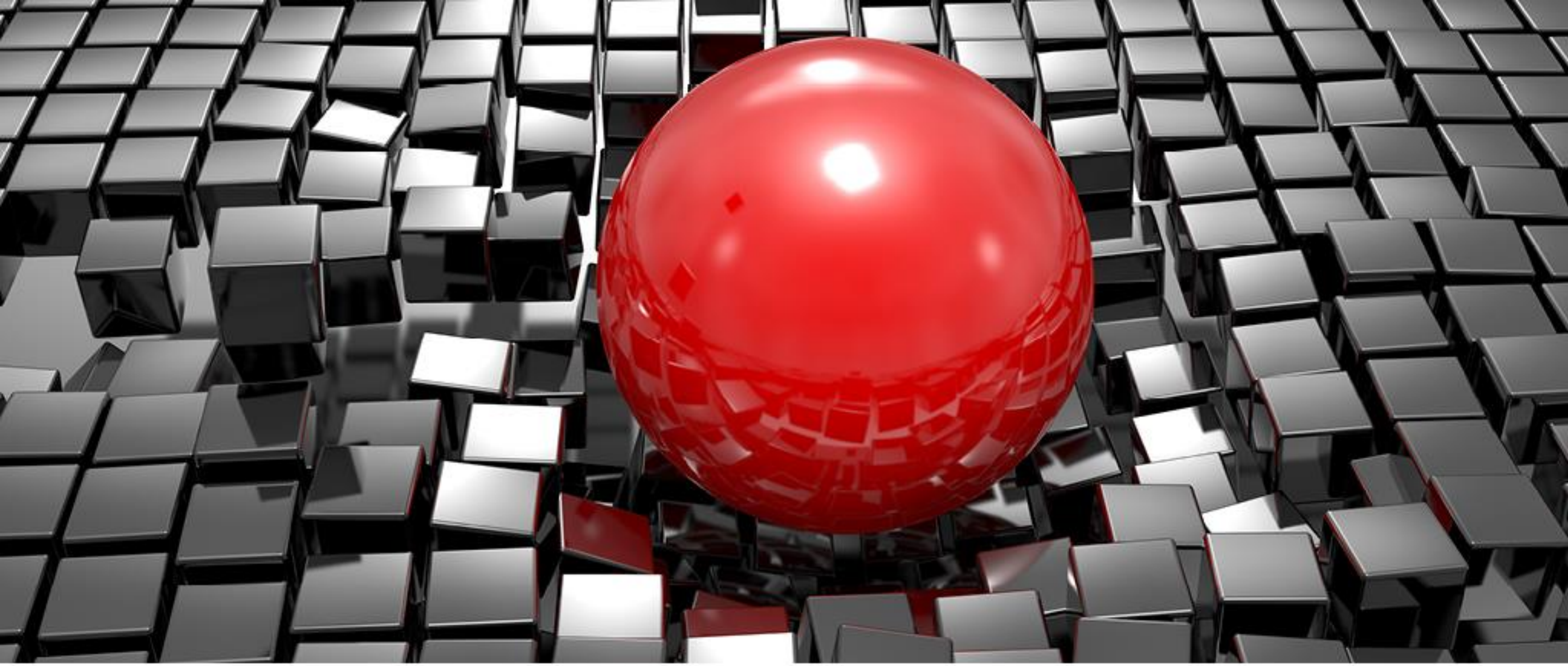


THE REVOLUTION OF Customer Optimized Datacenter

Feb, 2015

Seo, Young Seog

Architect / NAIM Networks

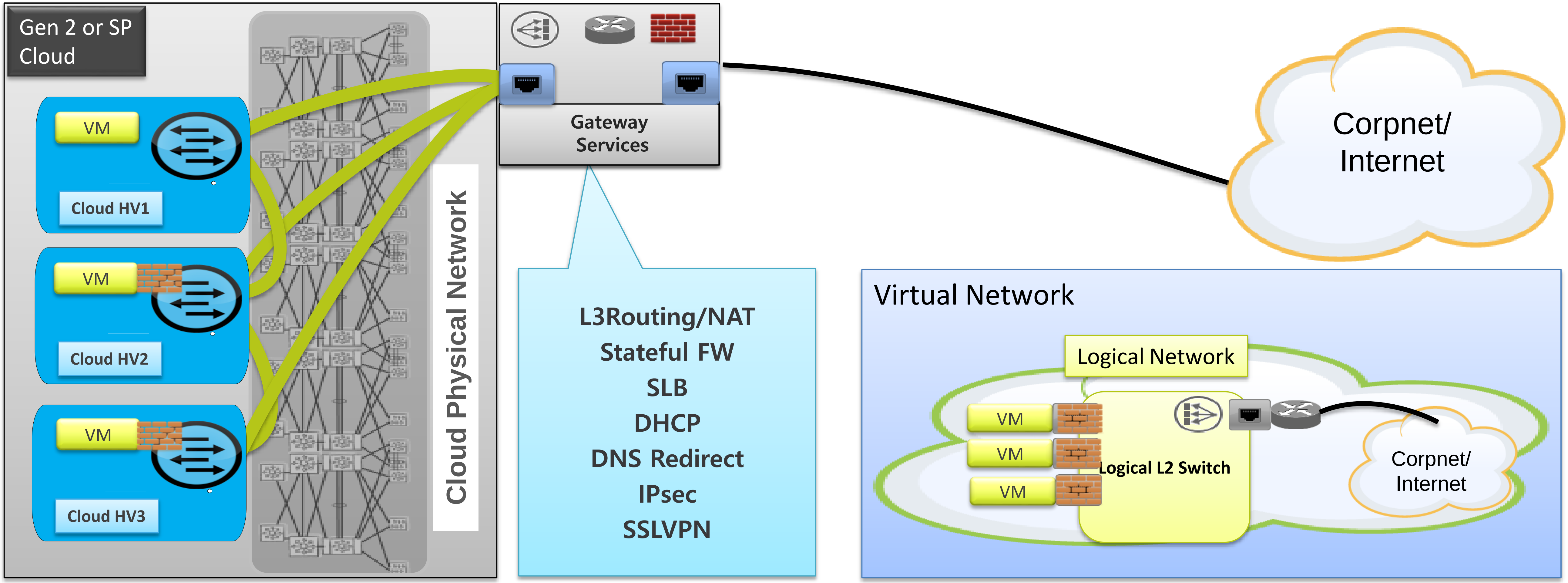
NAIM networks

DataCenter Design

1. Automate Network Services Provisioning

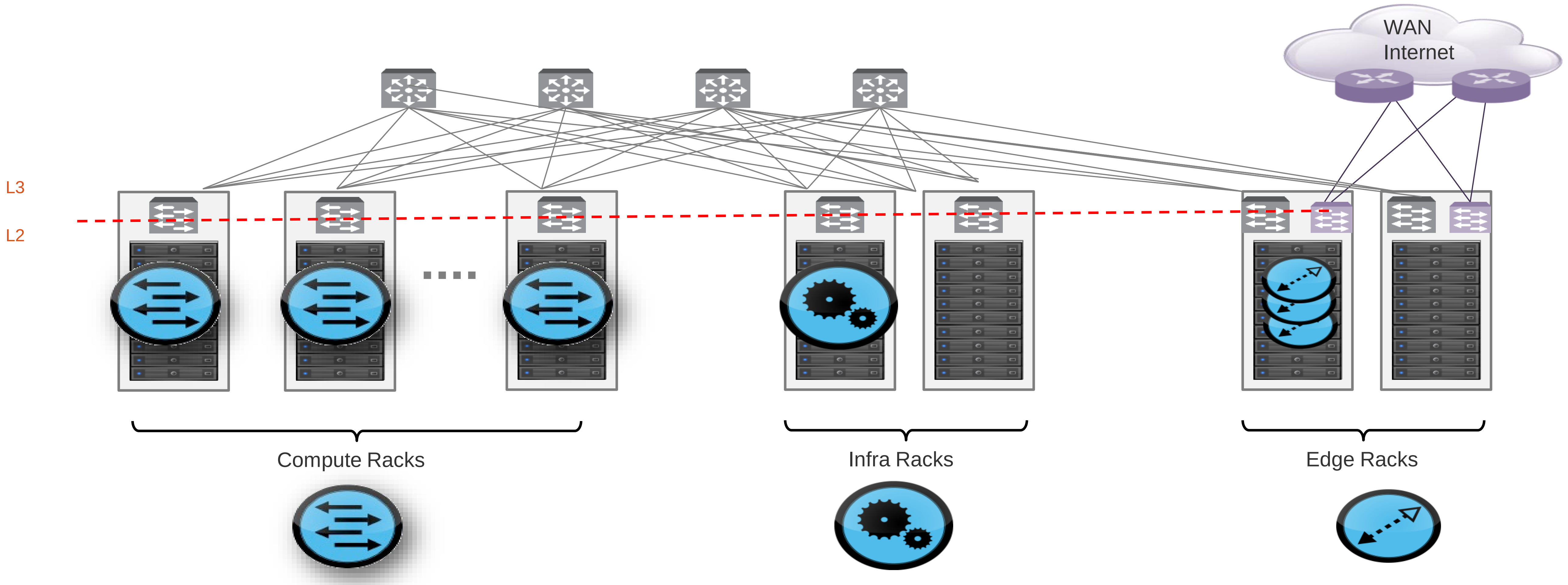
2. Micro segmentation to enhance security (DMZ / PCI, etc.)

