Part 1.

VXLAN EVPN의 이해

Cisco Systems Korea

TSS | Kyuhyun Lim

Cloud Infrastructure & Software Group

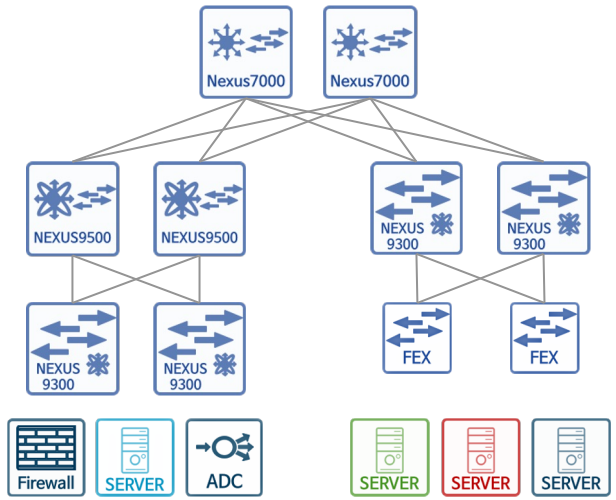
Data Center Networking

데이터센터 네트워크 구조 다양화

Nexus Switch의 디자인 모델

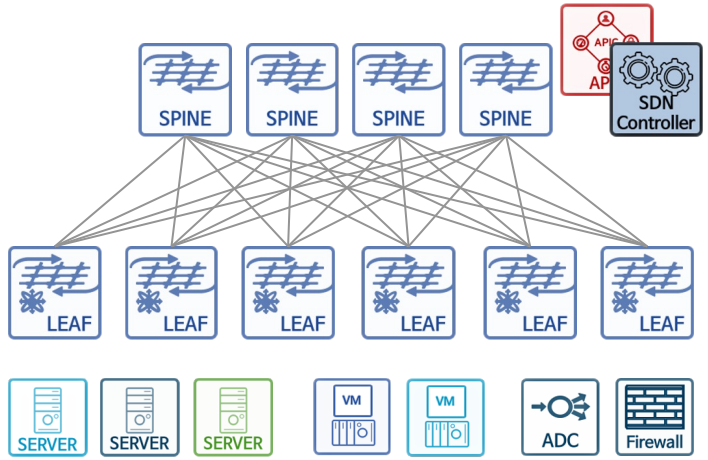
전통적 2/3 계층, VXLAN SDN, MSDC EVPN

일반적인 2/3 계층 모델



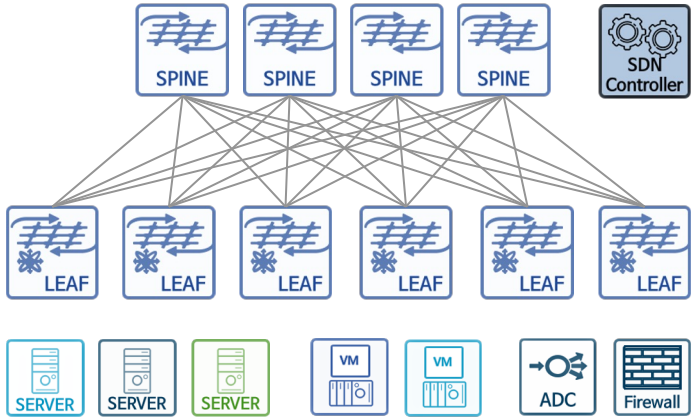
- 분산형 네트워크 모델링
- 2/3 계층의 전통적 네트워크 디자인 방식
- CLI 기반의 수동 프로비저닝 방식
  - 휴먼 에러, 긴 구성 변경 소요 시간
- 랙 중심의 어플리케이션 서버 구성
  - 제한된 확장, 가상화 이동성의 제약, 응답시간, 성능, 추적관리 제약

차세대 엔터프라이즈 가상화 모델



- 어플리케이션 중심의 네트워크 가상화 모델
- 논블러킹 Flat 구조의 Spine-Leaf CLOS 디자인 방식
- VxLAN 기술을 활용한 멀티 테넌트 기반 구성
- 테넌트 단위 Flat한 패브릭 구조로 물리적/가상화 서버 및 네트워크 서비스 구성의 다양성 제공
- 어플리케이션을 인지하여 서비스 형태로 구분 가능
- 가상화 이동성, 응답시간, 스케일 아웃, 추적 관리 최적화

