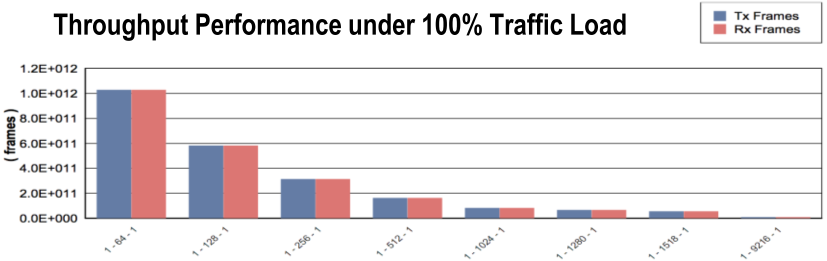
9700 Series

Nexus 9000 Series

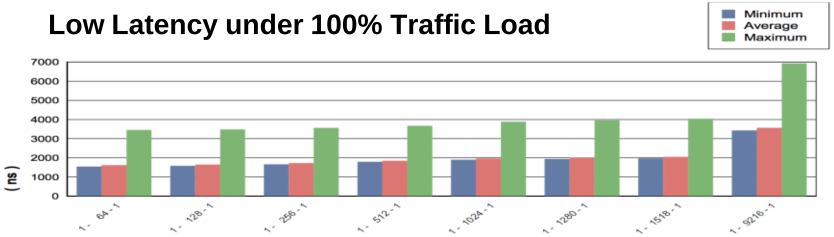
Full Line Rate Throughput Performance

- Line Rate, Low + Consistent Latency + High MTBF + Power Optimised
  - Platinum rated PS (90%-94% power efficiency across all work loads)
  - All ports are line rate at 100% unicast traffic load
  - All ports are line rate at 100% multicast traffic load
  - Full line rate for all packet sizes (64~9216 Bytes)
- Highly integrated switch and buffer functionality
  - Only 2 to 4 ASICs per line card
  - Mix of 28nm Cisco and 40nm Broadcom ASICs

Throughput Performance under 100% Traffic Load



Low Latency under 100% Traffic Load



Traffic type	Power (watts)	Fan Speed
No traffic	3233	0%
100% line-rate with IMIX packets	4746	20%
100% line-rate with 64 byte packets	5470	25%

Nexus 9300 Platform Architecture

Uplink Module

- 12-port 40 Gb QSFP+
- Additional 40 MB buffer
- Full VXLAN gateway, bridging and routing capability

Nexus® 9396PQ

- 960G
- 48-port 1/10 Gb SFP+ and 12-port 40 Gb QSFP+
- 2 RU

Nexus 9396TX (future)

- 960G
- 48-port 1/10 GBaseT & 12-port 40 Gb QSFP+
- 2 RU

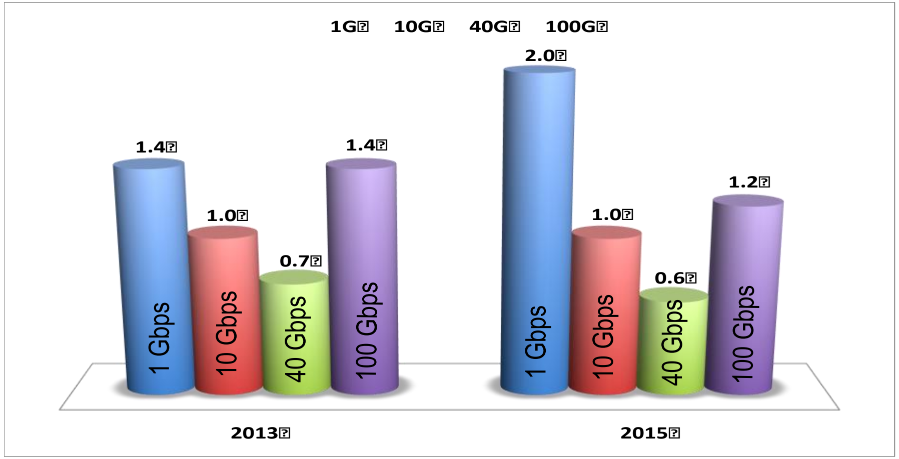
Nexus 93128TX

- 1,280G
- 96-port 1/10 G-T and 8-port 40 Gb QSFP+
- 3 RU

Nexus 9300 - Common

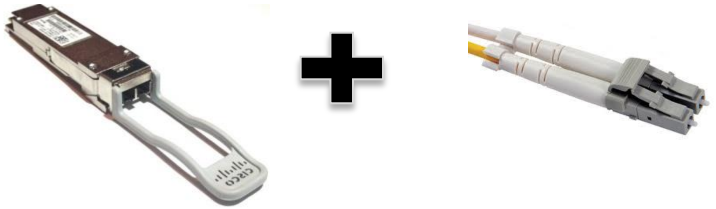
- Redundant fan and power supply
- Front-to-back and back-to-front airflow
- Dual- core CPU with default 64 GB SDD

Why a 40G Fabric?



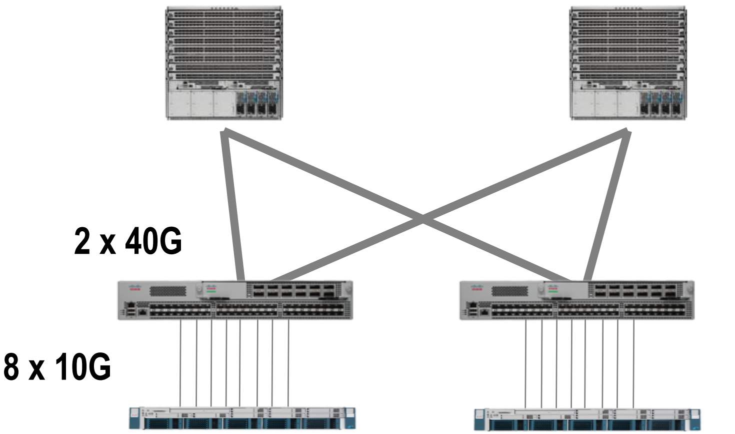
- Optimal Fabric Capacity and Cost
  - 40G provides the optimal cost point currently
  - Speed-up (higher speed transport than edge ports) necessary to achieve effective throughput in a switching network
  - 100G support (Future)

- 40G BiDi Optics
  - QSFP pluggable, MSA compliant
  - Dual LC Connector
  - Support for 100m on OM3 and 125m+ on OM4
  - TX/RX on 2 wavelength @ 20G each

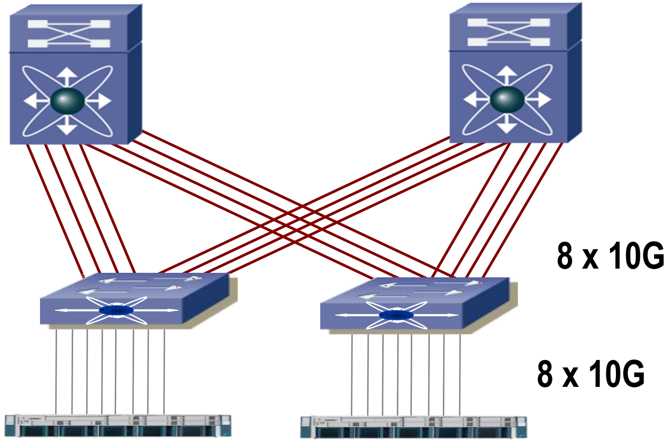


Why a 40G Fabric?

Increased BW Utilisation due to 40G speedup



Expected Max Effective Throughput = 86.33%



Expected Max Effective Throughput = 65.6%

Network Switching Designs have leveraged an uplink speed ups to avoid hashing collisions to the provide effective utilisation of available capacity

*A speedup of 40G on uplinks for 10G attached servers results in Flow Completion Times that are ~12–40% lower than that of a 10G fabric**

Without a speed up the capacity of the infrastructure will be diminished

- *http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6627738*

Nexus 9000

Common Platform: Two Modes of Operation

Standalone

NXOS w/ Enhancements

onePK

Bash Shell

python™

Per-Box Programmability

Network Ops Driven, Switch Automation

ACI

Policy Controller iNXOS

python™

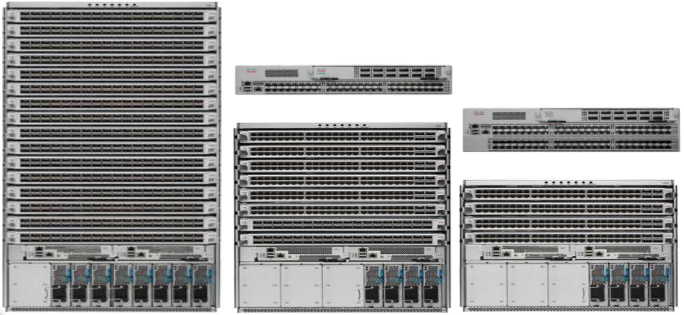
APIC

Bash Shell

Centralised Fabric Programmability

Policy Based Fabric Automation

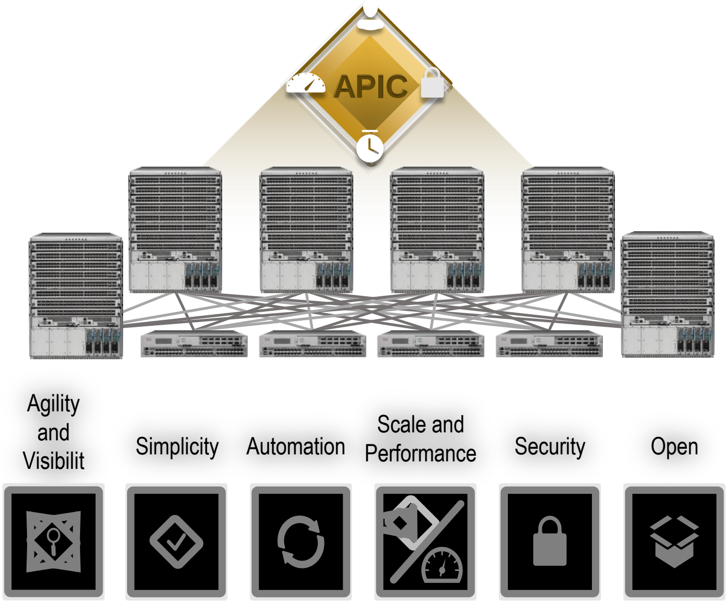
Open, Flexible, & Choice of Programmability Modes



1/10/40/100GE Common Platform

Agenda – Application Centric Infrastructure

- What is ACI - Concepts and Principles
  - Why, What & How
- Foundations of ACI
  - ACI Fabric
  - Nexus 9000
  - ACI Policy Model
  - Hypervisor Integration, VMWare, MSFT and KVM
  - Integration and Automation of L4-7 Services
  - APIC (The Controller)
- Integration, Migration and Co-Existence with Existing Infrastructure
- Open Standards, Open Source, Open API's



End-Points End EPG Membership

VM

Virtual Machine

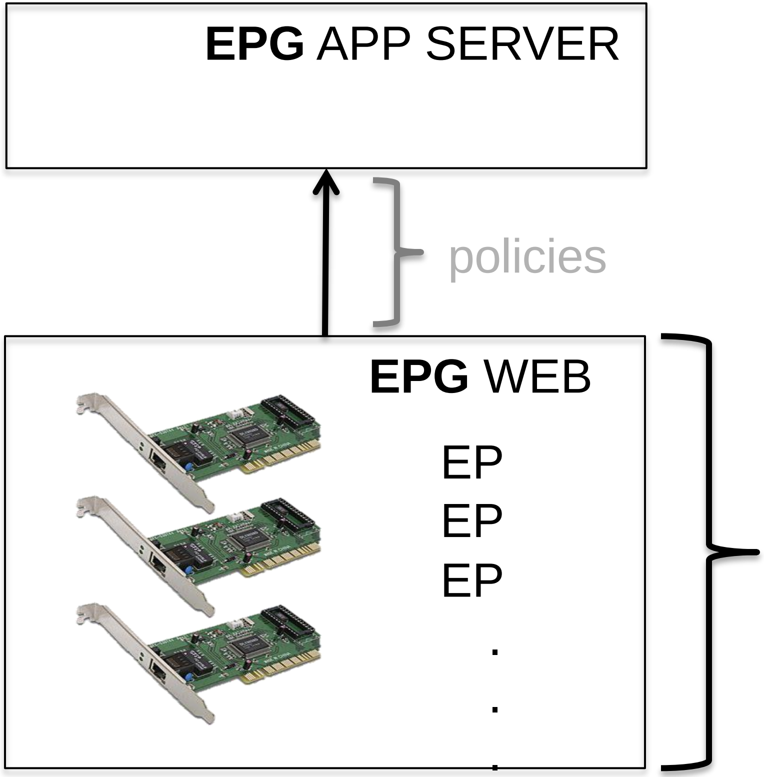
Server

Storage

Client

- Device connected to network directly or indirectly
- Has address (identity), location, attributes (version, patch level)
- Can be physical or virtual
- End Point Group (EPG) membership defined by:
  - Ingress physical port (leaf or FEX)
  - Ingress logical port (VM port group)
  - VLAN ID
  - VXLAN (VNID)
  - IP address (only applicable to external/border leaf connectivity at FCS)
  - IP Prefix/Subnet (only applicable to external/border leaf connectivity at FCS)
  - NVGRE (VSID) (future)
  - DNS/LDAP/RADIUS/… (future)
  - DSCP or Layer 4 ports (future)

End-point Groups EPGs



EPG APP SERVER

policies

EPG WEB

EP

EP

EP

.

.

.

Allows to specify rules and policies on groups of physical or virtual end-points without understanding of specific identifiers and regardless of physical location.