3. Augment Physical L4-7 Appliances



L3Routing/NAT
Stateful FW
SLB
DHCP
DNS Redirect
IPsec
SSLVPN

DataCenter Design



Hypervisor Service Modules

- Distributed network services

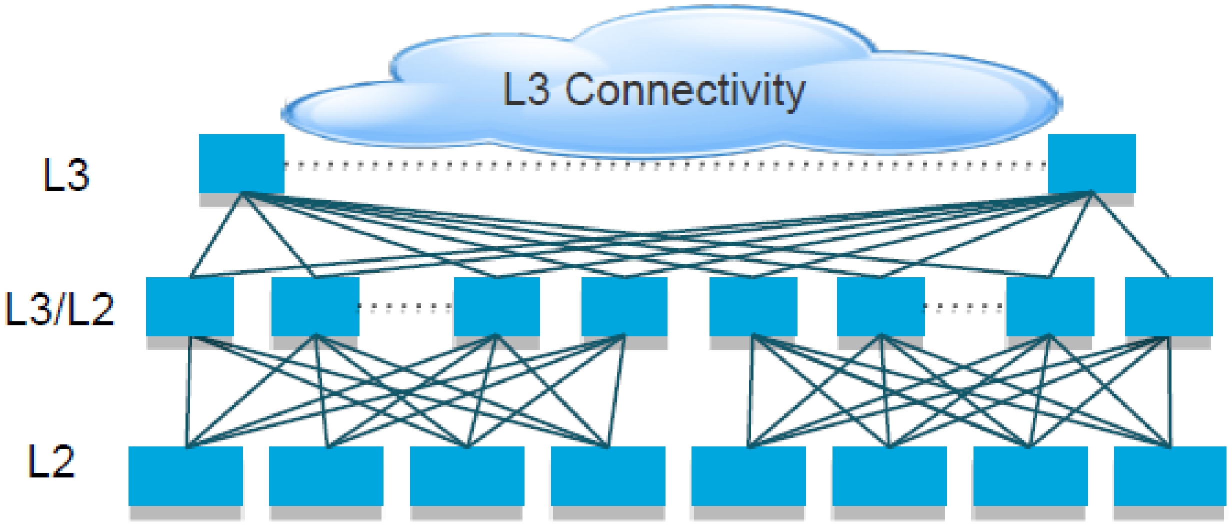
Controller Node

- Virtual network hypervisor
- Massive scale

Gateway Software

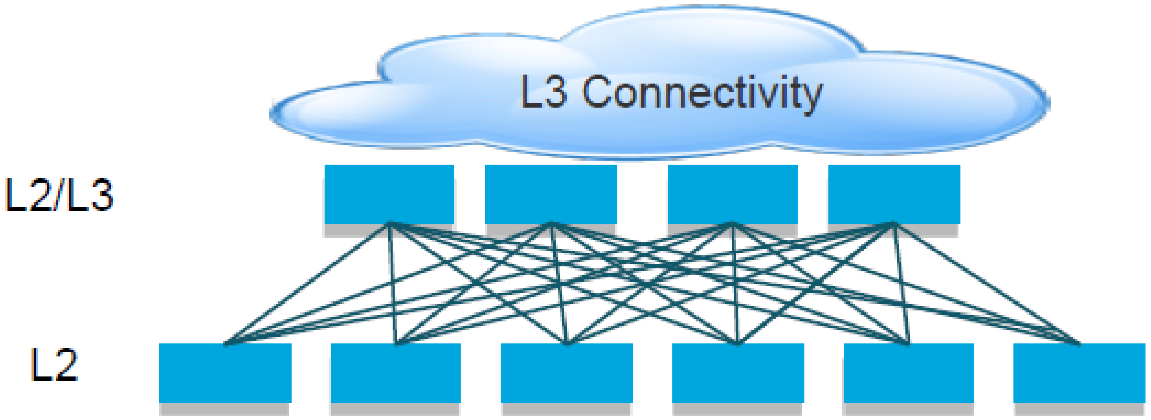
- Integration with existing physical infra.
- V to V / V to P

DataCenter Design



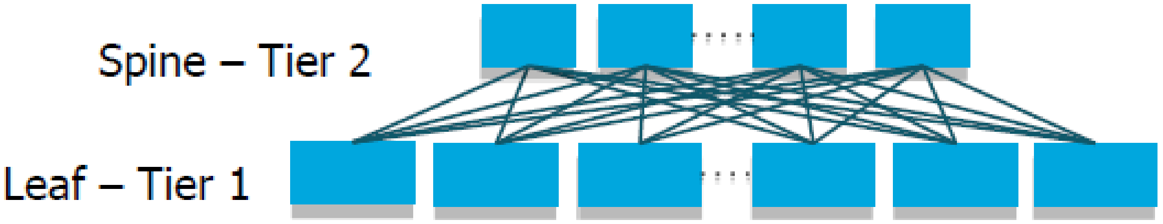
Multi-Tier

- Scalable 3-Tier design
- STP, VLAN spread
- Expensive, not ideal for Greenfields deployments



L2 Fabric - VLAN based

- Larger L2 domains, reliance on STP
- Comparatively limited in scalability – 2-tier design
- Generally industry is moving away from L2 fabrics



Leaf/Spine

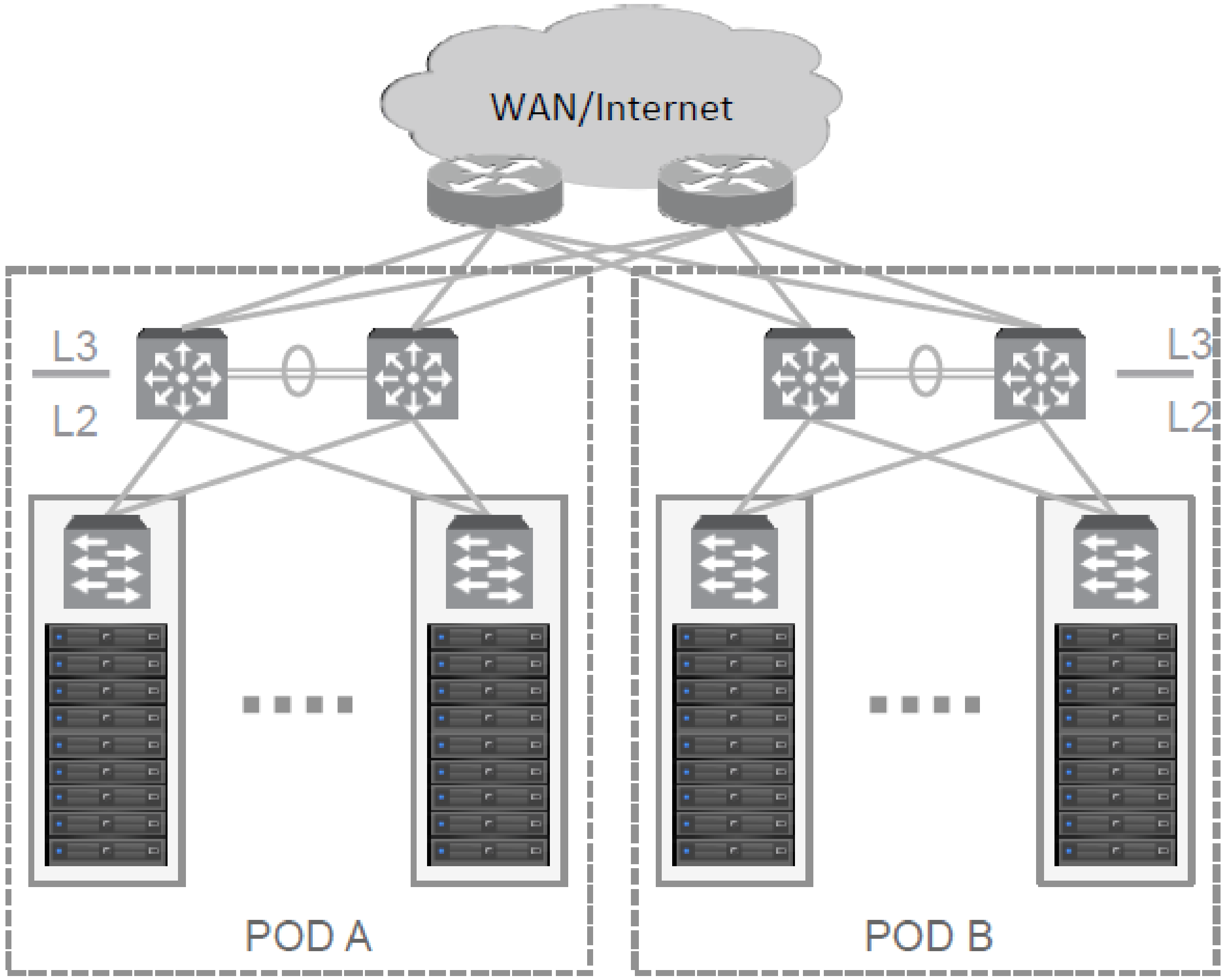
- Virtualization and Big Data applications are major contributors to East-West traffic growth – up to 75%
- Trill or L3
- Leaf-Spine design allows for:

  Uniform access and consistent latency
  N way ECMP – Link utilization and HA

DataCenter Design

Classical Access/Agg/Core Network

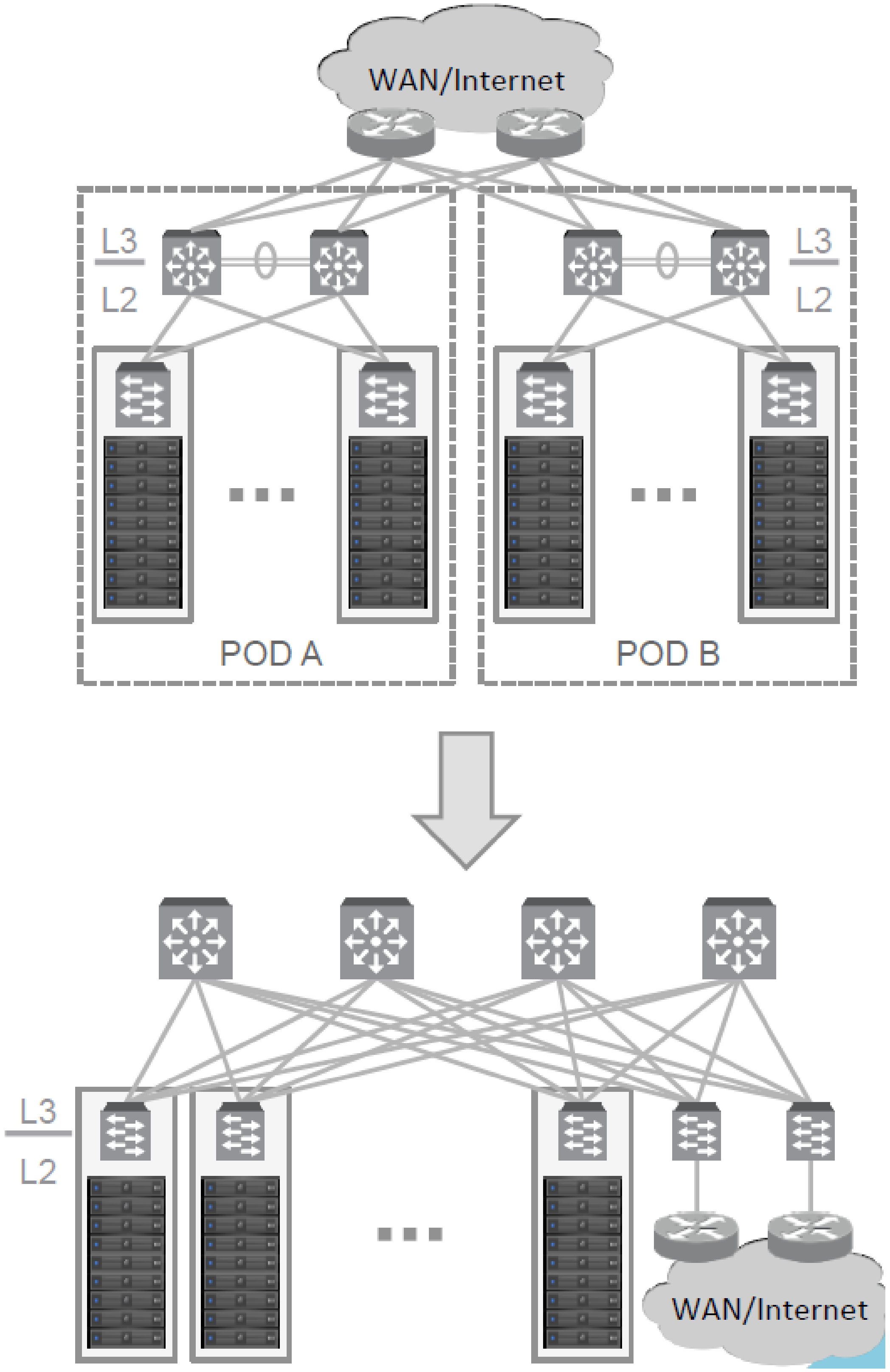
- VLANs carried throughout the Fabric
- Incremental configuration in the Fabric when adding/removing VLANs
- L2 application scope is limited to a single POD
- Layer 2 diameter is the failure domain size
- Multiple aggregation modules, to limit the Layer 2 domain size



