Designing a Scalable and Secure Lab Network: Engineering Labs with EVPN VXLAN

Yousuff Shaik
Network Engineer

Murat Mugan
Network Engineer Manager

Joe Waters
Network Engineer

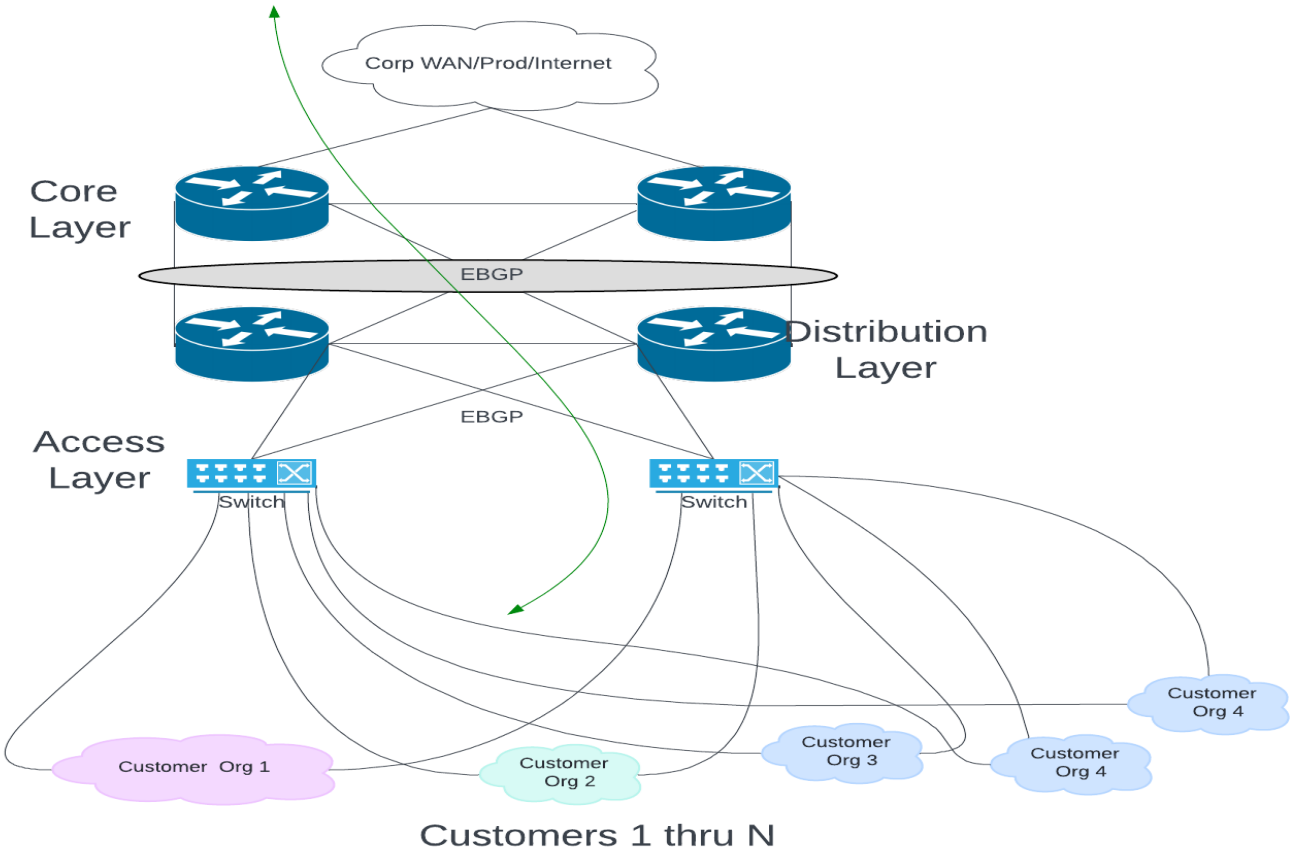
Network Infrastructure Services(NIS)

Meta

Meta’s Engineering Labs

NANOG™

Legacy Network



NANO

Legacy Network Challenges- 1

Reliability/Capacity

- L2 storm take down the lab network and VRRP split brain issues
- TCAM issues plague ACL programing
- Bandwidth goes wasted as one of the links stays in blocking state
- Lack of traffic manipulation capabilities between Distribution <> Access Layers
- Cost overhead from Distribution Layer redundancy requirements in large deployment
- Scalability challenges for WAN Northbound traffic

Legacy Network Challenges- 2

Tooling

- Provisioning and Operating overhead
- Lack of Derived Vs Desired states/DB for auditing/alerting/monitoring
- Non-standardized Port/Vlan/ACL/Exception policies

Security

- Limited to VLAN based traffic segmentation ***only***
- Large ACL configurations

Legacy Network Challenges- 3

Random MAC's/IP Space

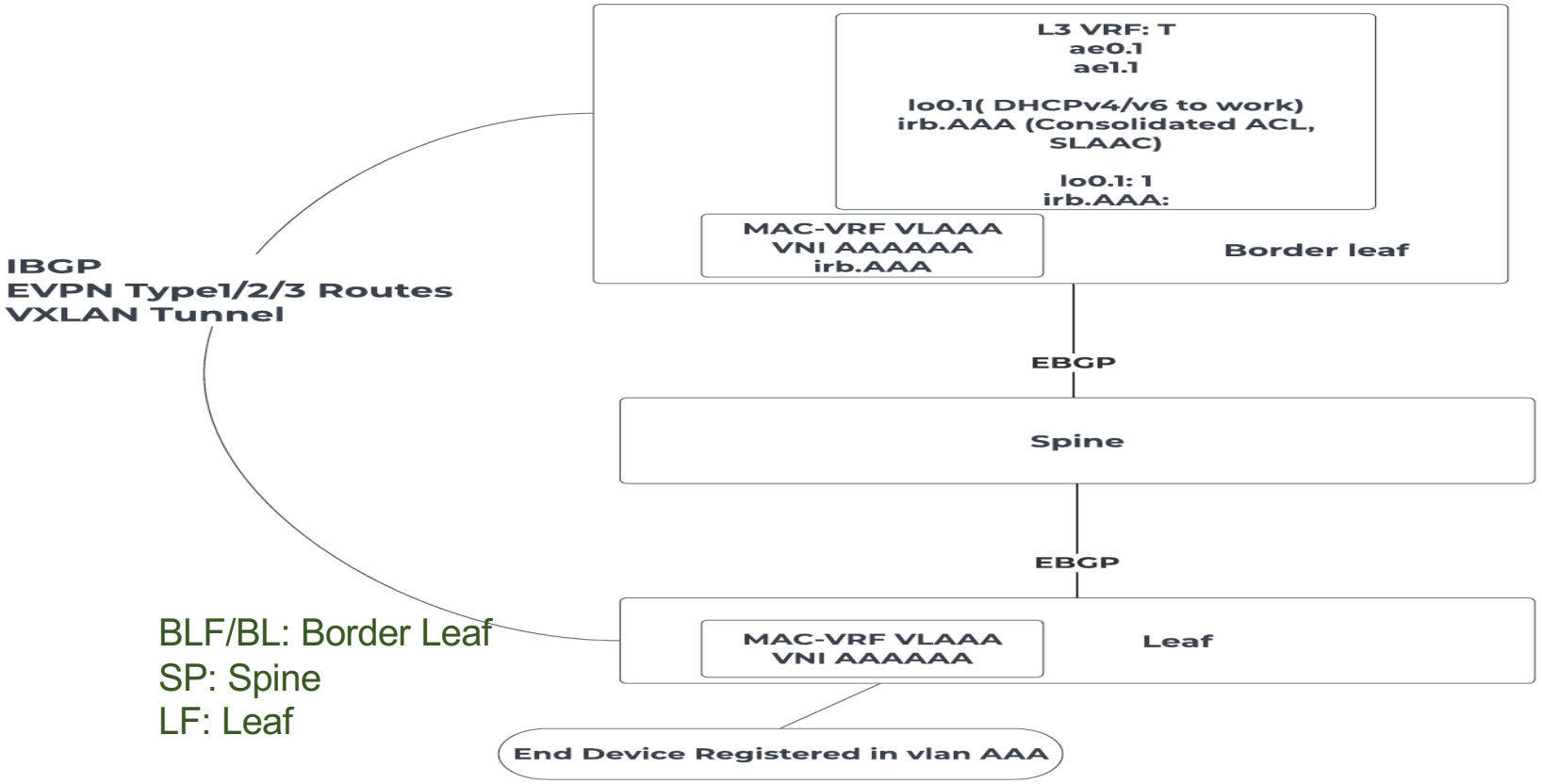
- Random MAC's eating up IP space
- Results in v4 address space exhaustion
- Security concerns
- NON-contiguous IP space allocation

