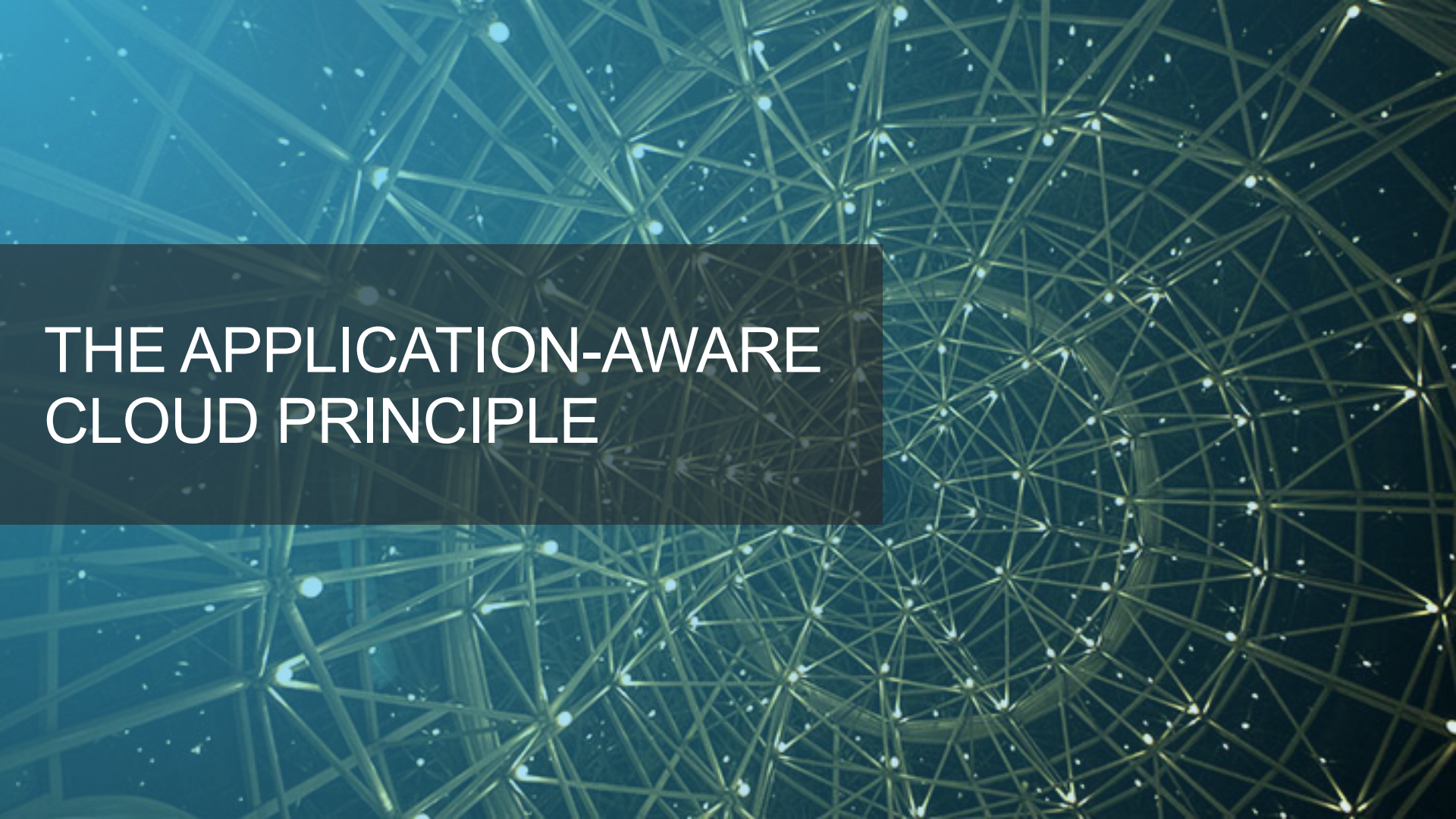
MULTICLOUD AS THE NEXT GENERATION CLOUD INFRASTRUCTURE

Deepti Chandra

Agenda

- The “Application-aware” Cloud Principle
- Problem Statement in Multicloud Deployment
- SDN in the Multicloud
- Building Blocks
- Building the Private Cloud – DC Fabric
- Building the Private Cloud – DC Interconnect (DCI)
- Building the Private Cloud – WAN Integration
- Building the Private Cloud – Traffic Optimization


THE APPLICATION-AWARE
CLOUD PRINCIPLE

The Big Picture: Cooperative Clouds

USE

REQUIRING

RUNNING

End Users
People, vehicles, appliances, devices

Applications
Made of software components

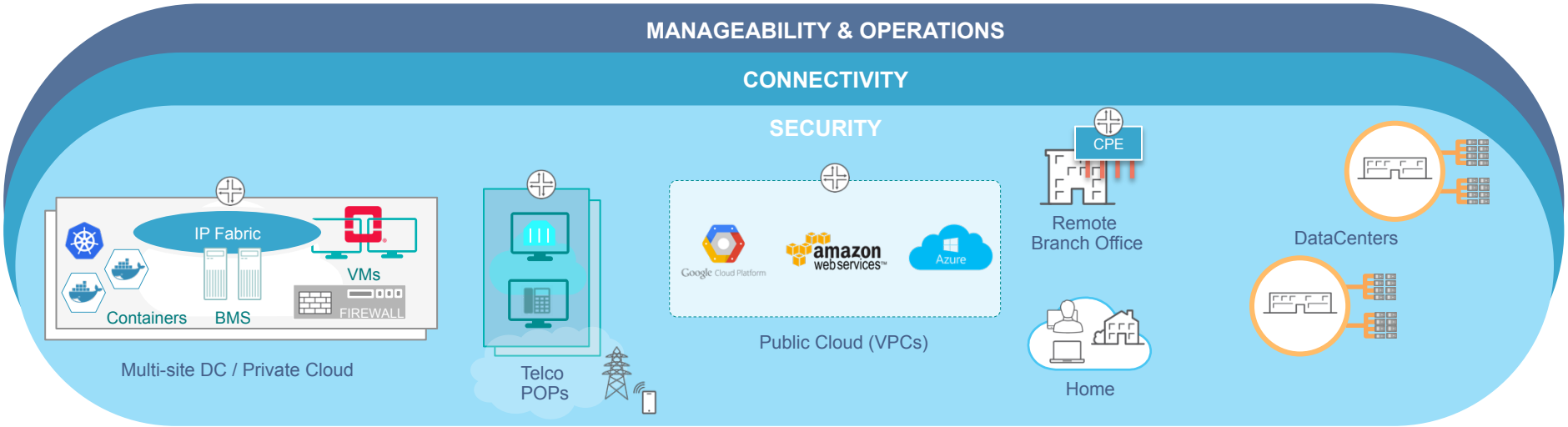
MANAGEABILITY & OPERATIONS

SECURITY

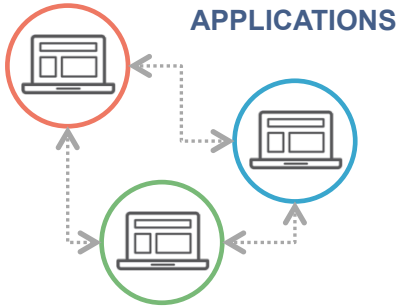
CONNECTIVITY

MULTIPLE LOCATIONS
Embedded (e.g. device or vehicle), in a Data Center

MULTIPLE ENVIRONMENTS
Containers, VMs, BMS



The Traditional Way



PRIVATE CLOUD

PUBLIC CLOUD

HOSTING

DECISIONS DRIVEN BY:

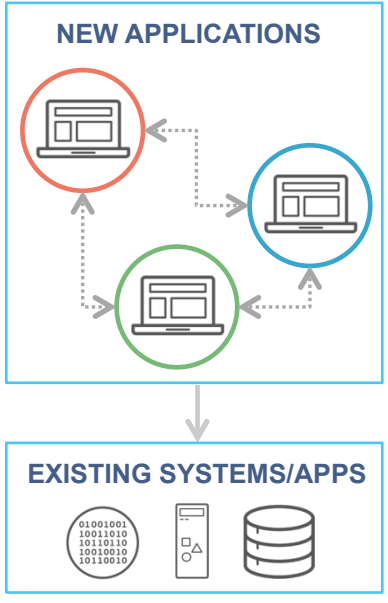
1. Existing assets
2. Skills and know-how
3. Security & Confidentiality
4. Costs and TCO control
5. Application specific requirements (scale, latency, performance, hypervisors,…)

EXISTING SYSTEMS/APPS

I need to deliver a service to my users

Build or buy or lease an Execution environment

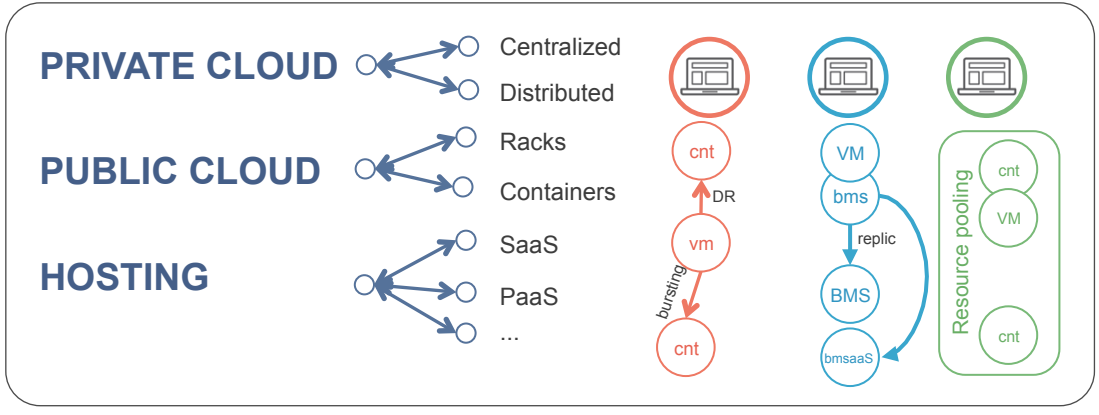
What Has Changed



I need to deliver a service to my users

Why Data Centers Need Multicloud?

Today most applications leverage cooperation between components deployed across multiple cloud infrastructure



Today – The New Cloud

DECISIONS DRIVEN BY:

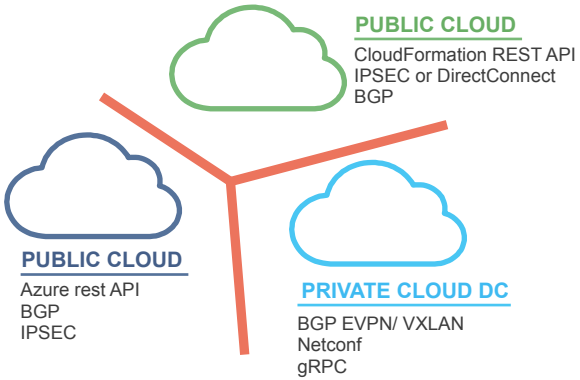
1. User experience
2. Costs and TCO control
3. Agility (time to change)
4. Security and confidentiality
5. Skills and know-how
6. Application specific requirements (scale, latency, performance, hypervisors, …)


PROBLEM STATEMENT IN THE
MULTICLOUD DEPLOYMENT

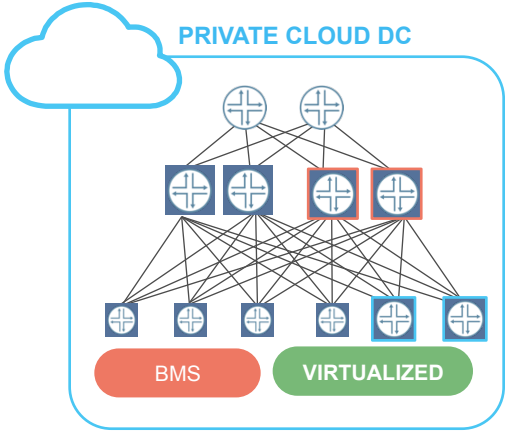
Challenges of the Multicloud

A set of independent ‘fabrics’