DataCenter Design

Physical Network Trends

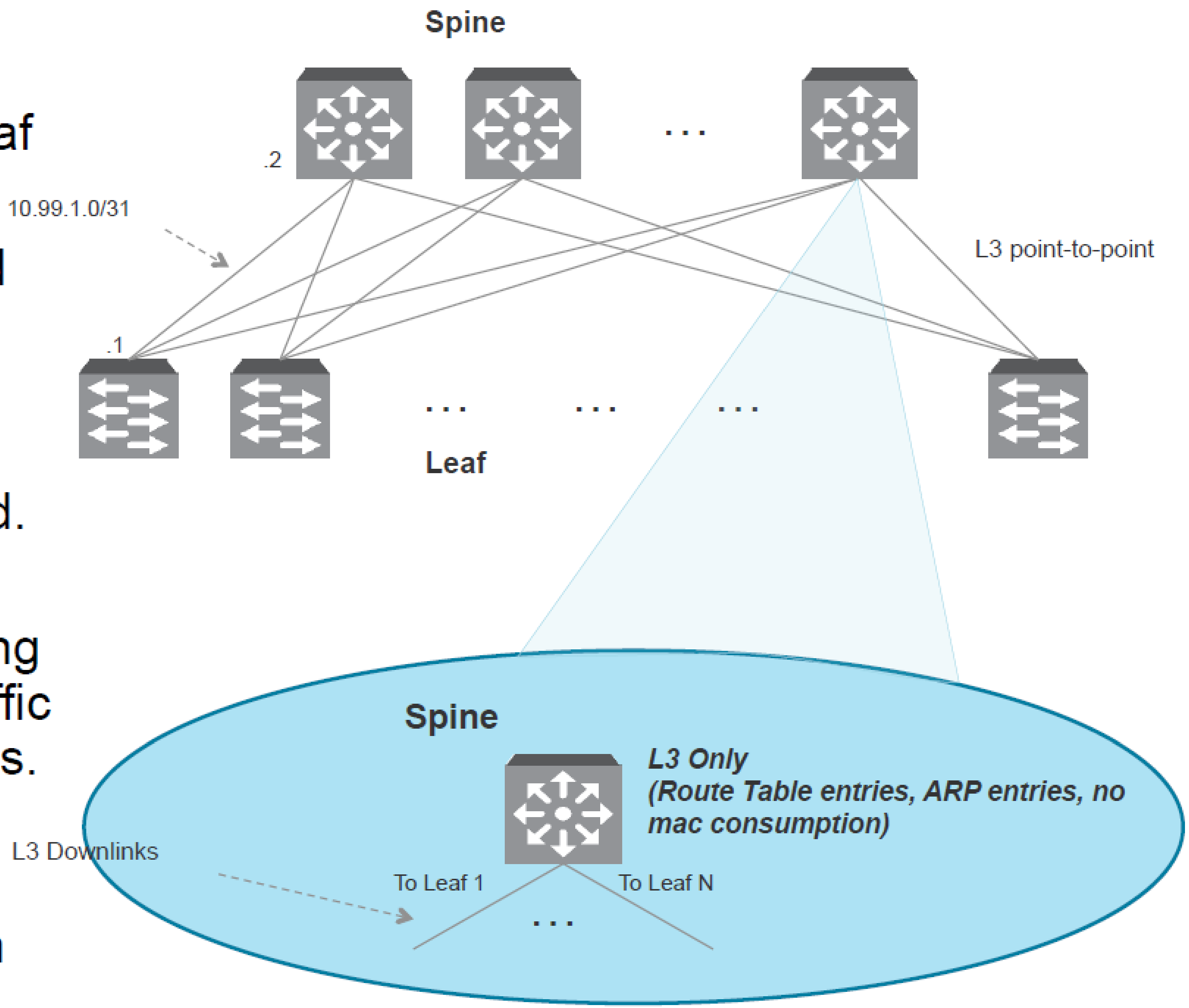
- From 2- or 3-tier to spine/leaf
- Density & bandwidth jump
- ECMP for layer 3 (and layer 2)
- Reduce network oversubscription
- Wire & configure once
- Uniform configurations



DataCenter Design

Spine Nodes

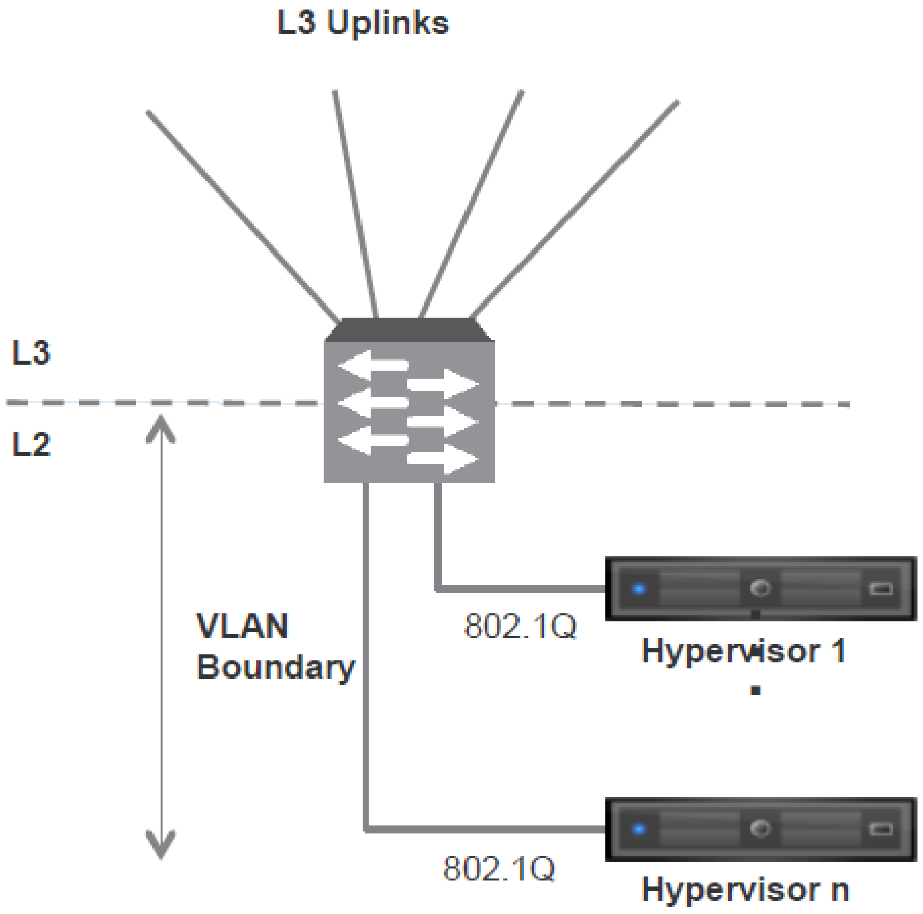
- Spine connects to leaf switches
- Interfaces configured as routed point-to-point L3 links.
- Links between spine switches not required.
- In case of a spine to leaf link failure, routing protocol reroutes traffic on the alternate paths.
- Aggregates the leaf nodes and provide connectivity between racks.



DataCenter Design

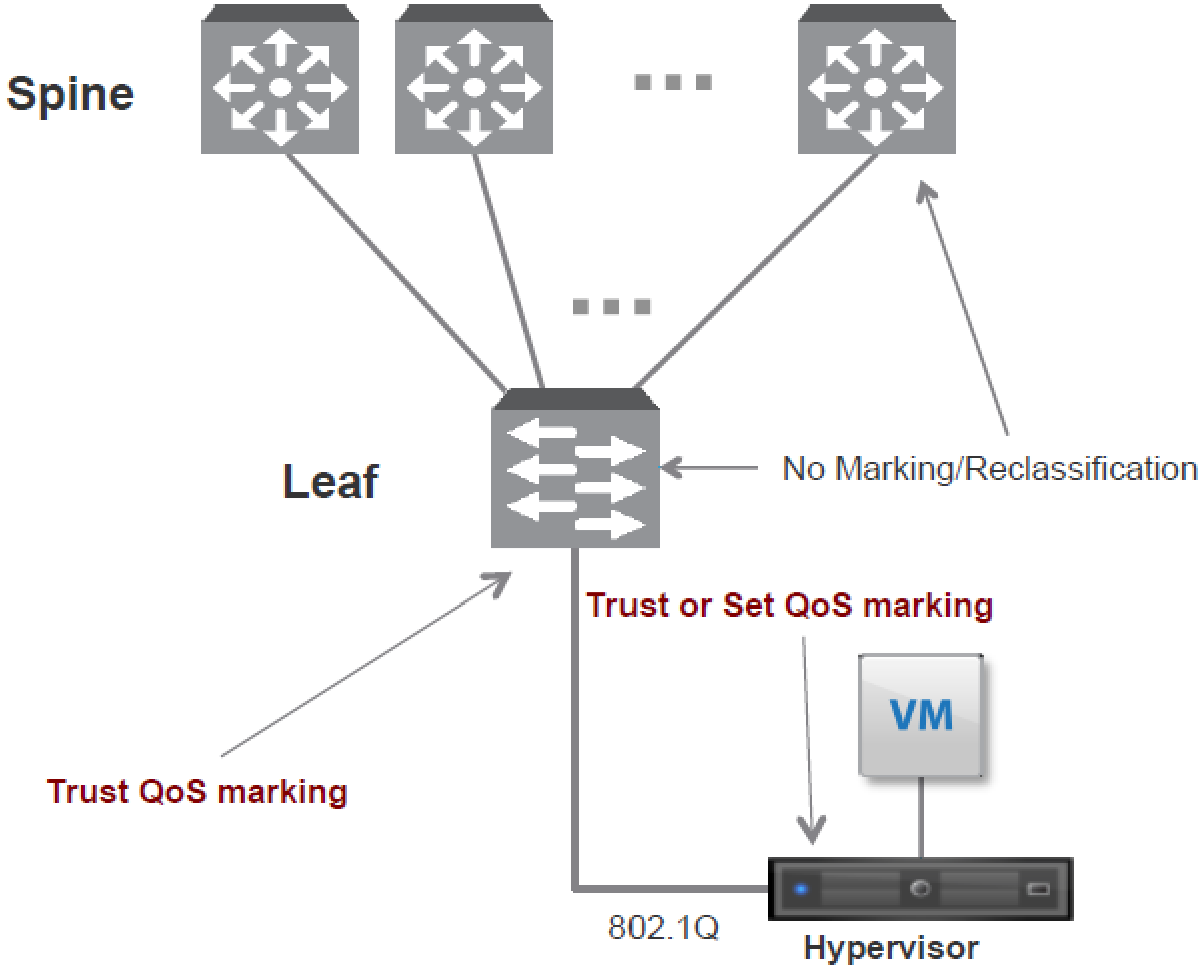
Leaf Nodes

- L3 ToR designs have dynamic routing protocol between leaf and spine.
- BGP, OSPF or ISIS can be used
- Rack advertises small set of prefixes (one per VLAN/subnet).
- Equal cost paths to the other racks prefixes.
- Switch provides default gateway service for each VLAN subnet
- Servers facing ports have minimal configuration
- 801.Q trunks with a small set of VLANs for VMkernel traffic

