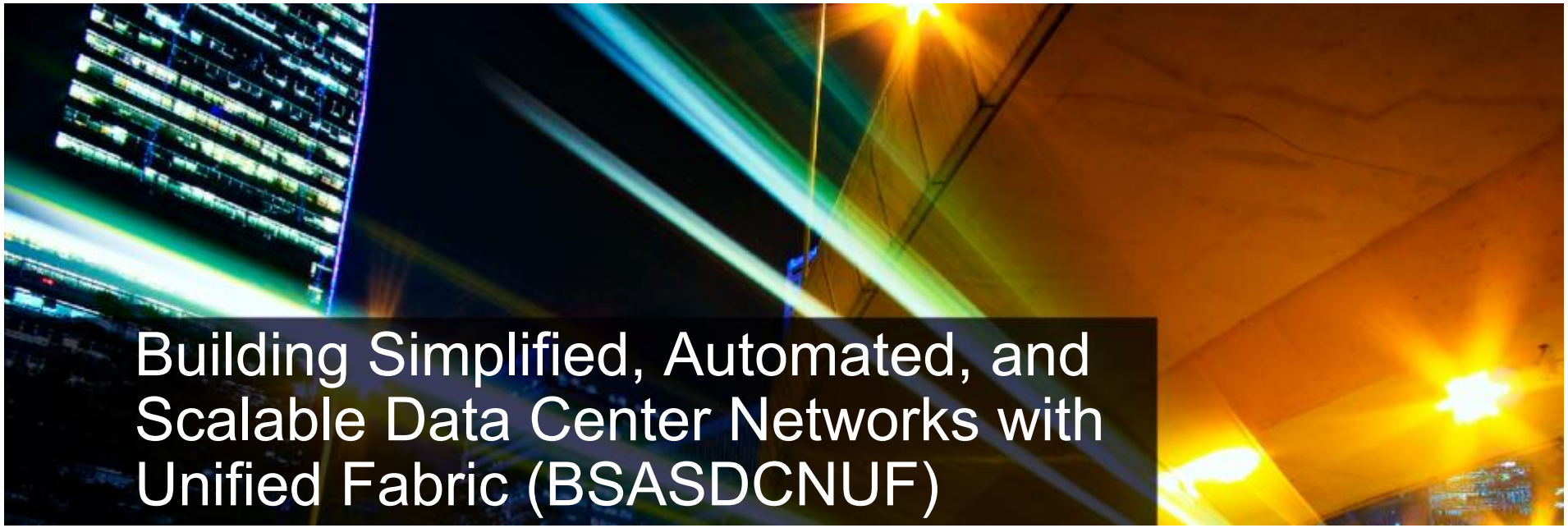


Building Simplified, Automated, and Scalable Data Center Networks with Unified Fabric (BSASDCNUF)

Packet Pushers PodCast – September 2014

Shyam Kapadia, Technical Leader, Engineering

Lukas Krattiger, Technical Marketing Engineer @CCIE21921

Who are we?

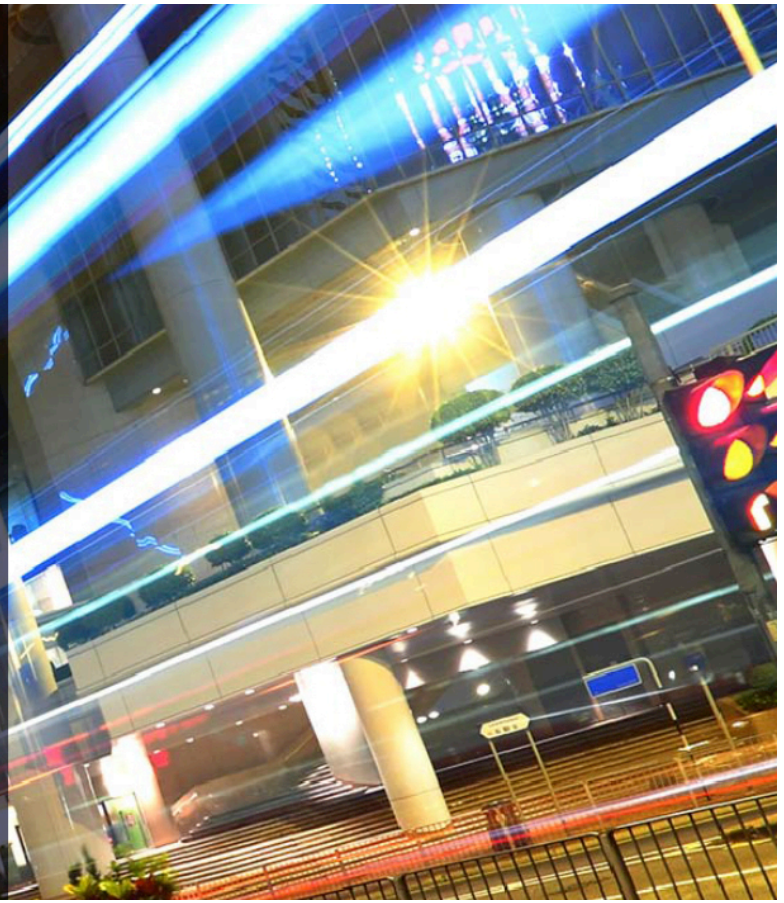
Shyam Kapadia
Technical Leader, Engineering

Lukas Krattiger @CCIE21921
Technical Marketing Engineer

© 2014 Cisco and/or its affiliates. All rights reserved.

Agenda

- **Requirements and Functions**
  - IDC Study
  - Data Center Challenges
  - Fabric Evolution
  - ACI Integration
- Building Blocks
- Optimized Network
- Virtual Fabric
- Workload Automation



© 2014 Cisco and/or its affiliates. All rights reserved.

IDC Digital Universe (2012)

Key Data Center Requirements

2012 IDC Digital Universe Study:

By 2020,

Server workloads to go to **70% Virtual** & will coexist with **Physical**

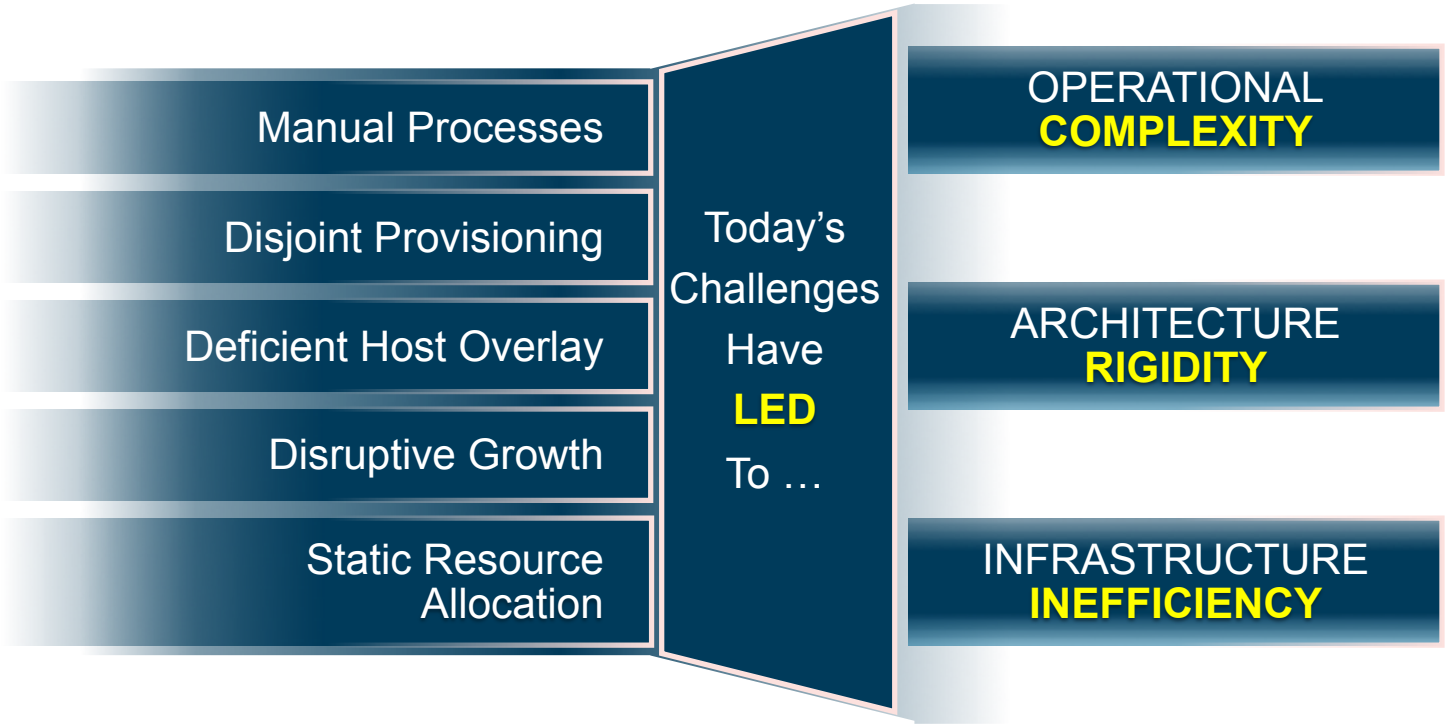
Over the next decade **the number of servers** (virtual and physical) worldwide will grow by **10 times.**

The **amount of information managed** by enterprise datacenters will grow by **14 times.**

Meanwhile, the **number of IT professionals** in the world will grow by less than **1.5 times.**

© 2014 Cisco and/or its affiliates. All rights reserved.

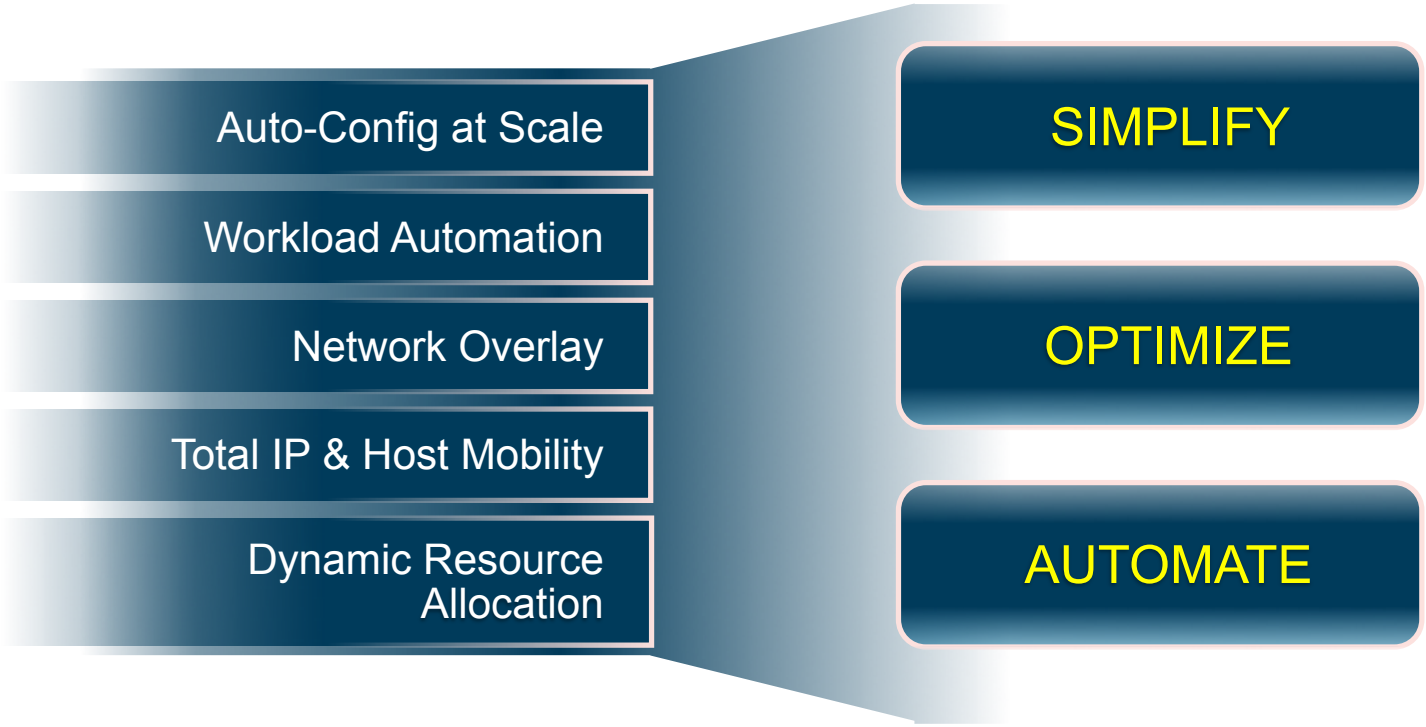
Today's Data Center Challenges



© 2014 Cisco and/or its affiliates. All rights reserved.

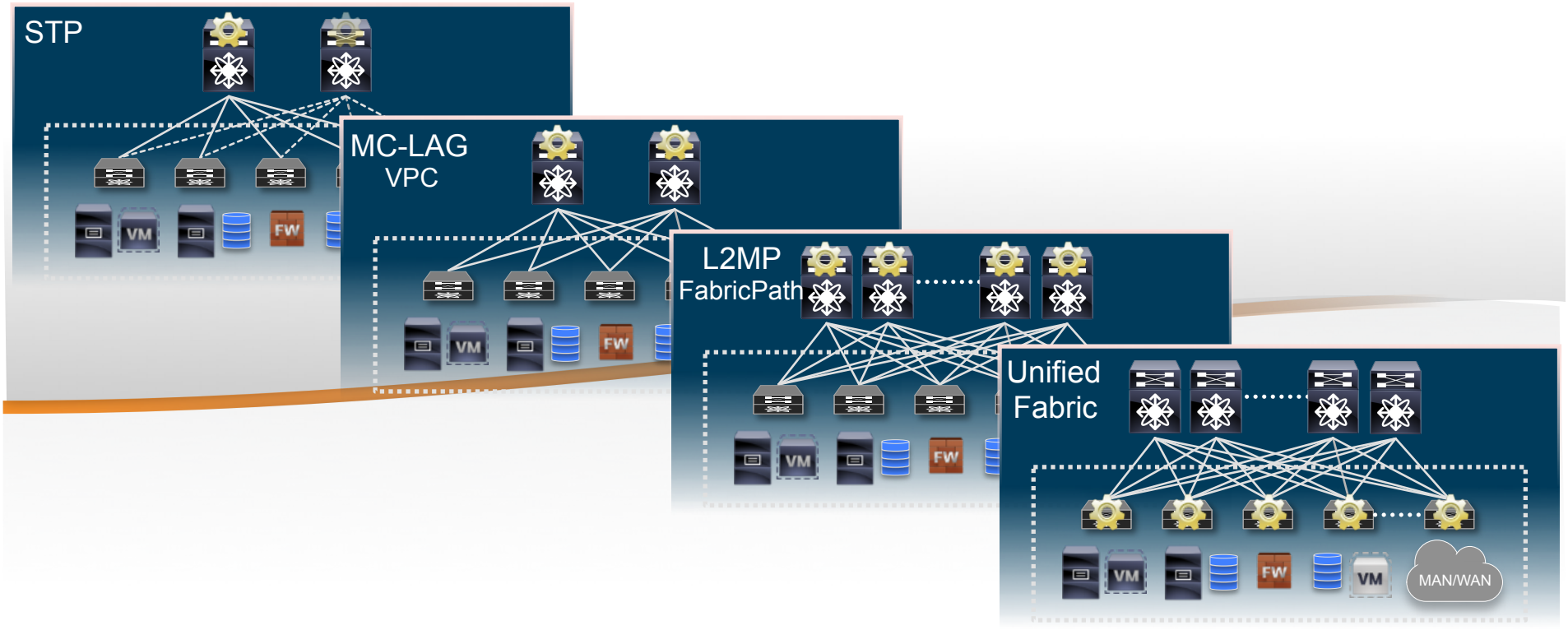
Today's Data Center Challenges

Solved!



© 2014 Cisco and/or its affiliates. All rights reserved.

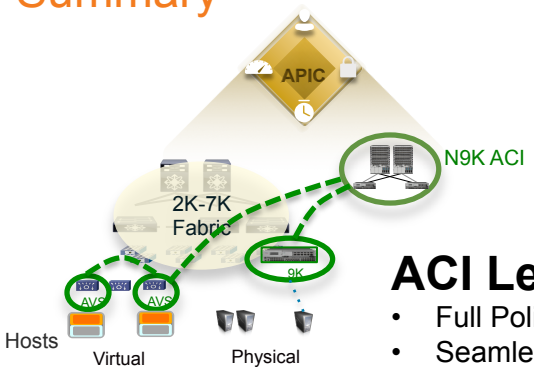
The Data Center Fabric Journey



 © 2014 Cisco and/or its affiliates. All rights reserved.

Cisco Unified Fabric: ACI Integration

Summary

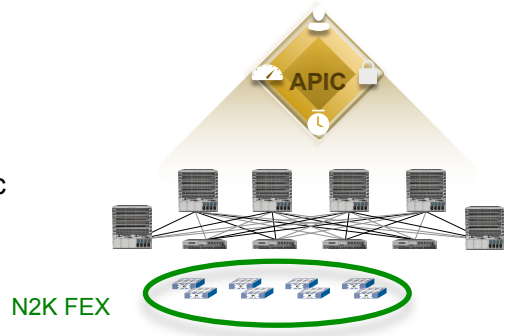


ACI Leaf Overlay

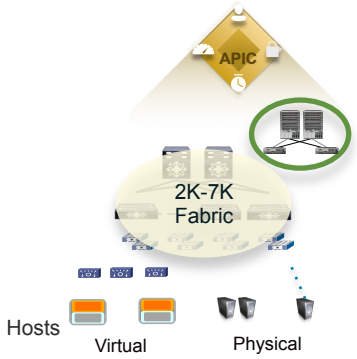
- Full Policy & Management Model
- Seamless HW GWY integration

N2K Integration in ACI Fabric

- Deploy N2K in ACI fabric



Extend **Integrate**

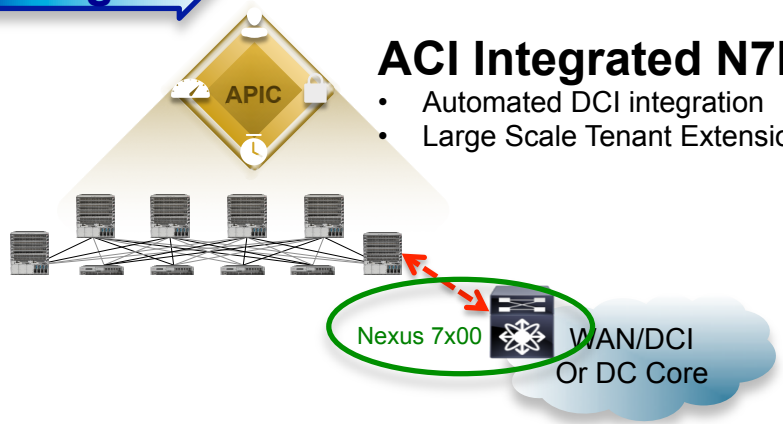


ACI Policy Engine

- Full Policy Model
- Zero impact to existing fabric
- Appliance style addition to fabric

ACI Integrated N7K DCI

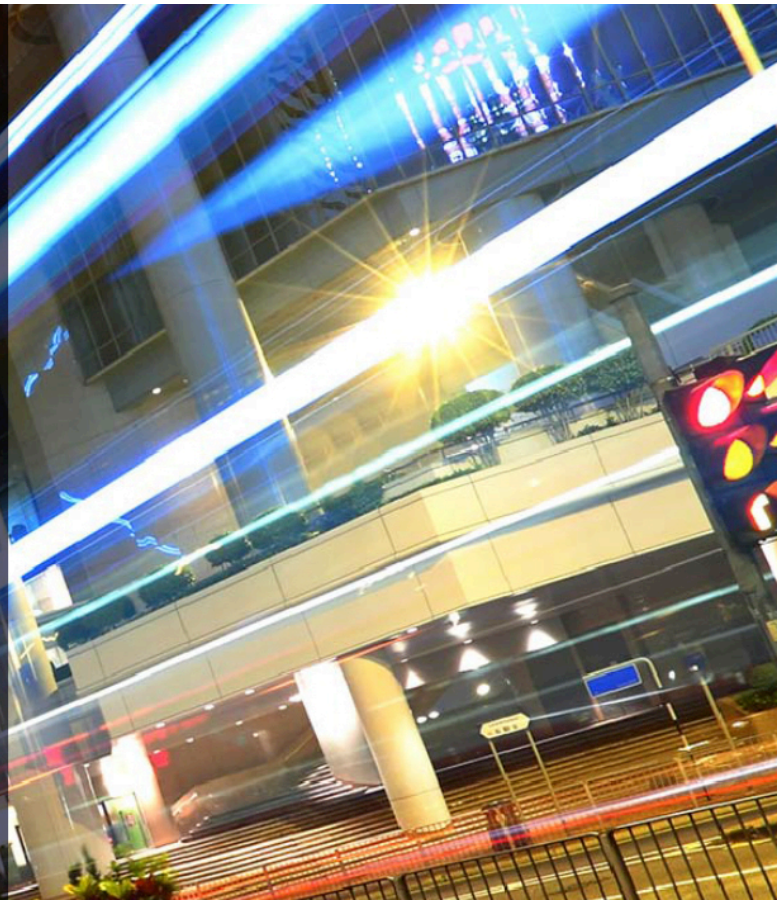
- Automated DCI integration
- Large Scale Tenant Extension



© 2014 Cisco and/or its affiliates. All rights reserved. Cisco Public

Agenda

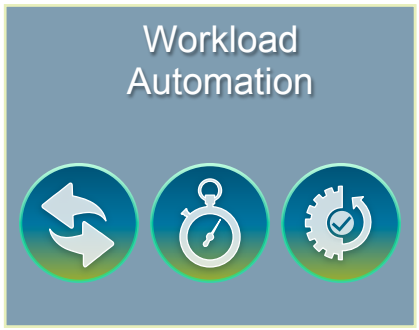
- Requirements and Functions
- **Building Blocks**
  - Hardware Support
  - Fabric Management
  - Workload Automation
  - Optimized Network
  - Virtual Fabrics
- Optimized Network
- Virtual Fabric
- Workload Automation

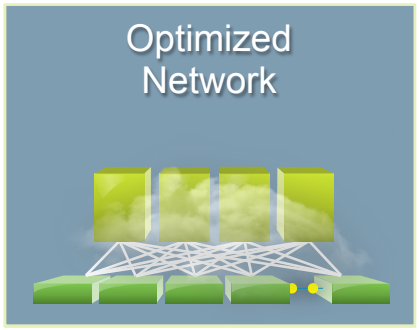


© 2014 Cisco and/or its affiliates. All rights reserved.

