TOMORROW starts here.

Cisco Dynamic Fabric Automation Architecture

BRKDCT-2385

Lukas Krattiger

Technical Marketing Engineer

Agenda

- DFA Requirements and Functions
- Fabric Management
- Workload Automation
- Optimised Network
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Hardware Support

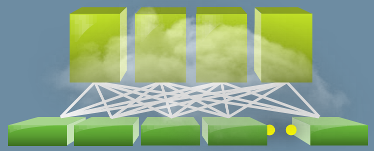
Dynamic Fabric Automation Architecture
Innovative Building Blocks

Bundled functions are modular and simplified for scale and automation

Fabric Management

Workload Automation

Optimised Network



Virtual Fabrics

Agenda

- DFA Requirements and Functions
- Optimised Network
    - Fabric Properties
    - Control Plane
    - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support

Today’s DC Challenges

Are the result of…

- Operational Complexity
- Architecture Rigidity
- Infrastructure Inefficiency

Dynamic
Fabric
Automation
Architecture

SIMPLIFY

Auto-configuration at scale

OPTIMISE

workload (physical/virtual) anywhere anytime

AUTOMATE

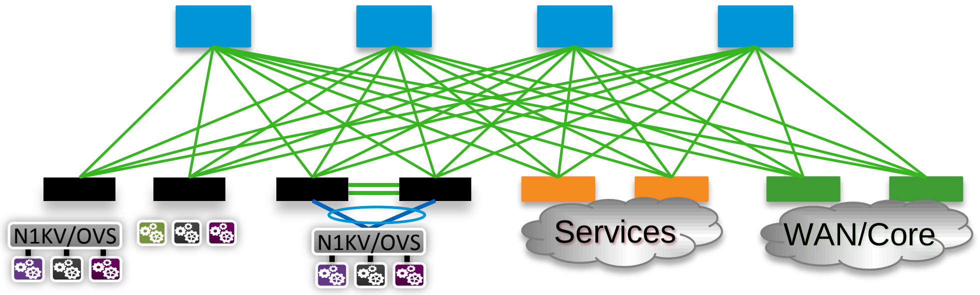
DFA Applicability and Use Cases

Cisco Dynamic Fabric Automation applies to any customer looking for solution to:

- ✓ DC Networks from the very small to the very large
- ✓ Environments with virtual and non-virtual workloads
- ✓ Looking to integrate with 3rd party Orchestration Tools
- ✓ Seeking flexibility on workload placement
- ✓ Looking for the Stability of small failure domains and flexibility or any app anywhere
- ✓ IPv4 and IPv6 aware Fabric technology

Dynamic Fabric Automation Architecture

Device Roles



- Spine
- Leaf
- Border Leaf
- Services Leaf
- Virtual Leaf* (N1KV/OVS)

Virtual Machines Physical Machines FEXs 3rd Party Switches UCS FIs Blade Switches Storage	Firewalls Load Balancers 3rd Party Appliances	Routers Switches 3rd Party Devices

Note: the different leaf roles are logical and not physical. The same leaf switch could perform all three functions (regular, services and border leaf)

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

Agenda

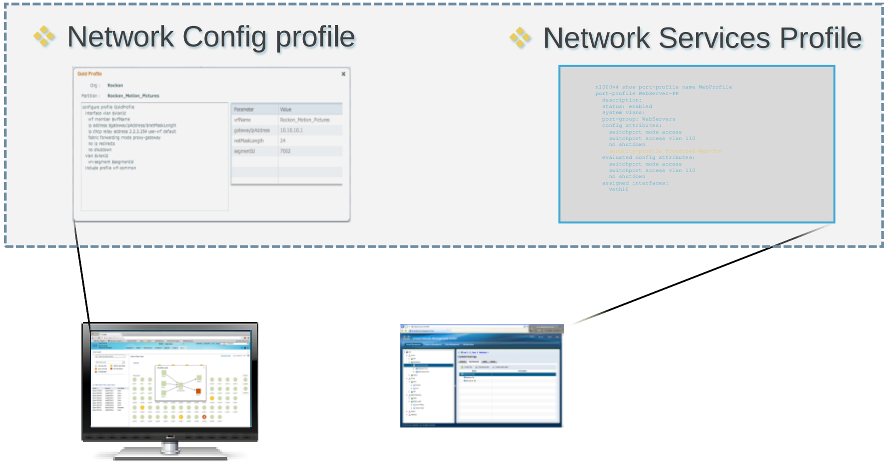
- DFA Requirements and Functions
- Optimised Network
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support

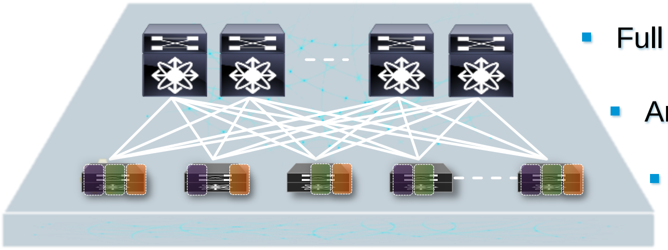
Cisco Dynamic Fabric Automation

Scale, Resiliency and Efficiency

Advantages

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



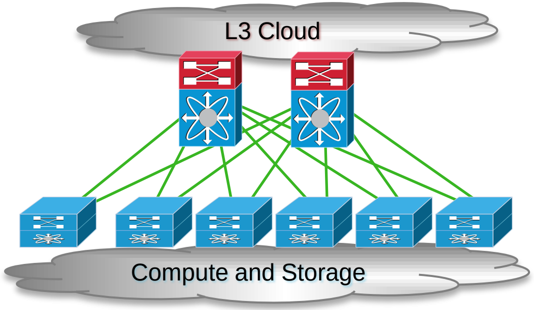


- Full bisectional bandwidth (N spines)
- Any/all Leaf Distributed Default Gateways
- Any/all subnets on any leaf

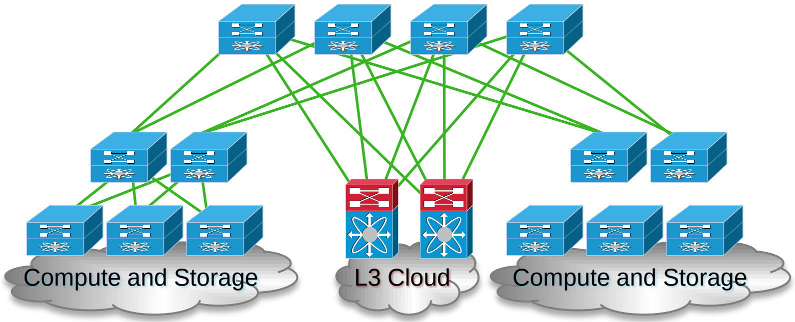
Cisco Dynamic Fabric Automation
Flexible Topologies Support

Optimized Network

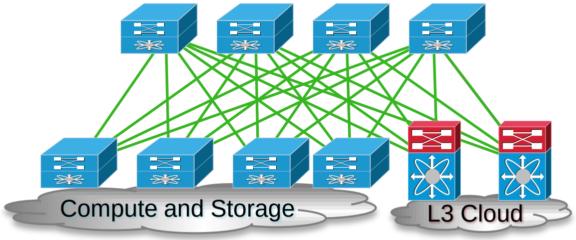
Traditional Access/Aggregation



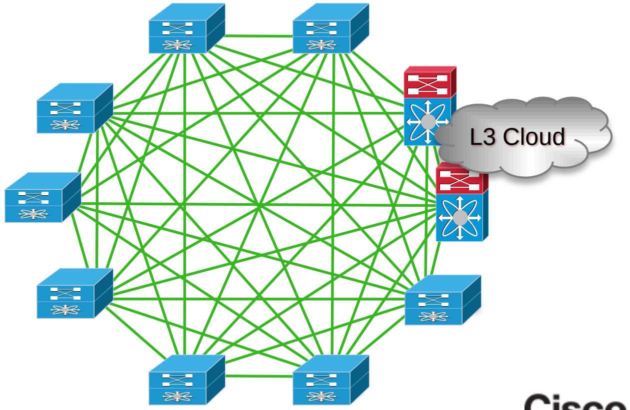
Three Tiers (Fat Tree)



Folded CLOS

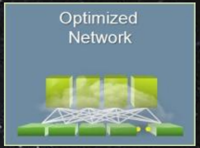


Full Mesh

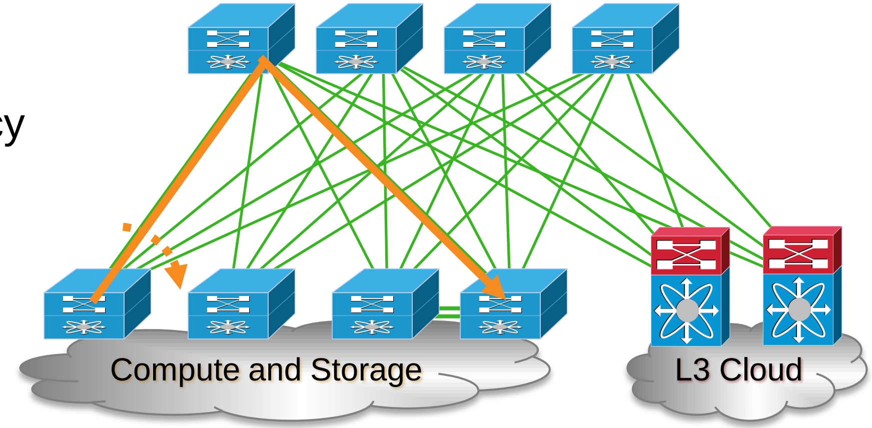


Cisco Dynamic Fabric Automation

Fabric Properties



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

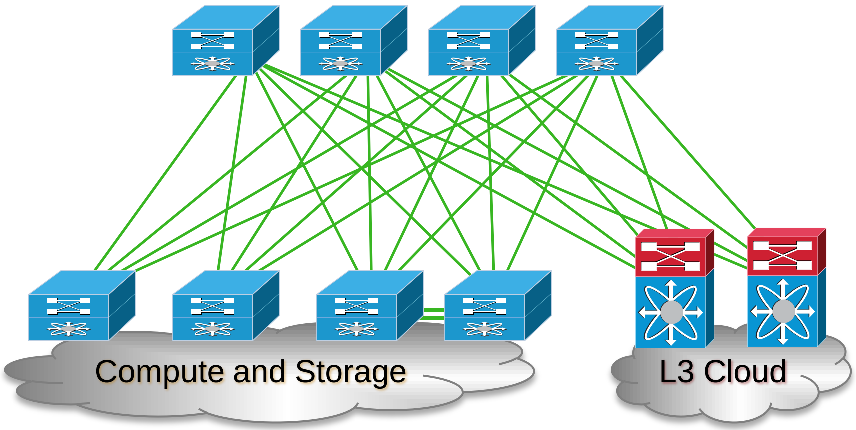


§§ Fabric properties applicable to all topologies §§

Cisco Dynamic Fabric Automation

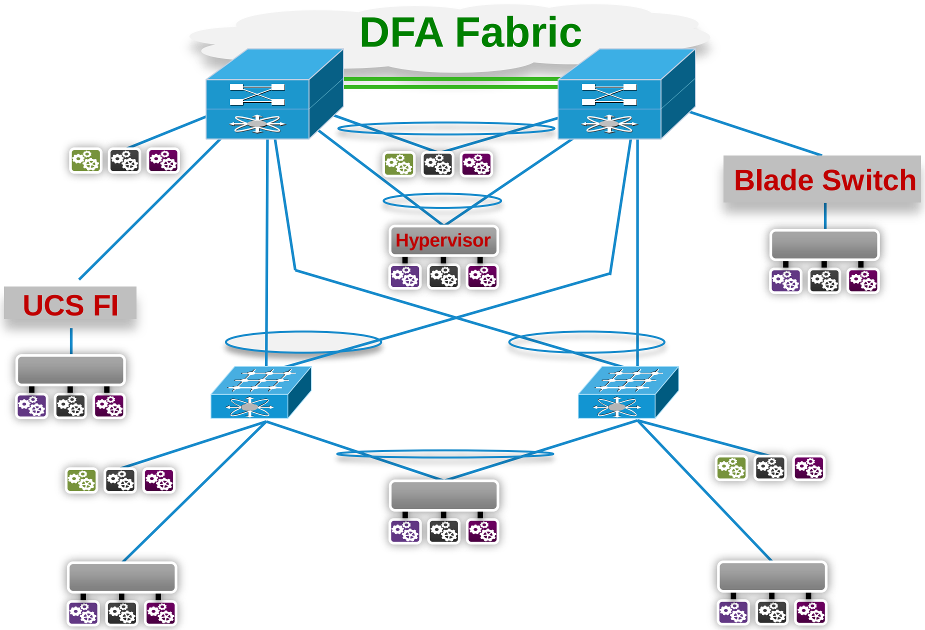
Variety of Fabric Sizes

- Fabric size: Hundreds to 10s of Thousands 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 need grows