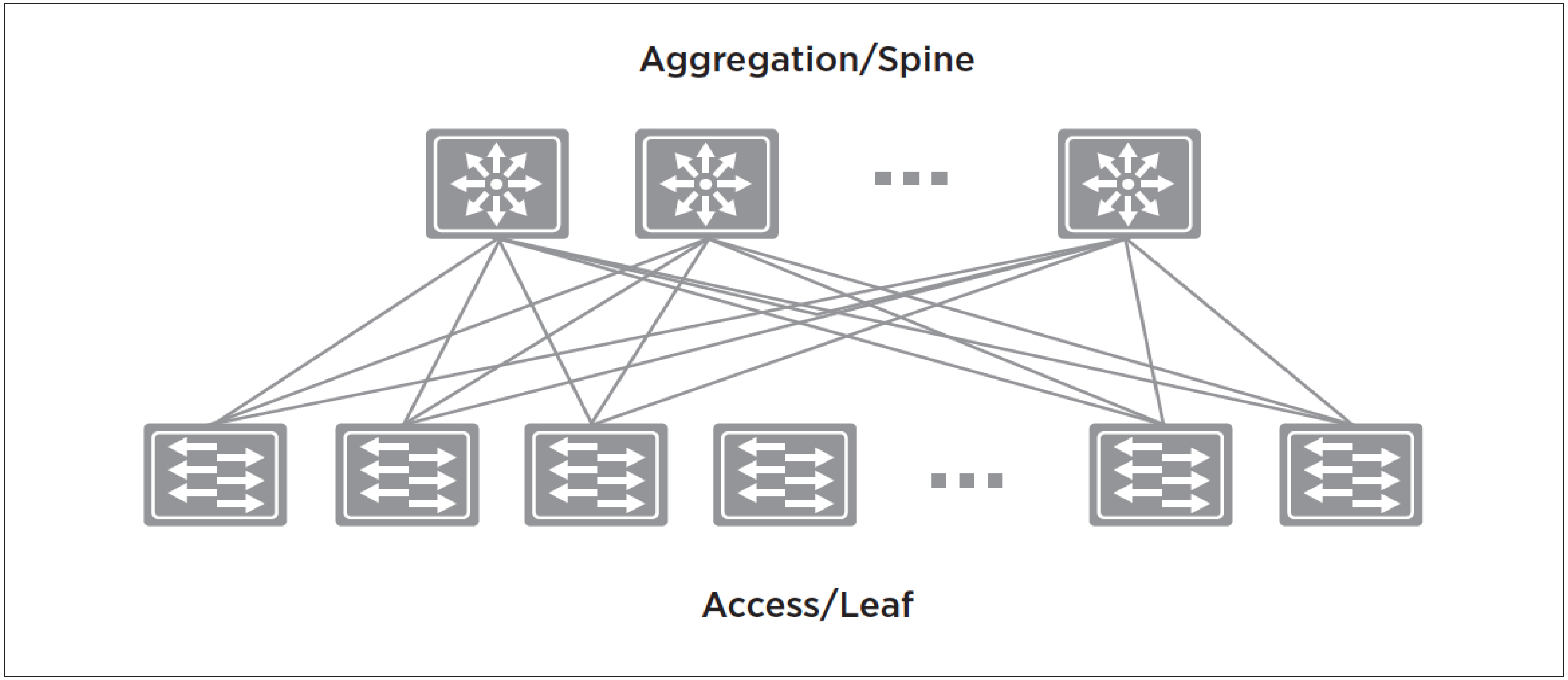


QoS in Virtualized Datacenter Designs

- Virtualized environments carries different types of traffic
- Hypervisor is trusted boundary, sets respective QoS values
- The physical switches “trust” values. No reclassification at Leaf
- QoS values determine what traffic to prioritize in case of congestion



DataCenter Consideration



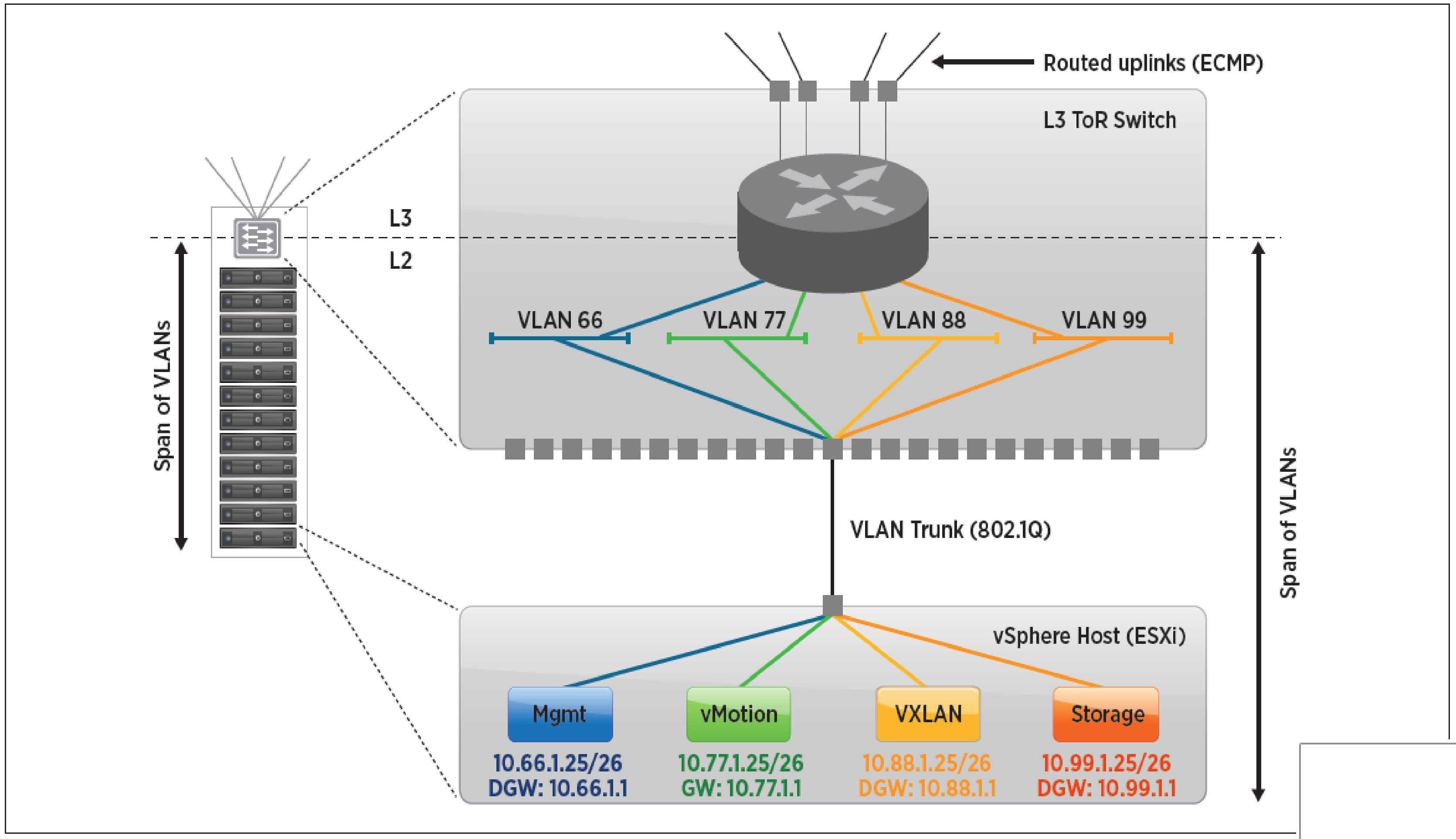
One of the key goals of network virtualization is to **provide virtual-to-physical network abstraction**.

- Simple
- Scalable

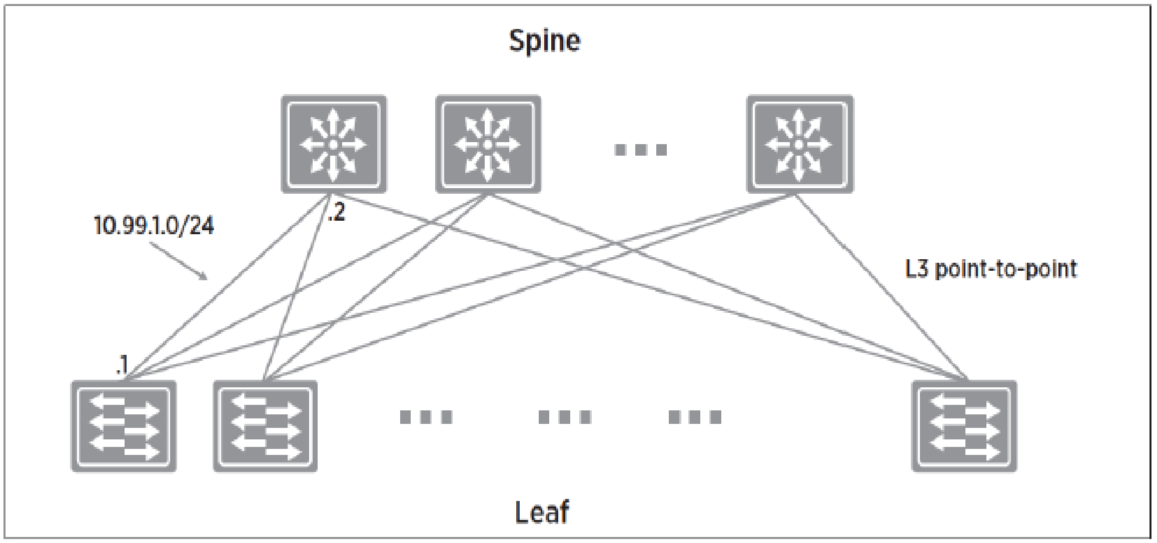
- High-bandwidth
- Fault-tolerant
- QoS-providing