MSDC 모델



- 개방형 API 를 이용한 DevOps 기반 네트워크 모델
- 논블러킹 Flat 구조의 Spine-Leaf CLOS 디자인 방식
- 테넌트 단위 Flat한 패브릭 구조로 물리적/가상화 서버 및 네트워크 서비스 구성의 다양성 제공
- MP-BGP 호스트 기반 라우팅 기술로 고도의 운영 기술
- 가상화 이동성, 응답시간, 스케일 아웃, 추적 관리 가능

구축 모델 별 Capacity

NDFC Architecture Deployment Model

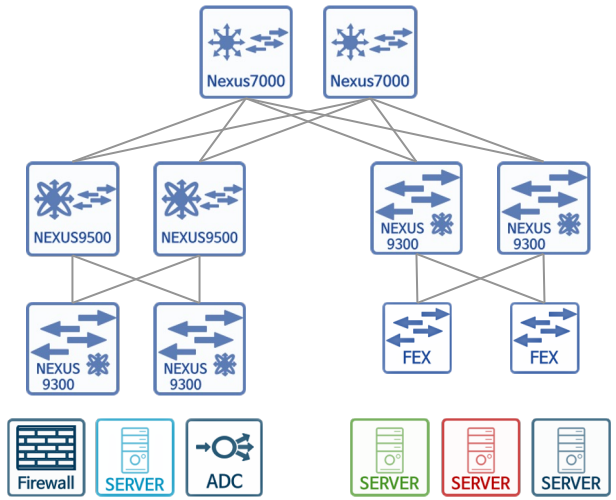
Model	MSDC/BGP	VXLAN w/ NDFC	VXLAN w/ ACI
물리 Port/Switch 수량	제한 없음 (BGP 구성에 따르며 ASIC Scale에 따라 상이	최대 약 30,000 포트 400대 스위치(금년 내 500대 수용 예정)	최대 24,000 포트 400대 스위치
지원 기능	• 표준 기반 라우팅 • Route Control • Telemetry (ASIC 제약 있음)	• 표준 기반 라우팅 • VXLAN/MP-BGP • Route Control • Telemetry (ASIC 제약 있음)	• Full Automation • Zero Trust Policy Model • VXLAN/MP-BGP • Route Control • Telemetry (ASIC 제약 있음)
L2 Extension	Overlay 구성은 선택	지원	지원
상호 호환성	연동 용이	VXLAN/EVPN 연동 어려움 Border Leaf/Spine 을 통해 표준 라우팅 방식 연동	Border Leaf 를 통해 표준 라우팅 방식 연동