Cisco Dynamic Fabric Automation

Variety of South-bound Topological 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 "remote linecards" and do not participate in DFA control plane and data plane encapsulation

Agenda

- DFA Requirements and Functions
- Optimised Network
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support

Control Plane

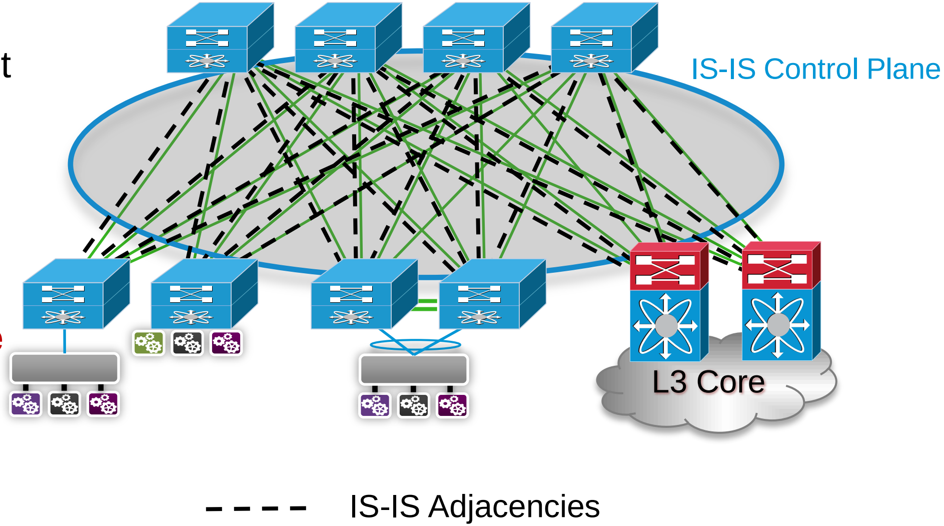
1 - IS-IS as Fabric Control Plane

ISIS for fabric link state distribution

- Fabric node reachability for overlay encap
- Building multi-destination trees for multicast and broadcast traffic
- Quick reaction to fabric link/node failure
- Enhanced for mesh topologies

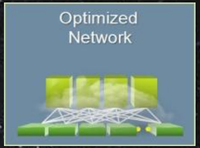
Fabric Control Protocol doesn’t distribute

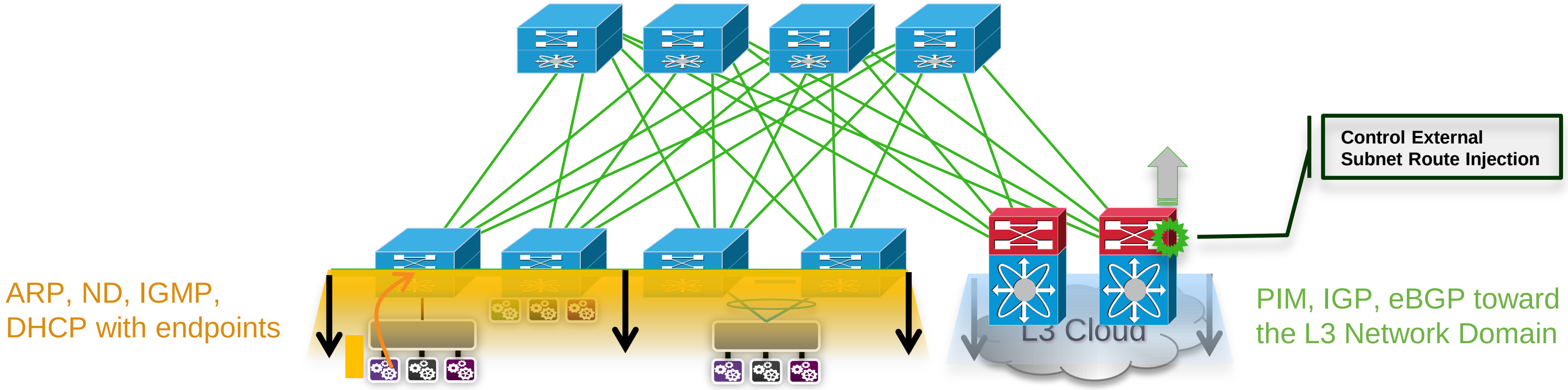
- Host Routes
- Host originated control traffic
- Server subnet information



Control Plane

2 – Host Originated Protocols Containment

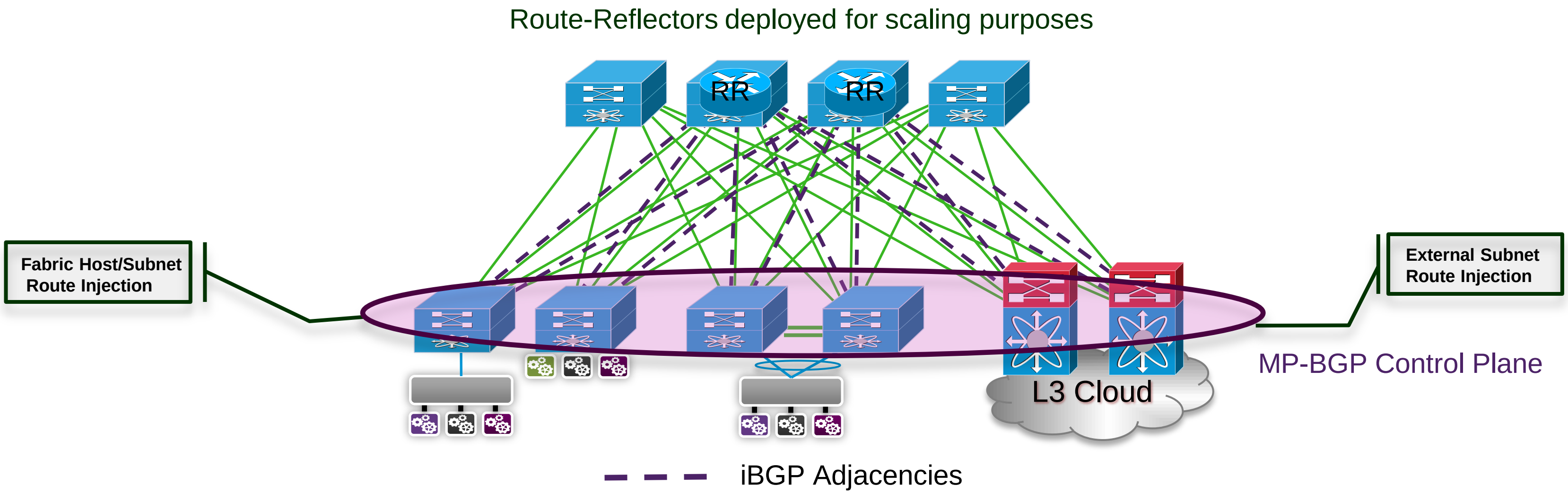




- ARP, ND, IGMP, DHCP originated on servers are terminated on Leaf nodes
- Contain floods and failure domains, distribute control packet processing
- Terminate PIM, OSPF, eBGP from external networks on Border Leafs

Control Plane

3 – 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 100s of thousands of routes and reduce convergence time

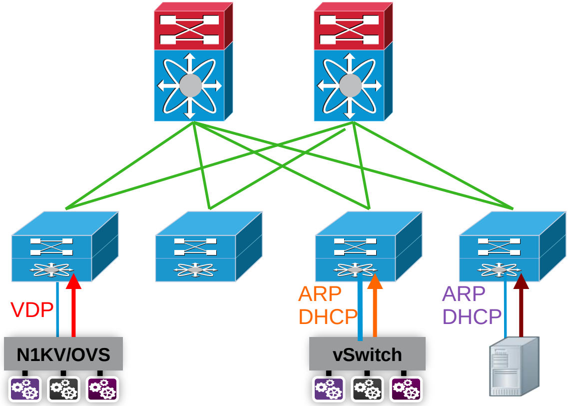
Control Plane

Hosts Detection and Deletion

- In order to advertise host reachability information, a leaf must discover first locally connected devices
- Detection of local hosts

  Based on VDP or ARP/DHCP
- Detection of remote hosts

  Received MP-BGP notifications

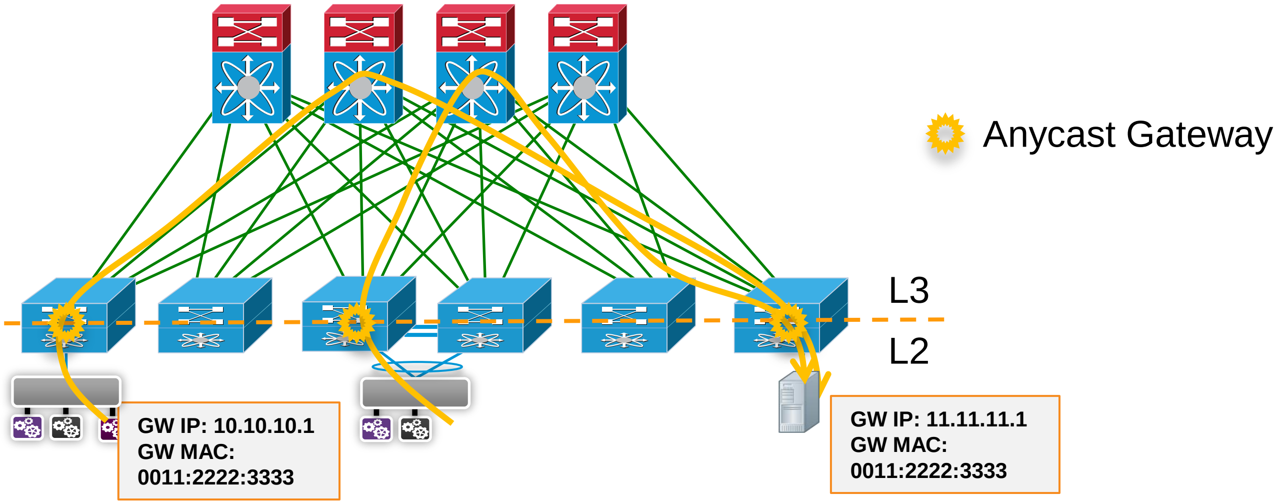


Agenda

- DFA Requirements and Functions
- Optimised Network
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support