DataCenter Consideration

Physical Network - Simple

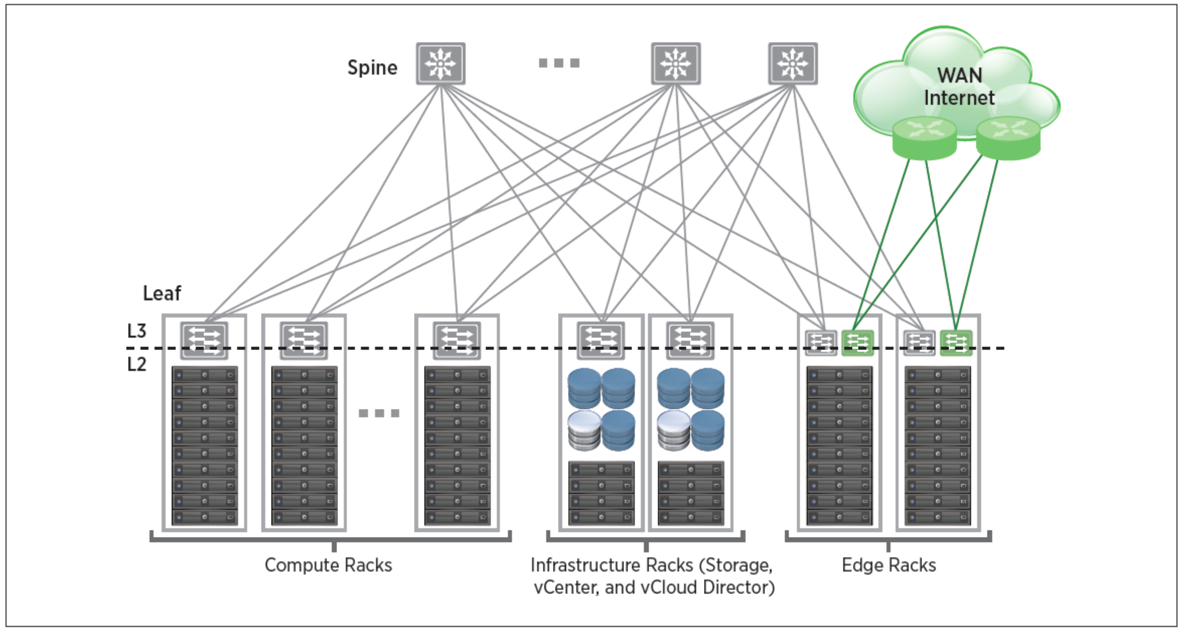


- Teaming
  - Load based
  - LACP
- L3 ToR/leaf
  - Default GW for VLANs
  - Allows dynamic routing



DataCenter Consideration

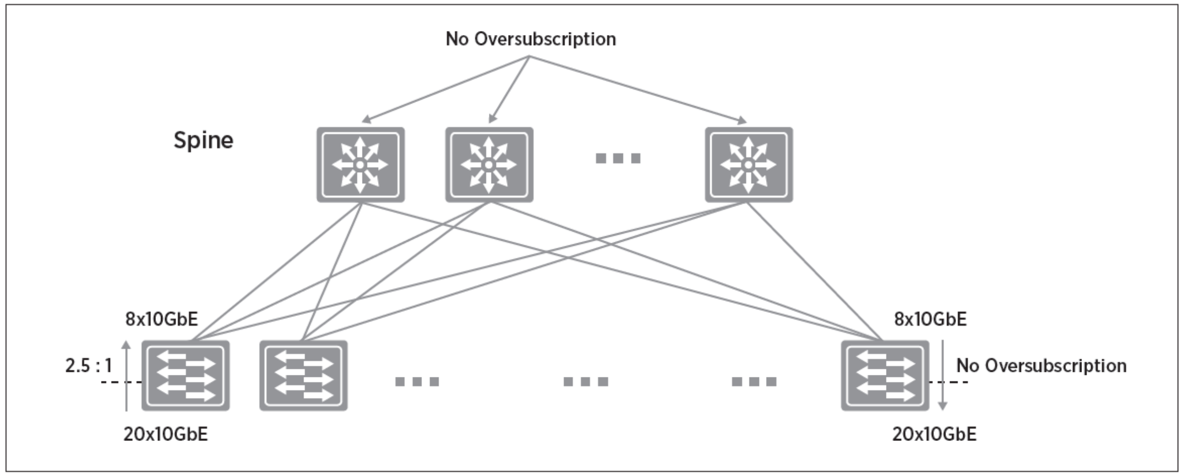
Physical Network - Scalable


Spine
WAN
Internet
Leaf
L3
L2
Compute Racks
Infrastructure Racks (Storage, vCenter, and vCloud Director)
Edge Racks

- Different type of racks
  - Compute: Hypervisors
  - Infrastructure: management
  - Edge: Connectivity
- Equal-Cost Multipathing
  - Fixed number of hops
  - Traffic is TCP/UDP

DataCenter Consideration

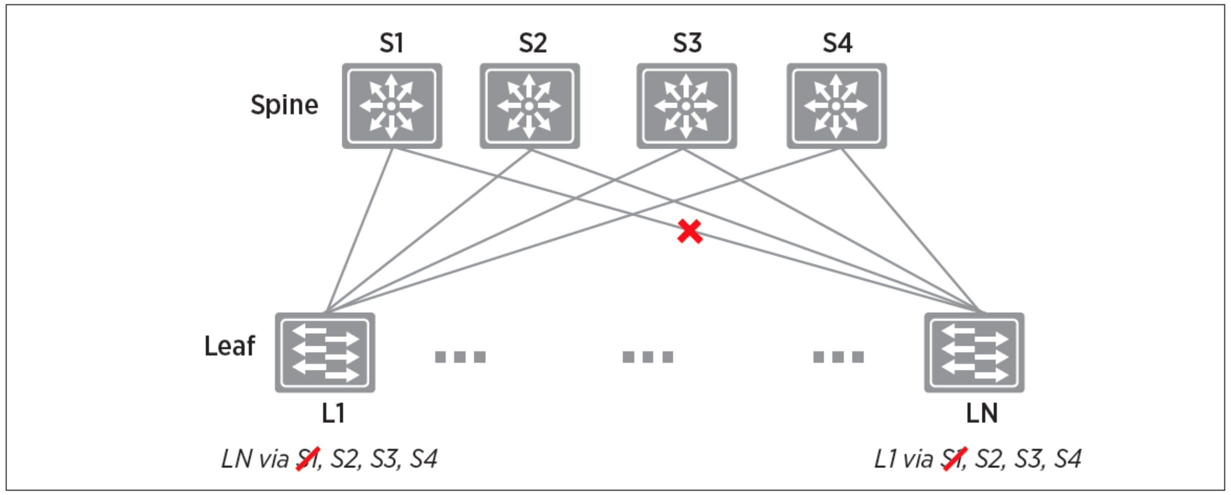
Physical Network - High-bandwidth

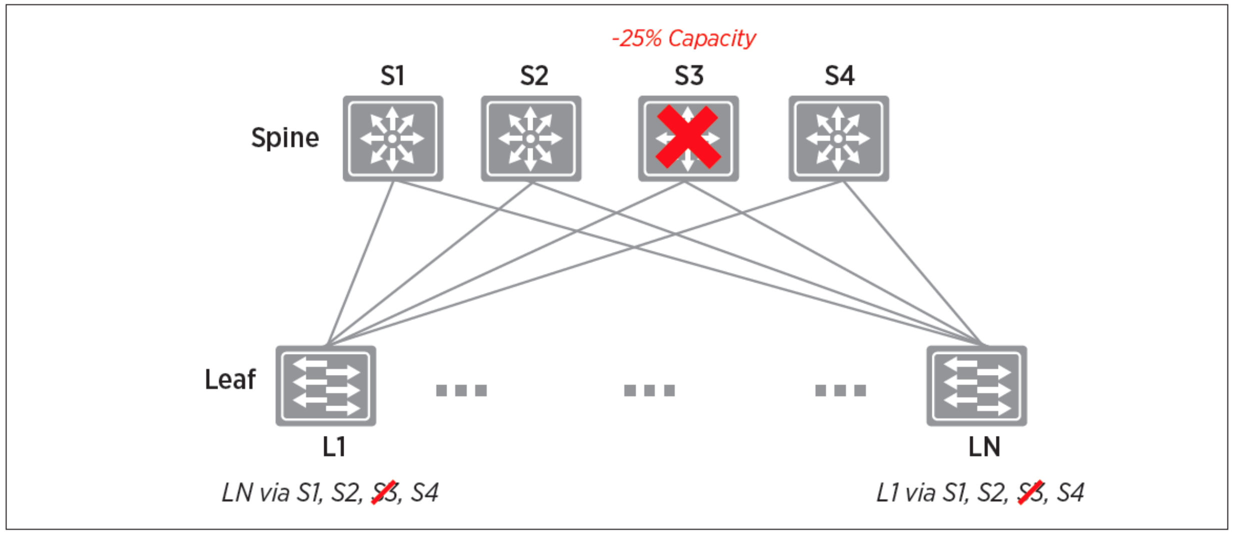


- Calculate oversubscription
  - 20x10Gbps servers – leaf
  - 8x10Gbps leaf – spine
  - = 2.5:1
- Add uplinks where necessary (ECMP)