Nexus Switch의 디자인 모델

전통적 2/3 계층, VXLAN SDN, MSDC EVPN

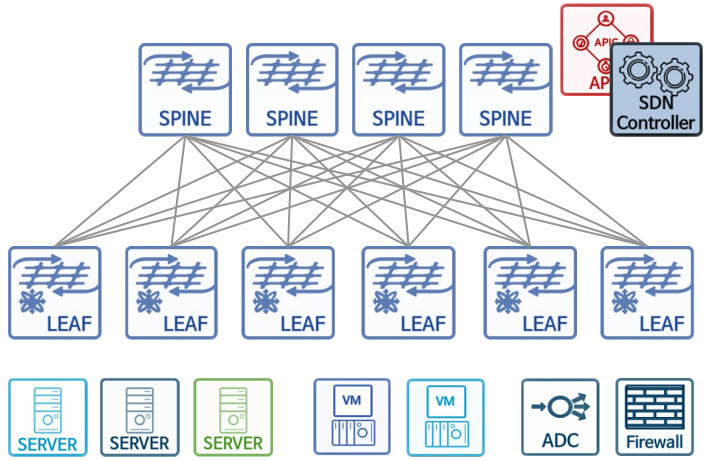
오늘의 주제

일반적인 2/3 계층 모델



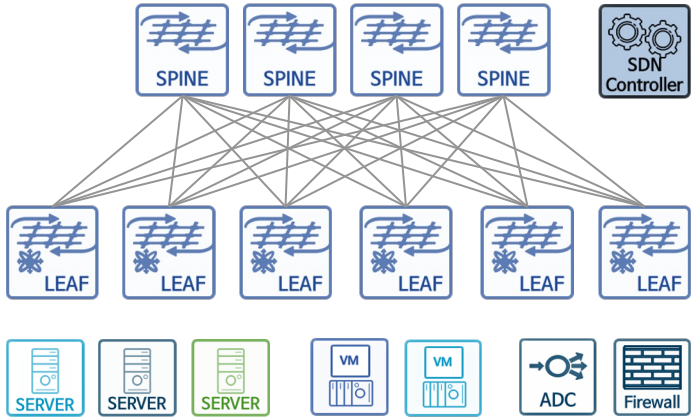
- 분산형 네트워크 모델링
- 2/3 계층의 전통적 네트워크 디자인 방식
- CLI 기반의 수동 프로비저닝 방식
  - 휴먼 에러, 긴 구성 변경 소요 시간
- 랙 중심의 어플리케이션 서버 구성
  - 제한된 확장, 가상화 이동성의 제약, 응답시간, 성능, 추적관리 제약

차세대 엔터프라이즈 가상화 모델



- 어플리케이션 중심의 네트워크 가상화 모델
- 논블러킹 Flat 구조의 Spine-Leaf CLOS 디자인 방식
- VxLAN 기술을 활용한 멀티 테넌트 기반 구성
- 테넌트 단위 Flat한 패브릭 구조로 물리적/가상화 서버 및 네트워크 서비스 구성의 다양성 제공
- 어플리케이션을 인지하여 서비스 형태로 구분 가능
- 가상화 이동성, 응답시간, 스케일 아웃, 추적 관리 최적화

MSDC 모델



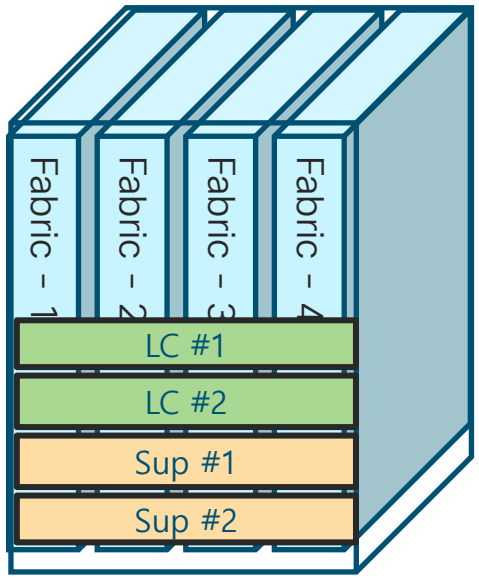
- 개방형 API 를 이용한 DevOps 기반 네트워크 모델
- 논블러킹 Flat 구조의 Spine-Leaf CLOS 디자인 방식
- 테넌트 단위 Flat한 패브릭 구조로 물리적/가상화 서버 및 네트워크 서비스 구성의 다양성 제공
- MP-BGP 호스트 기반 라우팅 기술로 고도의 운영 기술
- 가상화 이동성, 응답시간, 스케일 아웃, 추적 관리 가능

하이퍼스케일 인프라스트럭쳐의 정의

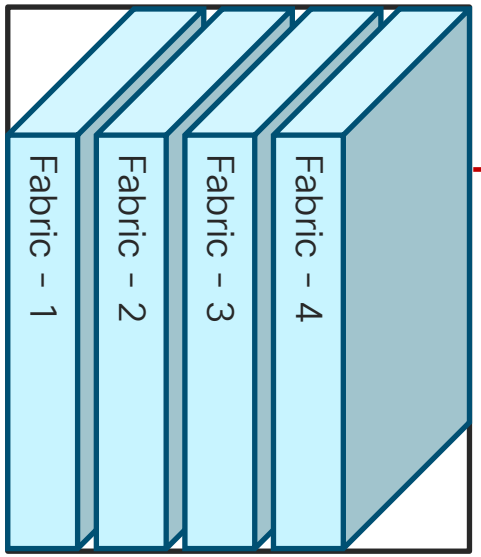
Hyper-Scale 인프라스트럭쳐란?