Optimised Network

Distributed Gateway at the Leaf

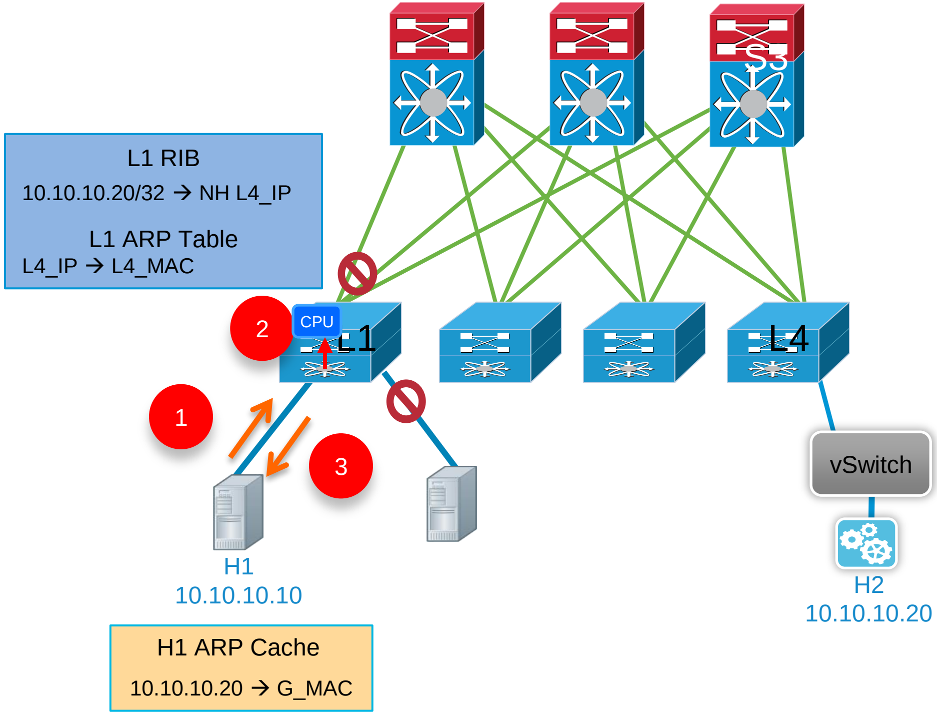


- Any subnet anywhere => Any leaf can instantiate any subnet
  - All leafs share gateway IP and MAC for a subnet (No HSRP)
  - ARPs are terminated on leafs, No Flooding beyond leaf
- Facilitates VM Mobility, workload distribution, arbitrary clustering
- Seamless L2 or L3 communication between physical hosts and virtual machines

Optimised Network

IP Forwarding within the Same Subnet

1. H1 sends an ARP request for H2 – 10.10.10.20
2. The ARP request is intercepted at the leaf L1 and punted to the Sup
3. Assuming a valid route to H2 does exist in the Unicast RIB, L1 sends the ARP reply with the G_MAC so that H1 can build its ARP cache

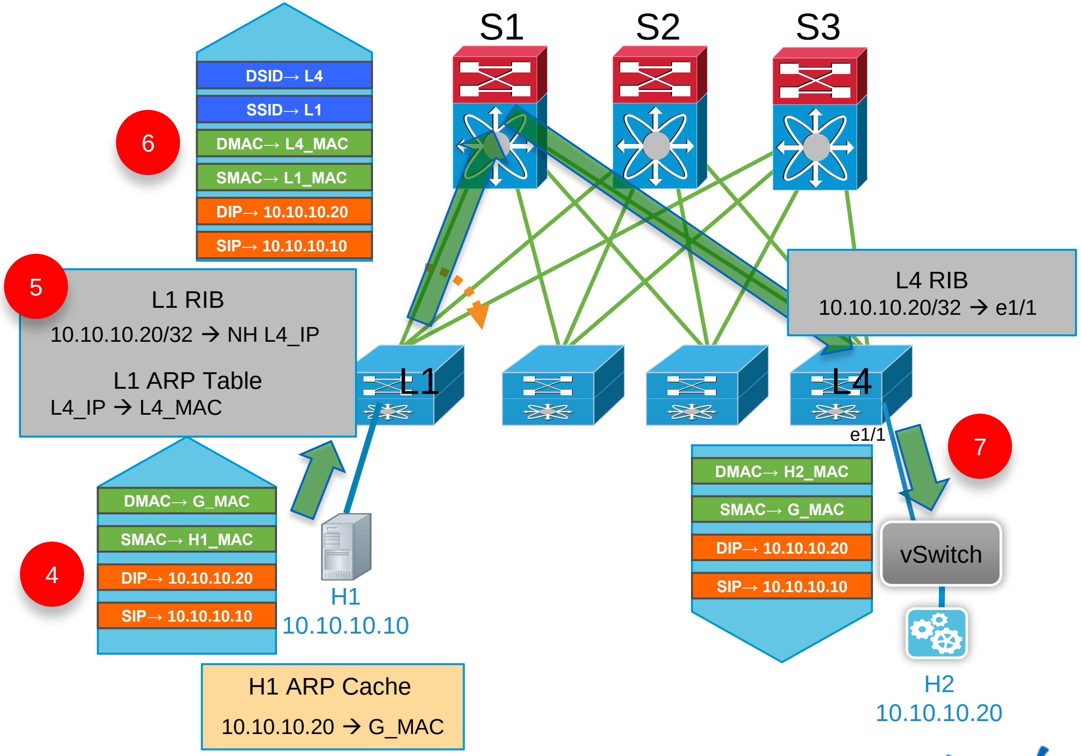


Note: the ARP request is NOT flooded across the Fabric, nor out of other local interfaces belonging to the same L2 domain

Optimised Network

IP Forwarding within the Same Subnet (2)

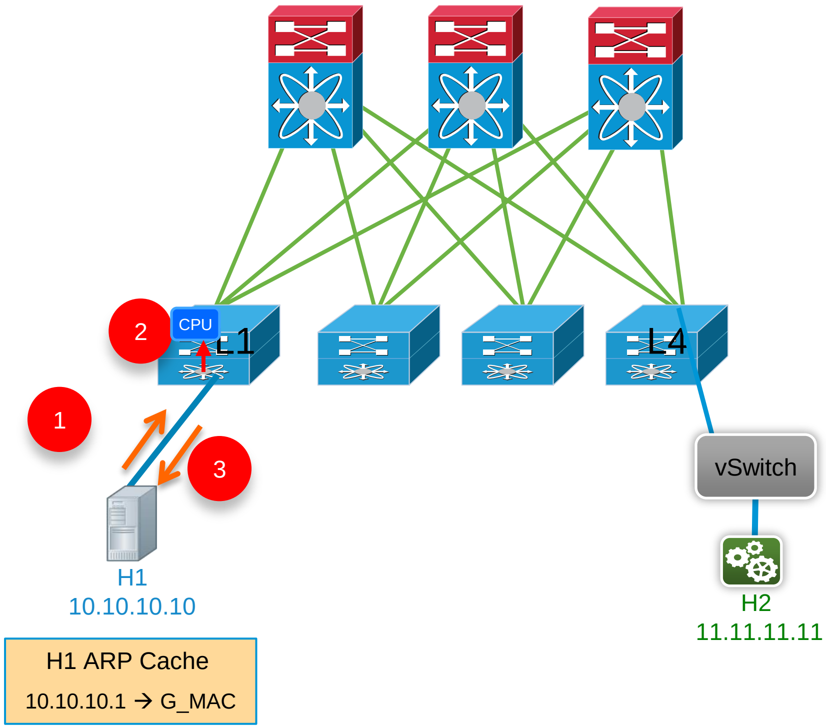
4. H1 generates a data packet with G_MAC as destination MAC
5. L1 receives the packet, remove the L2 header and performs Layer 3 lookup for the destination
6. L1 adds the Layer 2 and the FP headers and forwards the FP frame across the Fabric, picking one of the 3 equal cost paths available via S1, S2 and S3
7. L4 receives the packet, strips off the FP and L2 headers and performs L3 lookup and forwarding toward H2



Optimised Network

IP Forwarding Across Different Subnets

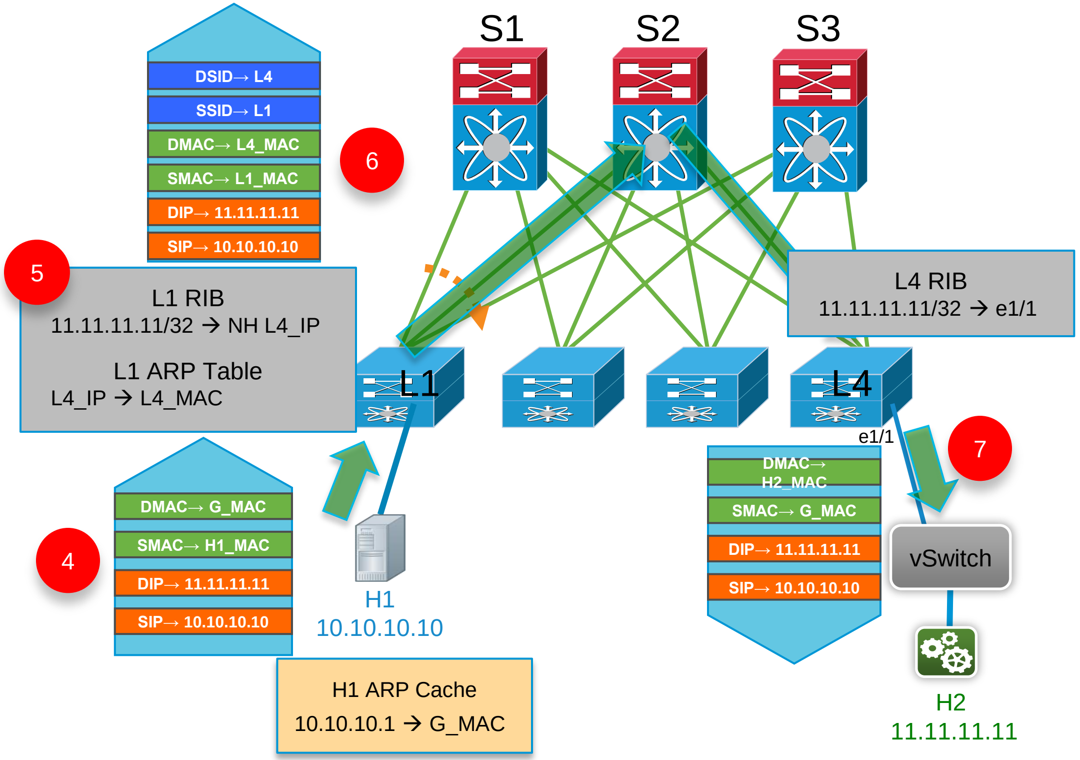
1. H1 sends ARP request for default gateway – 10.10.10.1
2. The ARP request is intercepted at the leaf and punted to the Sup
3. L1 acts as regular default gateway and sends ARP reply with G_MAC