Unified Fabric Innovations

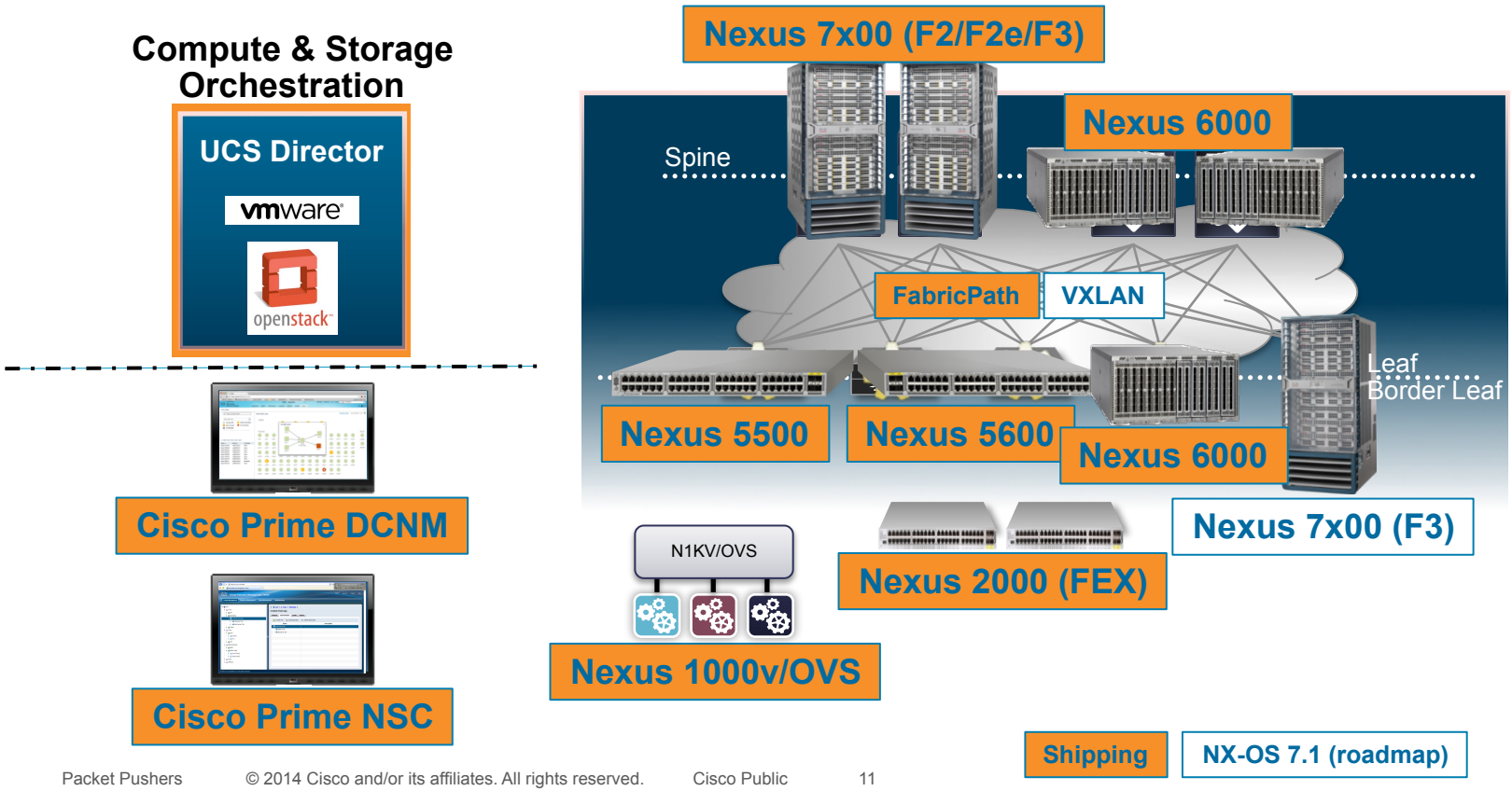
Building Blocks





 © 2014 Cisco and/or its affiliates. All rights reserved.

Platform Support



© 2014 Cisco and/or its affiliates. All rights reserved.

License Requirements

- Nexus 7000 / 7700
  - Enhanced Layer-2 (ENHANCED_LAYER2_PKG)
  - Enterprise Services (LAN_ENTERPRISE_SERVICES_PKG)
  - MPLS Services (MPLS_PKG)
- Nexus 6000 / Nexus 5600
  - Enhanced Layer-2 (ENHANCED_LAYER2_PKG)
  - Layer-3 Base (LAN_BASE_SERVICES_PKG)
  - Layer-3 Enterprise (LAN_ENTERPRISE_SERVICES_PKG)
- Nexus 5500
  - Enhanced Layer-2 (ENHANCED_LAYER2_PKG)

Product	Recommended License PID
Nexus 5500	• N55xx-EL2-SSK9
Nexus 5600	• N5672-DFA-BUN-P1 • N56128-DFA-BUN-P1
Nexus 6000	• N6001-DFA-BUN-P1 • N6004-DFA-BUN-P1
Nexus 7000	• N7K-DFA-BUN-P1 • N77-DFA-BUN-P1

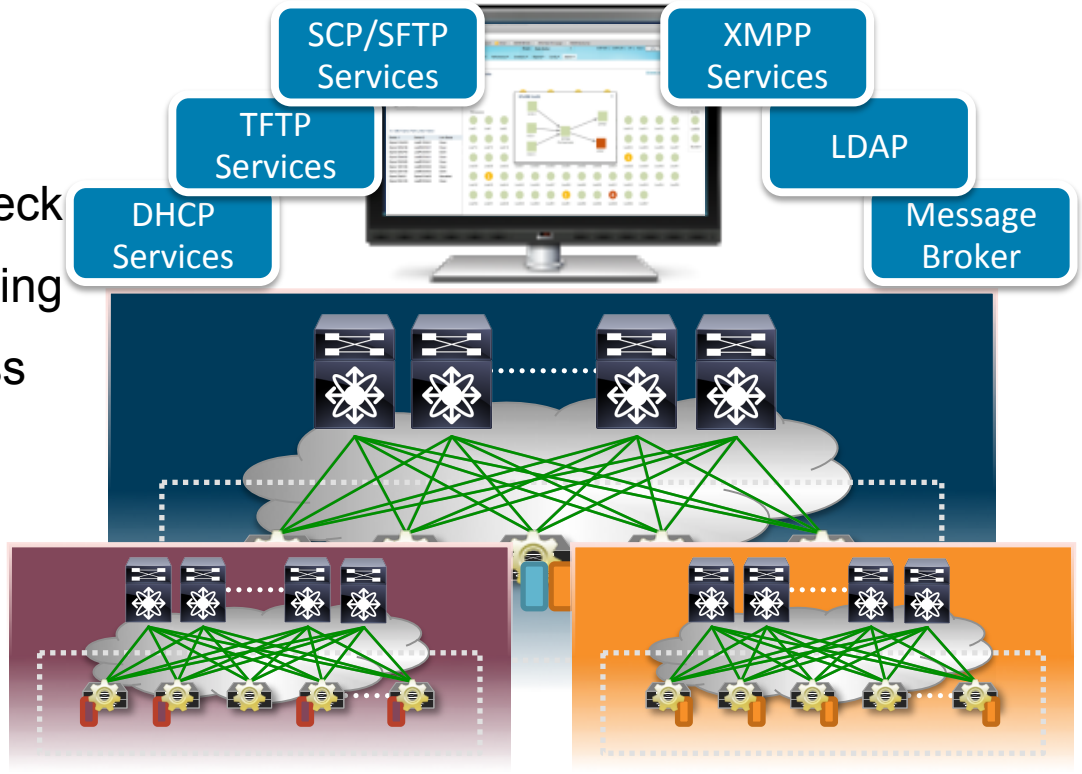
```
n6k# show license usage
Feature                      Ins  Lic   Status Expiry Date Comments
                                 Count
--------------------------------------------------------------------------------
FCOE_NPV_PKG                  No    -   Unused             -
FM_SERVER_PKG                 No    -   Unused             -
ENTERPRISE_PKG                No    -   Unused             -
FC_FEATURES_PKG               No    -   Unused             -
VMFEX_FEATURE_PKG             No    -   Unused             -
ENHANCED_LAYER2_PKG           Yes   -   In use Never       -
LAN_BASE_SERVICES_PKG         Yes   -   In use Never       -
LAN_ENTERPRISE_SERVICES_PKG   Yes   -   In use Never       -
--------------------------------------------------------------------------------
n6k#
```


© 2014 Cisco and/or its affiliates. All rights reserved.

Simplifying Fabric Management & Optimizing Fabric Visibility

Advantages

- Device Auto-Configuration
- Cabling Plan Consistency Check
- Automated Network Provisioning
- Common point of fabric access
- Tenant, Virtual Fabric, & Host Visibility



© 2014 Cisco and/or its affiliates. All rights reserved.

Cisco Prime Data Center Network Manager

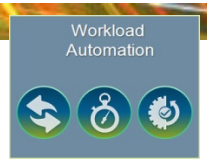
Fabric Features and Licensing

Feature		DCNM Essentials (FREE)	DCNM Advanced (Licensed)
Discovery and Inventory	Discover of Cisco Nexus Switches and related Inventory	✔	
Template Center	Templates for POAP and Push Deployment process	✔	
Fabric Visibility	View of Fabric Tenant information and VM Search. Optionally uses XMPP (prepackaged XCP-Server)	✔	
POAP	Power-on-Auto-Provisioning, integrated DHCP, POAP-Python-Script, TFTP/SCP-Server	✔	
Auto-Configuration	LDAP repository for with pre-populated Profiles for Auto-Configuration of End-Host and Services	✔	
Cable Plan and Consistency Check	Generating Cable-Plan and Visibility including Consistency Check	✔	
Display of historical statistical data	Switch and Interface Performance Collection (stats) and VMware vCenter integration		✔
DCNM-LAN Features (Java Client)	See Table in Link below for Details on Free vs. Licensed Features of DCNM-LAN (beyond Fabric)	✔	✔

Details: http://www.cisco.com/c/en/us/td/docs/switches/datacenter/sw/6_x/dcnm/installation/published/install/licensing_DCNM.html#48103

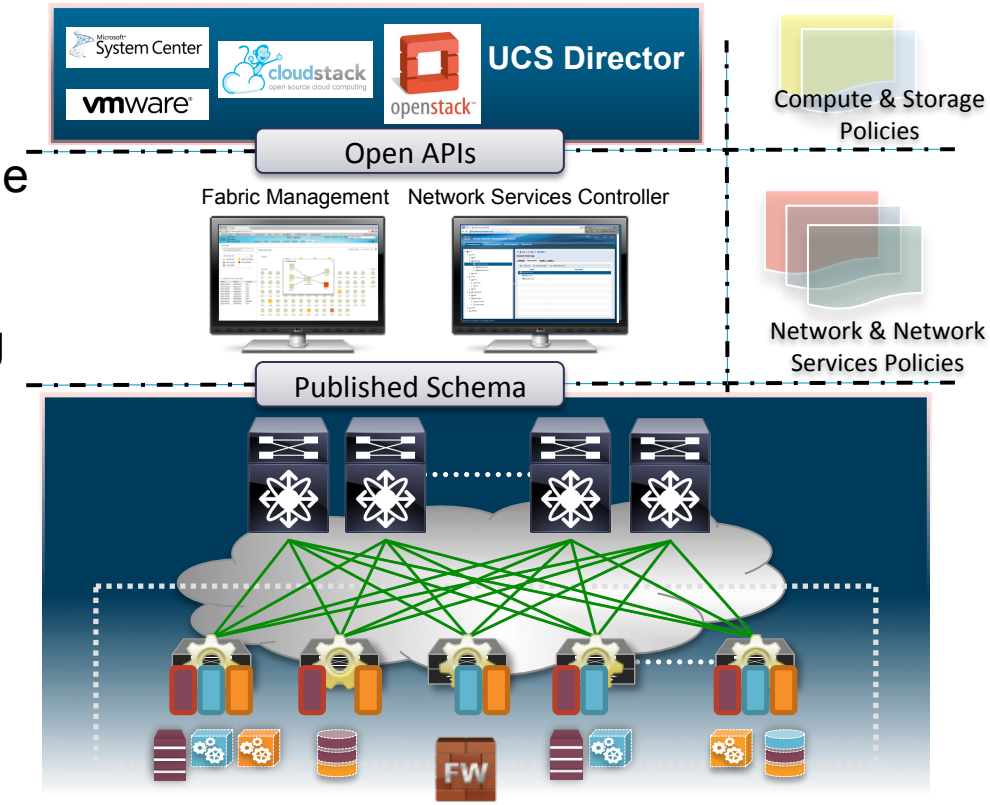
© 2014 Cisco and/or its affiliates. All rights reserved.

Workload Automation & Open Environment

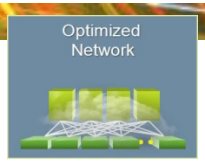


Advantages

- Any workload, Anywhere, Anytime
- Open Integration: Orchestration
- Automated Scalable Provisioning
- Workload aware fabric



© 2014 Cisco and/or its affiliates. All rights reserved.

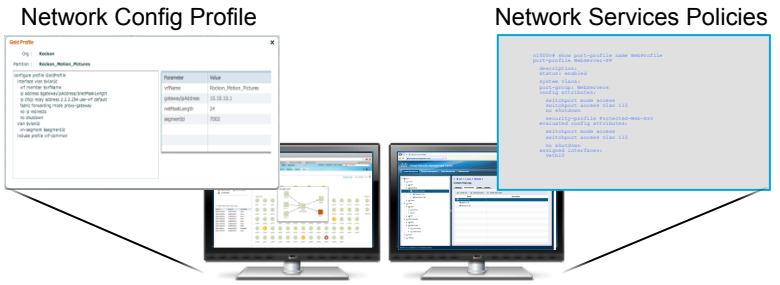


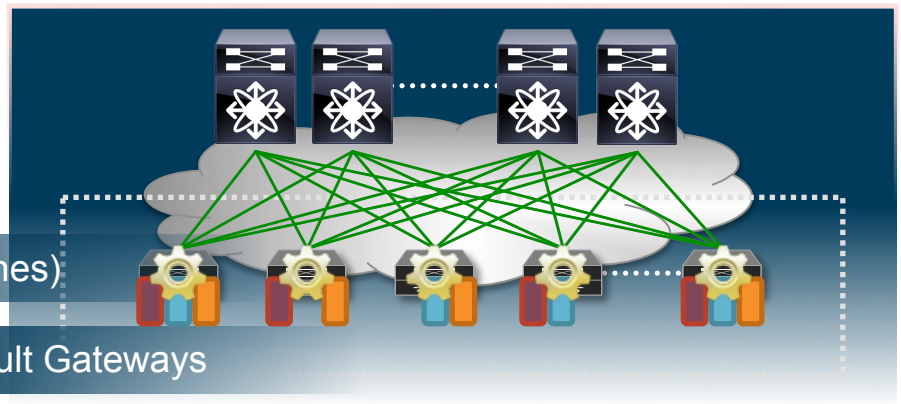
Optimized Network

Scale, Resiliency, and Efficiency

Advantages

- Any subnet, anywhere, rapidly
- Reduced Failure Domains
- Extensible Scale & Resiliency
- Profile Controlled Configuration





- Full Bi-Sectional Bandwidth (N Spines)
- Any/All Leaf Distributed Default Gateways
- Any/All Subnets on Any Leaf

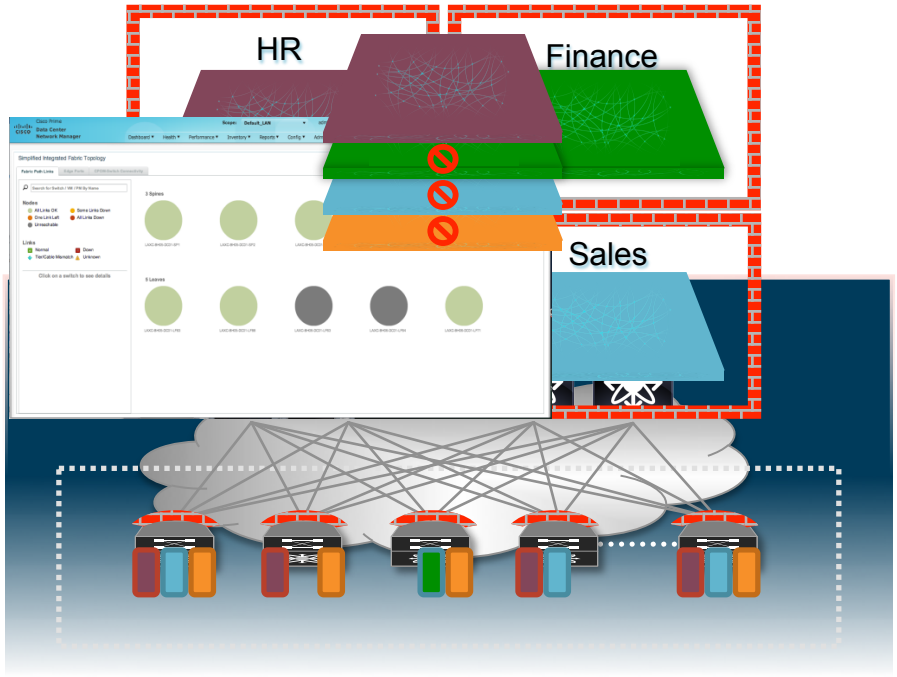
 © 2014 Cisco and/or its affiliates. All rights reserved.

Virtual Fabric for Separation

Virtual Fabrics for Public or Private Cloud Environments

Advantages

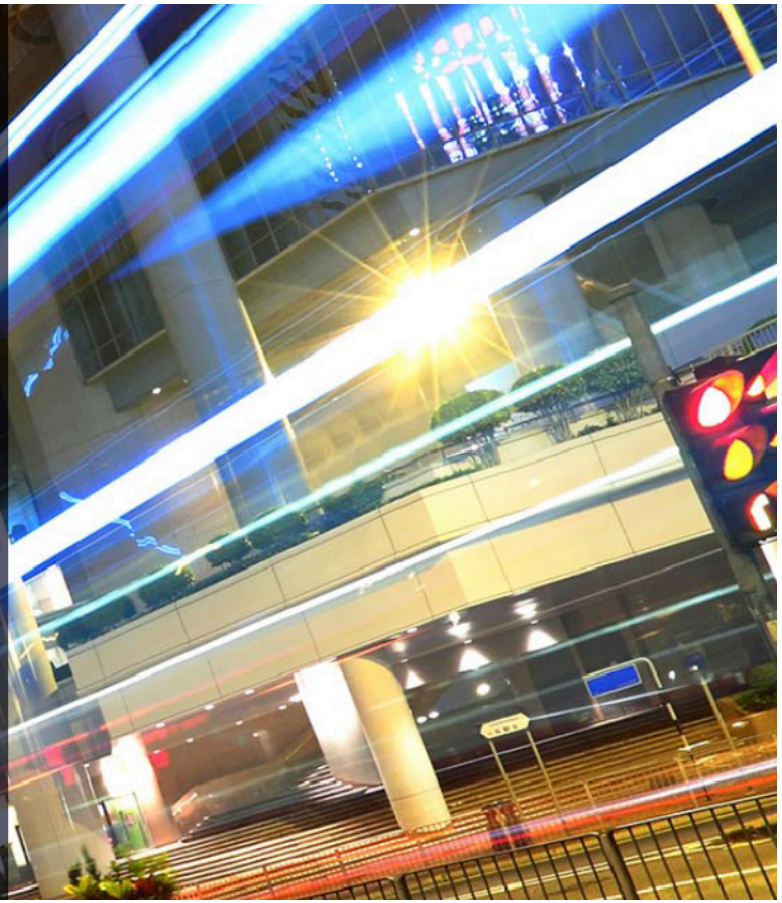
- Any workload, any Virtual Fabric, rapidly
- Scalable Secure Virtual Fabrics
- Virtual Fabric Tenant Visibility
- Routing/Switching Segmentation



 © 2014 Cisco and/or its affiliates. All rights reserved.

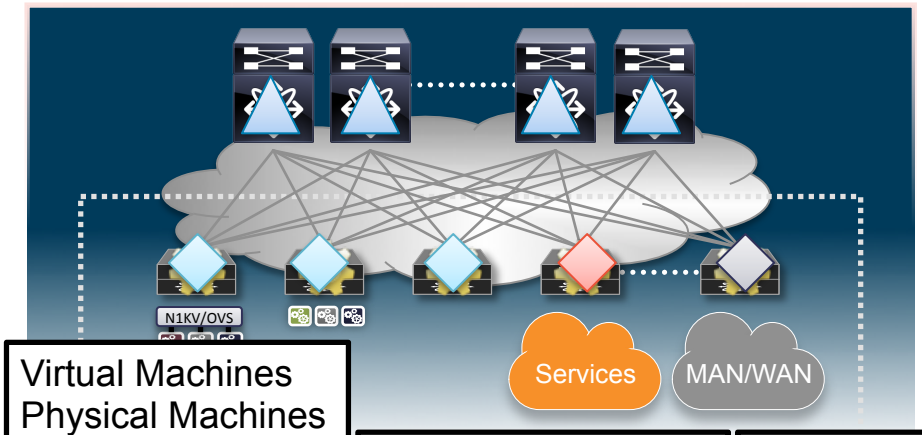
Agenda

- Requirements and Functions
- The Building Blocks
- **Optimized Network**
  - Fabric Properties
  - Control & Data Plane
  - Packet Walk
  - Border Leaf & DCI
- Virtual Fabric
- Workload Automation



© 2014 Cisco and/or its affiliates. All rights reserved.

Device Roles



Virtual Machines
Physical Machines
FEXes
3rd Party Switches
UCS FIs
Blade Switches
Storage

Firewalls
Load Balancers
3rd Party Appliances

Routers
Switches
3rd Party Devices

Spine

Leaf

Border Leaf

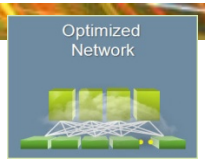
Services Leaf

N1KV/OVS Virtual Leaf*

*Virtual Leaf: N1KV/OVS being a "light" participant on the control plane protocol (supporting VDP)

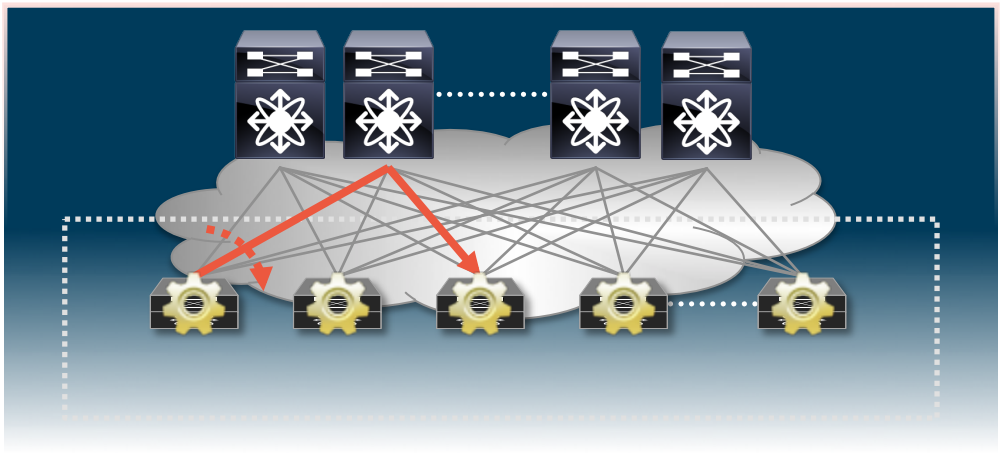
Note: the different Leaf roles are logical and not physical. The same Leaf Switch could perform all three functions (Regular, Services, and Border Leaf)

© 2014 Cisco and/or its affiliates. All rights reserved.



