
cisco
You make possible

Data Center Operations and Maintenance Best Practices

Arvind Durai, Director Solutions Integration

Anis Edavalath, Technical Leader Engineering

BRKDCN-2458

Arvind Durai

– 20 years with Cisco Advanced Services

– Has worked with 100+ customers in enterprise architecture, technology designs and operational simplification

– 11 years of Active Cisco live presenter

– Co-authored five Cisco Press Books

Cisco Firewall Services Module, Virtual Routing in the Cloud, TCL Scripting for Cisco IOS and IP Multicast vol1 & 2

– CCIE R/S and Security #7016

Anis Edavalath

- 7 years with Cisco Advanced Services

Enterprise Campus and Datacenter across different verticals

-Worked 10 years with BU engineering groups in Security , switching, datacenter and Network Management products

- Design and deployment of Next Gen Data center architecture enterprise and cloud customers
- AS team lead for ACI, VxLAN, Tetration, SDA (uniform policy)
- Worked with major telecom vendors and Cloud providers prior to Cisco

- CCIE Datacenter # 48152

Contributors: Satish Kondalam, Nick Garner Junmei Zhang and many others from the Nexus TME team.

Course Objective and Goal

- To help Data Center operations and engineering staff understand the operational best practices when maintaining a Cisco Nexus data center network deployment.

- Attendees should leave the session with a firm understanding of
  - Operational Best Practices & next gen tools
  - Nexus Graceful Insertion and Removal
  - Change Window Best Practices

Agenda

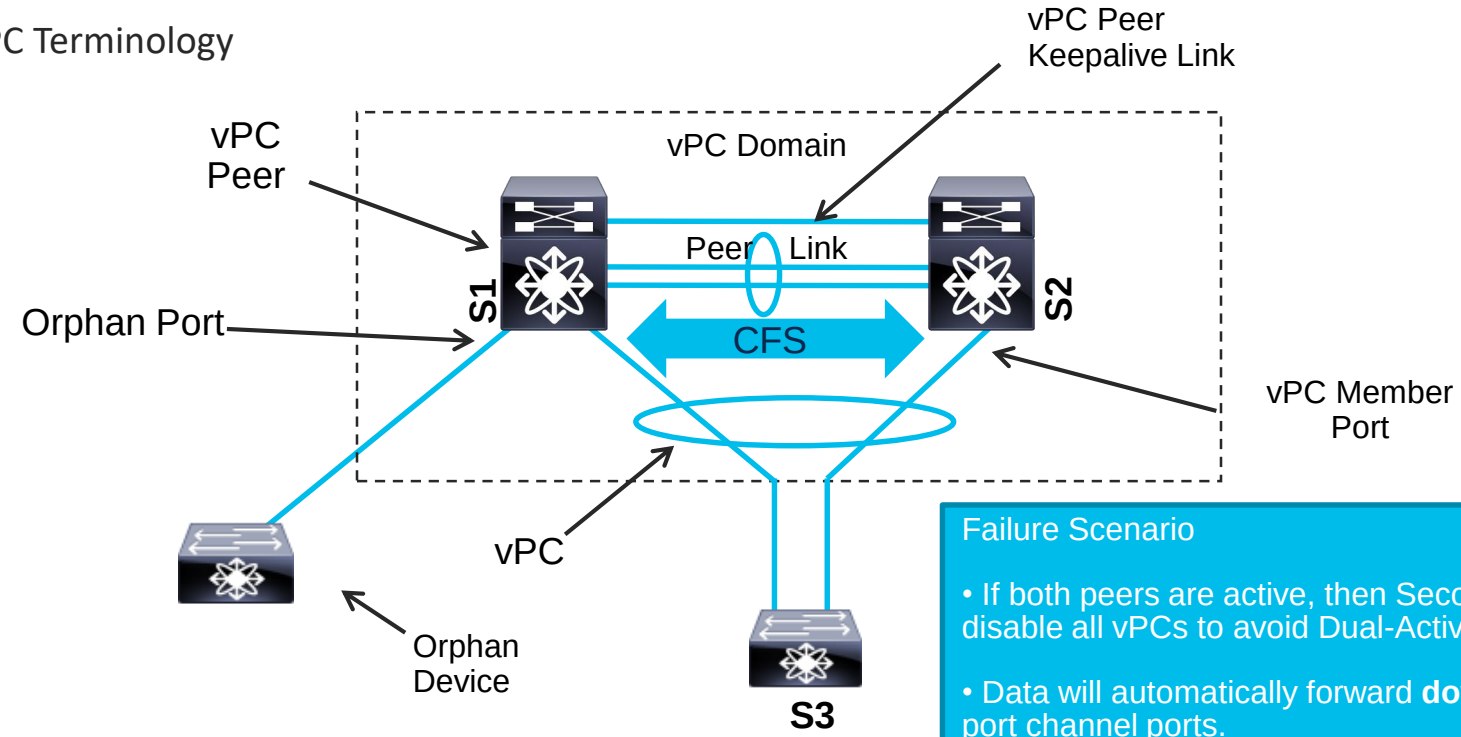
- vPC and VxLAN Refresher
- Operational Best Practices: Software
- Operational Best Practices: Hardware
- Node Isolation
- NX-OS Graceful Insertion and Removal
- ACI Operational Best Practices
- Data Center operation tool framework & use case demo
- Data Center Behavioral Monitoring - Tetration
- Change Window Best Practices

DC Baseline Refresher

vPC Feature Overview

For Your Reference

- vPC Terminology



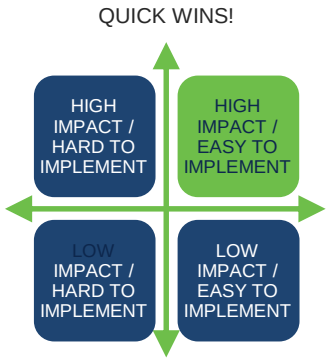
Failure Scenario

- If both peers are active, then Secondary vPC peer will disable all vPCs to avoid Dual-Active.
- Data will automatically forward **down** remaining active port channel ports.
- Loss of in-flight packets will depend on deployment of vPC best practice.

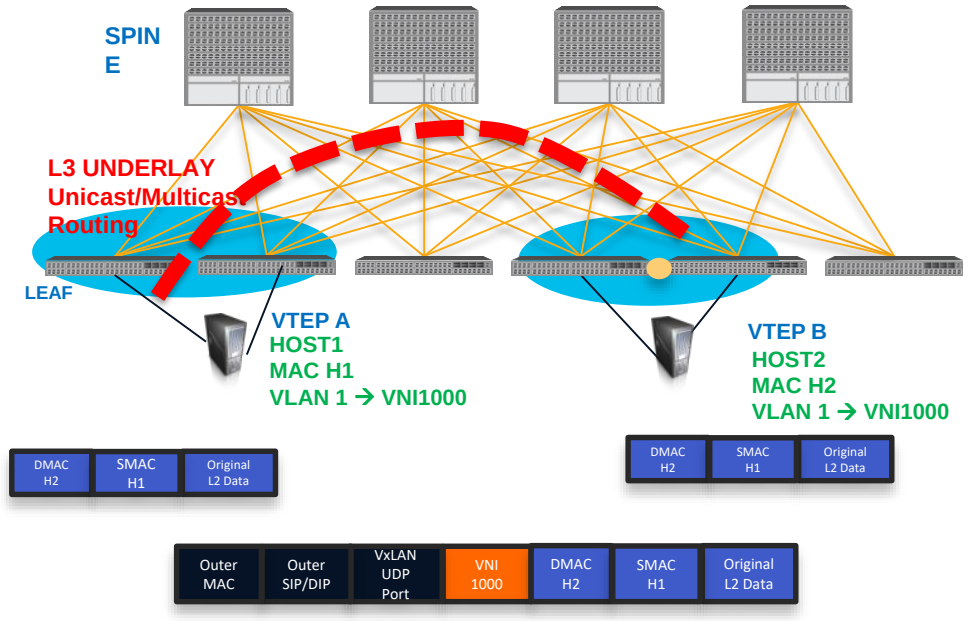
vPC Best Practice

vPC General Deployment Best Practice

- vPC Domain ID's
  - ✓ Use a unique vPC domain ID within a contiguous L2 domain to avoid MAC overlap.
- vPC Peer Link
  - ✓ Should be point-to-point connection & dedicated links.
- vPC Peer Keepalive Link
  - ✓ Dedicate a control plane in a dual-supervisor environment. Use a management switch.
- vPC peer-gateway
  - ✓ Acts as active gateway for frames addressed to peer switch. Avoid Peer Link forwarding.
- Use vPC peer-switch
  - ✓ Optimizes BPDU processing, single logical L2 entity
- Distribute port-channel member interfaces across line cards within the same chassis.
- Create a map for oversubscription aligned to current and future demand.
  - ✓ Deployment practice – 20:1 at access and 2:1 at Core.



VXLAN Overview



VTEP A or VTEP B in deployment will be a pair, and this pair will provide host redundancy for Layer 2 via VPC.

VPC is still NEEDED and VTEP will represent the VPC pair!

Layer 2 overlay on top of your Layer 3 underlay

- Each VXLAN Segment is identified by a unique 24-bit segment ID called a VXLAN Network Identifier (VNI)
- Only hosts on the same VNI are allowed to communicate with each other
- Original L2 packet is encapsulated with VXLAN header in a UDP->IP->Ethernet

Overcome 4094 VLAN Scale Limitation

- VLANs use a 10-bit VLAN ID

Better utilization of available network paths

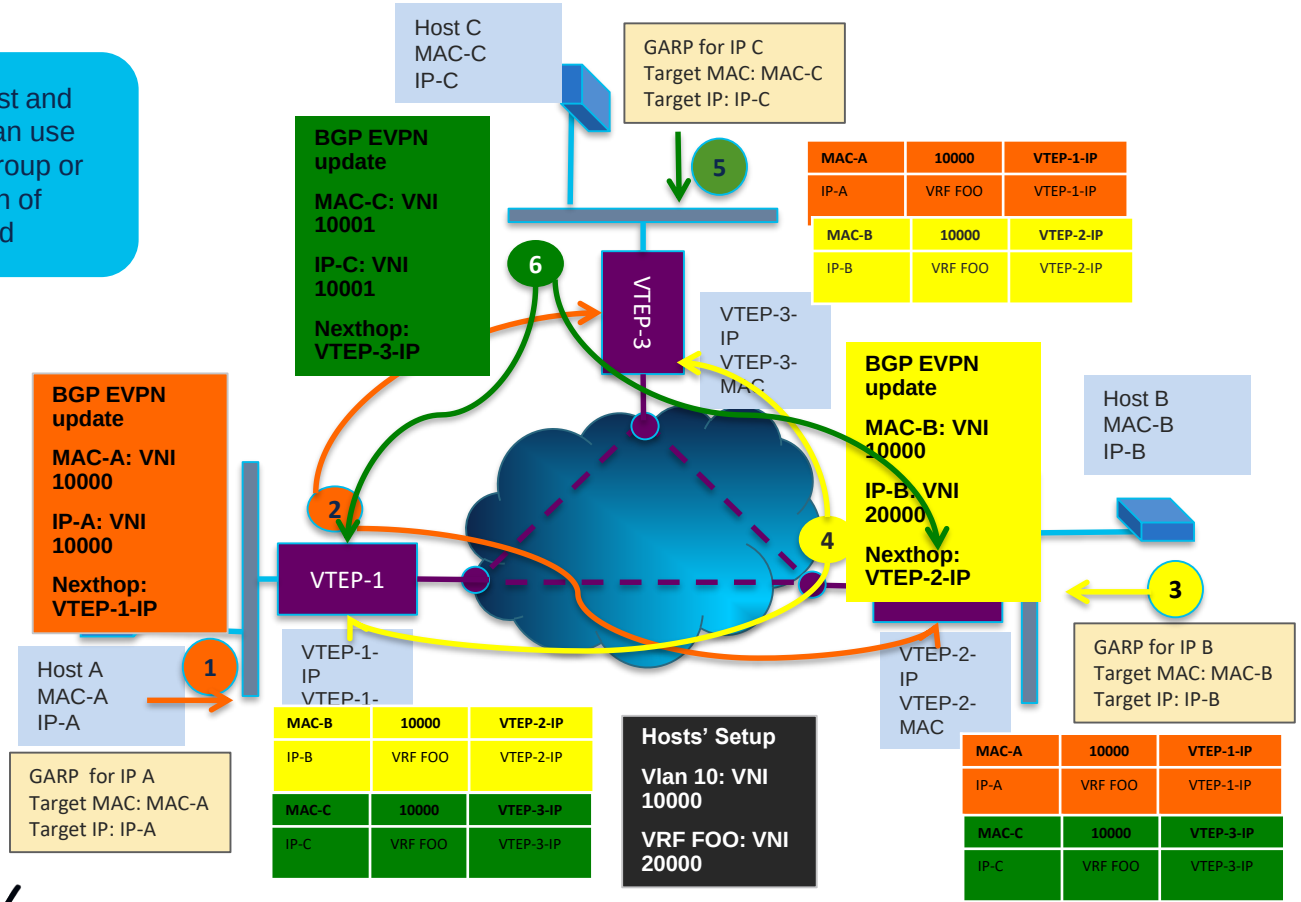
- No need of Spanning Tree (blocks paths)
- Utilize L3 underlay network (ECMP, Link Agg,…)

Multi-Tenant with virtualization

- Isolation of network traffic by a tenant and reusability of networking taxonomy for tenancy

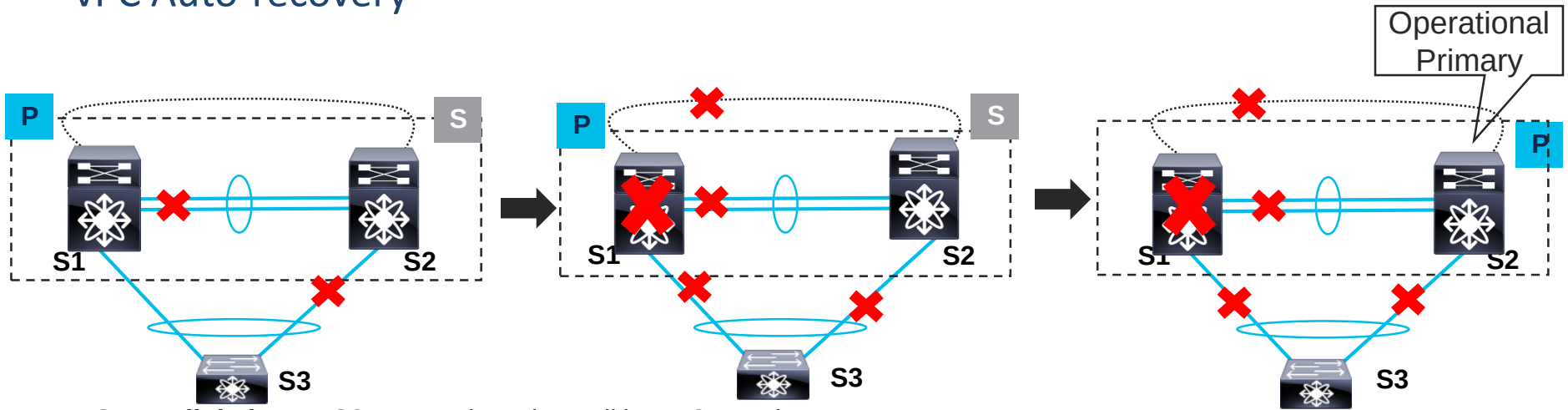
VxLAN Refresher With BGP EVPN Address Learning

Broadcast, Unicast and Multicast traffic can use either Multicast group or Ingress replication of traffic- not covered



vPC Configuration Best Practices

vPC Auto-recovery



1. **vPC peer-link down** : S2 - secondary shuts all its vPC member ports
2. **S1 down** : **vPC peer-keepalive link down** : S2 receives no keepalives
3. After 3 keepalive timeouts, S2 changes role and brings up its vPC

Auto-recovery addresses two cases of single switch behavior

- Peer-link fails and after a while primary switch (or keepalive link) fails
- Both VPC peers are reloaded and only one comes back up

P vPC Primary

S vPC Secondary

vPC Configuration Best Practices

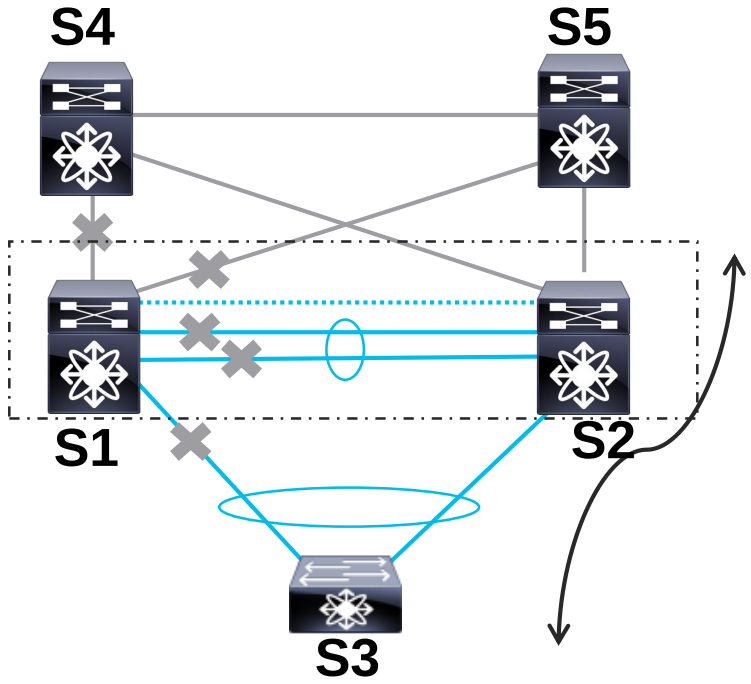
Object-tracking

- vPC object tracking, tracks both peer-link and uplinks in a list of Boolean OR
- Object Tracking triggered when the track object goes down
- Suspends the vPCs on the impaired device.
- Traffic forwarded over the remaining vPC peer.

```
! Track the vpc peer link
track 1 interface port-channel11 line-protocol
! Track the uplinks
track 2 interface Ethernet1/1 line-protocol
track 3 interface Ethernet1/2 line-protocol

! Combine all tracked objects into one.
! "OR" means if ALL objects are down, this object will go down
track 10 list boolean OR
object 1
object 2
object 3

! If object 10 goes down on the primary vPC peer,
! system will switch over to other vPC peer and disable all local vPCs
vpc domain 1
track 10
```

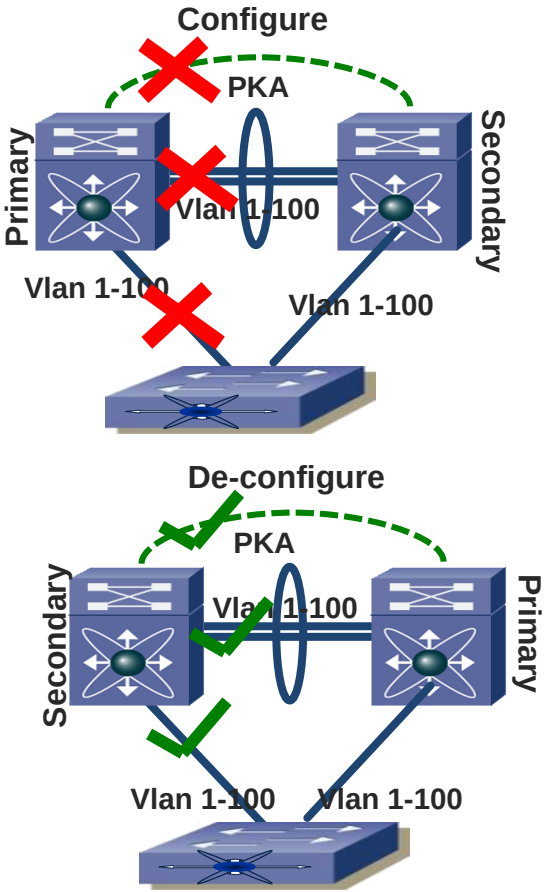


VPC Shutdown Feature

This feature allows customer to manually “isolate” a switch from vPC domain. This is a vPC configuration option.

Pre-VPC Shutdown	VPC Shutdown Behavior

• No “shutdown” command.
• Manual Shutdown Required
  • Down vPCs
  • Down Peer Link
  • vPC Members
  • Etc.
	• Local switch isolated from remote.
	• Cannot exit shutdown without manual intervention.
	• When exiting, PKA, PL, and vPCs will be re-initialized; vPC domain brought to normal state.