Optimised Network

IP Forwarding Across Different Subnets (2)

4. H1 generates a data packet destined to H2 IP with G_MAC as destination MAC
5. L1 receives the packet, remove the L2 header and performs Layer 3 lookup for the destination
6. If valid routing information for H2 are available in the unicast routing table, L1 adds the Layer 2 and the FP headers and forwards the FP frame across the Fabric, picking one of the 3 equal cost paths available via S1, S2 and S3
7. L4 receives the packet, strips off the FP and L2 headers and performs L3 lookup and forwarding toward H2



Optimised Network

Introducing L3 Conversational Learning

Optimized Network

- Use of /32 host routes may lead to scaling issues if all the routes are installed in the HW tables of all leaf nodes

  L3 conversational learning is introduced to alleviate this concern

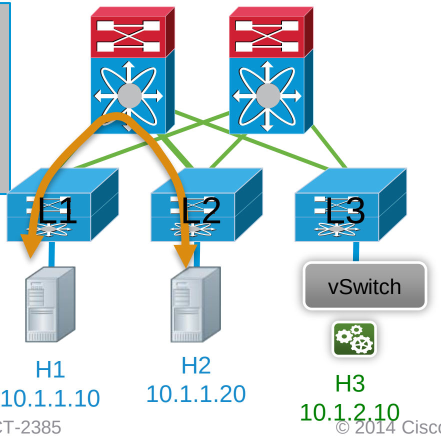
  Disabled by default → all host routes are programmed in the HW

- With L3 conversational learning, host routes for remote endpoints will be programmed into the HW FIB (from the SW RIB) upon detection of an active conversation with a local endpoint

Default Behaviour (No L3 Conversational Learning)

L1 RIB
10.1.1.20/32 → NH L2_IP
10.1.2.10/32 → NH L3_IP

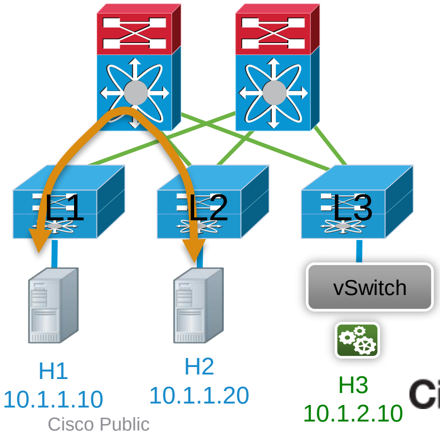
L1 FIB
10.1.1.20/32 → NH L2_IP
10.1.2.10/32 → NH L3_IP



After Enabling L3 Conversational Learning

L1 RIB
10.1.1.20/32 → NH L2_IP
10.1.2.10/32 → NH L3_IP

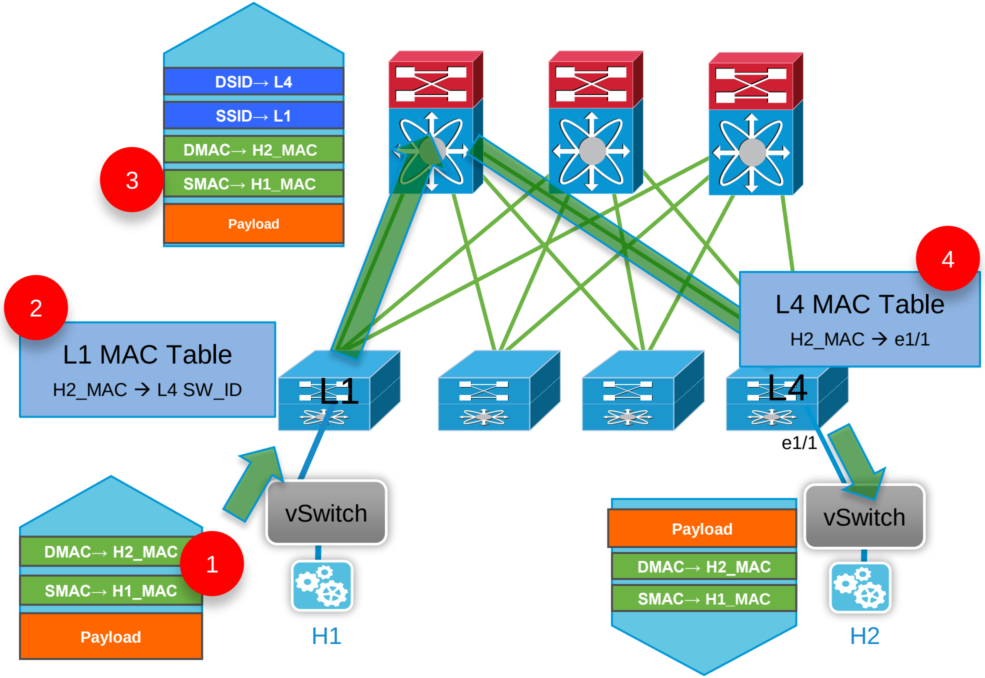
L1 FIB
10.1.1.20/32 → NH L2_IP



Forwarding

L2 non IP Flows

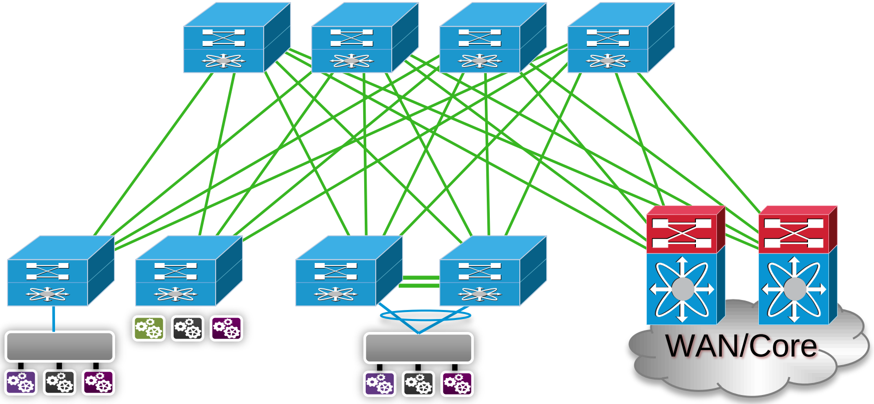
1. H1 originates a packet destined to H2 MAC address
2. L2 lookup is performed by L1 in the MAC Table for the VLAN the frame belongs to
3. L1 adds the FP header before sending the packet into the fabric
4. L4 receives the frame, decapsulates the FP header, performs the L2 lookup and then sends it to H2



Optimised Network

Multicast Forwarding

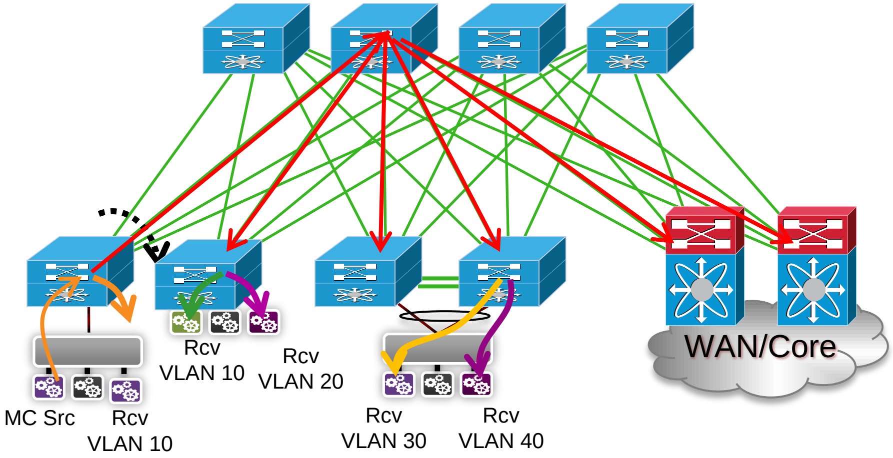
- Fabric supports computation of multiple distribution trees leveraging IS-IS
  - Used for multicast and broadcast traffic
  - No need for other multicast protocols (PIM, etc.) inside the fabric
- Multi Destination Trees (MDTs) Rooted on Spines
- Ingress Leaf load balances traffic across multiple paths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



Optimised Network

Multicast Forwarding

- Two tiers multicast replication across the fabric Ingress
  - Ingress Leaf always performs multicast routing functions and sends a single copy onto the fabric
  - Spine node replicates to the leaf nodes
  - Destination Leaf nodes locally replicate to server ports across subnets
- Optimisation possible to allow pruning on the spine (per tenant/VRF or per group)



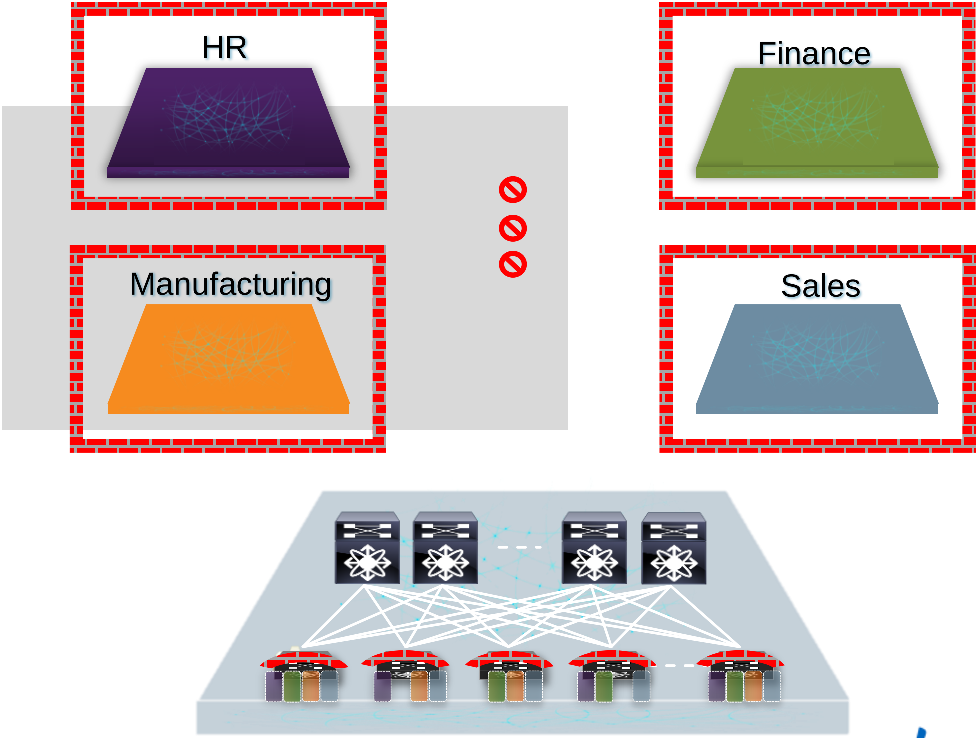
Agenda

- DFA Requirements and Functions
- Optimised Network
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support