CLOS Fabric Properties

- High Bi-Sectional Bandwidth
- Wide ECMP: Unicast or Multicast
- Uniform Reachability, Deterministic Latency
- High Redundancy: Node/Link Failure
- Line rate, low latency, for all traffic

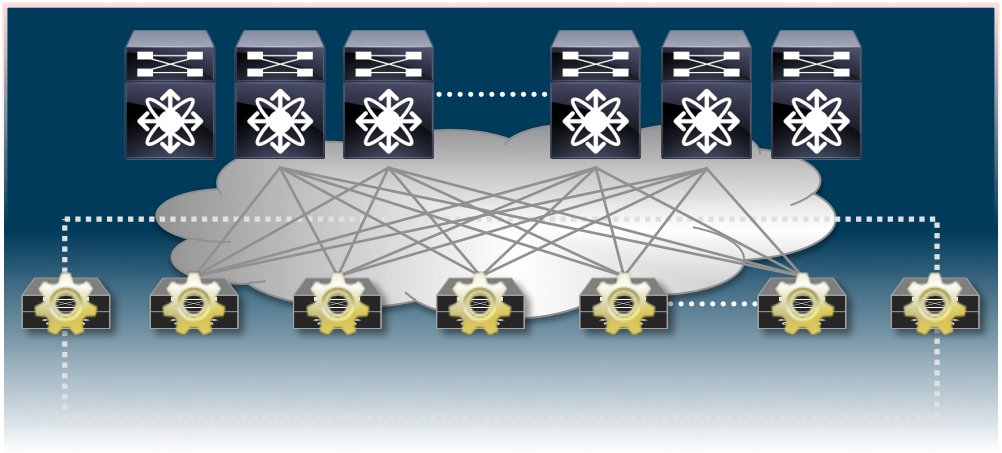


Fabric properties applicable to all topologies

 © 2014 Cisco and/or its affiliates. All rights reserved.

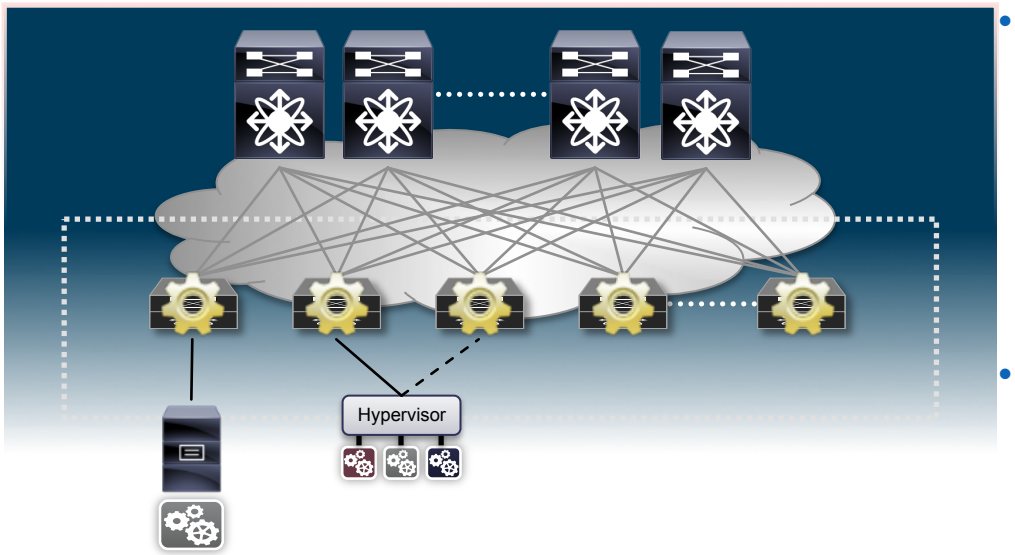
Variety of Fabric Sizes

- Fabric size: Hundreds to 10s of thousands of 10G ports
- Variety of Building Blocks:
  - Varying Size
  - Varying Capacity
  - Desired oversubscription
  - Modular and Fixed
- Scale Out Architecture
  - Add compute, service, external connectivity as the demand grows



 © 2014 Cisco and/or its affiliates. All rights reserved.

Variety of South-bound Topology Connectivity

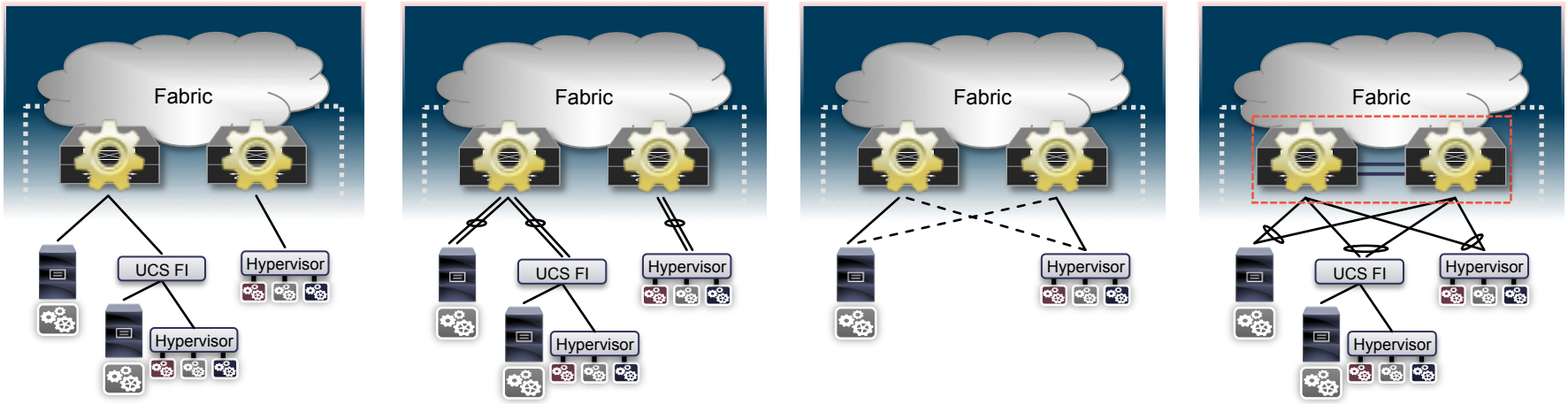


- Flexible connectivity options to the leaf nodes
  - FEX in straight-through or dual-active mode (eVPC)
  - UCS Fabric Interconnects
  - Hypervisors or bare-metal servers attached in vPC mode
- The FEX works as a “remote linecard” and does not participate in control plane and data plane encapsulation

© 2014 Cisco and/or its affiliates. All rights reserved.

Variety of South-bound Topological Connectivity

Server: Physical or Virtual

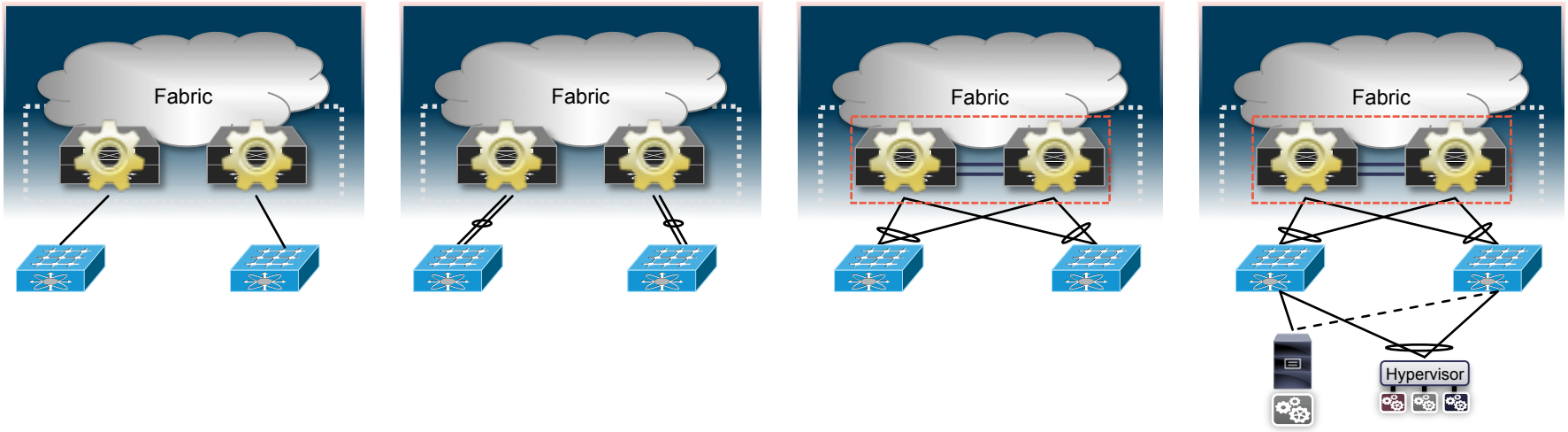


South-bound Switch connectivity is possible

 © 2014 Cisco and/or its affiliates. All rights reserved.

Variety of South-bound Topological Connectivity

FEX – Fabric Extender (Nexus 2000)



 © 2014 Cisco and/or its affiliates. All rights reserved.

Fabric Control Plane

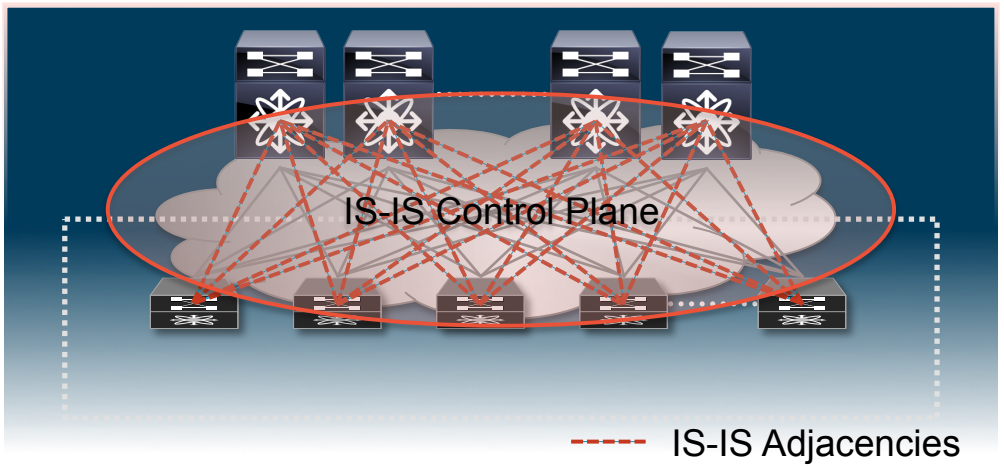
IS-IS as Fabric Control Plane

IS-IS for fabric link state distribution

- Fabric node reachability for overlay encapsulation
- Building multi-destination trees for multicast and broadcast traffic
- Quick reaction to fabric link/node failure (Layer-2 BFD)
- Enhanced for mesh topologies

Fabric Control Protocol doesn't distribute

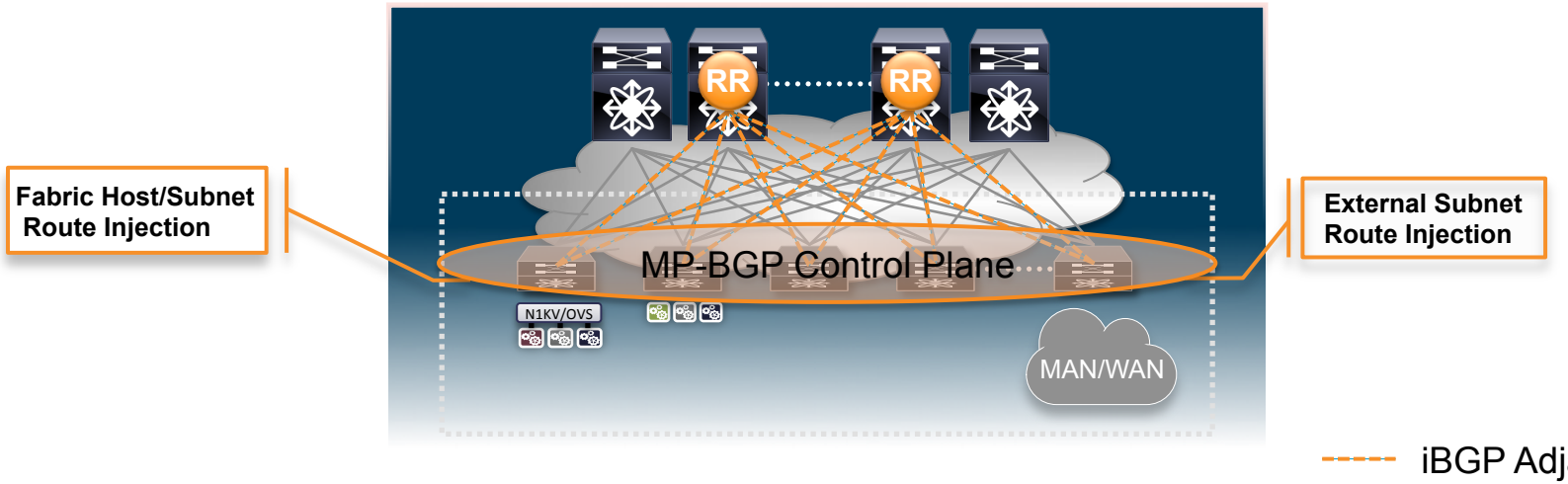
- Host Routes
- Host originated control traffic
- Server subnet information



 © 2014 Cisco and/or its affiliates. All rights reserved.

Fabric Control Plane

Host and Subnet Route Distribution



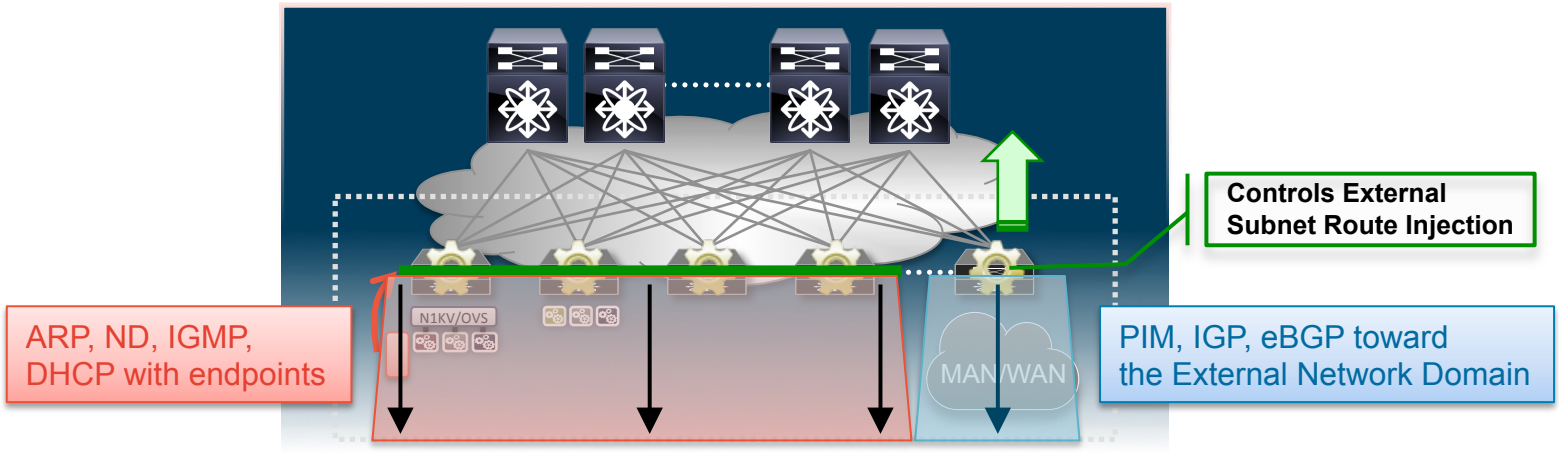
- Host Route Distribution decoupled from the Fabric link state protocol
- Use MP-BGP on the leaf nodes to distribute internal host/subnet routes and external reachability information
- MP-BGP enhancements to carry up to 1.2 Million routes and reduce convergence time

Note: Route-Reflectors deployed for scaling purposes

© 2014 Cisco and/or its affiliates. All rights reserved.

Fabric Control Plane

Host Originated Protocols Containment

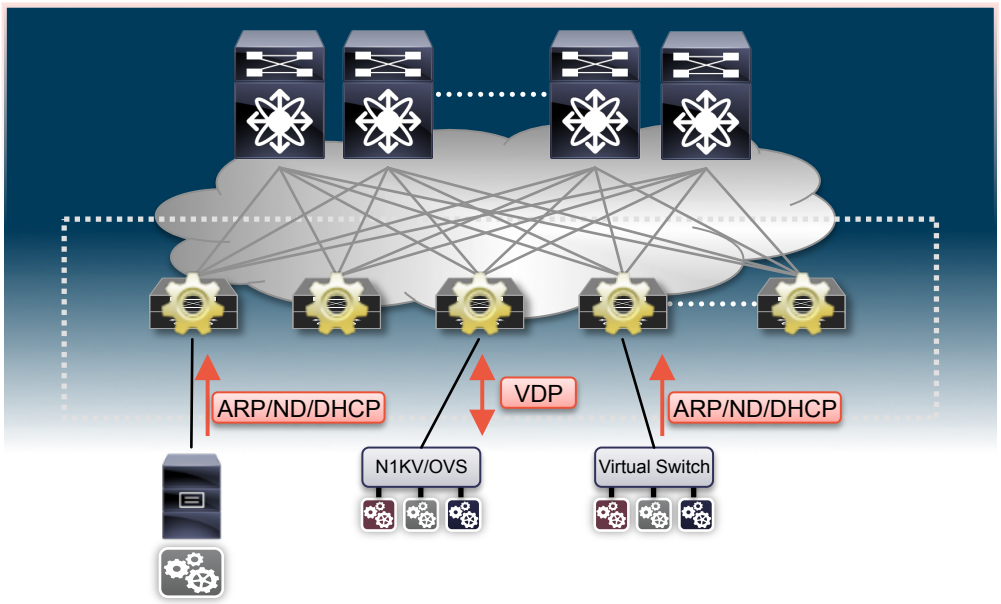


- ARP, ND, IGMP, DHCP originated on servers are terminated on Leaf nodes
- Contain floods and failure domains, distribute control plane processing
- Terminate PIM, OSPF, eBGP from external networks on Border Leafs

© 2014 Cisco and/or its affiliates. All rights reserved.

Fabric Control Plane

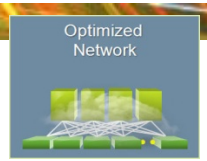
Host Detection and Deletion



- In order to advertise host reachability information, a leaf must first discover locally connected devices
- Detection of local hosts
  - Based on VDP* or ARP/ND/DHCP
  - *VDP (VSI Discovery and Configuration Protocol) is IEEE 802.1Qbg Clause 41
- Detection of remote hosts
  - Received MP-BGP notifications

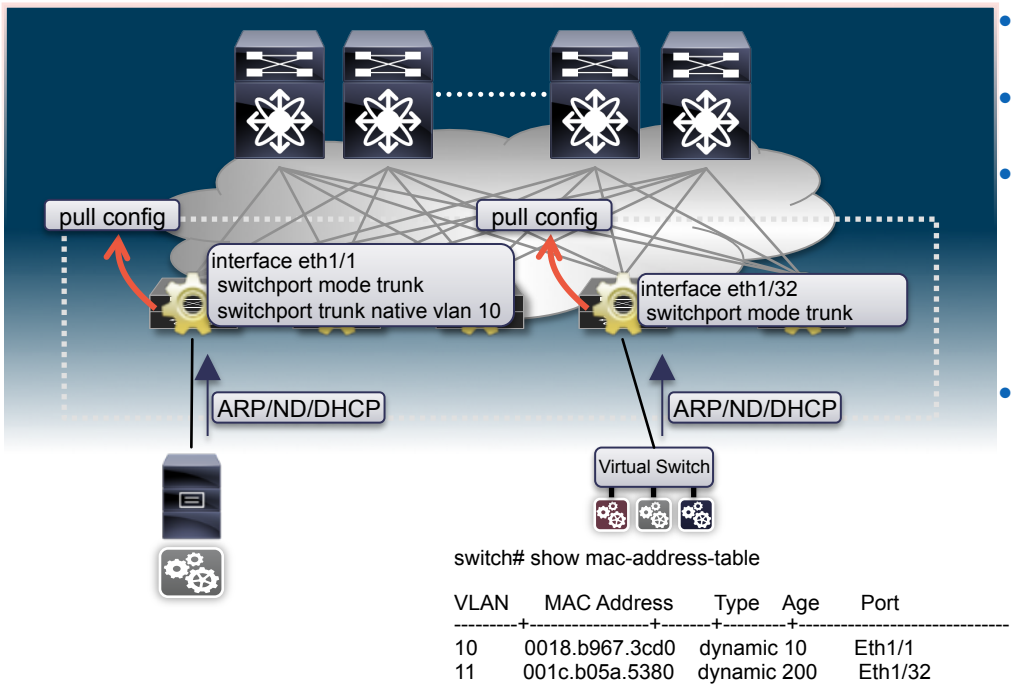
Note: Discovered IP address information from ARP/ND-Table get redistributed into MP-BGP Control Plane for End-Host reachability

© 2014 Cisco and/or its affiliates. All rights reserved.



Fabric Control Plane

Host Detection and Deletion (Detail on Data Plane Trigger)



switch# show mac-address-table

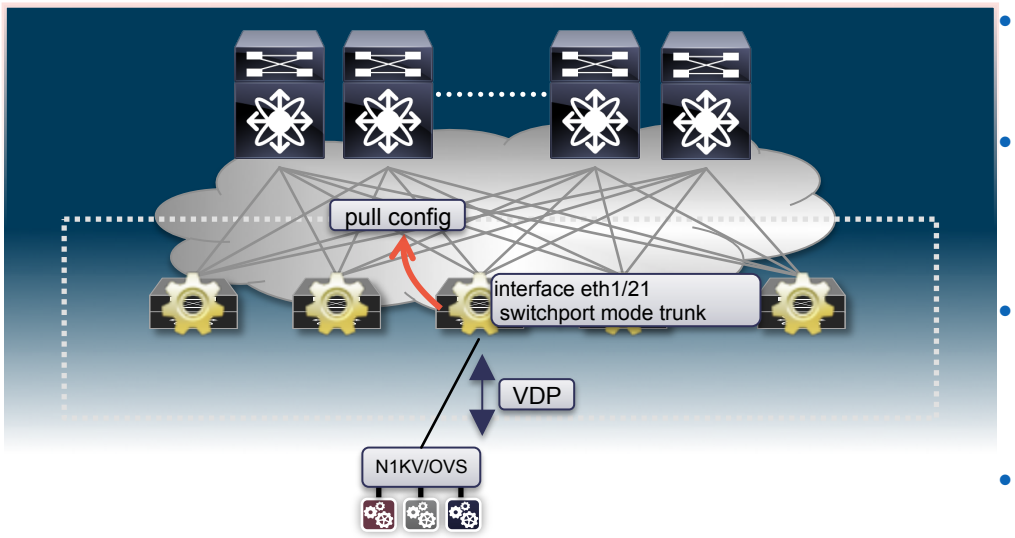
VLAN	MAC Address	Type	Age	Port
10	0018.b967.3cd0	dynamic	10	Eth1/1
11	001c.b05a.5380	dynamic	200	Eth1/32

- Data packet from Server reaches Leaf
- Leaf detects new MAC learn event
- VLAN detected based on:
  - IEEE 802.1q tag used between Server and Leaf
  - VLAN configured on Leaf port
- Based on learned VLAN and Leaf local configuration parameters, logical configuration get pulled, instantiated and applied on Leaf

© 2014 Cisco and/or its affiliates. All rights reserved.