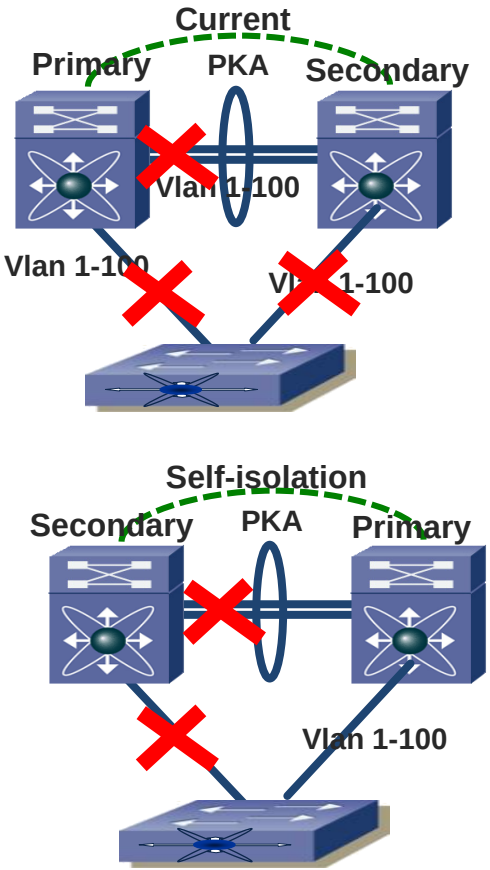


Availability 3k/5k/6k/7k/9k

VPC Self-isolation

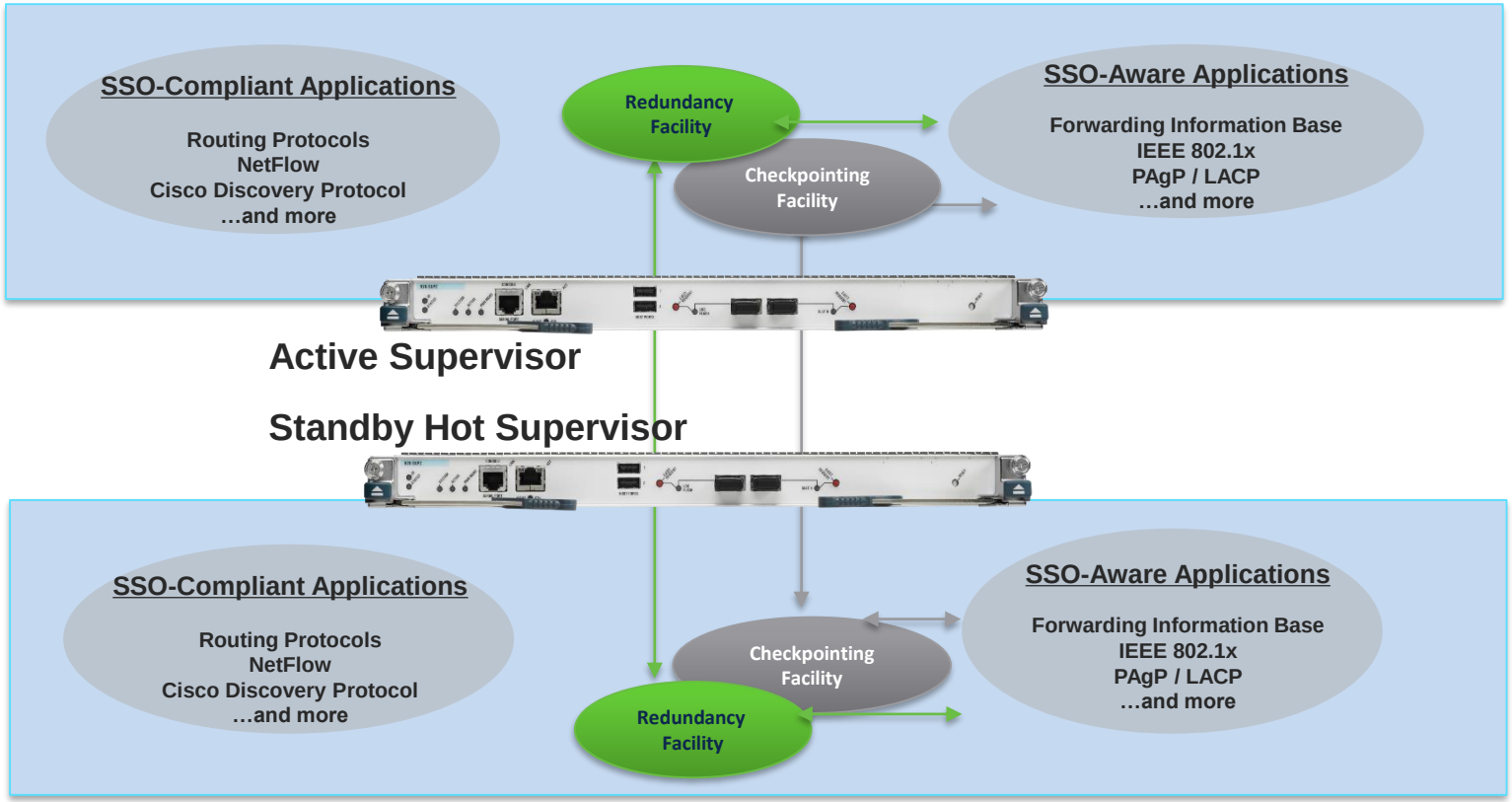
- Automatically triggered isolation
- Example Presented: All Line Cards Fail

Current Impact	Self-isolation Feature Behavior
• When this failure happens on primary, peer-link is brought down. • This causes the secondary to bring down all legs. • Traffic is completely blocked.	When this failure happens: •Physically bring down peer-link •Physically bring down all vPC legs •Send self-isolation through peer-keep-alive Peer switch: •Receive self-isolation from the peer through peer-keep-alive •Change role to Primary •Bring up all down vPC legs **BU Testing Results:** Sub-second Recovery (N>S) (S>N) (E>W)



NOTE: Available in NX-OS 7.2, 5k/6k/7k

Stateful Switchover Mode
SSO-Aware and SSO-Compliant Applications



Routing Protocol Redundancy With NSF
(Graceful Restart)

Active Supervisor Engine Slot 1

EIGRP RIB	
Prefix	Next Hop
10.0.0.0	10.1.1.1
10.1.0.0	10.1.1.1
10.20.0.0	10.1.1.1

OSPF RIB	
Prefix	Next Hop
192.168.0	192.168.0.1
192.168.55.0	192.168.55.1
192.168.32.0	192.168.32.1

ARP Table	
IP	MAC
10.1.1.1	aabbcc:ddee32
10.1.1.2	adbb32:d34e43
10.20.1.1	aa25cc:ddeee8

Standby Supervisor Engine Slot 2

EIGRP RIB	
Prefix	Next Hop
-	-
-	-
-	-

OSPF RIB	
Prefix	Next Hop
-	-
-	-
-	-

ARP Table	
IP	MAC
-	-
-	-
-	-

FIB Table	
Prefix	Next HOP
10.1.1.1	aabbcc:ddee32
10.1.1.2	adbb32:d34e43
192.168.0.0	aa25cc:ddeee8

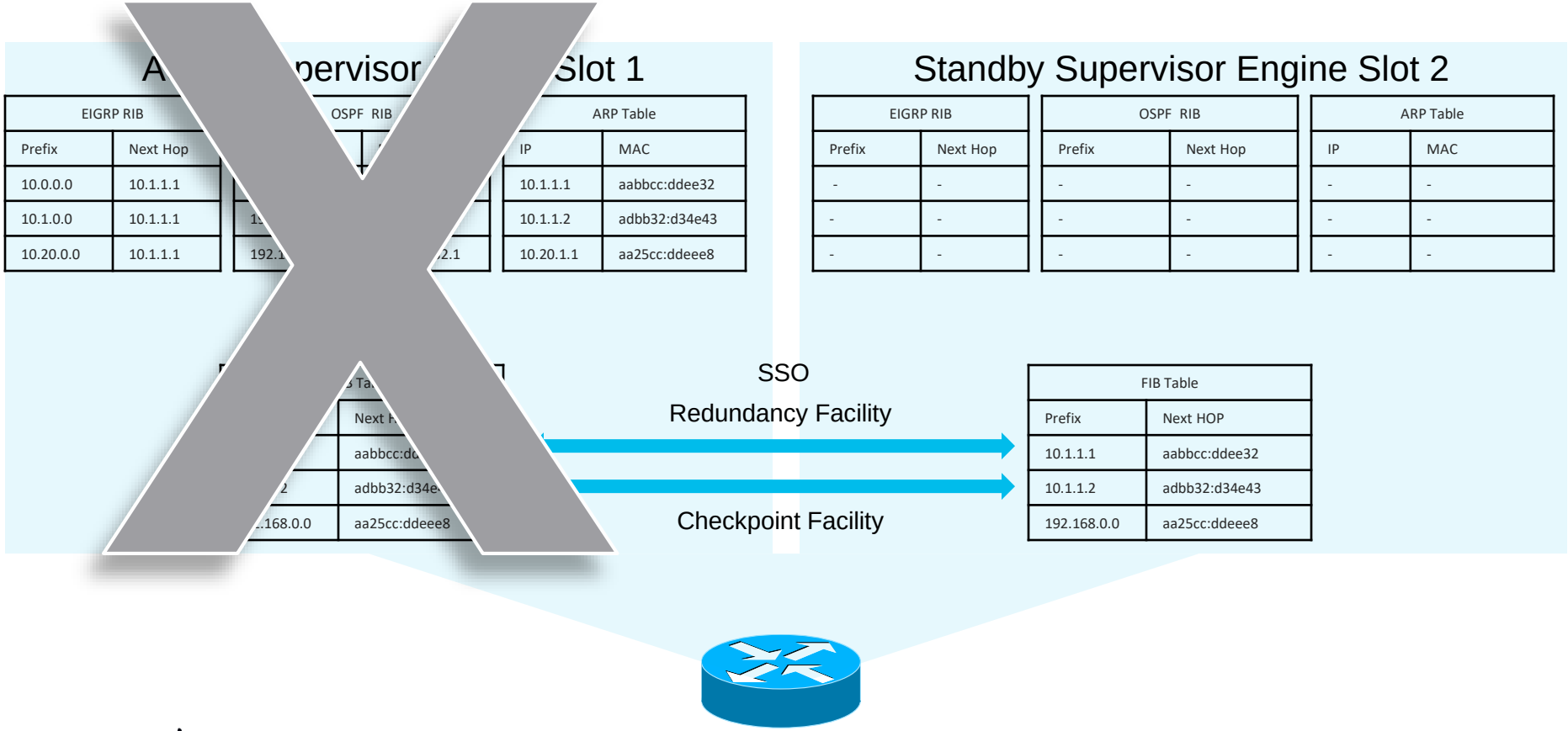
SSO

Redundancy Facility

Checkpoint Facility

FIB Table	
Prefix	Next HOP
10.1.1.1	aabbcc:ddee32
10.1.1.2	adbb32:d34e43
192.168.0.0	aa25cc:ddeee8

Routing Protocol Redundancy With NSF (Graceful Restart)



Routing Protocol Redundancy With NSF
(Graceful Restart)

Standby Supervisor Engine Slot 2

EIGRP RIB	
Prefix	Next Hop
10.0.0.0	10.1.1.1
10.1.0.0	10.1.1.1
10.20.0.0	10.1.1.1

OSPF RIB	
Prefix	Next Hop
192.168.0	192.168.0.1
192.168.55.0	192.168.55.1
192.168.32.0	192.168.32.1

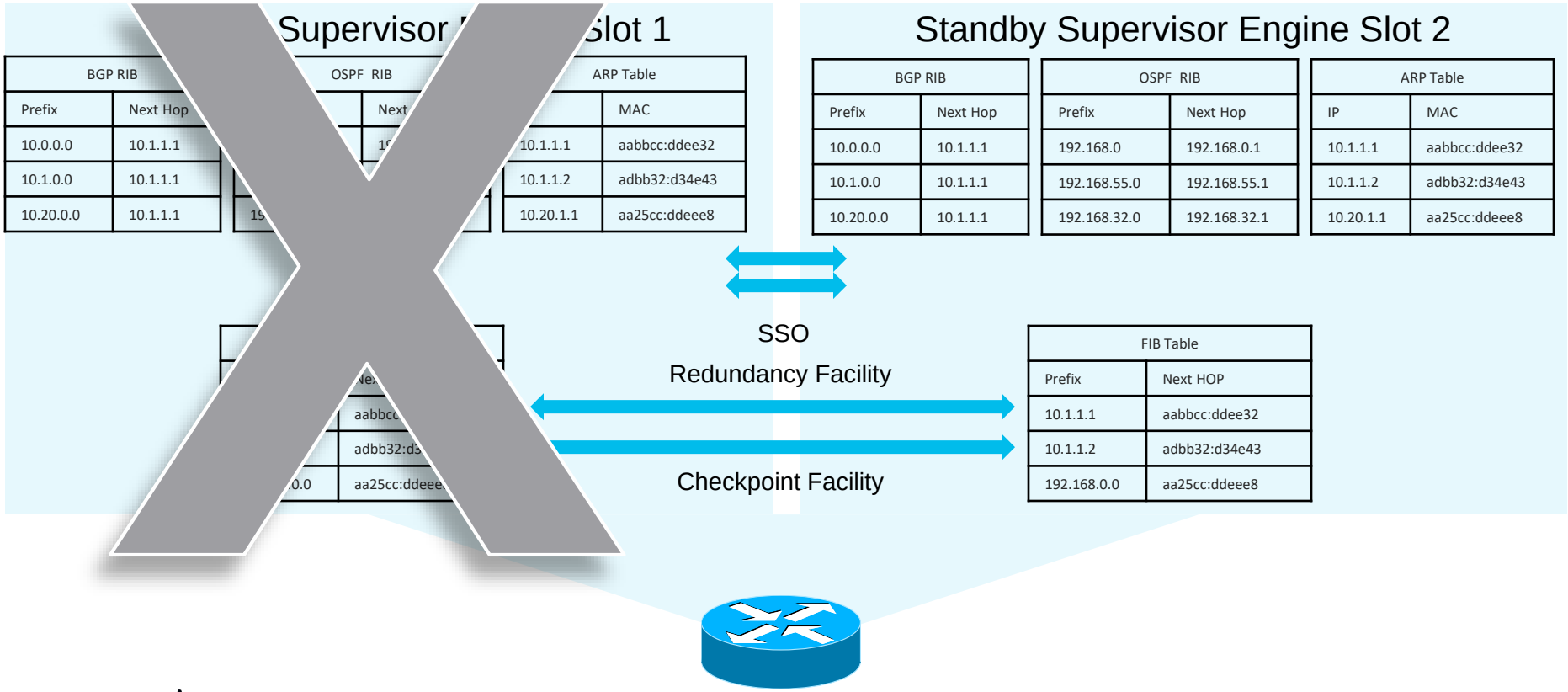
ARP Table	
IP	MAC
10.1.1.1	aabbcc:ddee32
10.1.1.2	adbb32:d34e43
10.20.1.1	aa25cc:ddeee8

FIB Table	
Prefix	Next HOP
10.1.1.1	aabbcc:ddee32
10.1.1.2	adbb32:d34e43
192.168.0.0	aa25cc:ddeee8

GR/NSF Signaling per protocol

Synchronization per protocol

Routing Protocol Redundancy With NSR
(Stateful Restart)



Routing Protocol Redundancy With NSR
(Stateful Restart)

Standby Supervisor Engine Slot 2

BGP RIB	
Prefix	Next Hop
10.0.0.0	10.1.1.1
10.1.0.0	10.1.1.1
10.20.0.0	10.1.1.1

OSPF RIB	
Prefix	Next Hop
192.168.0	192.168.0.1
192.168.55.0	192.168.55.1
192.168.32.0	192.168.32.1

ARP Table	
IP	MAC
10.1.1.1	aabbcc:ddee32
10.1.1.2	adbb32:d34e43
10.20.1.1	aa25cc:ddeee8

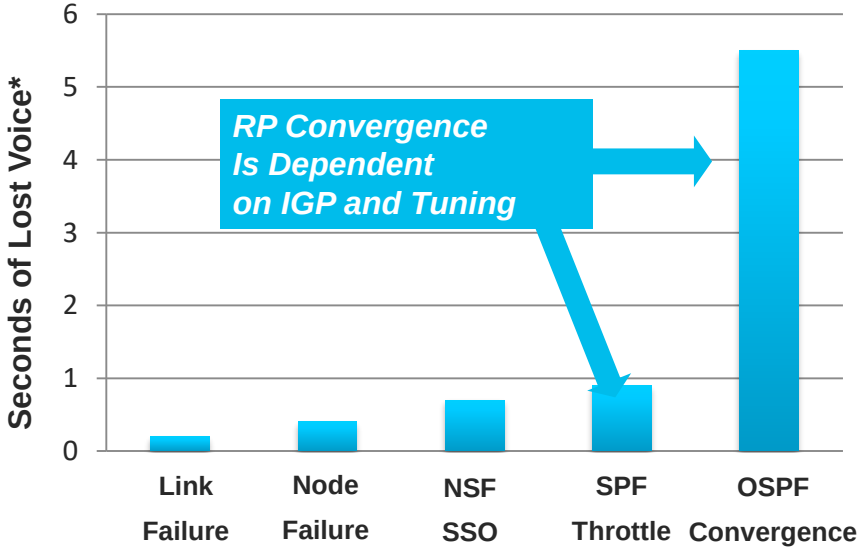
FIB Table	
Prefix	Next HOP
10.1.1.1	aabbcc:ddee32
10.1.1.2	adbb32:d34e43
192.168.0.0	aa25cc:ddeee8

No additional signaling required to maintain topology

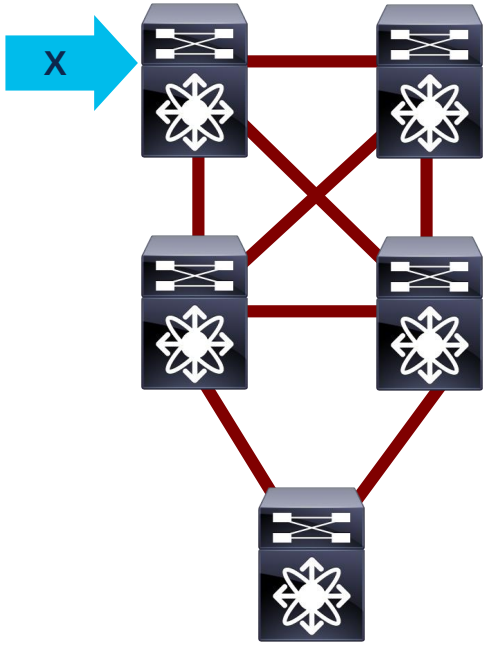
Standalone Chassis Redundant Core

Failure or Change at the Core

- Redundant topologies with equal cost paths provide sub-second convergence.
- NSF/SSO provides superior availability in environments with non-redundant paths.



* Route scale dependent.



Best practices:

- Layer 3
- Layer 2
- Hardware

- Enable BFD for all OSPF neighbor links
- Adjust OSPF spf-throttling timers with:

```
timers throttle spf
timers throttle lsa
timers lsa arrival
```

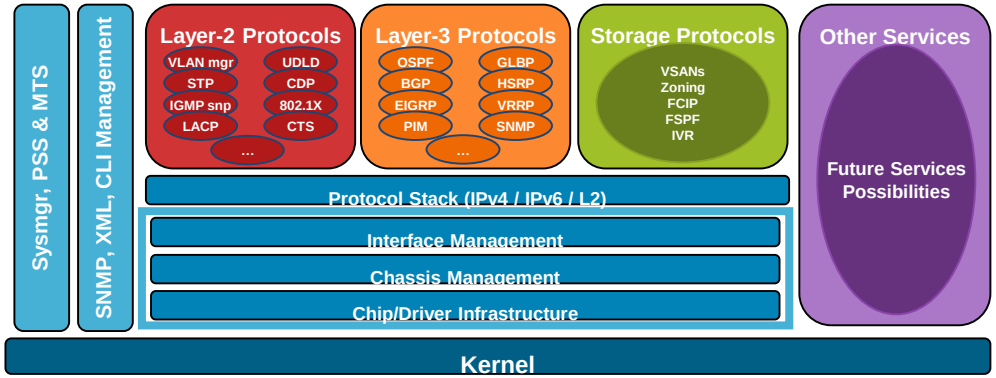
Operational Best Practices Software

NX-OS High Availability

Process Modularity

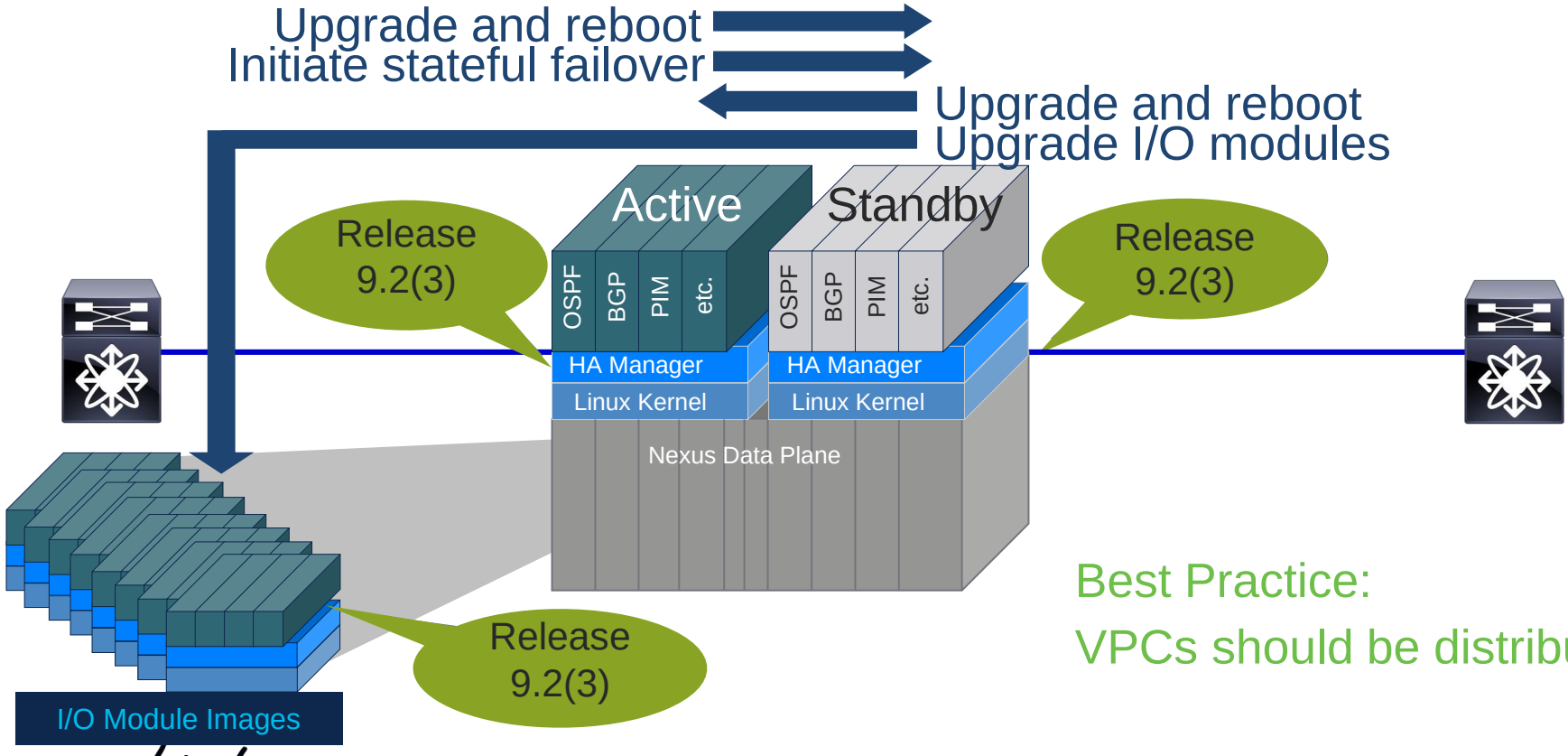
- Independent memory-protected restart-able processes
- Service Restart-ability
  - Stateful Restart with

    Persistent Storage Service (PSS)
    - Checkpoints states to PSS
    - Recover states from PSS upon restart.
  - Stateful Restart with Graceful Restart
    - Recover states based on information from other services and/or network.
    - Mainly Routing Protocols
  - Stateless Restart
    - Fresh start, no trace of former instantiation.



In-Service Software Upgrade

```
Nexus# install all nxos bootflash:nxos.9.2.3.bin
```



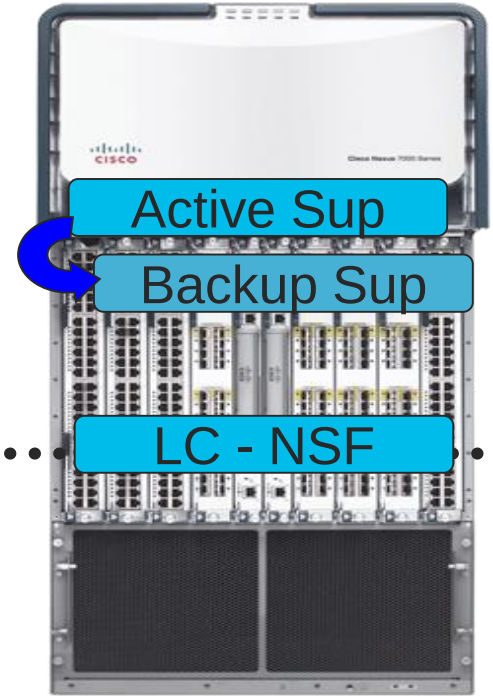
Best Practice:

VPCs should be distributed.

