



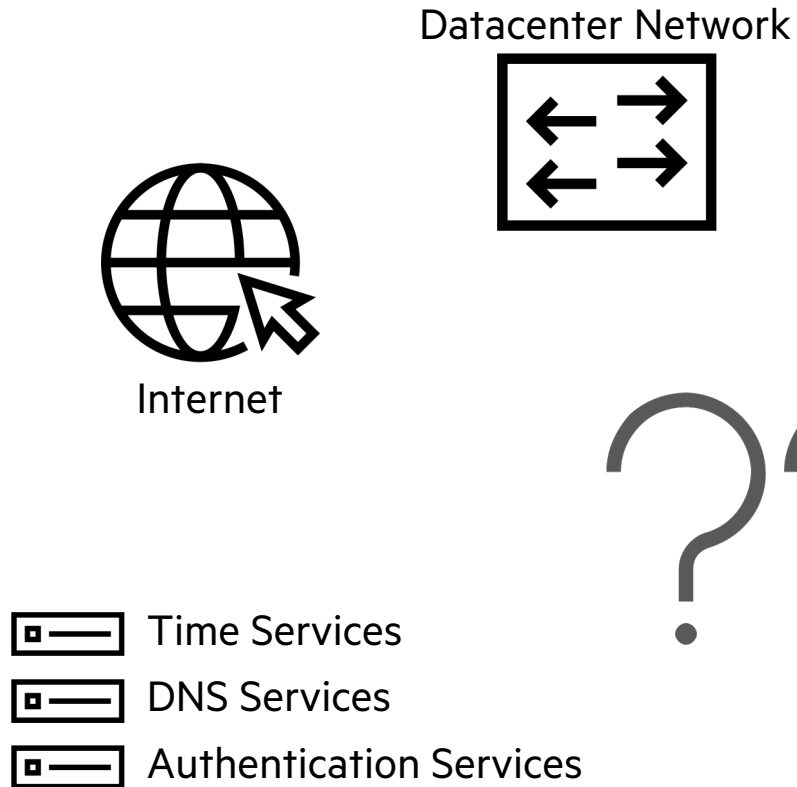
**Hewlett Packard
Enterprise**

USER AND ADMINISTRATIVE ACCESS FOR CSM- BASED SYSTEMS: NETWORK ARCHITECTURE EVOLUTION AND ACCESS CONTROL MECHANICS IN SHASTA V1.4 AND SHASTA V1.5

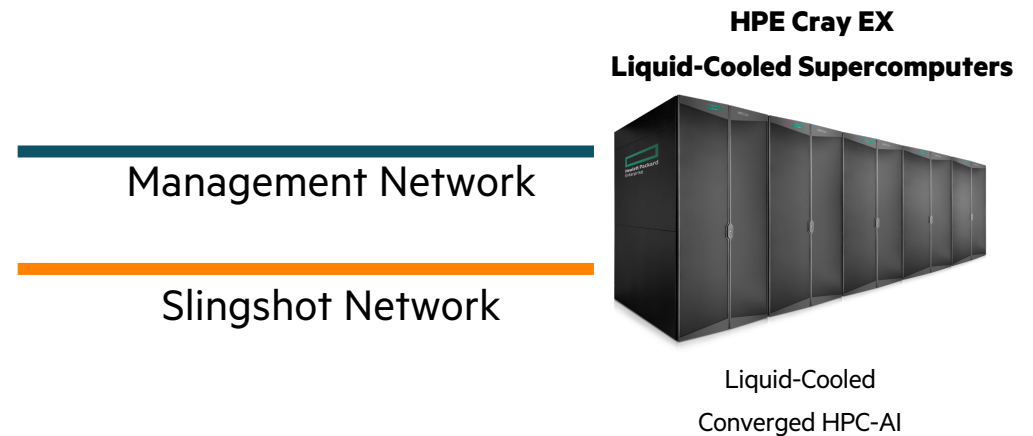
Alex Lovell-Troy and Sean Lynn

May 2021

HPE CRAY EX SUPERCOMPUTER AND THE DATACENTER



- What are the physical connection points?
- What routing does the system need?
- What site services are needed and why?
- How should Users and Admins connect?
- How flexible is it?



THE TWO BASIC NETWORKS

Management Network (Aruba Switches)

- VLANs limit broadcast domains
- Subnets separate BMC access from Node Management
- Additional VLANs and subnets for Customer Access Networking and to expose CSM services to the datacenter
- Leaf-Spine architecture with L3 routing mainly at spines
- ECMP for most traffic
- BGP for internal route resiliency (Not exposed to site)
- Can be used for default route

Slingshot Network (Rosetta Switches)