Fabric Control Plane

Host Detection and Deletion (Detail on VDP* Trigger)



switch# show evb host

Host Name	VNI	Vlan	BD	Mac-address	IP-Address	Interface
Server	31000	3000	3000	0050.56ac.1f71	192.168.131.103	Eth1/21

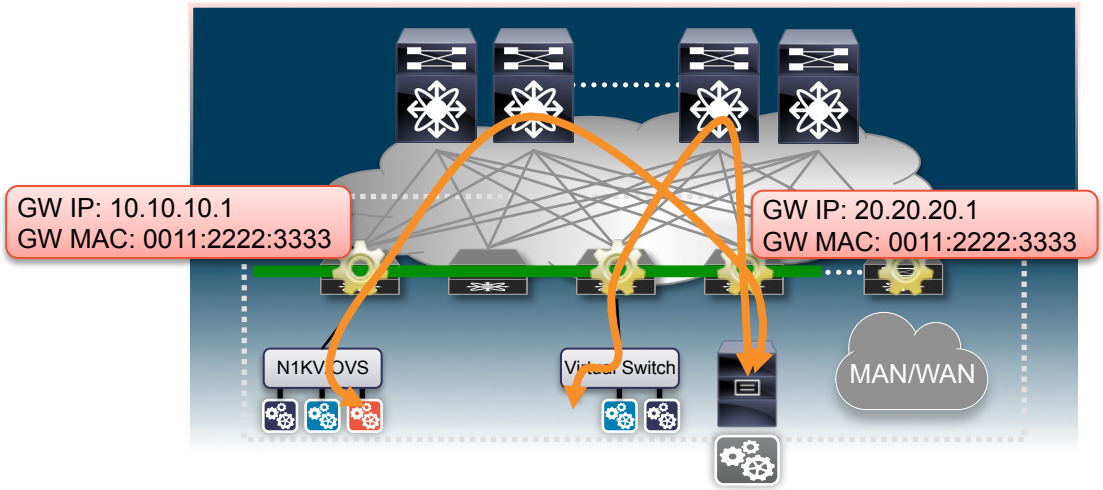
- VDP* session gets established between virtual switch and physical Leaf
- Segment ID (VNI) gets sent from virtual switch based on Virtual Machine configuration
- Physical Leaf responds with next available VLAN defined in Pool (system fabric dynamic-vlans xxx-yyy)
- Based on learned Segment ID (VNI), logical configuration get pulled, instantiated, and applied on Leaf

*VDP (VSI Discovery and Configuration Protocol) is IEEE 802.1Qbg Clause 41

© 2014 Cisco and/or its affiliates. All rights reserved.

Optimized Network

Distributed Anycast Gateway at the Leaf



Anycast Gateway

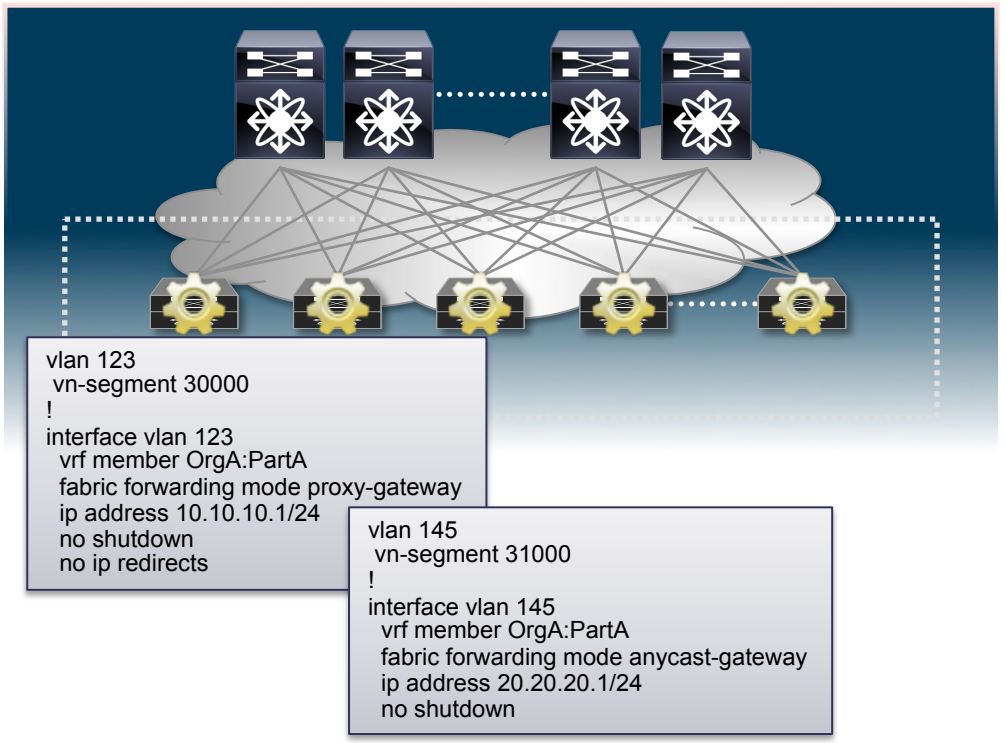
- Any Subnet anywhere => Any Leaf can instantiate ANY Subnet
  - All Leafs share gateway IP and MAC for a Subnet **(No HSRP)**
  - ARPs are terminated on Leafs, No Flooding beyond Leaf
- Facilitates VM Mobility, workload distribution, arbitrary clustering
- Seamless Layer-2 or Layer-3 communication between physical hosts and virtual machines

© 2014 Cisco and/or its affiliates. All rights reserved.

Optimized Networking

Distributed Gateway Mode

- Distributed Gateway exists on all Leafs where VLAN/Segment-ID is active
- There are different Forwarding Modes for the Distributed Gateway:
  - Proxy-Gateway (Enhanced Forwarding)
    - Leverages local proxy-ARP
    - Intra and Inter-Subnet forwarding based on Routing
    - Contain floods and failure domains to the Leaf
  - Anycast-Gateway (Traditional Forwarding)
    - Intra-Subnet forwarding based on Bridging
    - Data-plane based conversational learning for endpoints MAC addresses
    - ARP is flooded across the fabric



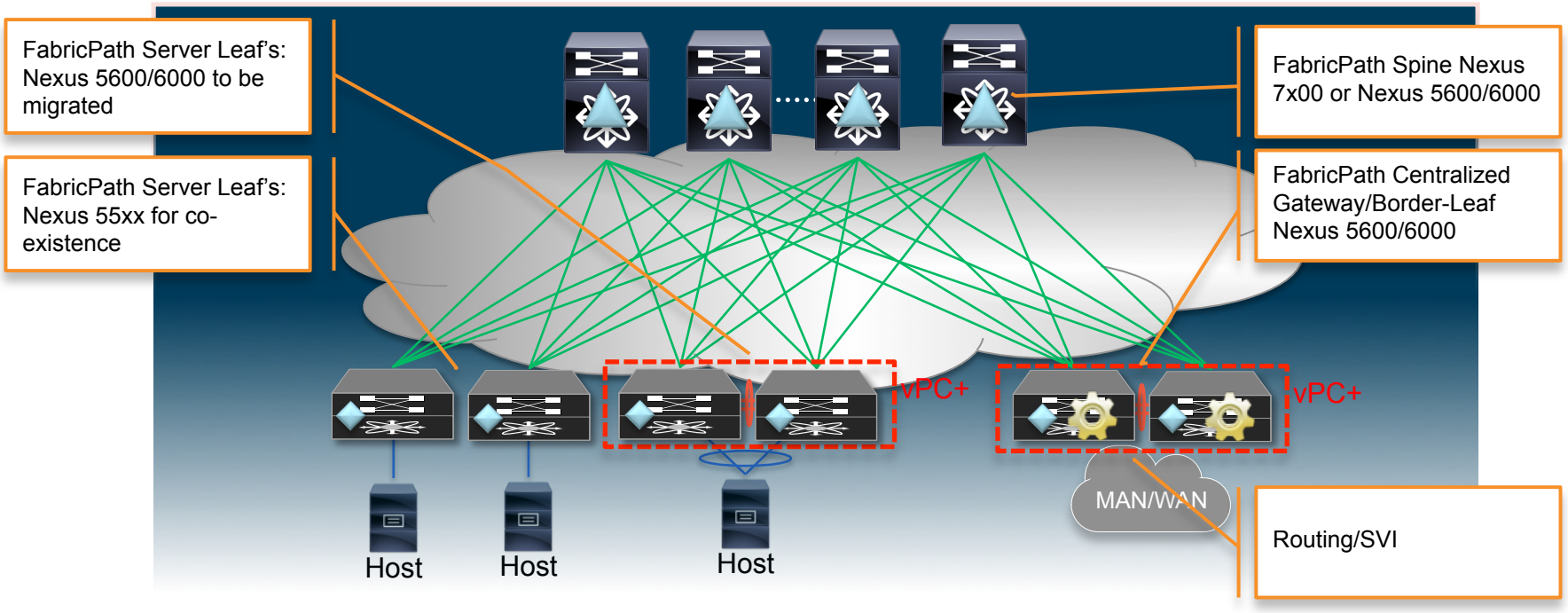
 © 2014 Cisco and/or its affiliates. All rights reserved.

Fabric Forwarding Mode Comparison

	Proxy-Gateway (Enhanced Forwarding)	**Anycast-Gateway** (Traditional Forwarding)	**Regular Layer-3 Mode** (Cisco FabricPath)
VLAN/Subnets stretched between Leafs	✓	✓	✓ (requires anchor Leaf)
Common Anycast GW IP across Leafs	✓	✓	✗
Common Anycast GW MAC across Leafs	✓	✓	✗
Use local Proxy-ARP/ND	✓ (respond to ARP/ND only if the destination is available in the RIB)	✗	✗
ARP Flooding in Layer-2 Domain	✗	✓ (floods also across Fabric)	✓ (flood within VLAN)
Intra-Subnet forwarding	Always routed (TTL decrement)	Bridged	Bridged
Silent Host Discovery	✗	✓	✓
Non-IP Forwarding	✓	✓	✓

© 2014 Cisco and/or its affiliates. All rights reserved.

Existing FabricPath Network Co-Existence



= FabricPath Spine

= Optimized Networking Spine

= FabricPath Leaf

= Optimized Networking Leaf

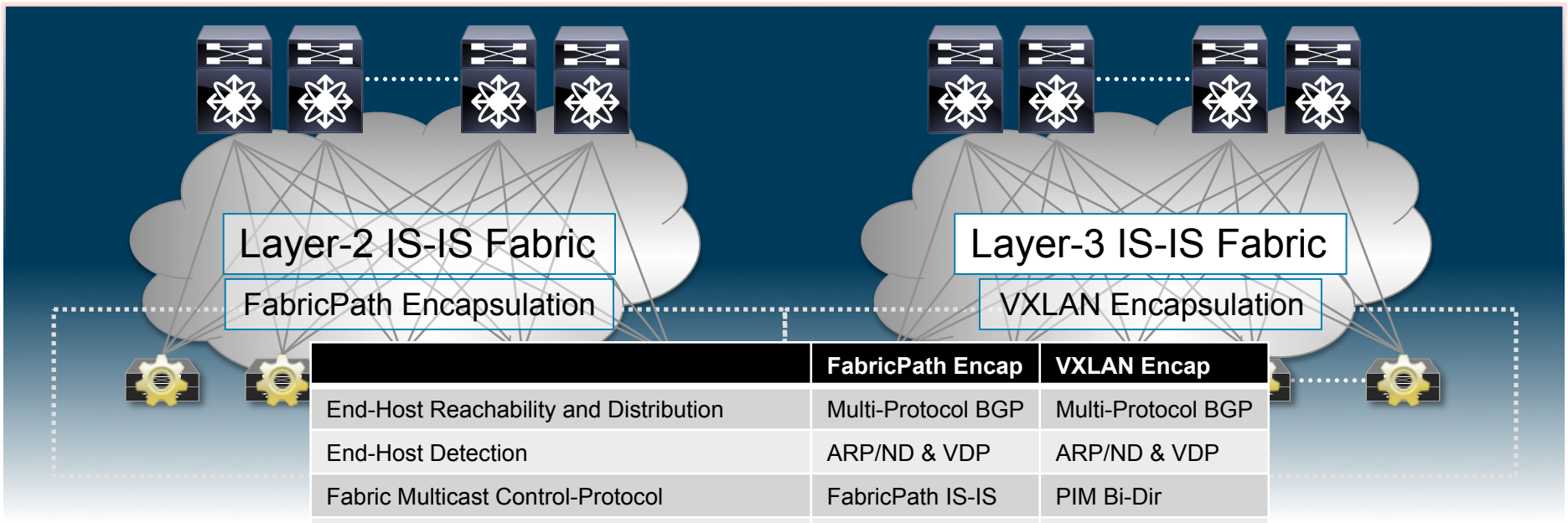
= SVI/Distributed GW

RR = Route Reflector

 © 2014 Cisco and/or its affiliates. All rights reserved.

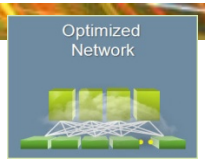
Encapsulation and Forwarding

What about Encapsulation?



	FabricPath Encap	VXLAN Encap
End-Host Reachability and Distribution	Multi-Protocol BGP	Multi-Protocol BGP
End-Host Detection	ARP/ND & VDP	ARP/ND & VDP
Fabric Multicast Control-Protocol	FabricPath IS-IS	PIM Bi-Dir
Fabric Topology Control-Plane	FabricPath IS-IS	Layer-3 IS-IS
Leaf IP/MAC binding distribution within Fabric	FabricPath IS-IS	Multi-Protocol BGP

© 2014 Cisco and/or its affiliates. All rights reserved.

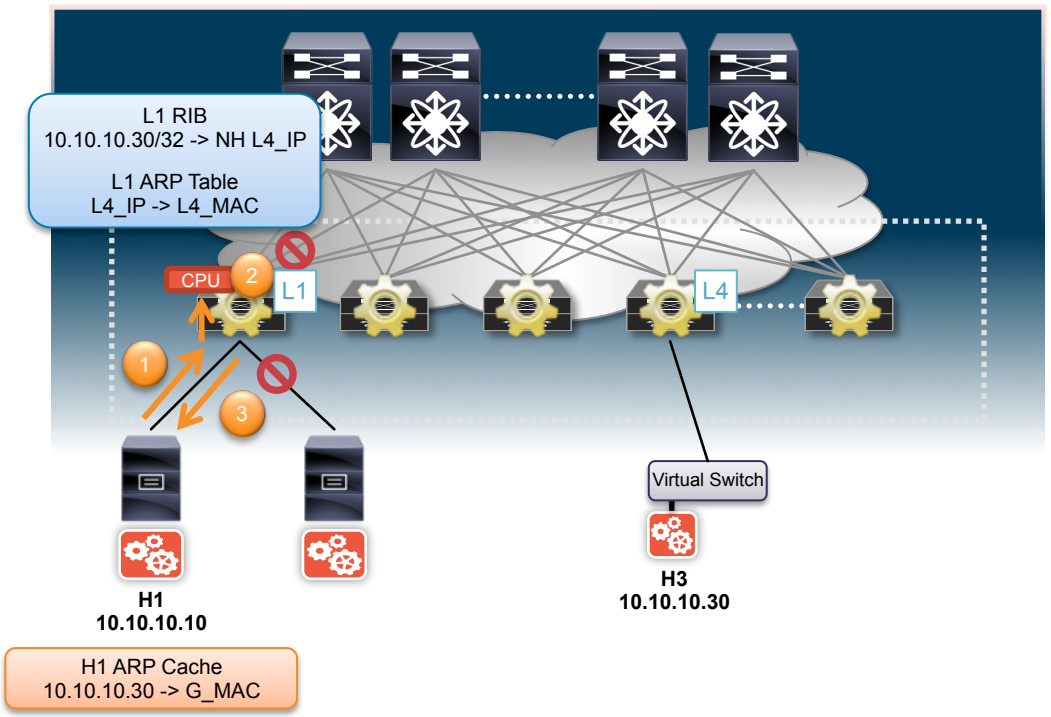


Optimized Networking

IP Forwarding within the Same Subnet

1. H1 sends an ARP request for H3 – 10.10.10.30
2. The ARP request is intercepted at Leaf1 (L1) and punted to the Supervisor
3. Assuming a valid route to H3 does exist in the Unicast RIB, Leaf1 (L1) sends the ARP reply with G_MAC so that H1 can build its ARP cache

Note: the ARP request is NOT flooded across the Fabric nor out of other local interfaces belonging to the same Layer-2 domain

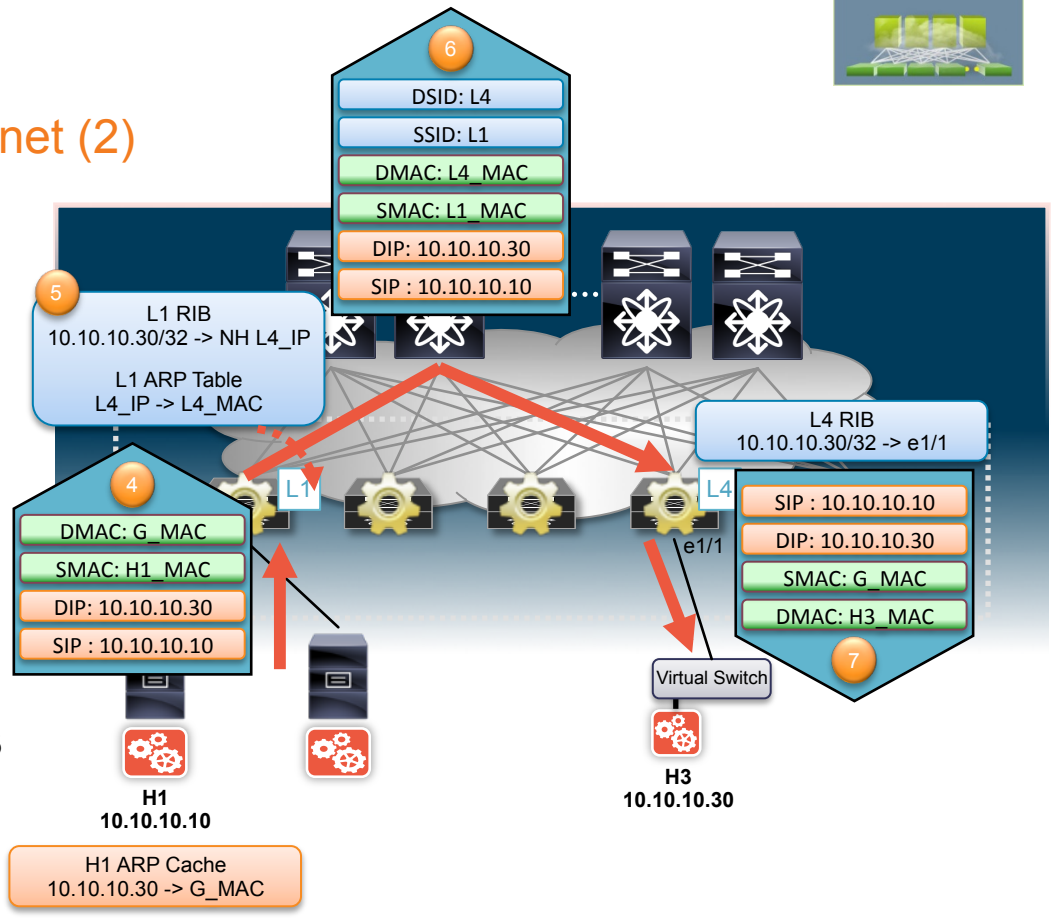


 © 2014 Cisco and/or its affiliates. All rights reserved.

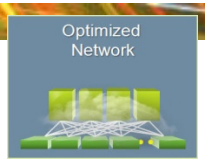
Optimized Networking

IP Forwarding within the Same Subnet (2)

4. H1 generates a data packet with G_MAC as destination MAC
5. Leaf1 (L1) receives the packet and performs Layer-3 lookup for the destination
6. Leaf1 (L1) adds the Layer-2 and the FabricPath headers and forwards the encapsulated frame across the Fabric, picking one of the equal cost paths available via the multiple Spines
7. Leaf4 (L4) receives the packet, strips off the FabricPath header and performs Layer-3 lookup and forwarding toward H3



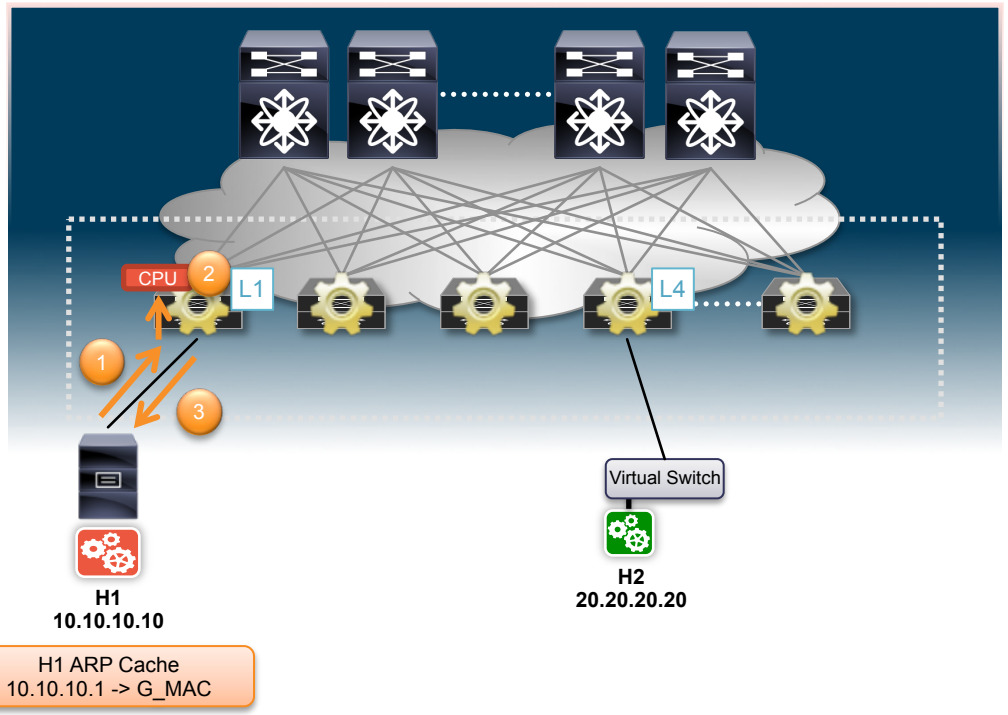
 © 2014 Cisco and/or its affiliates. All rights reserved.