DataCenter Consideration

Physical Network - Fault-tolerant

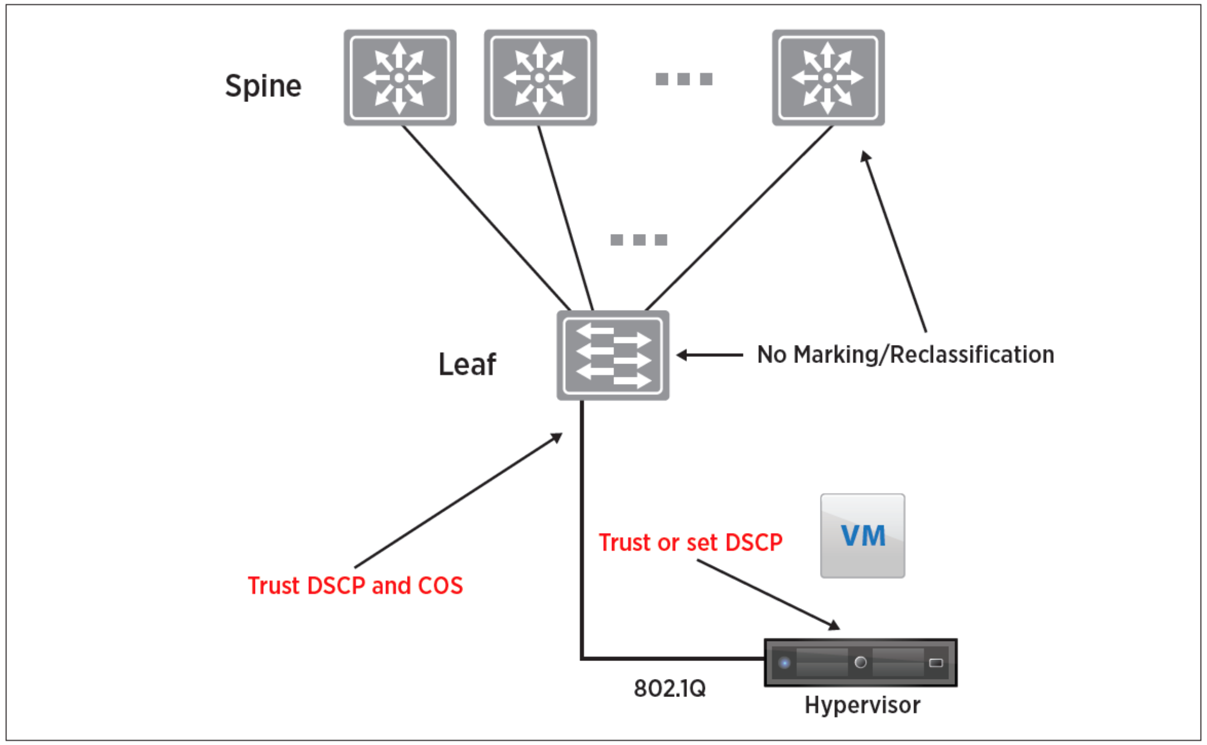




- Protection via routing protocol
  - Link failures
  - Switch downtime (upgrades)

DataCenter Consideration

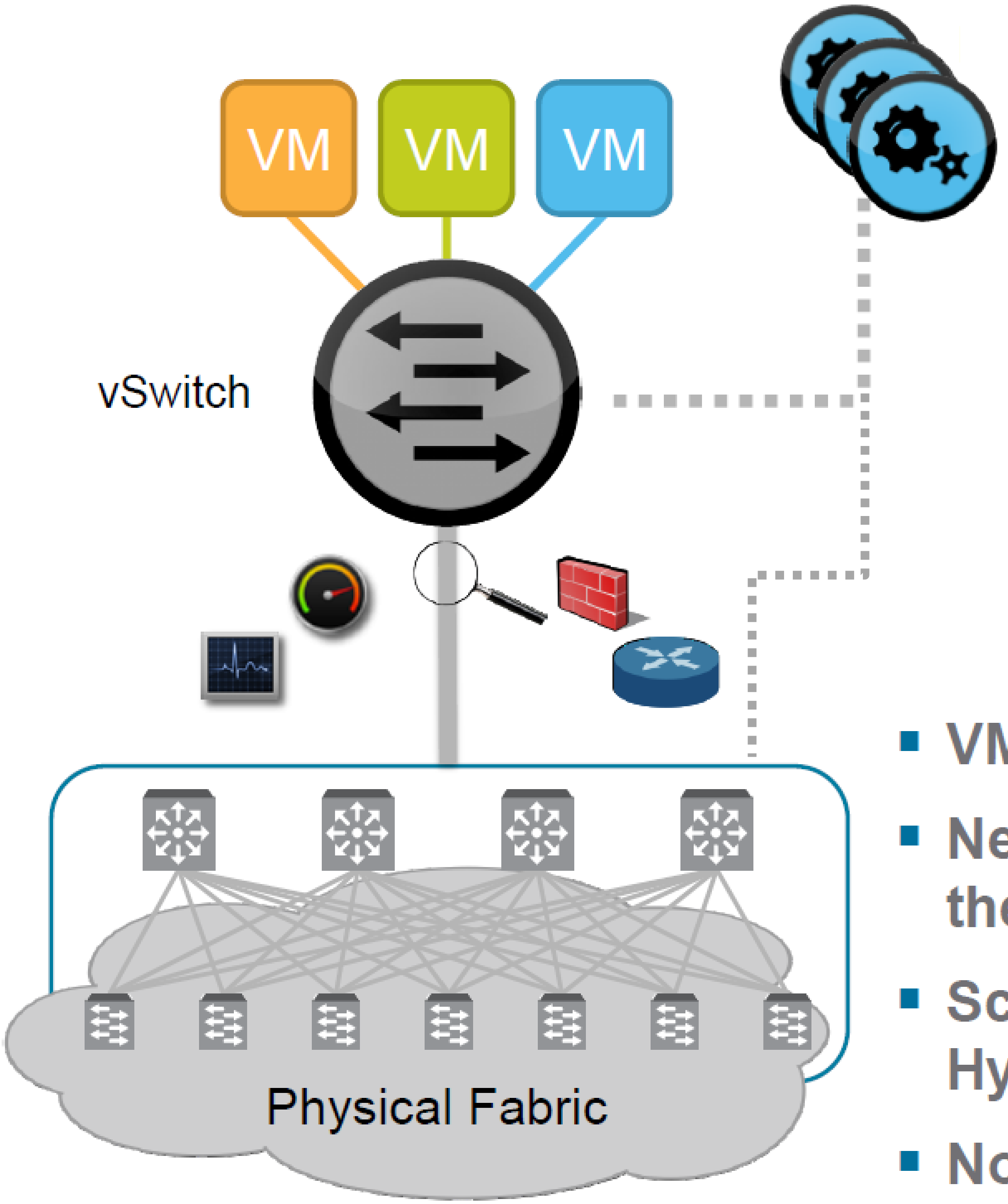
Physical Network - QoS-Providing

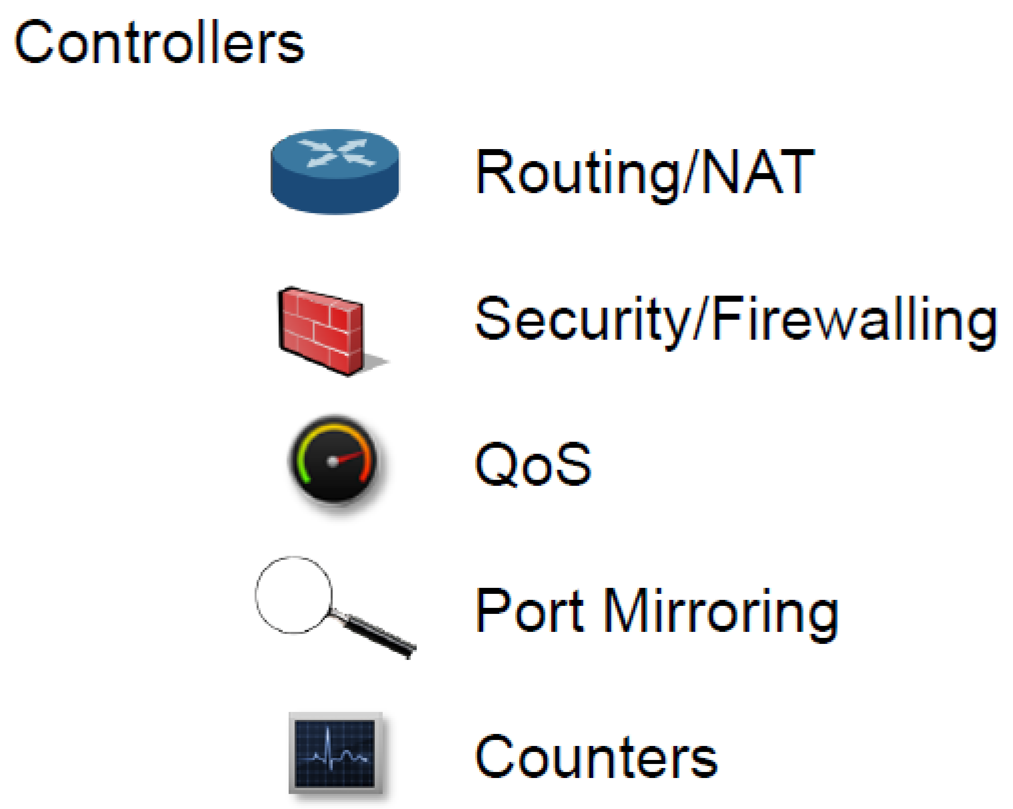


- Hypervisor trusted to set QoS
  - L2: CoS
  - L3: DSCP
- Values kept in VXLAN tunnel

DataCenter Consideration

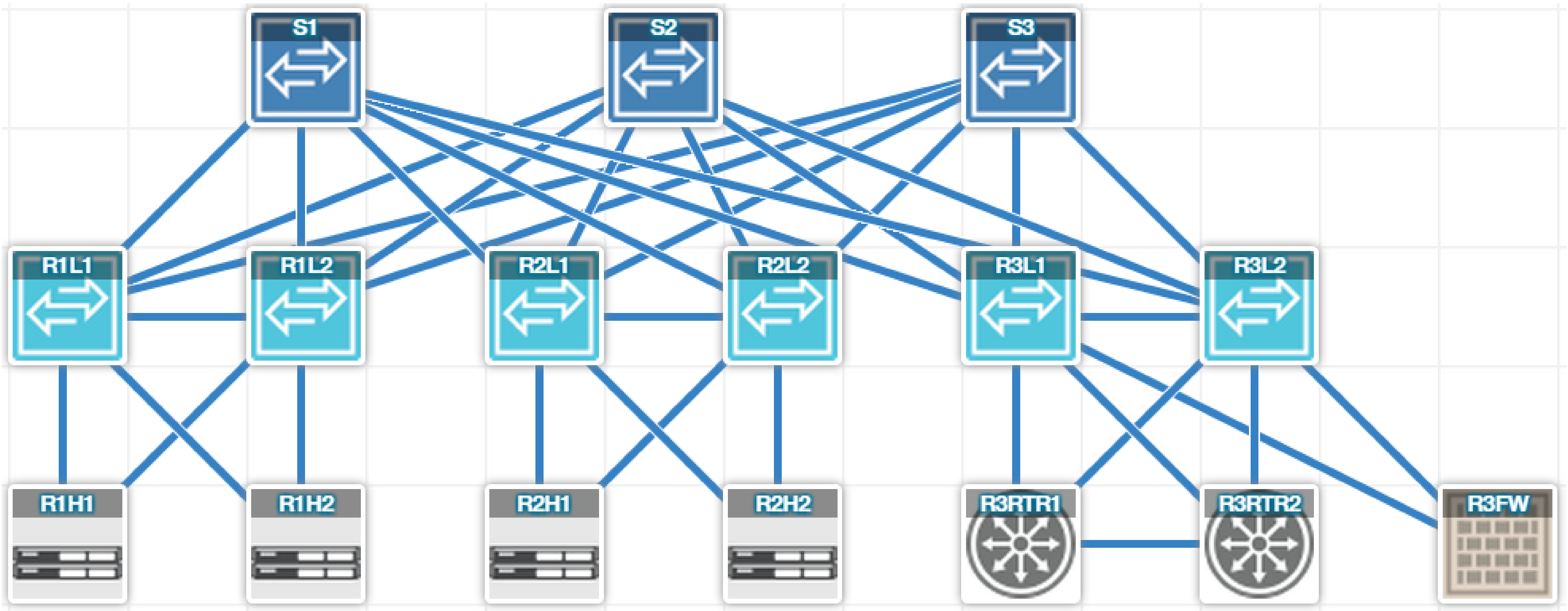
Networking functions in Virtual Space





- VM-Aware Networking services
- Networking services distributed at the edge
- Scale-out with the number of Hypervisors
- No choke points

DataCenter Consideration



ZERO TOUCH NETWORKING (ZTN)

The Big Cloud Fabric system provides Zero Touch Networking (ZTN) for switches.