Poor User Experience

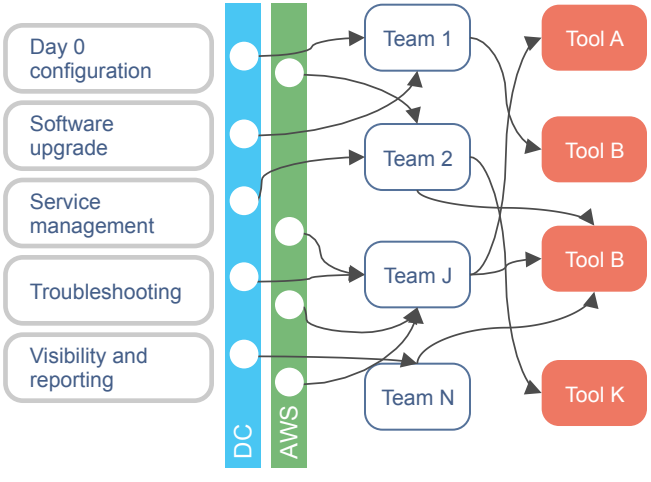


Operational Complexity



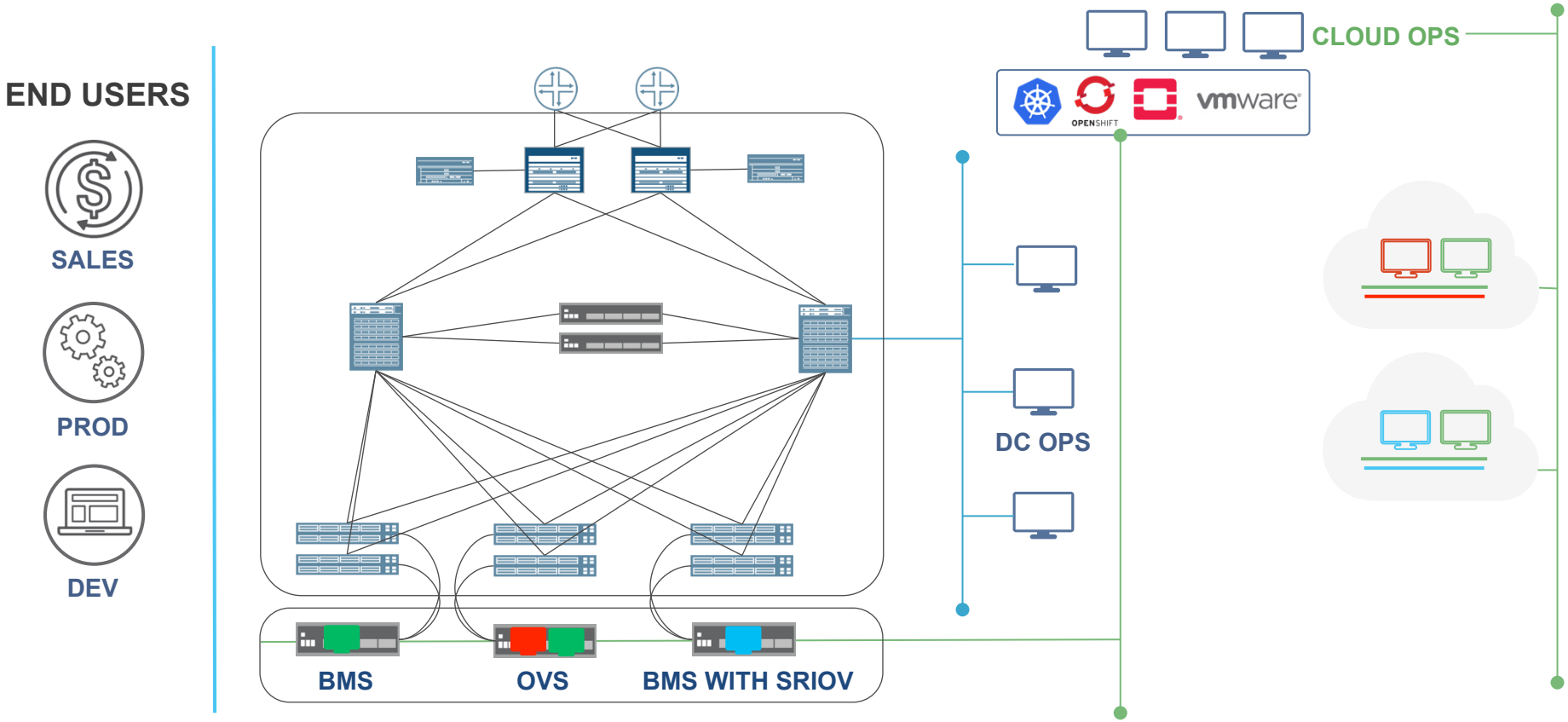
- Different skillsets for different clouds
- Manual operations for daily tasks

Lack of Automation



- Long lead-times for change management
- Inconsistent visibility for distinct environments

Clouds are Disconnected Environments



A Day in the Life of DC/Cloud Operations

THE END USER

"I need a two-tier application execution environment with these characteristics"

"Can I have my DB cluster up and running by next week and connected to the Web front end ?"

"DB Cluster can't talk to Web server"

THE OPERATOR

1. Provision tenant
2. Image servers
3. Create containers
4. Create/select policies
5. Request networking service to public cloud
6. Create EC2 instance
7. ...

Need to correlate & contextualize: "which IP1/MAC1 on VNI X on Switch A can't talk to IP2/MAC2 on VNI Y on Switch B?"

COMPLEXITY

INCONSISTENCY

REVENUE-LOSS

LONG LEAD TIME

PROVISIONING | **MANAGEMENT** | **VISIBILITY**


SDN IN THE
MULTICLOUD

WHAT DOES SDN OFFER IN THE MULTICLOUD?

MULTICLOUD NETWORKING AS A SERVICE

Single pane of glass orchestration across clouds

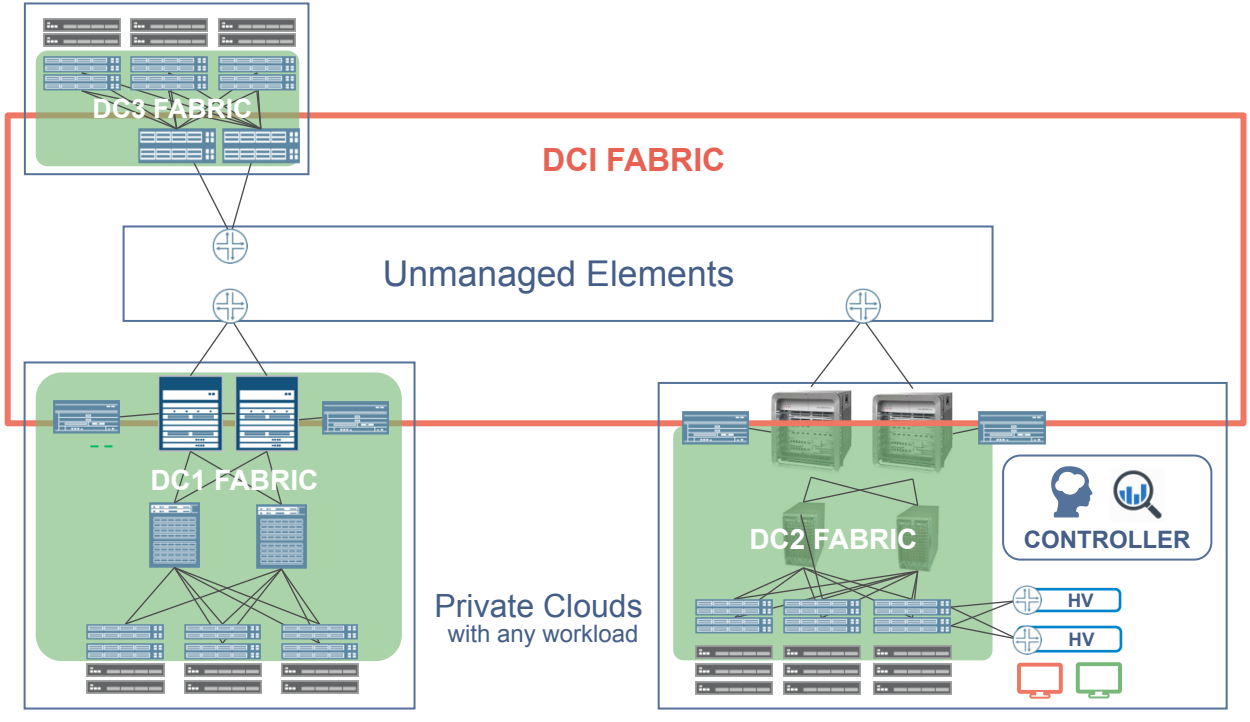
Visibility and unified management across clouds

Secure service delivery across clouds

Building for containerization

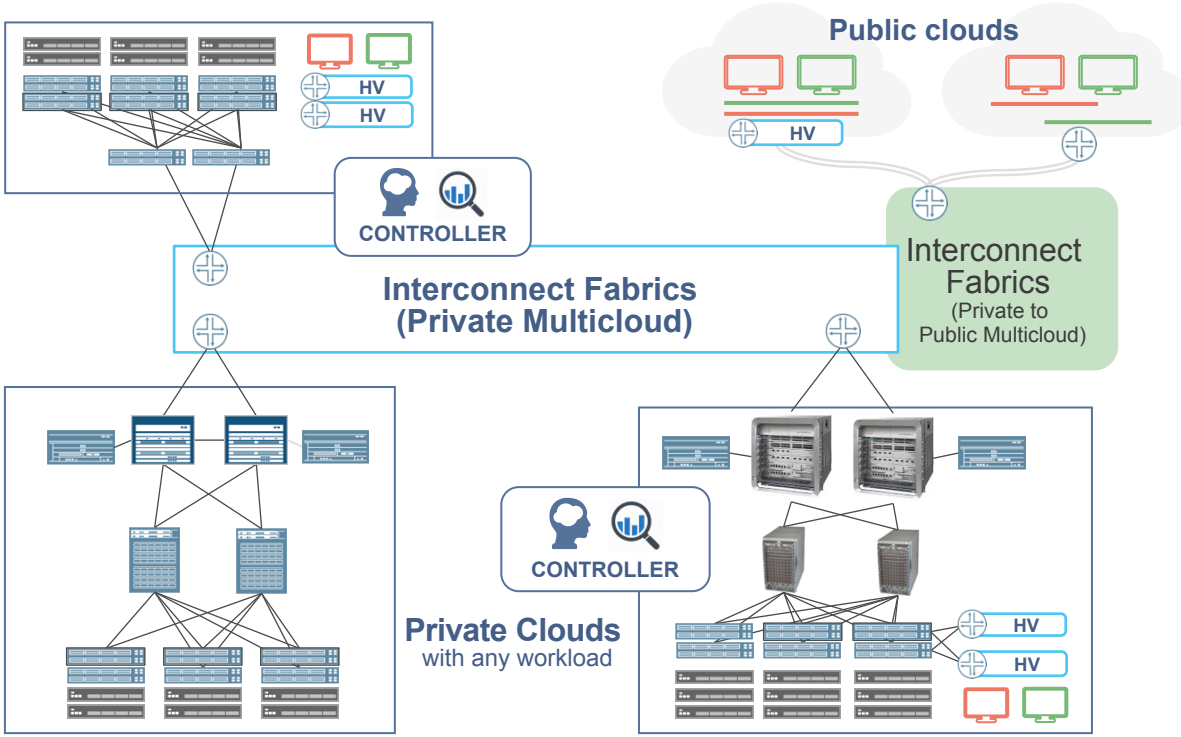
Federation to unify controllers across clouds

Administration of Fabrics

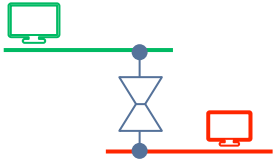


- A fabric is an independently administered IP network managed by the controller for configuration, and eventually routing control plane and analytics
- The controller has IP reachability to ALL endpoints (e.g. a device) which belongs to the fabric managed by the controller
- Devices/endpoints can belong to multiple fabrics