NX-OS High Availability

Supervisor Switchover

- Stateful Switchover (SSO)
  - Active-backup supervisors synchronized at all times
- Routing Protocols: → PSS Stateful Restart
  → NSF Graceful Restart failover
- Other components: → PSS Stateful Restart
- Triggers:
  - HA Policy Initiated – e.g. 3 component crashes → SSO
  - User Initiated – system switchover
  - ISSU initiated SSO



NX-OS High Availability

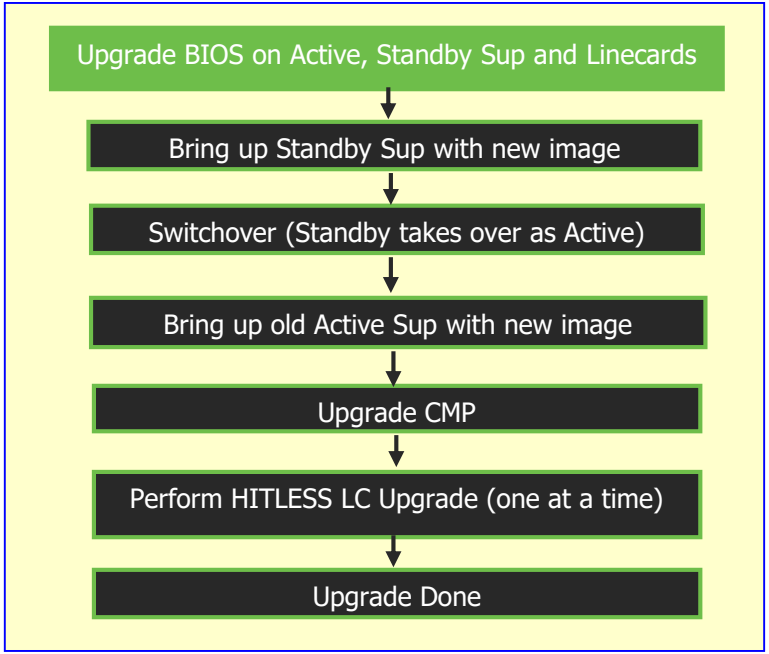
ISSU

- Dual-supervisor failover only
- ISSU is user initiated:
  - Compatibility Check: `show install all impact …`
  - Through CLI

    ```
    For N7k: install all kickstart <kickstart image>
    system <system image> cmp <cmp image>

    Fir N9k: install all nxos <system image>
    ```

  - Components upgraded:
  - Supervisor: BIOS, System image
  - Linecard: BIOS and Linecard image
- System wide upgrade
- Single-supervisor ISSU is not possible on the modular n9k. Service disruption *might* occur.*



NX-OS High Availability - Innovation

ISSU

Enhanced ISSU or LXC ISSU on N9k ToR (Single Sups)

By Creating Virtual instances on Sup and LC
Separate standby sup is brought up inside LXC
6s Control plane down time

Enhanced ISSU or LXC ISSU on N9k ToR

By Creating Virtual instances on Sup and LC
Separate standby syp is brought up inside LXC
6s Control plane down time

```
switch(config)# boot mode lxc
Using LXC boot mode
Please save the configuration and reload system to switch into the LXC mode.
switch(config)# copy r s
[##################################] 100%
Copy complete.
```

ISSU on EoR

Defect Impact

TAC: You've encountered defect CSCxy12345. It's operationally impacting and, I'm sorry to say, there's no workaround. You'll need to **upgrade**.

Belay my last. We have a SMU for that.

You: Fine. Let's just get it fixed.

Bill, start up a war room.

John, get our AS NCE on the phone.

Sally, schedule testers in two hours.

Where's my $#@! coffee?

What?

Gesundheit.

Sally: You know how Richard gets when we call him at 2 AM...

Software Patching in NX-OS

Who's familiar with Software Maintenance Updates (SMU)?

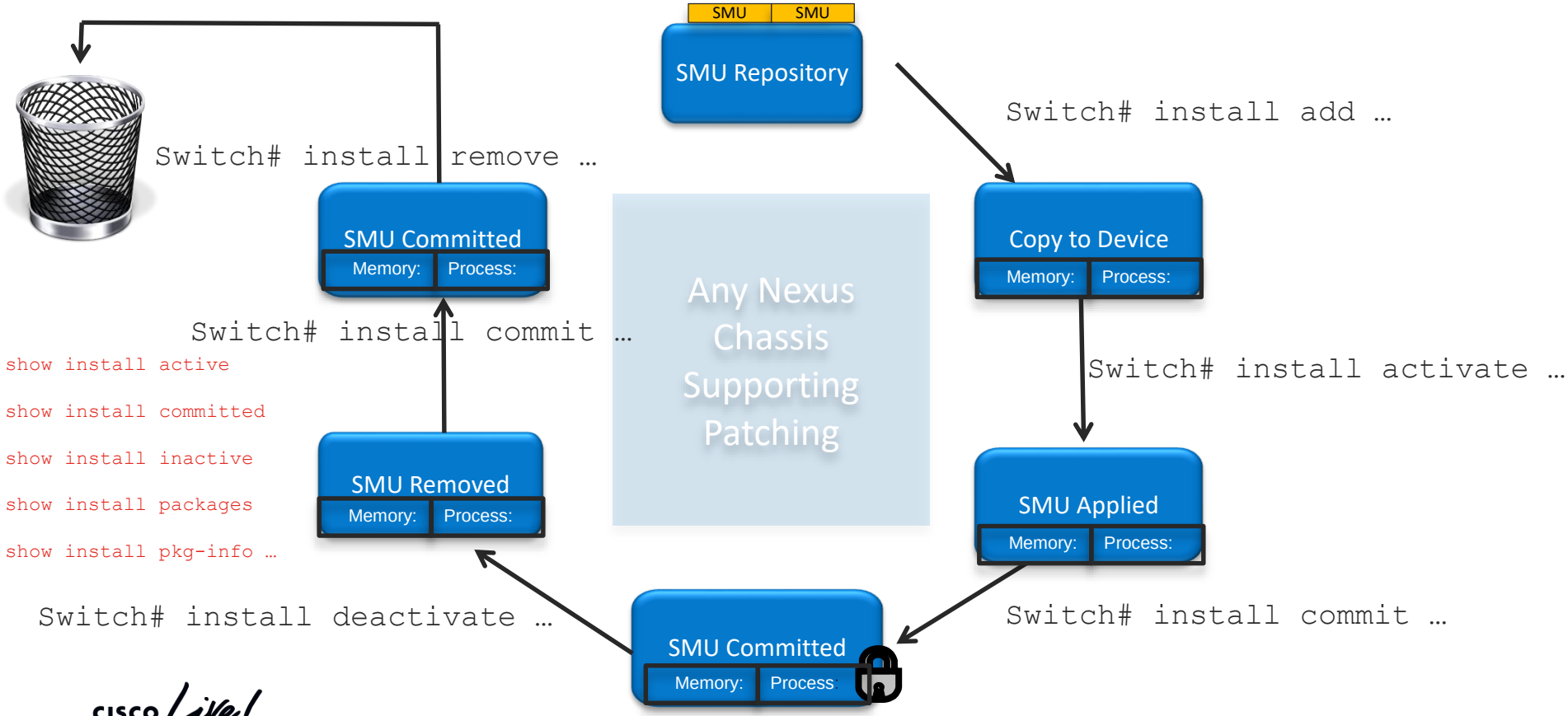
Overview

- Software Patching is Platform Independent
- Available on Nexus 9000 (6.1(2)I2)
- FCS NX-OS 7.2 (5/6/7k)
- Fully supported with ISSU

Benefits

- Reduce time to resolution in your network.
- SMUs in NX-OS build upon years of experience in IOS XR.
- Simplify customer operations for defect resolution and code qualification.
- Better utilize the software HA capabilities of NX-OS.
- Provide a common cross-platform experience (N9K/N7K/N6K/N5K).

SMU Lifecycle – CLI



Patching Highlights

SMU Types

- Restart: Restarts affected process
  - Process restarted in all VDCs where running.
- ISSU SMU:
  - Dual Sup -> ISSU
  - Single Sup -> Reload

- Patching is for operationally impacting bugs without a ***workaround.***
  - ***Cannot patch to next release.***
- Patching is done in default/admin VDC and applies to all VDCs.
  - ***Patching is not available per-VDC.***
- ISSU will work with all, or a subset of patches applied.
  - ***You don't need to apply all patches.***
- Some SMUs may only have a single fix, others may have multiple packaged.

Patching Highlights

- SMUs are TAC supported.
- SMUs are synched to standby supervisor.
- On Sup replacement, patch(es) will be synchronized.
- SMUs are not for feature implementation. A SMU cannot change the configuration.

Operational Best Practices

Hardware Maintenance

NX-OS >= 6.1: Parallel EPLD Upgrades!

Electronic Programmable Logic Device Upgrade Example

The following example upgrades the EPLD image for module 1. The EPLD image should be local when the upgrade is performed.

This procedure is typically not required during an NX-OS upgrade.

```
n7000# install module 1 epld bootflash:n7000-s1-epld.4.0.1.img

EPLD image file , built on Mon Mar 31 10:31:48 2008

EPLD                                  Curr Ver    New Ver
-----------------------------------------------------------
Power Manager                          4.1         5.3
IO                                     2.6         2.10
Forwarding Engine                      1.4         1.6
WARNING: Upgrade process could take up to 30 minutes.
Module could be powered down and up.

Module 1 will be powered down now!!
Do you want to continue (y/n) ?  [n] y

Module 1 EPLD upgrade is successful.
```

EPLD upgrades are intrusive and may take up to 30 minutes per module!

The "install" command highlights the EPLD version differences

The user is prompted to continue

Hardware Maintenance

- **Scenario: Line Card Hardware Upgrade or Replacement**
  - Power down line card prior to removal.

    ```
    Nexus# out-of-service module <module-number>
    ```
  - Hitless with VPC provided sufficient bandwidth and port-channel distribution.
  - Mixed line card deployment between VPC peers is not supported.

NOTE: Evaluate the VDC interface assignments to verify which VDCs will experience a service impact.

Line card support matrix:

http://www.cisco.com/c/dam/en/us/td/docs/switches/datacenter/nexus7000/sw/matrix/technical/reference/Module_Comparis

However:
#conf t
#vpc domain <*id*>
#bypass module-check
Not BP, only corner case, change window.

Hardware Maintenance

Scenario: Chassis Hardware Upgrade

- Bring switch being replaced into Graceful Insertion and Removal mode or manually isolate prior to power down.
- Gas up your fork lift.

Scenario: Fabric Module Hardware Upgrade

- Don't oversubscribe the fabric when replacing fabric modules.
- `n7000# show hardware fabric-utilization`

Scenario: Power Supply Hardware Upgrade

- Online Insertion and Removal (OIR) is supported.
- Be mindful of power budget.

NX-OS Graceful Insertion and Removal

Protocol Isolation in Nexus

- IGPs

	OSPF	IS-IS	EIGRP

Option 1 Isolate  Recommended	**Advertise as Stub Router, LSInfinity**
	`max-metric router-lsa [ on-startup [ seconds | wait-for bgp tag ]]`
		Advertise as with LSP Database Overload Bit set
		`set-overload-bit {always | on-startup {seconds | wait-for bgp as-number}} [suppress [interlevel | external]]`
			Manipulate Metrics
			`interface e1/1`
			`ip delay eigrp instance-tag seconds`

Option 2 Shutdown Protocol	`router ospf 1` `  shutdown`	`router isis 1` `  shutdown`	`router eigrp 1` `  shutdown`

Option 3 Interface Disable	`interface e1/1`
	`  ip ospf shutdown`
		`interface e1/1`
		`  isis shutdown`
			`interface e1/1`
			`  ip eigrp 1 \`
			`  shutdown`

Protocol Isolation in Nexus

BGP

Option 1: Advertise prefixes with longer AS path / higher local-preference


```
switch(config)# route-map prepend
switch(config-route-map)# match as-path 1
switch(config-route-map)# set as-path prepend last-as 3
switch(config)# router bgp 65000
switch(config-router)# neighbor 192.168.10.2 remote-as 20
switch(config-router-neighbor)# address-family ipv4 unicast
switch(config-router-neighbor-af)# route-map prepend out
```

Option 2: Shutdown BGP (Process)

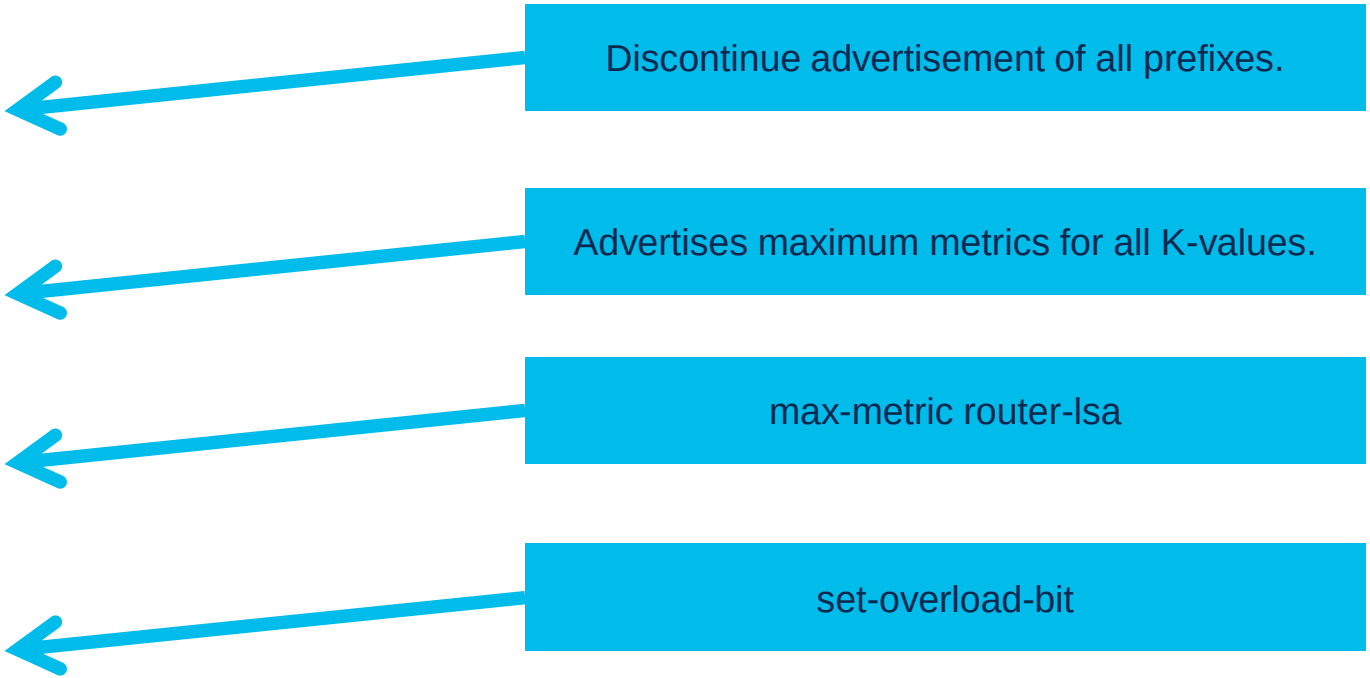
```
router bgp 65010
shutdown
```

NOTE: This is a not a graceful shutdown such as you would achieve with GSHUT / RFC 6198.

Nexus 9k/7k/6k

Graceful Removal

```
router bgp 33
   isolate
 router eigrp 1
   isolate
 router ospf 1
   isolate
 router isis 1
   isolate
```



Nexus 9k/7k/6k/5k

Graceful Insertion

- Move the switch from Maintenance mode to Normal mode.
- Control plane maintained throughout isolation of the switch.
- Protocols advertise routes only after it is installed in hardware.

```
N9372(config)# no system
mode maintenance
Following configuration
will be applied:
  router bgp 33
    no isolate
  router eigrp 1
    no isolate
  router ospf 1
    no isolate
  router isis 1
    no isolate
```

Protocol Isolation in Nexus

- All Protocols

Option 4: System Interface Shutdown

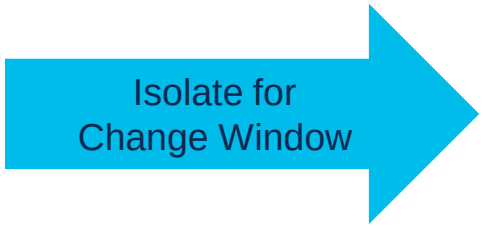
```
system interface shutdown
```

For many, this is good enough.
And, easy!

Graceful Insertion and Removal

```
feature ospf

feature vpc
```



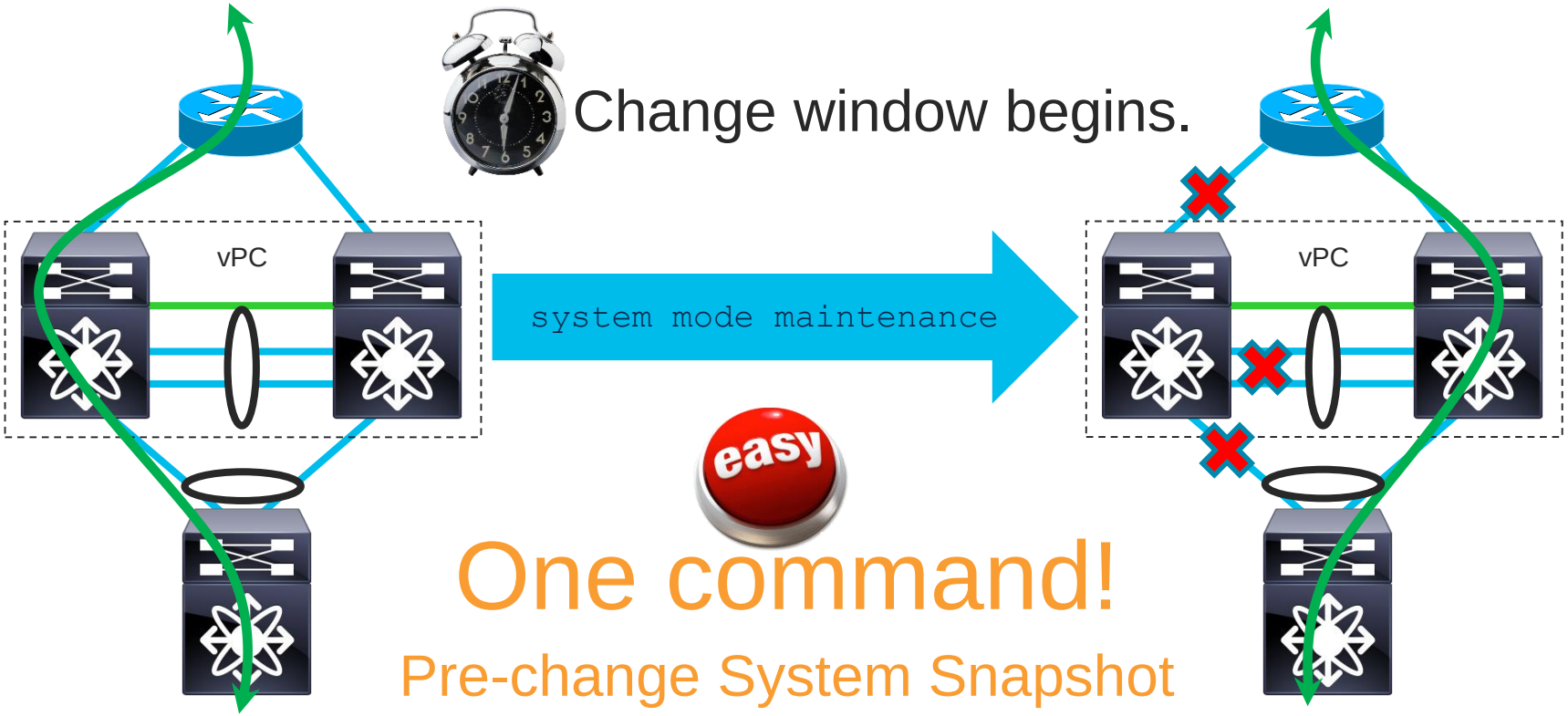
OSPF:
max-metric router-lsa

VPC:
shutdown

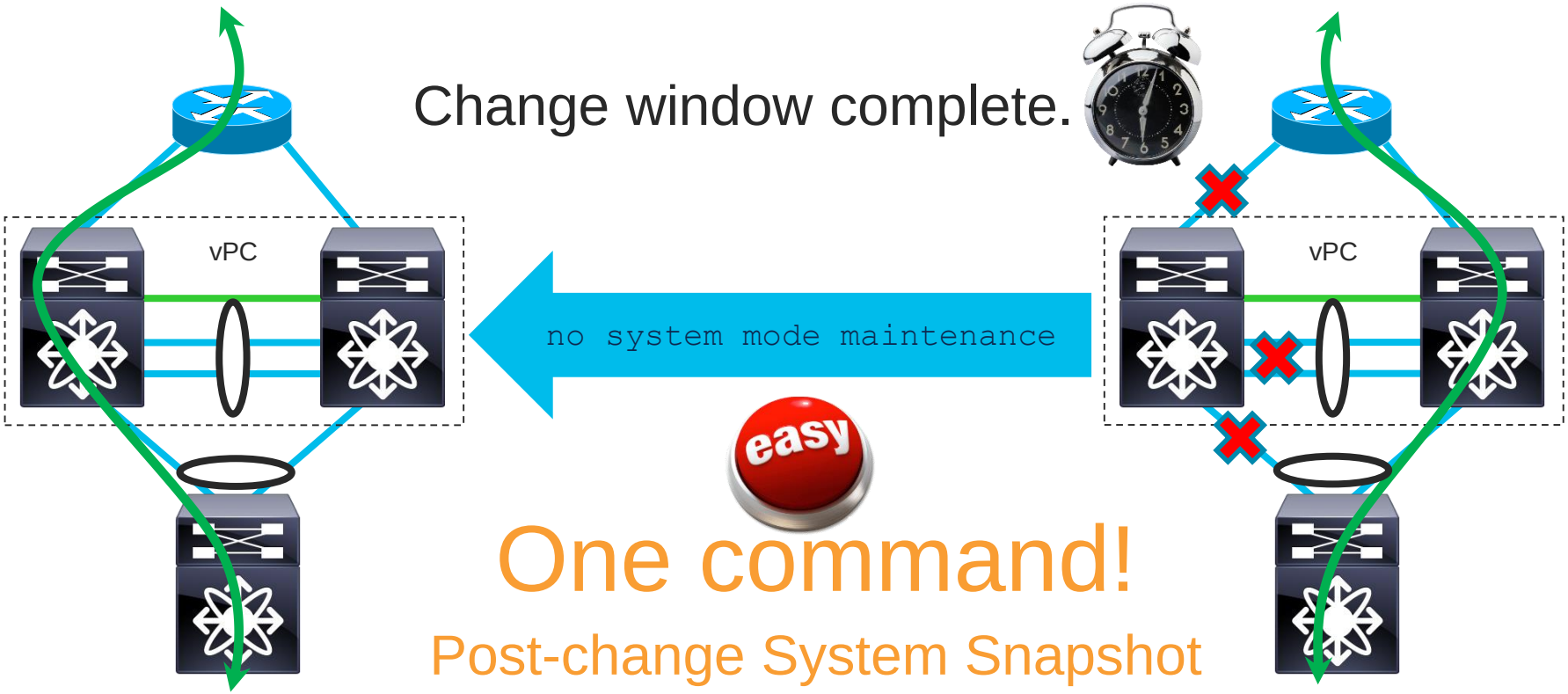
Scripting takes time.

It'd be nice to automate this…

Graceful Insertion and Removal



Graceful Insertion and Removal



One command!

Post-change System Snapshot

Graceful Insertion and Removal

- Flexible framework providing a comprehensive, systemic method to isolate a node.
- Configuration profile foundation in NX-OS
- Initial support for:
  - vPC/vPC+
  - ISIS
  - OSPF
  - EIGRP
  - BGP
  - Interface
- Per VDC on Nexus 7x00

Configuration Profiles

- Maintenance-mode profile is applied when entering GIR mode,
- Normal-mode profile is applied when GIR mode is exited.

Automatic Profiles	Manual Profiles
• Generated by default • Parses configuration to determine changes going into and out of GIR • Changes based on base protocol configuration settings. • Use: Maintenance Windows	• User created profile for maintenance-mode and normal-mode • Flexible selection of protocols for isolation • Use: maintenance windows and isolation during troubleshooting using preconfigured scripts.

Enabling Graceful Insertion and Removal
Automatic Profile Generation

```
N7K-1-Core# show system mode
System Mode : Normal
N7K-1-Core# config
Enter configuration commands, one per line.  End with
CNTL/Z.
N7K-1-Core(config)# system mode maintenance

BGP is not enabled, nothing to be done

EIGRP is not enabled, nothing to be done

OSPF is up..... will be shutdown
        OSPF TAG = 100, VRF = default
                config terminal
                router ospf 100
                shutdown
                end

OSPFv3 is not enabled, nothing to be done

ISIS is not enabled, nothing to be done

vPC is not enabled, nothing to be done

Interfaces will be shutdown
Do you want to continue (y/n)? [n] y

Generating maintenance-mode profile
Progressing.....................Done.

System mode operation completed successfully

N7K-1-Core# show system mode
System Mode : Maintenance
N7K-1-Core#
```

NOTE: Custom profile generation requires “dont-generate-profile”.