Optimized Networking

IP Forwarding Across Different Subnet

1. H1 sends an ARP request for Default Gateway – 10.10.10.1
2. The ARP request is intercepted at the Leaf1 (L1) and punted to the Supervisor
3. Leaf1 (L1) acts as a regular Default Gateway and sends ARP reply with G_MAC

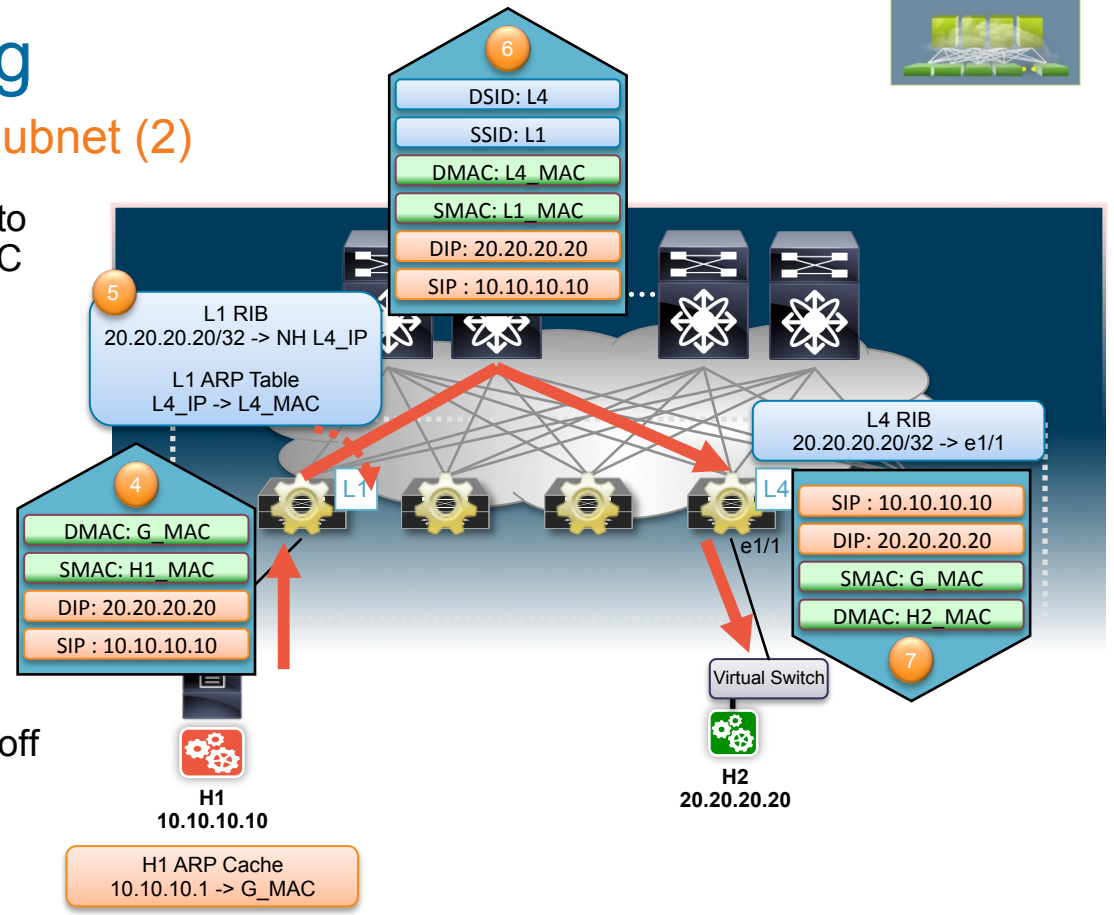


© 2014 Cisco and/or its affiliates. All rights reserved.

Optimized Networking

IP Forwarding Across Different Subnet (2)

4. H1 generates a data packet destined to H2 IP with G_MAC as destination MAC
5. Leaf1 (L1) receives the packet and performs Layer-3 lookup for the destination
6. If valid routing information for H2 is available in the unicast routing table, Leaf1 (L1) adds the Layer-2 and the FabricPath headers and forwards the FabricPath frame across the Fabric, picking one of the equal cost paths available via the multiple Spines
7. Leaf4 (L4) receives the packet, strips off the FabricPath header and performs Layer-3 lookup and forwarding toward H2



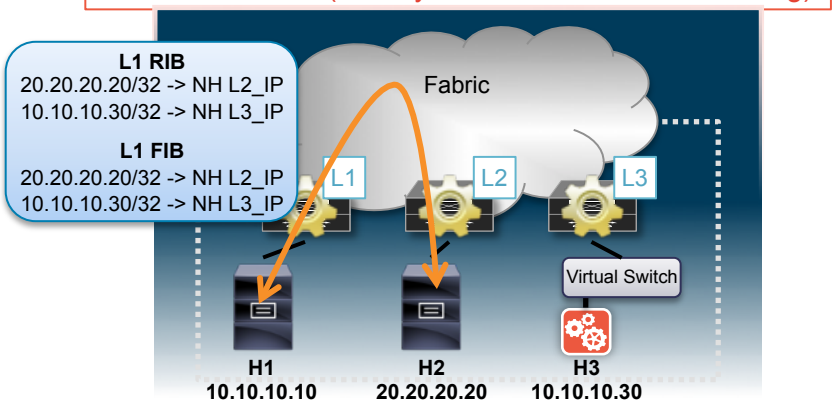
 © 2014 Cisco and/or its affiliates. All rights reserved.

Optimized Networking

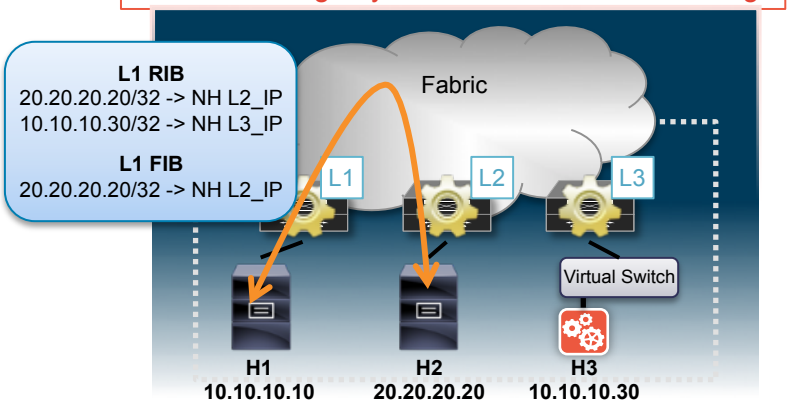
Introducing Layer-3 Conversational Learning

- Use of /32 host routes may lead to scaling issues if all the routes are installed in the Hardware tables of all Leaf nodes
  - Layer-3 conversational learning is introduced to alleviate this concern
  - Disabled by default -> all host routes are programmed in the Hardware
- With Layer-3 conversational learning, host routes for remote endpoints will be programmed into the Hardware FIB (from the Software RIB) upon detection of an active conversation from a local endpoint

Default Behavior (No Layer-3 Conversational Learning)



After Enabling Layer-3 Conversational Learning

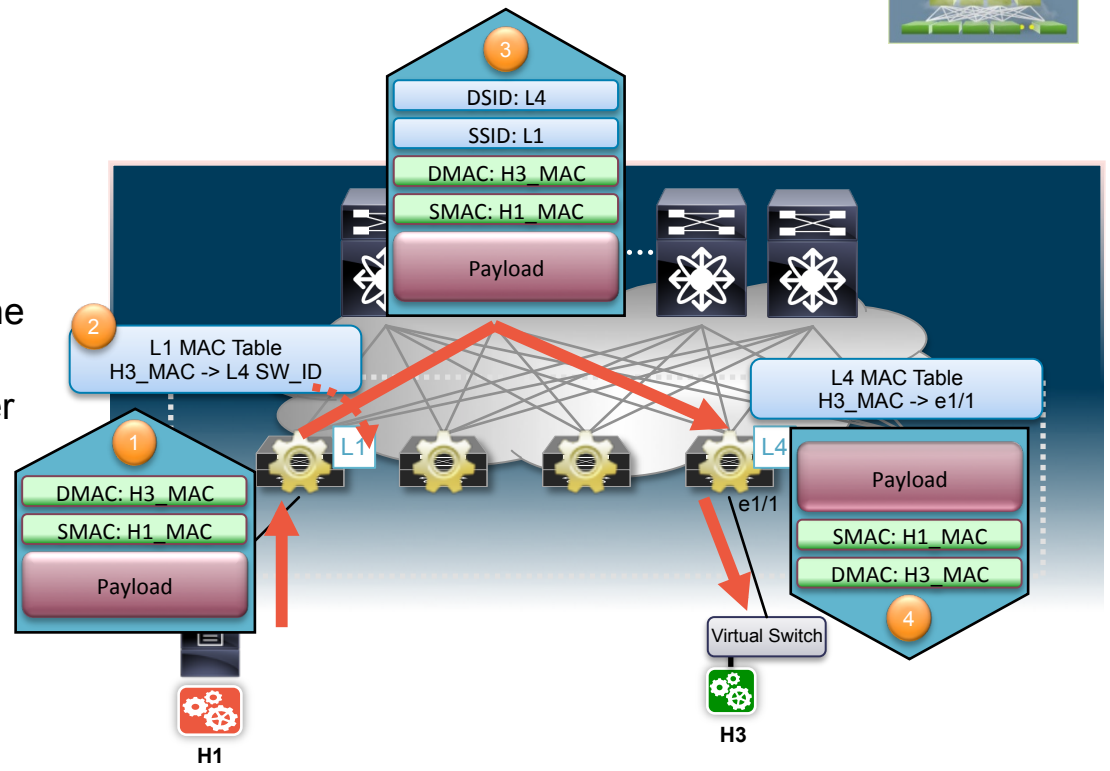


© 2014 Cisco and/or its affiliates. All rights reserved.

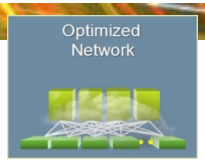
Forwarding

Layer-2 non IP Flows

1. H1 originates a packet destined to H3 MAC address
2. Layer-2 lookup is performed by Leaf1 (L1) in the MAC Table for the VLAN the frame belongs to
3. Leaf1 (L1) adds the FabricPath header before sending the packet into the fabric
4. Leaf4 (L4) receives the frame, decapsulates the FabricPath header, performs the Layer-2 lookup and then sends the frame to H3



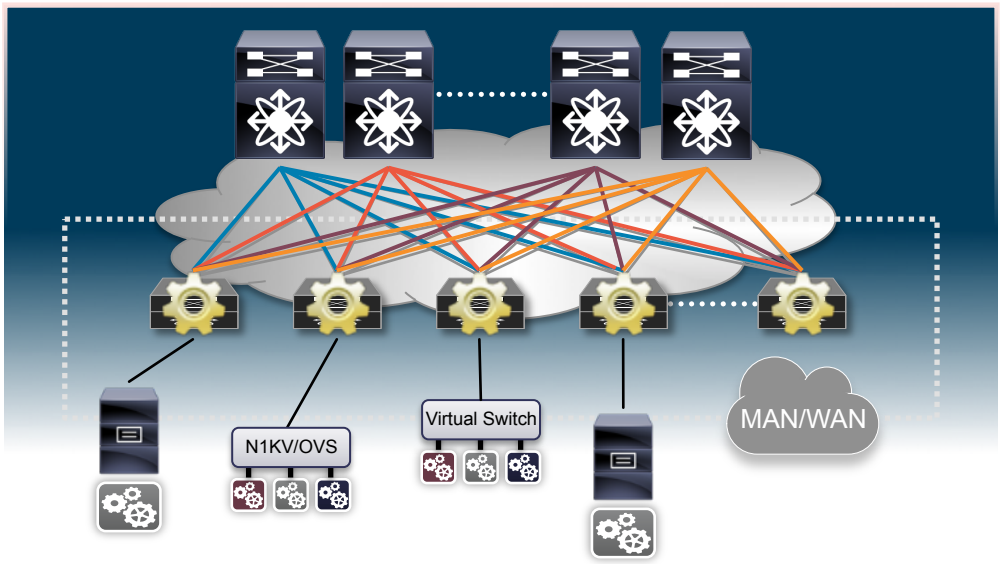
 © 2014 Cisco and/or its affiliates. All rights reserved.



Optimized Networking

Multicast Forwarding

- Fabric supports computation of multiple distribution trees leveraging IS-IS
  - Used for multicast and broadcast traffic
  - No need for other multicast protocols (PIM, etc.) inside the fabric
- Multi Destination Trees (MDTs) Rooted on Spines
- Ingress Leaf load balances traffic across multiple paths
  - Efficient use of fabric links

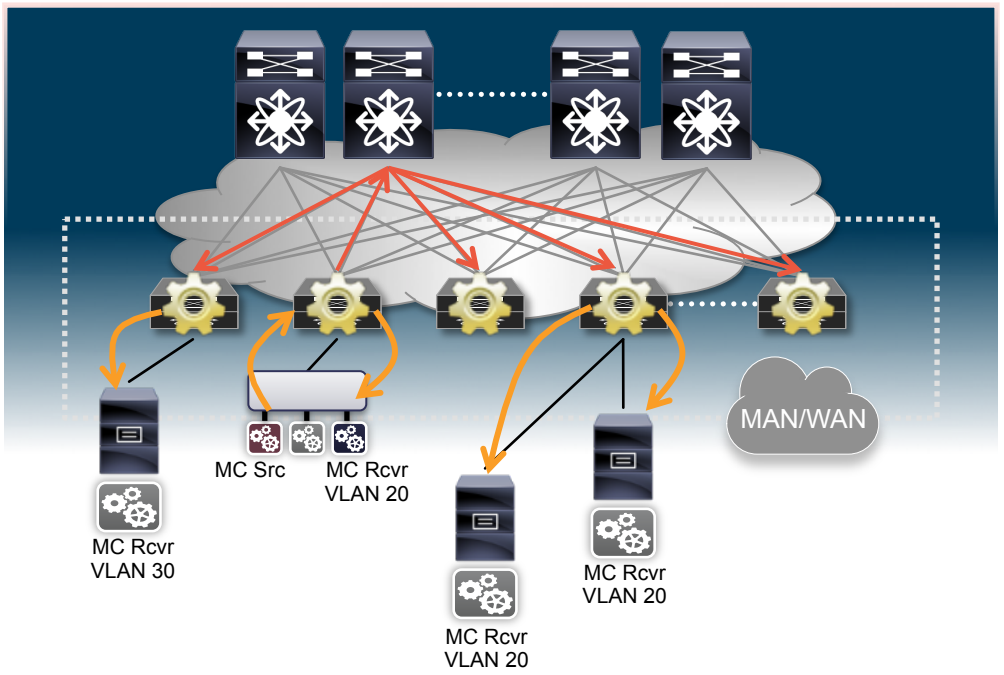


 © 2014 Cisco and/or its affiliates. All rights reserved.

Optimized Networking

Multicast Forwarding

- Two tiers multicast replication across the fabric
  - Ingress Leaf always performs multicast routing functions and sends a single copy to the fabric
  - Spine node replicates to the leaf nodes
  - Destination Leaf nodes locally replicate to the local receivers
- Optimization possible to allow pruning on the spine (per tenant VRF)

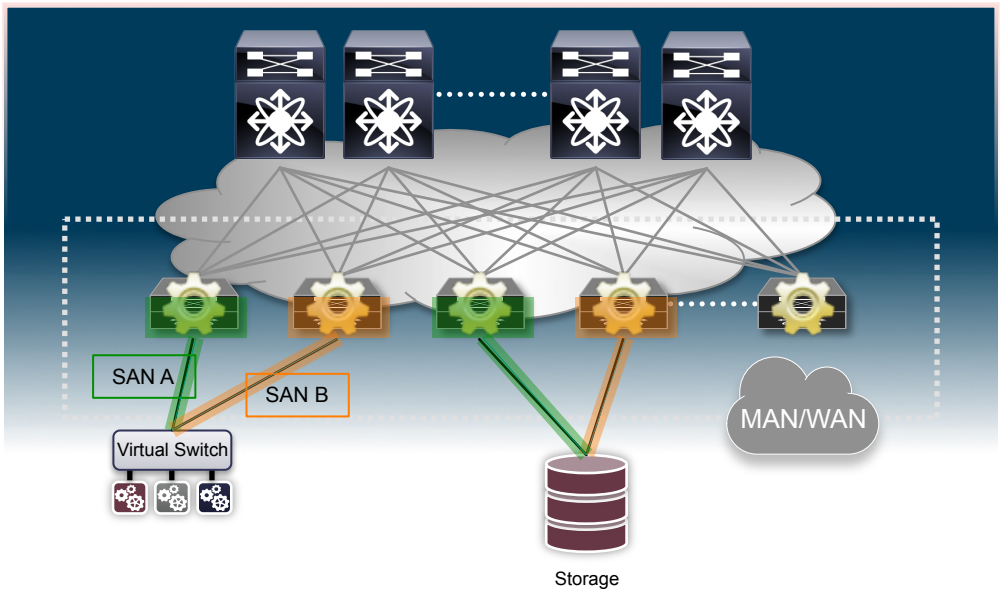


© 2014 Cisco and/or its affiliates. All rights reserved.

Dynamic FCoE

How It Works

- Dynamic FCoE modifies the way we see storage network
  - It uses FCoE as an “overlay” on top of Ethernet forwarding technology
  - Using Equal-Cost Multipathing (ECMP) capabilities
  - Leaf Switch can capitalize bandwidth to each Spine for east-west traffic
- Increases resiliency and robustness to our storage networks
  - Every Link can be used for FCoE

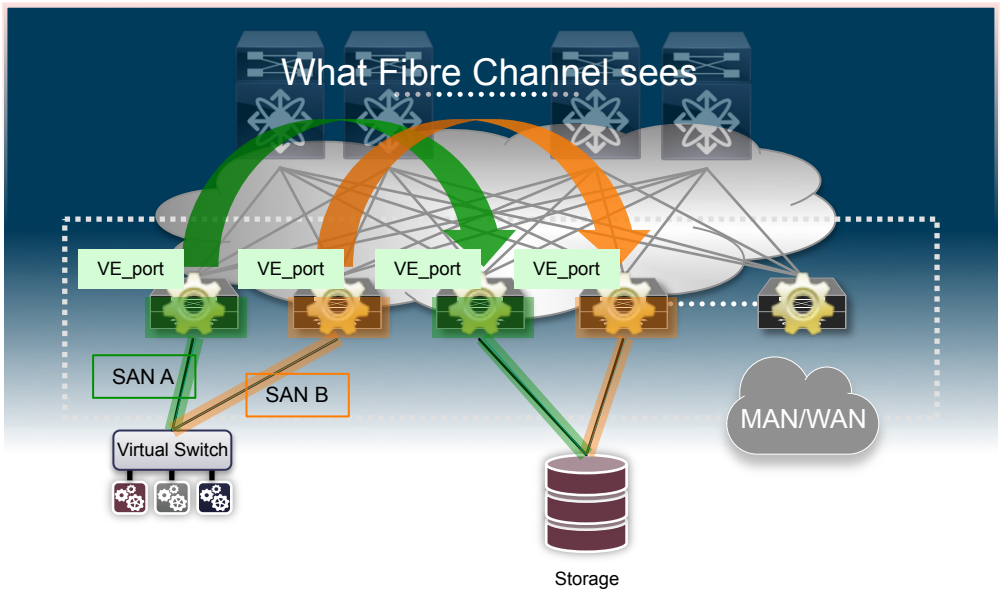


© 2014 Cisco and/or its affiliates. All rights reserved.

Dynamic FCoE

SAN A/B Separation

- Dynamic FCoE, physical SAN A/B separation occurs at the edge layer
- VE_Ports are automatically created and discovered across all the Spines to the Storage Leaf
- Traffic load-balanced across all the spines
  - FCF Leaf does not lose connectivity to destination storage Leaf
  - Increased resiliency



 © 2014 Cisco and/or its affiliates. All rights reserved.

Dynamic FCoE

Summary

- Continue to use standards-based solutions to improve Data Center performance for compute, network and storage

AND

- Dynamically create the individual component pieces for storage ISLs
- Dynamically create the appropriate port types
- Dynamically discover and initialize new Leafs when they come online
- Increase our bandwidth efficiency and reduce our failure domains
- Improve our economies of scale by not overprovisioning “just-in-case” strategies
- “Bolt-on” new topological considerations into existing, Classical FC/FCoE storage environments
  - Increase the availability through better architectures
  - Increase our overall bandwidth capability through the use of 40G and 100GbE

© 2014 Cisco and/or its affiliates. All rights reserved.

Simplified Underlay

- No IP addressing on Point-2-Point connections*
- No ARP flooding on Underlay
- Optimized Multi-Destination Topology for Scale and Convergence

Optimized Overlay

- FabricPath/VXLAN with distributed Default-Gateway
- End Host Discovery and Distribution (aka Control-Plane)
- Minimized Flood & Learn across Overlay

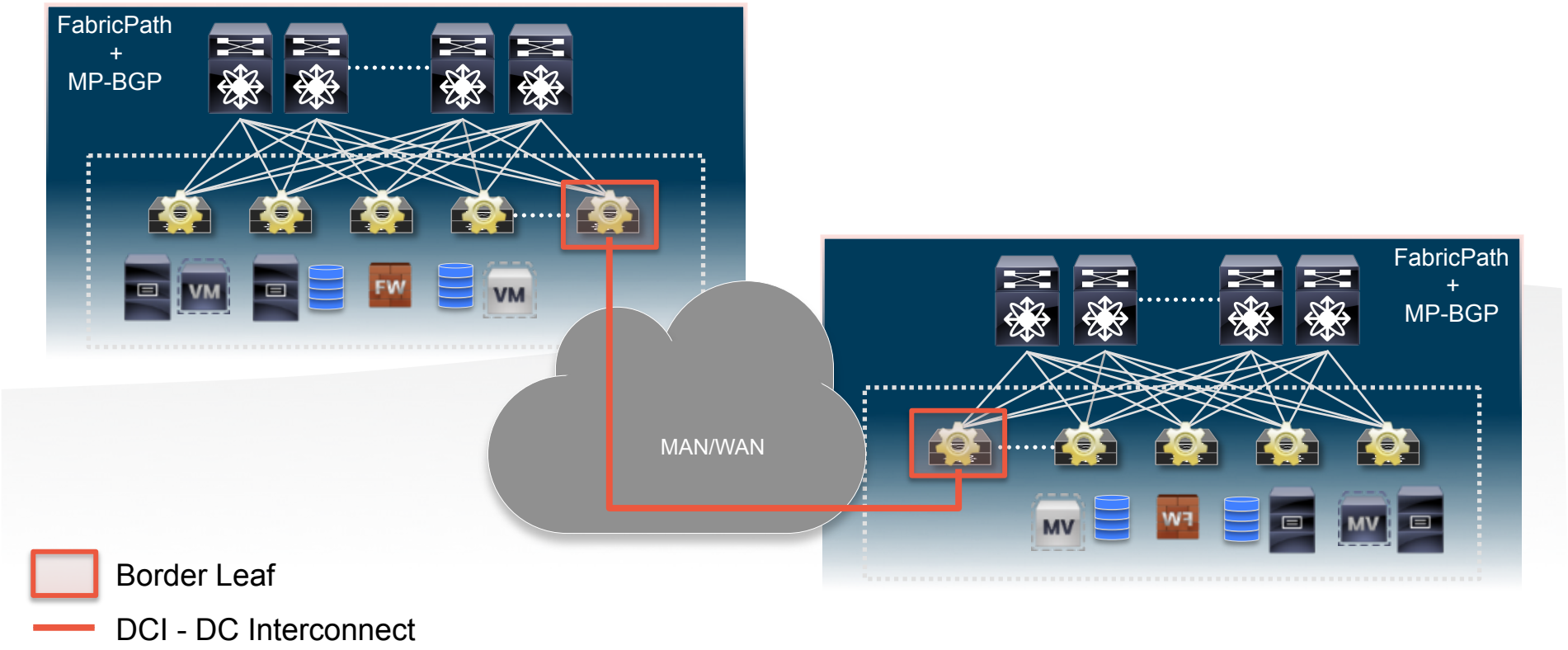
Automated Configuration

- Auto-Configuration of Tenant and Network
- DCNM aided Fabric Management and Troubleshooting

*Related to VXLAN Layer-3 Underlay (IP Un-Numbered)

© 2014 Cisco and/or its affiliates. All rights reserved.