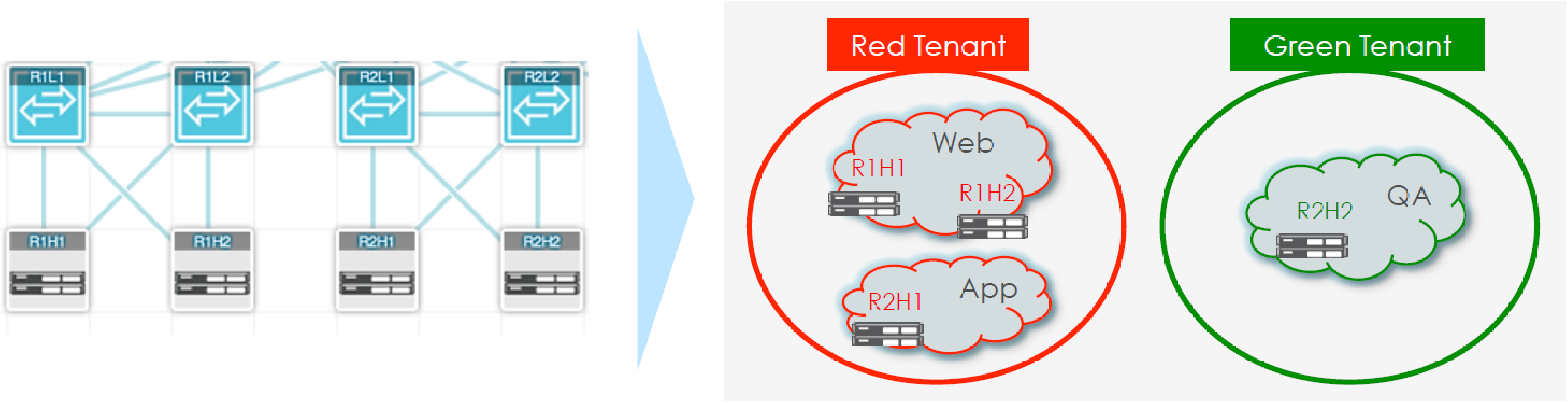
ZTN uses the Open Network Install Environment (ONIE) boot loader to automate switch installation and configuration.

Fabric switches in the Big Cloud Fabric run the Switch Light OS software, which is bundled with the Big Cloud Fabric software.

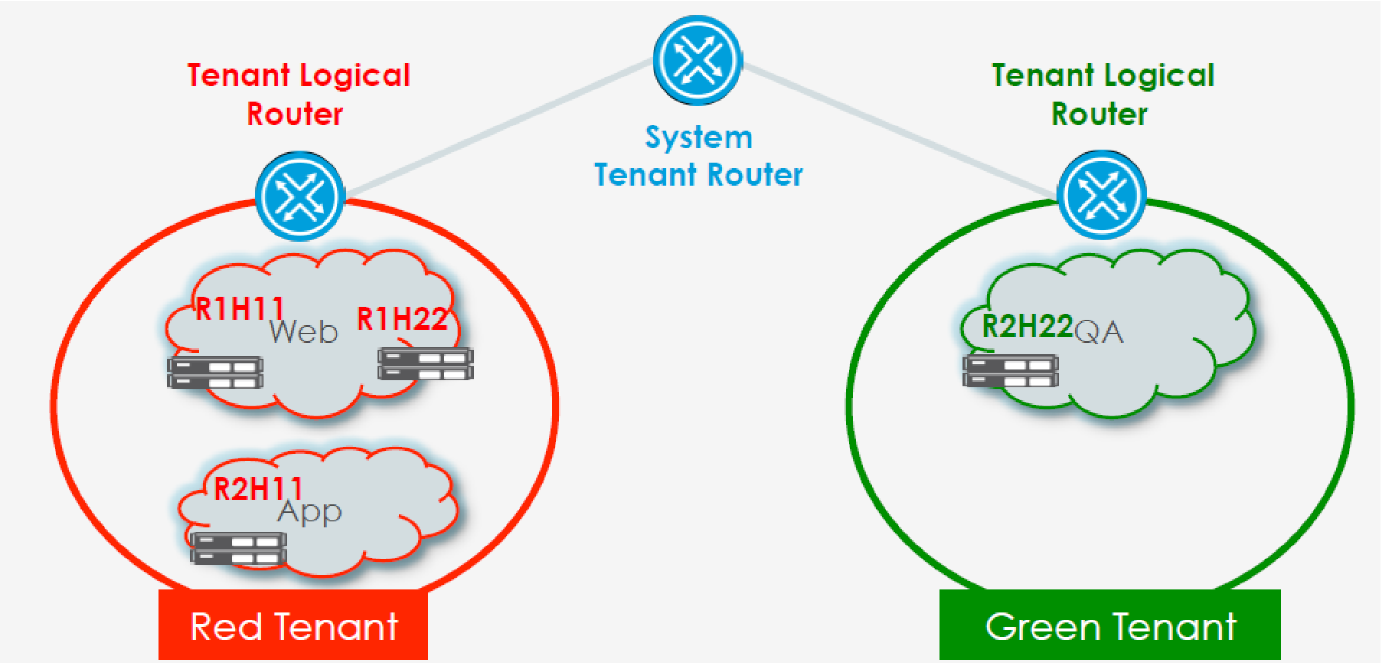
DataCenter Consideration

- Configure two tenants: Green & Red
- Create logical segments for the tenants
- Assign workloads to respective logical segments shown below:

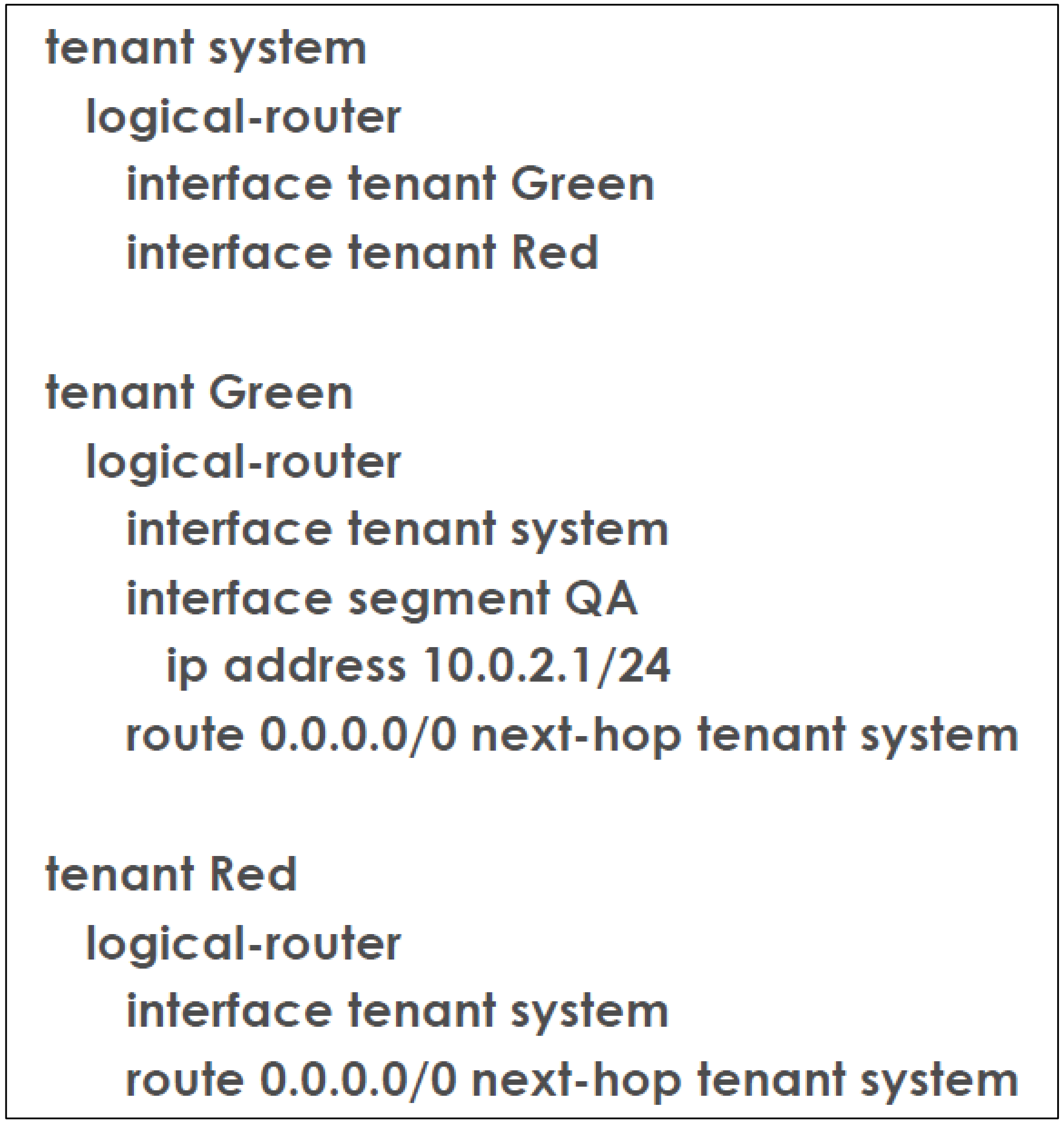
```
tenant Green
  segment QA
    member port-group R2H2 vlan untagged
tenant Red
  segment App
    member port-group R2H1 vlan untagged
  segment Web
    member port-group R1H1 vlan untagged
    member port-group R1H2 vlan untagged
```



DataCenter Consideration



- Create Layer 3 interface for tenant Green, segment QA
- Create System Tenant Router and add default routes on Red and Green tenant routers to point to System Tenant Router
- Add system interface to Green and Red tenant routers
- Add tenant interfaces to System Tenant Router