Multi-cloud Networking-as-a-service for Any Workload and Any Cloud

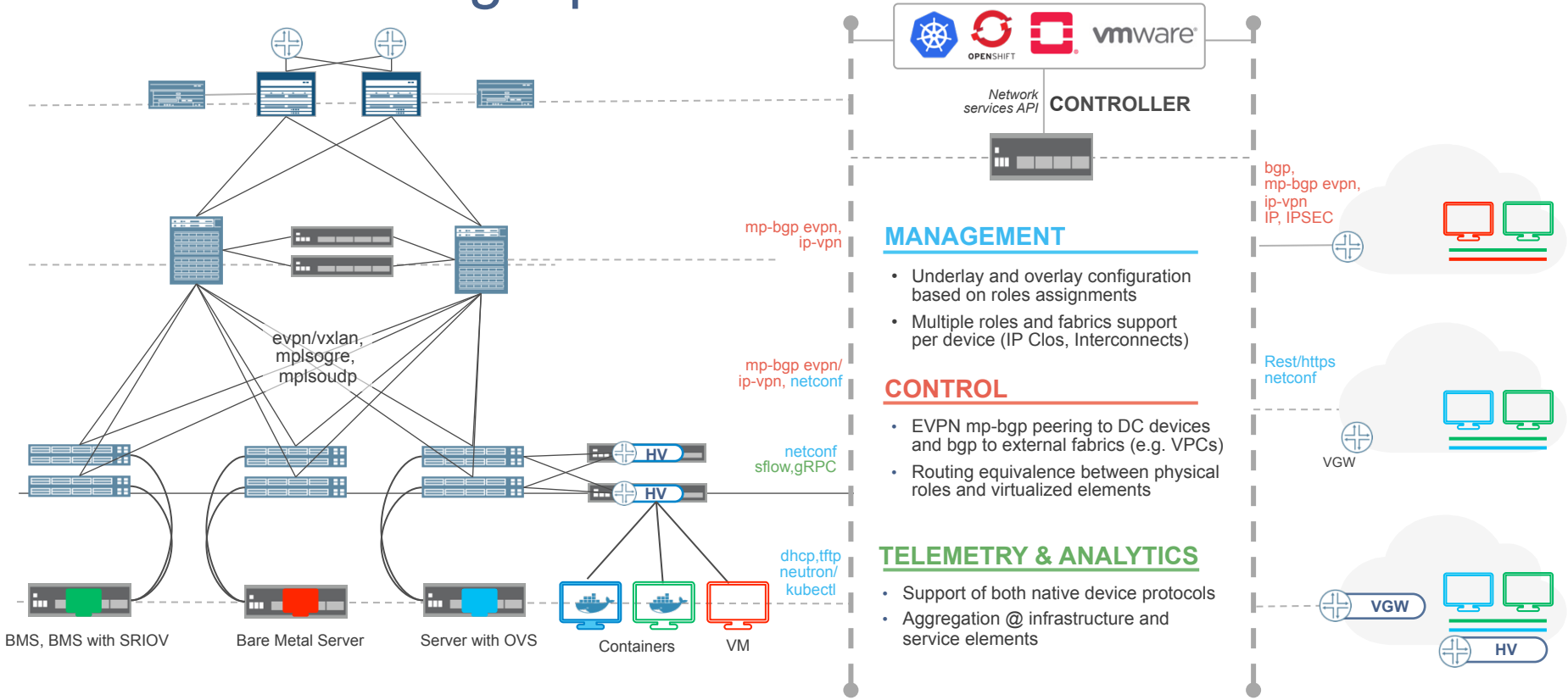


- Automate Private Clouds / Multi-cloud infrastructure
- One-click Application Services



- Predictive Analytics and Visibility

A Unified View Across Cloud and Networking Operations



MANAGEMENT

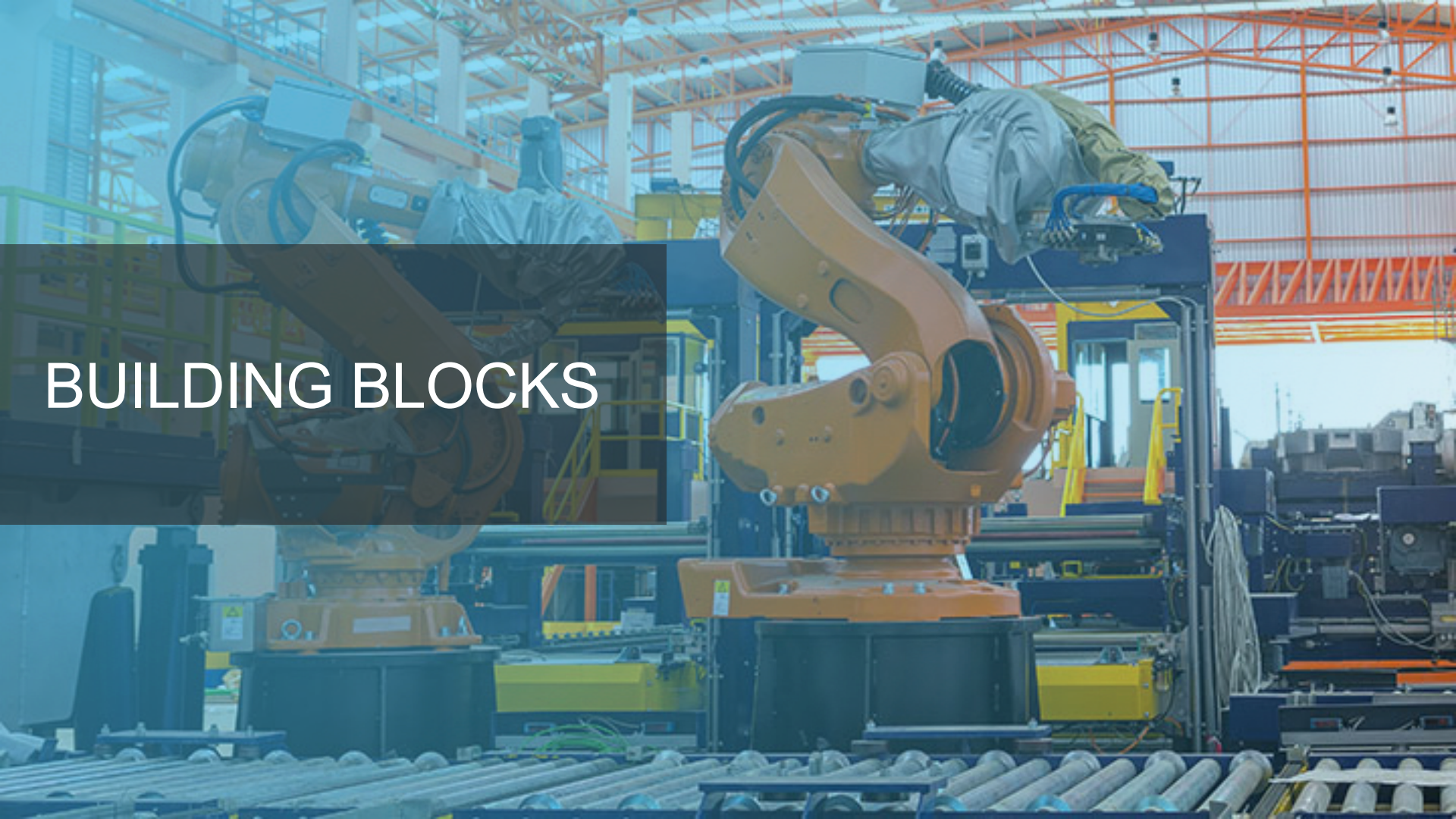
- Underlay and overlay configuration based on roles assignments
- Multiple roles and fabrics support per device (IP Clos, Interconnects)

CONTROL

- EVPN mp-bgp peering to DC devices and bgp to external fabrics (e.g. VPCs)
- Routing equivalence between physical roles and virtualized elements

TELEMETRY & ANALYTICS

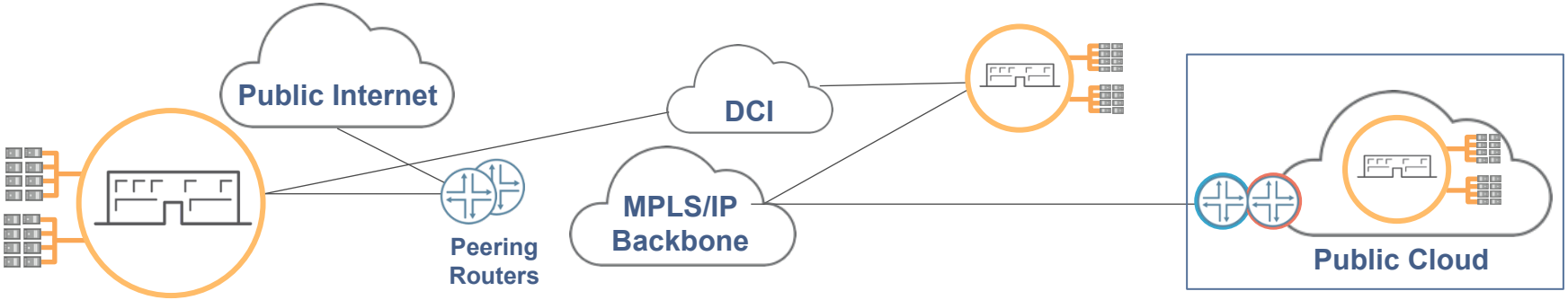
- Support of both native device protocols
- Aggregation @ infrastructure and service elements


BUILDING BLOCKS

Data Center Requirements

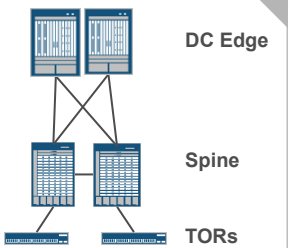
Design Requirement	Technology Attribute
Rising EW traffic growth	Easy scale-out
Resiliency and low latency	Non-blocking, fast fail-over
Agility and speed	Any service anywhere
Open architecture	No vendor lock-in
Design simplicity	No steep learning curve
Architectural flexibility	EW, NS & DCI

Common Building Blocks for Data Centers



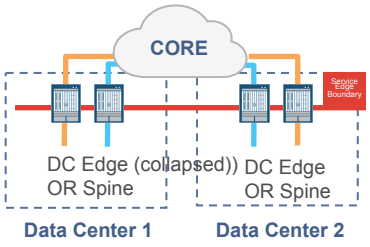
DATA CENTER FABRIC

CLOS Fabric (IP, MPLS)