Enabling Graceful Insertion and Removal Custom Profile Generation

```
config-profile maintenance-mode type admin
router bgp 65001
 isolate
 sleep instance 1 10
router ospf 100
  isolate
  sleep instance 3 20
  vpc domain 20
   shutdown
  system interface shutdown exclude fex-fabric
```

```
config-profile normal-mode type admin
router bgp 65001
 no isolate
 sleep instance 1 10
router ospf 100
  no isolate
  sleep instance 3 20
  vpc domain 20
   no shutdown
  no system interface shutdown
```

- By default, GIR Mode will automatically generate profiles.
- CLI to disable automatic profile generation: `dont-generate-profile`
- If you enter GIR mode with automatic profile, it will overwrite your custom profile.

Graceful Insertion and Removal Mode for Unplanned Outages

system mode maintenance on-reload reset-reason *reason*

HW_ERROR-Hardware error,

SVC_FAILURE-Critical service failure,

KERN_FAILURE-Kernel panic,

WDOG_TIMEOUT-Watchdog timeout,

FATAL_ERROR-Fatal error,

MANUAL_RELOAD---Manual reload,

MATCH_ANY-Any of the above reasons,

ANY_OTHER-Any reload reason not specified above.

Nexus GIR Snapshots

- Used before and after a GIR mode to compare pre/post change operation.
- Snapshots are automatically generated when entering GIR mode.

```
switch# snapshot create snap1 For testing
Executing show interface... Done
Executing show bgp sessions vrf all... Done
Executing show ip eigrp topology summary... Done
Executing show vpc... Done
Executing show ip ospf vrf all... Done
Feature 'ospfv3' not enabled, skipping...
Snapshot 'snap1' created
Switch#
```

Nexus GIR Snapshots Comparison

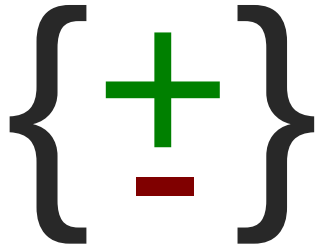
```
Nexus# sh snapshots compare before_maintenance after_maintenance
================================================================================
Feature Tag           before_maintenance           after_maintenance
================================================================================

[bgp]
--------------------------------------------------------------------------------

[neighbor-id:100.120.1.221]
connectionsdropped 2                               **3**
lastflap              P1DT21H5M12S                 **P1DT21H25M47S**
lastread              P1DT21H25M12S                **PT0S**
lastwrite             P1DT21H25M14S                **PT0S**
state                 Established                  **Idle**
localport             52737                        **0**
remoteport            179                          **0**
notificationssent     2                            **3**
<...>
```

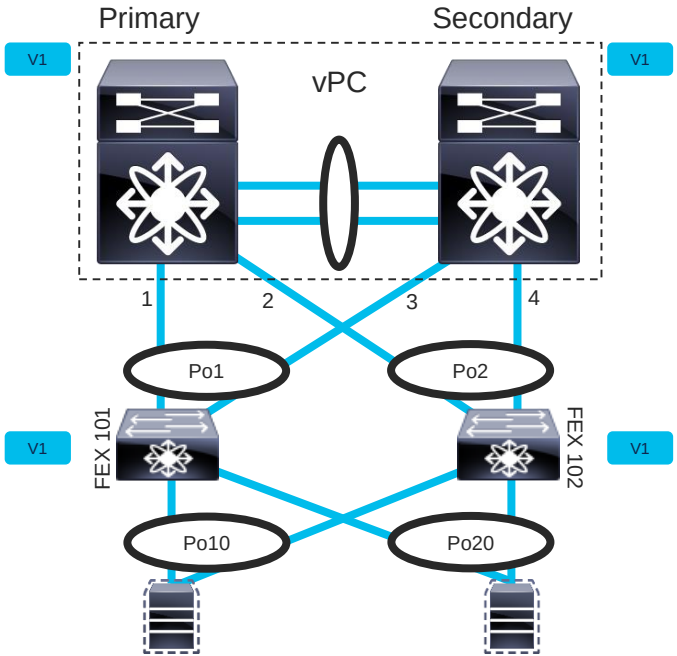
```
switch# show snapshots compare snapshot1 snapshot2 ipv4routes
metric                snapshot1    snapshot2     changed
# of routes           33           3             *
# of adjacencies      10           4             *

Prefix          Changed Attribute
------          -----------------
23.0.0.0/8      not in snapshot2
10.10.10.1/32   not in snapshot2
21.1.2.3/8      adjacency index has changed from 29 (snapshot1) to 38
(snapshot2)
```



Nexus 5k Scenario: Dual-homed FEX w/ VPC

Software Upgrade



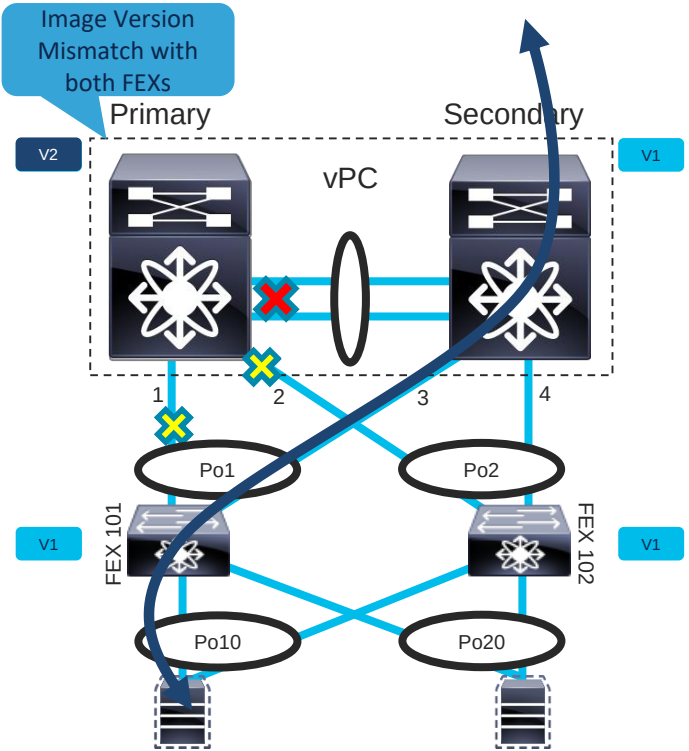
Overview

- Highly Redundant Design
- Dual-attached FEX
- Dual-attached Hosts

How do we upgrade this environment with minimal disruption?

Nexus 5k Scenario: Dual-homed FEX w/ VPC

Software Upgrade



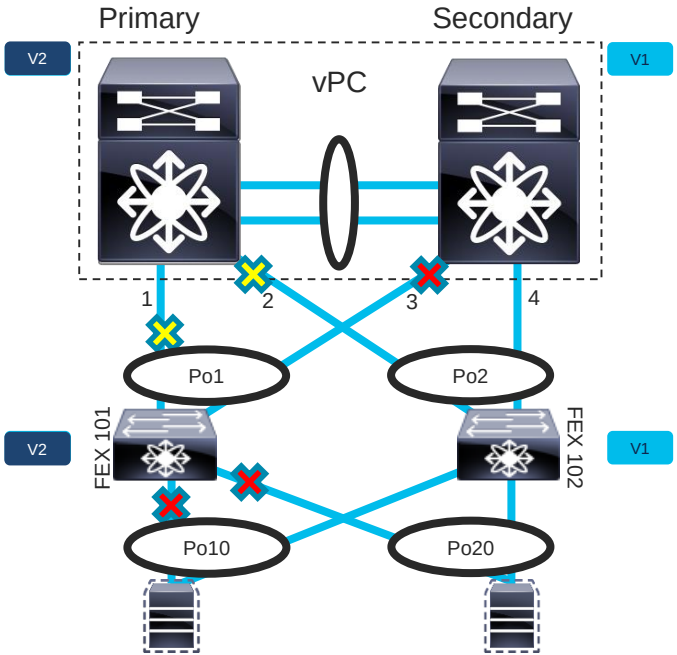
- Enter GIR Mode on N5k1
  Traffic flow through N5k2
- Upgrade N5k1
- Exit GIR on N5k1

IF Down

IF Up, No Forwarding

Nexus 5k Scenario: Dual-homed FEX w/ VPC

Software Upgrade



- Manually shut down IF3 on N5k2

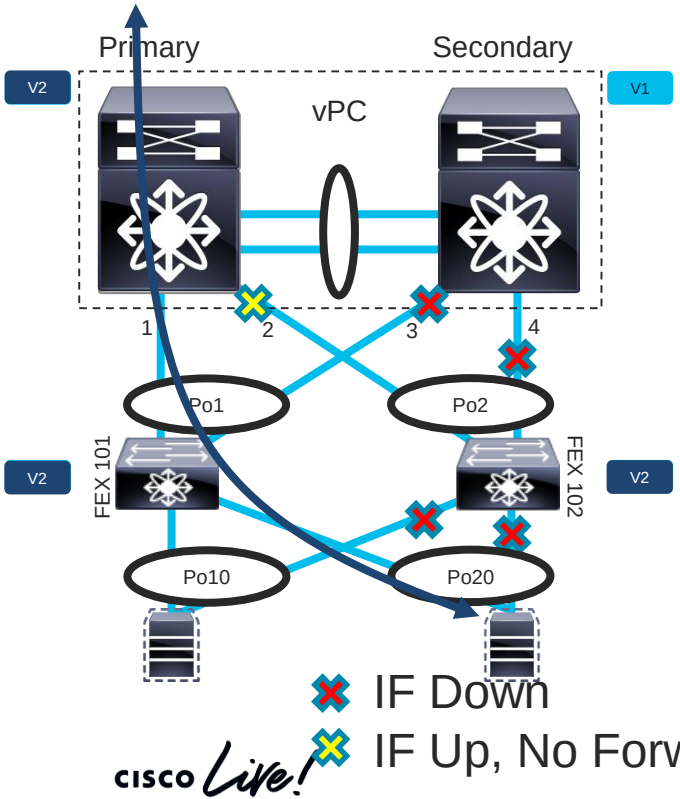
FEX 101 goes offline.
FEX 101 HIFs go down.

FEX 101 starts pairing process with N5k1.

FEX 101 upgrades to V2.

Nexus 5k Scenario: Dual-homed FEX w/ VPC

Software Upgrade



- Manually shut down IF4 on N5k2

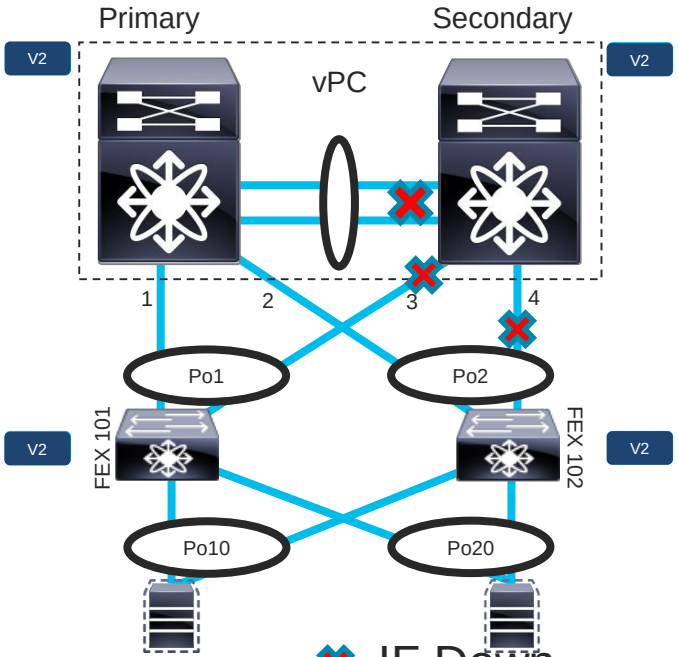
FEX 102 goes offline.
FEX 102 HIFs go down.

FEX 102 starts pairing process with N5k1.

FEX 102 upgrades to V2.

Nexus 5k Scenario: Dual-homed FEX w/ VPC

Software Upgrade



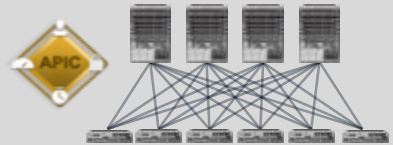
- Enter GIR Mode on N5k2
  IF 3 & 4 Still Admin Down
- Upgrade N5k2
- Exit GIR on N5k2
- Manual Up of IF 3 & 4

Environment upgrade completed with minimal traffic disruption.

ACI Operational Practices

SDN ‘with’ FCAPS ‘and’ Automation

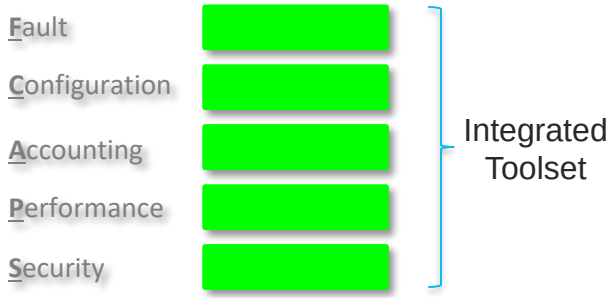
Application Centric Infrastructure



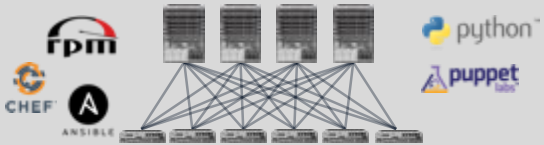
Turnkey integrated solution with security, centralized management, compliance and scale

Automated application centric-policy model with embedded security

Broad and deep ecosystem



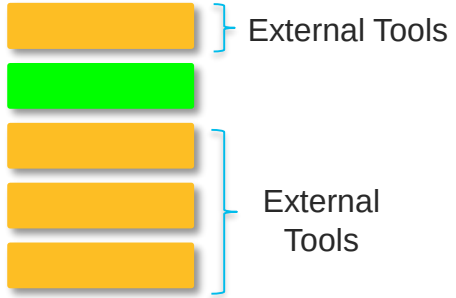
Programmable Network



Modern NX-OS with enhanced NX-APIs

DevOps toolset used for Network Management
(Puppet, Chef, Ansible etc.)

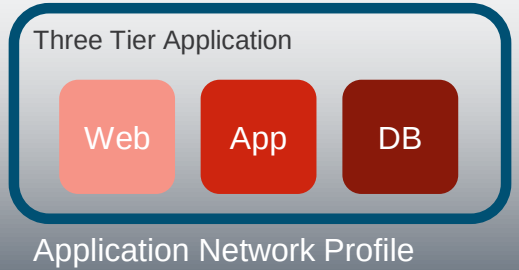
Custom Script based Operations and Workflows



Application Centric Infrastructure (ACI)

Rapid Deployment of Applications onto
Networks with Scale, Security and Full Visibility

Nexus 9k



App Centric Policy

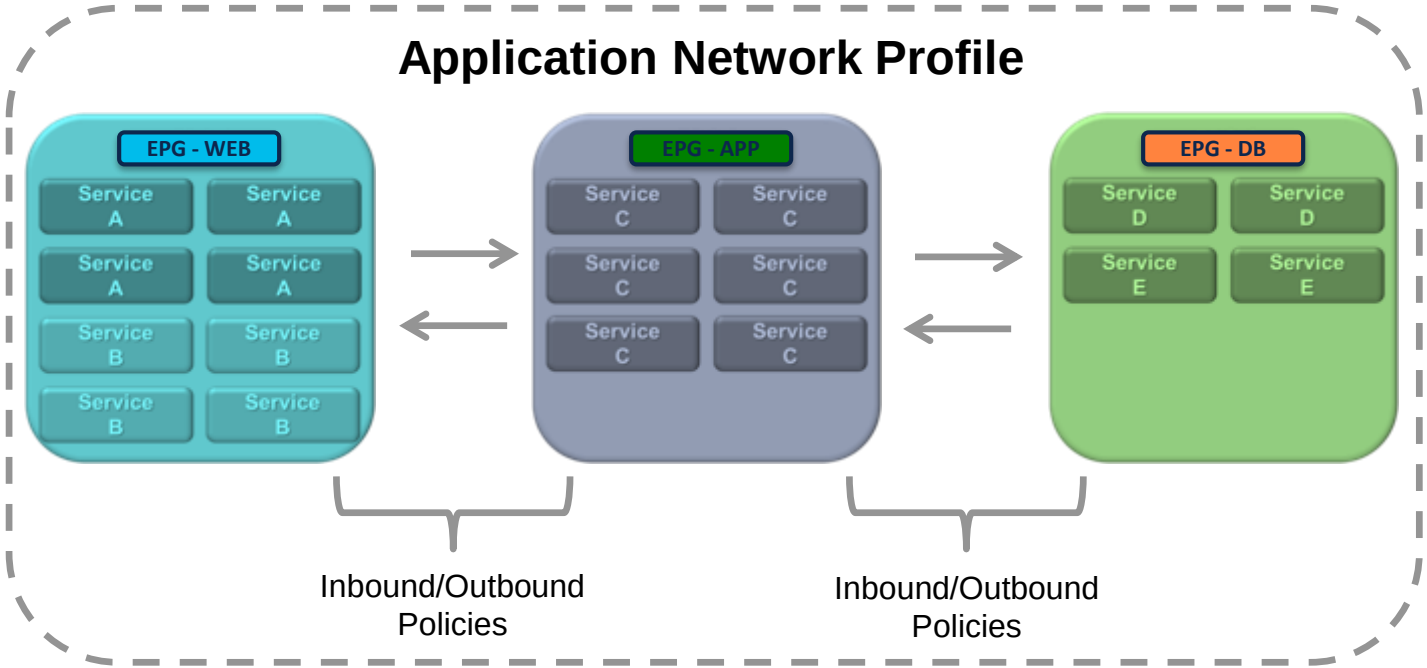
APIC

ACI

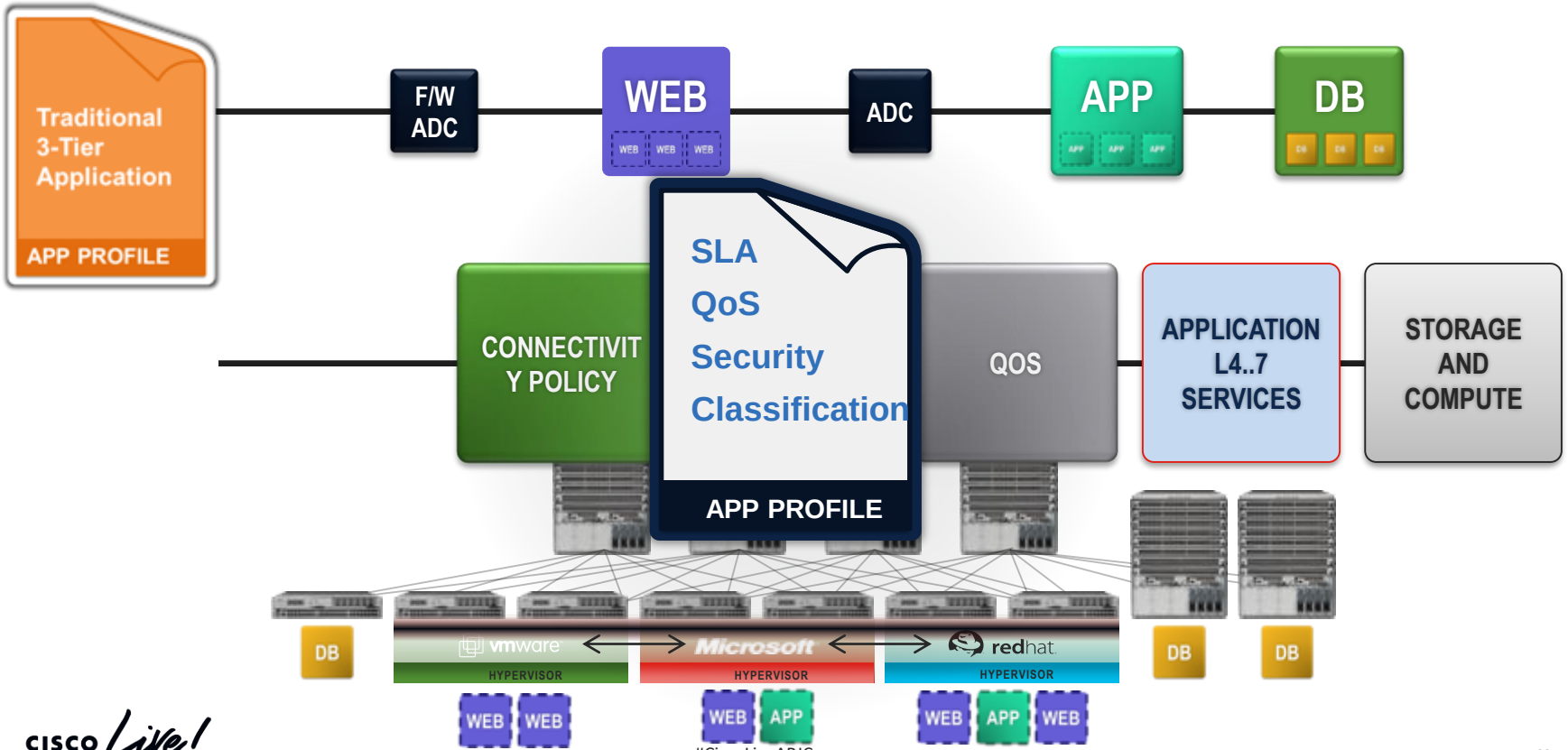
Application Network Profiles (ANP) – What's That?

Application Network profiles are a group of EPGs **and** the policies that define the communication between them.

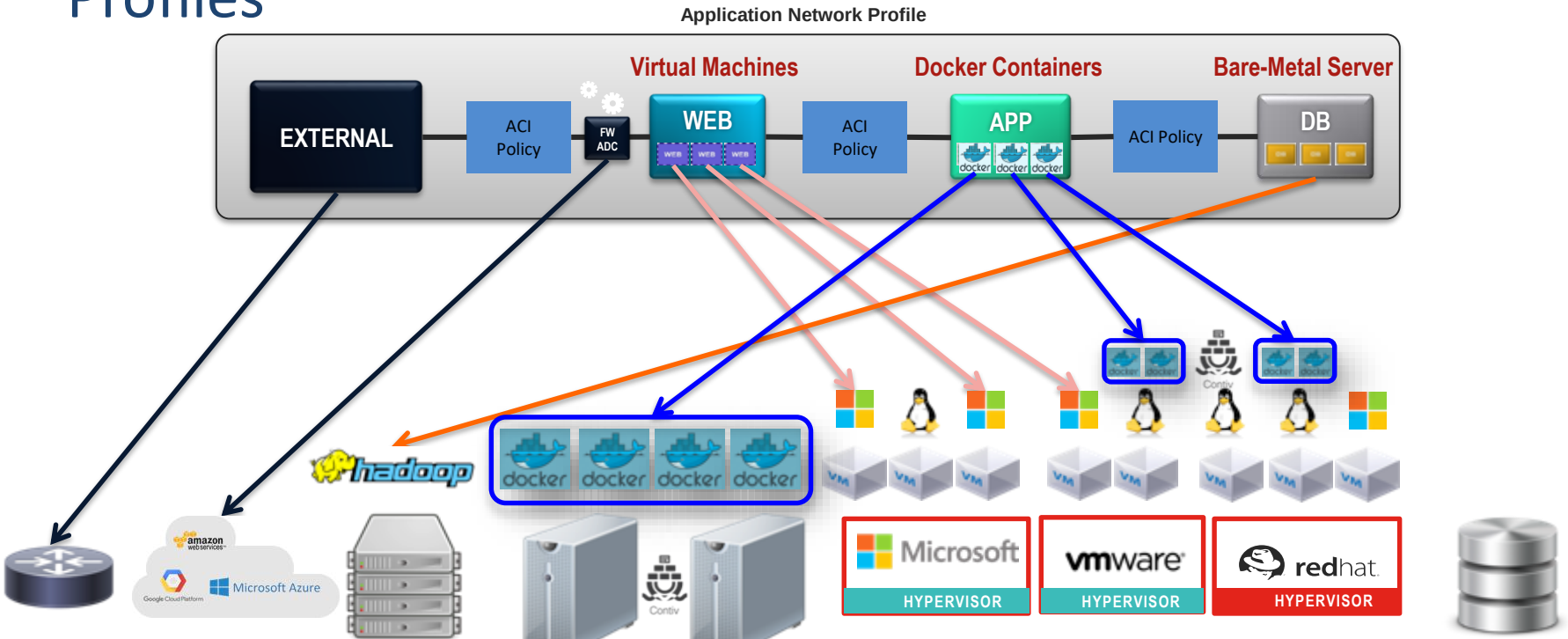
=



Application Network Profiles (ANP) & ACI: how it works?