SLAAC

- SLAAC Multicast Floods
- Devices sticking to more than One SLAAC IP (service vs quarantine pods)

Path to EVPN VXLAN

Architecture Change



Architecture Change

Design

- 3 Stage CLOS- EVPN VXLAN / Vendor Agnostic
- Utilize Anycast Gateway and eliminate the use of any other First Hop Routing Protocol (FHRP)

Security

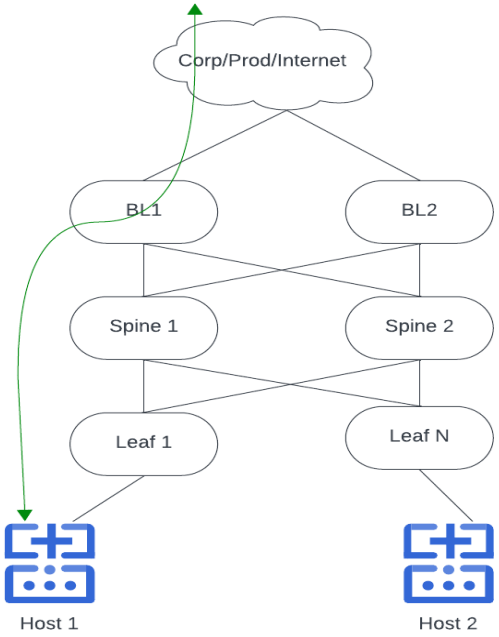
- BLF Consolidated ACL Layer
- Monitor TCAM utilization to enable robust traffic manipulation and layer 3 ECMP load balancing
- Better segmentation of traffic flow using VRFs, MAB, 802.1x

Architecture Change

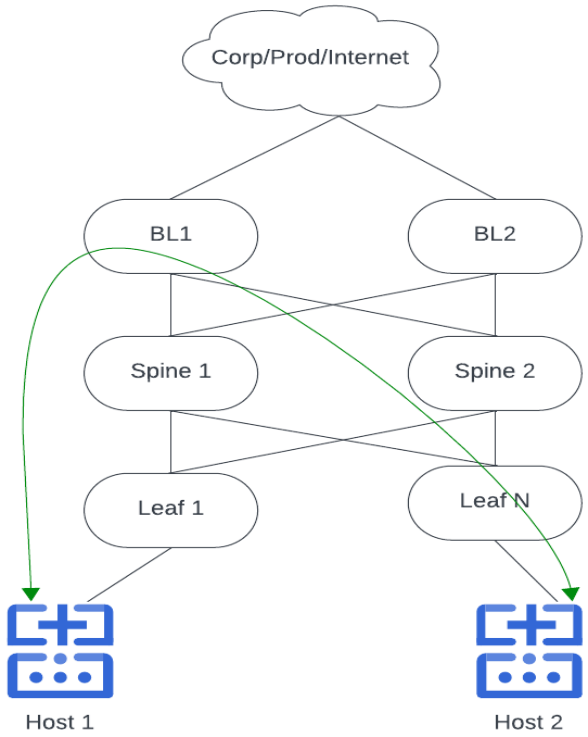
Reliability

- Better DHCPv4/v6 and SLAAC performance
- Multifold capacity, availability, resiliency
- L3 handoff possibility
- Emerging capabilities such as PTP