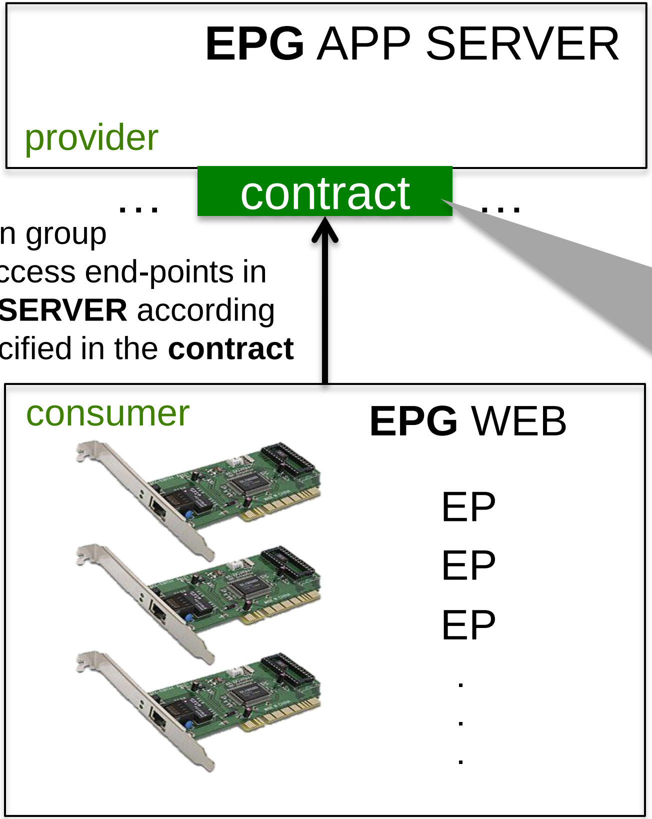
Can flexibly map into

→ application tier of multi-tier app
→ segmentation construct (ala VLAN)
→ a security construct
→ ESX port group
→… **end-point group [*EPG*]**

***All* EPs *share* common properties**

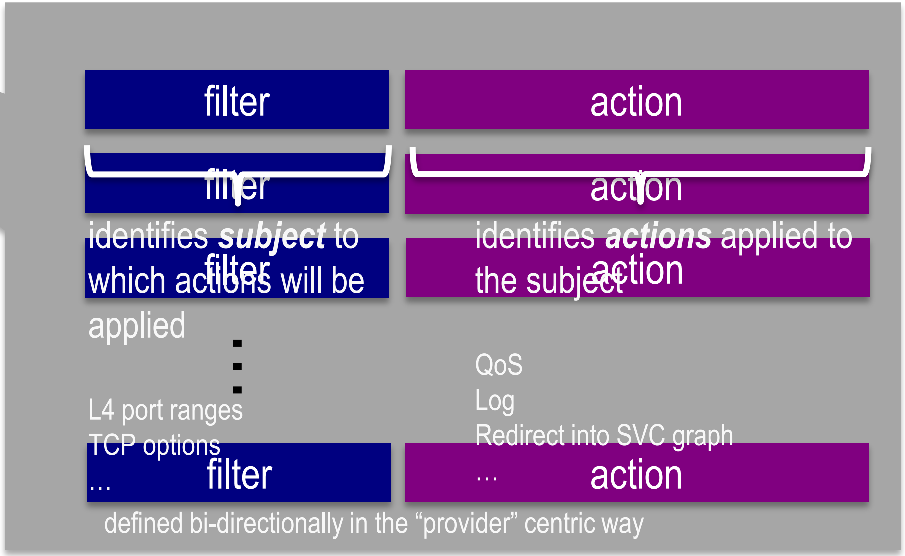
→ Connectivity
→ Security/Access control
→ QoS
→ Services
→ …

End Point Group Contracts



End points in group **WEB** can access end-points in group **APP SERVER** according to rules specified in the **contract**

Allows to specify rules and policies on groups of physical or virtual end-points without understanding of specific identifiers and regardless of physical location.



identifies ***subject*** to which actions will be applied

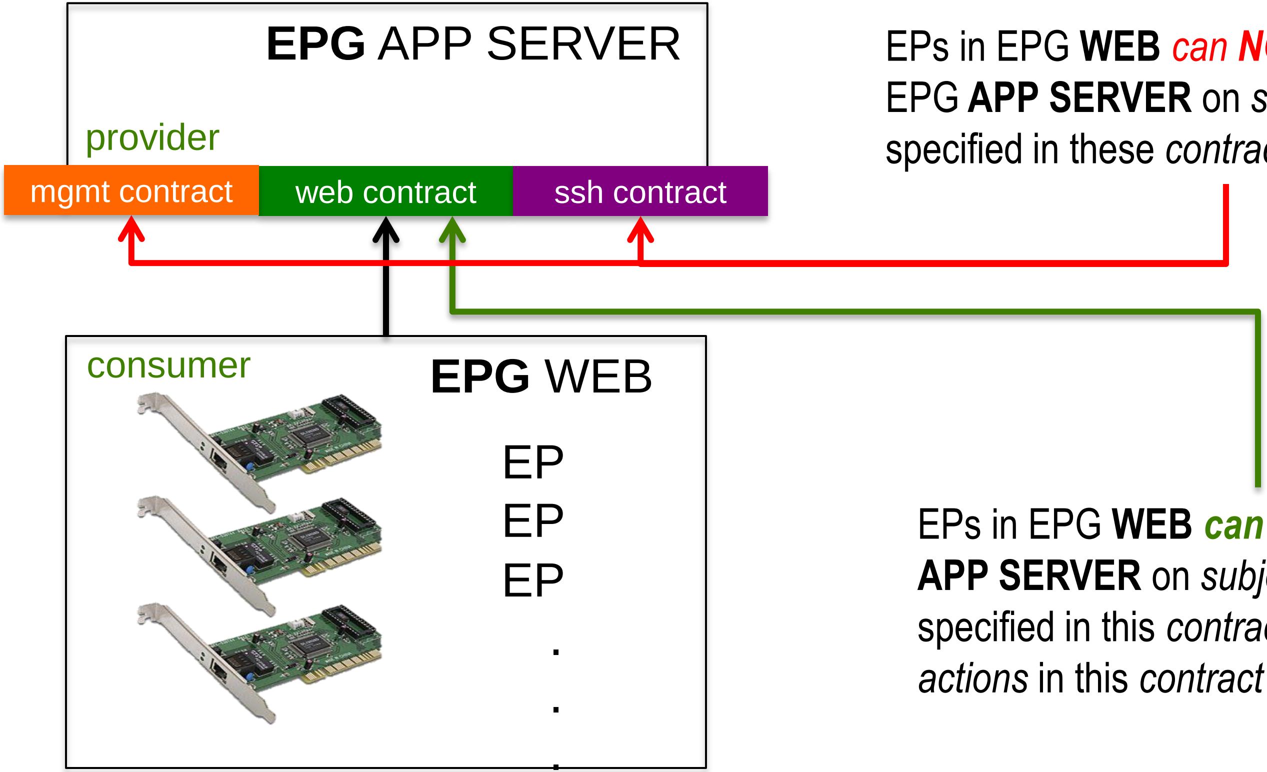
identifies ***actions*** applied to the subject

L4 port ranges
TCP options
…

QoS
Log
Redirect into SVC graph
…

defined bi-directionally in the “provider” centric way

Multiple Contracts

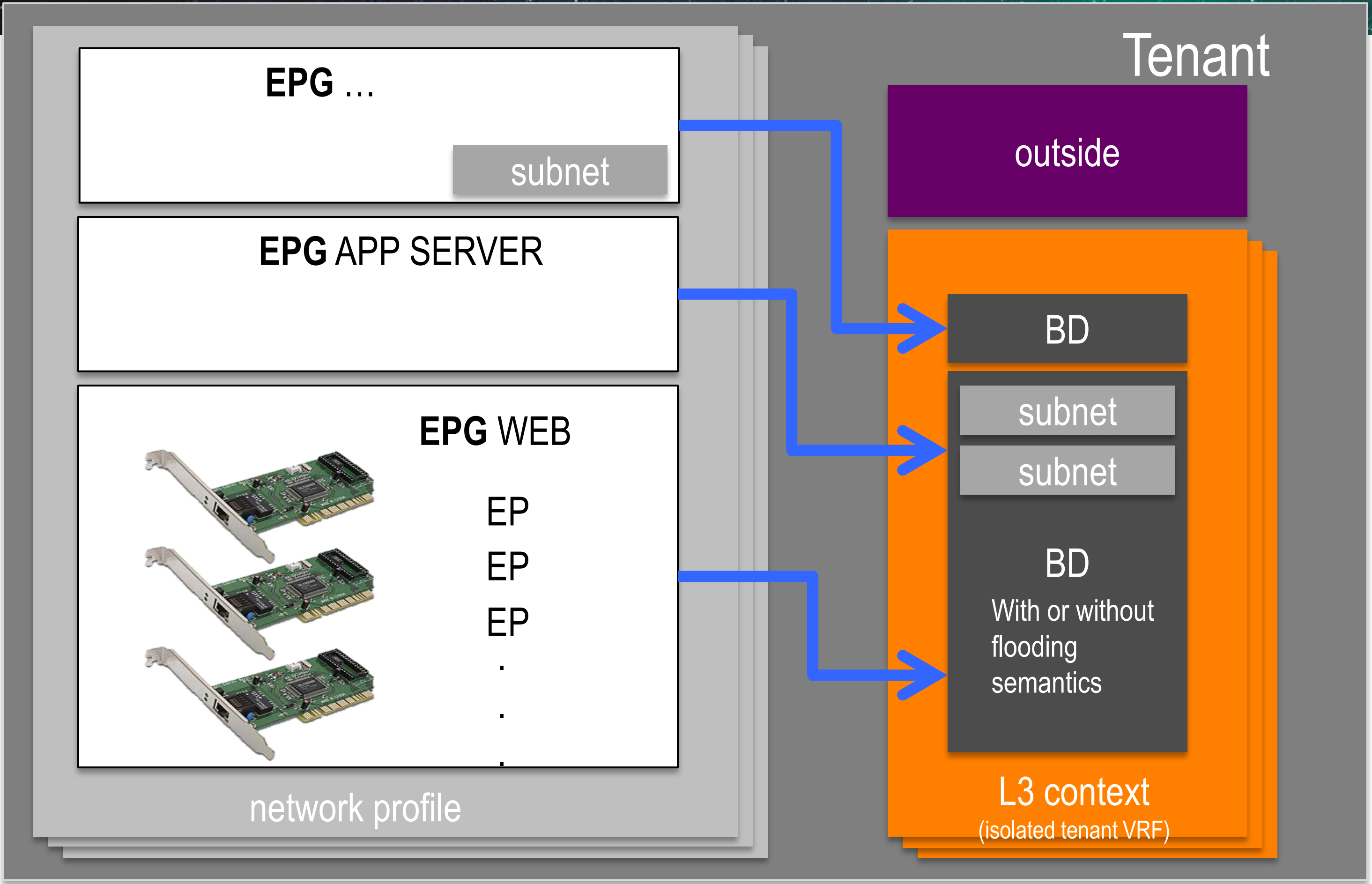


EPs in EPG **WEB** *can* ***NOT*** access EPs in EPG **APP SERVER** on *subjects* (L4 ports) specified in these *contracts*

EPs in EPG **WEB** ***can*** access EPs in EPG **APP SERVER** on *subjects* (L4 ports) specified in this *contract*, subjected to *actions* in this *contract*

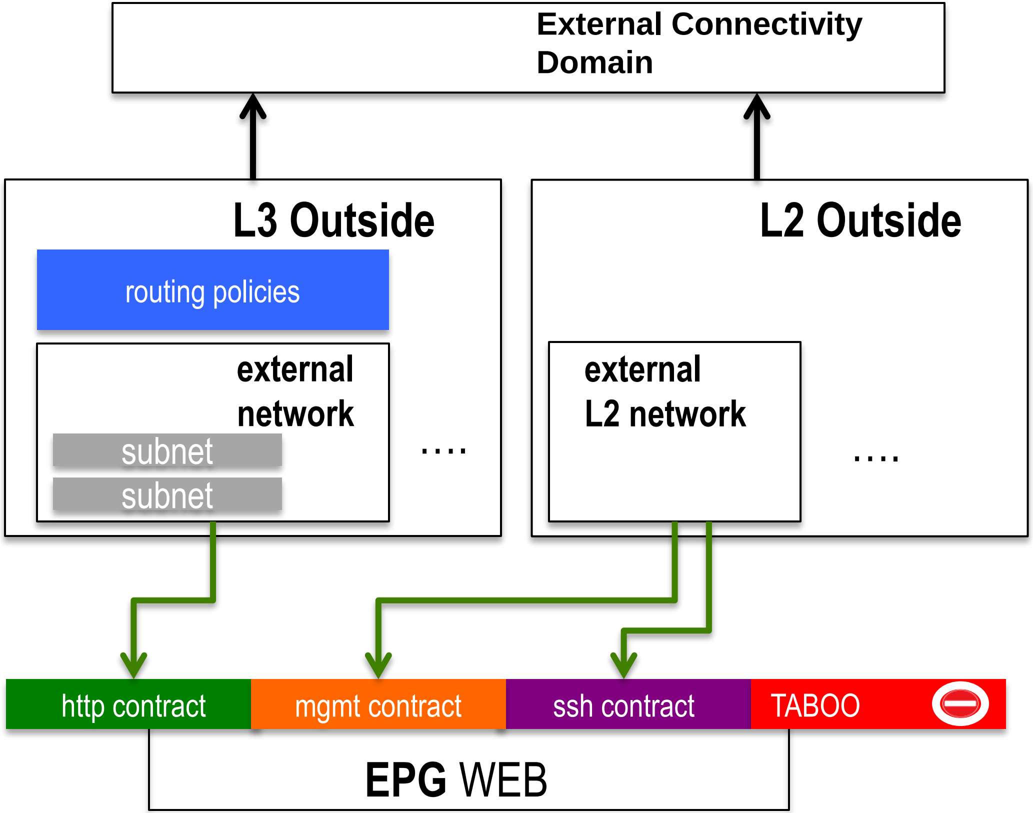
→ Explicit white-list like model for specifying rules between groups

Tenant L3, L2 Isolation



self-contained tenant definition representable as a recursive structured text document

Connecting to the Outside



Connects to a set of *border* leaf ports facing towards an external L2 or L3 datacentre interconnect