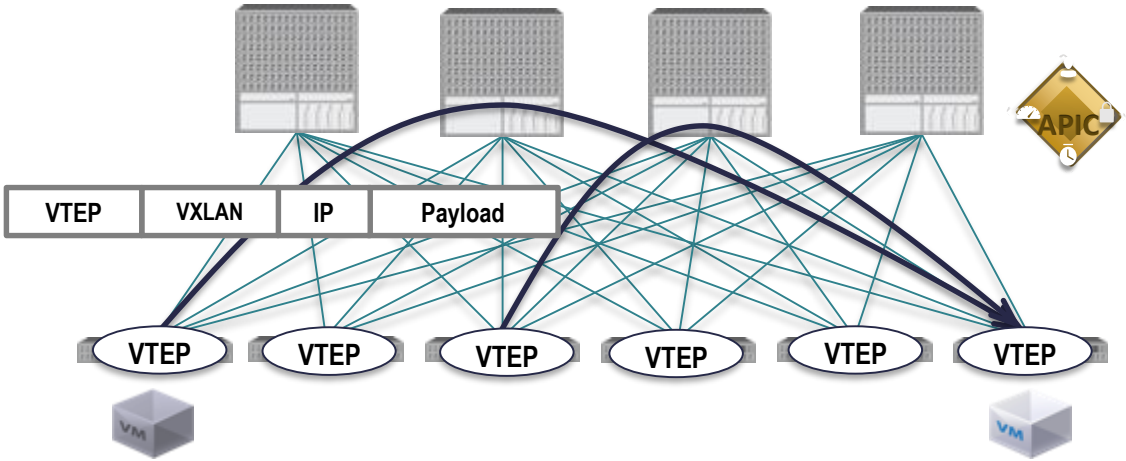


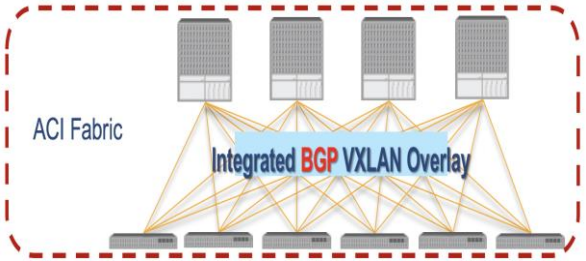
Abstracting / Mapping via ACI's Application Network Profiles



Did you notice ? There is no network device represented here ☺

ACI Fabric – Integrated Overlay
Decoupled Identity, Location & Policy



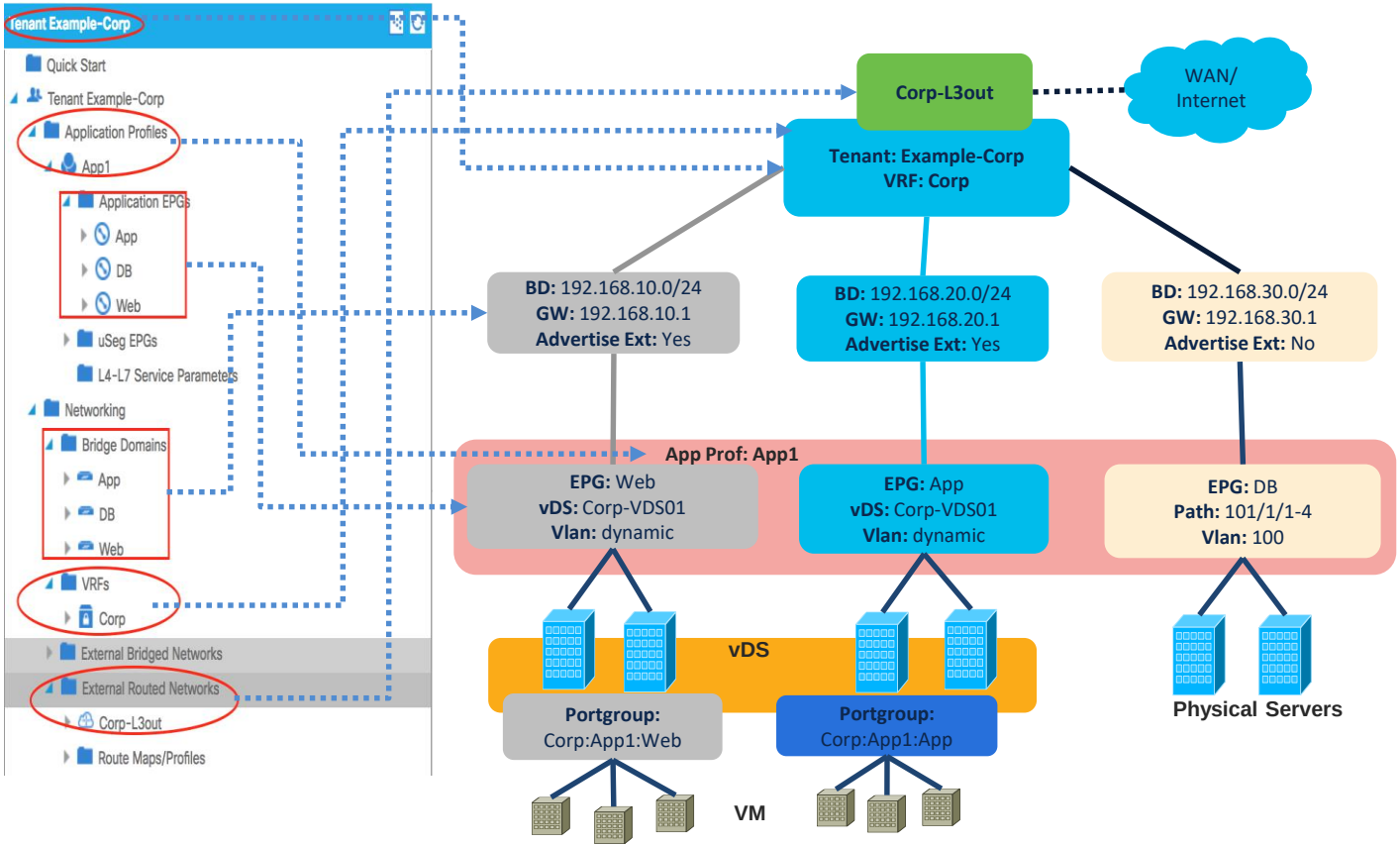


- ACI Fabric decouples the tenant end-point address, it's "identifier", from the location of that end-point which is defined by it's "locator" or VTEP address
- Forwarding within the Fabric is between VTEPs (ACI VXLAN tunnel endpoints) and leverages an extender VXLAN header format referred to as the ACI VXLAN policy header
- The mapping of the internal tenant MAC or IP address to location is performed by the VTEP using a distributed mapping database
- The control plane managed by COOP (Council of Oracle Protocols)

ACI Network Centric Deployment

Network configuration

- VRF CORP …. vrf configuration
- Interface VLAN 100 192.168.10.0/24), VIP 192.168.10.1, VRF corp
- Trunk the switch ports with respective vlans
- VMware port Group Assignment
- Routing Configuration for subnets



Use Case For Complex Deployment Made Simple
Cisco AS DAFE- Deploy ACI from Excel

1) Fill in the excel sheet

node_provisioning | switch_profile | vpc_domain | fex_provisioning | vlan_pool | phys_domain | vmm_domain | aaep | interface_policy | interface_policies | int_pol_group | interface_profile

2) Select the tasks (ACI Objects) you want to deploy

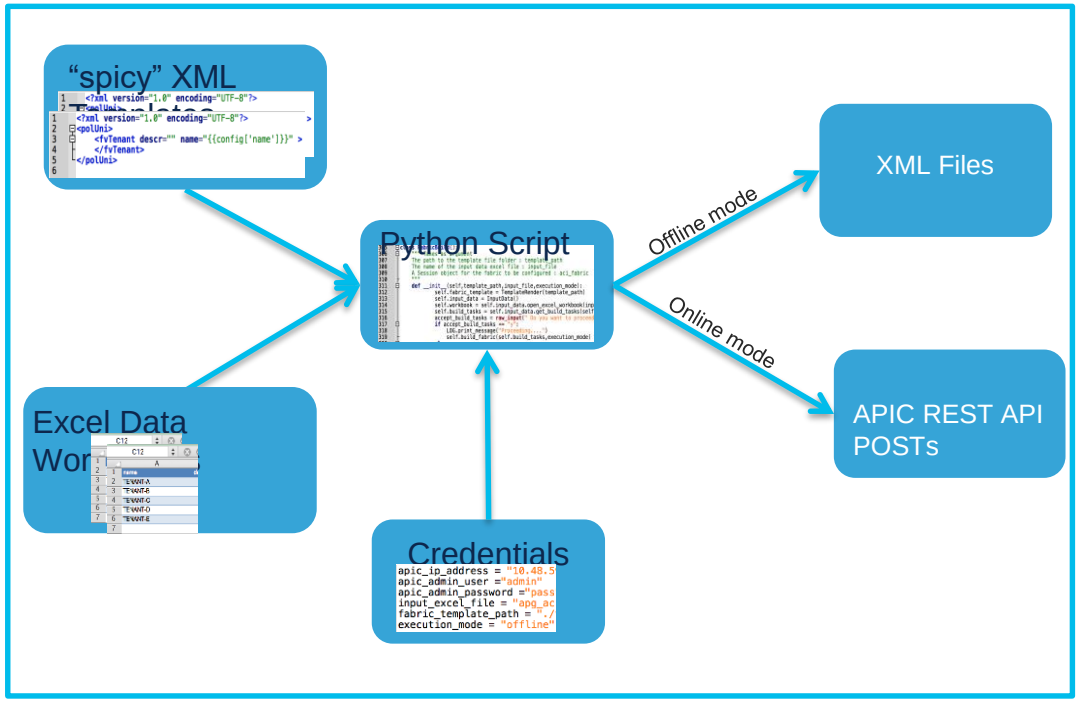
include	description	input_worksheet
yes	Node Registration	node_provisioning
yes	Node Addressing	node_provisioning
yes	Switch Profile	switch_profile
yes	vPC Domain	vpc_domain
no	Fex Provisioning	fex_provisioning
yes	Vlan Pool	vlan_pool
no	Physical Domain	phys_domain
no	VMM Domain	vmm_domain
no	AAEP	aaep
no	Interface Policies	interface_policies
no	Interface Policy Group	int_pol_group
no	Interface Profile	interface_profile
no	Fex Interface Profile	fex_interface_profile
no	Associate Interface Profile to Switch Profile	sw_prof_int_prof

3) Edit The credentials sheet

```
apic_ip_address = ""    # Do not delete leave empty if you use offline mode
apic_admin_user =""     # Do not delete leave empty if you use offline mode
apic_admin_password =""# Do not delete leave empty if you use offline mode
input_excel_file = ""
fabric_template_path = "./templates/"
execution_mode = "offline"    # Possible choices are    offline  or online
```

4) Run the Script

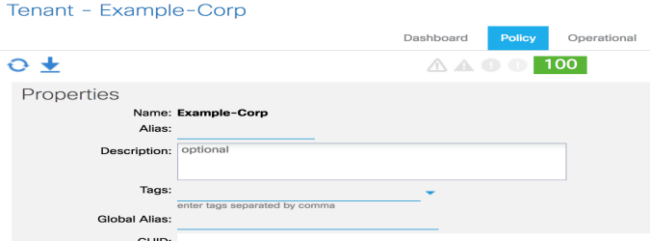
Automation-srvr$python aci_deploy_fabric_from_excel.py >> output.xml



Operational Best Practices

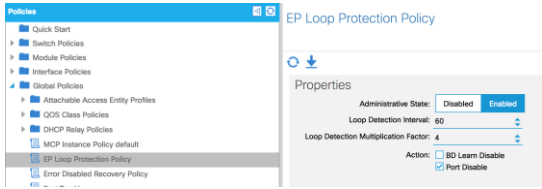
- **MO Naming Convention**
  - Develop and plan the MO(Managed Objects) Naming Convention according to Organizations best Practice

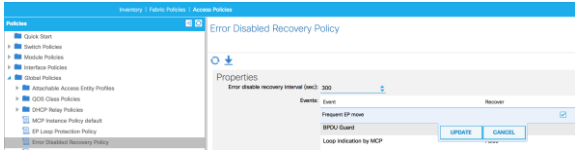
- **Tags and Aliases**
  - Workaround to Rename Objects
  - Objects can be grouped to make query easier
  - Tags/Aliases have no functional impact- Where as Labels have



- **AAA Fallback to Local Auth**
  - Fallback domain should be set to local to avoid lockout

- **EP Loop Prevention**



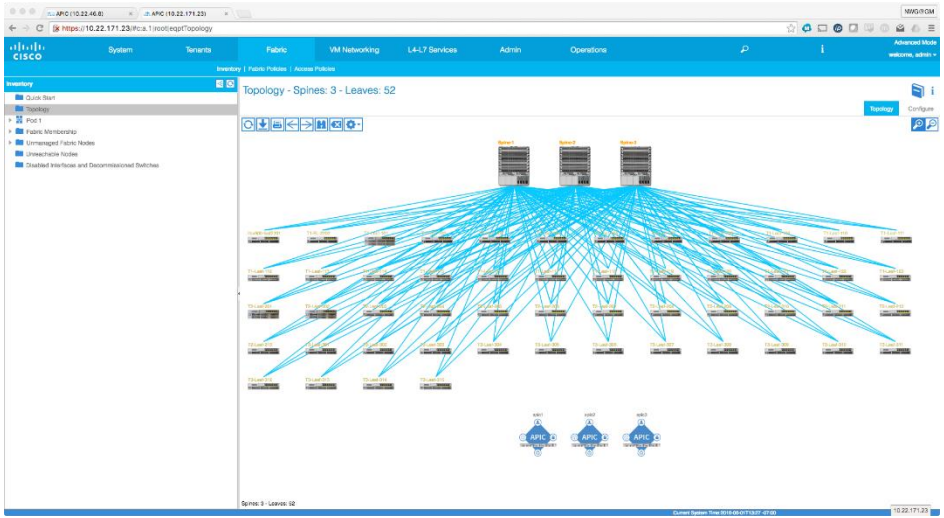


- **BD Level Configuration**
  - Limit IP Learning to Subnet

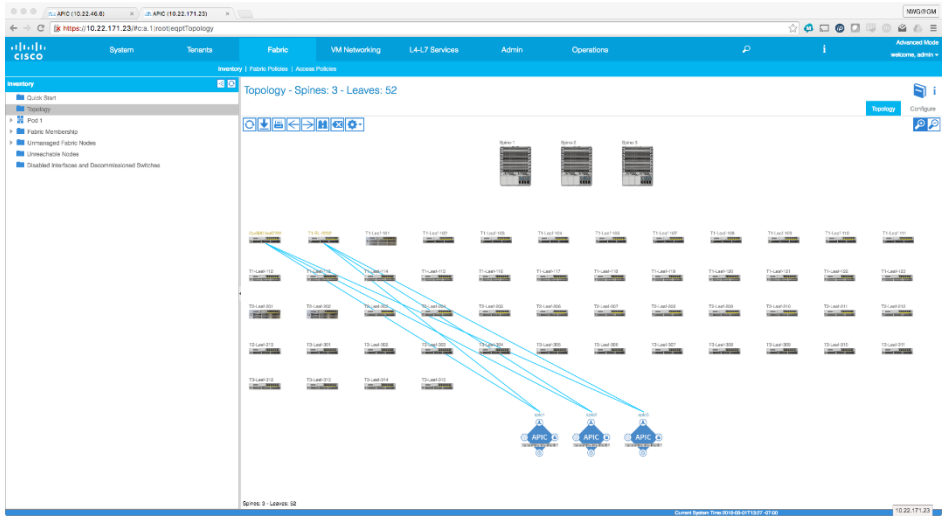
- **Fabric Wide Configuration**
  - IP Aging Policy
  - Disable Remote EP Learning – On Border Leaf
  - Enforce Subnet Check

Cisco ACI Fabric

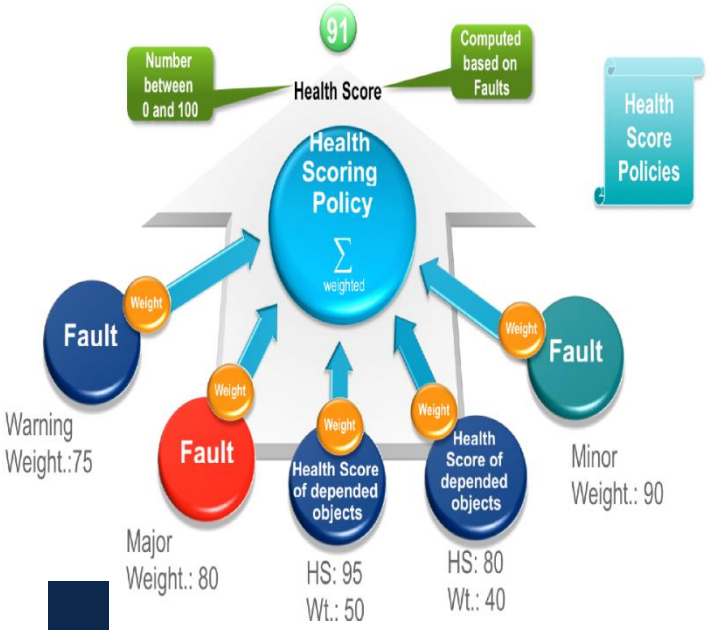
Fabric View



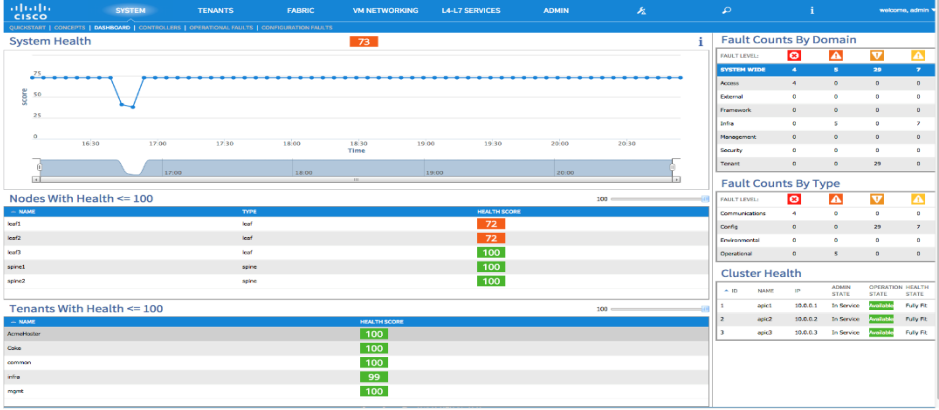
Controller Connectivity



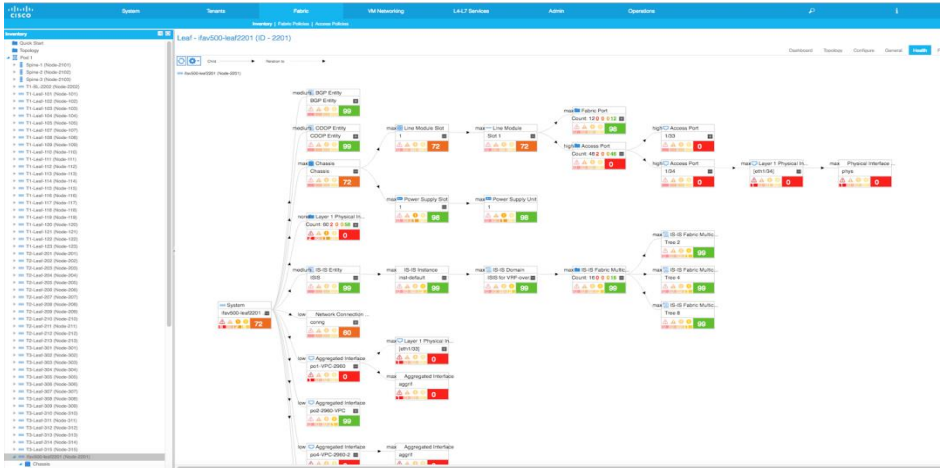
Health Score



Aggregated View



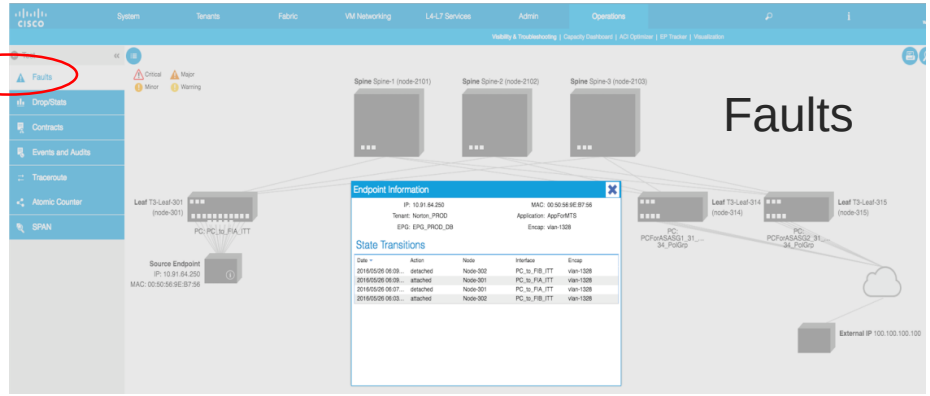
Fabric Topology View

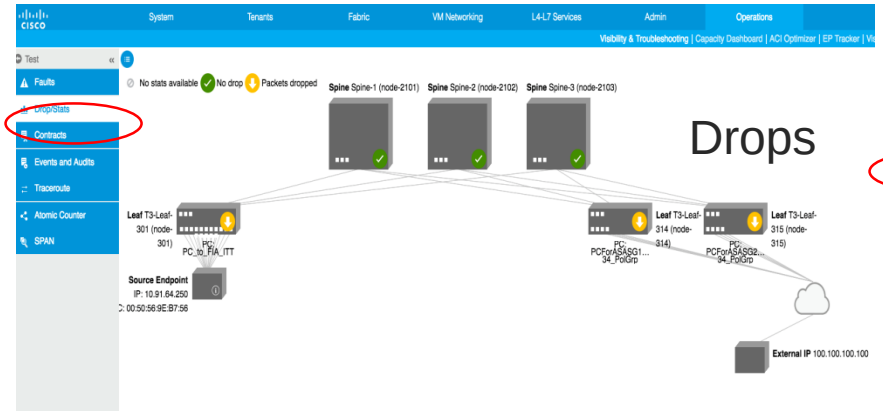


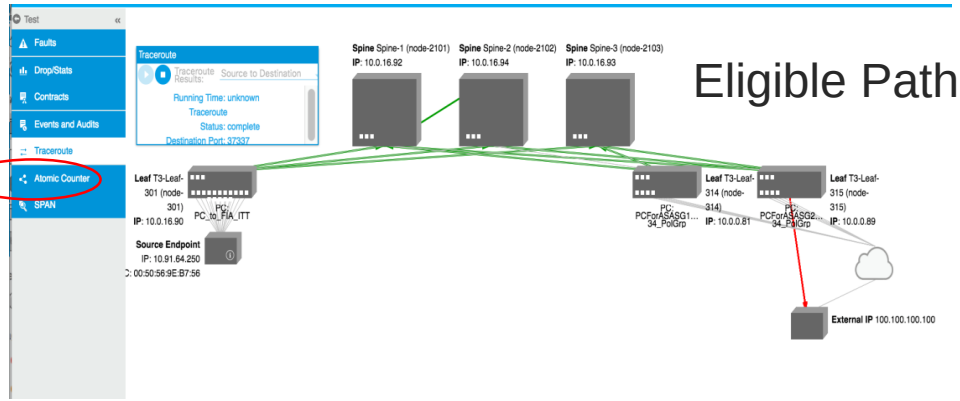
Aggregation of system-wide health, including pod health scores, tenant health scores, system fault counts domain and type and the APIC cluster health state.