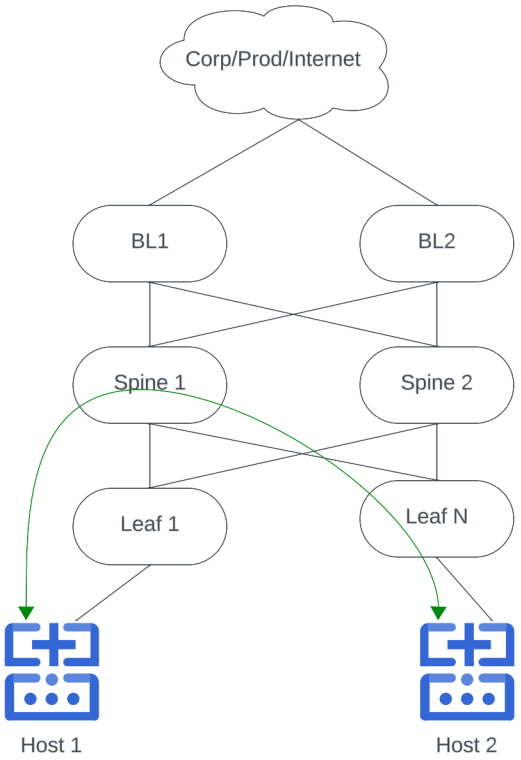
Traffic Flow



Scenario 3: External traffic



Scenario 2: Inter/VNI traffic



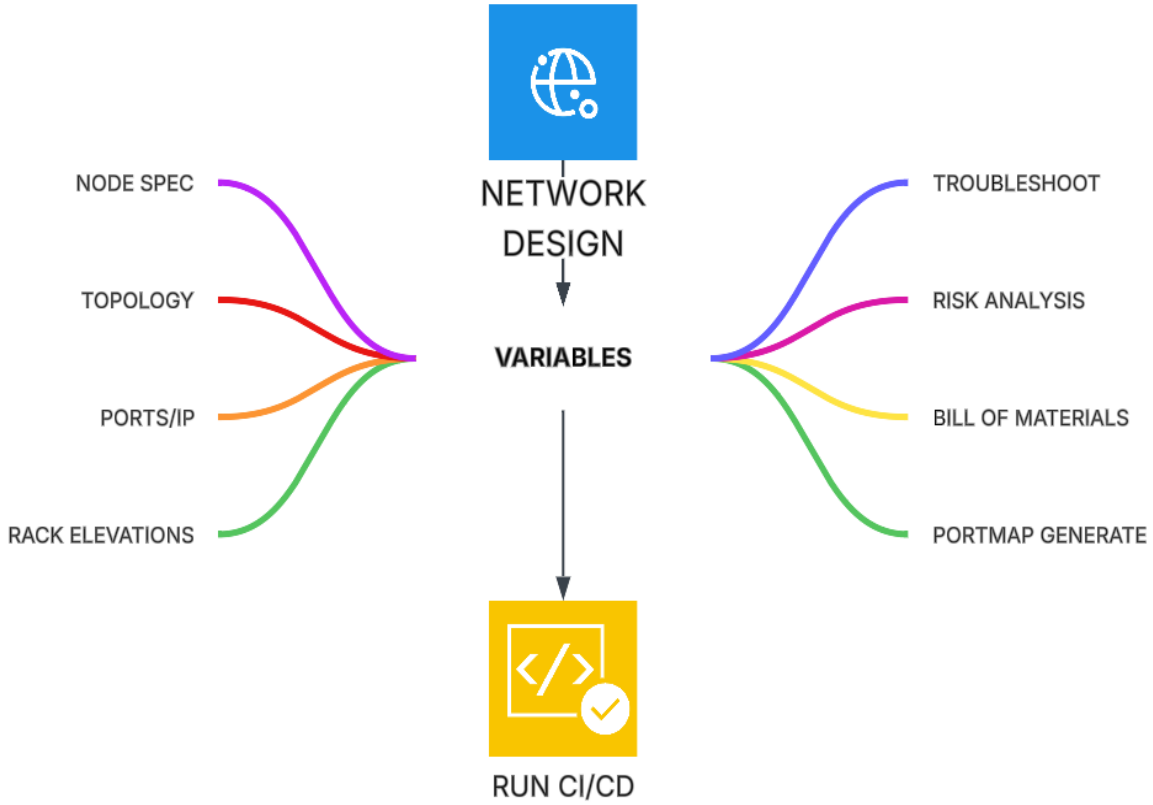
Scenario 1: Intra-vni

BLF/BL: Border Leaf
SP: Spine
LF: Leaf

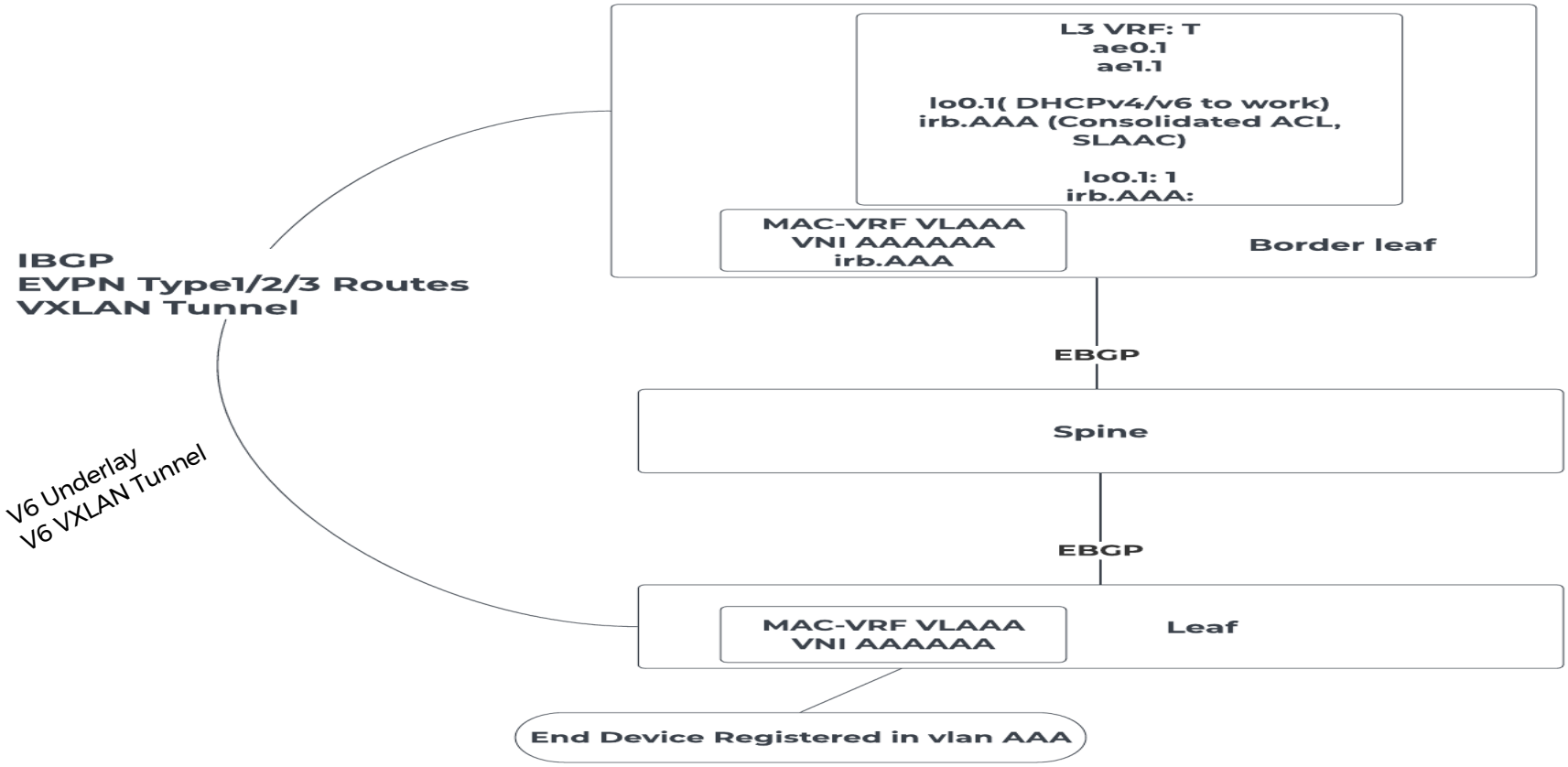
NANOG™

Tooling Spec On Mind

- Network Design
- Intent-based network modeling
- Intent generation
- Mapping and device targeting (VLANs, Platforms)
- Templating for CLOS Layers
- Bootstrap and configure (ZTP), Operate, Run



Config Snippet



Config Snippet

Border Leafs [MAC VRF/IRB/L3-VRF] <> EVPN BGP config excluded [Can be any vendor] with Supported ACL Scale**

set routing-instances MAC-VRF-VLAAA instance-type mac-vrf
set routing-instances MAC-VRF-VLAAA protocols evpn encapsulation vxlan
set routing-instances MAC-VRF-VLAAA protocols evpn default-gateway no-gateway-community
set routing-instances MAC-VRF-VLAAA vtep-source-interface lo0.0
set routing-instances MAC-VRF-VLAAA bridge-domains VLAAA-BD vlan-id AAA
set routing-instances MAC-VRF-VLAAA bridge-domains VLAAA-BD routing-interface irb.AAA
set routing-instances MAC-VRF-VLAAA **bridge-domains VLAAA-BD vxlan vni 10AAA**
set routing-instances MAC-VRF-VLAAA service-type vlan-based
set routing-instances MAC-VRF-VLAAA route-distinguisher 192.168.0.5:10AAA
set routing-instances MAC-VRF-VLAAA vrf-target target:AAA:AAA