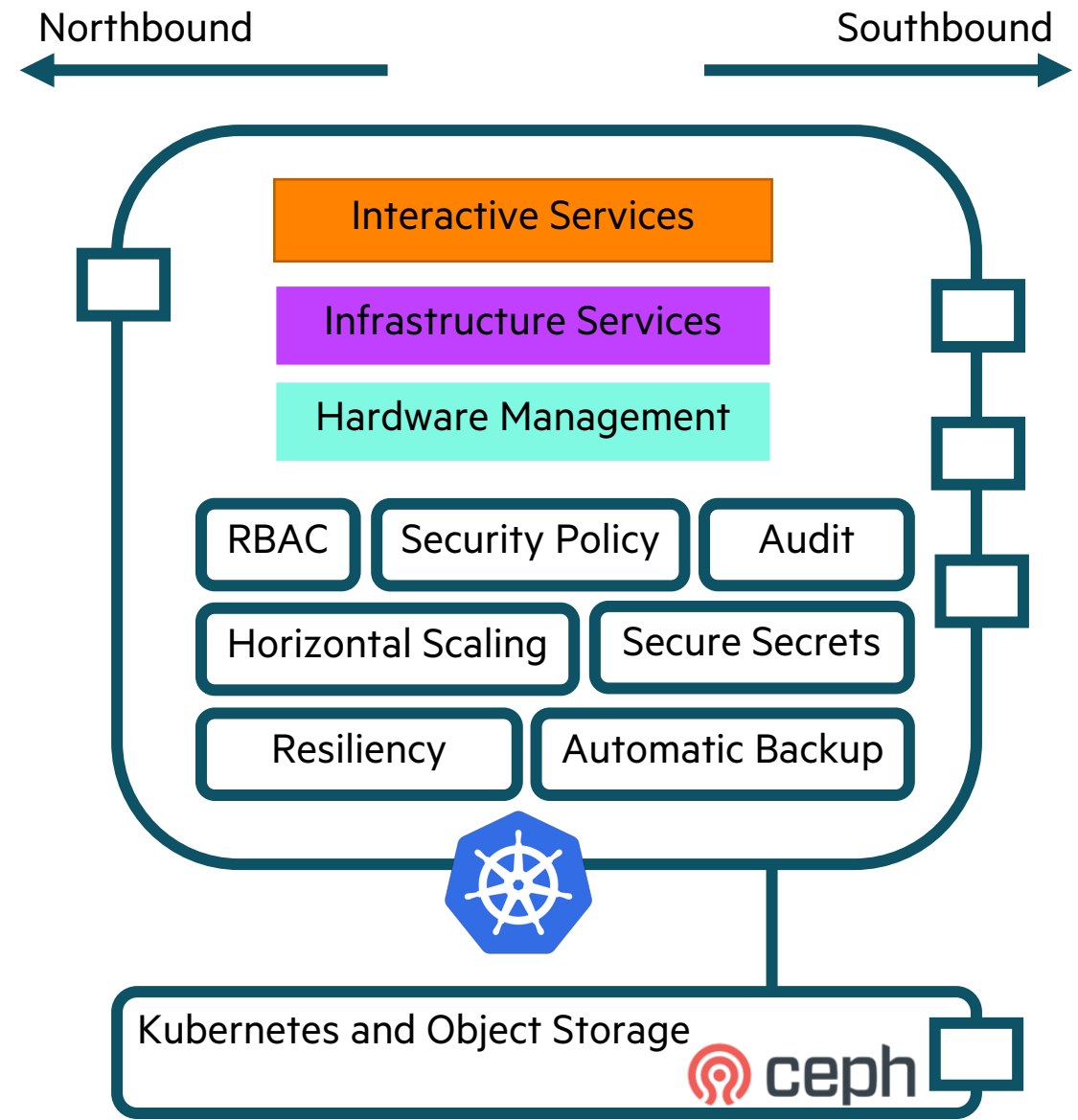
- Connects all HPC nodes
- Effectively flat ethernet network
- L3 routing provided by edge router(s)
- Main HPC network for TCP/IP and HPC Protocols
- Access to shared storage
- Can be used for default route