Virtual Fabrics for Public or Private Cloud Environments

Advantages

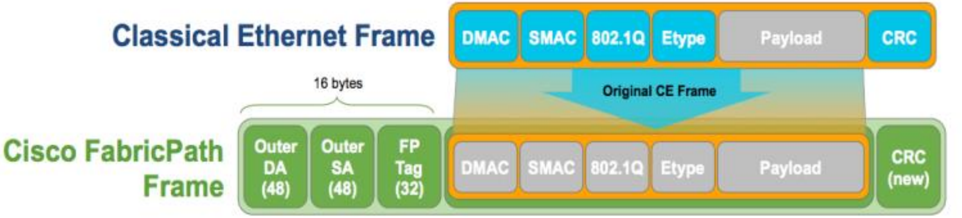
- Any workload, any vFabric, rapidly
- Scalable Secure vFabrics
- vFabric Tenant Visibility
- Routing/Switching Segmentation



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 centre to 4096 VLANs

FabricPath Frame Format

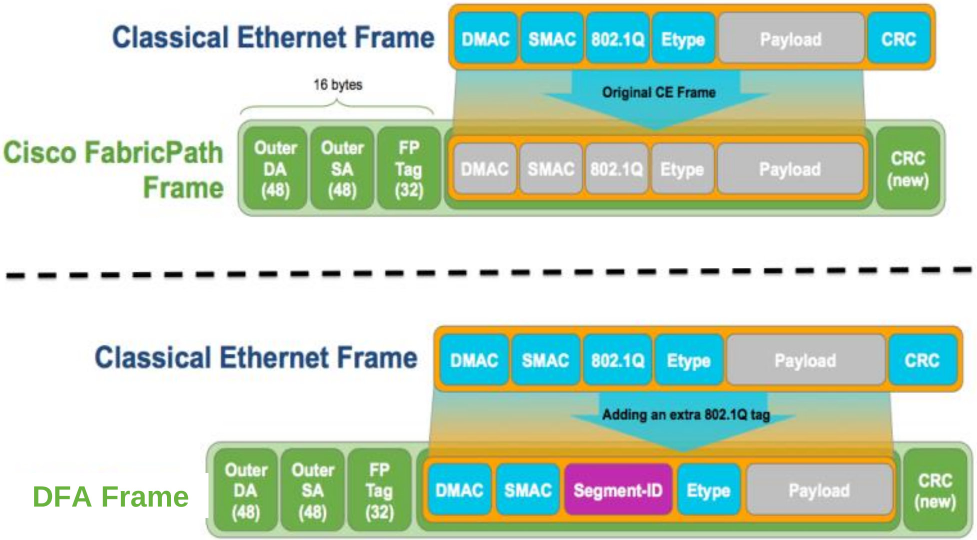


Virtual Fabrics

Introducing Segment-ID Support

- Traditionally VLAN space is expressed over 12 bits (802.1Q tag)
  - Limits the maximum number of segments in a data centre to 4096 VLANs
- DFA leverages a double 802.1Q tag for a total address space of 24 bits
  - Support of ~16M L2 segment (10K targeted at FCS)
- Segment-ID is hardware-based innovation offered by leaf and spine nodes part of the Integrated Fabric

FabricPath Frame Format



Integrated Fabric Frame Format

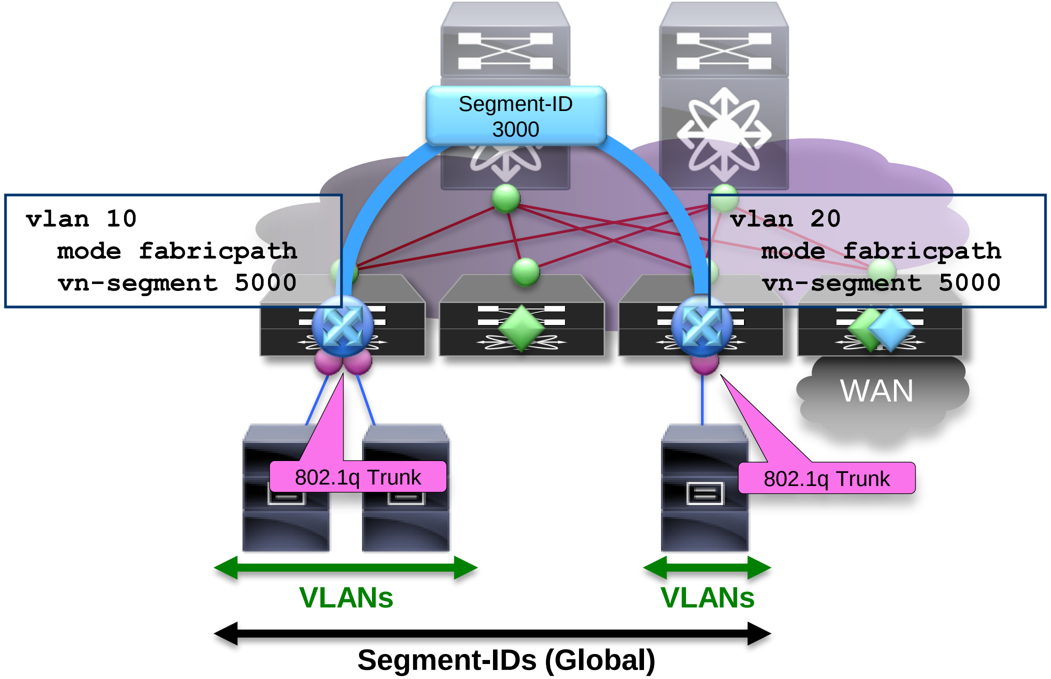
Virtual Fabrics

802.1Q Tagged Traffic to Segment-ID Mapping

Virtual Fabrics

- Segment-IDs are utilised for providing isolation at L2 and L3 across the Integrated 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

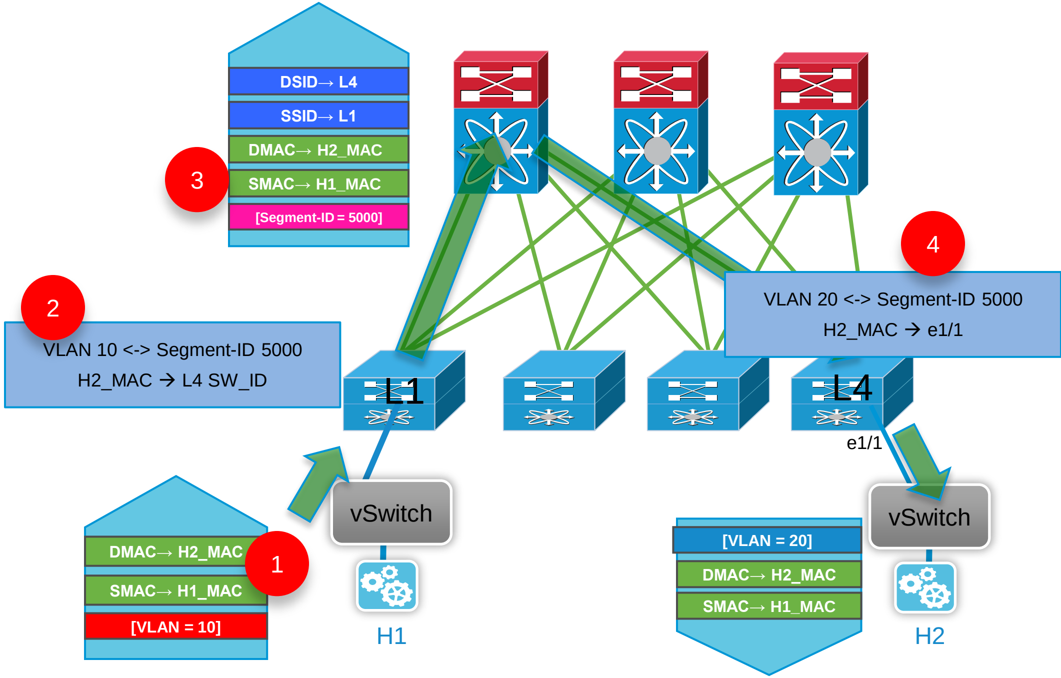
  VLANs become locally significant on the leaf node and 1:1 mapped to a Segment-ID
- Segment-IDs are globally significant, VLAN IDs are locally significant



Virtual Fabrics

L2 non IP Flows

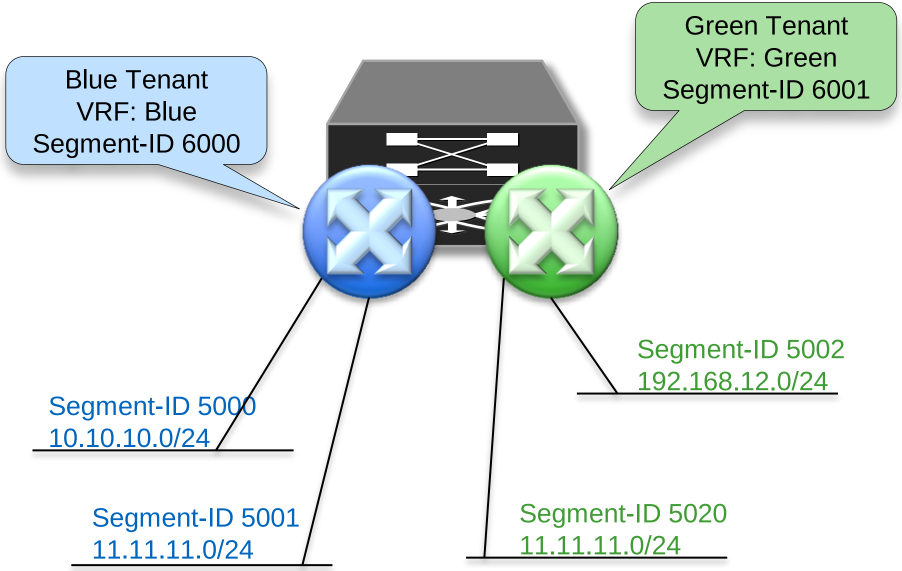
1. H1 sends a packet to H2 → traffic between the vSwitch and the Leaf is tagged with a local VLAN-ID 10
2. L2 lookup is performed by L1 in the MAC Table for the Segment-ID associated to VLAN 10 (5000)
3. L1 adds the L2 and FP headers before sending the packet into the fabric. The Segment-ID associated to VLAN 10 is added inside the L2 header
4. L4 receives the frame and performs the L2 lookup by looking at the Segment-ID value. It then sends it to H2 using a local VLAN-ID 20



Virtual Fabrics

How are Segment-IDs Utilised?