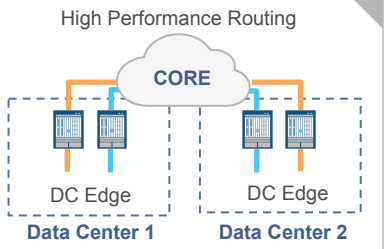
DATA CENTER INTERCONNECT

WAN or Dark fiber



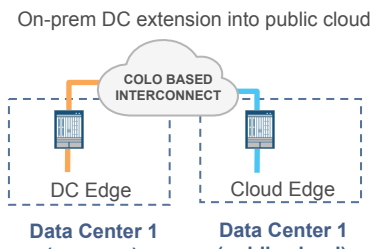
WAN INTEGRATION

Private/Public WAN



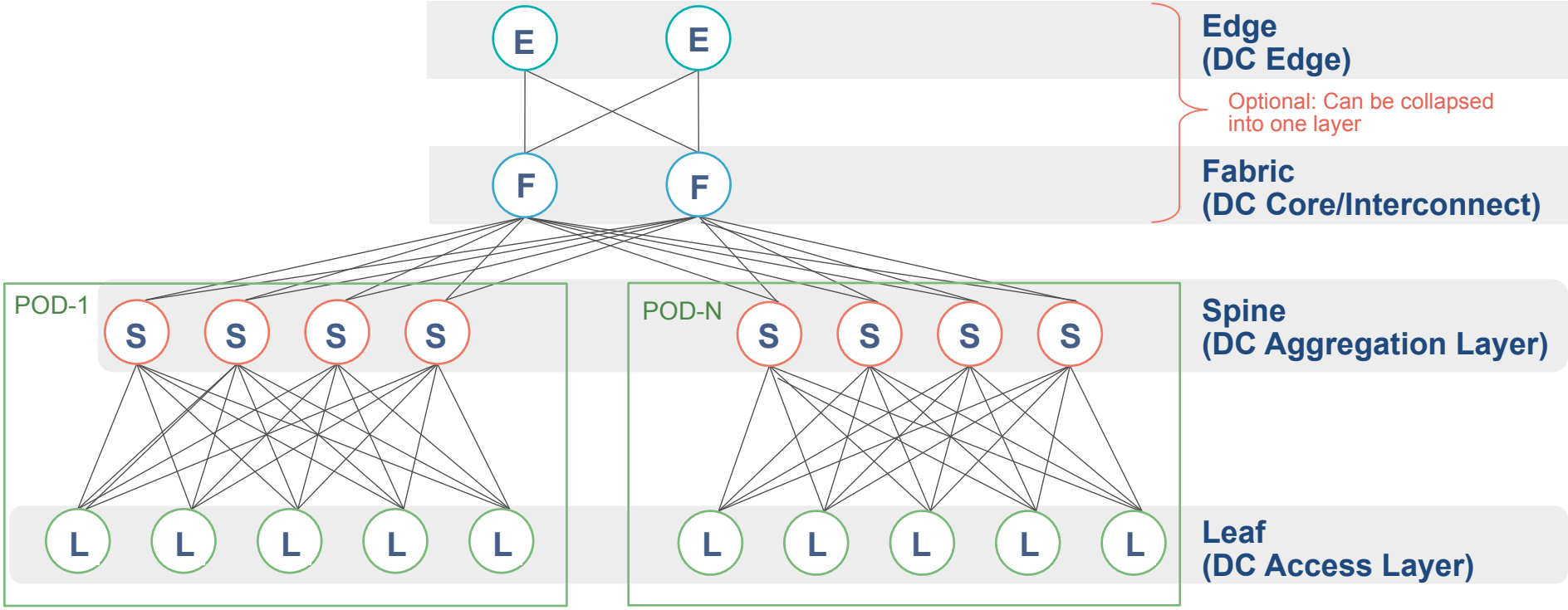
HYBRID CLOUD CONNECTIVITY

Private/Public WAN



BUILDING THE PRIVATE CLOUD – DC FABRIC

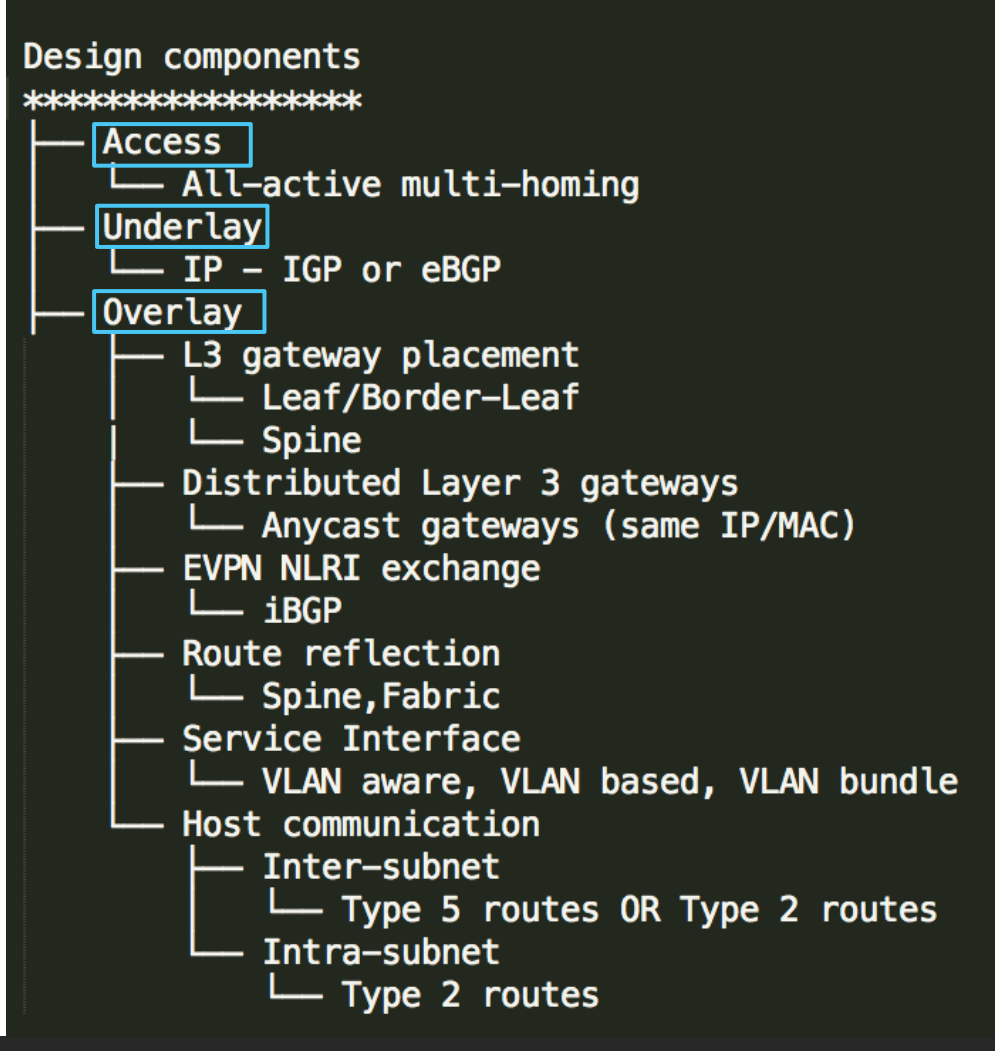
Defining Terminology…



Building the DC Fabric

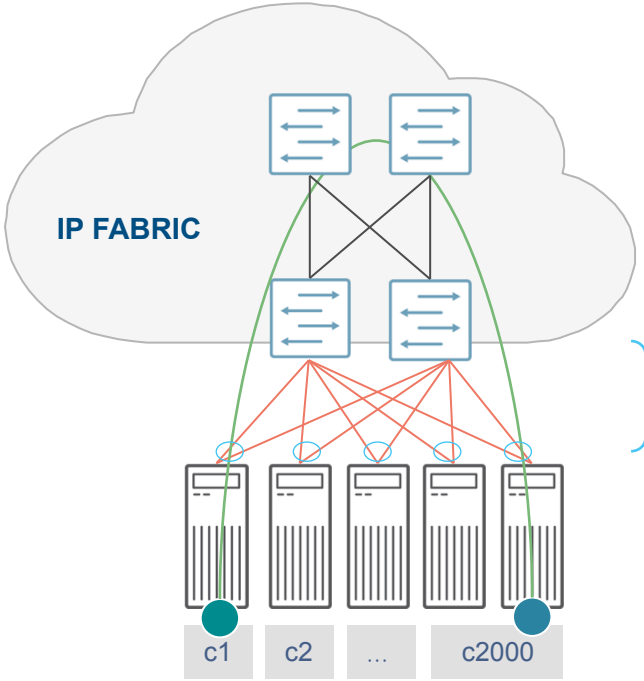
Let's start with the smallest unit – the POD

- Leverage BGP constructs to achieve L2/L3 traffic and multi-tenancy
- L3 gateway placement can be at the leaf or spine
- Hierarchical route-reflection for reduced control plane state and redundancy
- Easy integration with L3VPN with no added provisioning
- Service insertion for EW and/or NS traffic (inter-tenant inter-subnet)



Architectural Flexibility

Containerization influence on network infrastructure

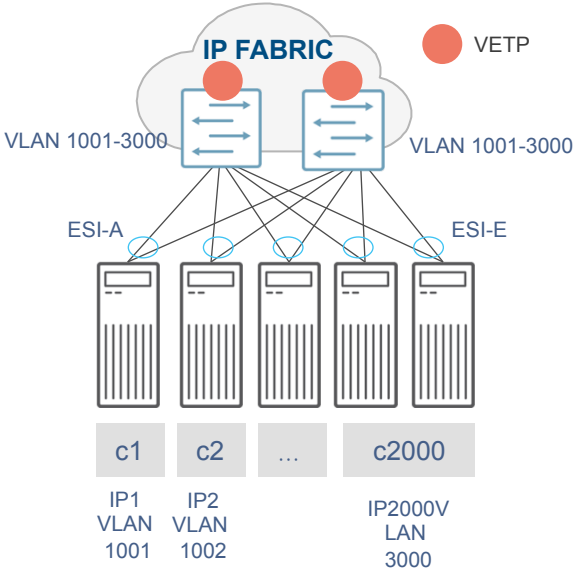


PROBLEM STATEMENT

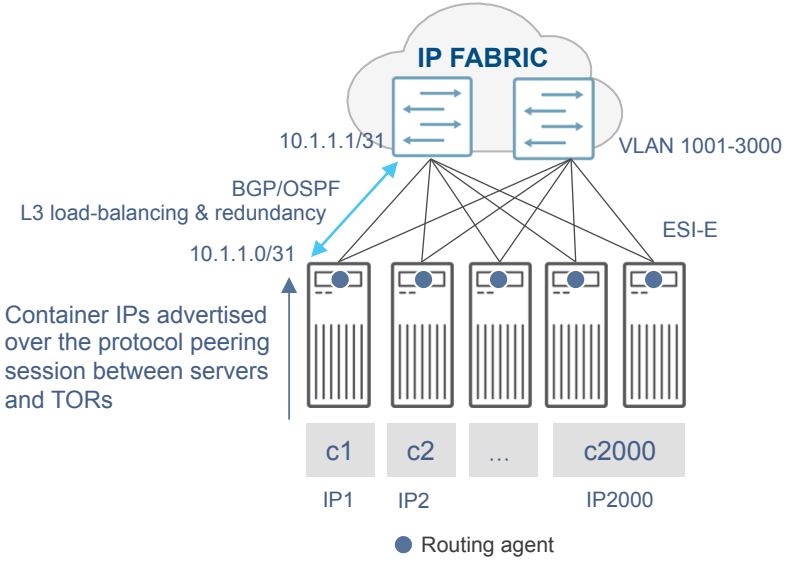
Communication needs to be enabled between 2000 containers residing on servers spread across racks

Connection between servers and TORs can be Layer 2 or Layer 3

Building a Fabric for Containers

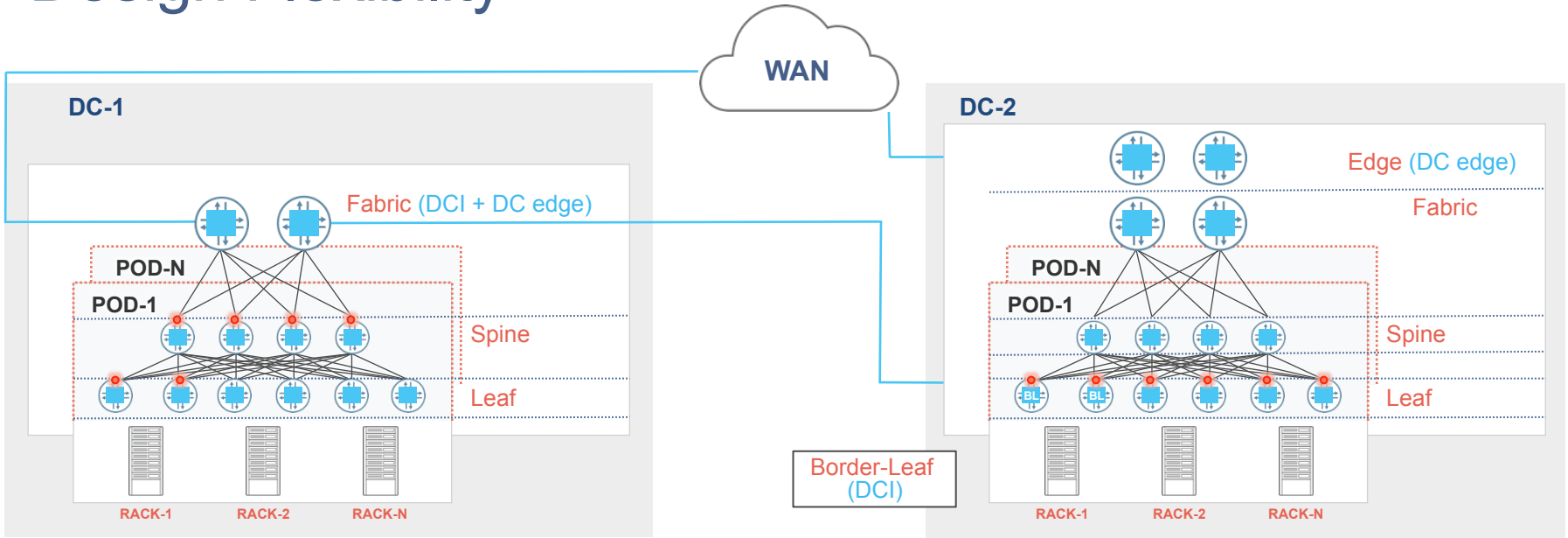


- Layer 2: Trunk ports with each app container identified by a separate VLAN mapped to VNIs on hardware VTEPs
- Benefits: Higher scale capabilities, active-active load balancing with open standards (EVPN N-way multihoming)