CSM ARCHITECTURE REVIEW

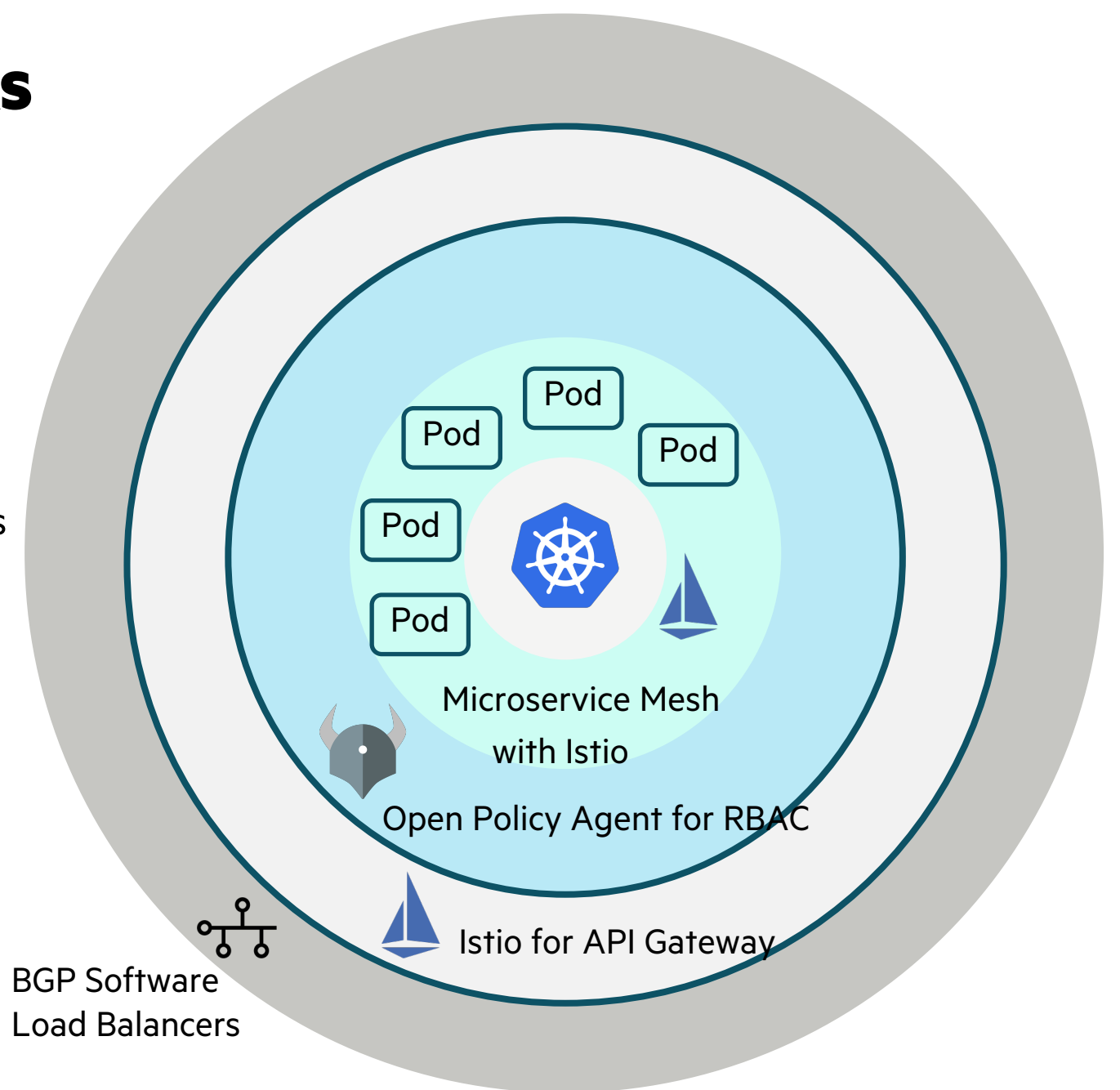
- Kubernetes as Platform-Building-Platform
- Kubernetes, Istio, and Operators for infrastructure
- Layered microservices for managing HPC clusters
- HPC-enablement only in the upper layers
- Northbound APIs for Users and Admins
- Southbound APIs for interacting with Compute hardware