- Each IP subnets defined at the edge of the DFA Fabric is associated to a Layer 2 domain, which is represented by a Segment-ID
- Multiple Segments can be defined for a given tenant and are usually mapped to a L3 VRF uniquely identifying that tenant
- A dedicated Segment-ID value uniquely identifies each VRF defined in the DFA Fabric

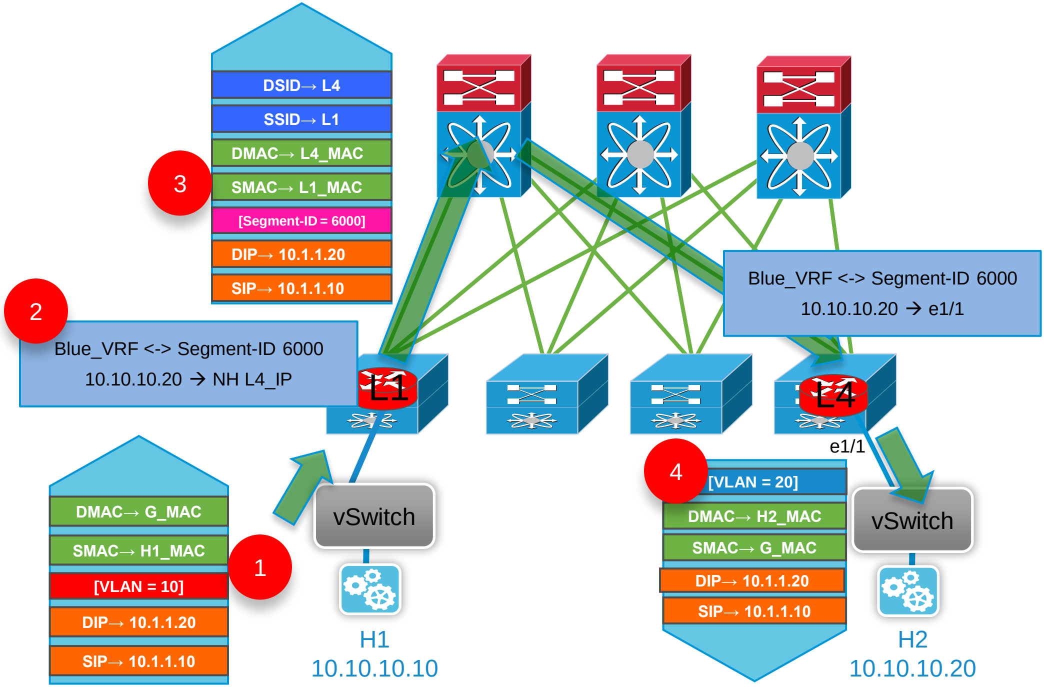


Virtual Fabrics

Fabric Routed Flows

1. H1 sends a packet to H2 → traffic between the vSwitch and the Leaf is tagged with a local VLAN-ID 10
2. L3 lookup is performed by L1 in the context of the Red VRF
3. L1 adds the L2 and FP headers before sending the packet into the fabric. The Segment-ID identifying the Red VRF is added inside the L2 header
4. L4 receives the frame and associates it to the Red VRF by looking at the Segment-ID value. It then sends it to H2 using a local VLAN-ID 20

Note: this behaviour applies to all fabric routed flows (intra-subnet or inter-subnet)



Agenda

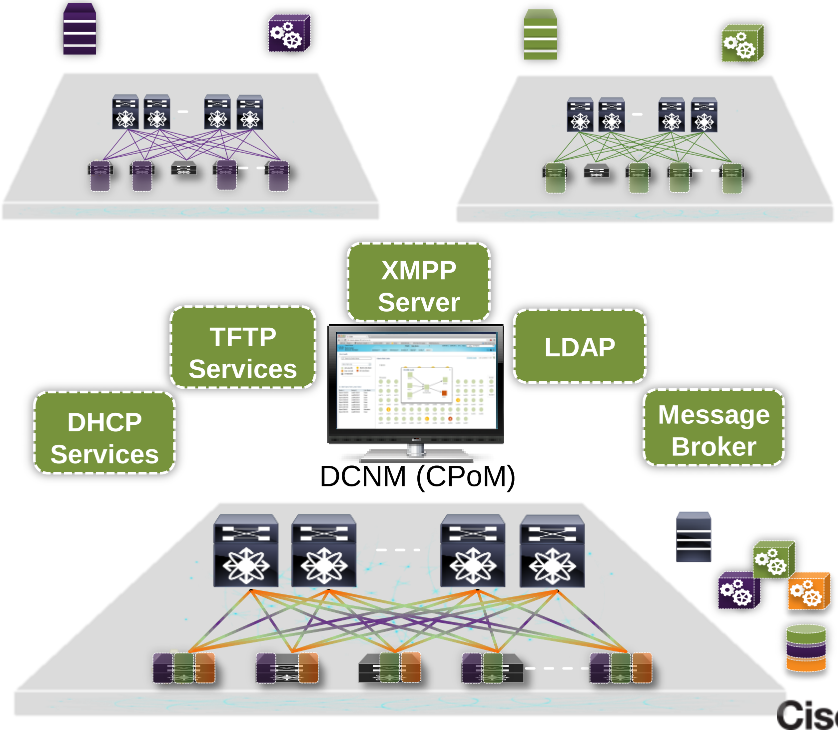
- DFA Requirements and Functions
- Optimised Network
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support

Simplifying Fabric Management & Optimising Fabric Visibility

Fabric Management

Advantages

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



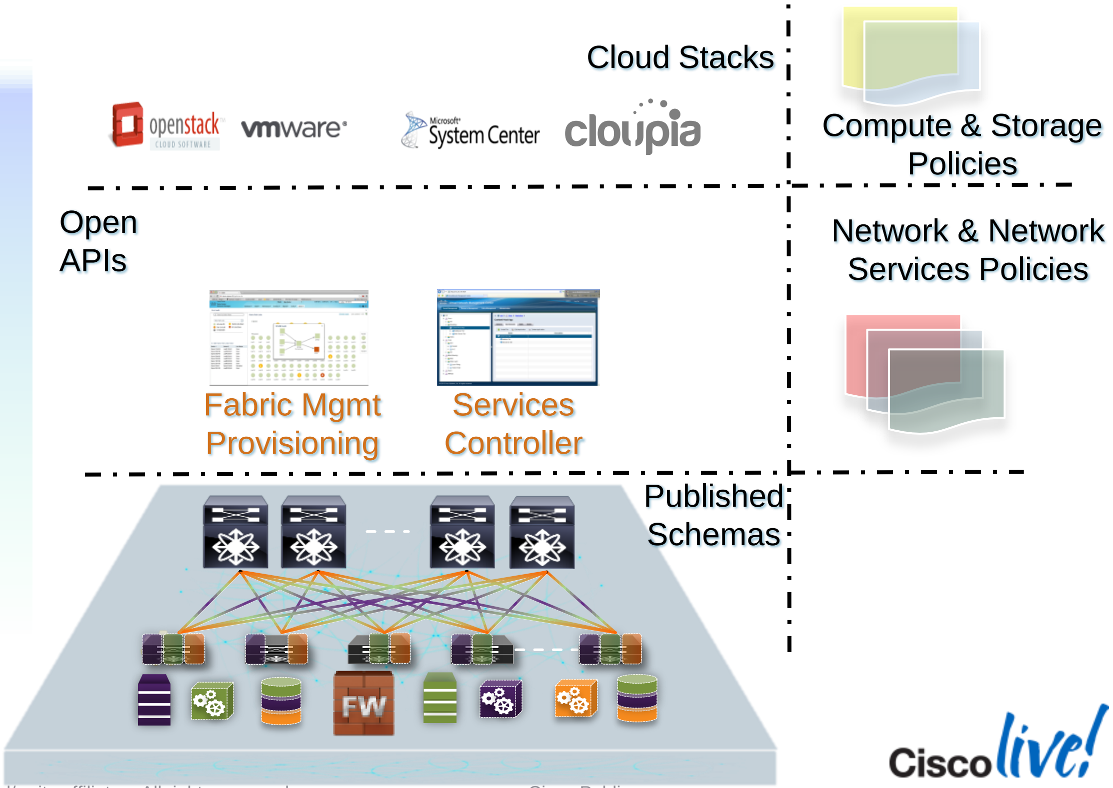
Agenda

- DFA Requirements and Functions
- Optimised Network
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support

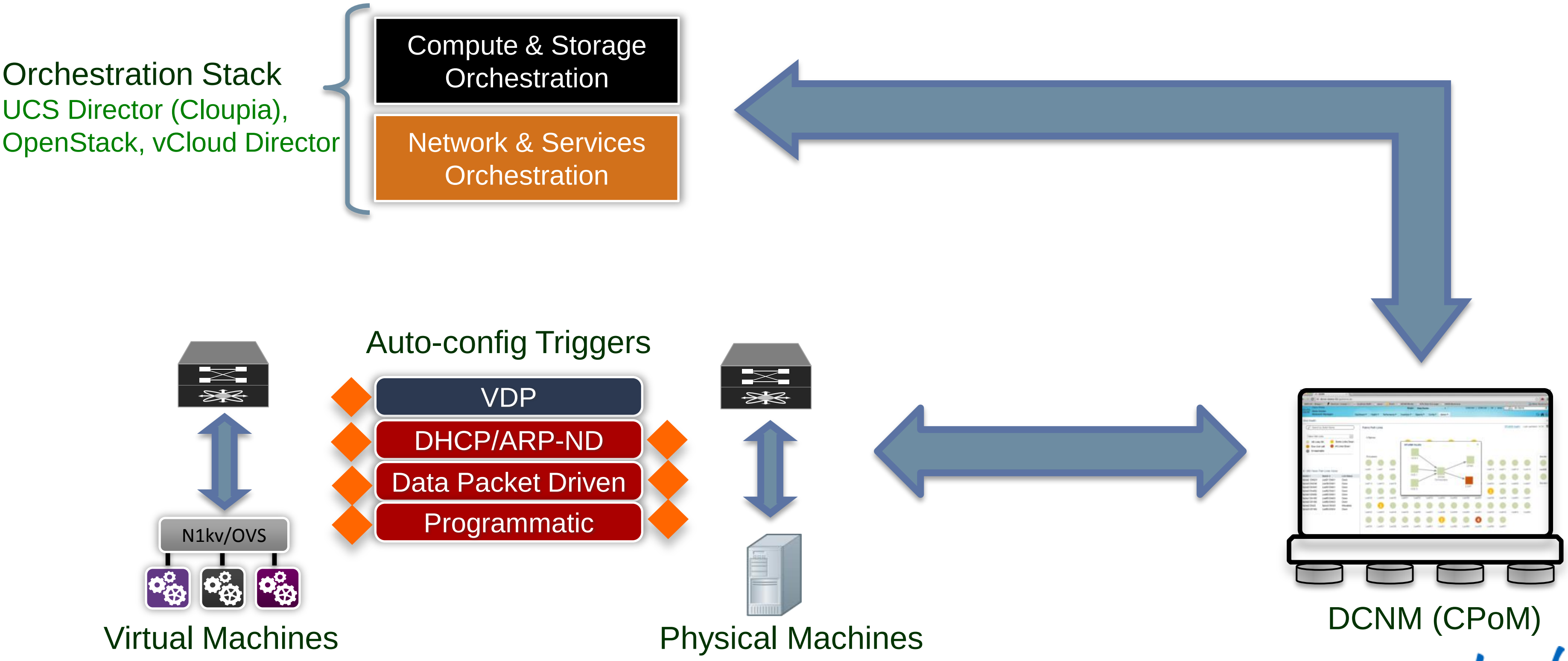
Workload Automation & Open Environment

Advantages

- Any workload, anywhere, anytime
- Open Integration: orchestration
- Automated scalable provisioning
- Workload aware fabric