A special construct representing external connectivity

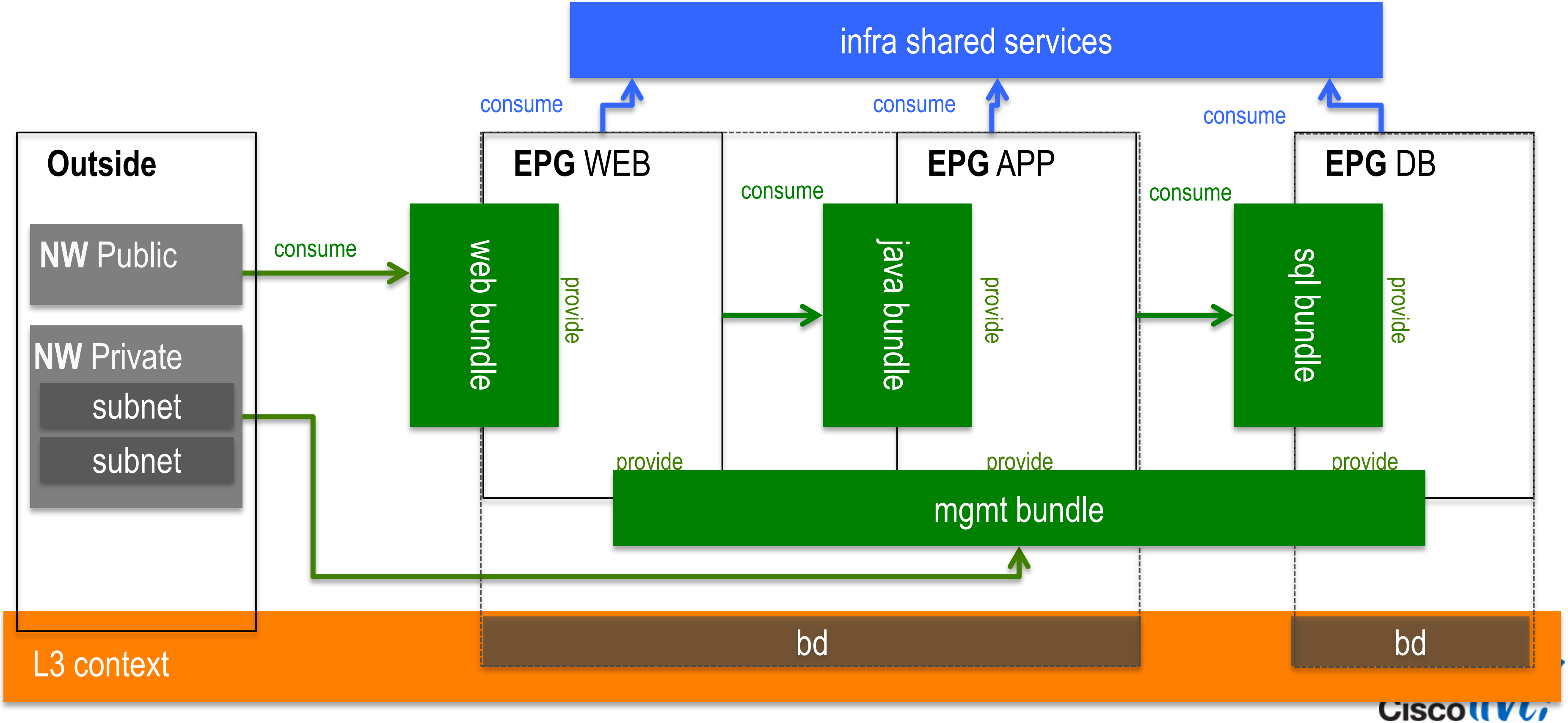
Can be **L2** or **L3**

Contains several **external networks**

An **EPG-like** construct representing external private or public network

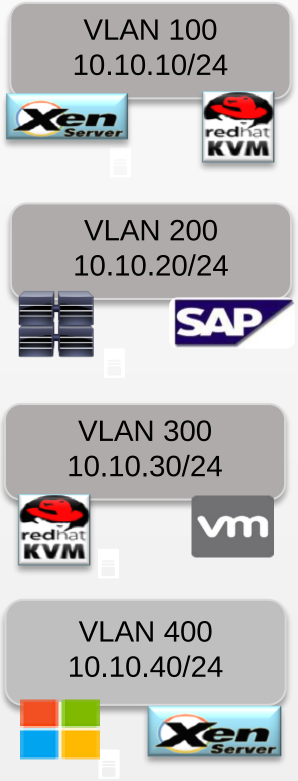
All the policy/contract concepts apply

EXAMPLE: Three-tier APP



EPGs @ ACI Bring True Network Sbstraction, as Needed

Traditional Network Model



Apps Coupled to Location

Visibility At Network or VLAN Level

ACL-based Policy Per Interface

No Address Independence or Policy Mobility

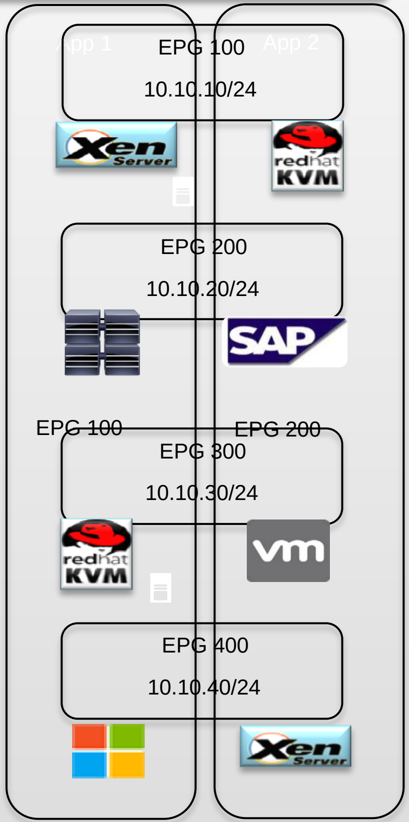
Application Centric Infrastructure

Apps Decoupled from Location

Visibility At App or Group Level

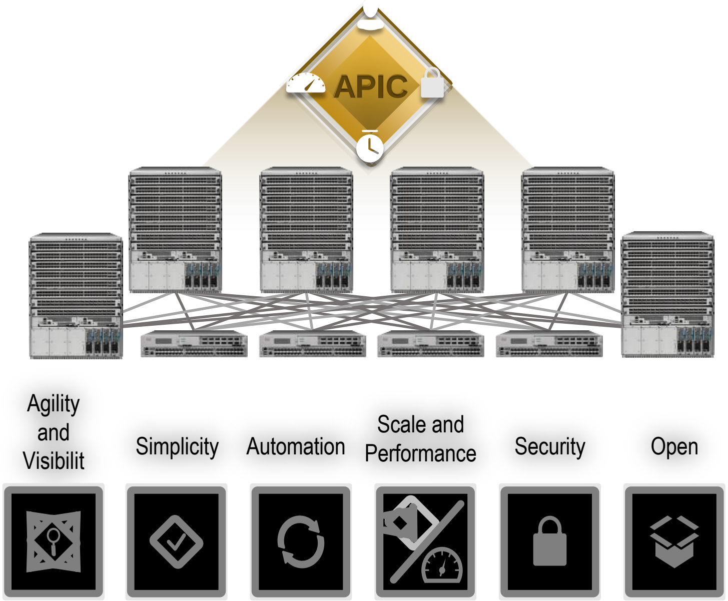
Policy Between Groups

Complete Address Independence & Policy Mobility



Agenda – Application Centric Infrastructure

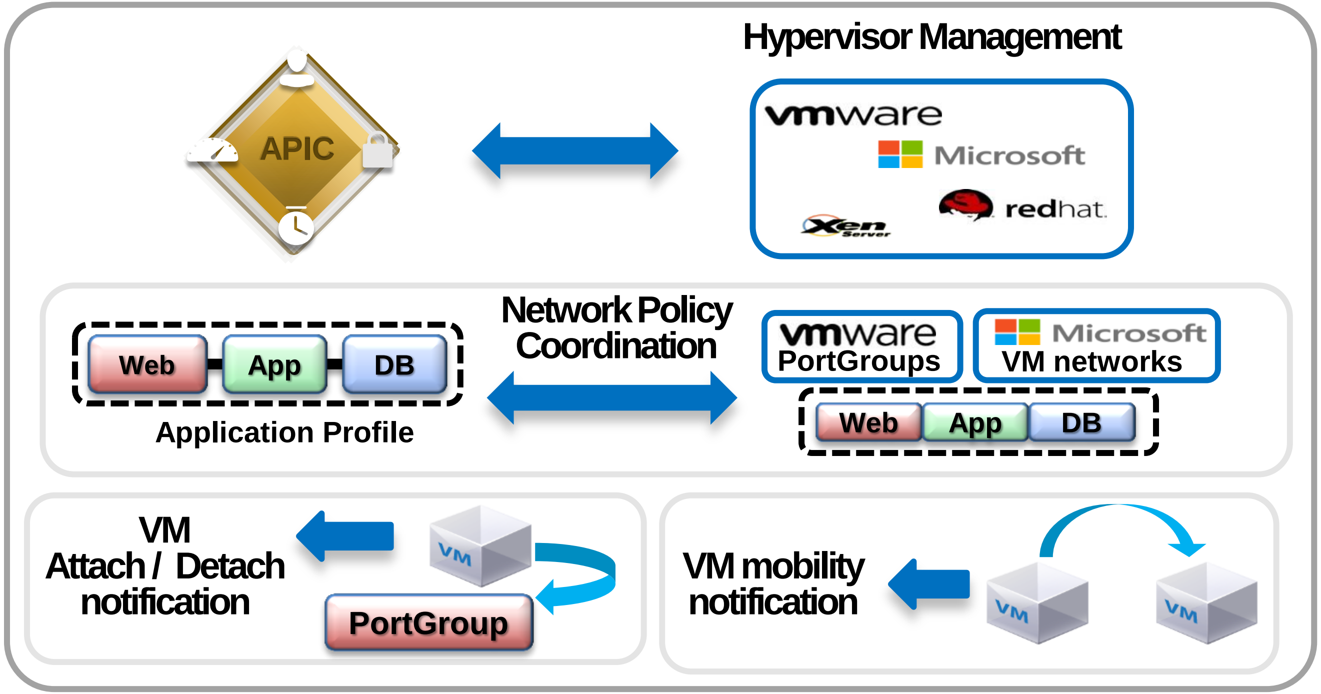
- What is ACI - Concepts and Principles
  - Why, What & How
- Foundations of ACI
  - ACI Fabric
  - Nexus 9000
  - ACI Policy Model
  - Hypervisor Integration, VMware, MSFT and KVM
  - Integration and Automation of L4-7 Services
  - APIC (The Controller)
- Integration, Migration and Co-Existence with Existing Infrastructure
- Open Standards, Open Source, Open API’s



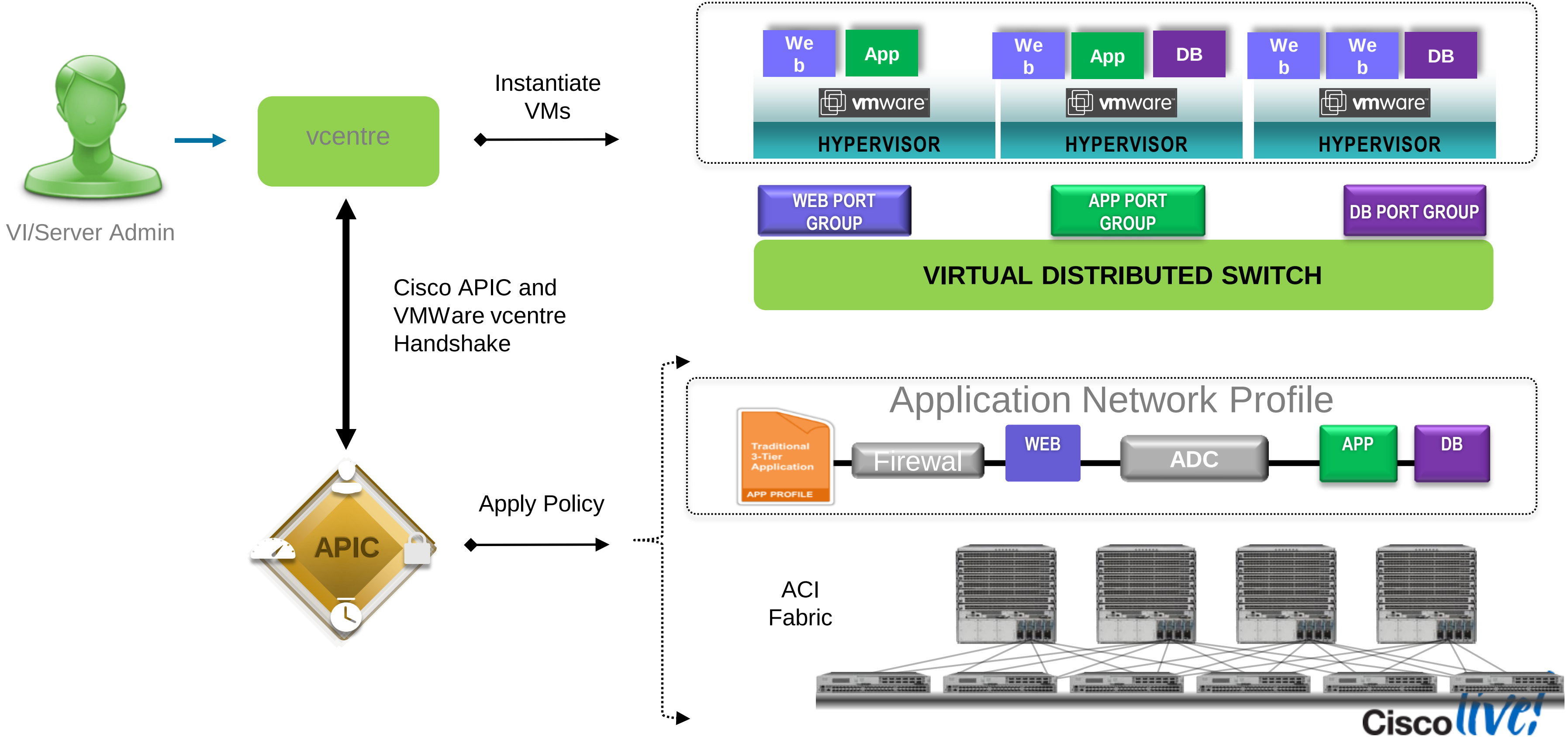
Policy Coordination with VM Managers

Leveraging the Native vSwitch

- Unified point of Data centre network automation and management for 'virtual' and physical
- Network Policies coordination with virtualisation managers
- Automatic virtual end point detection and policy placement
- Multi-Hypervisor capable