All User/Admin interactions protected by TLS 1.3 and OIDC authentication



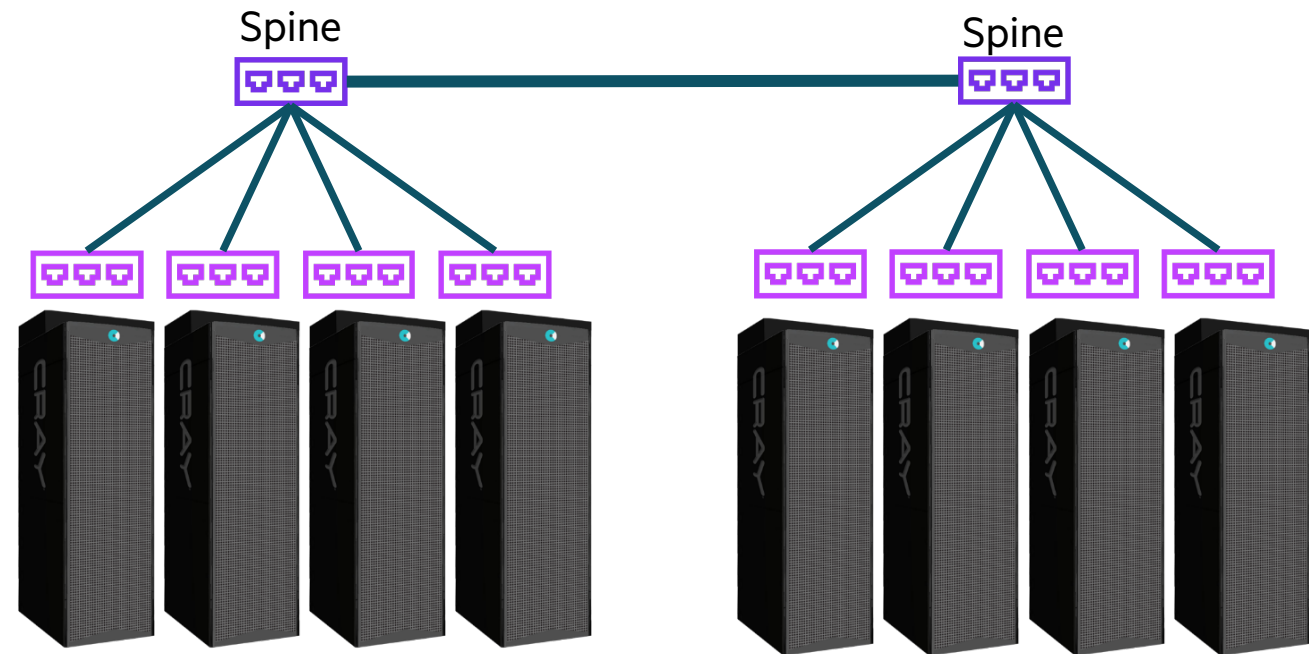
MICROSERVICE SECURITY LAYERS

- Pod to Pod Traffic is secured by Istio with mTLS and Kubernetes Policy
- Ingress and Egress traffic is regulated by OPA
- Istio provides gateway services to expose collections of services
- MetalLB allocates Virtual IP addresses that pass traffic to Istio Gateways
- Keycloak handles authentication and issues refreshable bearer tokens, required for API Access
- Keycloak federates with upstream LDAP or Kerberos for user directories



MANAGEMENT NETWORKING

- Spine switches are mainly for Layer 3 Routing between Subnets
- Leaf switches connect directly to nodes and node controllers
- Olympus/Mountain Cabinets have embedded network switches that connect to leaf switches

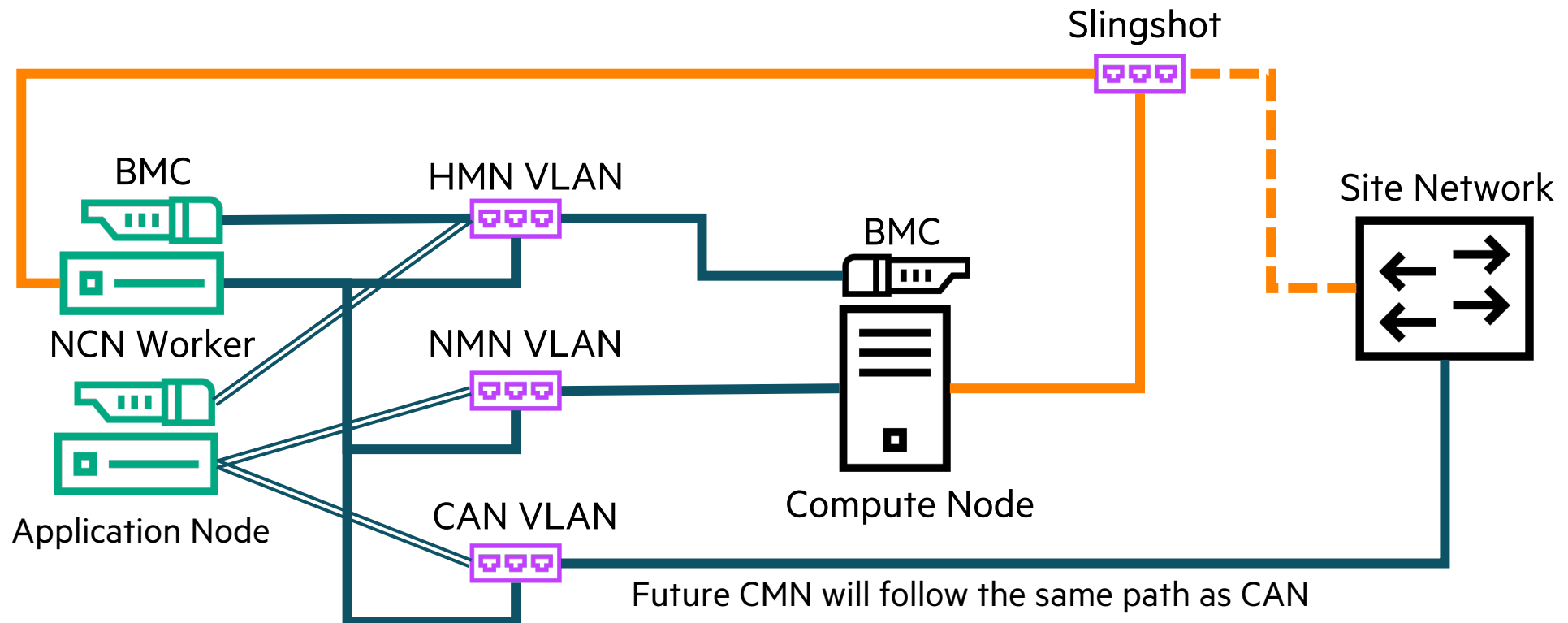


Per Cabinet Subnets

- /22 of private ipv4 space for BMCs
- /22 of private ipv4 space for Node Management
- One VLAN per cabinet subnet