- Layer 3: Routing agent can be residing on the server hypervisor
- Benefits: Routing table scale, greater provisioning benefits by using unnumbered addresses for peering between servers and TORs

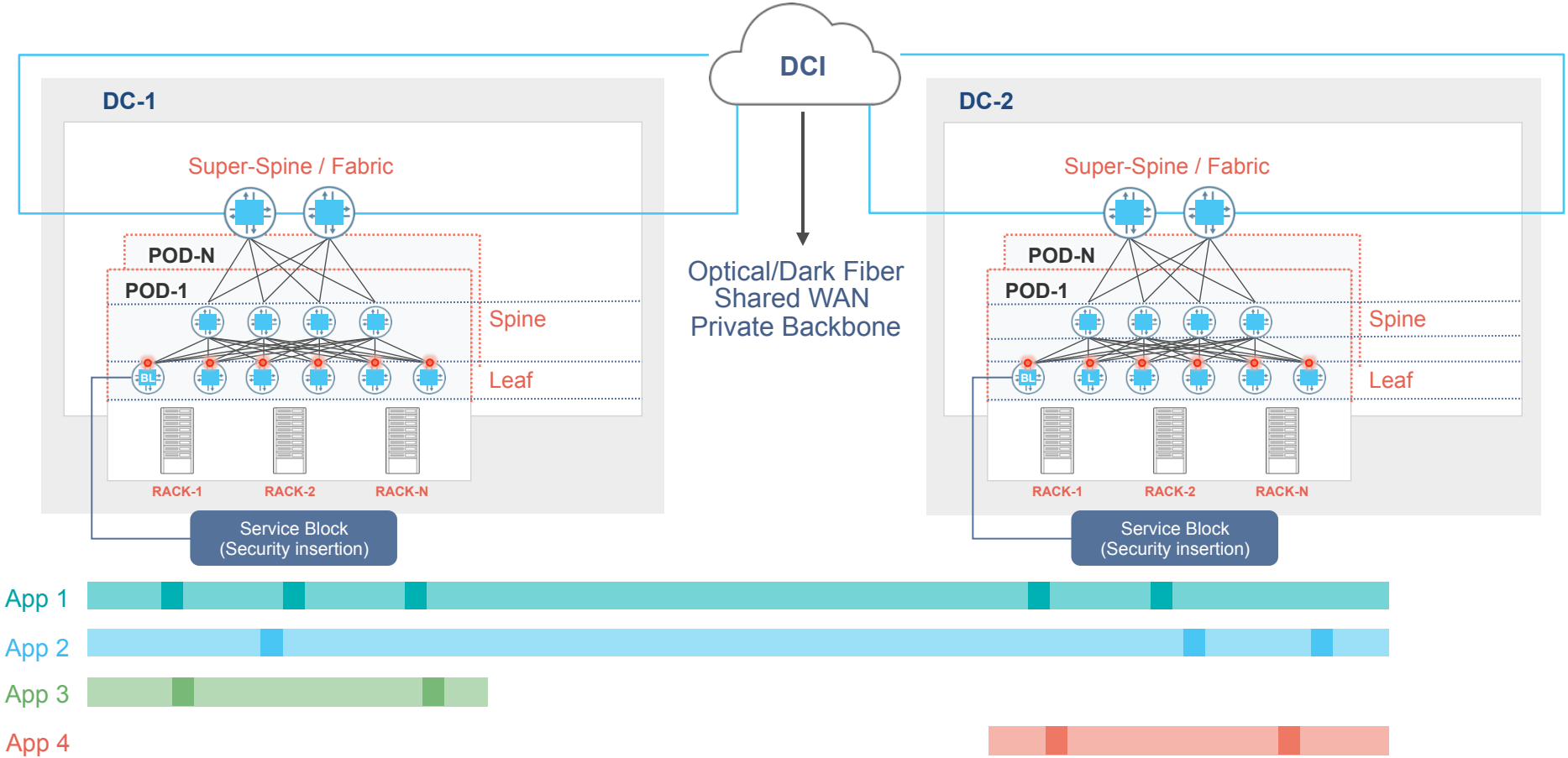
Design Flexibility



Centralized or distributed routing – design choices based on requirements

BUILDING THE PRIVATE CLOUD – DC INTERCONNECT (DCI)

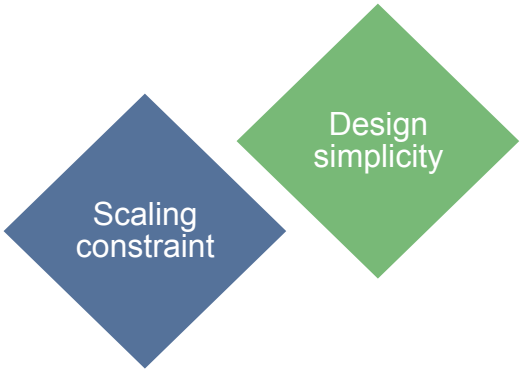
Let's Draw a Picture – DCI



EVPN – DCI Design Options

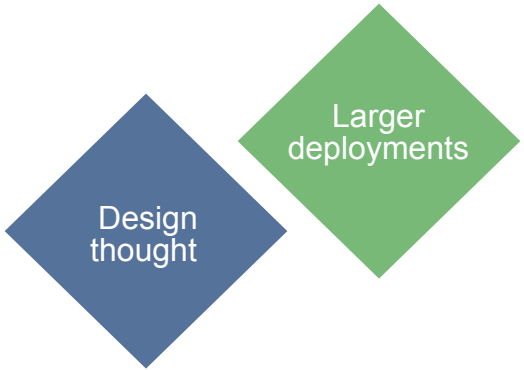
Over The Top (OTT)

- Extended control plane
- Interconnect used as transport (EVPN unaware)



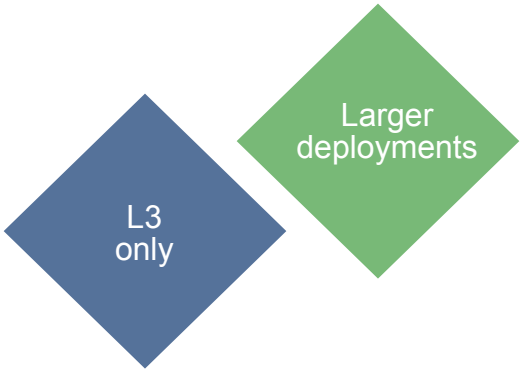
Segmented Approach

- Clear demarcation
- Interconnect EVPN aware
- MPLS TE in core, L2 stretch

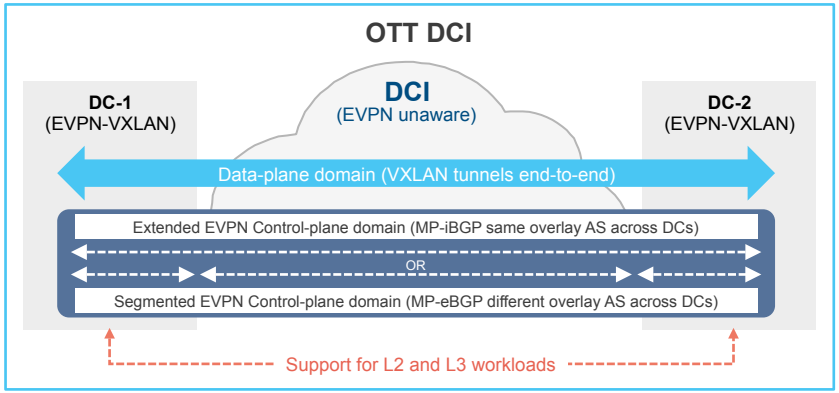


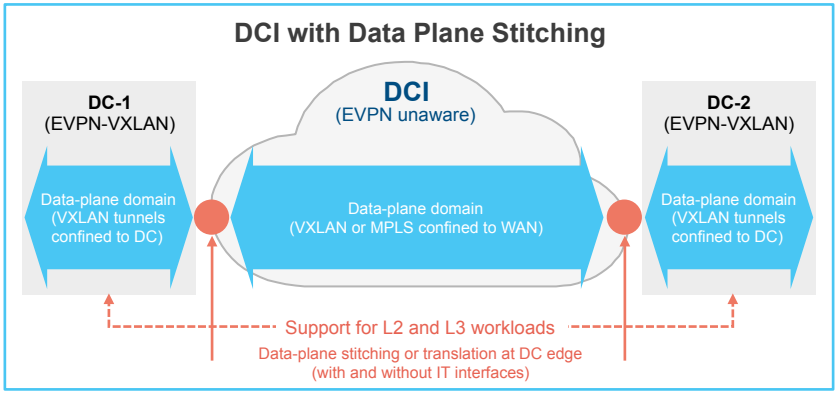
Layer 3 DCI

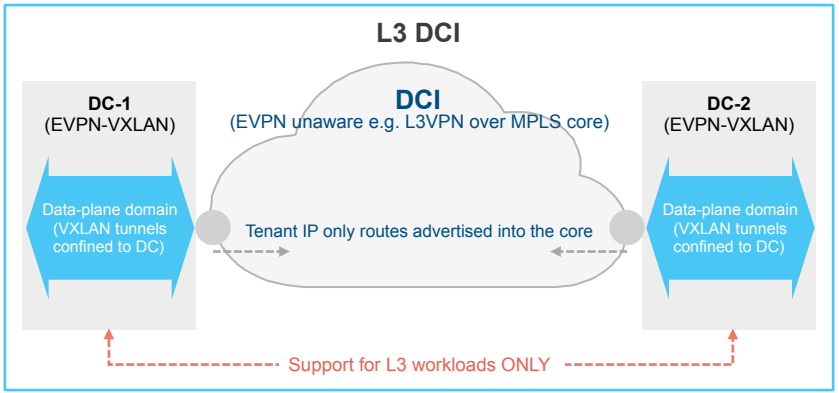
- Clear demarcation
- Interconnect EVPN unaware
- MPLS TE in core, NO L2 stretch