- 수요의 증가에 맞춰 적합한 규모로 인프라를 확장할 수 있는 환경을 의미 → 확장성
- 하이퍼스케일에 대한 역량이 커질 수록 시스템 확장이 유연해짐 → 확장성
- 대규모 컴퓨팅 환경에서 서버, 스토리지, 네트워크 자원을 탄력적으로 추가하고 매끄럽게 자원할당이 가능한 것이 특징 → 민첩성
- 기존 인프라 아키텍쳐로는 하이퍼 스케일 환경을 효율적으로 구현하기 어려움

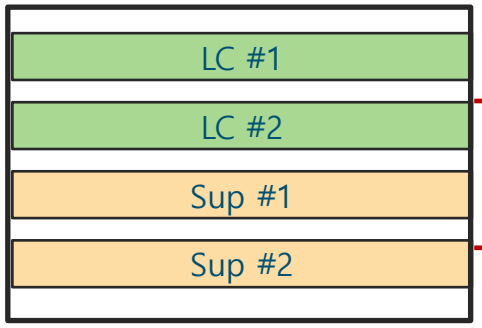
Fabric구축의 목적 = 거대한 모듈러 스위치를 만드는 것



모듈러 스위치



Chassis(Back)

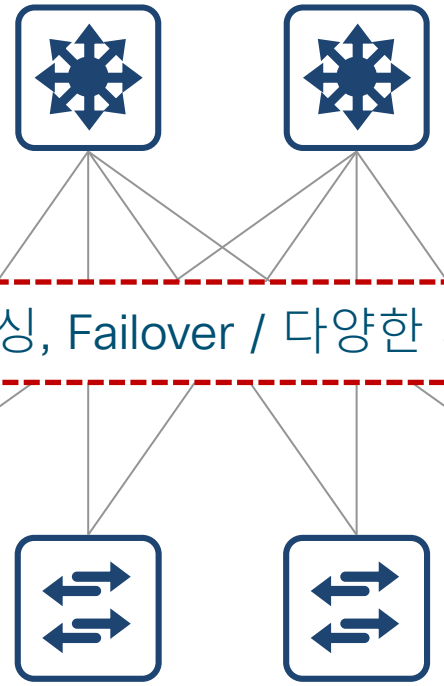


Chassis(Front)