DETAILED VLAN DIAGRAM

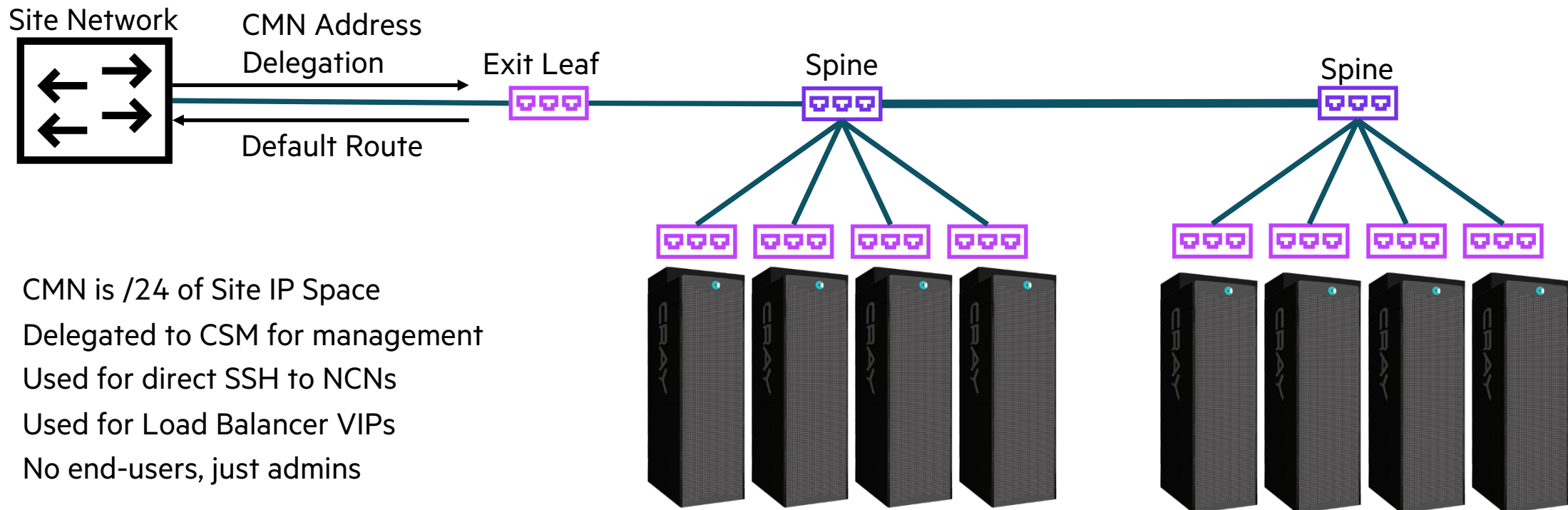
- NCN Workers connect to all VLANs
- Application Nodes connect to all VLANs except HMN
- Compute Nodes only connect to the NMN VLAN
- BMCs only connect to HMN VLAN
- No routing between VLANs



SITE NETWORKING OVER THE MANAGEMENT NETWORK

Customer Access Network (**CAN**)

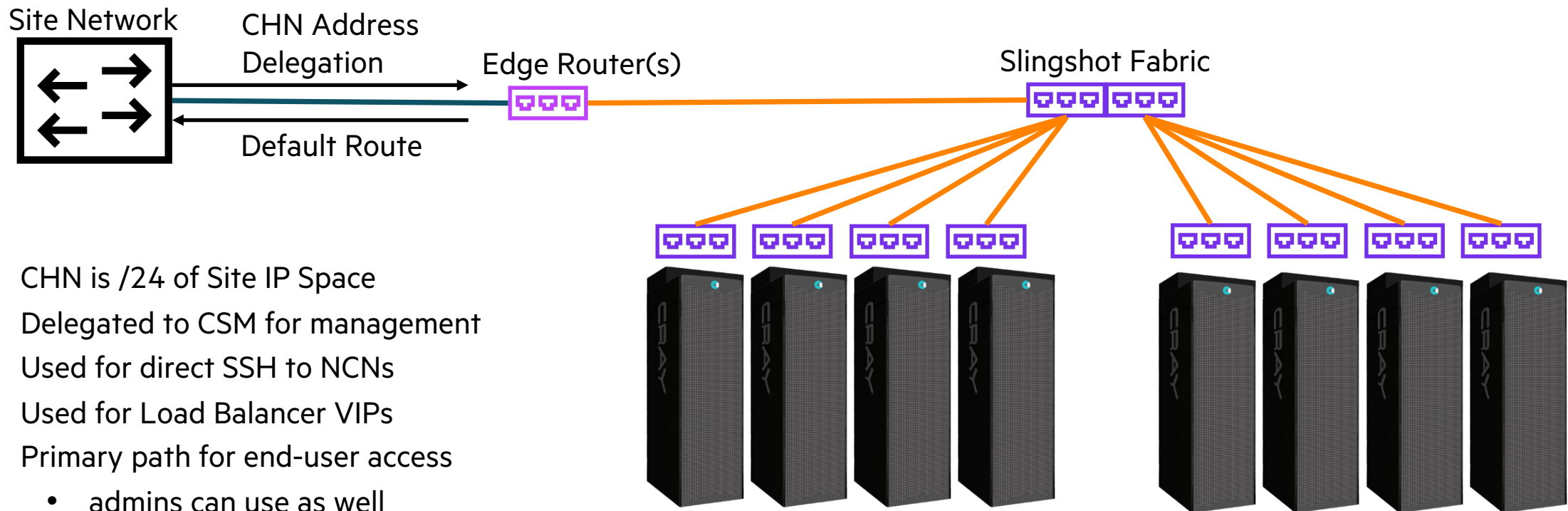
- Overloaded term to describe L3 IP Space, L2 Vlans, L1 Connectivity, and intentional use
- Replaced with more specific language in Shasta 1.5