DCI Options

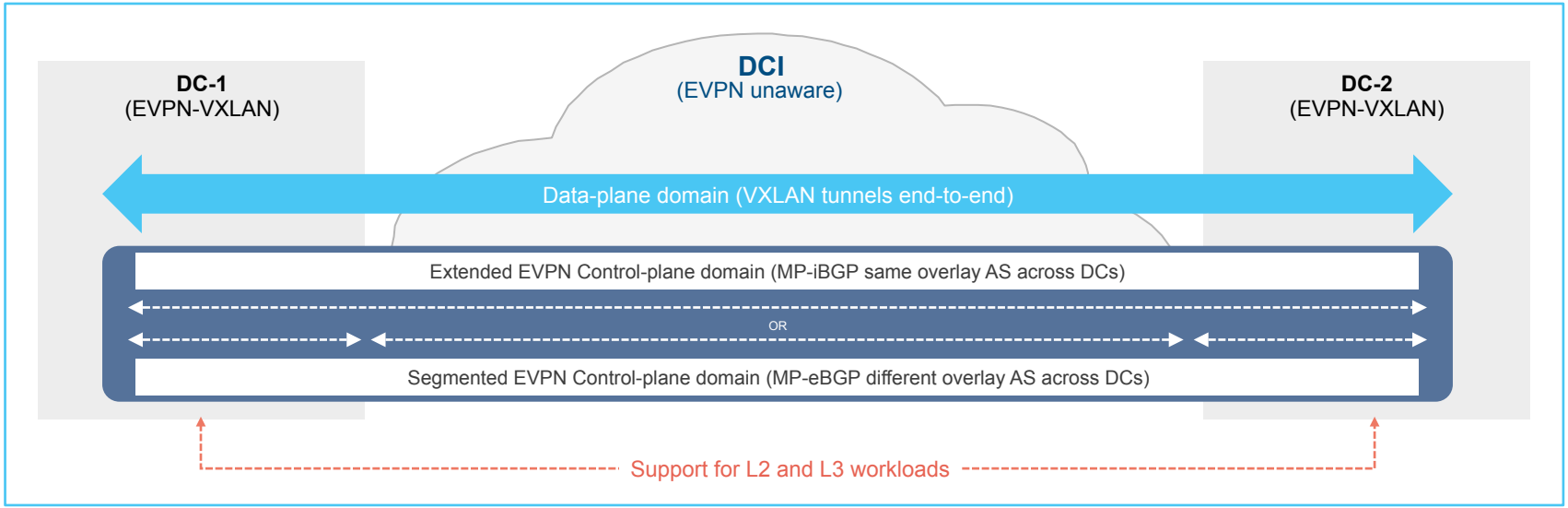

OTT DCI
DC-1
(EVPN-VXLAN)
DCI
(EVPN unaware)
DC-2
(EVPN-VXLAN)
Data-plane domain (VXLAN tunnels end-to-end)
Extended EVPN Control-plane domain (MP-iBGP same overlay AS across DCs)
OR
Segmented EVPN Control-plane domain (MP-eBGP different overlay AS across DCs)
Support for L2 and L3 workloads


DCI with Data Plane Stitching
DC-1
(EVPN-VXLAN)
DCI
(EVPN unaware)
DC-2
(EVPN-VXLAN)
Data-plane domain
(VXLAN tunnels
confined to DC)
Data-plane domain
(VXLAN or MPLS confined to WAN)
Data-plane domain
(VXLAN tunnels
confined to DC)
Support for L2 and L3 workloads
Data-plane stitching or translation at DC edge
(with and without IT interfaces)


L3 DCI
DC-1
(EVPN-VXLAN)
DCI
(EVPN unaware e.g. L3VPN over MPLS core)
DC-2
(EVPN-VXLAN)
Data-plane domain
(VXLAN tunnels
confined to DC)
Tenant IP only routes advertised into the core
Data-plane domain
(VXLAN tunnels
confined to DC)
Support for L3 workloads ONLY

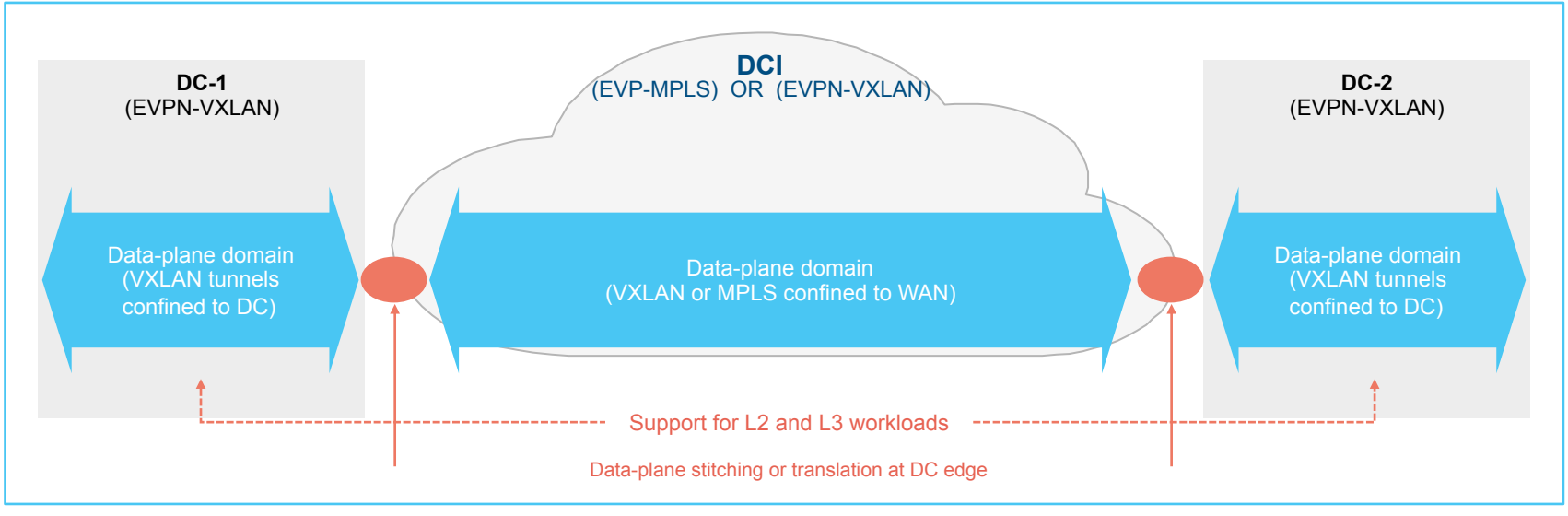
Over the Top (OTT) – DCI

Control plane is extended across sites with the connecting infrastructure used as transport only (EVPN unaware)



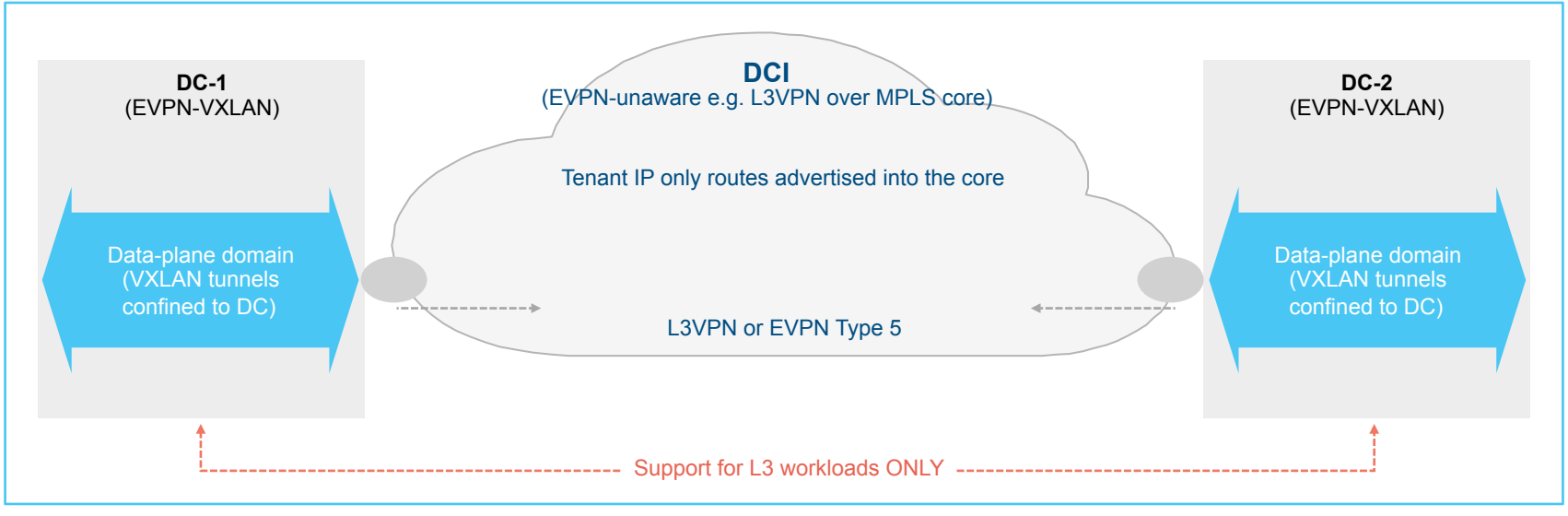
Segmentation of DC & WAN domains

Clear demarcation of DC and WAN boundaries, connecting infrastructure is EVPN aware



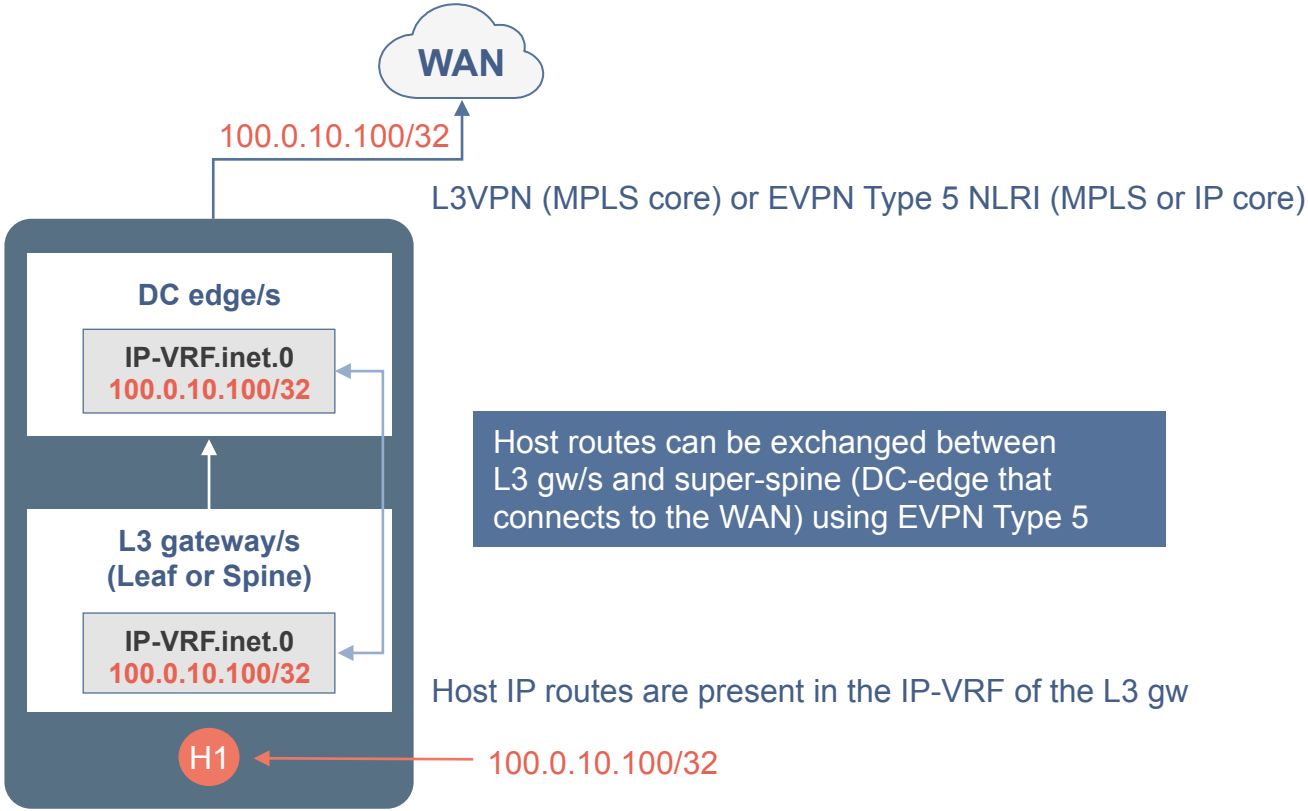
Layer 3 DCI

Only Layer 3 connectivity is extended across DCs (no Layer 2).
Data plane domain is confined within DC and not extended across DCs.