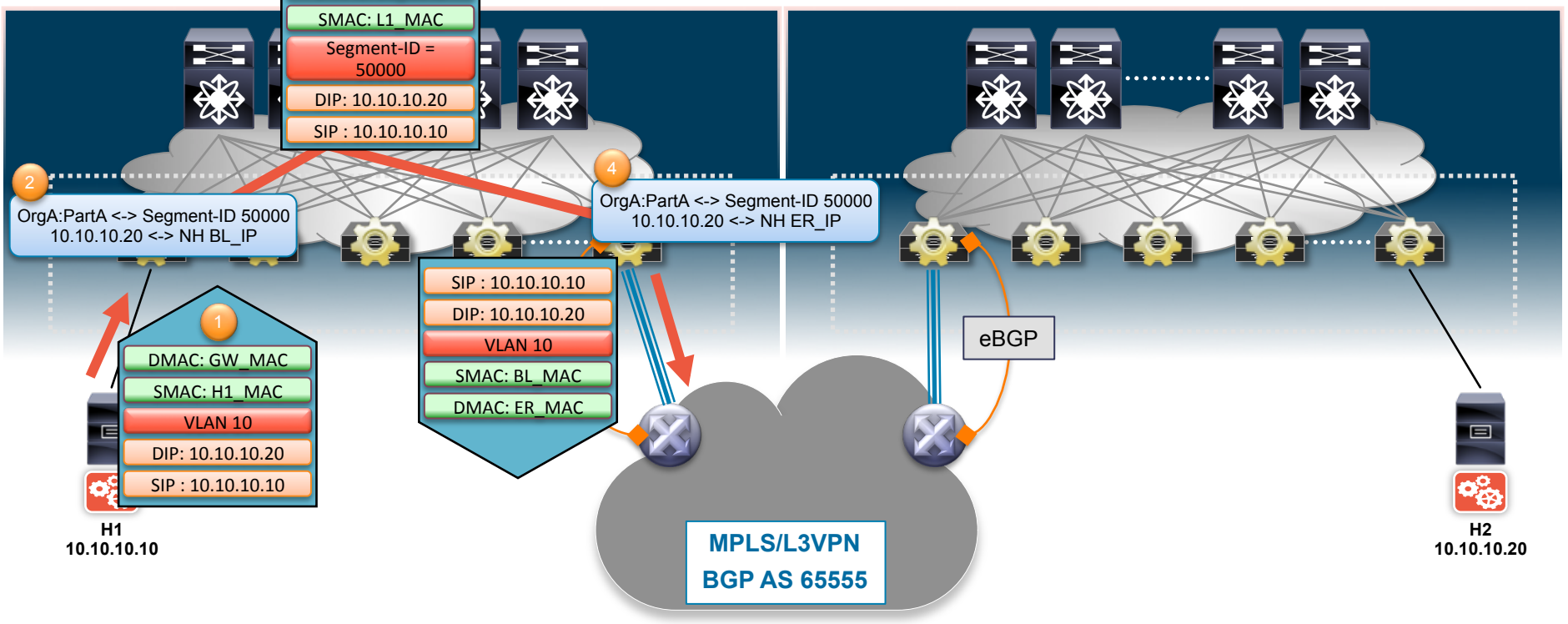
Border-Leaf and Data Center Interconnect (DCI)



 © 2014 Cisco and/or its affiliates. All rights reserved.

IP Forwarding within same Subnet across DCI

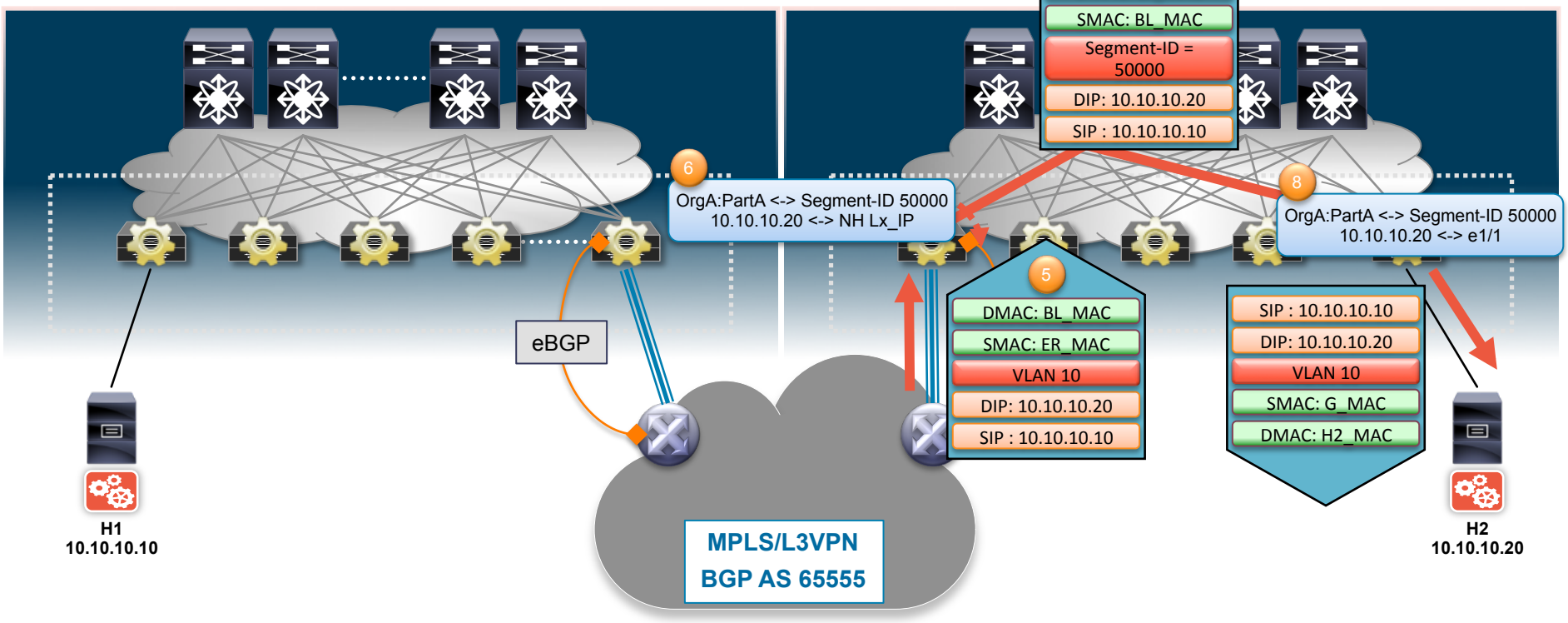
Enhanced F… proxy-gateway) & Layer-3 based DCI (e.g. MPLS L3VPN)



 © 2014 Cisco and/or its affiliates. All rights reserved.

IP Forwarding within same Subnet ac[illegible] DCI

Enhanced Forwarding (proxy-gateway) & Layer-3 based [illegible]MPLS L3VPN)



 © 2014 Cisco and/or its affiliates. All rights reserved.

IP Forwarding within same Subnet across DCI

Enhanced Forwarding (proxy-gateway) & Layer-3 based DCI (e.g. MPLS L3VPN)

- In the previous packet-walk the Source and Destination VLAN, Segment-IDs were the same in both Fabrics

Fabric	Subnet	VLAN	Layer-2 Segment-ID	VRF	Layer-3 Segment-ID
#1	10.10.10.0/24	10	30000	OrgA:PartA	50000
#2	10.10.10.0/24	10	30000	OrgA:PartA	50000

- The Source and Destination VLAN, Segment-IDs can be different, as traffic is crossing the Fabric boundary via the Border-Leaf

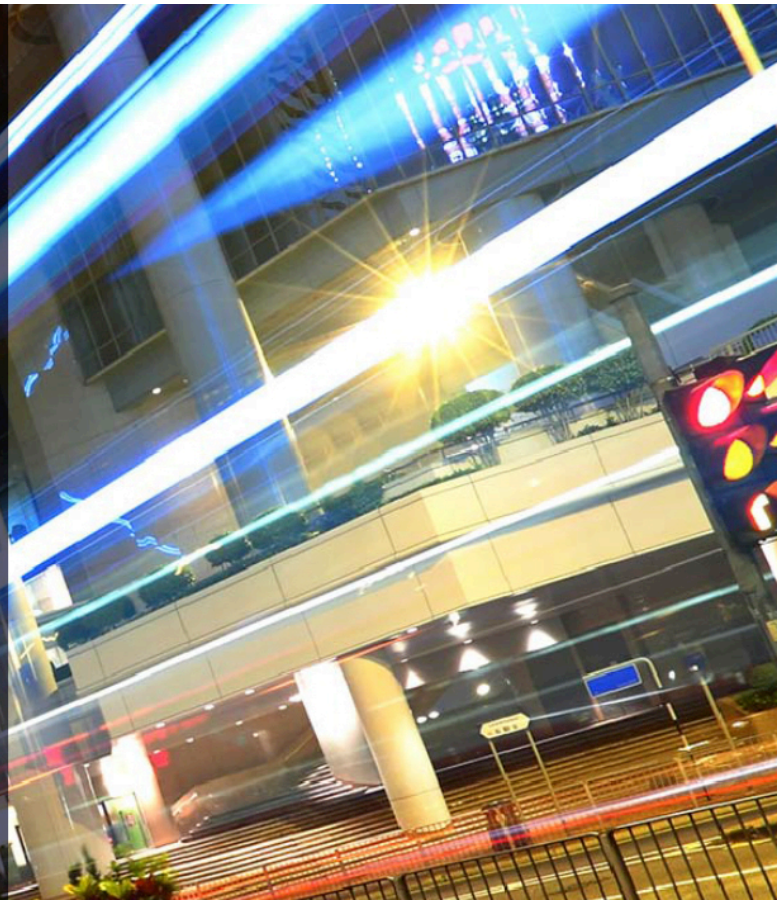
Fabric	Subnet	VLAN	Layer-2 Segment-ID	VRF	Layer-3 Segment-ID
#1	10.10.10.0/24	10	30000	OrgA:PartA	50000
#2	10.10.10.0/24	99	30001	OrgX:PartX	50001

- VLANs are Switch local significant and can vary at every stage of IEEE 802.1q encapsulation

© 2014 Cisco and/or its affiliates. All rights reserved.

Agenda

- Requirements and Functions
- Building Blocks
- Optimized Network
- **Virtual Fabric**
  - Segment ID
  - Packet Walk
- Workload Automation

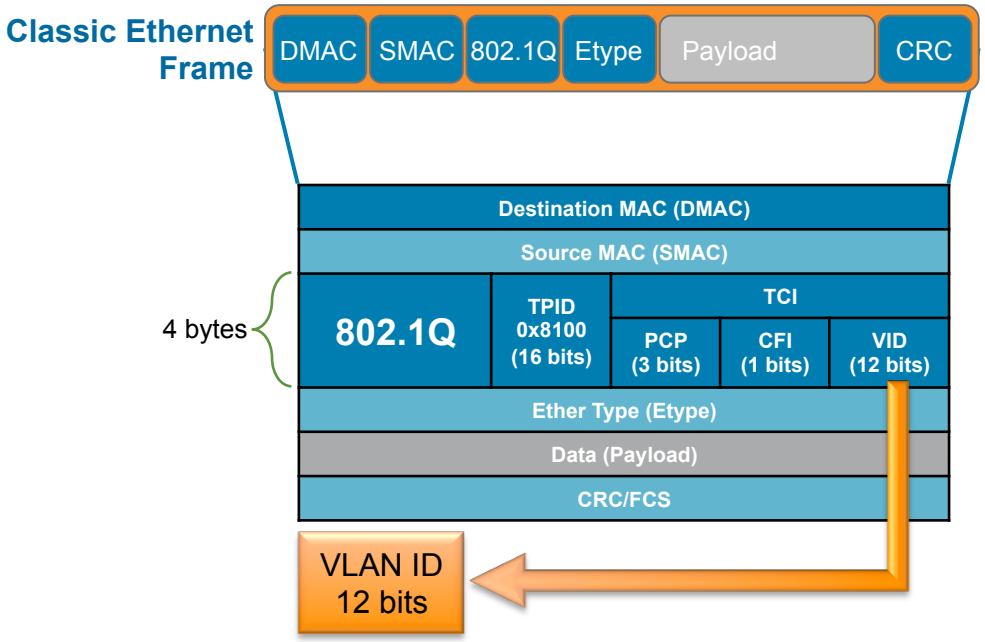


© 2014 Cisco and/or its affiliates. All rights reserved.

Virtual Fabrics

Classic Ethernet IEEE 802.1Q Format

- Traditionally VLAN space is expressed using 12 bits (802.1Q tag)
  - Limits the maximum number of segments in a Data Center to 4096 VLANs

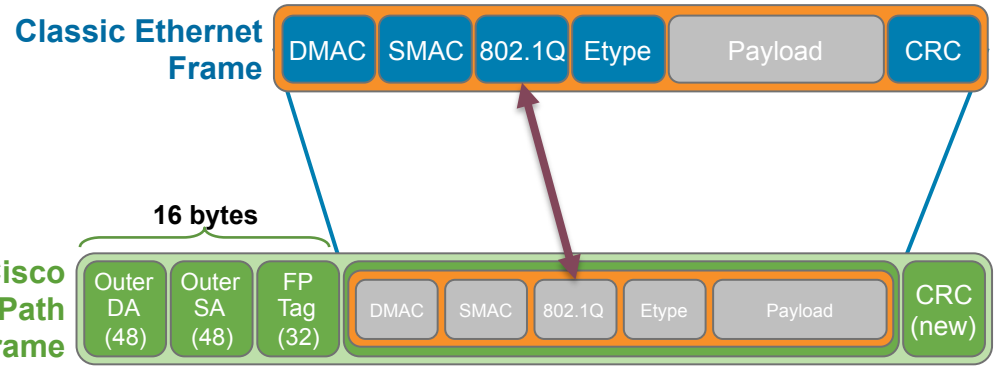


TPID = Tag Protocol Identifier, TCI = Tag Control Information, PCP = Priority Code Point, CFI = Canonical Format Indicator, VID = VLAN Identifier

Virtual Fabrics

Introducing FabricPath Encapsulation

- Cisco FabricPath introduces Classic Ethernet Frame Encapsulation by adding 16 bytes
- Traditionally VLAN space, expressed over IEEE 802.1Q tag, stays the same

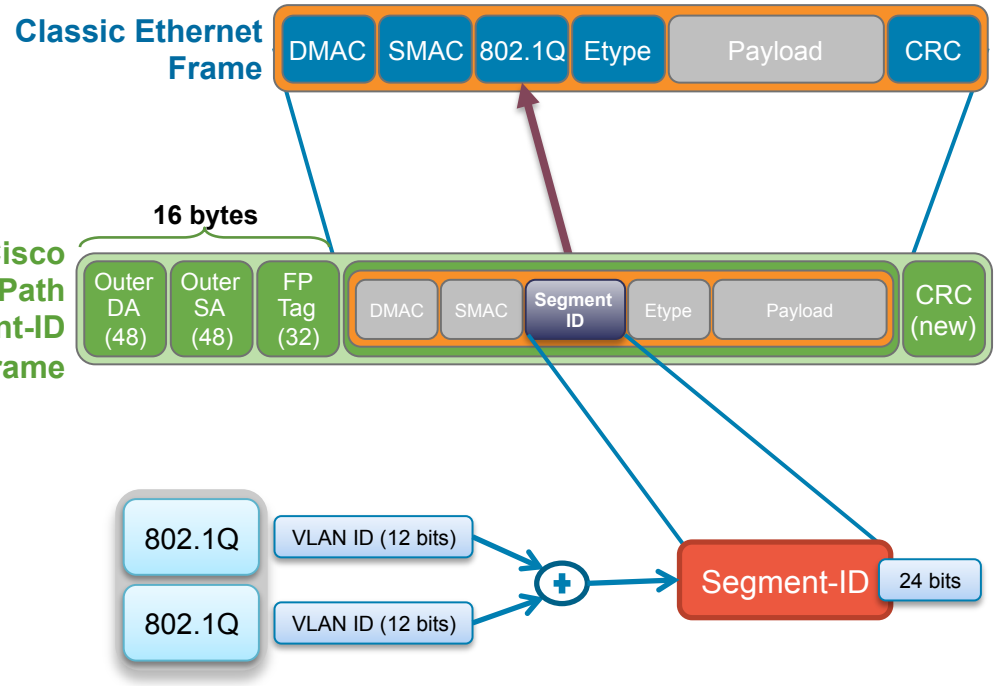


 © 2014 Cisco and/or its affiliates. All rights reserved.

Virtual Fabrics

Introducing Segment-ID Support

- Cisco FabricPath introduces Classic Ethernet Frame Encapsulation by adding 16 bytes
- Traditionally VLAN space, expressed over IEEE 802.1Q tag, stays the same
- The Fabric leverages a double 802.1Q tag for a total address space of 24 bits
  - Support of ~16M segments
- Segment-ID is a hardware-based innovation offered by Leaf and Spine nodes that are part of the Fabric

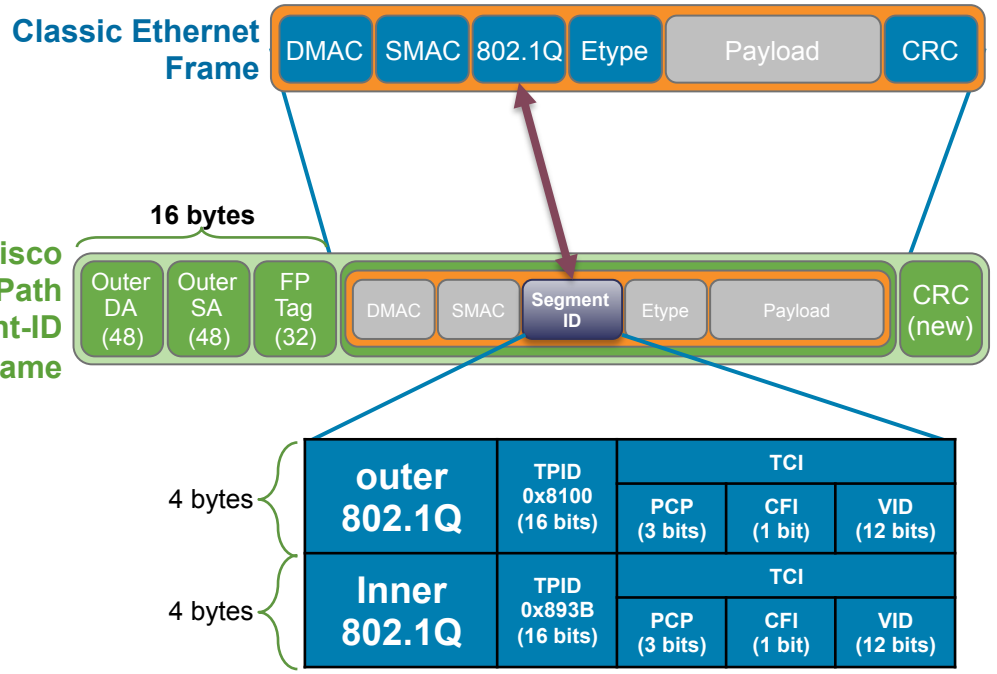


© 2014 Cisco and/or its affiliates. All rights reserved.

Virtual Fabrics

Introducing Segment-ID Support

- Cisco FabricPath introduces Classic Ethernet Frame Encapsulation by adding 16 bytes
- Traditionally VLAN space, expressed over IEEE 802.1Q tag, stays the same
- The Fabric leverages a double 802.1Q tag for a total address space of 24 bits
  - Support of ~16M segments
- Segment-ID is a hardware-based innovation offered by Leaf and Spine nodes that are part of the Fabric

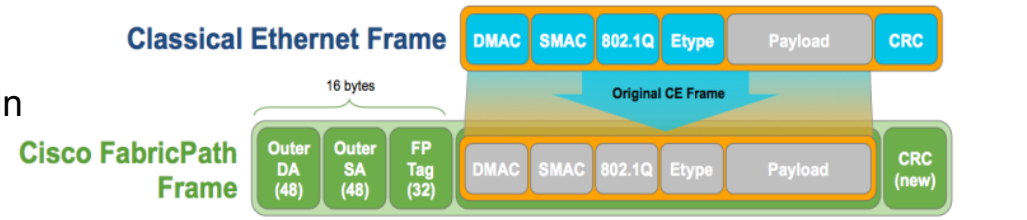


TPID 0x8100 = VLAN Tagged Frame with IEEE 802.1Q / TPID 0x893B = TRILL Fine Grained Labeling

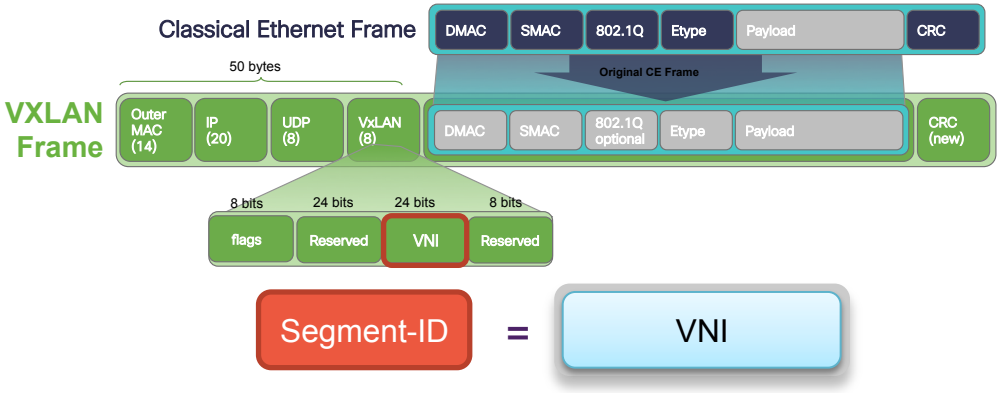
Virtual Fabrics

Introducing Segment-ID Support

- Traditionally VLAN space is expressed over 12 bits (802.1Q tag)
  - Limits the maximum number of segments in a Data Center to 4096 VLANs



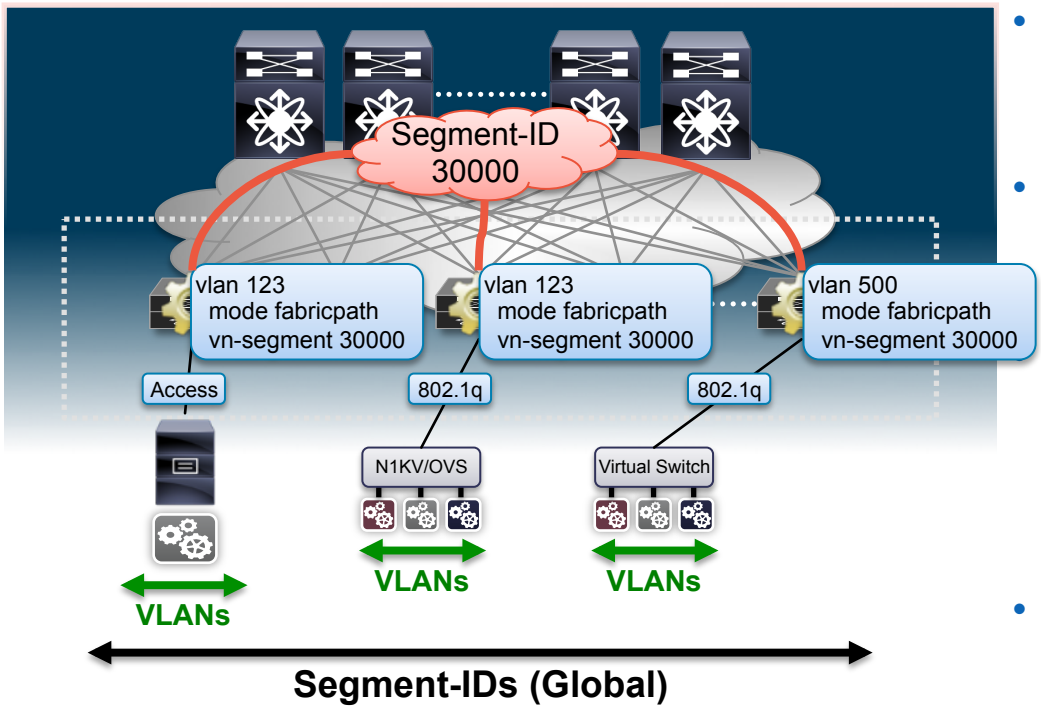
- The Fabric leverages the VNI field with a total address space of 24 bits
  - Support of ~16M segments
- Segment-ID (VNI) is part of the VXLAN header and supported in hardware by Leaf and Spine nodes that are part of the Fabric



 © 2014 Cisco and/or its affiliates. All rights reserved.