Config Snippet

```
set routing-instances LAB-L3-VRF instance-type vrf
set routing-instances LAB-L3-VRF interface irb.AAA
set routing-instances LAB-L3-VRF route-distinguisher 192.168.0.5:888

set interfaces irb unit 100 proxy-macip-advertisement
set interfaces irb unit 100 family inet filter output-chain BOILERPLATE
set interfaces irb unit 100 family inet filter input-chain BOILERPLATE
set interfaces irb unit 100 family inet filter output-chain IRB100-FW
set interfaces irb unit 100 family inet address 10.0.0.254/24 virtual-gateway-
address 10.0.0.1
set interfaces irb unit 100 virtual-gateway-v4-mac 00:00:00:00:10:00
```

Config Snippet

Spine <>Underlay<> <EBGP config excluded>[Can be any vendor]**

Leaf <> EVPN BGP config excluded [Can be any vendor]**

```
router bgp ABCDE
   router-id 1.1.1.101
      vlan AAA
      rd 1.1.1.101:10AAA
      route-target both 10AAA:10AAA
      redistribute learned
      redistribute dot1x
   !

interface Vxlan1
   vxlan source-interface Loopback0
   vxlan udp-port 4789
   vxlan vlan AAA vni 10AAA
!
```

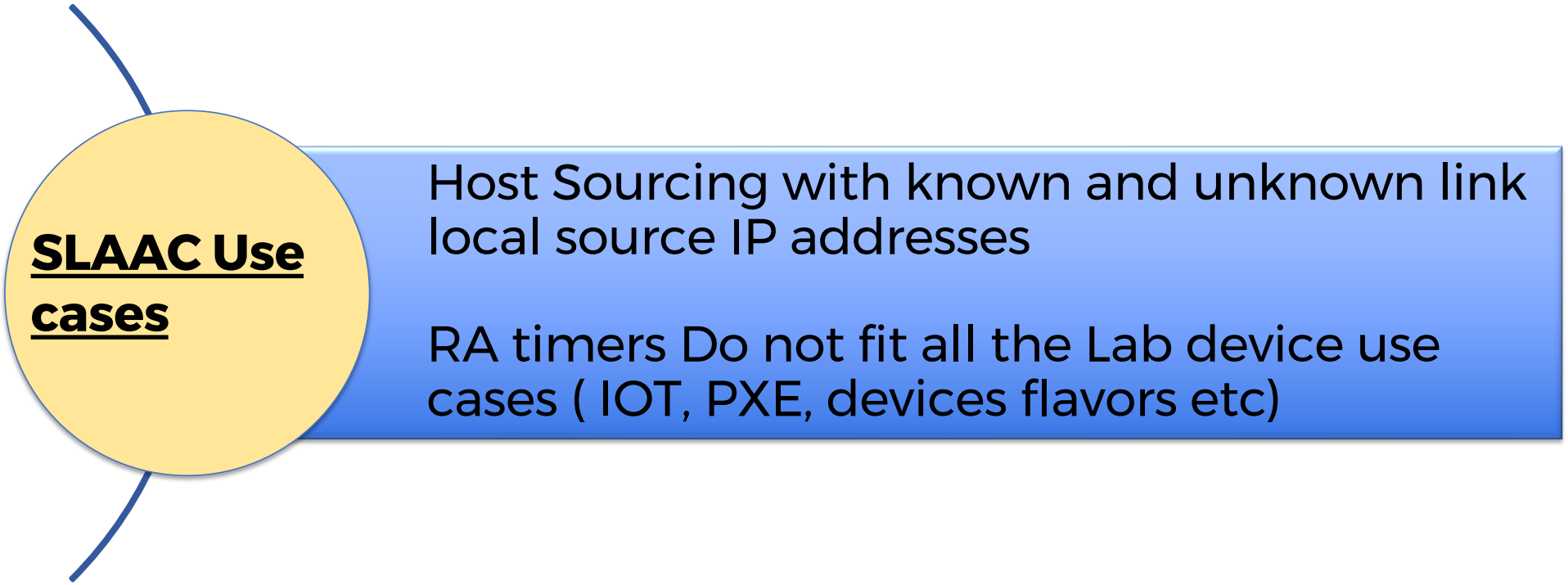
Time to Migrate!

Migration Challenges - 1

limited to local LAN for contiguous v4 allocations"- Making it difficult to manage an IP object and act on it

IPv4 data is stored in multiple places with inconsistencies and non-persistent data causing challenges in tracking actual IP usage.

Migration Challenges - 1

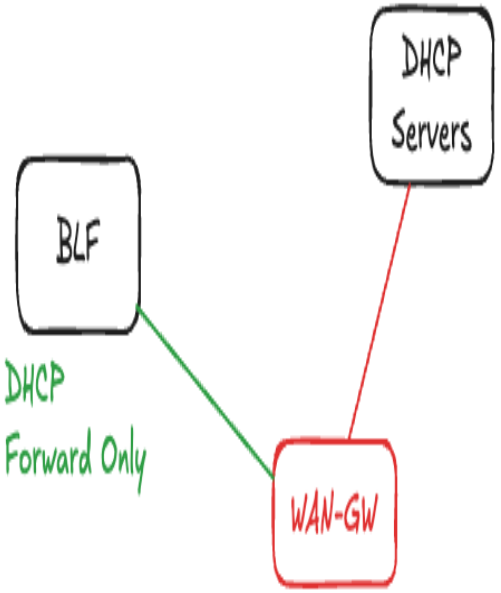


Migration Challenges – 2

DHCP v4/V6 reliability- DUAL RELAY SCENARIO

Lab infra uplinks to WAN-GW, WAN-GW snoop DHCP packets only on trusted interfaces

DHCP traffic arriving on any other non trusted interface is **punted to RE and dropped**