Cisco ACI And VMware Integration

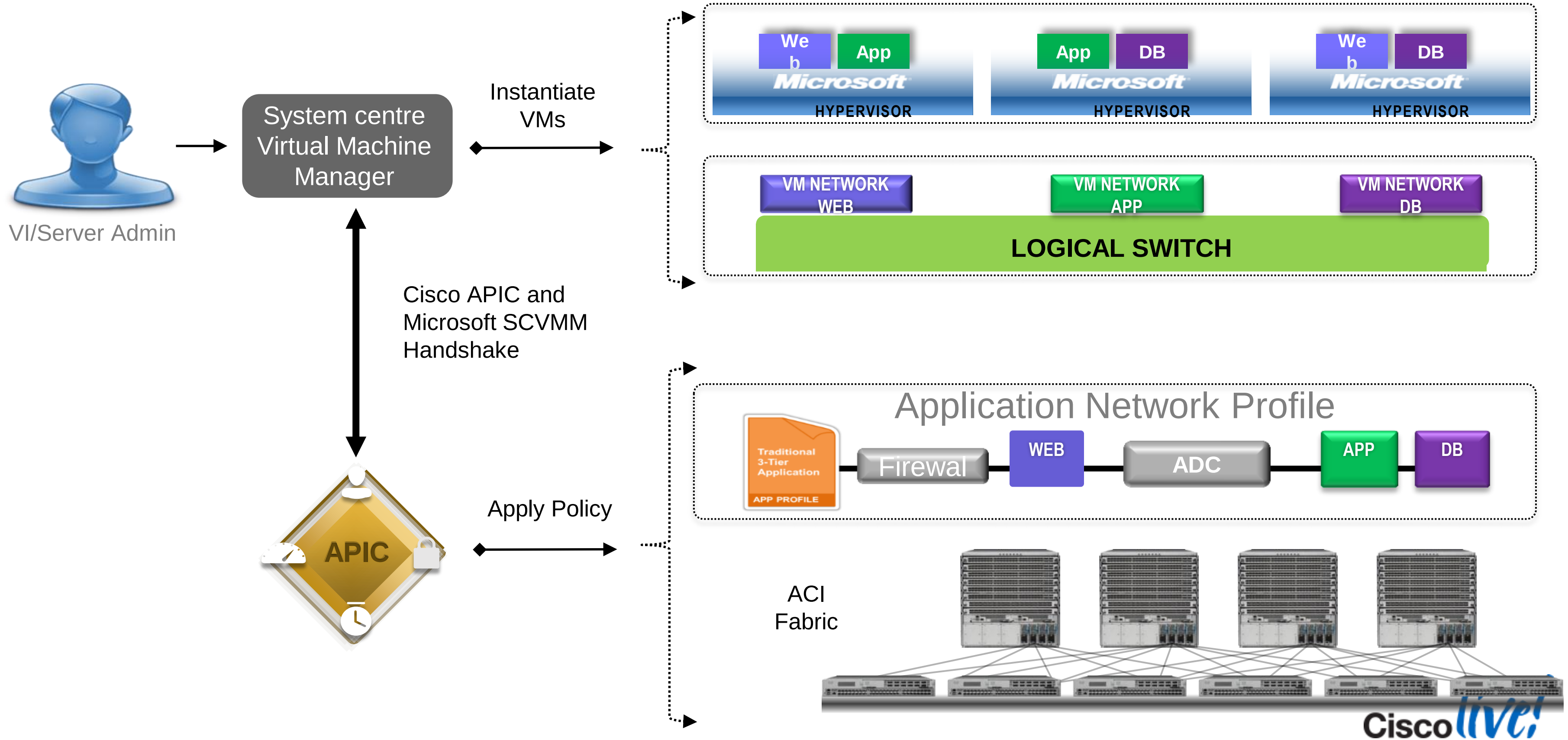


Instantiate VMs

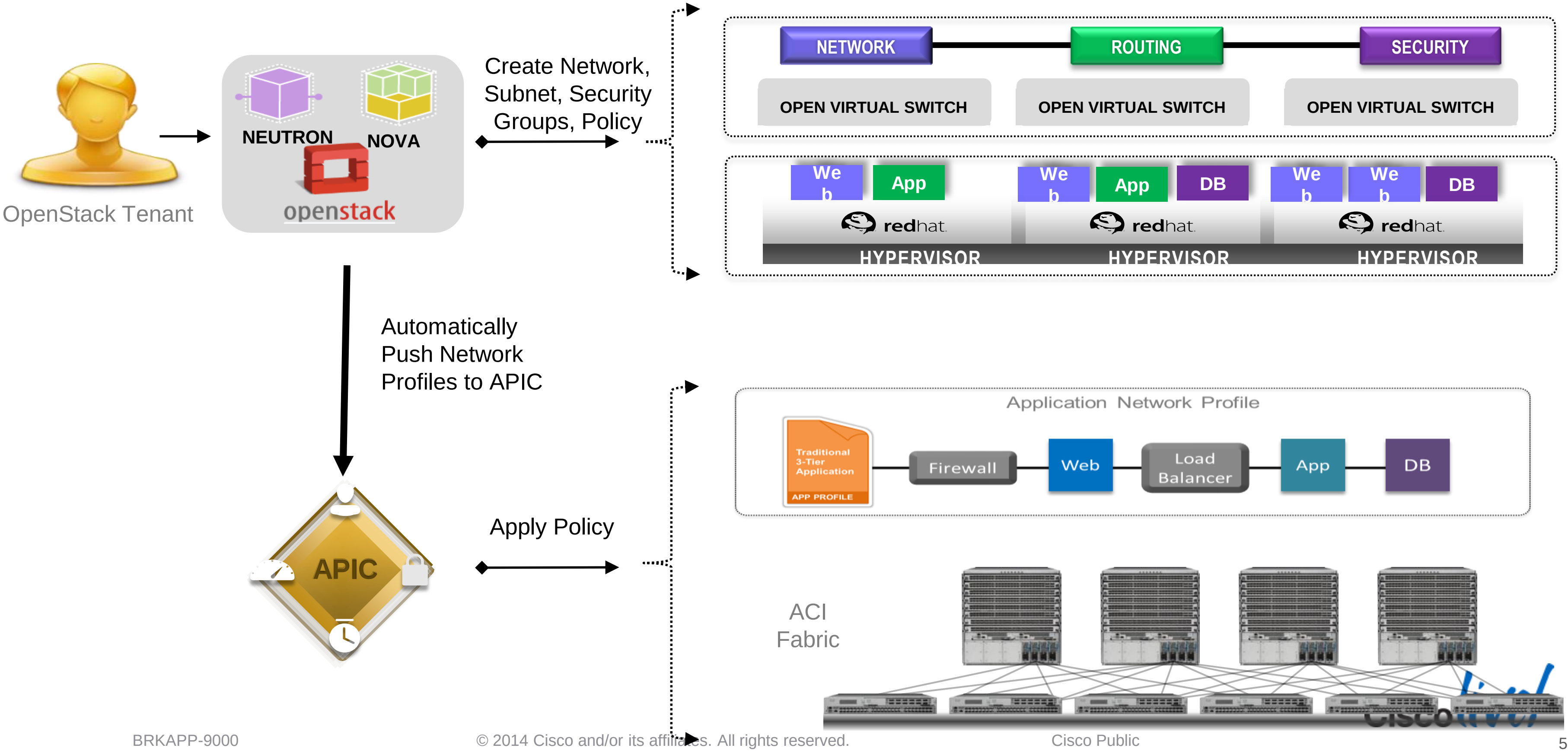
Cisco APIC and VMWare vcentre Handshake

Apply Policy

Cisco ACI And Microsoft Integration

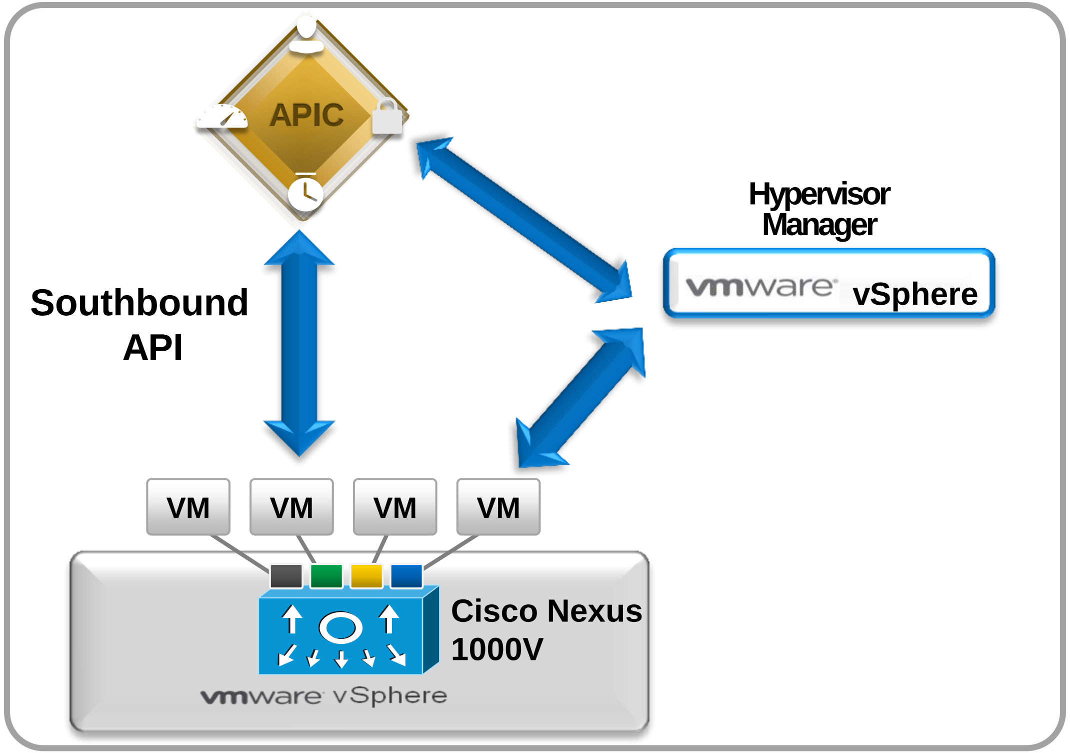


Cisco ACI And RHEL OS Integration



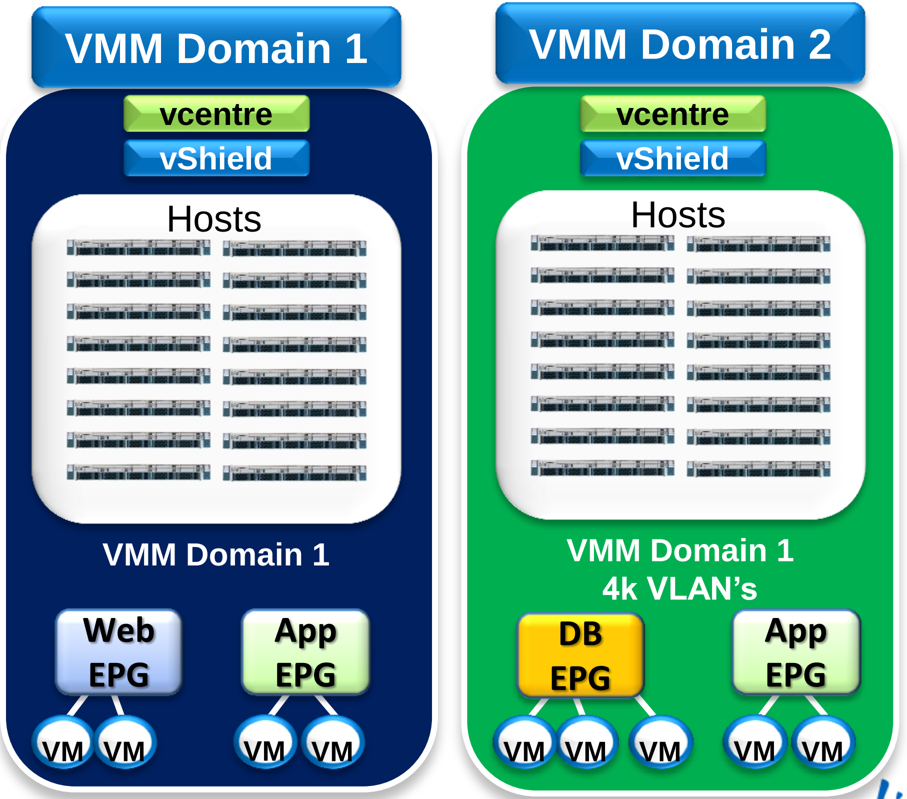
Nexus 1000V Integration Overview

- Nexus 1000v VEM Supported at FCS
- Control channel in Port Channel, VPC modes
- VM attach/detach, link states notifications via control channel
- vMotion Supported
- vSphere 5.0 and above (4.1 under consideration)
- BPDU Filter/BPDU Guard
- SPAN/ERSPAN
- Port level stats collection



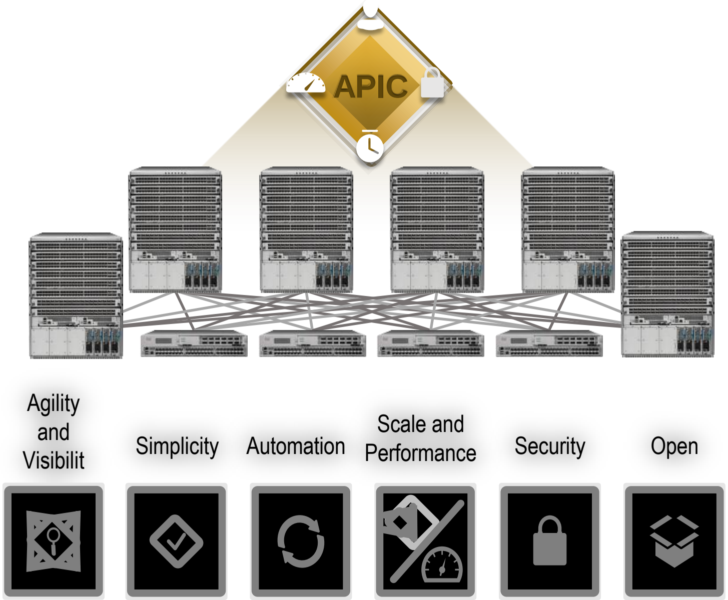
EPG Spanning Across VMM Domains

- ACI provides a number of mechanisms for addressing mobility scope and scaling
- Performs overlay offload (maintain VLAN configuration on vSwitch)
- Stretch subnets and application end points (EPGs) across VMM Domains
- EPG's can take different network identities across VMM Domain
- Applications can be deployed across VMM Domains
- Note: VM Mobility is not allowed between VMM Domain due to vcentre/SCVMM limitation



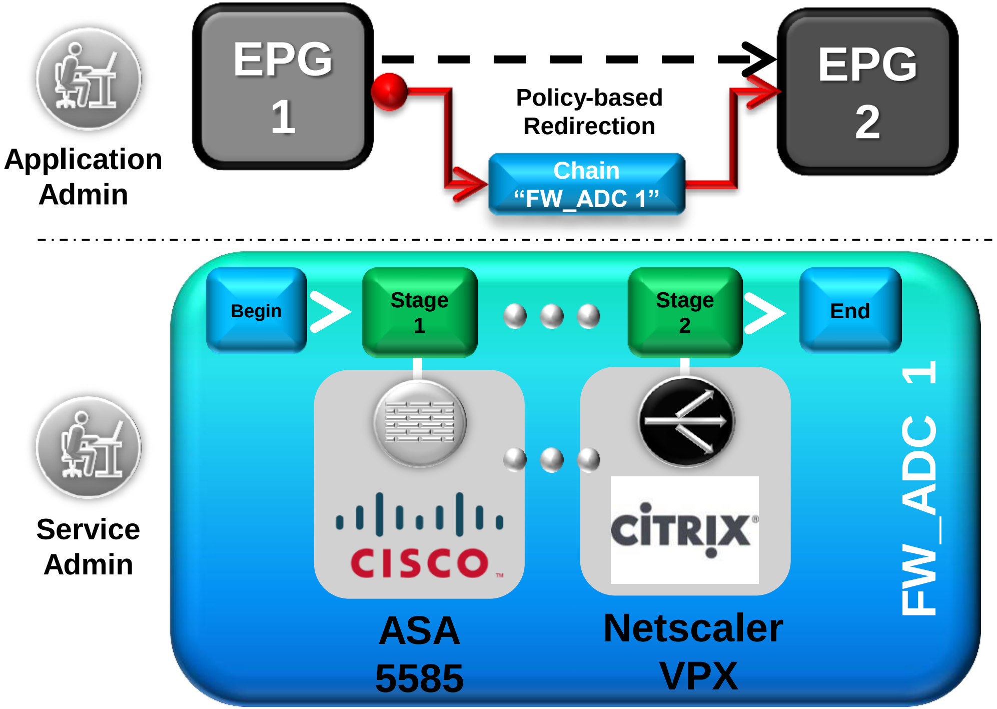
Agenda – Application Centric Infrastructure

- What is ACI - Concepts and Principles
  - Why, What & How
- Foundations of ACI
  - ACI Fabric
  - Nexus 9000
  - ACI Policy Model
  - Hypervisor Integration, VMWare, MSFT and KVM
  - Integration and Automation of L4-7 Services
  - APIC (The Controller)
- Integration, Migration and Co-Existence with Existing Infrastructure
- Open Standards, Open Source, Open API's



ACI Service Redirection via Policy

- Automated and scalable L4-L7 service insertion
- Packet match on a redirection rule sends the packet into a services graph.
- Service Graph can be one or more service nodes pre-defined in a series.
- Service graph simplifies and scales service operations



Service Graphs - Extensibility of the Data Path

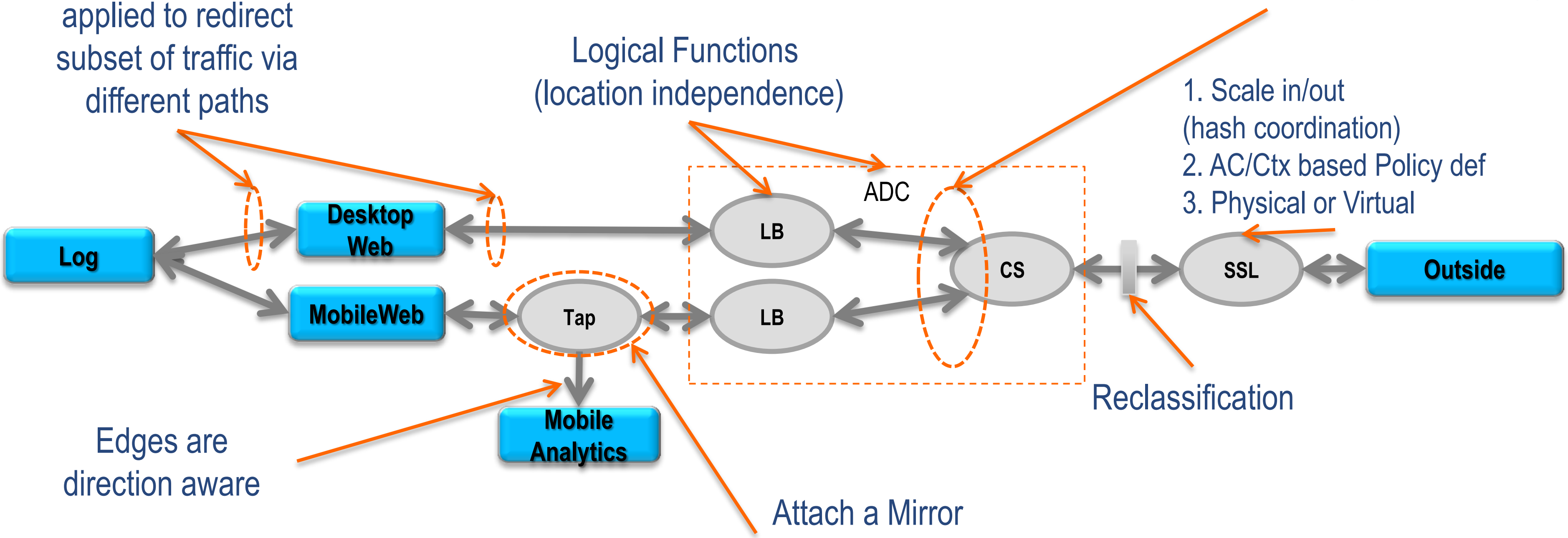
Insertion of NFV elements in the Data Path