Virtual Fabrics

802.1Q Tagged Traffic to Segment-ID Mapping



- Segment-IDs are utilized for providing isolation at Layer-2 and Layer-3 across the Fabric
- 802.1Q tagged frames received at the Leaf nodes from edge devices must be mapped to specific Segments

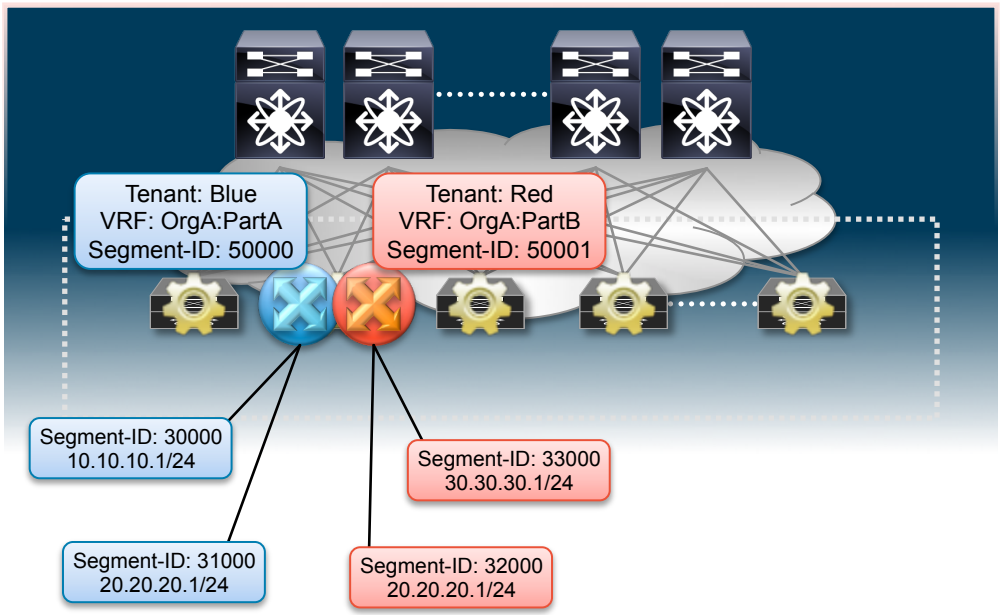
  The VLAN-Segment mapping can be performed on a leaf device level
  - VLANs become locally significant on the leaf node and is 1:1 mapped to a Segment-ID
- Segment-IDs are globally significant, VLAN IDs are locally significant

© 2014 Cisco and/or its affiliates. All rights reserved.

Virtual Fabrics

How are Segment-IDs Utilized?

- Each IP Subnet, defined at the edge of the Fabric is associated to a Layer-2 domain, which is represented by a Segment-ID
- A Segment-ID will also be used to uniquely identify a VRF within the Fabric
- Multiple Layer-2 domains can be defined for a given Tenant and are mapped to a Layer-3 VRF

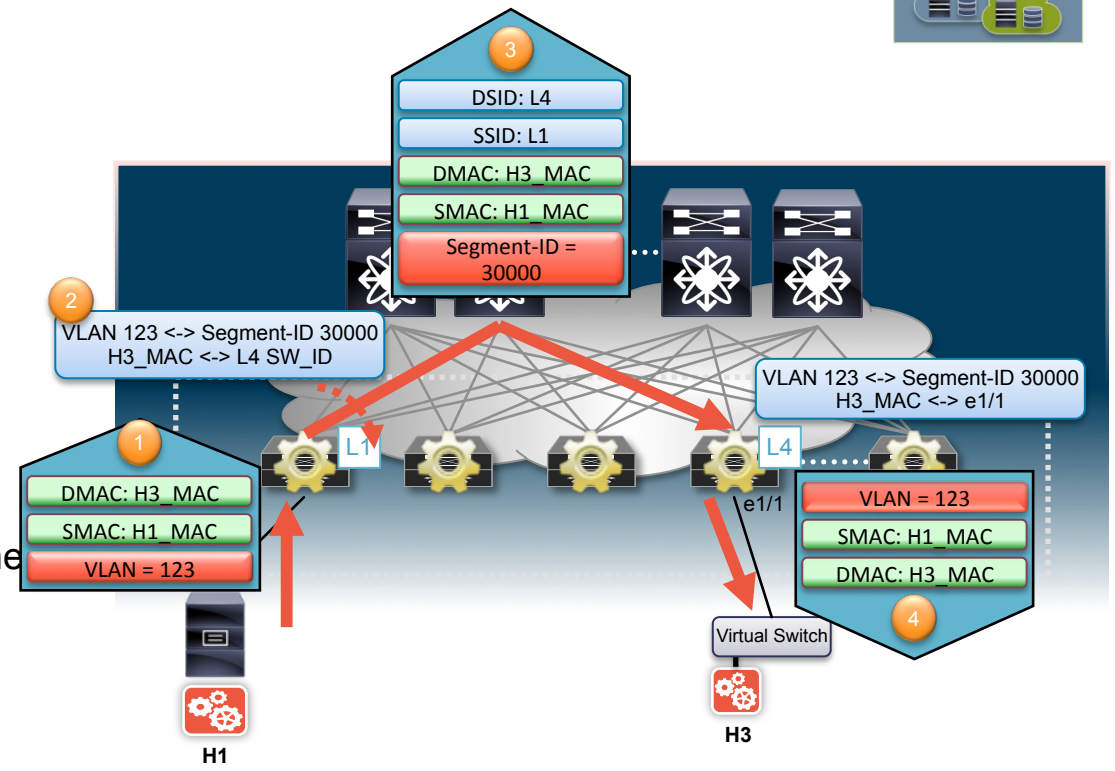


 © 2014 Cisco and/or its affiliates. All rights reserved.

Virtual Fabrics

Layer-2 non IP Flows

1. H1 sends a packet to H3. Traffic between the Server and the Leaf is tagged with a **local VLAN-ID 123**
2. Layer-2 lookup is performed by Leaf1 (L1) in the MAC Table for the Segment-ID associated to VLAN 123 (30000)
3. Leaf1 (L1) adds the Layer-2 and FabricPath headers before sending the packet to the fabric. The Segment-ID associated to VLAN 123 is added to the header.
4. Leaf4 (L4) receives the frame and performs the Layer-2 lookup by looking at the Segment-ID value. It then sends it to H3 using a **local VLAN-ID 123**

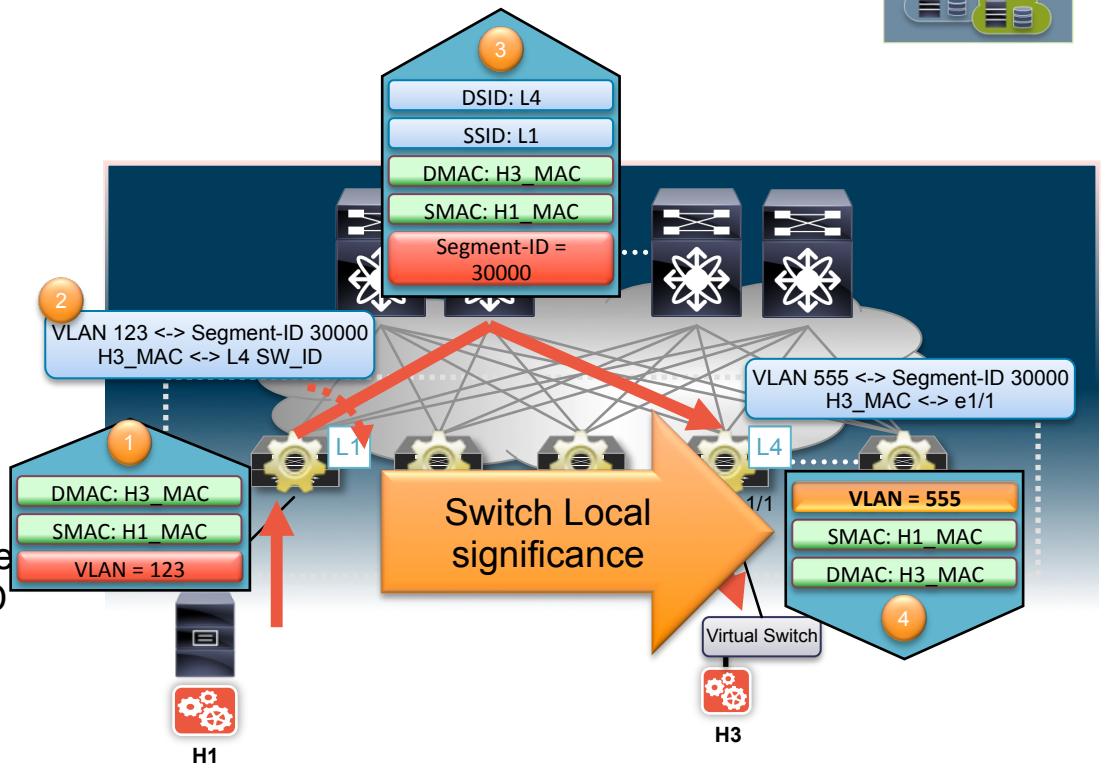


 © 2014 Cisco and/or its affiliates. All rights reserved.

Virtual Fabrics

Layer-2 non IP Flows

1. H1 sends a packet to H3. Traffic between the Server and the Leaf is tagged with a **local VLAN-ID 123**
2. Layer-2 lookup is performed by Leaf1 (L1) in the MAC Table for the Segment-ID associated to VLAN 123 (30000)
3. Leaf1 (L1) adds the Layer-2 and FabricPath headers before sending the packet into the fabric. The Segment-ID associated to VLAN 123 is added to the header
4. Leaf4 (L4) receives the frame and performs the Layer-2 lookup by looking at the Segment-ID value. It then sends it to H3 using a **local VLAN-ID 555**



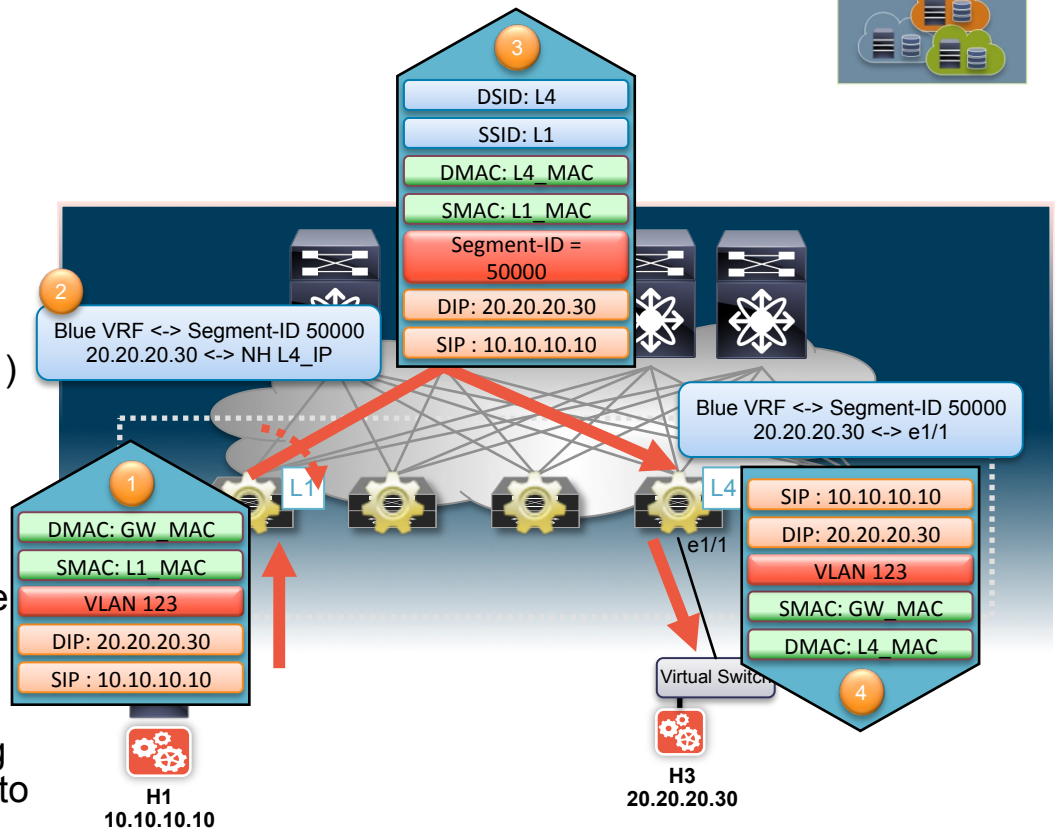
 © 2014 Cisco and/or its affiliates. All rights reserved.

Virtual Fabrics

Fabric Routed Flows

1. H1 sends a packet to H3; traffic between the Server and the Leaf is tagged with a **local VLAN-ID 123**
2. Layer-3 lookup is performed by Leaf1 (L1) in the context of the **BLUE VRF**
3. Leaf1 (L1) adds the Layer-2 and FabricPath headers before sending the packet into the fabric. The Segment-ID identifying the **BLUE VRF** is added inside the Layer-2 header
4. Leaf4 (L4) receives the frame and associates it to the **BLUE VRF** by looking at the Segment-ID value. It then sends it to H3 using a **local VLAN-ID 123**

Note: this behavior applies to all Fabric routed flows (intra-subnet or inter-subnet)



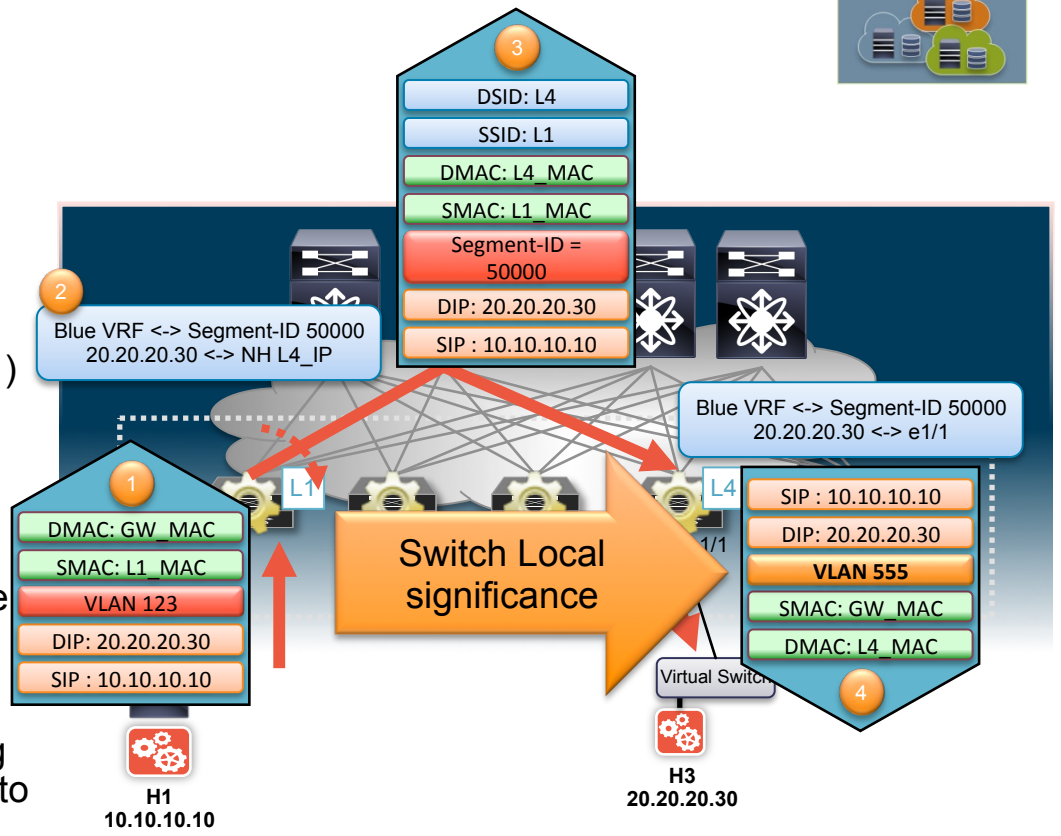
 © 2014 Cisco and/or its affiliates. All rights reserved.

Virtual Fabrics

Fabric Routed Flows

1. H1 sends a packet to H3 traffic between the Server and the Leaf is tagged with a **local VLAN-ID 123**
2. Layer-3 lookup is performed by Leaf1 (L1) in the context of the **BLUE VRF**
3. Leaf1 (L1) adds the Layer-2 and FabricPath headers before sending the packet into the fabric. The Segment-ID identifying the **BLUE VRF** is added inside the Layer-2 header
4. Leaf4 (L4) receives the frame and associates it to the **BLUE VRF** by looking at the Segment-ID value. It then sends it to H3 using a **local VLAN-ID 555**

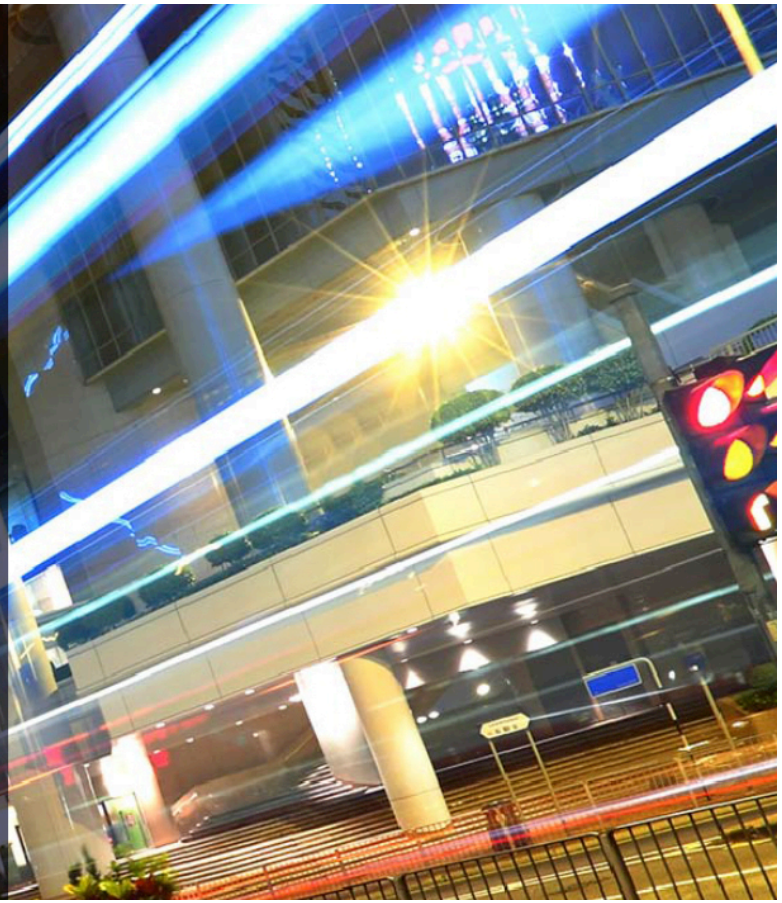
Note: this behavior applies to all Fabric routed flows (intra-subnet or inter-subnet)



 © 2014 Cisco and/or its affiliates. All rights reserved.