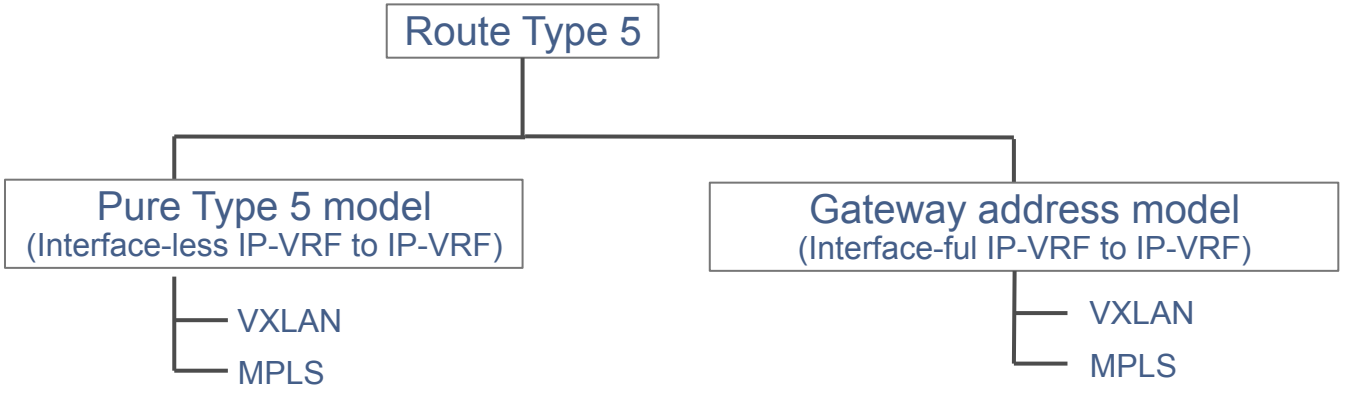
BUILDING THE PRIVATE CLOUD – WAN INTEGRATION

How are host IP prefixes exchanged between L3 gateway/s and DC edge/s so as to be advertised out of the DC?



EVPN Route Type 5 – Classification

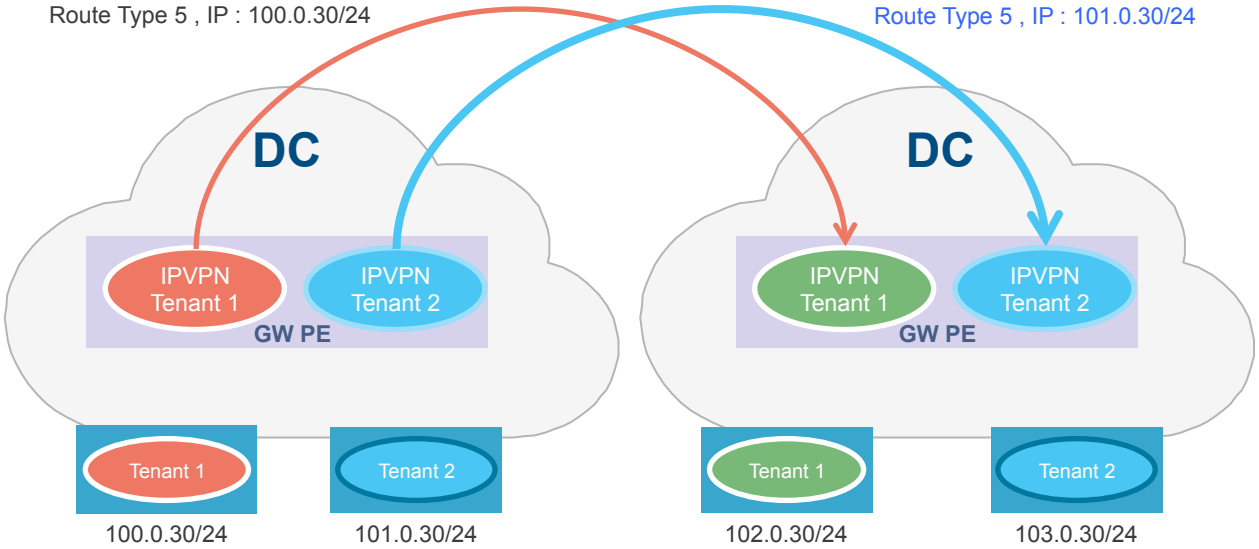
draft-ietf-bess-evpn-prefix-advertisement



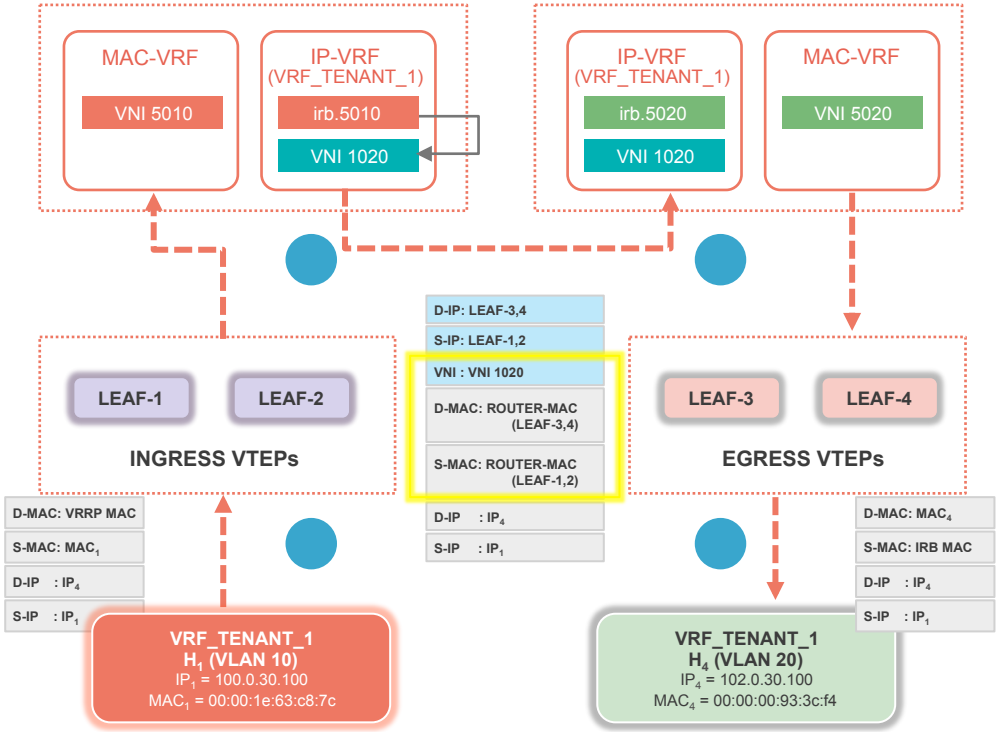
Type 5 route provides all necessary forwarding information

Type 5 route needs recursive route resolution for forwarding. The lookup is for an IP prefix but forwarding information is extracted from Type 2 route

Pure Route Type 5 Model



Packet Walk – Pure Route Type 5



EVPN Route Type 5 vs L3VPN NLRI

EVPN Pure Route Type 5 NLRI

* 5:10.1.1.30:30::0::100.0.30.0::24/304

RD Prefix info

Similar information carried across both NLRI types

L3VPN NLRI

10.1.1.30:30: 100.0.30.0/24

RD Prefix info

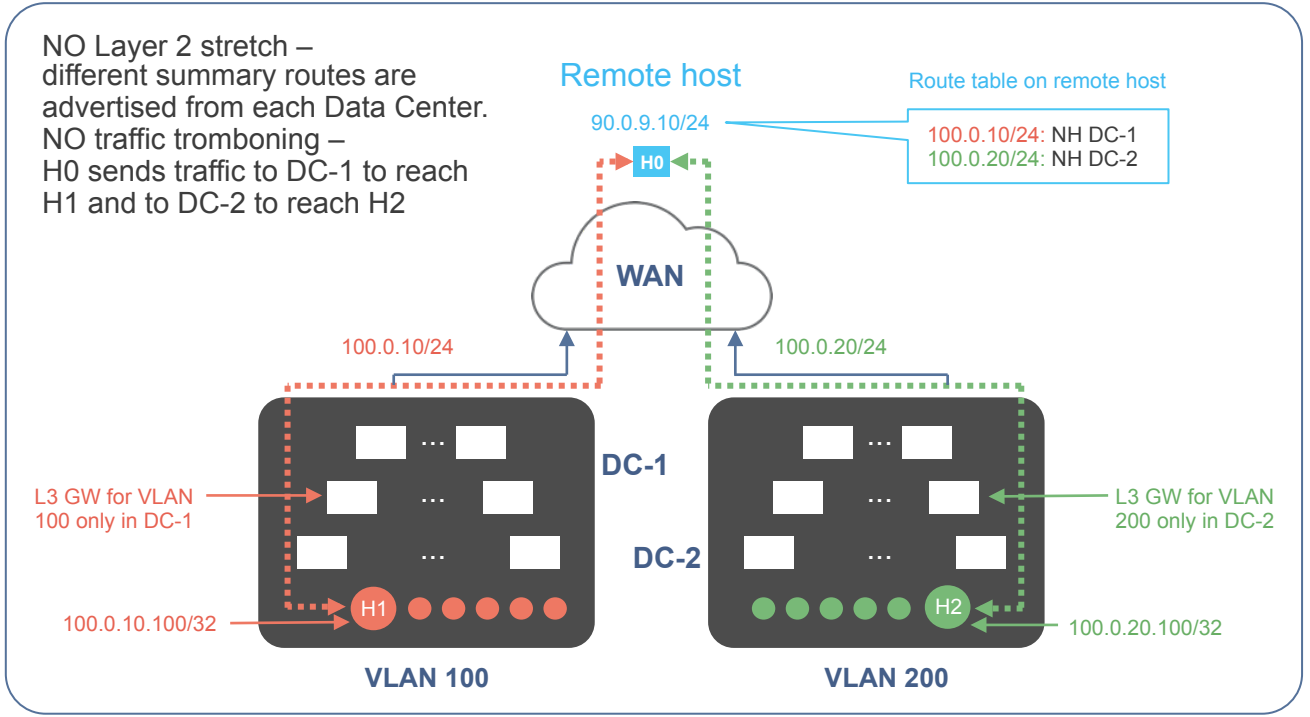
Benefits with EVPN Type 5

- Unified solution end to end with one address family inside the DC and outside
- Data plane flexibility with EVPN – use over MPLS or IP core
- If you do not have MPLS between DCs for DCI
  - It is not possible to run L3VPN over VXLAN
  - For control plane, Route Type 5 is the only option
- Hybrid cloud connectivity (Type 5 with VXLAN over GRE/IPsec)

BUILDING THE PRIVATE CLOUD – TRAFFIC OPTIMIZATION

What is VMTO ?
Virtual **M**achine **T**raffic **O**ptimization

Resolves ingress and egress traffic tromboning, focusing on north-south traffic optimization