- CMN is /24 of Site IP Space
- Delegated to CSM for management
- Used for direct SSH to NCNs
- Used for Load Balancer VIPs
- No end-users, just admins

SITE NETWORKING OVER SLINGSHOT

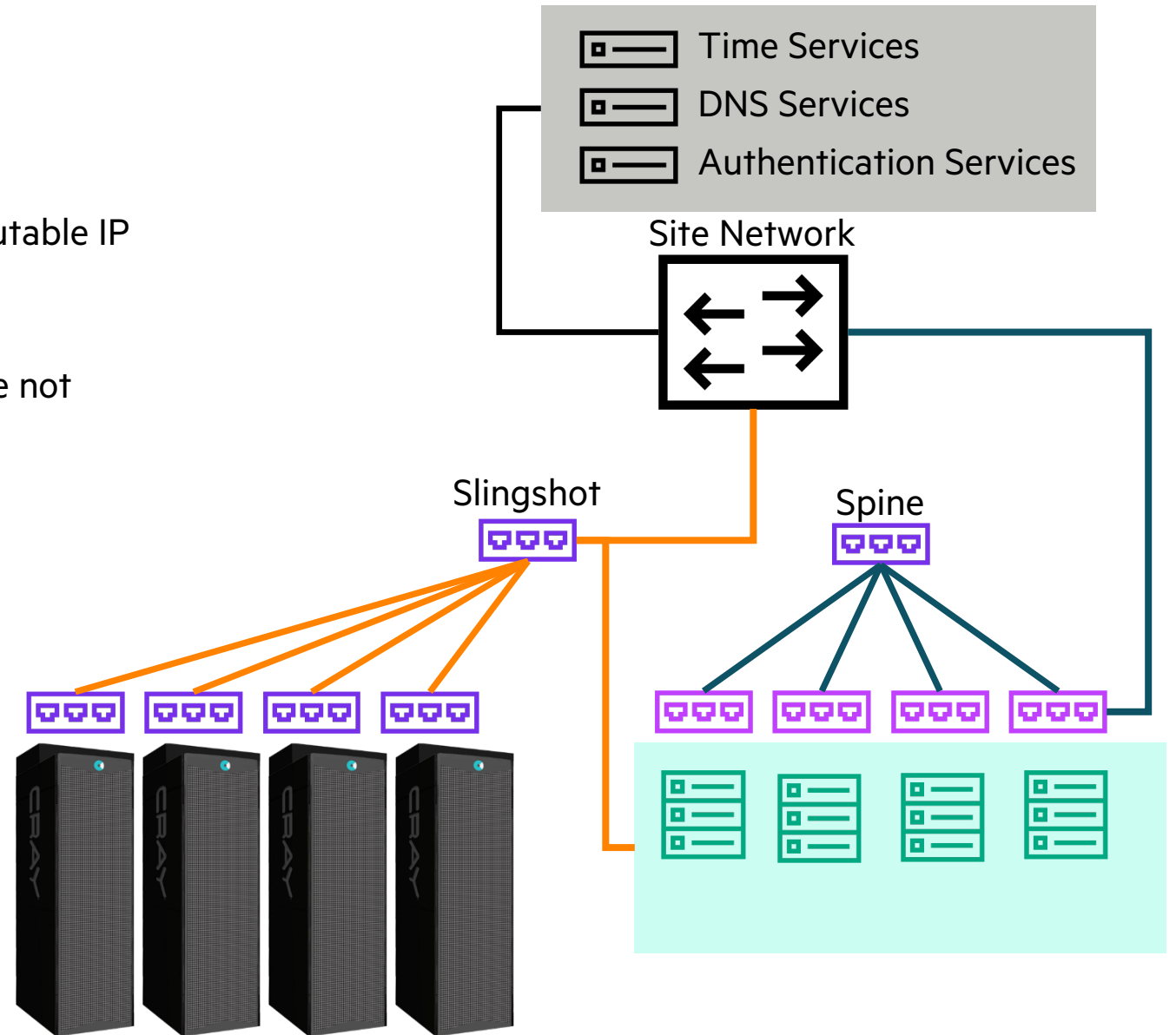
The slingshot fabric acts like a very large ethernet switch with all routing happening at the edge router(s). Architecturally, the edge router acts like an exit-leaf in this configuration as well as providing L3 routing similar to a spine switch



- CHN is /24 of Site IP Space
- Delegated to CSM for management
- Used for direct SSH to NCNs
- Used for Load Balancer VIPs
- Primary path for end-user access
 - admins can use as well