L4 Filters can be applied to redirect subset of traffic via different paths

Logical Functions (location independence)

Split/Join chain based on pkt/flow/transaction Context (eg HTTP hdrs)

1. Scale in/out (hash coordination)
2. AC/Ctx based Policy def
3. Physical or Virtual



Edges are direction aware

Reclassification

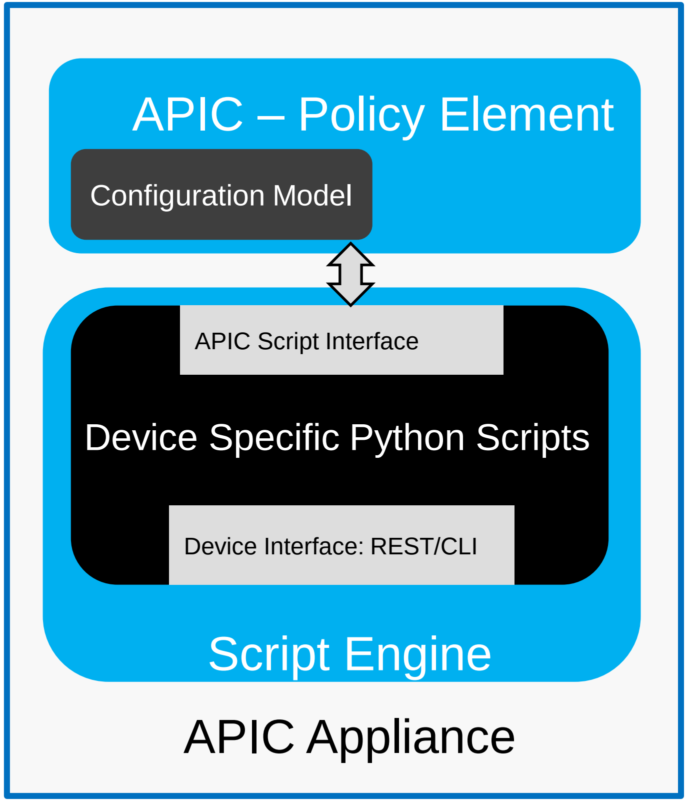
Attach a Mirror (HW mirroring)

ACI Device Package

Automation of the Appliances

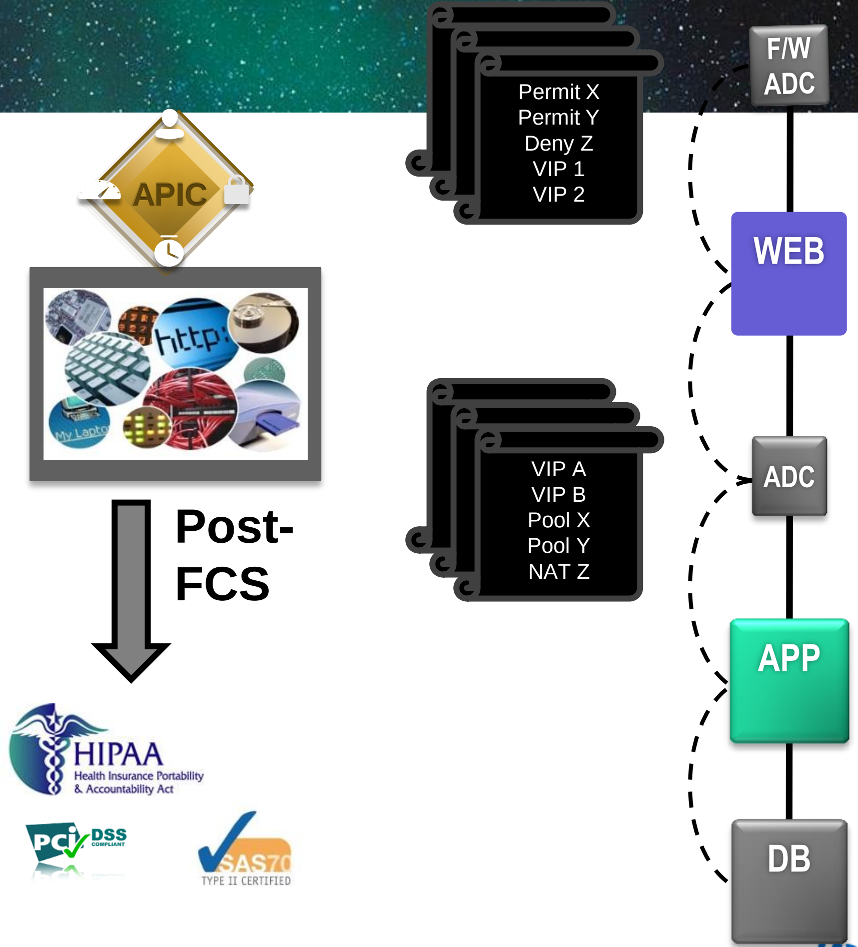
ACI SERVICE AUTOMATION ARCHITECTURE

- Defines services appliances
- Lists service functions offered by the services appliance
- Provides scripts for driving service configuration
- Plan is to open the API so that anyone can create a device package and have a community similar to Puppet manifests or Chef recipes



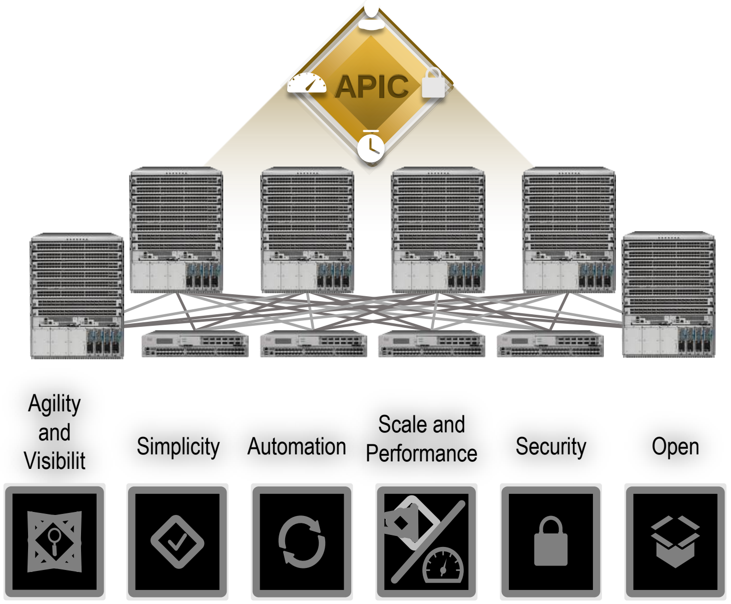
Audits and Compliance

- Application policy and state is stored within the APIC as metadata
- Contracts specify service insertion between EPGs
- Services configuration metadata associated with the contract and application profile is available in APIC
- API can be used to pull the metadata and create a compliance report
- Future goal is to have compliance reports automatically available (i.e. PCI, SOX, SAS70, etc.)



Agenda – Application Centric Infrastructure

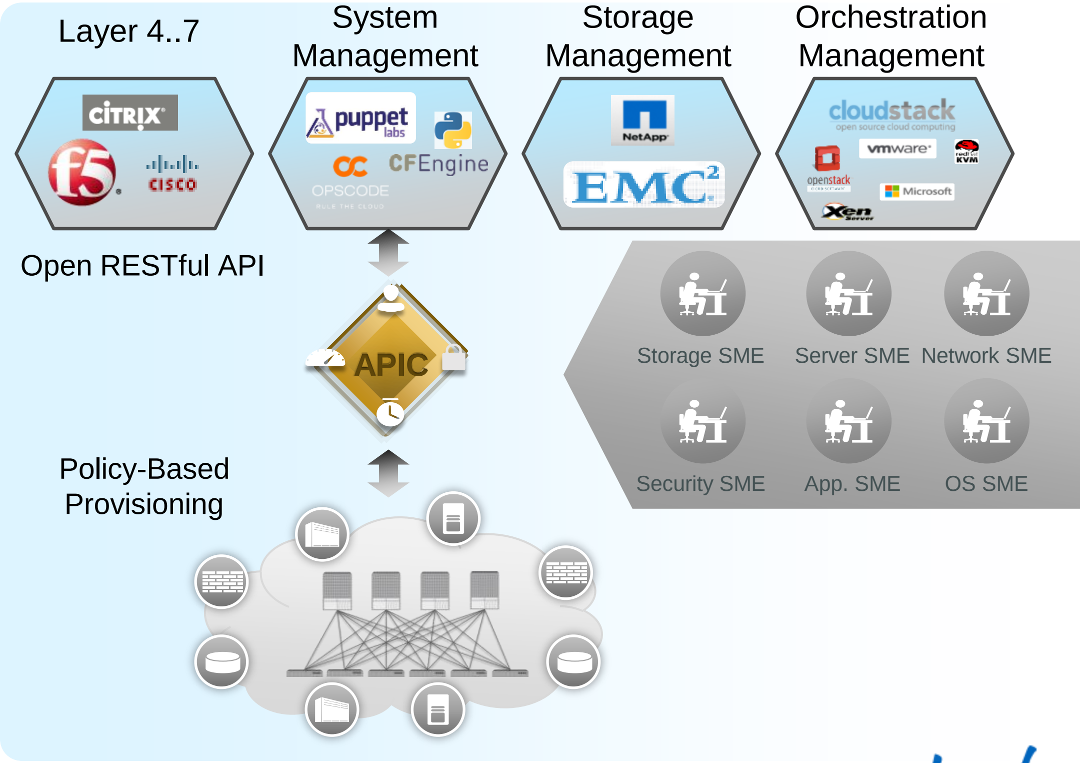
- What is ACI - Concepts and Principles
  - Why, What & How
- Foundations of ACI
  - ACI Fabric
  - Nexus 9000
  - ACI Policy Model
  - Hypervisor Integration, VMWare, MSFT and KVM
  - Integration and Automation of L4-7 Services
  - APIC (The Controller)
- Integration, Migration and Co-Existence with Existing Infrastructure
- Open Standards, Open Source, Open API’s



Application Policy Infrastructure Controller

Centralised Automation and Fabric Management

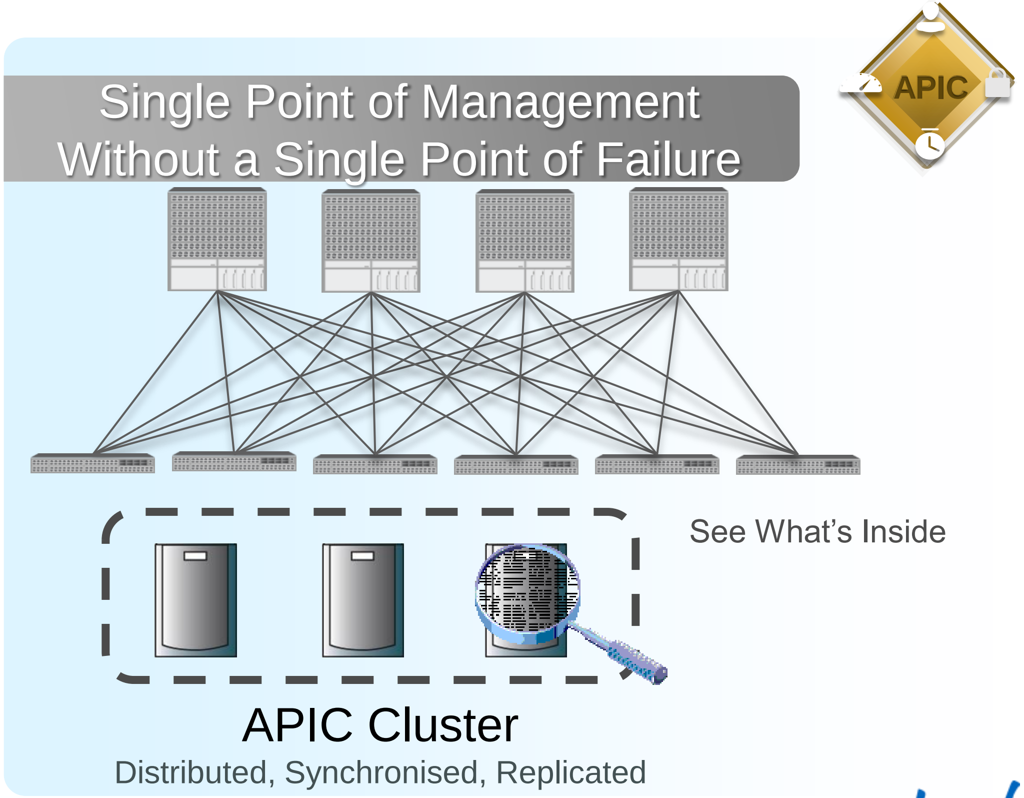
- Unified point of data centre network automation and management:
  - Data Model based declarative provisioning
  - Application, Topology Monitoring, & Troubleshooting
  - 3rd party Integration (L4-L7 Services, Storage, Compute, WAN, …)
  - Image Management (Spine / Leaf)
  - Fabric Inventory
- Single APIC cluster supports one million+ end points, 200,000+ ports, 64,000+ tenants
- Centralised Access to 'all' Fabric information - GUI, CLI and RESTful API's
- Extensible to compute and storage management



Application Policy Infrastructure Controller

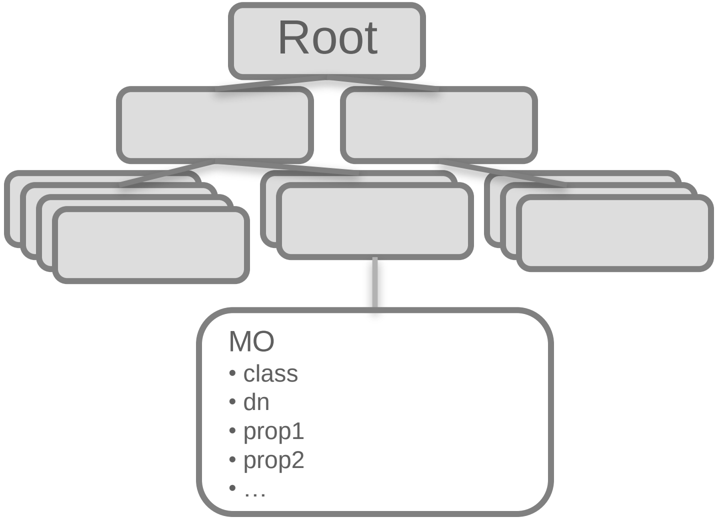
Algorithmically Sharded Cluster

- Applications fully use clustered and replicated controller (N+1, N+2, etc.)
- Any node is able to service any user for any operation
- Seamless APIC node adds and deletes
- Fully automated APIC software cluster upgrade with redundancy during upgrade
- Cluster size driven by transaction rate requirements
- APIC is not in the data path


APIC
Single Point of Management
Without a Single Point of Failure
See What's Inside
APIC Cluster
Distributed, Synchronised, Replicated

APIC Data Structures

Distributed MIT Managed Objects



Root

MO
- class
- dn
- prop1
- prop2
- …

Full unified description of entities.

No artificial separation of configuration, state, runtime data.