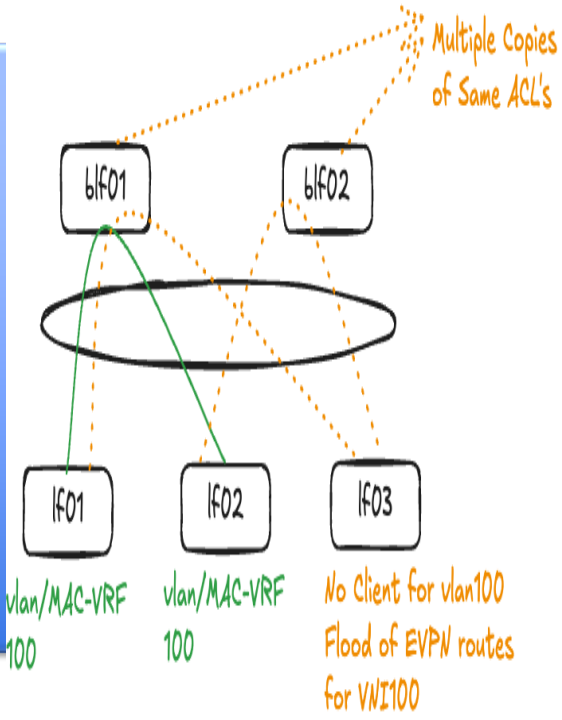


Migration Challenges – 2 Continued

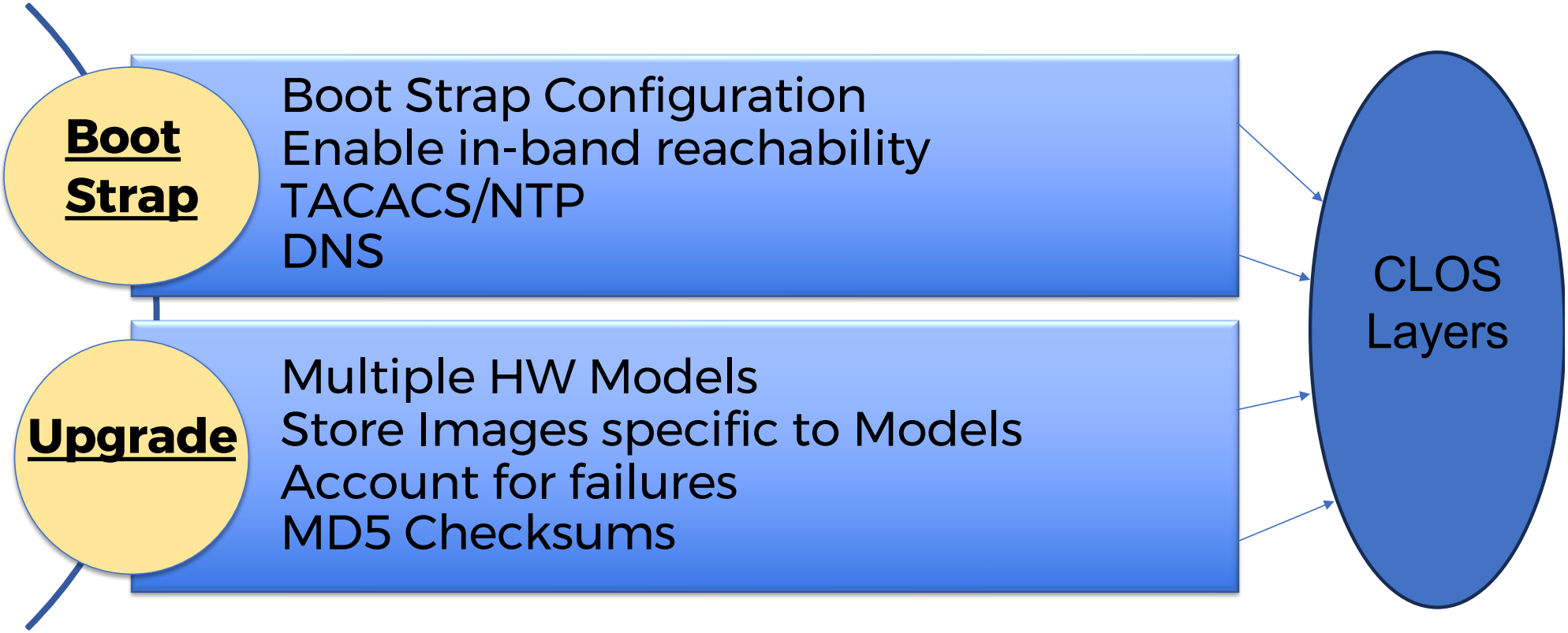
TCAM/ ACL's

EVPN Routes flooding via VX1 interface eating up TCAM on leafs

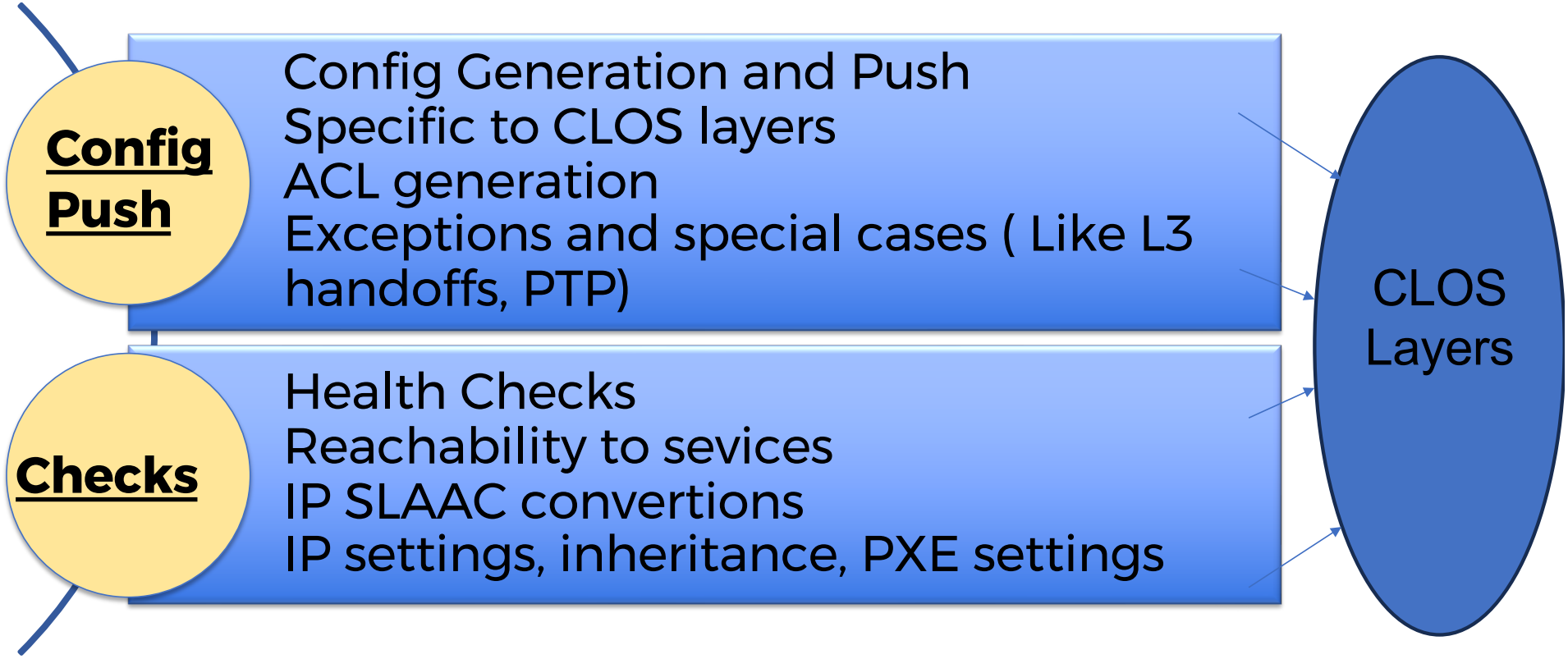
Multiple Copies of SAME ACL sit on the router BLF (L3 VXLAN gateway) layer eating up TCAM



Migration Challenges - 3



Migration Challenges – 3 Continued



Migration Challenges - 4

Operate Phase

•Managing high volume of tickets with prioritization and timely resolution.
•Ensuring accurate ticket categorization and routing to reduce delays.
•Achieving comprehensive observability across distributed systems.
•Handling alert fatigue due to excessive or noisy alerts.
•Ensuring reliable and secure OOB access for emergency troubleshooting.