Troubleshoot a Flow

Use ACI Inbuilt Visibility Engine

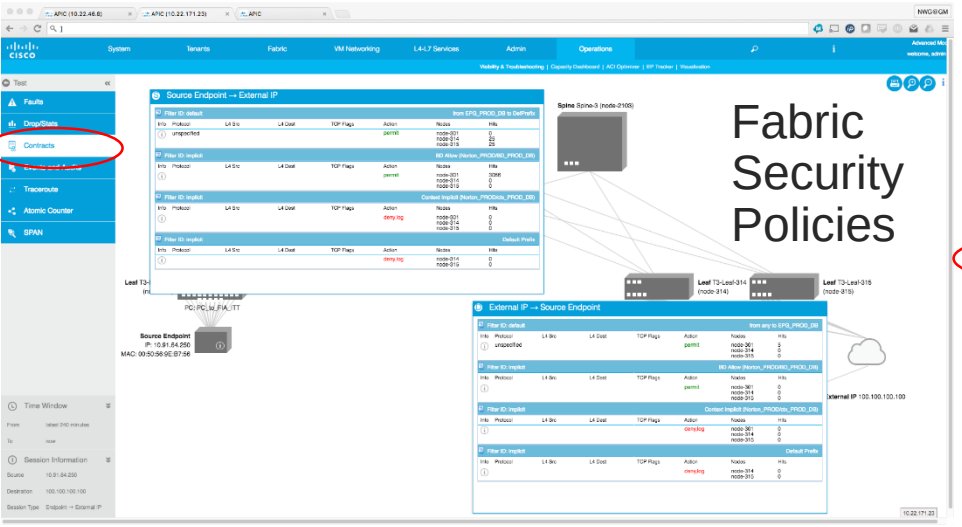


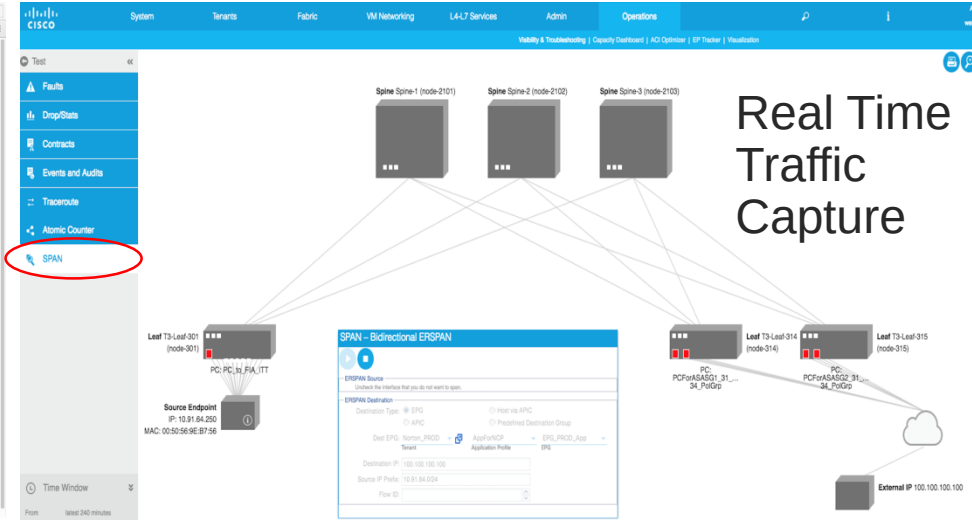




Troubleshoot a Flow

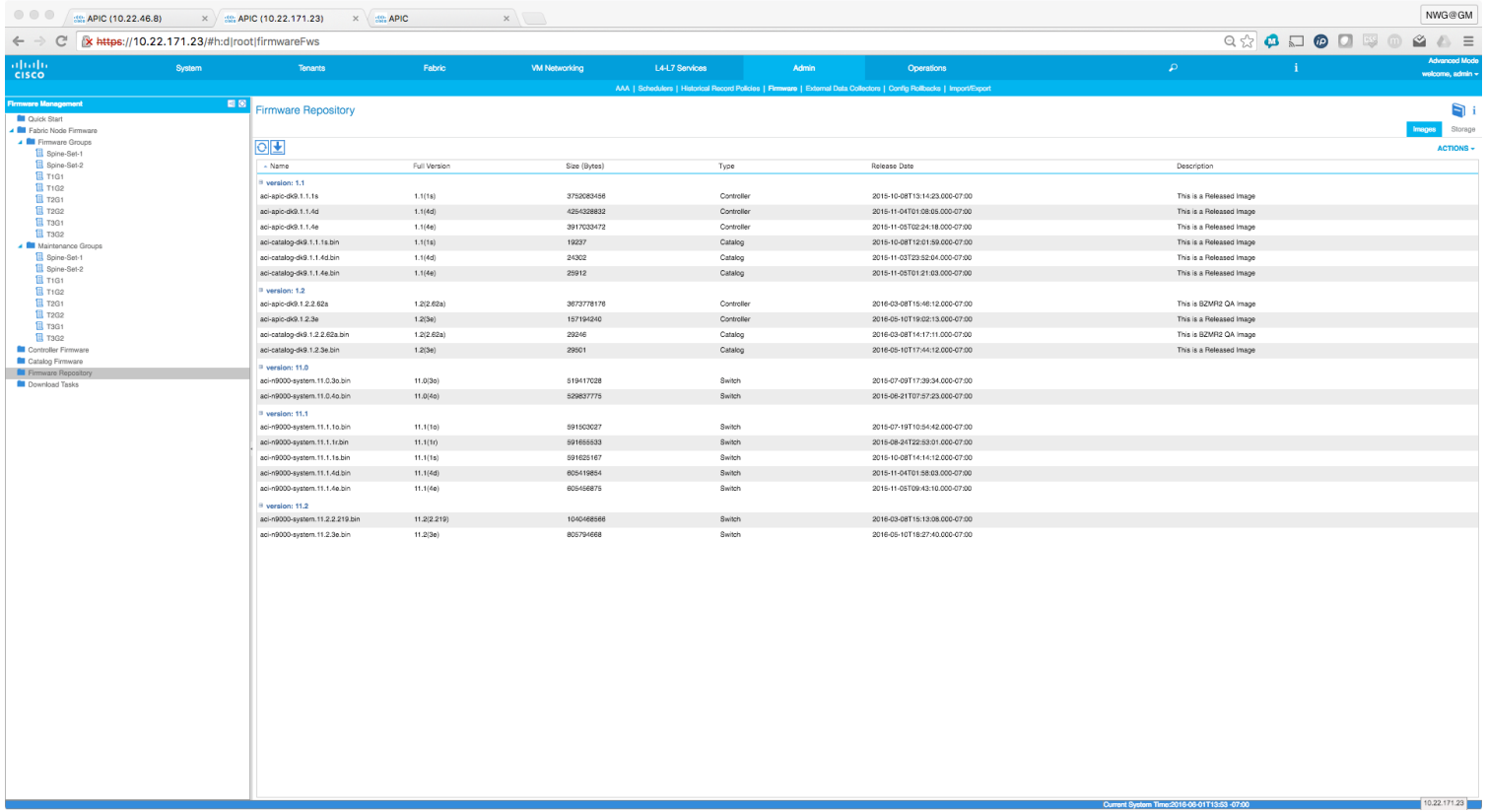
Use ACI Inbuilt Visibility Engine





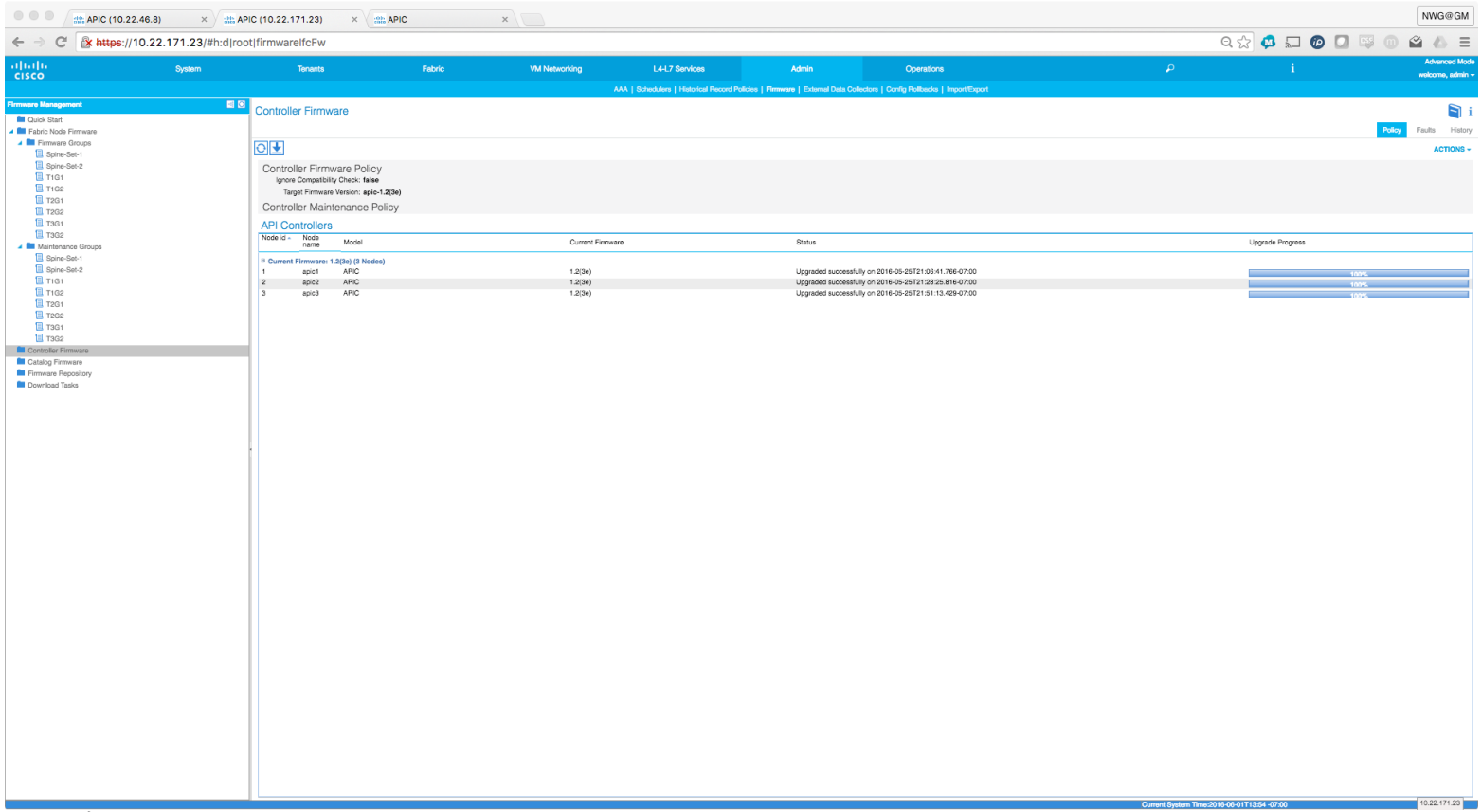
Maintenance Upgrade #1

Download the Release on the APIC



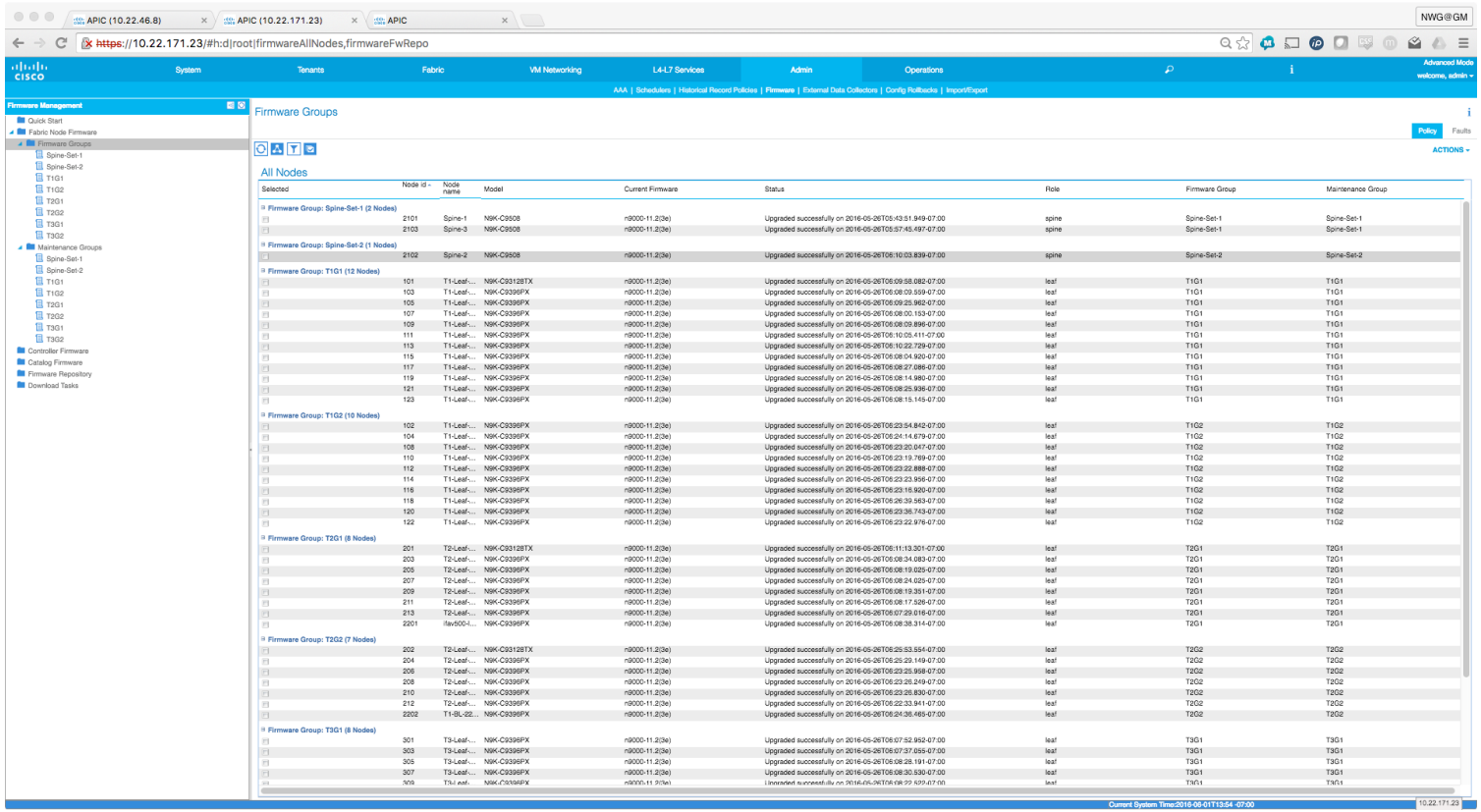
Maintenance Upgrade #2

Upgrade APIC



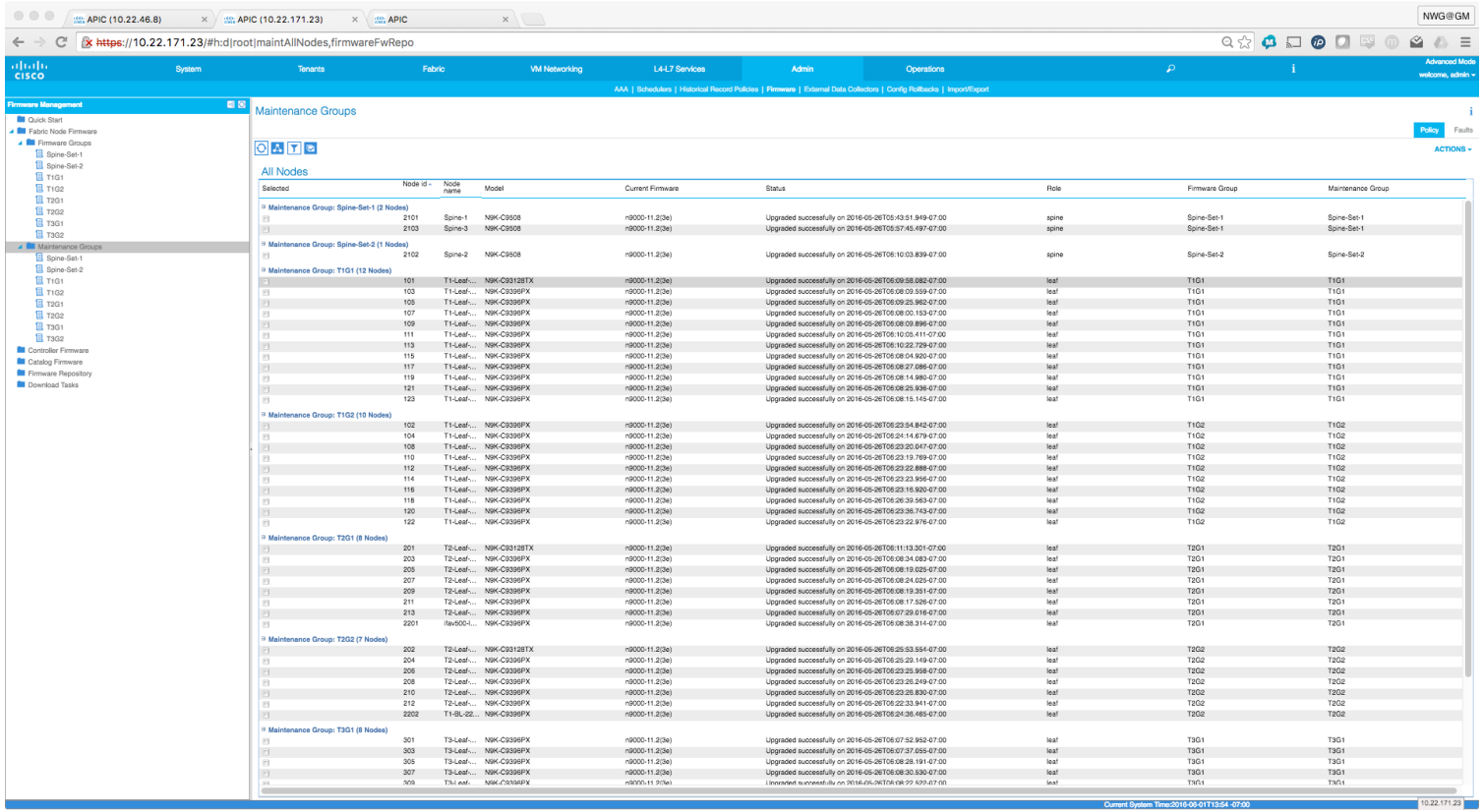
Maintenance Upgrade #3

Create Groups



Maintenance Upgrade #4

Upgrade the Maintenance Groups



Capacity Dashboard
View the Capacity of Data Center Fabric

Cisco ACI Deployment Lifecycle

Proactive

- Faults
- Events
- **Health Score**
- Atomic Counter
- Contract deny logs
- Statistics
- **Capacity Dashboard**

Preemptive

- Monitor
- Image Management
- Config Export / Import
- Fabric Inventory
- Show Usage
- Configuration Rollback

Reactive

- Troubleshoot
- Manage
  - Audit Logs
  - iPing
  - iTraceroute
  - Endpoint Tracker
  - ERSPAN
  - Traffic Map
  - On Demand Counter View
  - CLI option

Recommended Live sessions for ACI :
BRKACI-2210, LTRACI-2143

DCNM

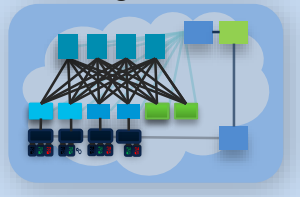
DCNM : Functionality

Health Monitoring, Inventory & Diagnostics

Configuration

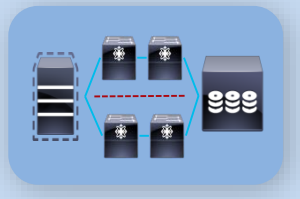
Automation

Trend Analysis



Visualization & Troubleshooting

Alert/Notifications

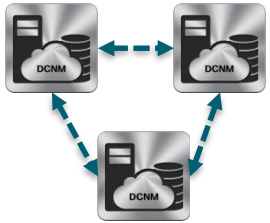


Storage Management

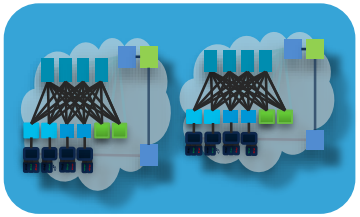
DCNM Infrastructure & LAN Fabric Updates

Turn-Key Virtual Appliance Enterprise DB,H/A & Large Scale

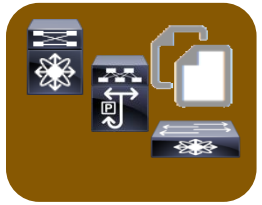
HTML5 GUI Topology Driven



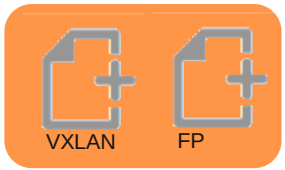
Multi-Site



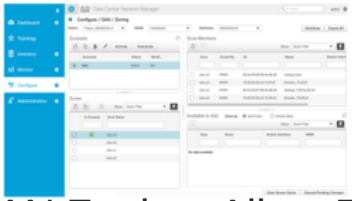
Multi-Fabric



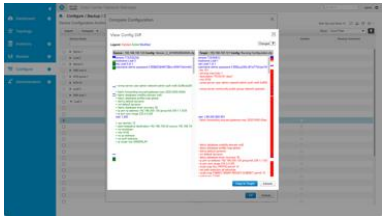
Device Packs



Solution Templates & Automation



SAN Zoning, Alias, PMon



Image, Config, Patch, GIR

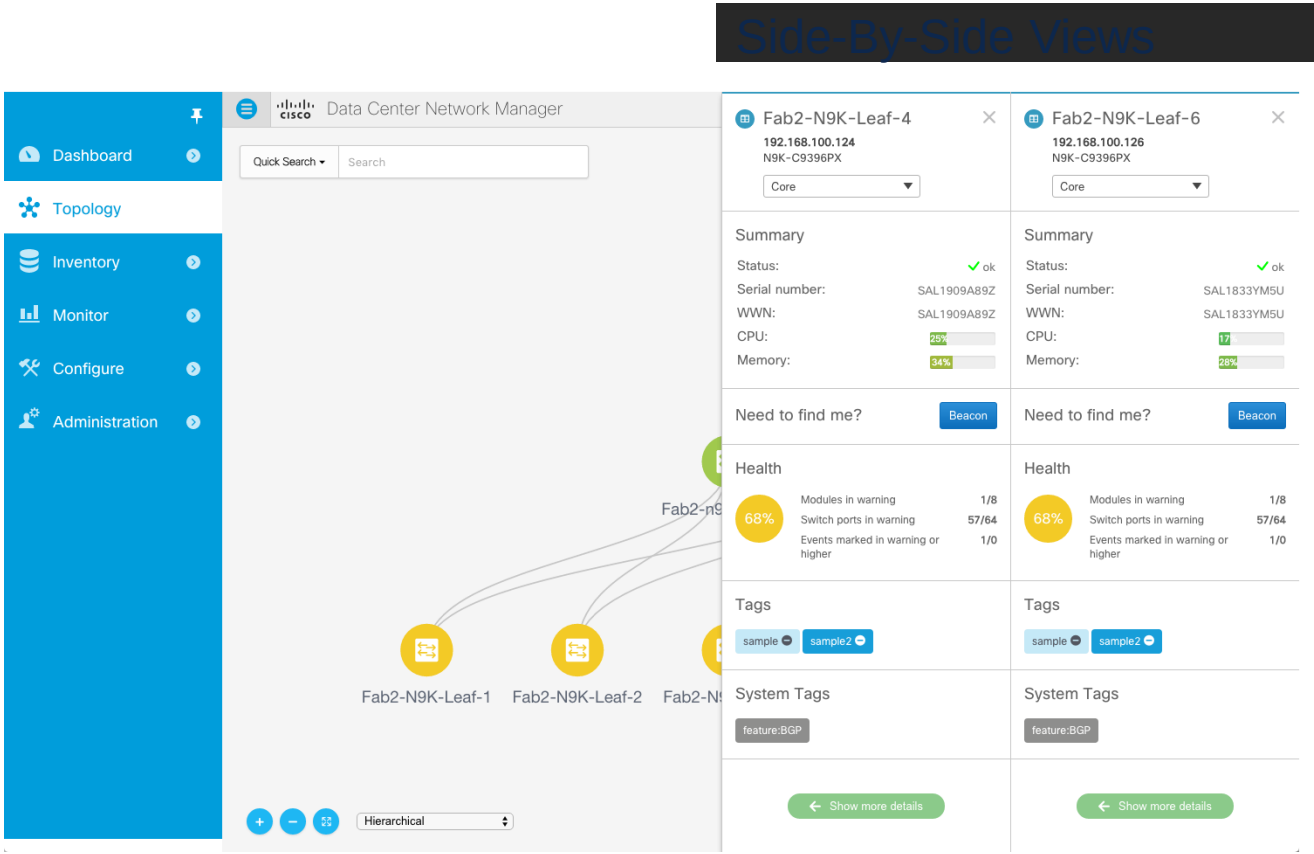
Nexus & MDS Platforms

^POAP = “Power-On Auto Provisioning”
*PM = “Physical Machine”