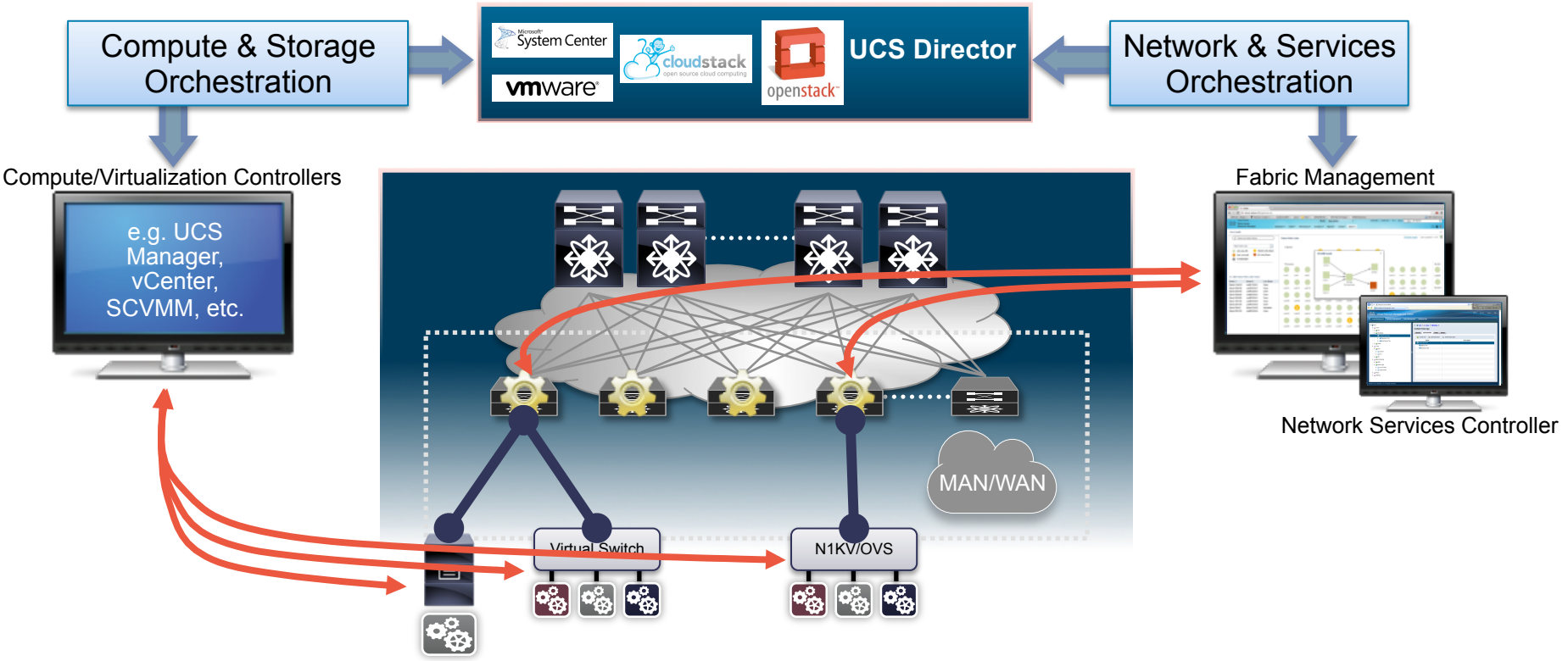
Agenda

- Requirements and Functions
- Building Blocks
- Optimized Network
- Virtual Fabric
- **Workload Automation**
  - Workload Deployment with REST API
  - Auto-Config PULL for Physical and Virtual Machines



© 2014 Cisco and/or its affiliates. All rights reserved.

Workload Automation & Open Environment

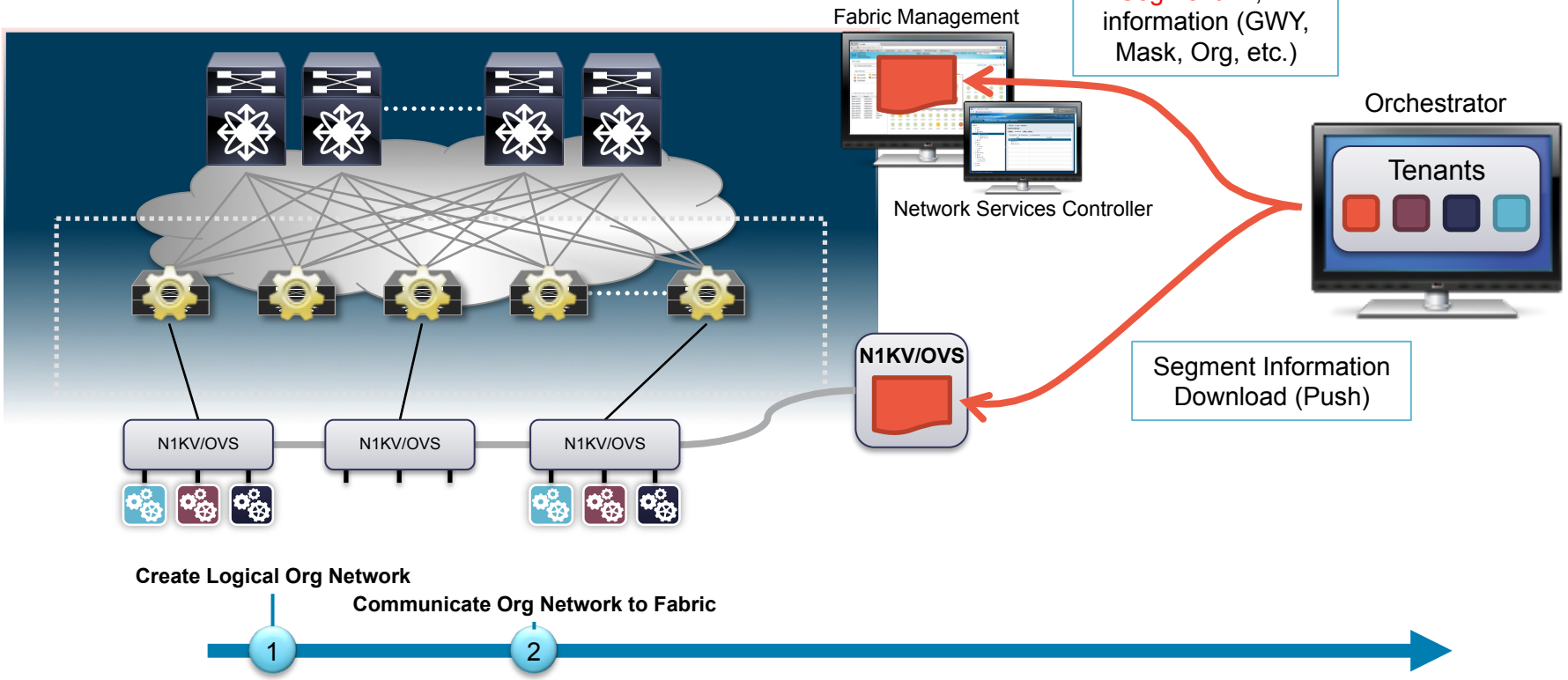


*VDP (VSI Discovery and Configuration Protocol) is IEEE 802.1Qbg Clause 41

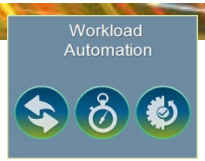
© 2014 Cisco and/or its affiliates. All rights reserved.

Workload Automation

Leveraging VDP for Leaf Auto-Configuration

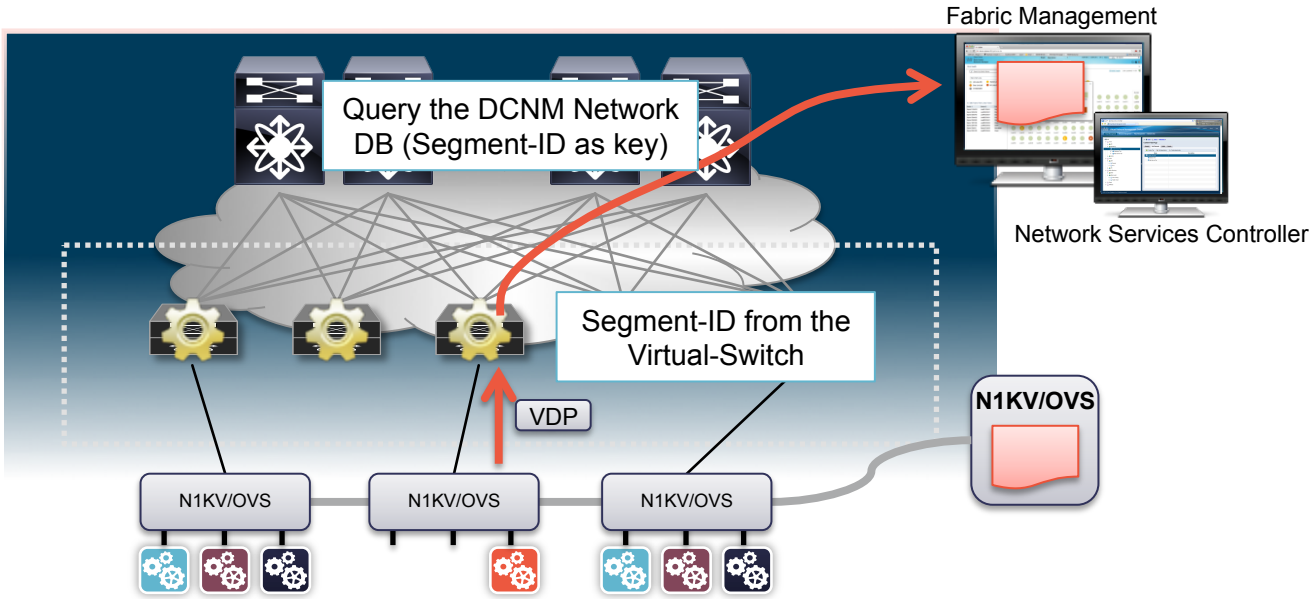


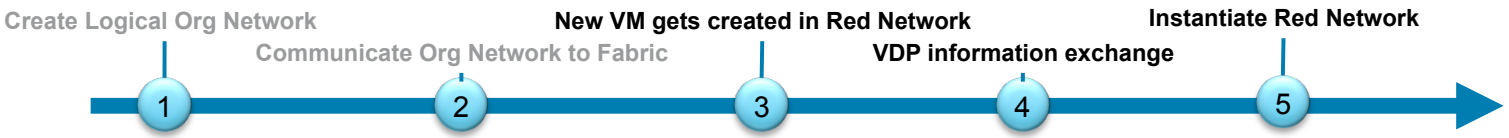
© 2014 Cisco and/or its affiliates. All rights reserved.



Workload Automation

Leveraging VDP for Leaf Auto-Configuration (2)

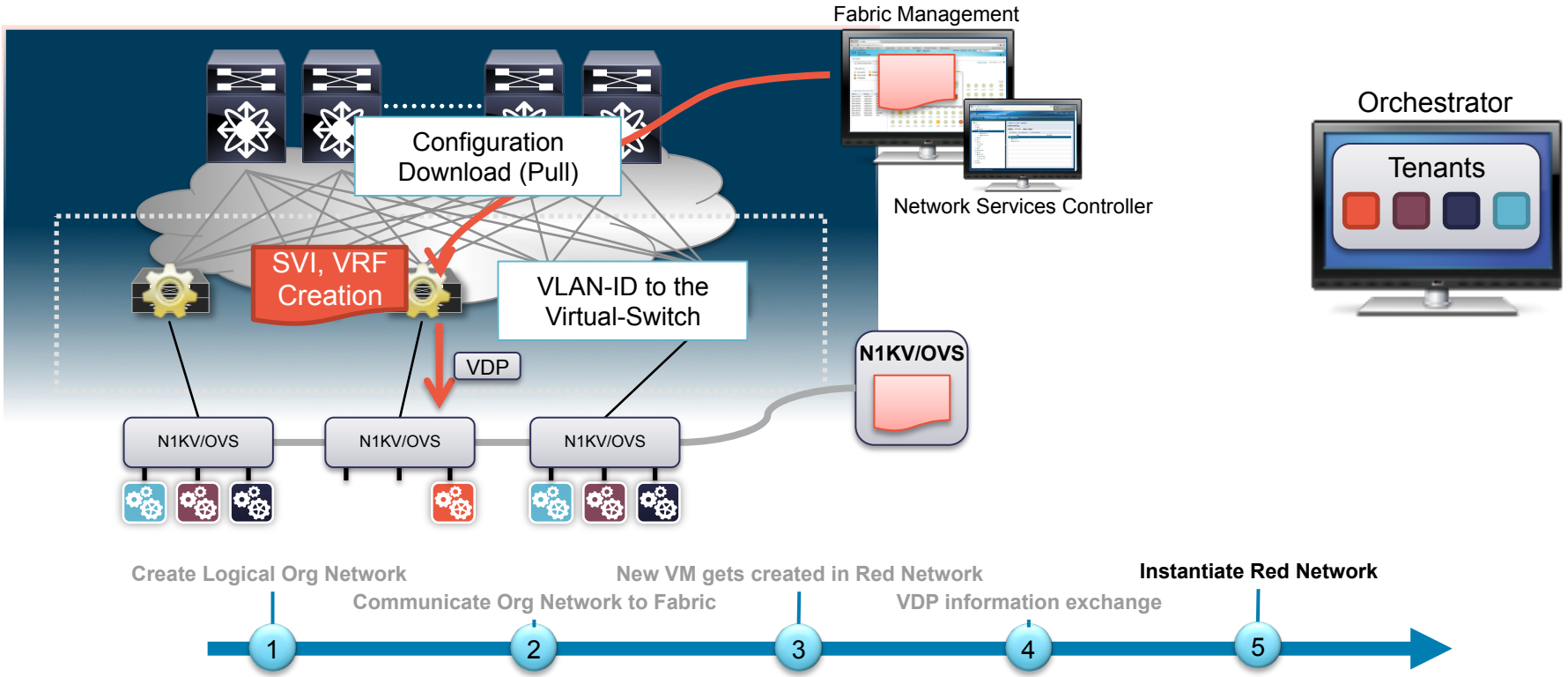




 © 2014 Cisco and/or its affiliates. All rights reserved.

Workload Automation

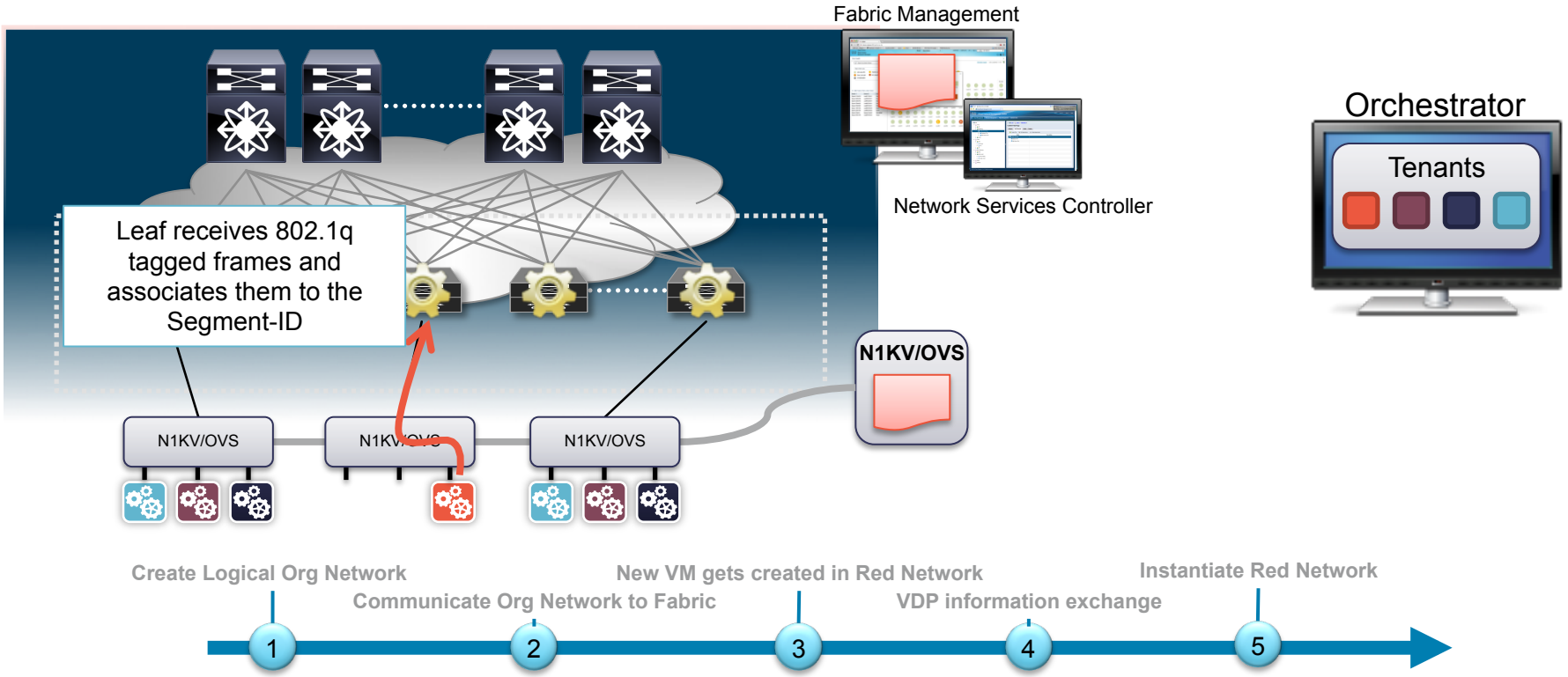
Leveraging VDP for Leaf Auto-Configuration (3)



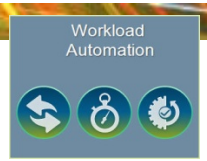
 © 2014 Cisco and/or its affiliates. All rights reserved.

Workload Automation

Leveraging VDP for Leaf Auto-Configuration (4)

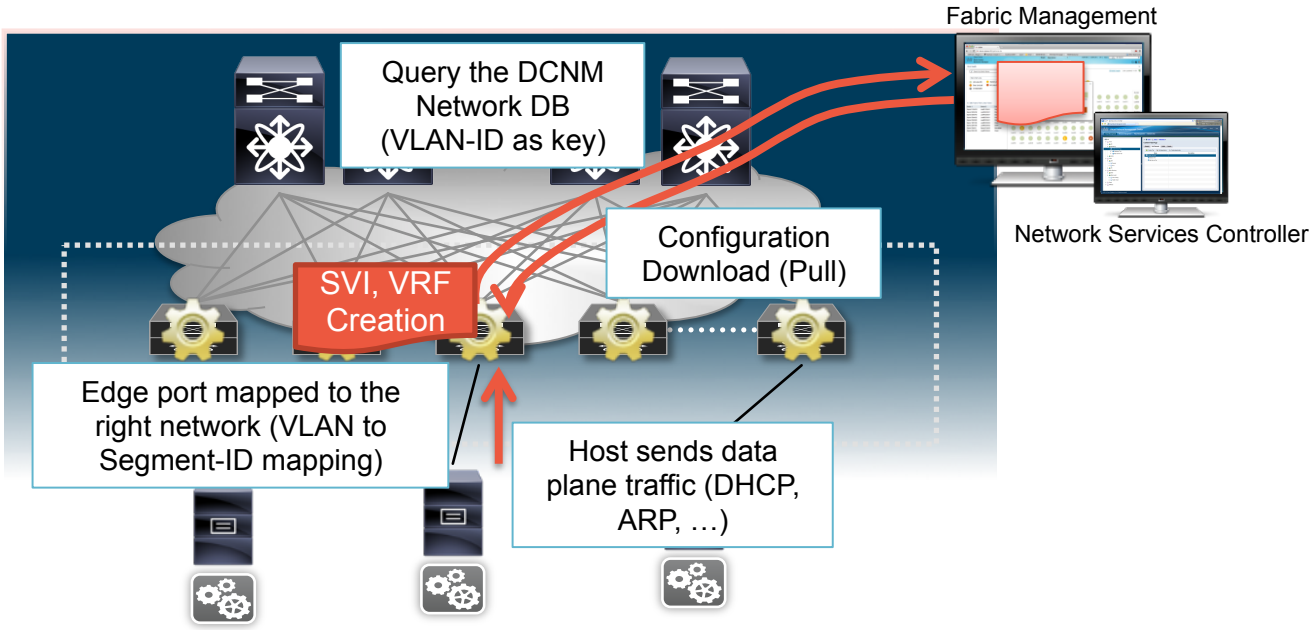


© 2014 Cisco and/or its affiliates. All rights reserved.



Workload Automation

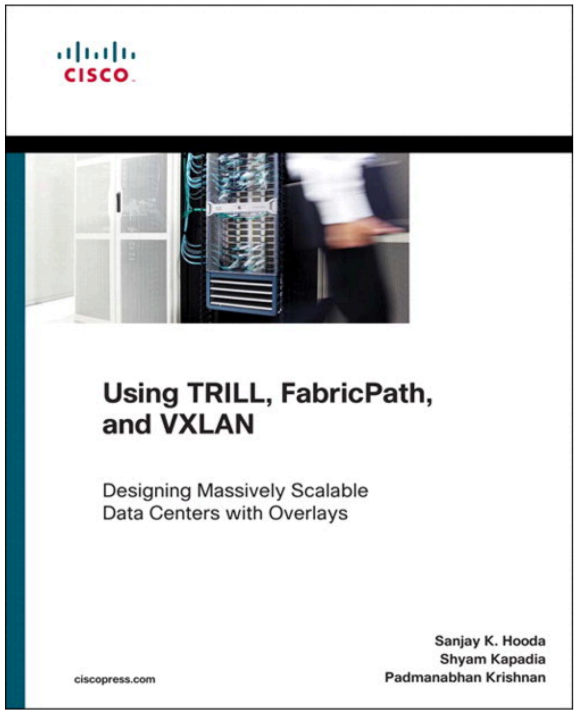
What about Auto-Configuration for Physical Hosts?



Note: For identifying the Switch or Port-Local VLAN namespace, we introduced the term “mobility-domain”. Data packet driven auto-config uses VLAN + Mobility Domain to download (pull) the configuration

© 2014 Cisco and/or its affiliates. All rights reserved.

Recommended Reading



Using TRILL, FabricPath, and VXLAN: Designing Massively Scalable Data Centers (MSDC) with Overlays

- Sanjay K. Hooda
- Shyam Kapadia
- Padmanabhan Krishnan

ISBN-10: 1-58714-393-3

ISBN-13: 978-1-58714-393-9

 © 2014 Cisco and/or its affiliates. All rights reserved.

Verified Scalability

Full Fabric

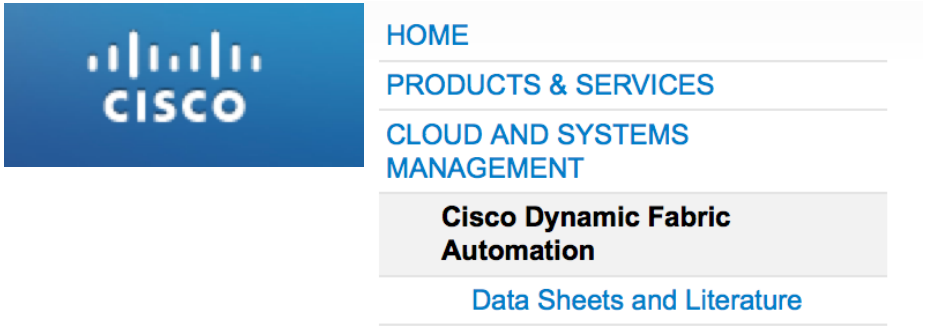
Feature		Verified Topology	Verified Maximum
Number of Spines	Multi-Level Tier tested and verified)	8	16
Number of Leaf	Concludes of Server, Services and Border Leaf	384	384
Number of Tenants	1 Tenant = 2 VRF	10’000	10’000
Number of VRF	Layer-3 Segments (VNI)	20’000	20’000
Number of Segments	Layer-2 Segments (vn-segment)	50’000	50’000
IPv4 Routes	/32 + Subnet Route	800’000	1.2 Million
IPv6 Routes	/128 + Subnet Route	192’000	384’000
Virtual Machines	Real VMs Deployed within Scale testing (1 VM has multiple vNIC)	12’000	300’000

© 2014 Cisco and/or its affiliates. All rights reserved.

Where to get more Information

CCO

- http://www.cisco.com/go/dfa



Cisco Community / Devnet

- https://communities.cisco.com/community/technology/datacenter/dfa
- https://developer.cisco.com/site/data-center/converged-infrastructure/dfa/index.gsp

 © 2014 Cisco and/or its affiliates. All rights reserved.


Thank you.