Fabric Module → Spine

Line Card→ Leaf

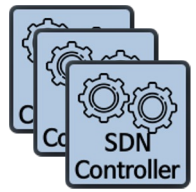
Supervisor → Controller



성능 담당

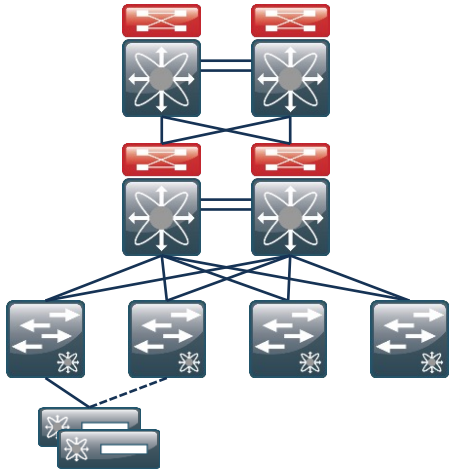
로드발란싱, Failover / 다양한 기술 접목

포트 수용담당

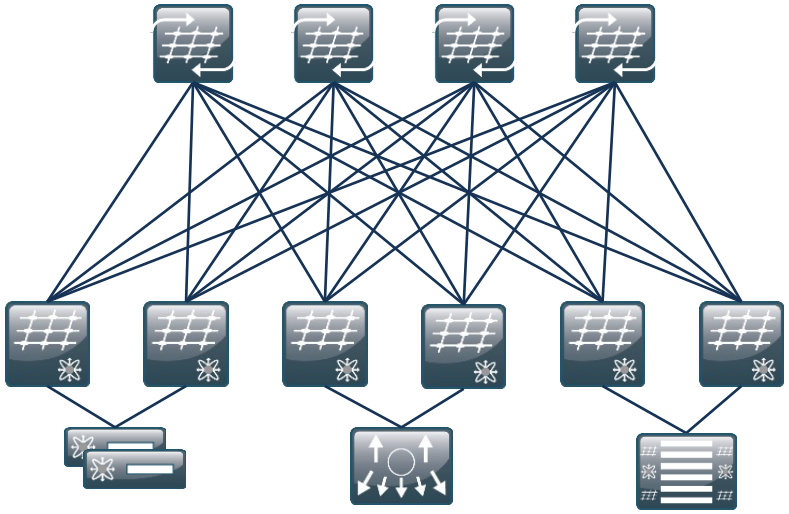


제어담당

하이퍼스케일 인프라스트럭쳐의 정의

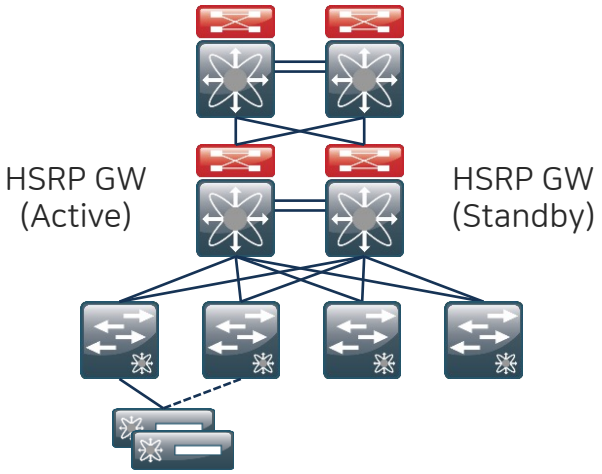


전통적 3-Tier Architecture

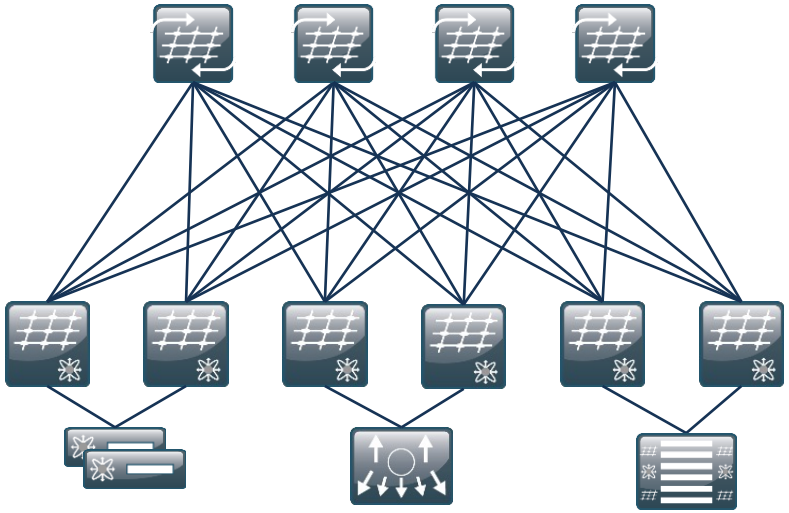


Scalable Spine/Leaf

하이퍼스케일 인프라스트럭쳐의 정의