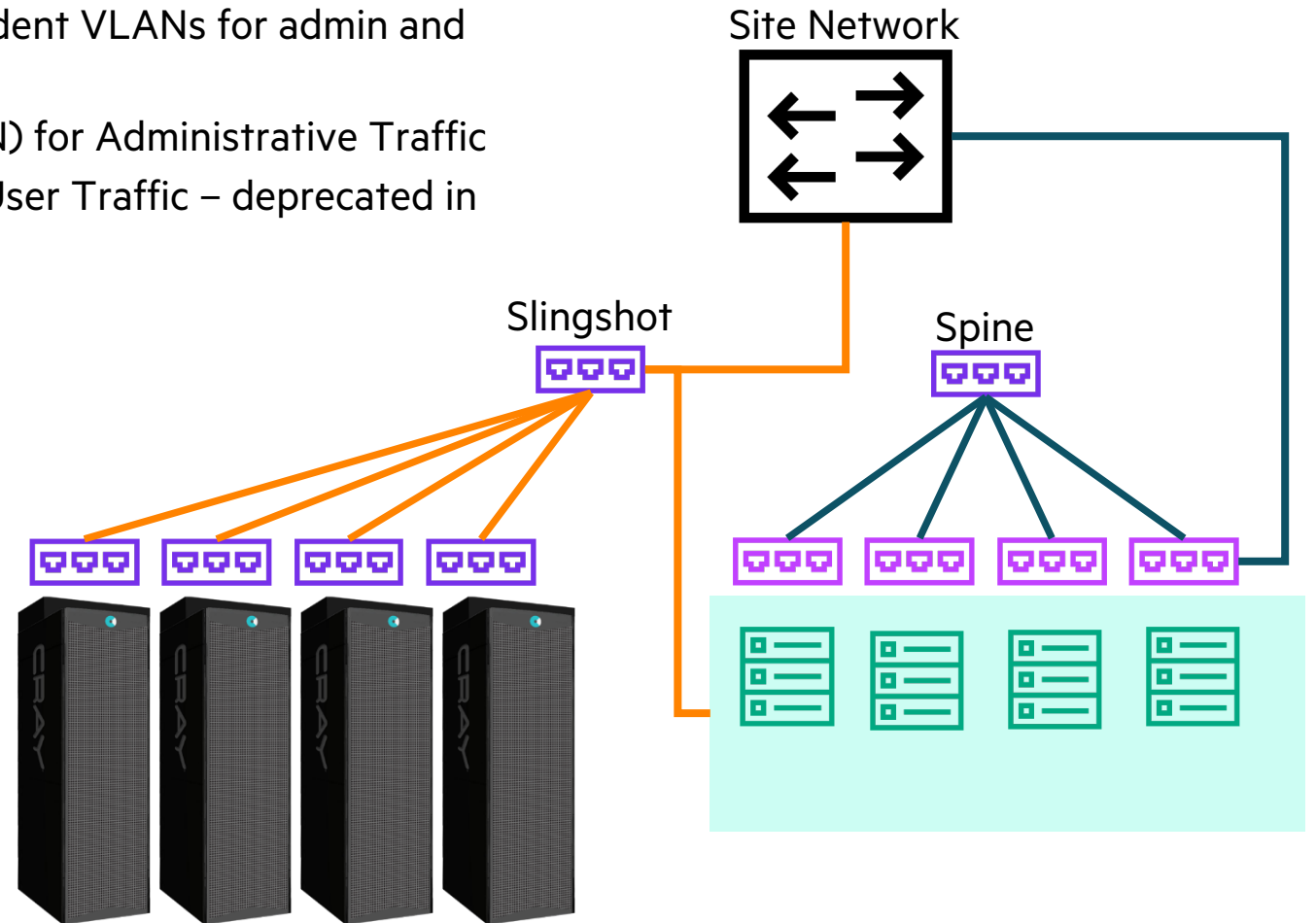
ACCESSING SITE SERVICES

- CSM does not provide NAT services
- All compute nodes and NCNs must have routable IP addresses to directly access site services
- DNS and NTP services are cached in CSM
- LDAP and other Authentication Services are not cached