Everything is an object

Objects are hierarchically organised

Distributed Managed Information Tree (dMIT) contains comprehensive system information
- discovered components
- system configuration
- operational status including statistics and faults

Class identifies object type

Card, Port, Path, EPG…

Class Inheritance

Access port is a subclass of port.

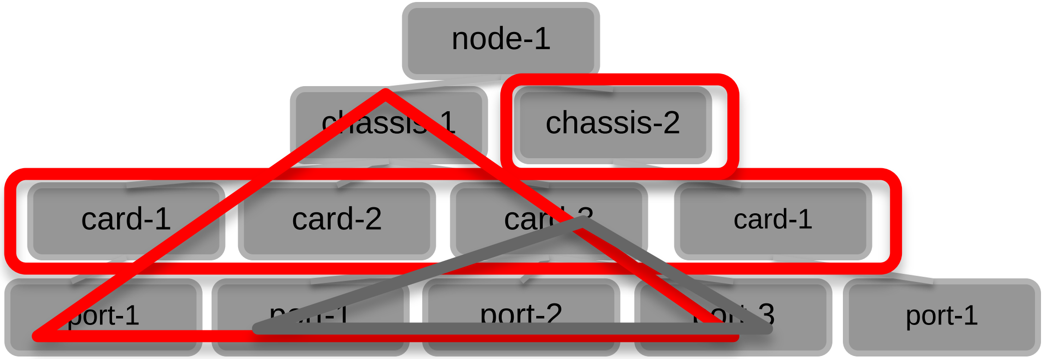
A leaf node is a subclass of fabric node.

Set of attributes

- identity
- states
- descriptions
- references
- lifecycle

APIC Data Structures

Queries



Returned are a set of objects or sub-trees

→ Option to return entire or partial sub-tree per each object in resolution scope

Role and Domain based Access Control

→ Privileges define what type of objects can be accessed
→ Domains identifies what sub-trees

Class-level queries

- Find all members of this object class that match given criteria
  - Class or Superclass
  - Property filter

Object-level queries

- Find a managed object by DN
  - Distinguished name

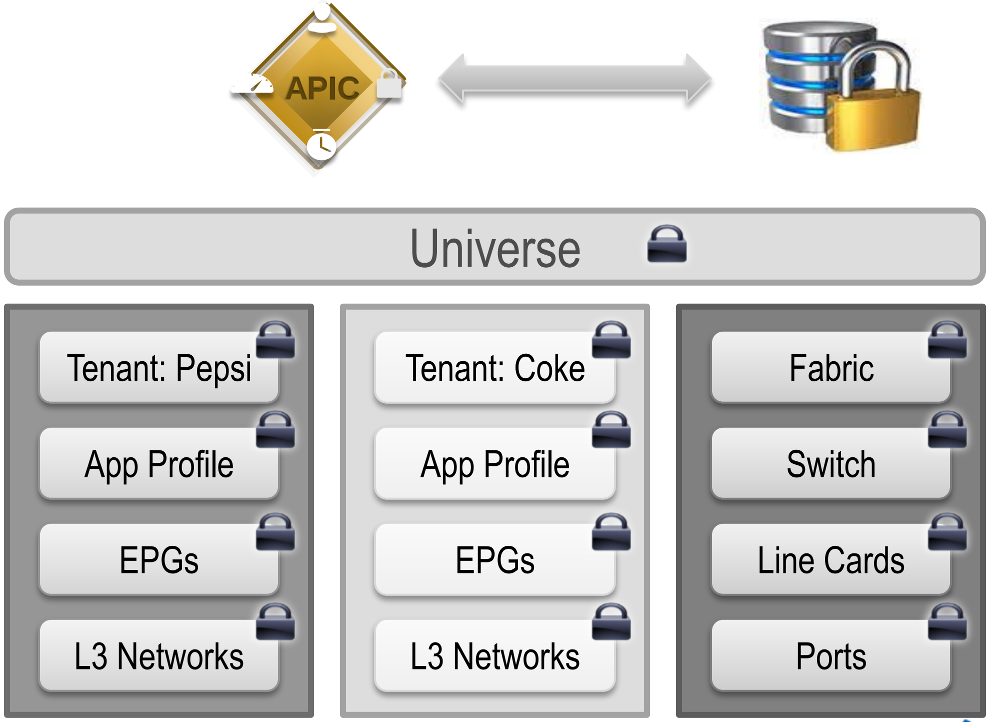
Tree-level queries

- Sub-tree-Scope: On a given sub-tree, find all members of this object class that match given criteria
  - Distinguished name
  - Class or Superclass
  - Property filter

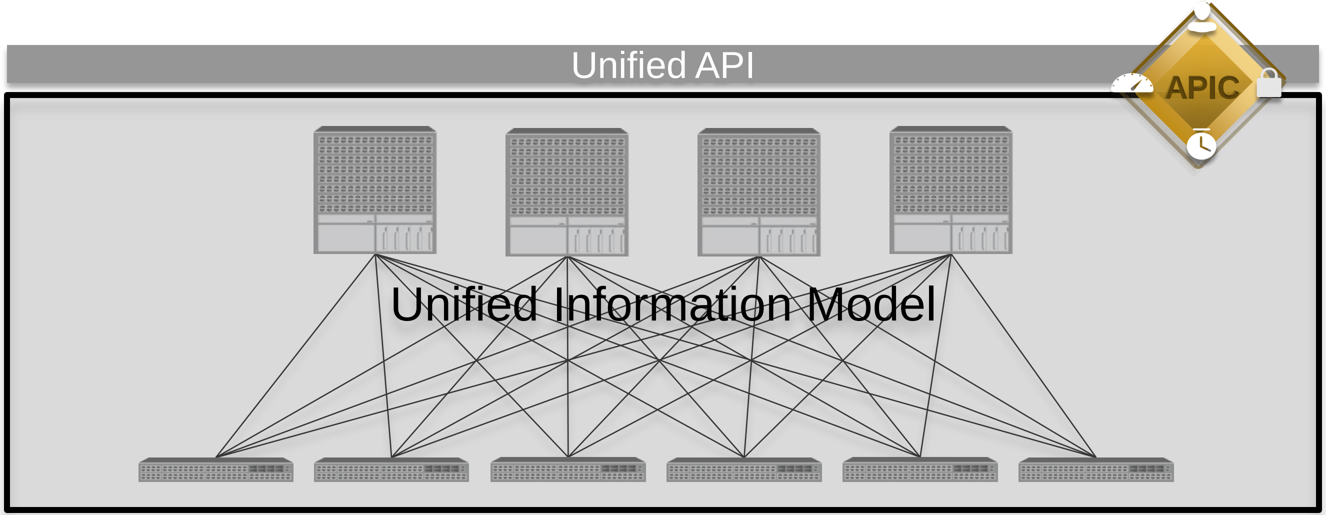
System Access

Authentication, Authorisation, RBAC

- Local & External AAA (TACACS+, RADIUS, LDAP) Authentication & Authorisation
- RBAC to control READ and WRITE for ALL Managed Objects
- RBAC to enforce Fabric Admin and per-Tenant Admin separation



Fully Exposed System, Fully Programmable



RESTFul over HTTP(s)

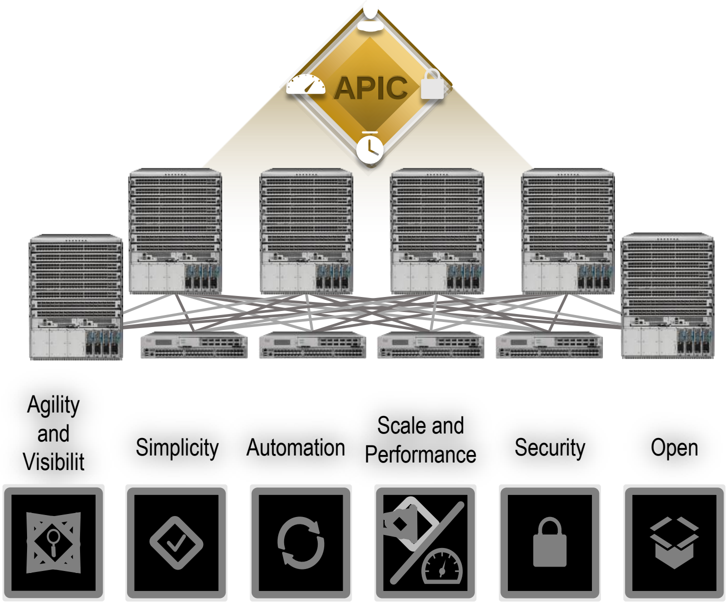
- JSON + XML
- ***Unified***: automatically delegates request to corresponding components
- ***Transactional***
- Single Management Entity yet fully independent components

Object Oriented

- ***Comprehensive*** access to underlying information model
- Consistent object naming directly mapped to URL
- Supports object, sub-tree and class-level queries

Agenda – Application Centric Infrastructure

- What is ACI - Concepts and Principles
  - Why, What & How
- Foundations of ACI
  - ACI Fabric
  - Nexus 9000
  - ACI Policy Model
  - Hypervisor Integration, VMWare, MSFT and KVM
  - Integration and Automation of L4-7 Services
  - APIC (The Controller)
- Integration, Migration and Co-Existence with Existing Infrastructure
- Open Standards, Open Source, Open API's



Two Big Questions

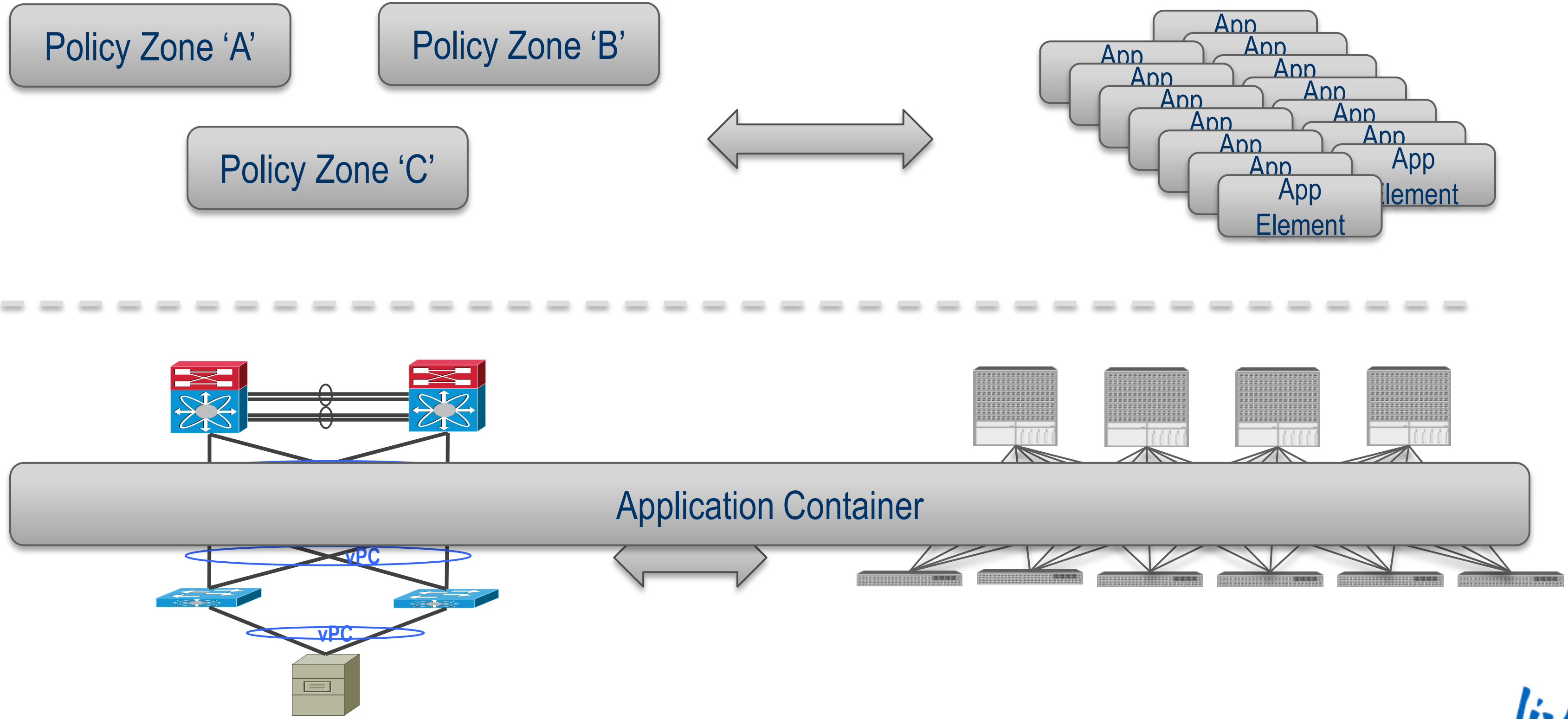
“Do I need to have a complete knowledge of my current application environment to fully use, benefit or leverage Cisco ACI ?”

Do I need to replace all of my existing infrastructure to begin leveraging ACI?

ABSOLUTELY NOT !!!