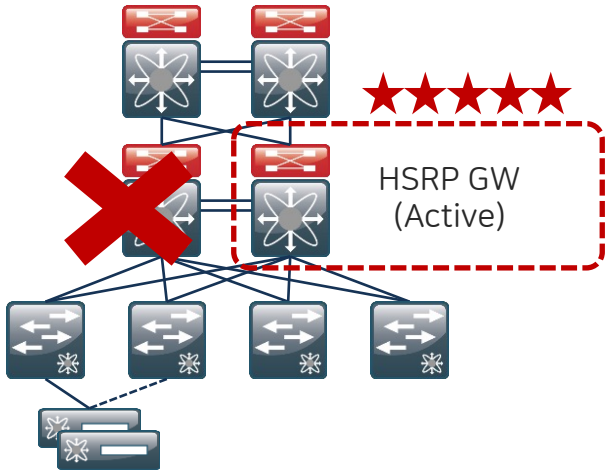


전통적 3-Tier Architecture

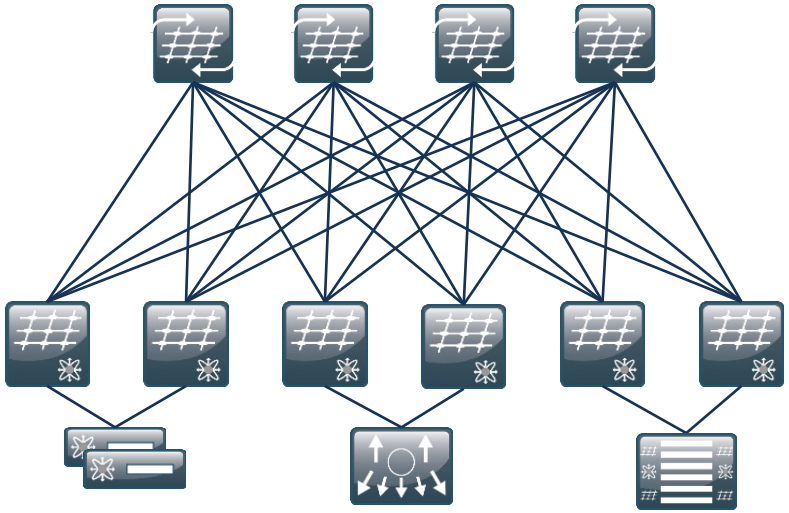


확장가능한 Spine/Leaf Architecture

하이퍼스케일 인프라스트럭쳐의 정의

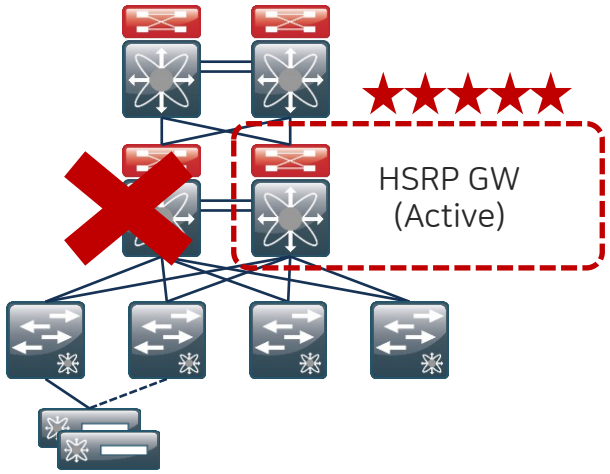


전통적 3-Tier Architecture

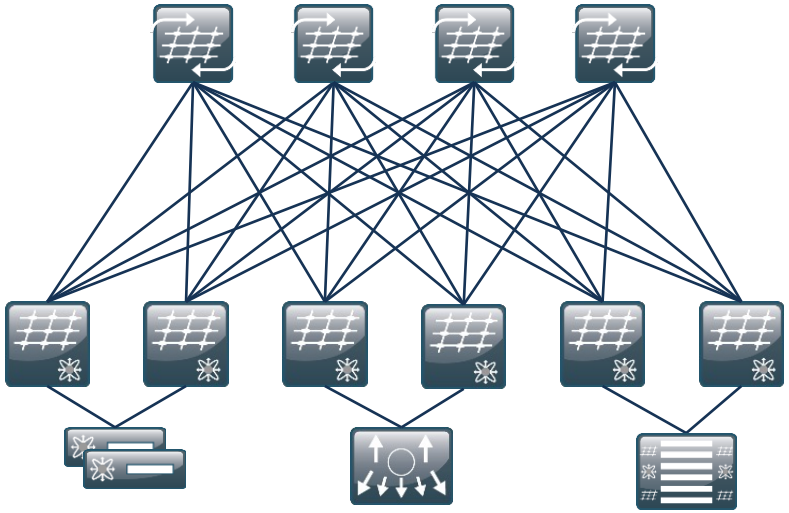


확장가능한 Spine/Leaf Architecture

하이퍼스케일 인프라스트럭쳐의 정의

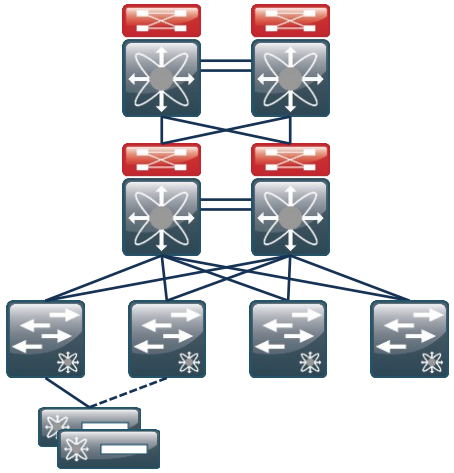


전통적 3-Tier Architecture

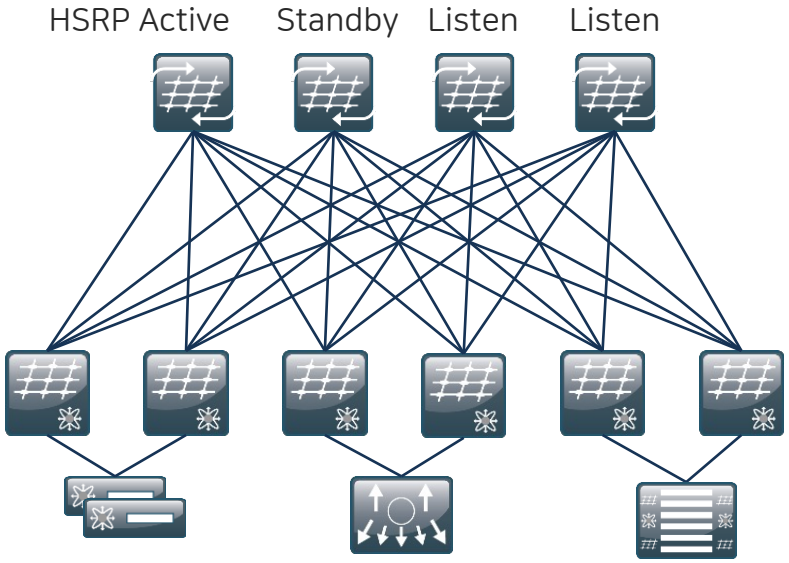