Migration Challenges – 4 Continued

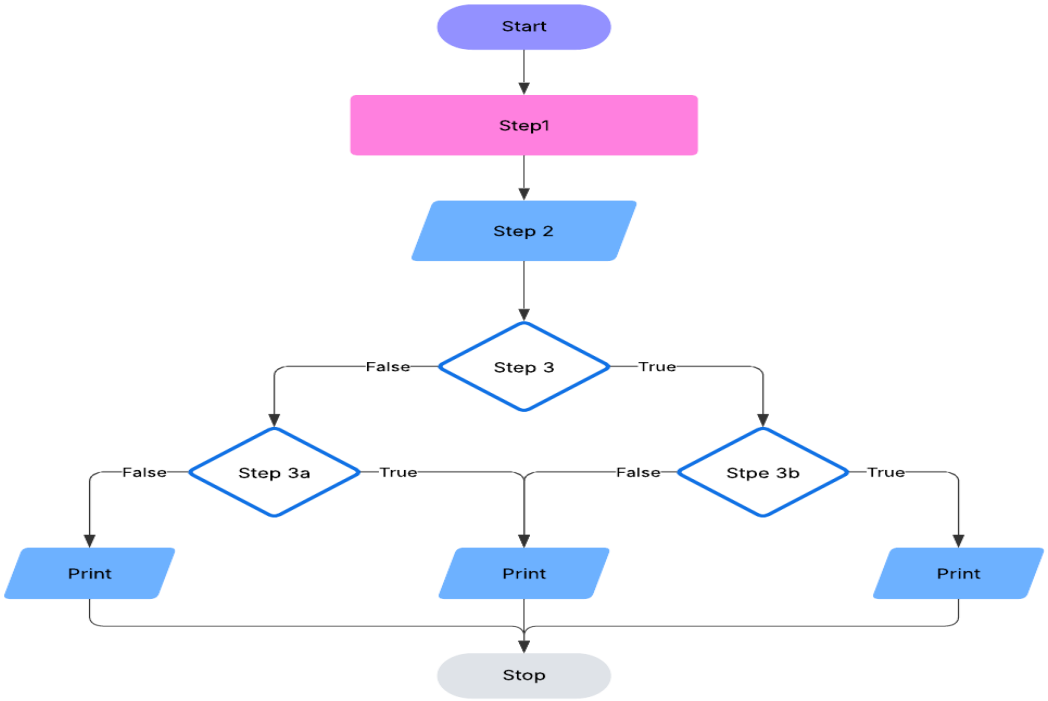
Run Phase

•Rapidly detecting and diagnosing incidents to minimize downtime.
•Prioritizing remediation efforts based on impact and risk.
•Maintaining accurate and up-to-date asset inventories at scale.
•Scheduling maintenance without disrupting critical operations.

Migration Challenges Solved - 1

Python workflow executing steps

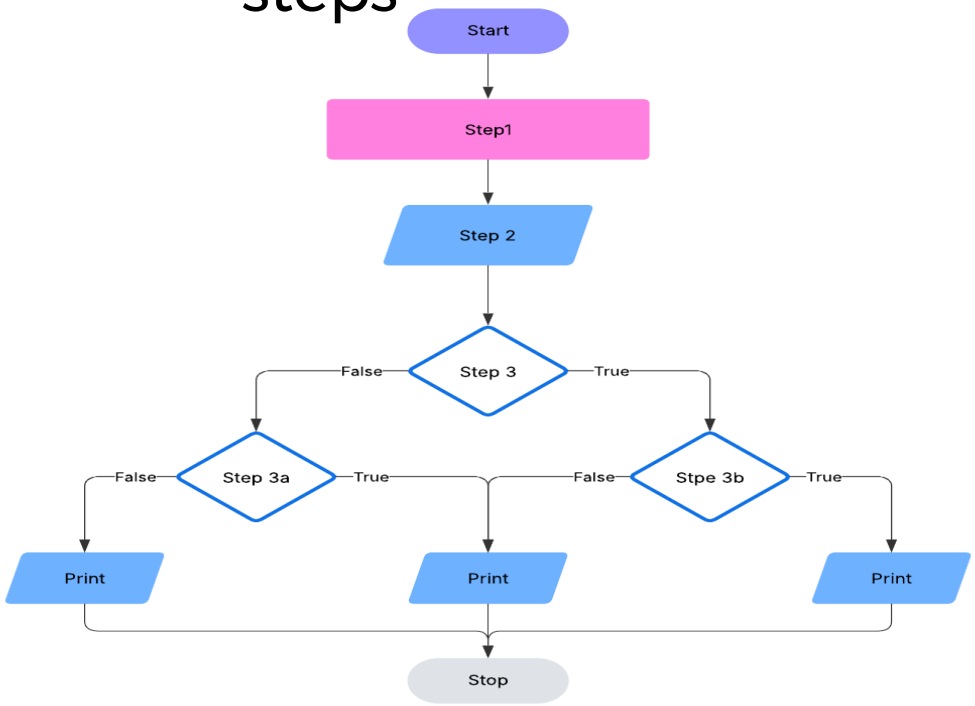
- Steps look at the intent fed per Site
  - Subnet in use,
  - Vlan/Port
- Get the right size
  - Subnet, P2P/service
- Allocate and finish



NANOG™

Migration Challenges Solved - 1

Python workflow executing steps



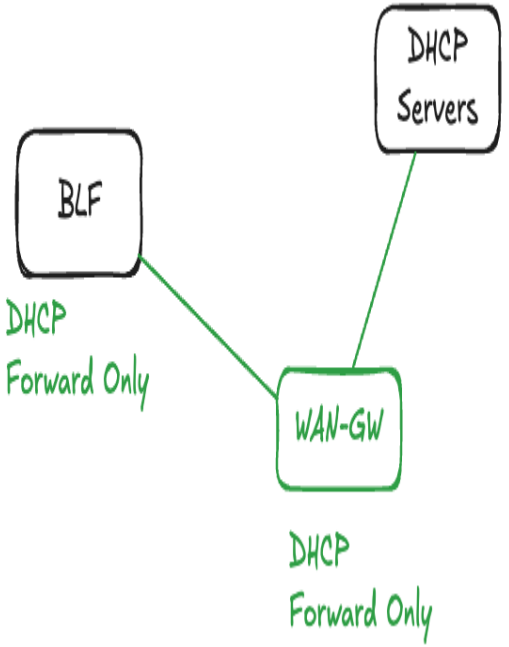
SLAAC Use cases Addressed

- Steps look at the intent fed per Site
  - RA timer requirements based on service vlan / locatior
  - Unique Knob's
- Get the right size
  - Service subnet
- Render config and finish