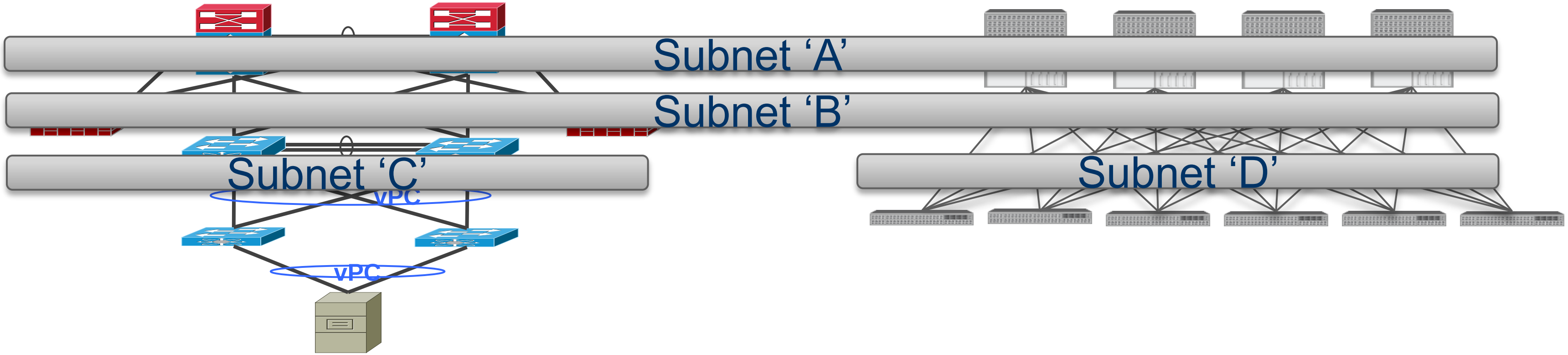
Let’s see WHY and HOW …

Transitioning Business Logic Independently from Infrastructure Changes

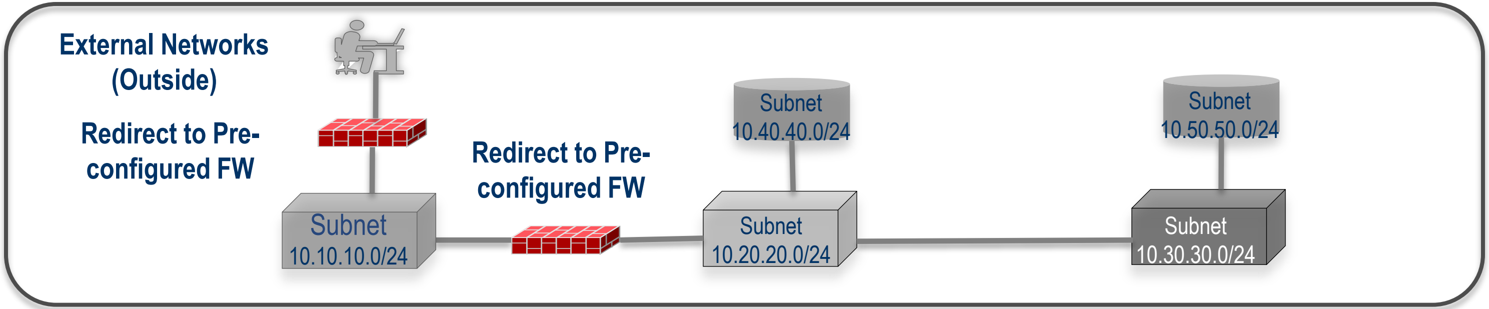


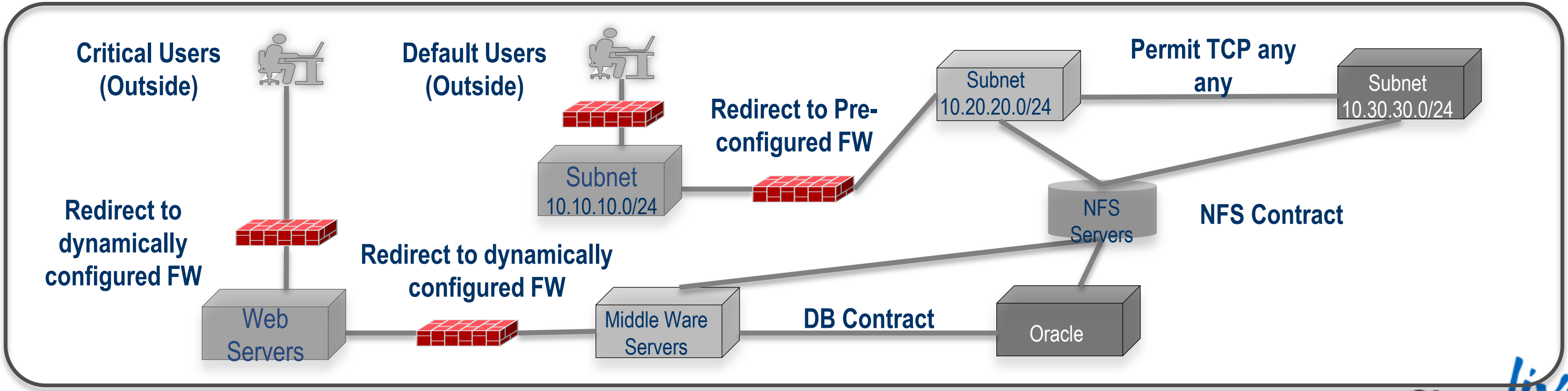
Transitioning Business Logic Independently from Infrastructure Changes

- Layer 2 and Layer 3 interoperation between ACI Fabric and Existing Data centre builds
- Layer 3 interconnect via standard routing interfaces, OSPF, MP-BGP, EIGRP, …
- Layer 2 interconnect via standard STP or via VXLAN overlays



Transitioning Business Logic Independently from Infrastructure Changes



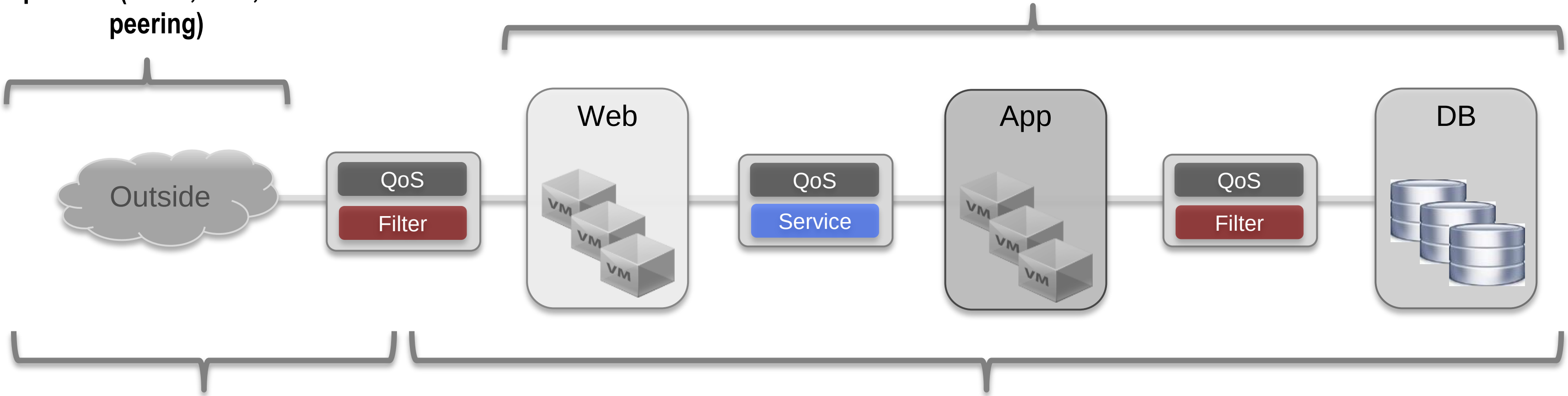


Fabric Infrastructure

Policy and the Network

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

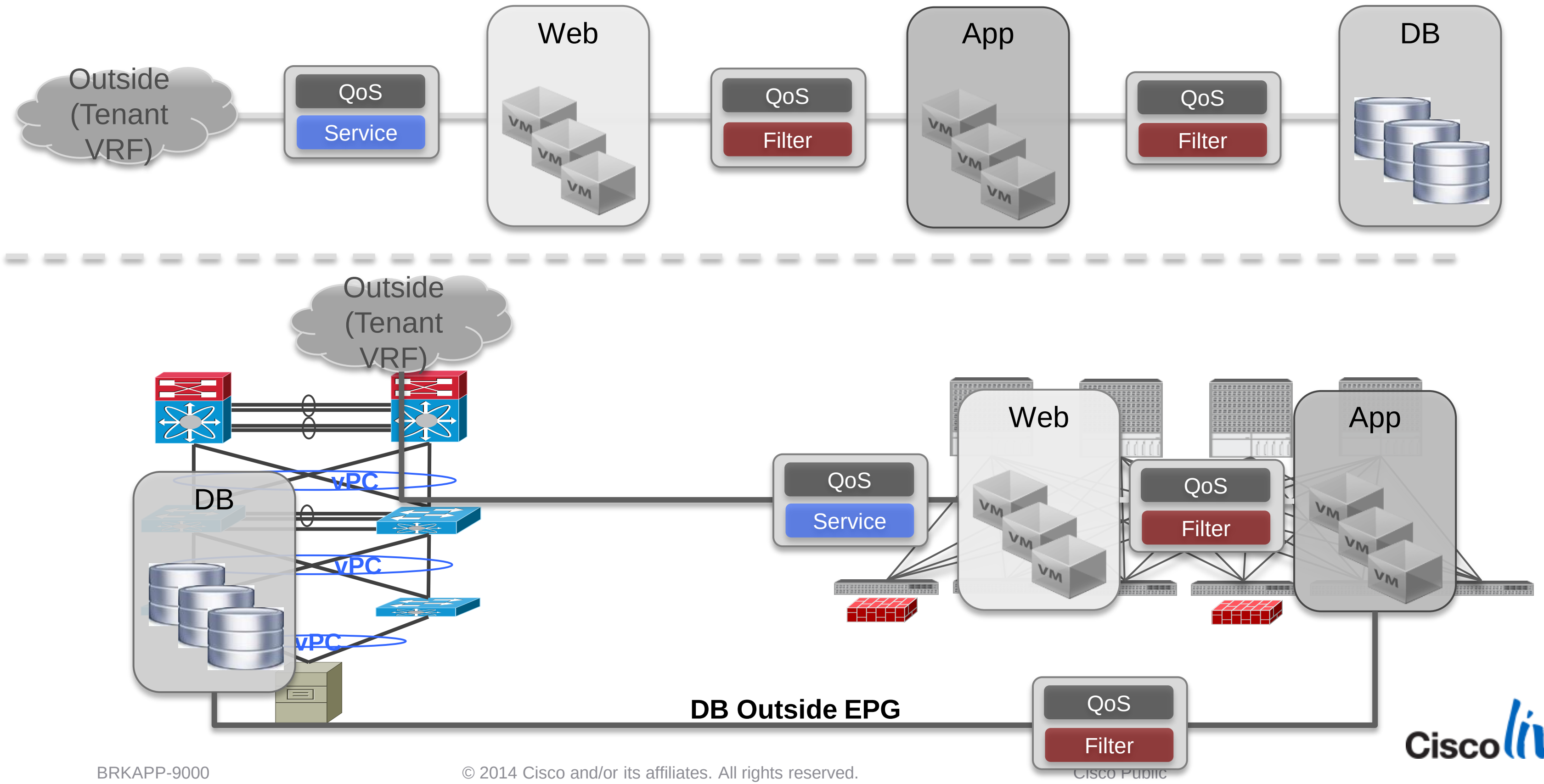
Forwarding Policy for 'inside' EPG's defined by associated Bridge Domain network policies



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

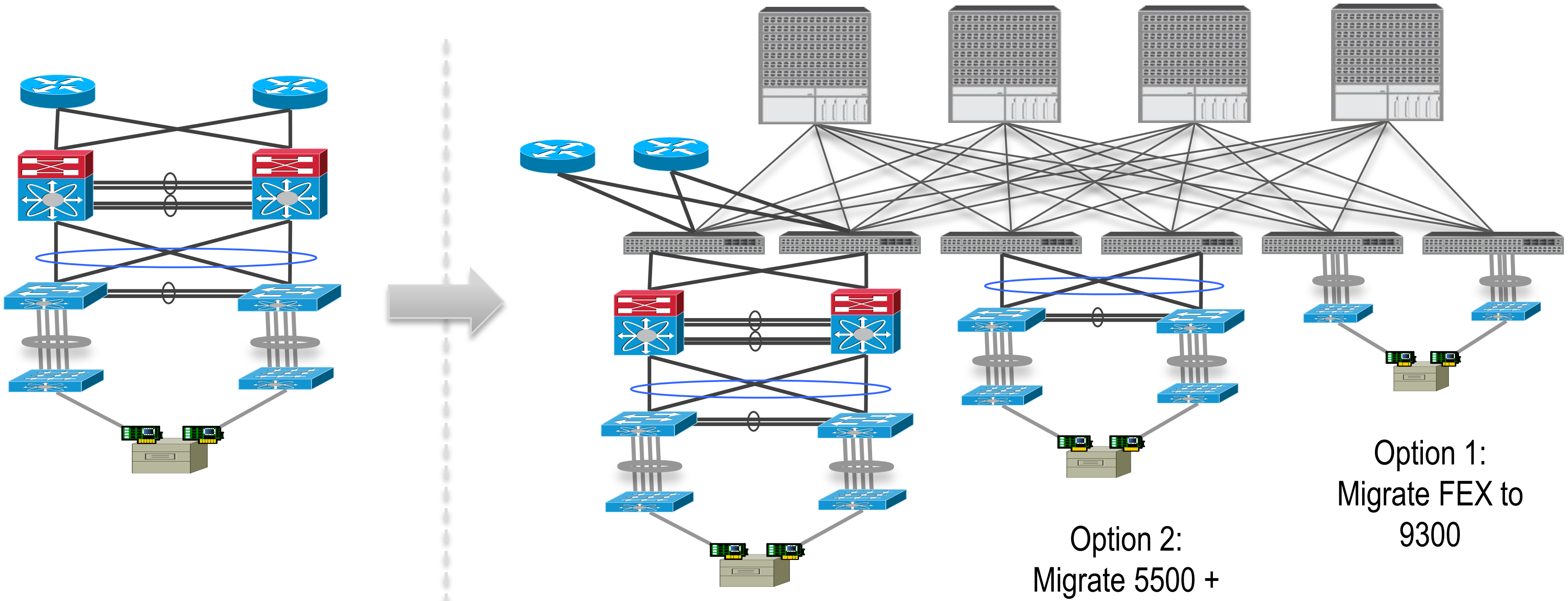
Location for Endpoints that are 'Inside' the Fabric are found via the Proxy Mapping DB (Host Level Granularity)

Transitioning Business Logic Independently from Infrastructure Changes



Integration of Existing DC Network Assets

Migration ‘and/or’ Interconnection of Existing Nexus



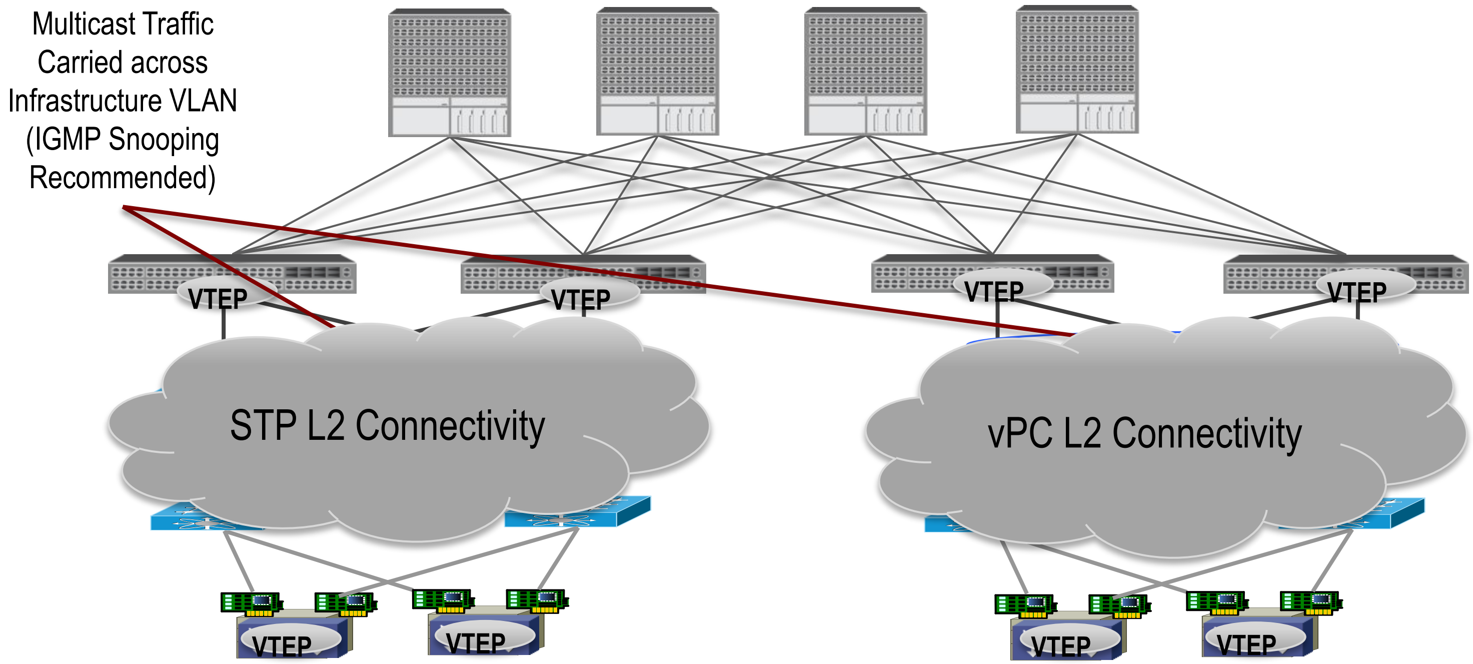
Option 3: Interconnect existing POD to Fabric