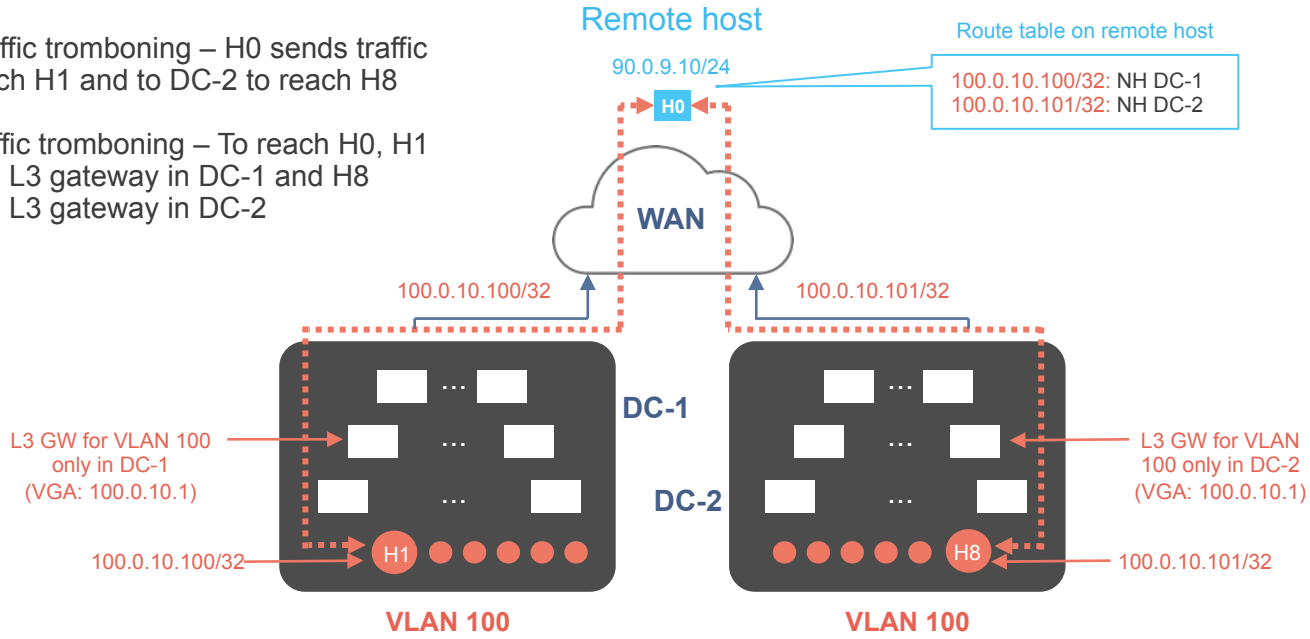


Ingress and Egress North-South Traffic Optimization

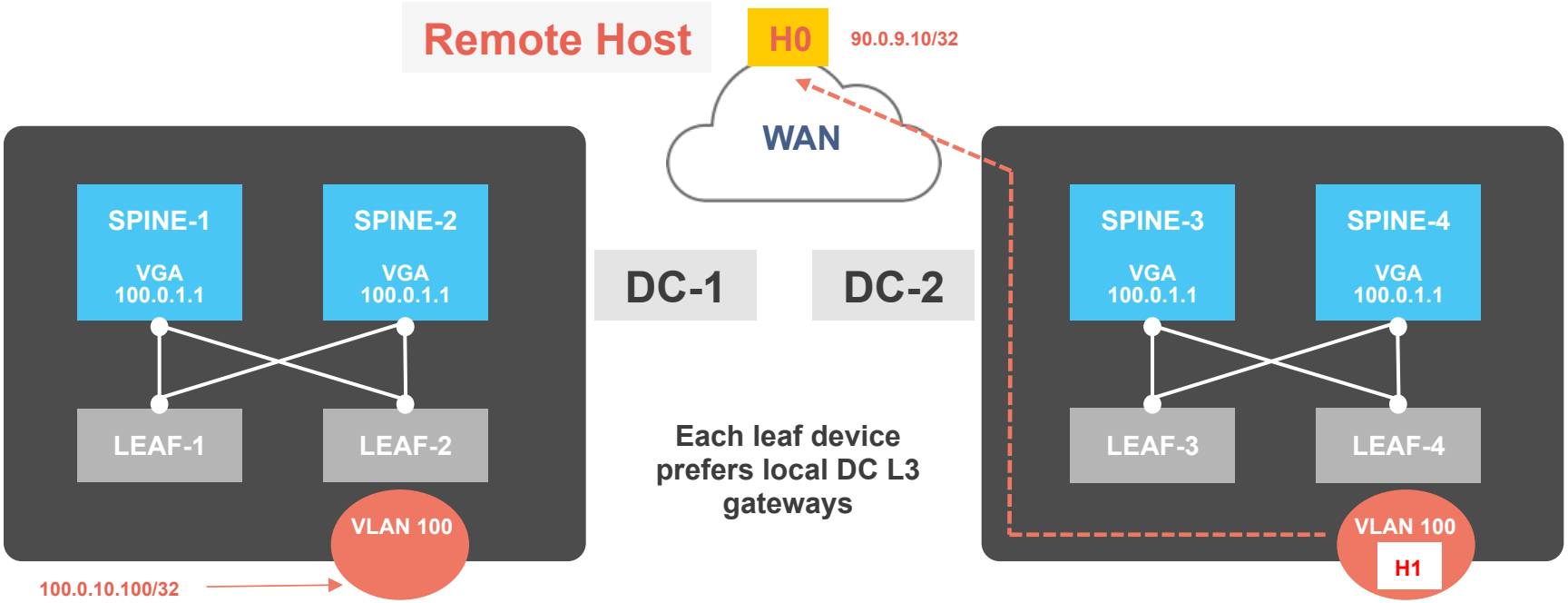
Layer 2 stretch – Host prefix routes are advertised from each Data Center

NO ingress traffic tromboning – H0 sends traffic to DC-1 to reach H1 and to DC-2 to reach H8

NO egress traffic tromboning – To reach H0, H1 sends traffic to L3 gateway in DC-1 and H8 sends traffic to L3 gateway in DC-2

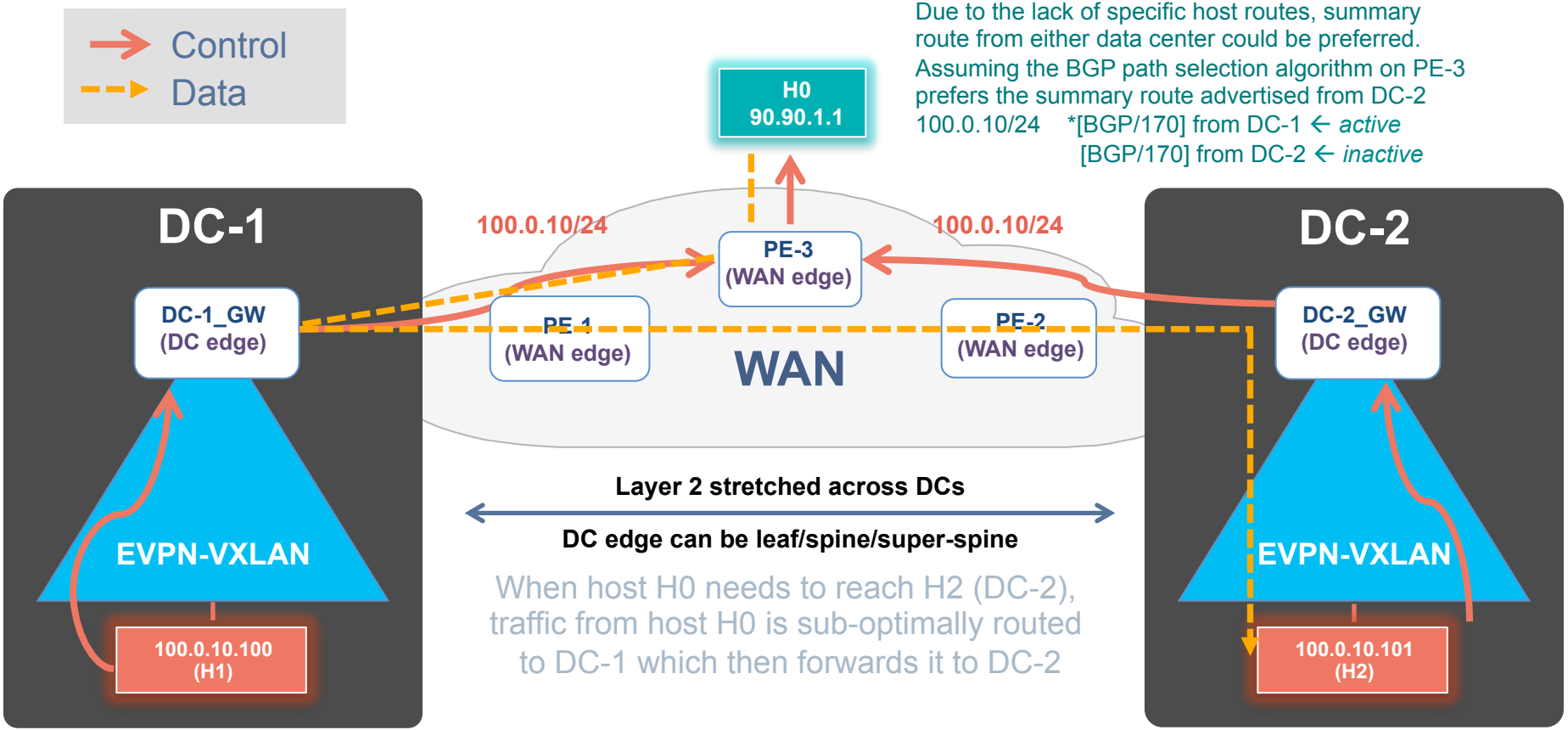


How to Avoid Egress Tromboning?

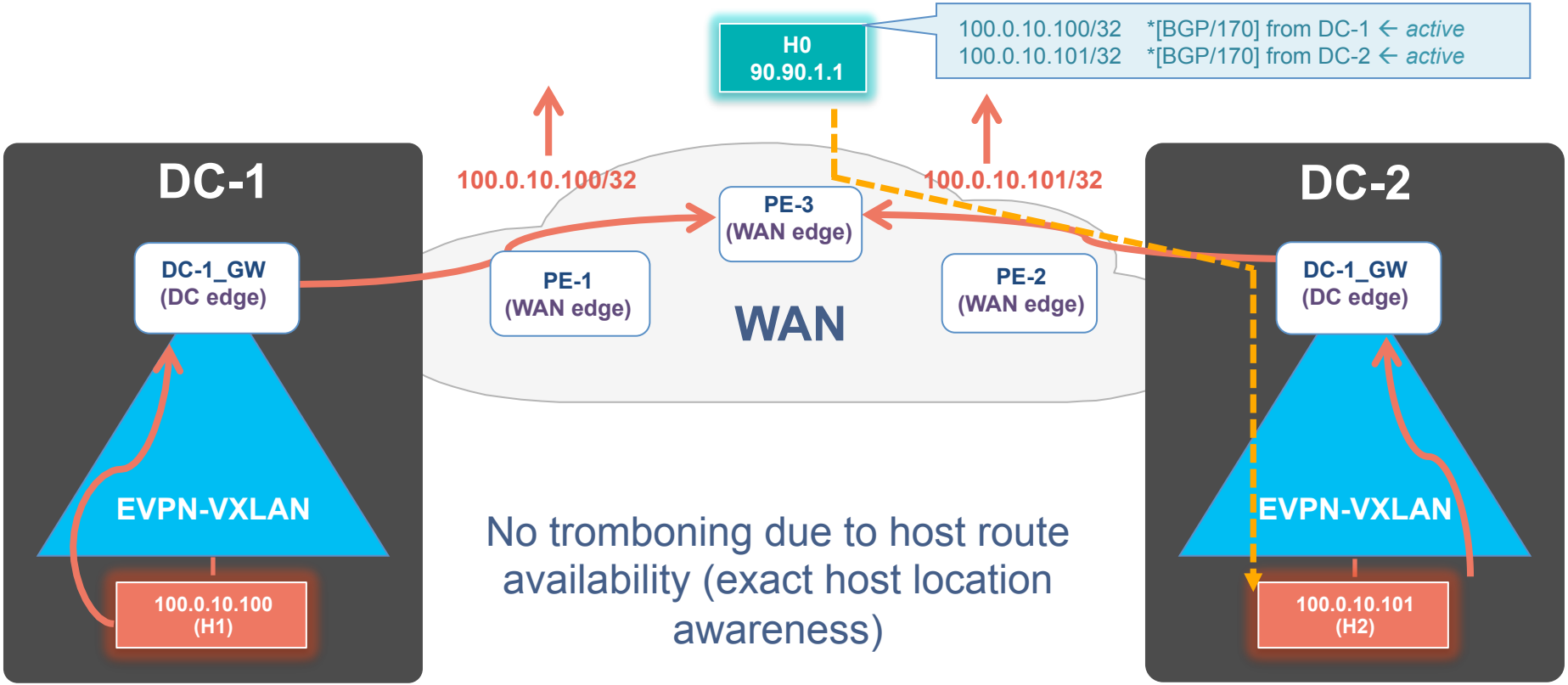


Distributed layer 3 anycast gateway function ensures, local DC gateway preferred (even on host migration)

How to Avoid Ingress Tromboning?



No Ingress Tromboning



No tromboning due to host route availability (exact host location awareness)

Thank You