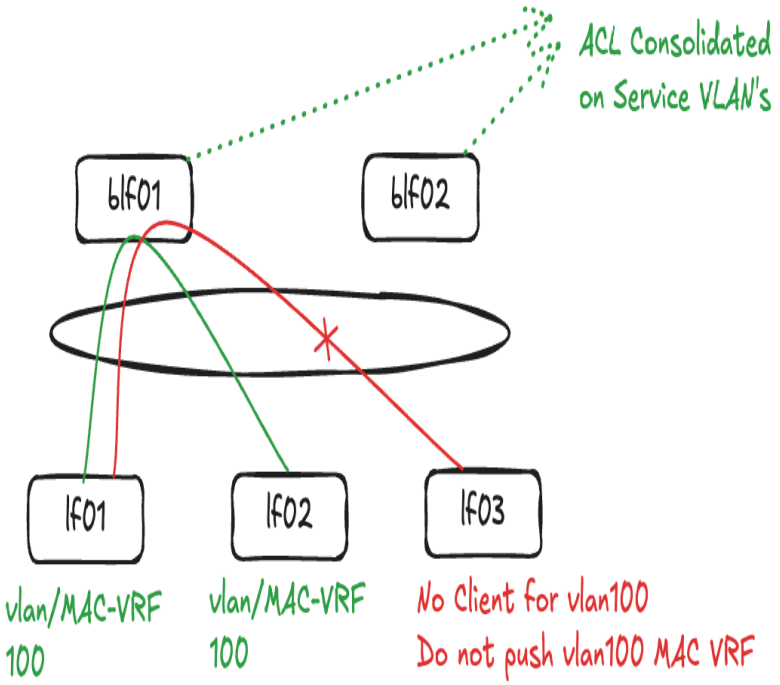
Migration Challenges Solved - 2

DHCP v4/V6 reliability- DUAL RELAY SCENARIO Addressed

- Lab infra uplinks to WAN-GW, WAN-GW snoop DHCP packets
- In-house tooling to check WAN-GW is acting as forward-only role and not punting the packets to CPU



Migration Challenges Solved - 2



TCAM/ACL's Ulitization Addressed

In-house tooling executions.

- ACL on the routed BLF layer consolidated and applied
- Only Configure MAC VRF where the service VLAN exists
- EVPN routes no longer flooded to all L2VTEPs

Migration Challenges Solved - 3

BootStarp

- Boot Startp Configuration
  - Jinja2 templates
  - Render
  - Push Configuration
  - Via Console

Upgrade

- Multiple HW Models
  - Centralized repo to store OS<>Model maps
  - Policies to push builds based on Model
  - MD5 checks
  - Automated upgrade steps via workflow

Migration Challenges Solved - 3

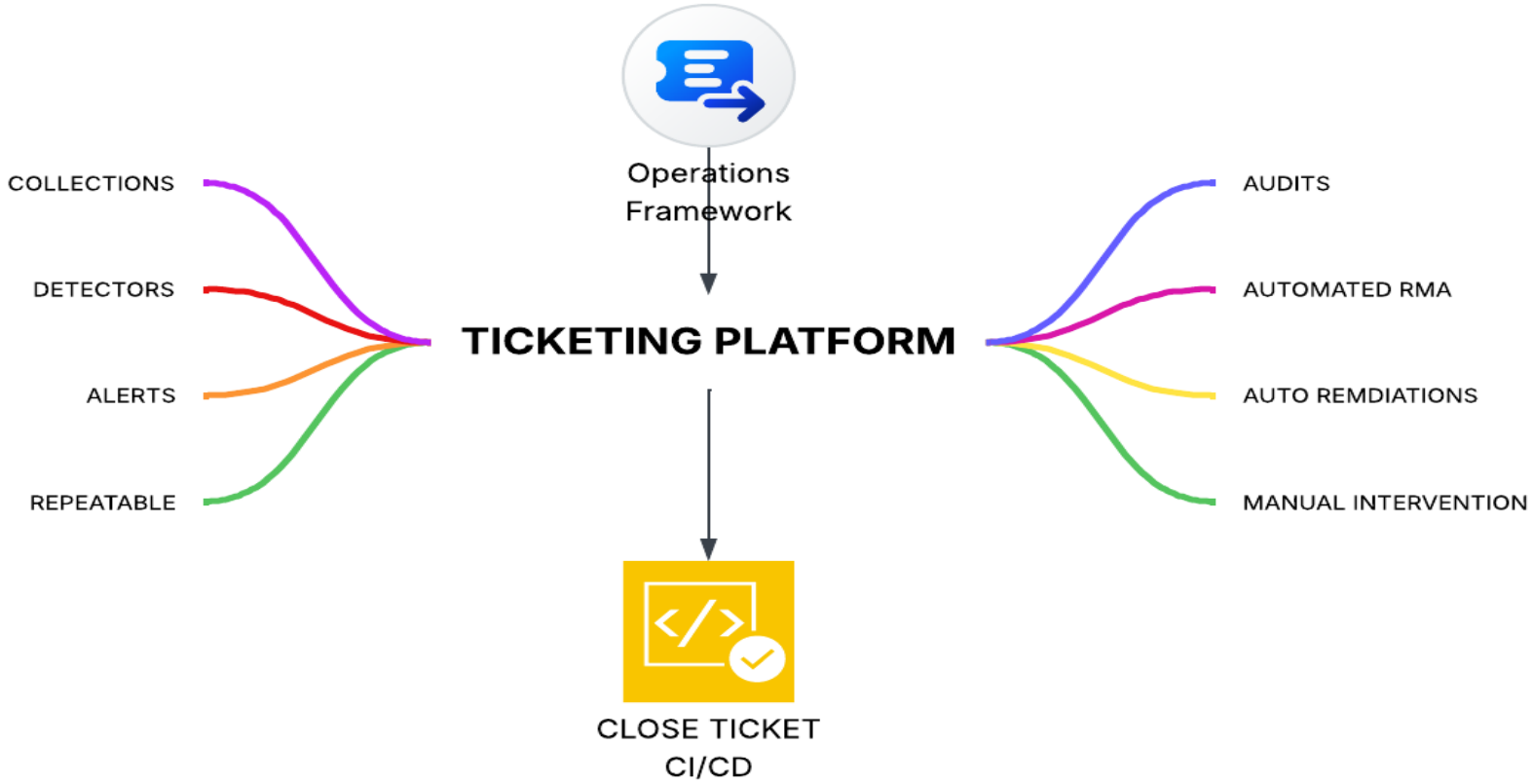
Config Push

- Config render and Push
  - Match the CLOS layers and jinja2 templates
  - ACL generation a seprate Push to maintain a copy of latest ACL's
  - Policies to handle Exceptions and special cases (Like L3 handoffs, PTP)

Pre/Post Checks

- Health Checks
  - Internal tools executing steps to check if the end devices and network devices are alive
  - Autmated IP SLAAC convertions, based on link local NDP table
  - Workflow steos updating IP settings, DNS/DHCP inheritance, PXE settings