Option 2: Migrate 5500 + FEX to 9300

Option 1: Migrate FEX to 9300

Integration of Existing DC Network Assets

Integrating the Hypervisor VTEP



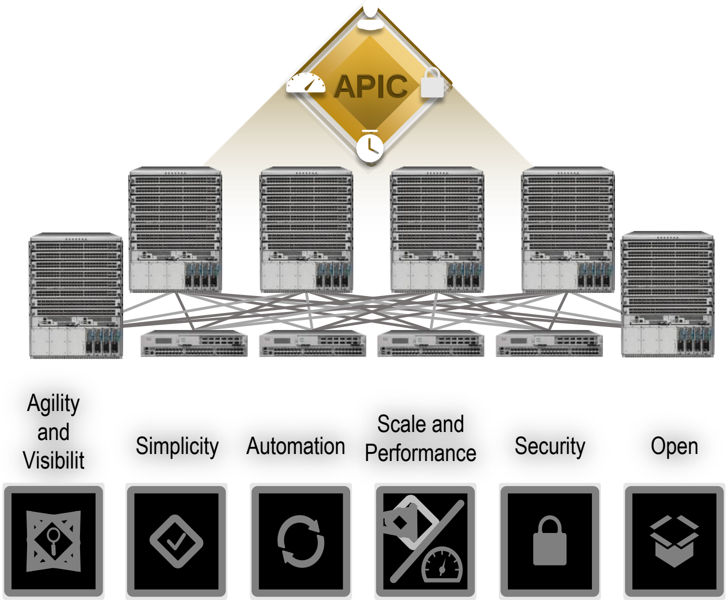
Multicast Traffic Carried across Infrastructure VLAN (IGMP Snooping Recommended)
VTEP
VTEP
VTEP
VTEP
STP L2 Connectivity
vPC L2 Connectivity
VTEP
VTEP
VTEP
VTEP

VXLAN attached
N1KV VEM or ESX
DVS

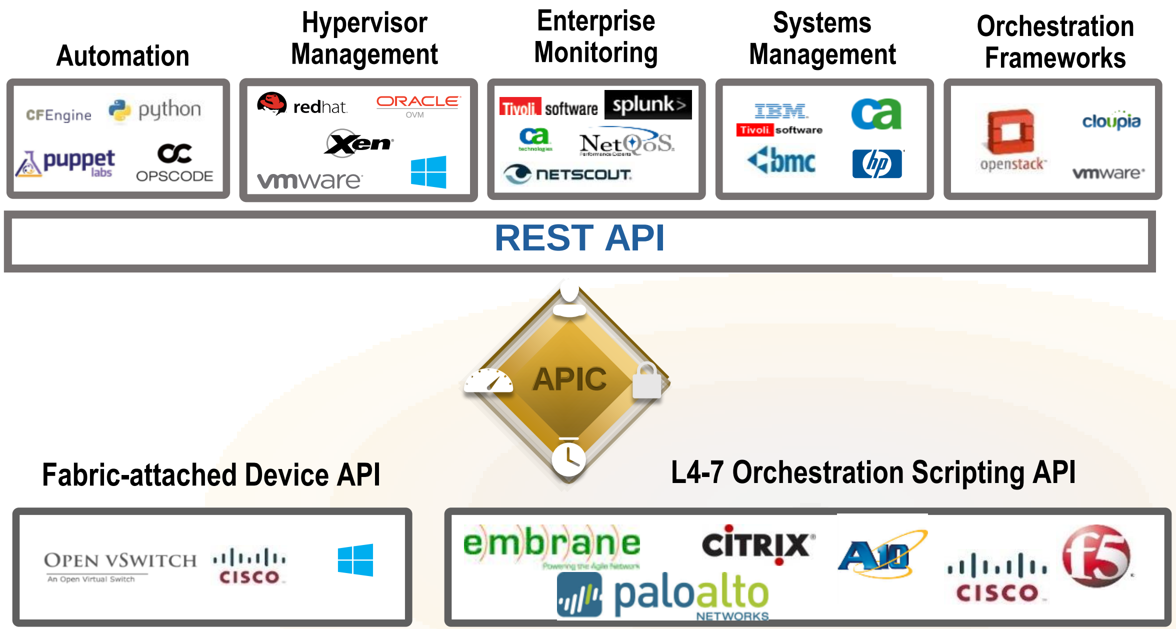
VXLAN attached
N1KV VEM or ESX
DVS

Agenda – Application Centric Infrastructure

- What is ACI - Concepts and Principles
  - Why, What & How
- Foundations of ACI
  - ACI Fabric
  - Nexus 9000
  - ACI Policy Model
  - Hypervisor Integration, VMWare, MSFT and KVM
  - Integration and Automation of L4-7 Services
  - APIC (The Controller)
- Integration, Migration and Co-Existence with Existing Infrastructure
- Open Standards, Open Source, Open API's



ACI Open APIs and Ecosystem



Automation

Hypervisor Management

Enterprise Monitoring

Systems Management

Orchestration Frameworks

REST API

Fabric-attached Device API

L4-7 Orchestration Scripting API

NORTHBOUND PROGRAMMABILITY LAYER

SOUTHBOUND PROGRAMMABILITY LAYER

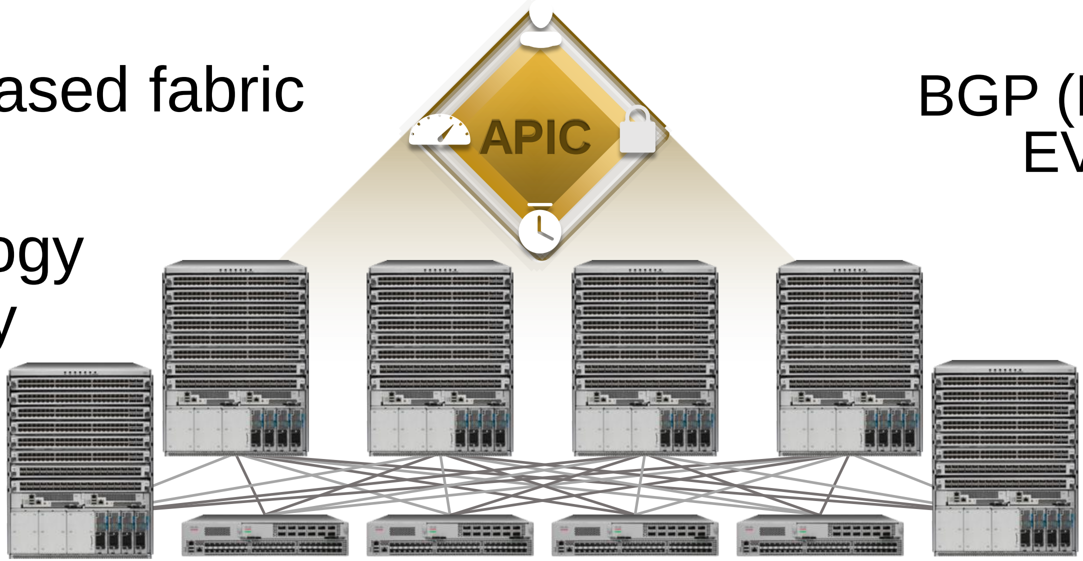
APIC SUPPORTS A RICH ECOSYSTEM BUILT AROUND OPEN NORTHBOUND AND SOUTHBOUND APIS

Standards Based Architecture

VXLAN-based fabric

BGP (MP-BGP EVPN)

IS-IS Topology Discovery



ACI Fabric Attached Device Protocol*

Network Service Header (NSH) Protocol*

OPEN POLICY MODEL EXPOSED THROUGH OSS TOOLS

Traditional 3-Tier Application
APPLICATION NETWORK PROFILE
F/W ADC
WEB
ACC
APP
DB

APP CENTRIC POLICY MODEL

Cloud Orchestration

openstack™ CLOUD SOFTWARE

- Neutron Group Policy Extensions Working Group
- Future extensions to Heat / Nova

Network Controller

OPEN DAYLIGHT

- Group Policy Northbound and Southbound API
- Yang-based API

Hypervisor / vSwitch

OPEN VSWITCH An Open Virtual Switch

- Fabric-attached Device API agent support
- Policy enforcement modules

POLICIES ARE OPEN AND WILL BE REUSABLE THROUGH A COMPLETE OPEN SOURCE STACK.

Community Code Development

- Visit us on GitHub: https://github.com/datacentre/nexus9000
- ACI and NX-OS code examples and libraries
- Open source and community developed tools by partners and 3rd party developers

github
SOCIAL CODING

IBM

redhat.

cloudstack

OPSCODE

CITRIX

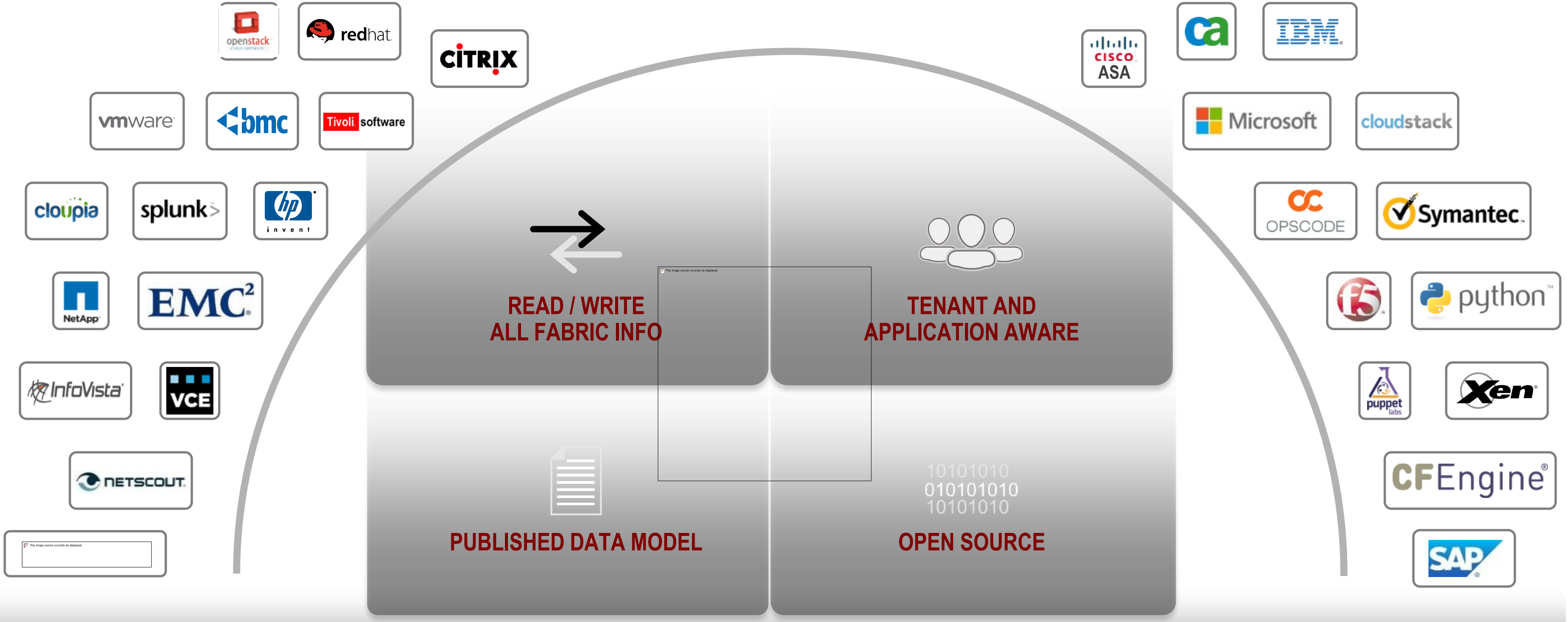
puppet labs

Xen

CFEngine

Open Ecosystem, Open APIs, Open Source

Comprehensive access to underlying information model



READ / WRITE
ALL FABRIC INFO

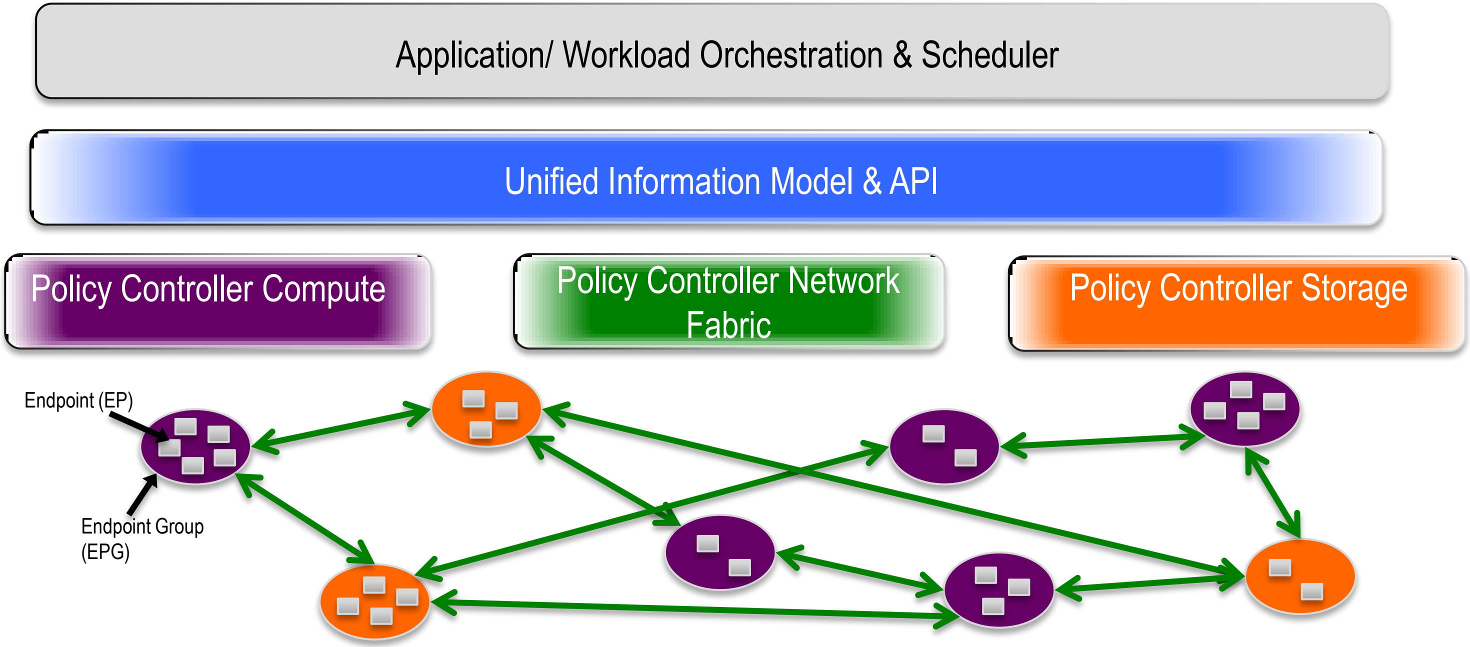
TENANT AND
APPLICATION AWARE

PUBLISHED DATA MODEL

OPEN SOURCE

Industry Standard Compliant and Certified

User Driven, Policy Based IT Infrastructure



Application/ Workload Orchestration & Scheduler

Unified Information Model & API

Policy Controller Compute

Policy Controller Network Fabric

Policy Controller Storage

Application Graph (EP, EPG, graph edges)

Application Profile = Compute Service Profile + Network Connectivity Profile + Storage Service Profile

Designed from the Ground-Up to be Application Centric

Q & A

Complete Your Online Session Evaluation

Give us your feedback and receive a Cisco Live 2014 Polo 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 www.ciscoliveaustralia.com/mobile
- Visit any Cisco Live Internet Station located throughout the venue

Polo Shirts can be collected in the World of Solutions on Friday 21 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 cisco TM