“Infrastructure ++” Updates

New Topology

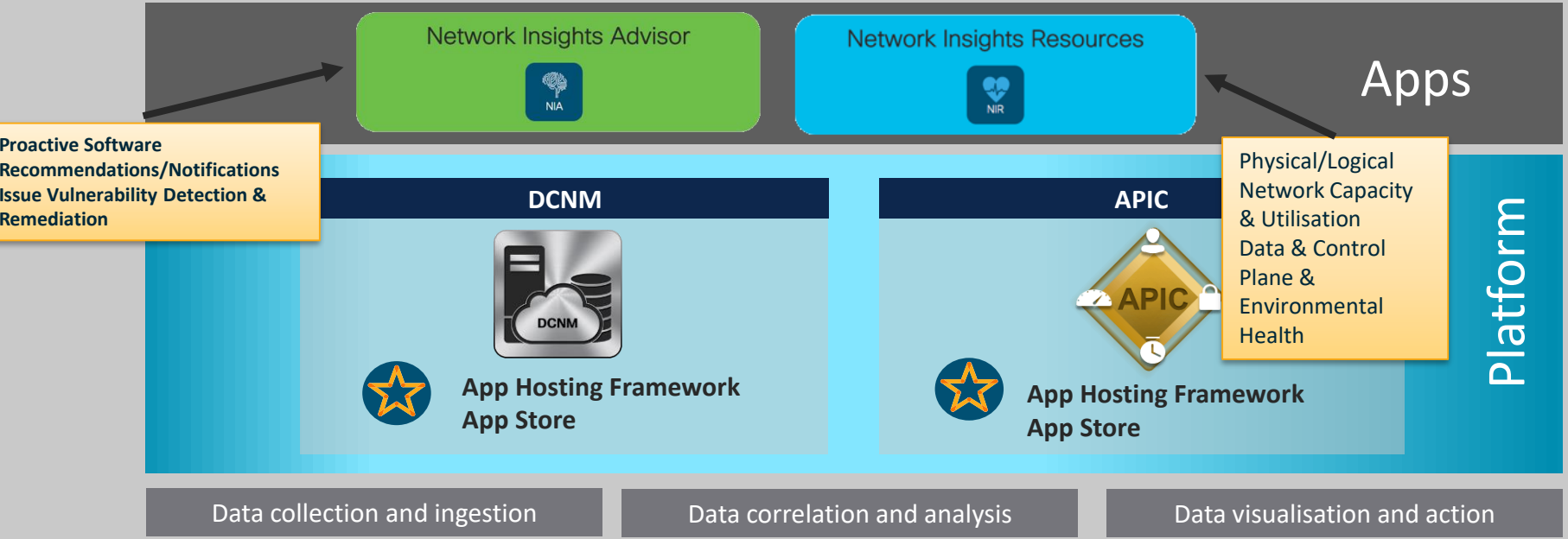
- Dynamic Arrangement
- Multi-Fabric/Overlay
- Arrange by Tier
  - [Core, Ag, Access Leaf, Spine etc..]
- Metadata Tags
- Show FEX links
- Device Pop-Over
  - Side-By Side View



Demo
DAFE & DCNM

Network Insights

Network Insights Applications



Visibility
Learn from your network and recognize anomalies

Insights
See problems before your end users do

Proactive Troubleshooting
Find root cause faster with granular details

Using Network Insights to Deliver Outcomes

Ingest and Process Telemetry Data

- Config File
- Syslog
- Tech-Support
- RIB
- FIB
- Accounting Logs
- Debug Logs
- Encapsulation Tables
- Streaming Telemetry
- Environmental
- Event History
- Topology
- Cores
- Consistency Checkers
- Mac Table
- TCAM Tables

Derive Insights

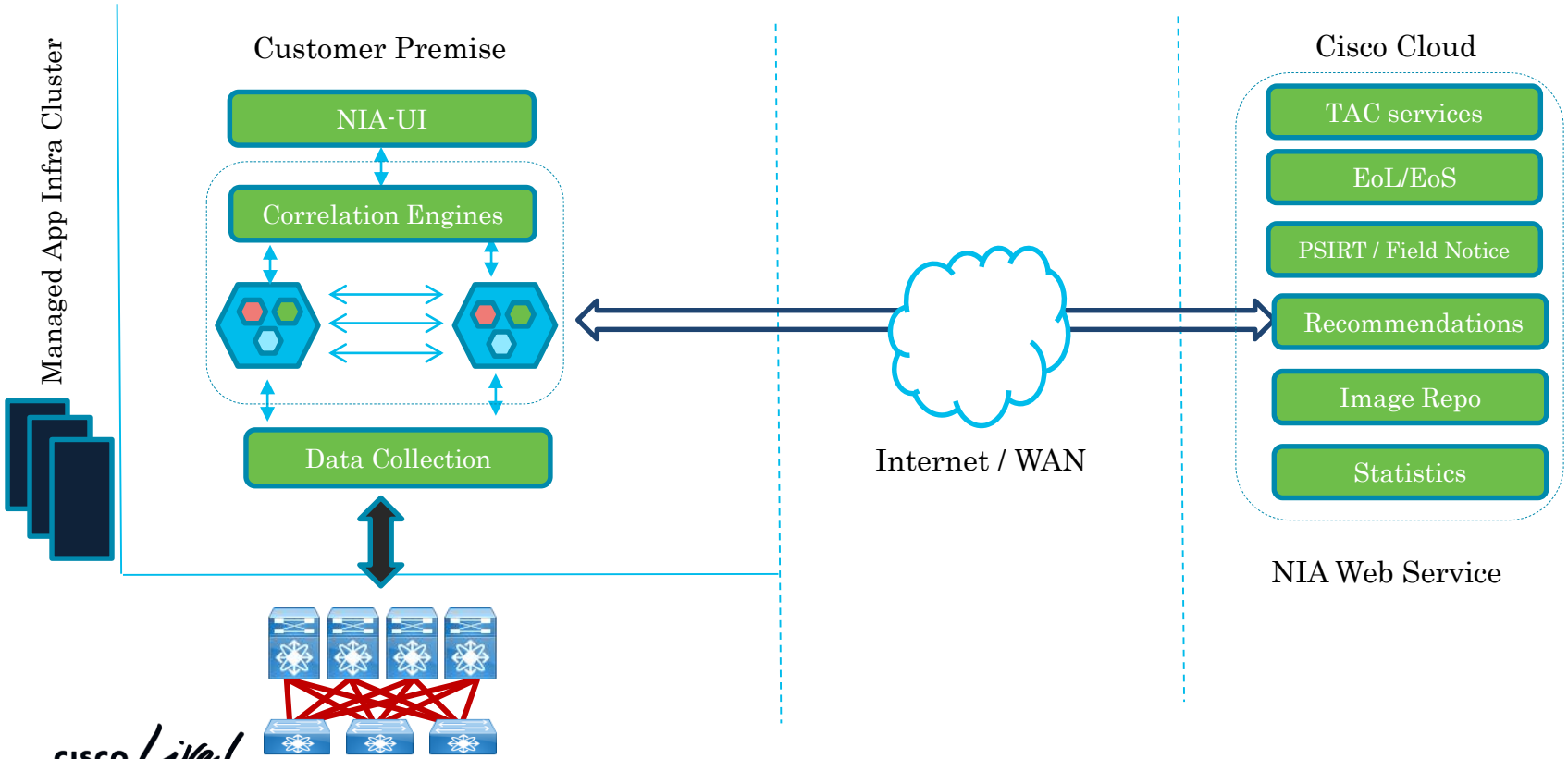
Performance

Capacity

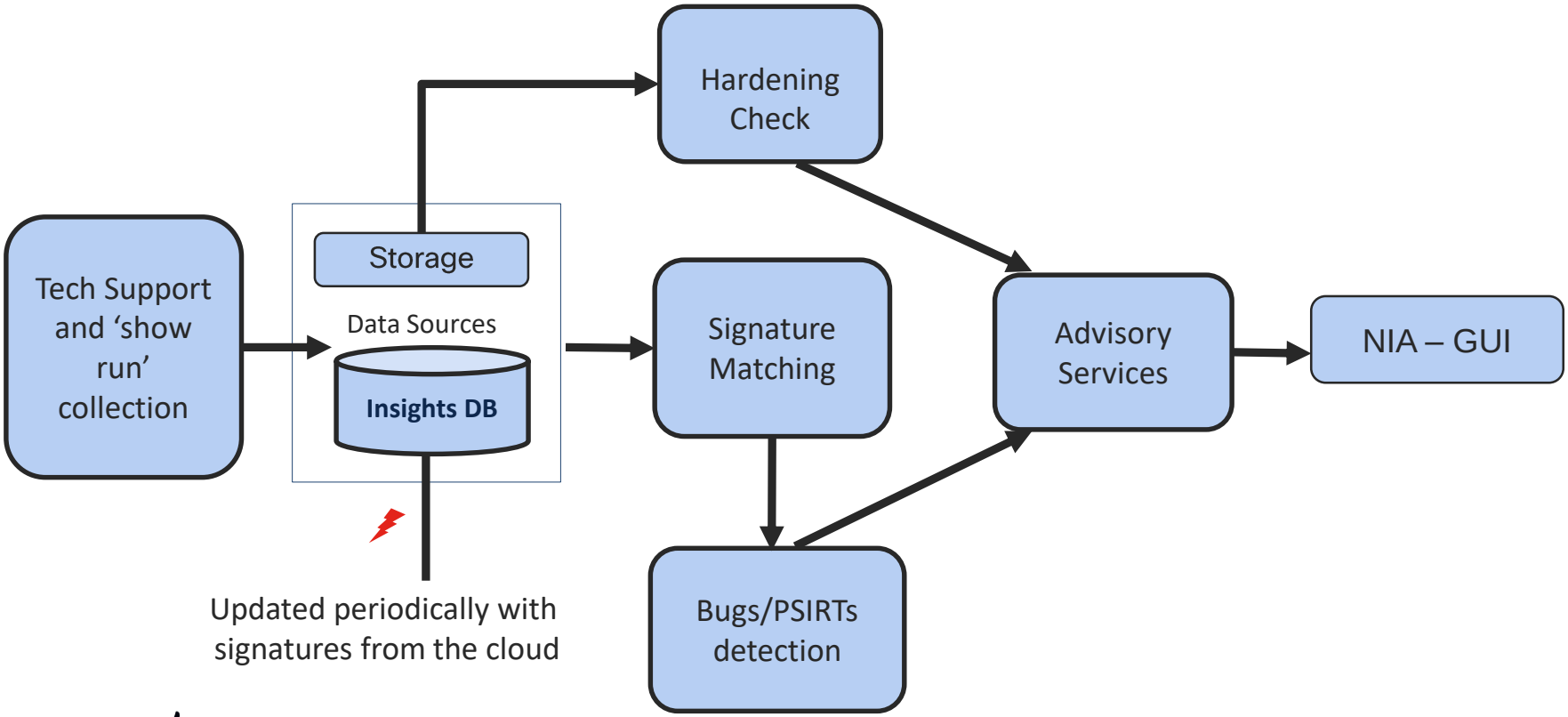
SW Integrity

System Health

Network Insights Advisor High Level Architecture

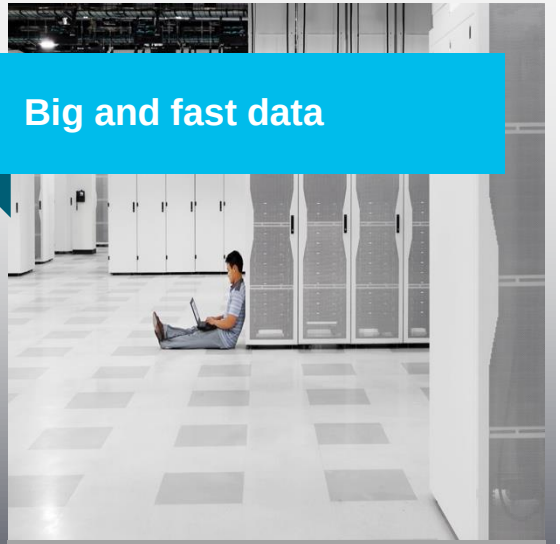


How Does NIA Detect Known Issues?



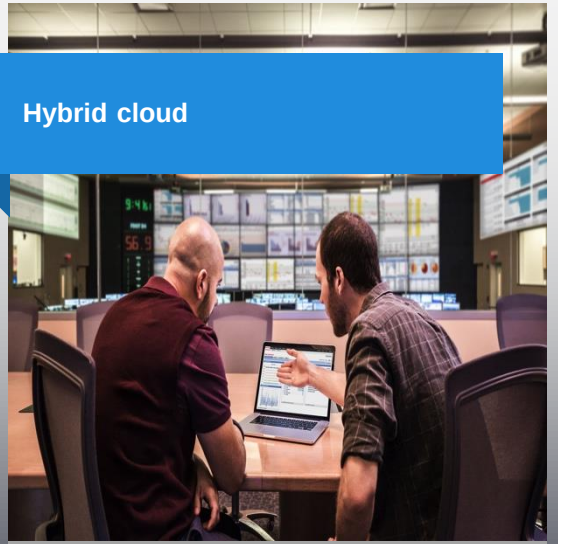
Data Center Behavioral Monitoring

Modern Data Centers Are Getting Increasingly Complex



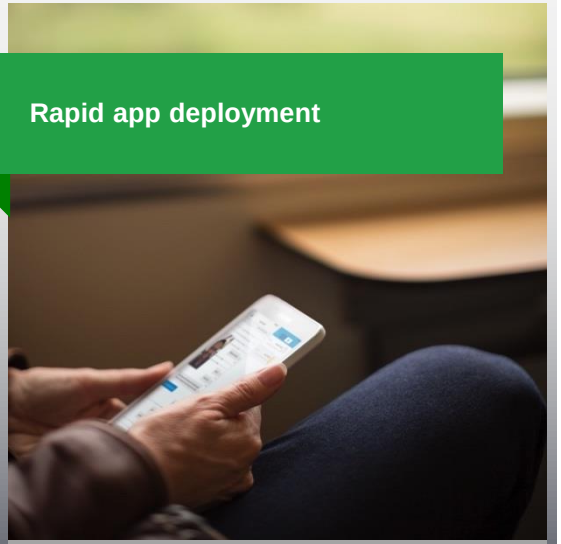
Big and fast data

- Increase in east-west traffic
- Expanded attack surface
- Open source



Hybrid cloud

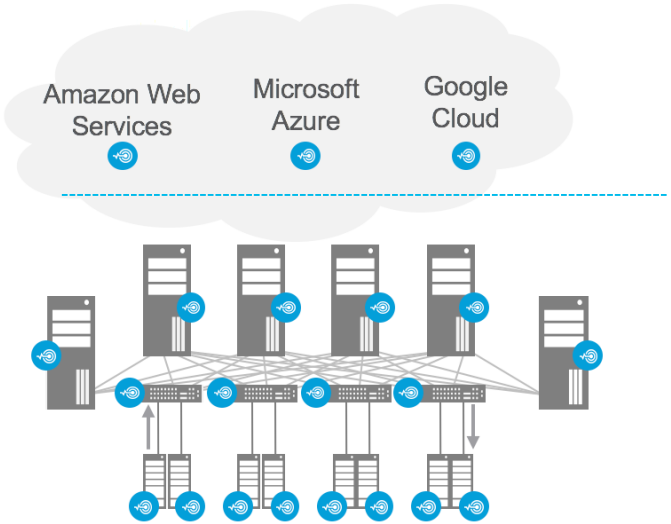
- Zero-trust model
- Multicloud orchestration
- Application portability



Rapid app deployment

- Continuous development
- Application mobility
- Micro services

Evolving Landscape and Monitoring



Data Creation

Real time tools: TAPs, NETFLOW
Non Real time tools: SNMP, Syslog, CLI (scripts)

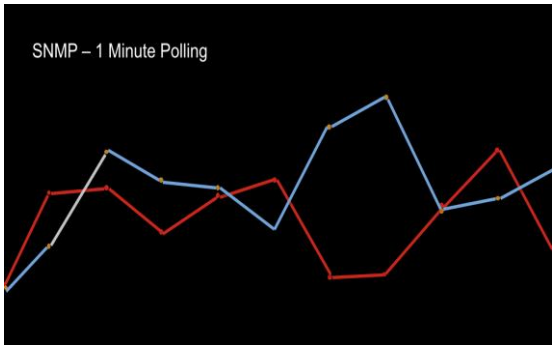
Data Analysis

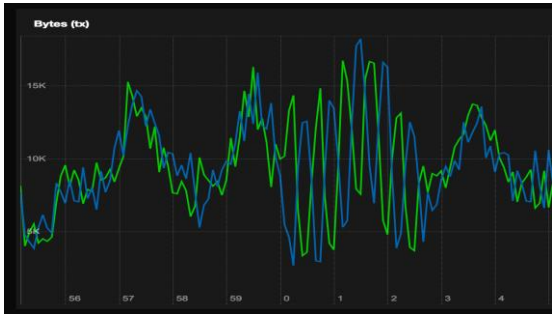
Storage & Analysis

Strong burden on back-end

Normalize different encodings, transports, data models, timestamps

Need For **Data** Analytics





10 Second SW Process Push

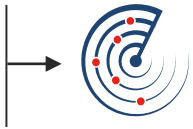
Cisco Tetration

Profile and Context Driven Application Segmentation/ Behavioral Assessment

1. Real-time Asset Tagging

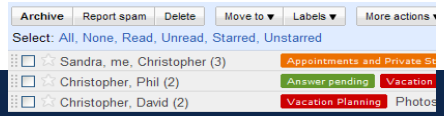
Cisco Tetration Sensors

Customer Defined



2. Policy Workflows

Cisco Tetration Application Insights (ADM)
+
Tag and Label-Based Add-on Policy
(For Example, Mail Filters)



3. Policy Enforcement
(Role Based and Hierarchical)

No Need to Tie Policy to IP Address and Port

Cisco Tetration Platform Performs the Translation

Compliance Monitoring

Enforcement

Public Cloud

Bare Metal

Virtual

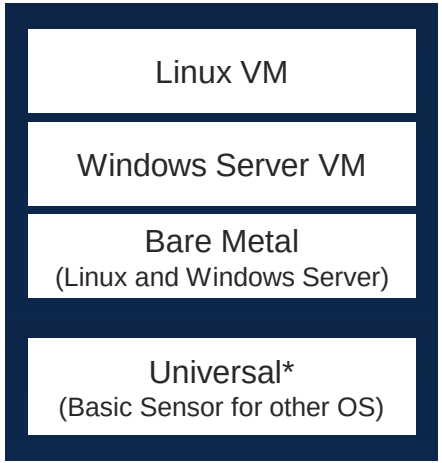
Cisco ACI™*

Traditional Network*

Pervasive Sensors

Software Sensors

Available Now



*Note: No per-packet Telemetry, Not an enforcement point

- ✓ Low CPU Overhead (SLA enforced)
- ✓ Low Network Overhead (SLA enforced)

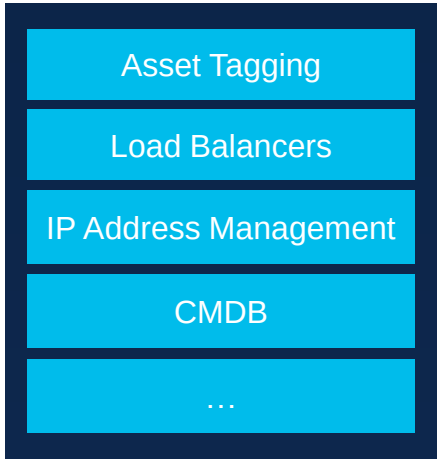
Network Sensors

Next Generation 9K switches

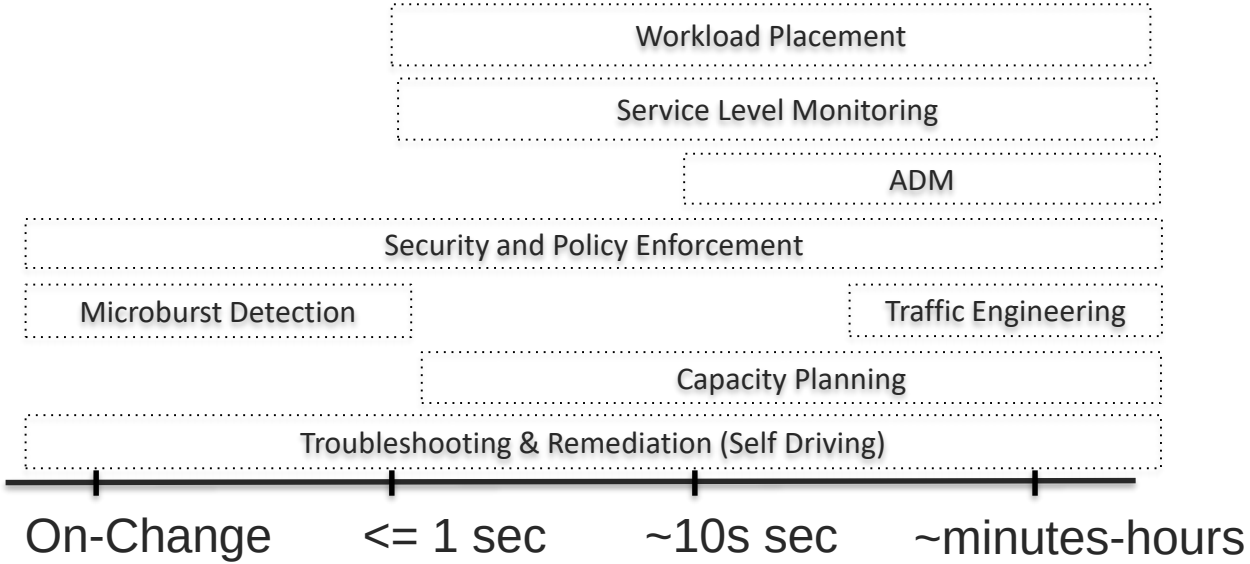
- ✓ Enforcement Point (Software agents)
- ✓ Highly Secure (Code Signed, Authenticated)
- ✓ Every Flow (No sampling), NO PAYLOAD