Workload Automation & Open Environment

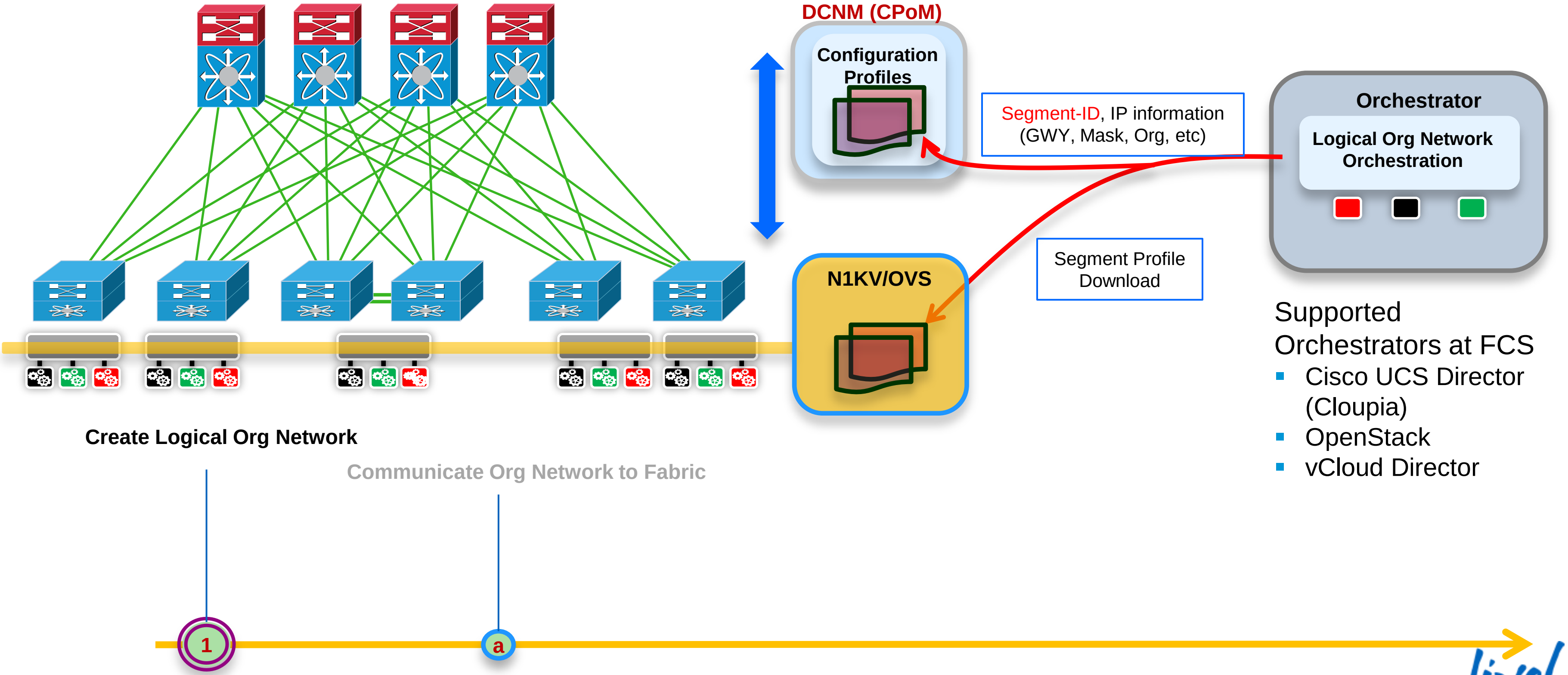


*VDP (VSI Discovery and Configuration Protocol) is IEEE 802.1Qbg Clause 41

DFA Demo

Workload Automation

Leveraging VDP for Leaf Auto-Configuration

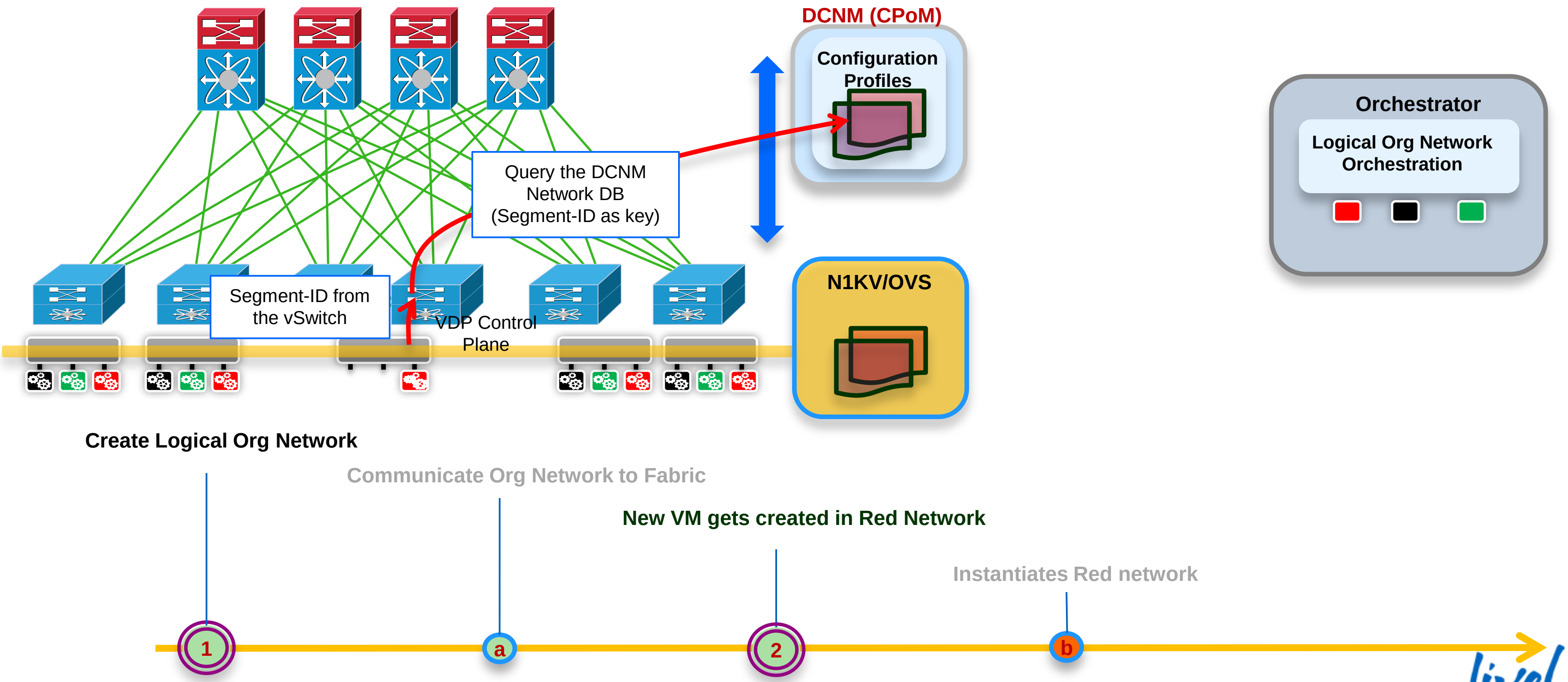


Supported Orchestrators at FCS

- Cisco UCS Director (Cloupia)
- OpenStack
- vCloud Director

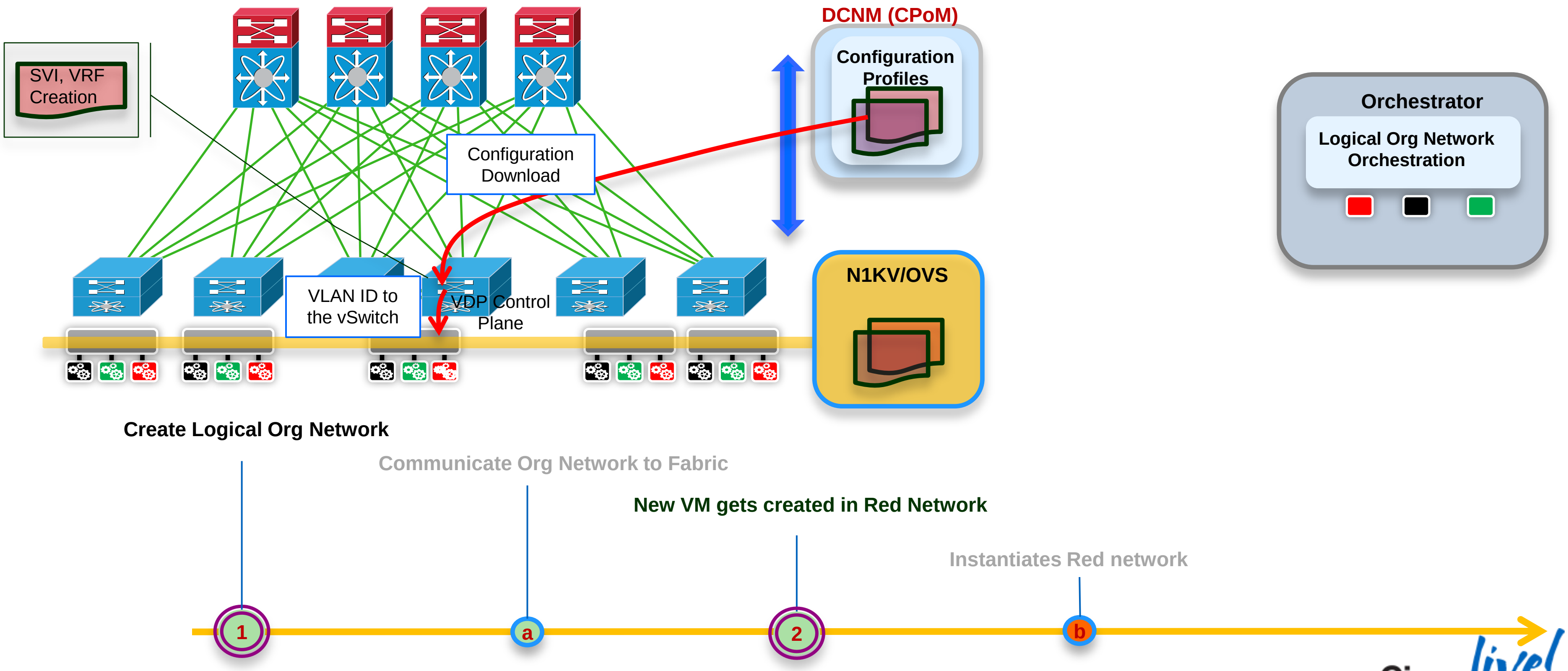
Workload Automation

Leveraging VDP for Leaf Auto-Configuration (2)