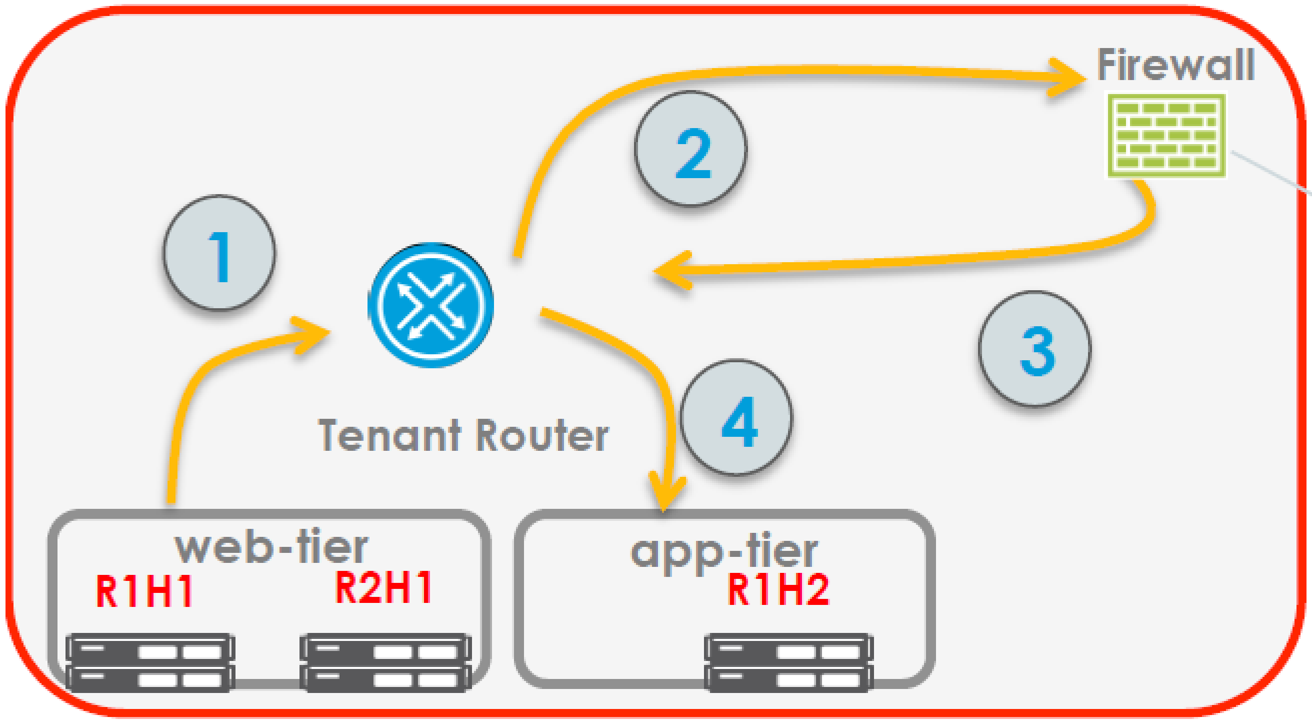


```
tenant system
   logical-router
      interface tenant Green
      interface tenant Red

tenant Green
   logical-router
      interface tenant system
      interface segment QA
         ip address 10.0.2.1/24
      route 0.0.0.0/0 next-hop tenant system

tenant Red
   logical-router
      interface tenant system
      route 0.0.0.0/0 next-hop tenant system
```

DataCenter Consideration



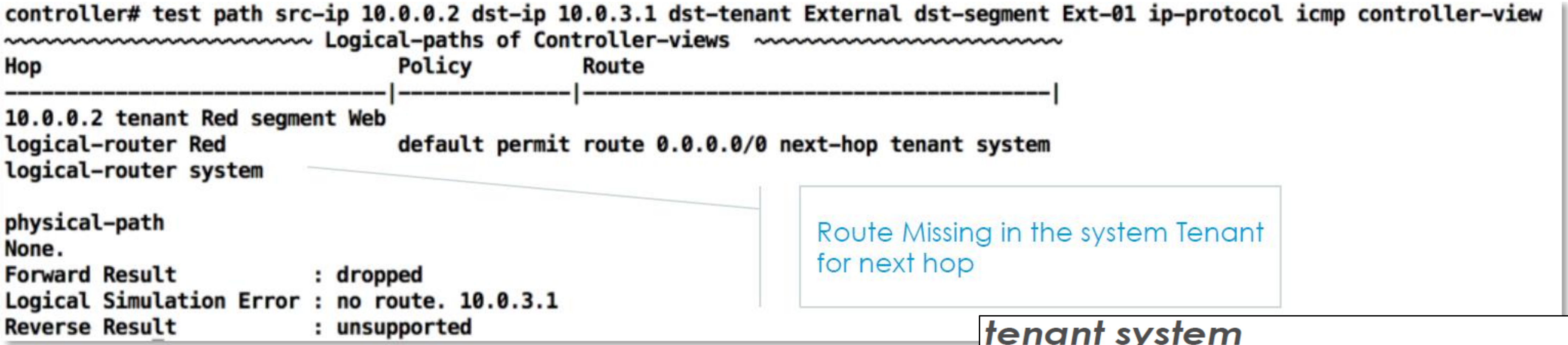
Rule: deny icmp from 10.0.0.0/24 to 10.0.1.0/24

Red Tenant

```
tenant Red
segment FW-01
    member port-group FW-01 vlan untagged
logical-router
    interface segment FW-01
      ip address 10.0.5.1/24
    policy-list FireWall
     10 permit segment-interface Web any to tenant Red segment App next-hop ServiceNode
     11 permit any to any
    apply policy-list FireWall
    next-hop-group ServiceNode
     ip 10.0.5.2
```

Creating another segment in tenant Red
Associate firewall port-groups to the new segment
Add interface to the tenant logical router
Create a policy for Tenant Red to redirect the interesting traffic to firewall node (10.0.5.2)

DataCenter Consideration

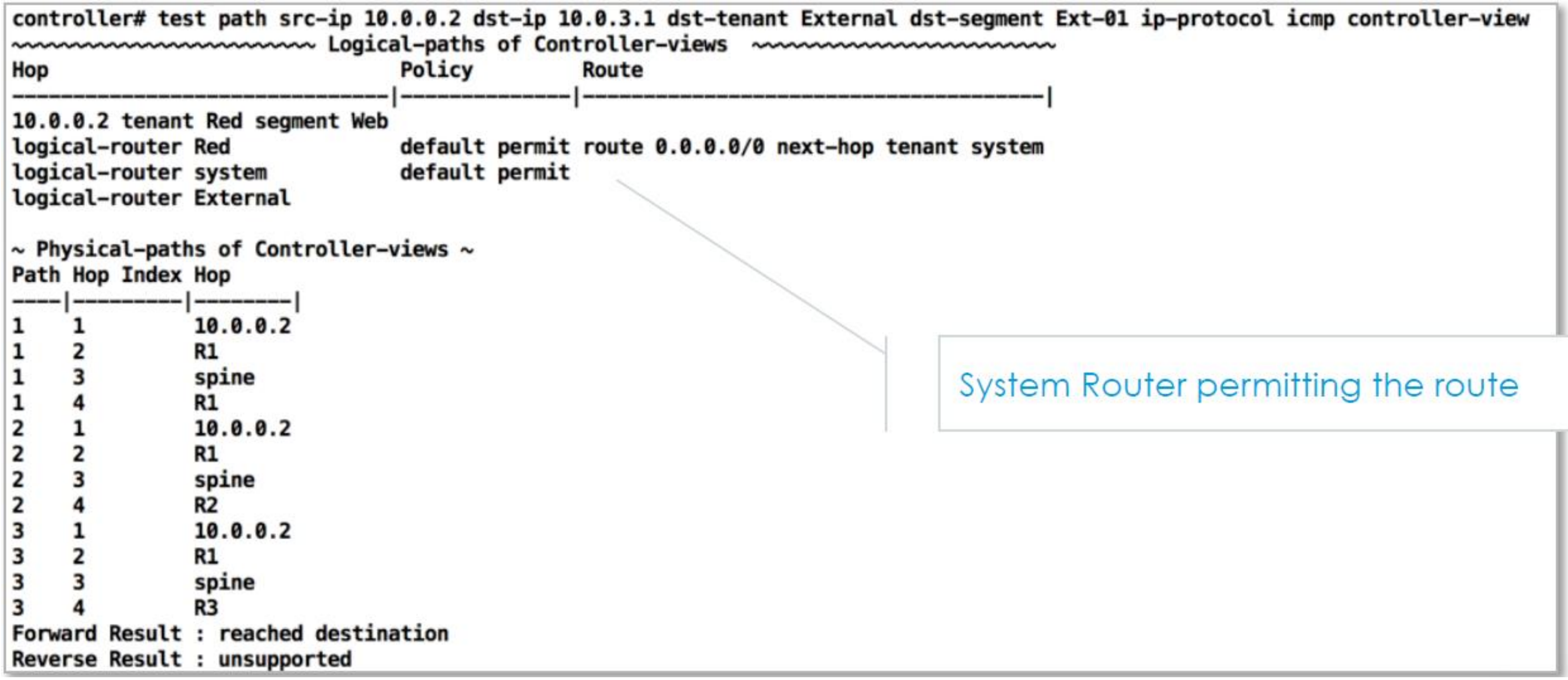


```
controller# test path src-ip 10.0.0.2 dst-ip 10.0.3.1 dst-tenant External dst-segment Ext-01 ip-protocol icmp controller-view
~~~~~~~~~~~~~~~~~~~~~~~~~~ Logical-paths of Controller-views  ~~~~~~~~~~~~~~~~~~~~~~~~~~
Hop                              Policy         Route
--------------------------------|--------------|--------------------------------------|
10.0.0.2 tenant Red segment Web
logical-router Red               default permit route 0.0.0.0/0 next-hop tenant system
logical-router system

physical-path
None.
Forward Result             : dropped
Logical Simulation Error : no route. 10.0.3.1
Reverse Result             : unsupported
```

Route Missing in the system Tenant for next hop

tenant system
logical-router
route 0.0.0.0/0 next-hop tenant External



```
controller# test path src-ip 10.0.0.2 dst-ip 10.0.3.1 dst-tenant External dst-segment Ext-01 ip-protocol icmp controller-view
~~~~~~~~~~~~~~~~~~~~~~~~~~ Logical-paths of Controller-views  ~~~~~~~~~~~~~~~~~~~~~~~~~~
Hop                              Policy         Route
--------------------------------|--------------|--------------------------------------|
10.0.0.2 tenant Red segment Web
logical-router Red               default permit route 0.0.0.0/0 next-hop tenant system
logical-router system            default permit
logical-router External

~ Physical-paths of Controller-views ~
Path Hop Index Hop
----|---------|--------|
1    1          10.0.0.2
1    2          R1
1    3          spine
1    4          R1
2    1          10.0.0.2
2    2          R1
2    3          spine
2    4          R2
3    1          10.0.0.2
3    2          R1
3    3          spine
3    4          R3
Forward Result : reached destination
Reverse Result : unsupported
```

System Router permitting the route

Thank you very much

NAIM