NETWORK ACCESS CONTROL

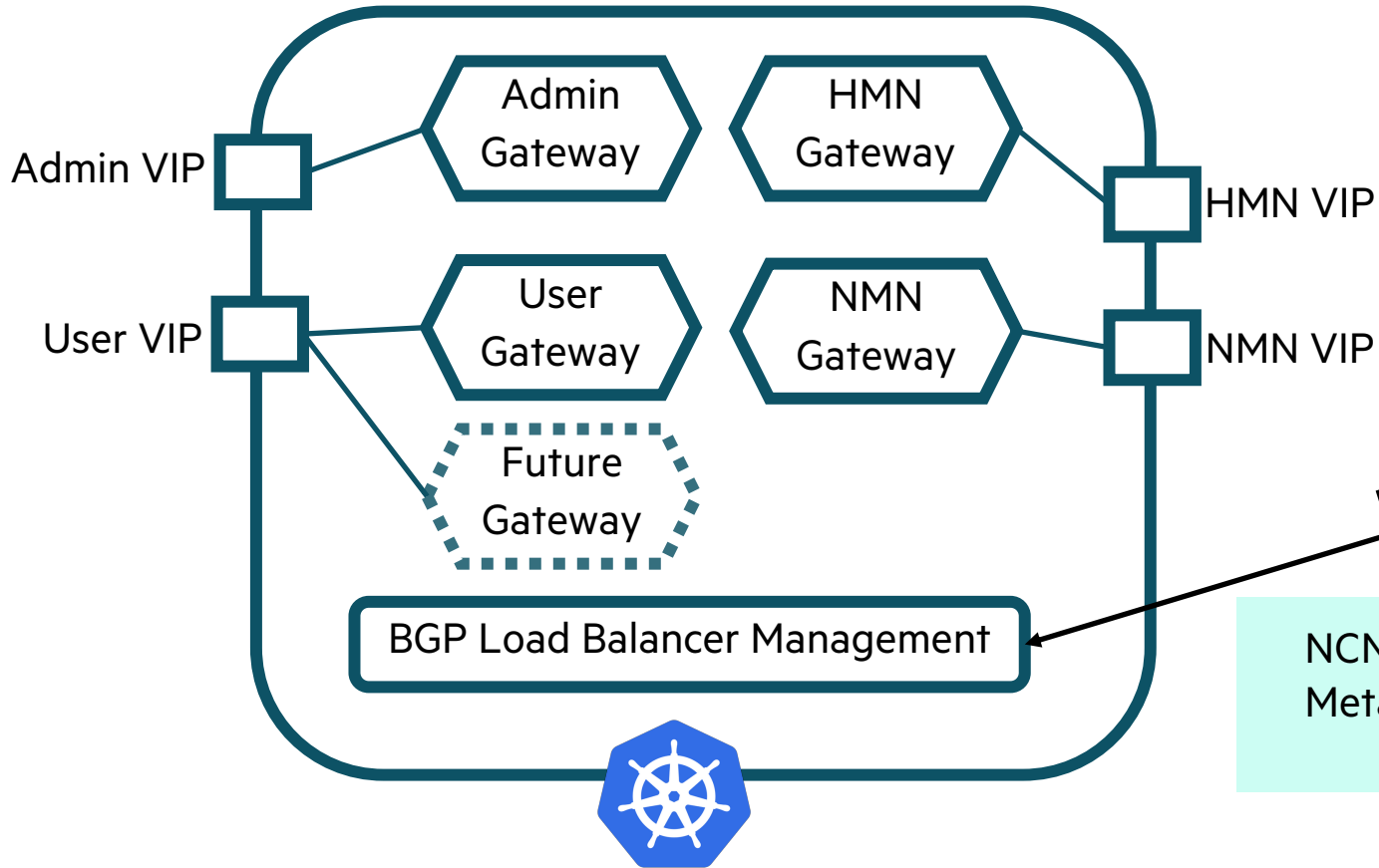
- “Bifurcated CAN”
- Management network path carries independent VLANs for admin and user traffic
 - Customer Management Network (CMN) for Administrative Traffic
 - Customer Access Network (CAN) for User Traffic – deprecated in favor of Slingshot in 1.5
- Slingshot network path carries user traffic
 - Customer High Speed Network (CHN)
- SSH to Management NCNs limited to CMN
- Administrative APIs limited to CMN





CSM DYNAMIC SERVICE ADVERTISEMENT

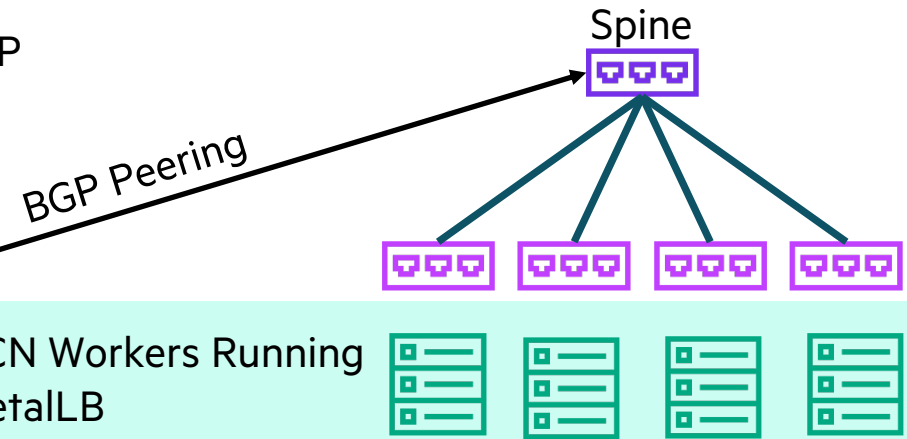
**Northbound APIs
for admins and users**

**Southbound APIs
for Nodes and Jobs**



 Virtual IP Addresses (VIPs) are independent of the individual NCN

 Each API gateway can be connected to one or more VIPs



*BGP Peering may descend to management leaf as scale suggests



THANK YOU

alt@hpe.com

sean.lynn@hpe.com