Workload Automation

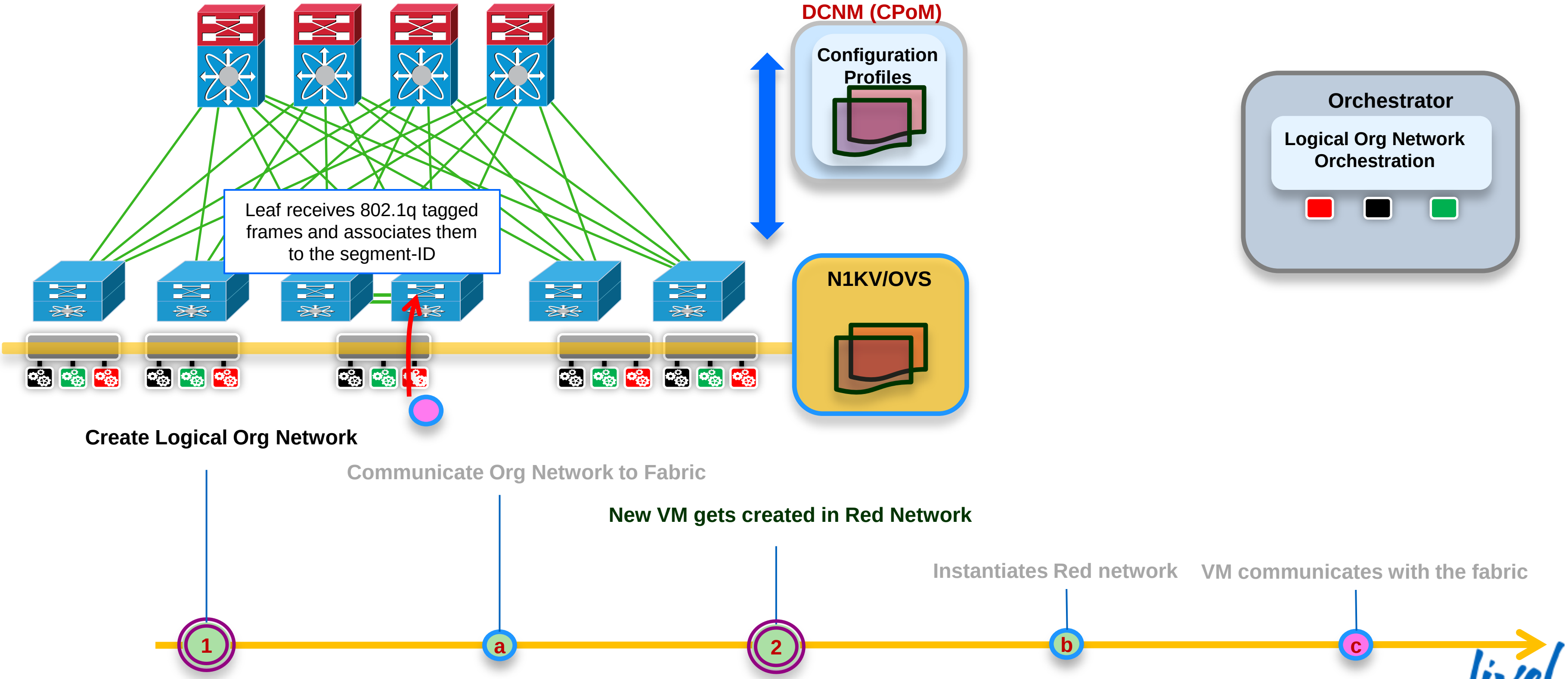
Leveraging VDP for Leaf Auto-Configuration (3)



*VDP (VSI Discovery and Configuration Protocol is part of 802.1Qbg Draft

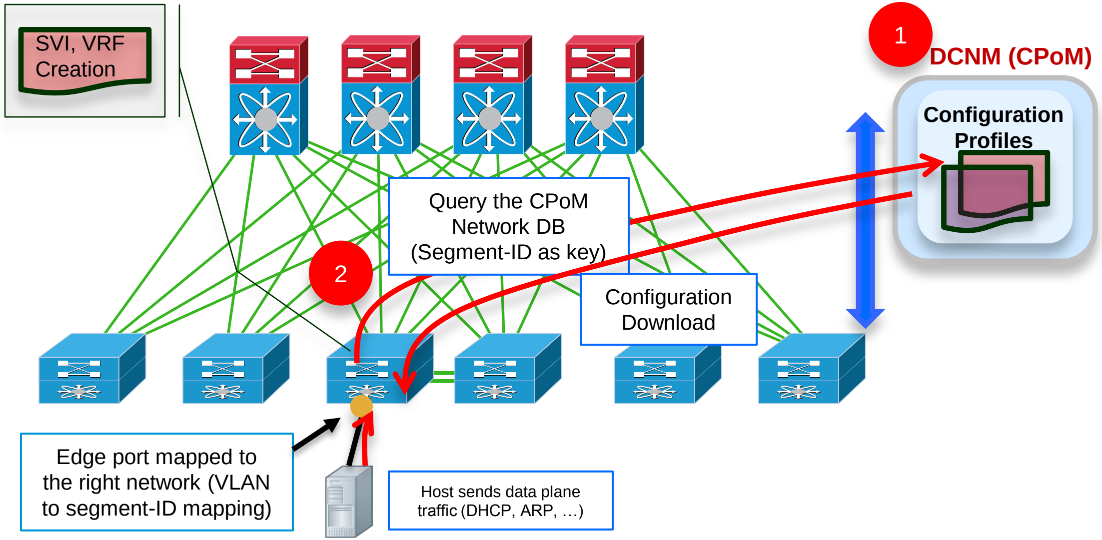
Workload Automation

Leveraging VDP for Leaf Auto-Configuration (4)



Workload Automation

What about Auto-Configuration for Physical Hosts?



Two steps required to provide connectivity into the fabric to a physical host

1. Adding configuration profile to the CPOM network database
2. Detecting when the host connects to query the database and instantiate the configuration on the leaf

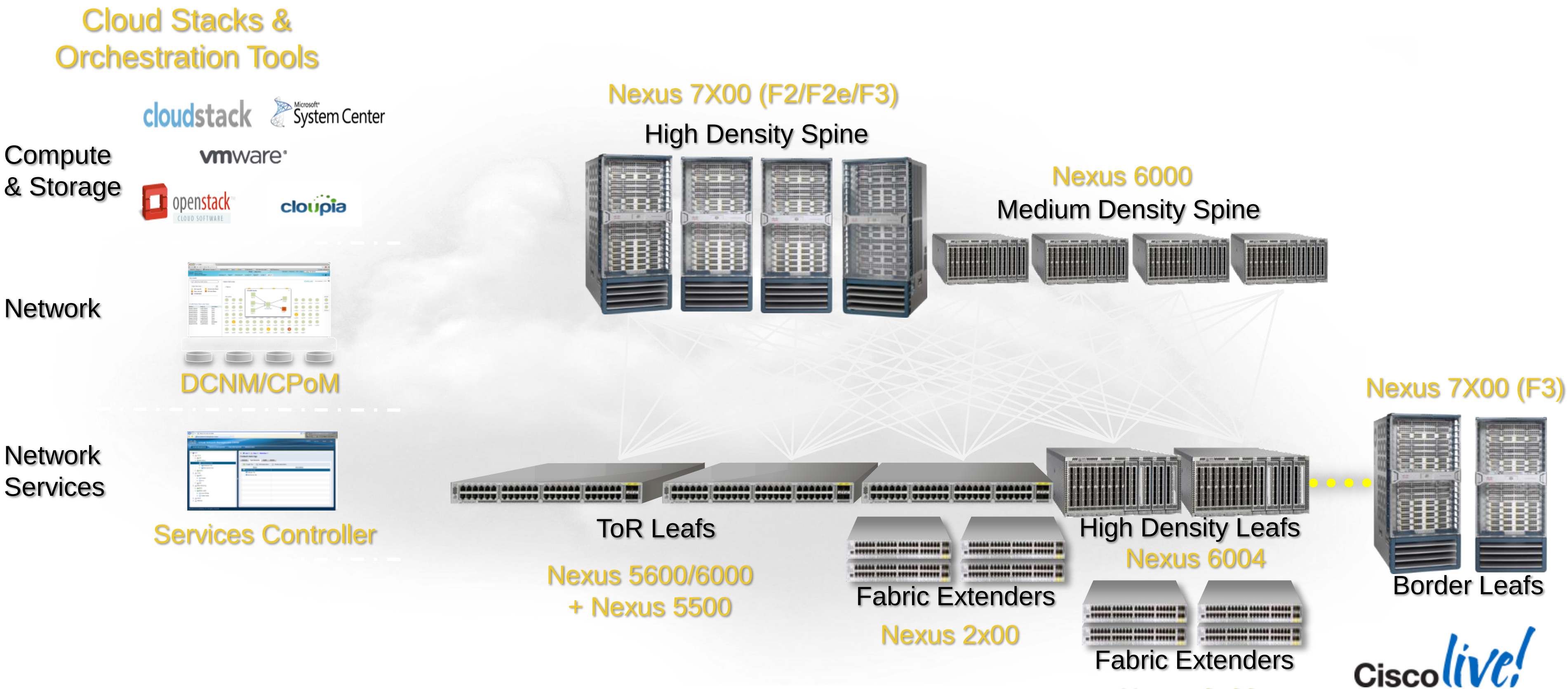
Same model could apply to VMs deployed on vSwitches not supporting VDP

Agenda

- DFA Requirements and Functions
- Optimised Network
  - Fabric Properties
  - Control Plane
  - Forwarding Plane
- Virtual Fabrics
- Fabric Management
- Workload Automation
- Hardware Support

Cisco Dynamic Fabric Automation

Platform Support at FCS



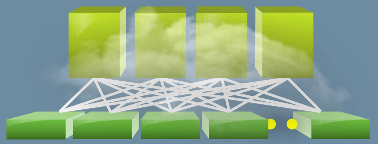
Cisco Dynamic Fabric Automation Architecture

Where to Get More Information

Fabric Management

Workload Automation

Optimised Network



Virtual Fabrics

Check out the DFA Booth at the World of Solutions (live demo)

www.cisco.com/go/dfa




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.

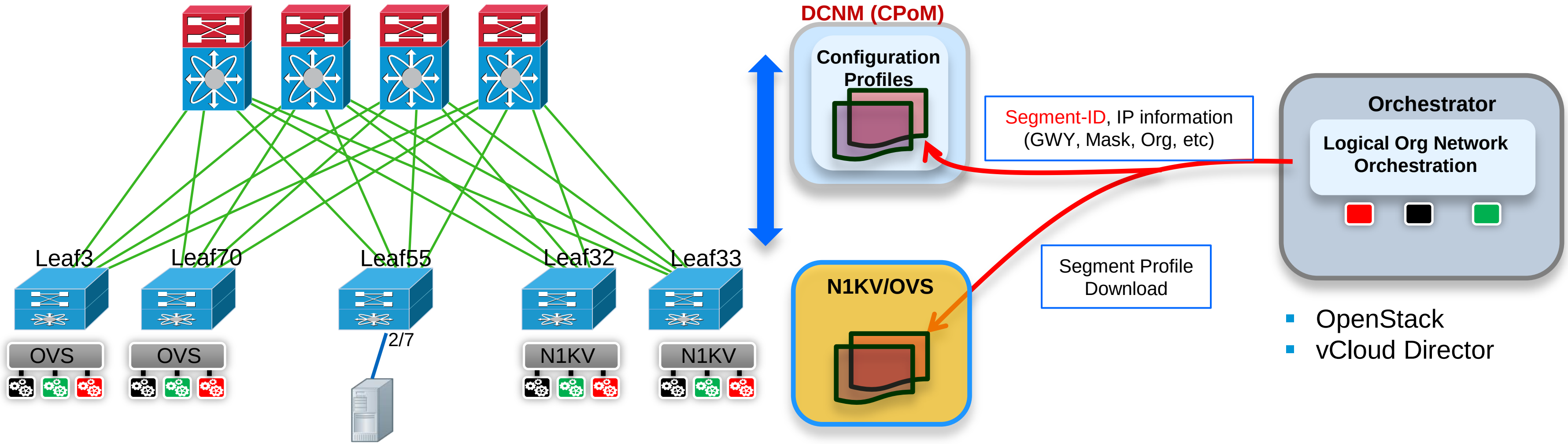
- Directly from your mobile device on the Cisco Live Mobile App
- By visiting the Cisco Live Mobile Site www.ciscoliveaustralia.com/mobile
- Visit any Cisco Live Internet Station located throughout the venue

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™

Demo Setup



- OpenStack
- vCloud Director

CISCO™