Migration Challenges Solved - 4



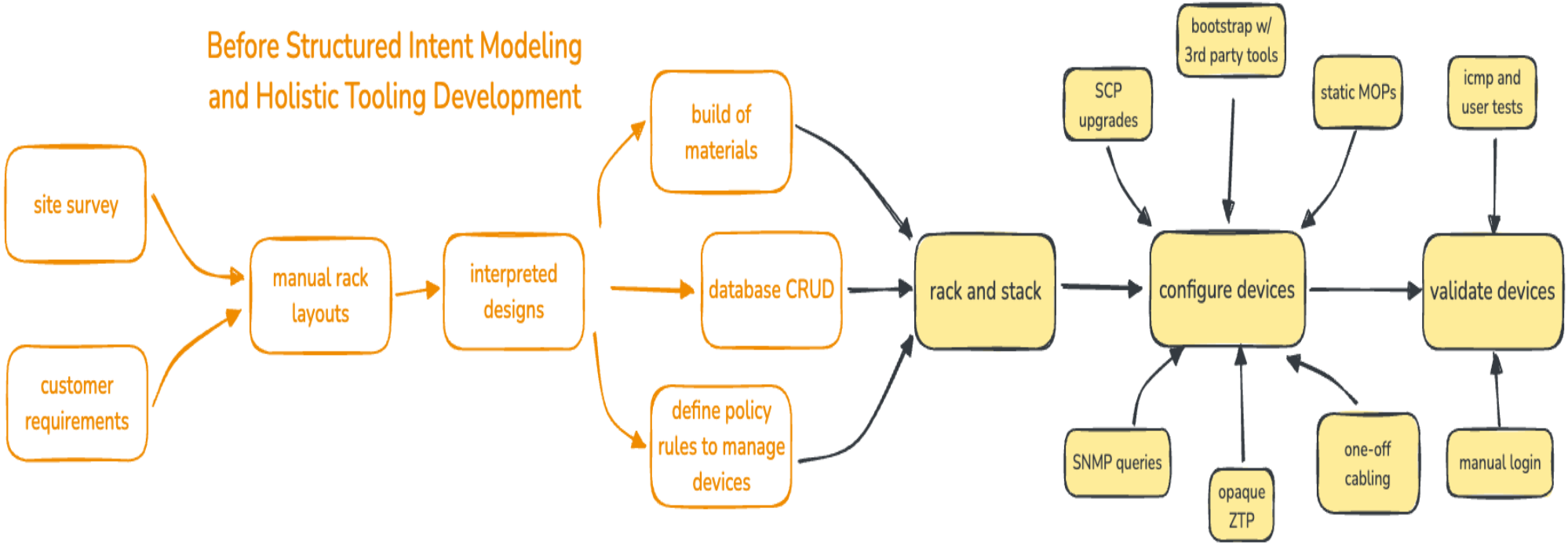
Migration Challenges Solved - 4

- Incident Response Automation: CI/CD pipelines to automate remediation scripts and updates, speeding up recovery from incidents and reducing manual intervention.

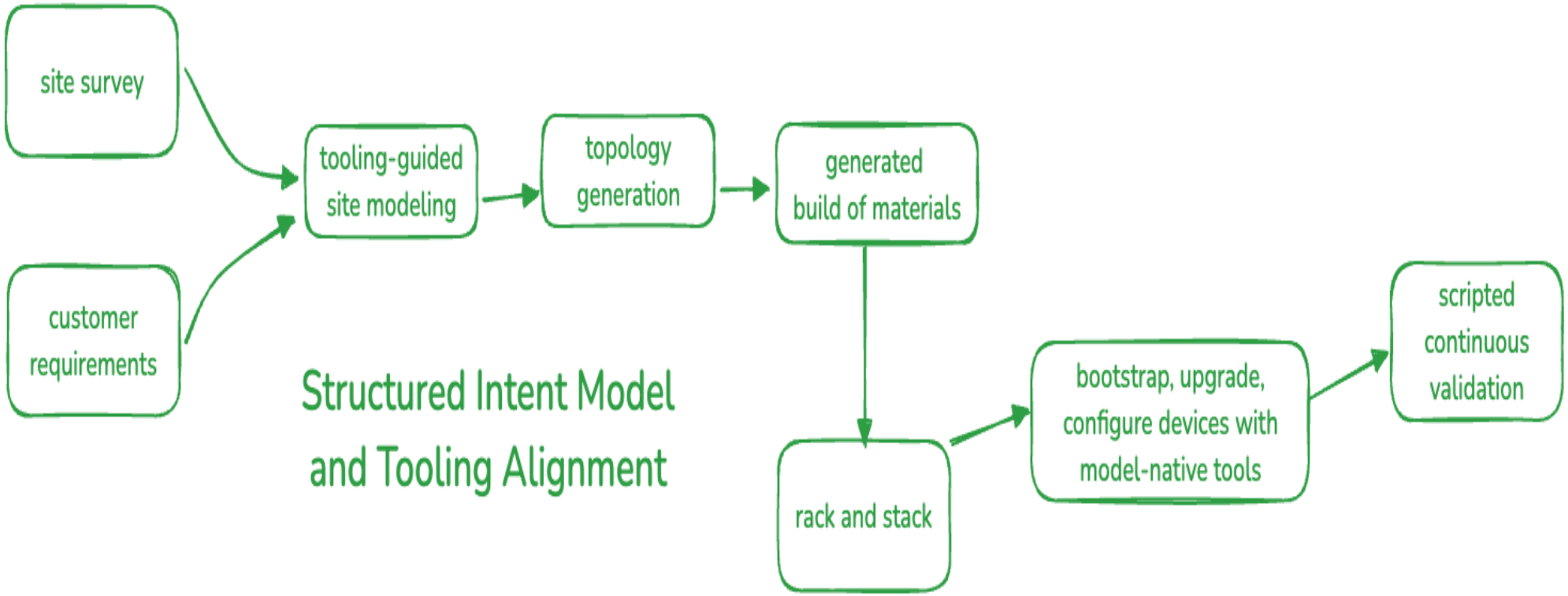
- Consistent Configuration Updates: CI/CD to ensure that asset inventories and equipment configurations updated and maintained.

- Monitoring and Observability Integration: CI/CD with monitoring tools to automatically deploy updates that improve observability and alerting, prevent issues proactively.

Rolling out at Scale- Before Intent



Rolling out at Scale- After Intent



Conclusion: SCALE SCALE SCALE Deployment and Operations

Tooling Continuous Integration / Continuous Deployment (CI/CD)
Integrate automated testing and deployment pipelines for network configurations.
Failing quick: Immediate feedback loop for early detection and correction.

Documentation & Knowledge Sharing
Maintain up-to-date runbooks and network diagrams.
Encourage documentation updates alongside code/config changes.

Conclusion: SCALE SCALE SCALE Deployment and Operations

Monitoring & Observability
Implement centralized logging, real-time monitoring, and alerting.
Use telemetry and analytics to predict and prevent issues.

Change Management
Utilize formal change windows and review processes.
Implement rollback plans for fast recovery from failures.

Thank you