Third Party Sources

3rd party Data Sources



Data Granularity Needs to Improve
Type of Problems Customers are Looking to Address

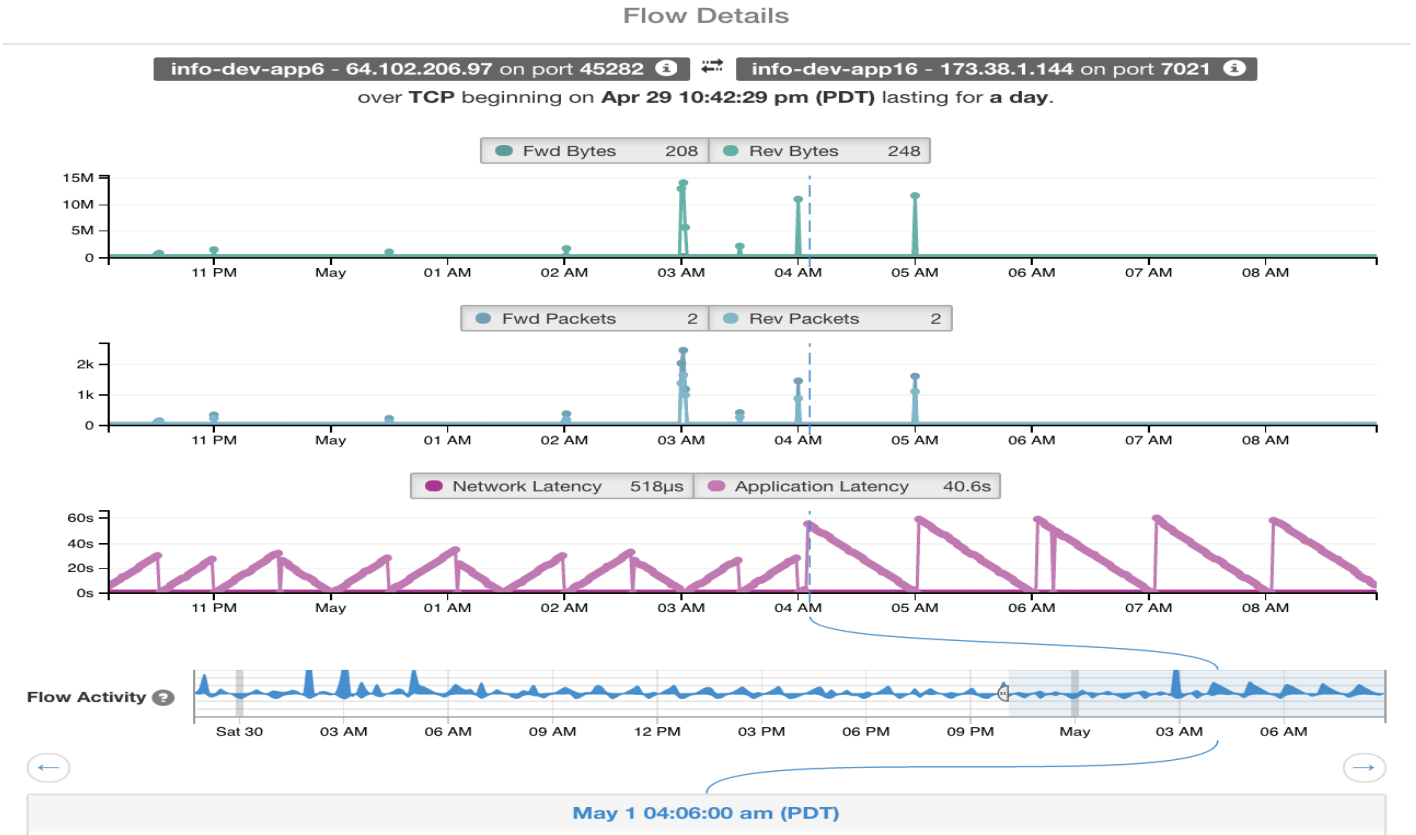


Resolution = Frequency of Data Collection

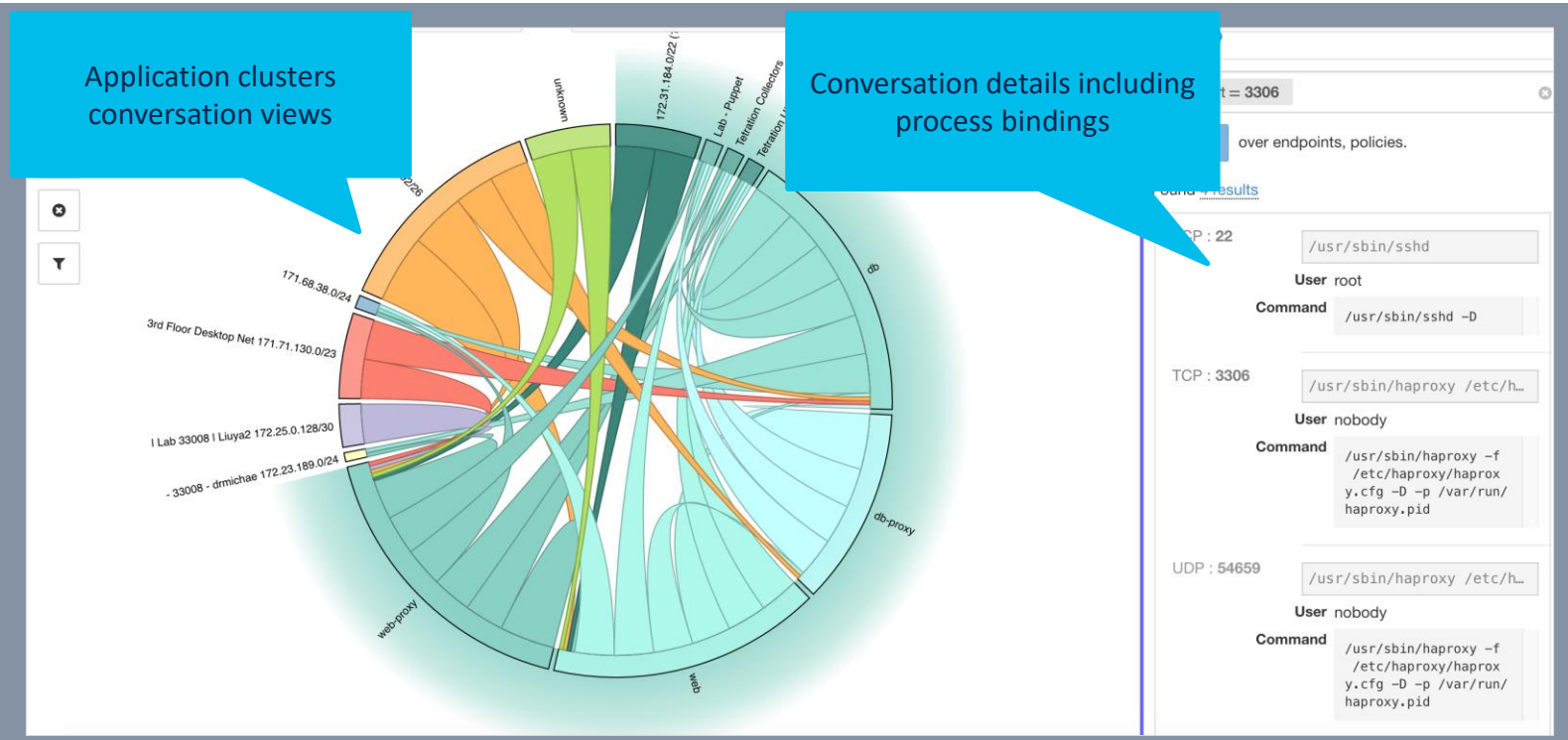
Flow Cache with streaming export

- Maintain a small 'cache' and export the cache at a high data rate
- Shift the cost of aggregation to backend resources
- Aggregate 'Flow Table' can be much larger

Data Granularity Needs to Improve
Sub Second HW/SW Push – Use Case 1

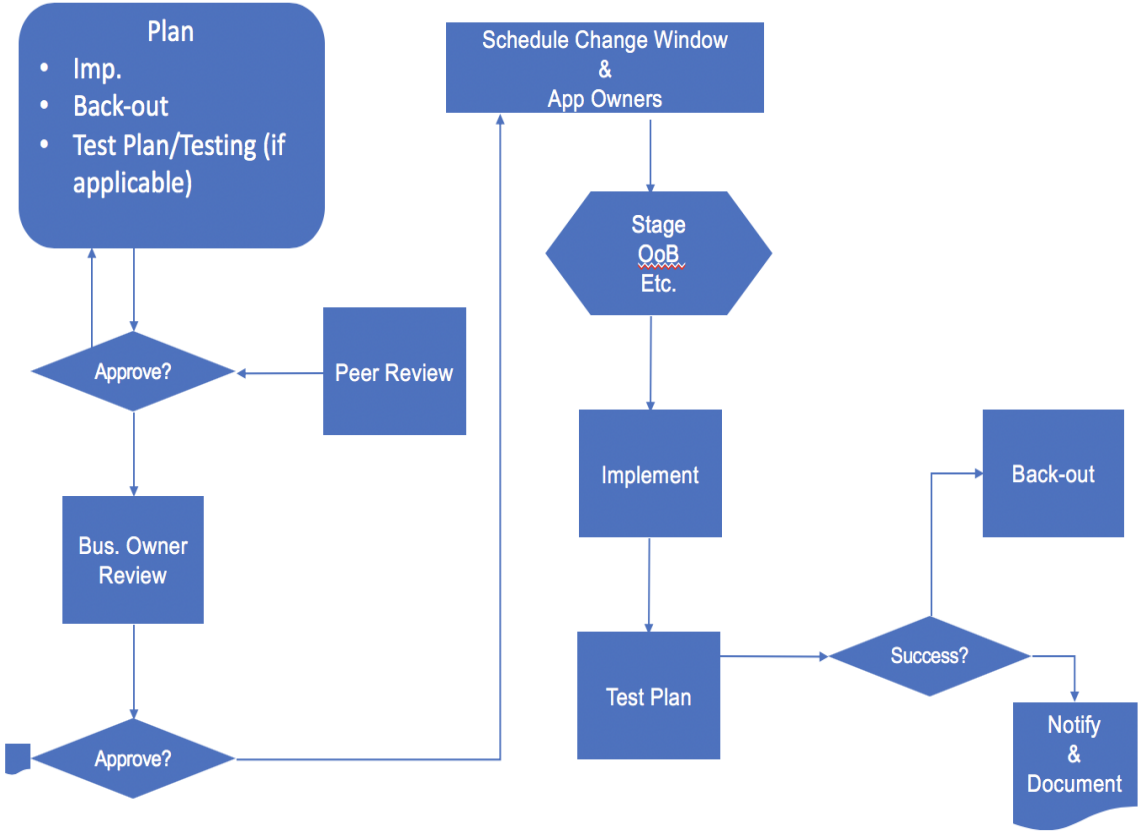


Application Conversation View – Use Case 2



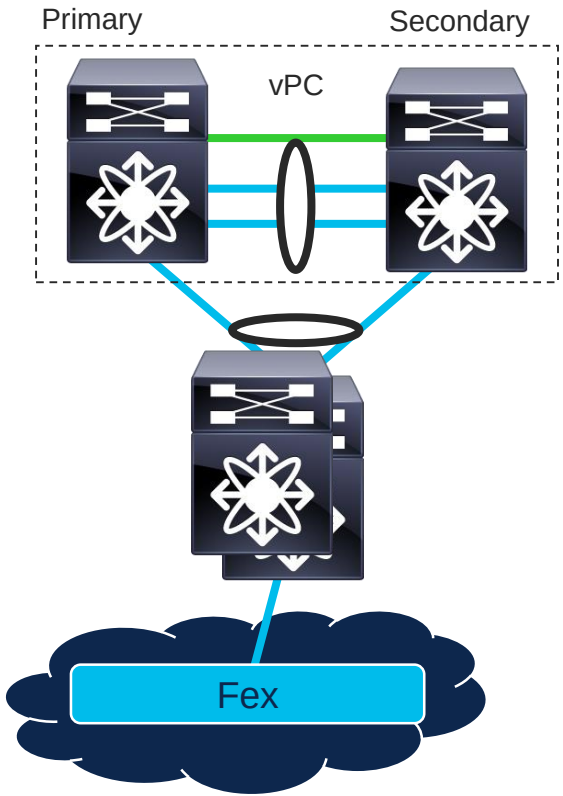
Maintenance Windows – Golden Rules

- Change Review Board
- Schedule when environment will be least impacted.
- Software Staging
- Verify out of band.
- Test! After *and* before.



Traditional vPC Environment Change

Change Best Practice and Window



Core Isolation

1. Graceful L3 Protocol Isolation
2. Layer 2 Isolation
   - VPC
3. Interface Isolation

Using GIR Mode Steps 1-3 could be achieved prescriptively.

Access Isolation

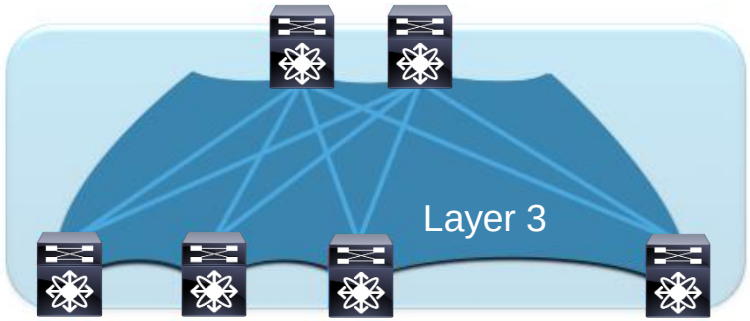
1. Layer 2 Isolation
   - VPC
2. Interface Isolation
   1. Fex-fabric (include/exclude)
   2. Dual-attached FEX Procedure * Recommended

Using GIR Mode Steps 1-2 could be achieved prescriptively.

NOTE: Maintenance mode consideration should be based on Fex-fabric connectivity.

If change window is for software upgrade or spot fix, consider ISSU or SMU feasibility.

L3 Environment
Change Best Practice and Window



Core Isolation

1. Graceful L3 Protocol Isolation
2. Interface Isolation

Using GIR Mode Steps 1-2 could be achieved prescriptively.

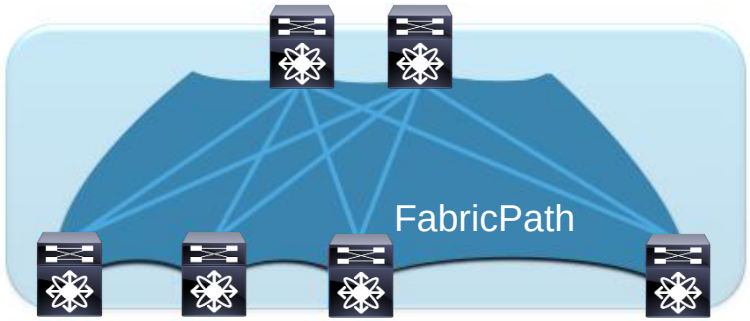
Access Isolation

1. L3 Protocol isolation
2. Layer 2 Isolation
   - vPC
3. Interface Isolation
   1. Fex-fabric (include/exclude)
   2. Dual-attached FEX Procedure * Recommended

Using GIR Mode, prescriptive isolation is possible.

If change window is for software upgrade or spot fix, consider ISSU or SMU feasibility.

FabricPath Environment
Change Best Practice and Window



Spine Isolation

1. Use FabricPath IS-IS Overload Bit

Using GIR Mode with isolate configuration, Step 1 could be achieved prescriptively.

Leaf Isolation

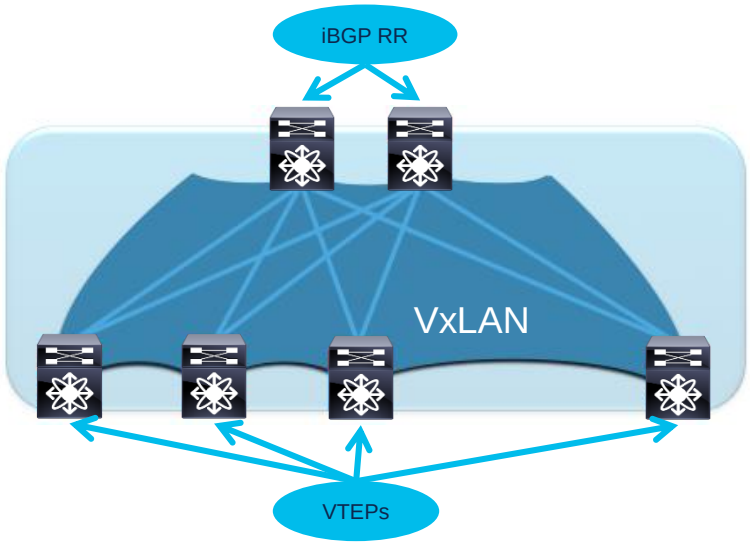
1. Use FabricPath IS-IS Overload Bit
2. Shutdown the VPC+ domain.

Using GIR Mode with manual profile, step 1 could be achieved prescriptively.

If change window is for software upgrade or spot fix, consider ISSU or SMU feasibility.

VxLAN EVPN Environment

Change Best Practice and Window



Spine Isolation

1. L3 Protocol isolation
   - If iBGP EVPN, consider IGP isolation
   - If eBGP EVPN, consider BGP isolation
2. Interface Isolation

Using GIR Mode Steps 1-2 could be achieved prescriptively.

Leaf Isolation

1. L3 Protocol isolation
   - If iBGP EVPN, consider IGP isolation
   - If eBGP EVPN, consider BGP isolation
2. Layer 2 Isolation
   - vPC
3. Interface Isolation
   1. Fex-fabric (include/exclude)
   2. Dual-attached FEX Procedure * Recommended

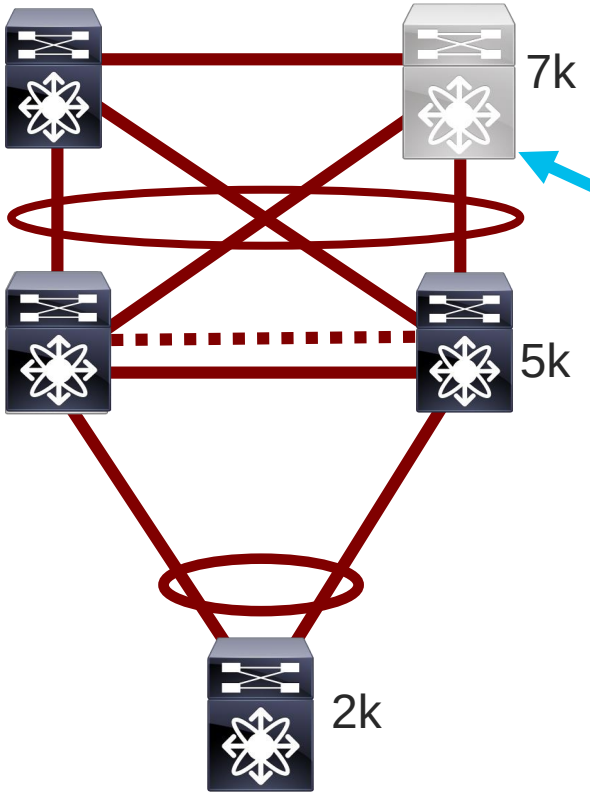
Using GIR Mode, prescriptive isolation is possible.

If change window is for software upgrade or spot fix, consider ISSU or SMU feasibility.

NX-OS 6.x -> 7.x
Use Case

NX-OS 6.x -> 7.x Use Case - Secondary

Manual Effort



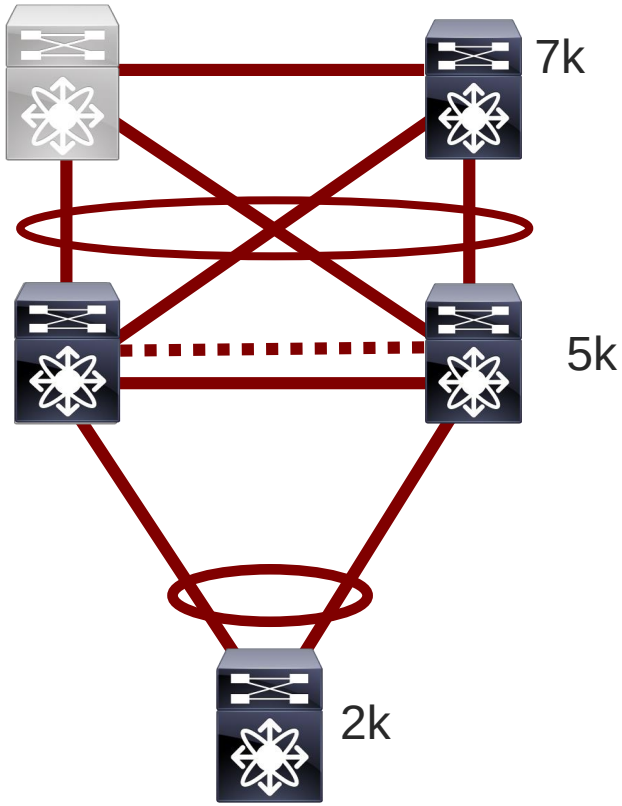
- Peer Switch
- Peer Gateway
- Auto-recovery
- L3 Link between vPC pairs
- BFD
- Routing Protocol Convergence Tuning

7k Upgrade

- Prerequisites
  - Code Staging
  - VPC Best Practices
- Manual Isolation of Secondary
  - Protocol Isolation
    - Max-metric LSA, etc. -> No service impact (0-20ms)
  - VPC Isolation
    - Down vPCs-> No service impact (0-20ms)
    - Down Peer Link-> No service impact
- Reload Upgrade->No service impact
  - Peer link is brought UP-No Service impact
  - South links UP – No Service impact
  - North protocol Max-metric LSA removal – UP –No Service impact

NX-OS 6.x -> 7.x Use Case Primary

7k Single Supervisor



Manual Effort

Prerequisites
- ✓ Code Staging
- ✓ VPC Best Practices

7k Upgrade

- Manual Isolation of Secondary
  - Protocol Isolation
    - Max-metric LSA, etc. -> No service impact (0-20ms)
  - VPC Isolation
    - Down vPCs-> No service impact (0-20ms)
    - VPC peer priority changes The secondary should have a lower priority to become the primary incase of flapping.
    - Down Peer Link-> No service impact
- Peer link & KPA is brought down & Reload initiated for Upgrade->No service impact to 0-50ms impact in traffic based on traffic pattern (this switch comes as secondary)
  - Peer link is brought UP-> No Service impact
  - South links UP ->No Service impact
  - North protocol UP ->No Service impact

Note: The System did not have firewall or LB connected directly to it.

Summary

Putting it all Together

- What to use? GIR Mode? Patching? ISSU? All of them?

Option \ Situation	Critical Bug Fix & PSIRT	Hardware Upgrade	New Features
ISSU	✓	X	✓
GIR + Cold Boot	✓	X	✓
GIR + Disruptive Installer	✓	X	✓
SMU Restart	✓	X	X
GIR + SMU ISSU	✓	X	X
GIR	X	✓	X

Summary

- Verify environment conforms to data center networking best practices.
- Follow your *documented* change management process.
- Isolate nodes during maintenance to minimize disruption.
  Use GIR Mode where possible to ease isolation configuration.

Thank you


cisco
You make possible


CISCO