확장가능한 Spine/Leaf Architecture

하이퍼스케일 인프라스트럭쳐의 정의

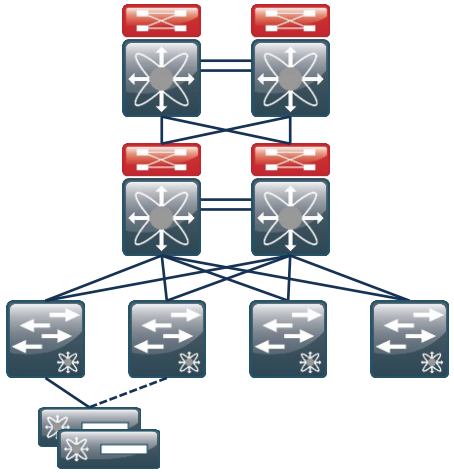


전통적 3-Tier Architecture

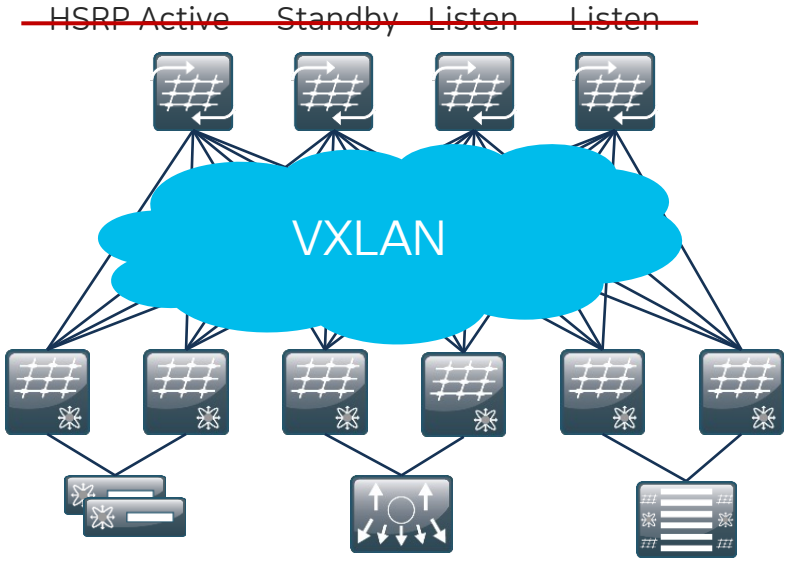


확장가능한 Spine/Leaf Architecture

하이퍼스케일 인프라스트럭쳐의 정의

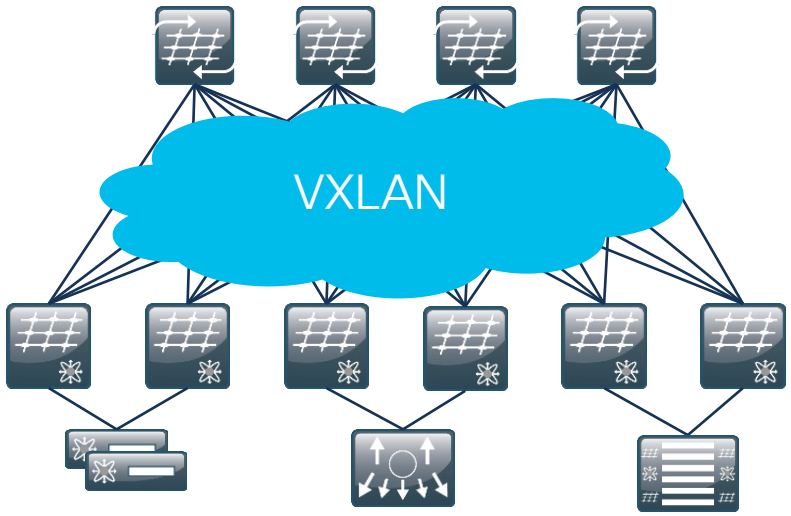


전통적 3-Tier Architecture



확장가능한 Spine/Leaf Architecture

하이퍼스케일 인프라스트럭쳐의 정의



확장가능한 Spine/Leaf Architecture

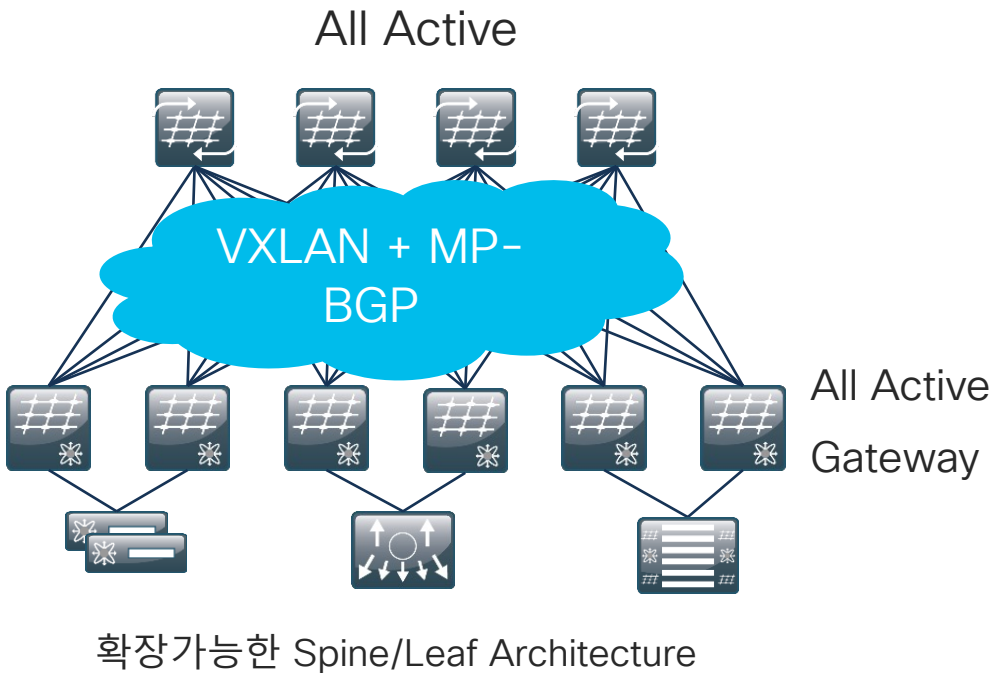
VXLAN 기술을 통한 L2 확장 완료

VXLAN 기술 = L2 확장

정보 교환과 통신 – 최초에 Flood & Learn 방식 사용
- IP 중복의 방지 – 확장된 L2
- 통신을 위한 End-point 위치파악 – 정보 교환

Flood & Learn 방식의 문제점 발생
- ARP와 유사
- 데이터 센터 내 전 영역에서
  Broadcast, Multicast, Unknown Unicast Flooding

하이퍼스케일 인프라스트럭쳐의 정의



확장가능한 Spine/Leaf Architecture

Flood & Learn 방식의 문제점 발생
- ARP와 유사
- 데이터 센터 내 에서 Broadcast

L3 네트워크 구성을 통해
Broadcast와 STP Loop 제거

L3 구성으로 Flood & Learn 방식의
IP 중복 체크와 정보교환 불가

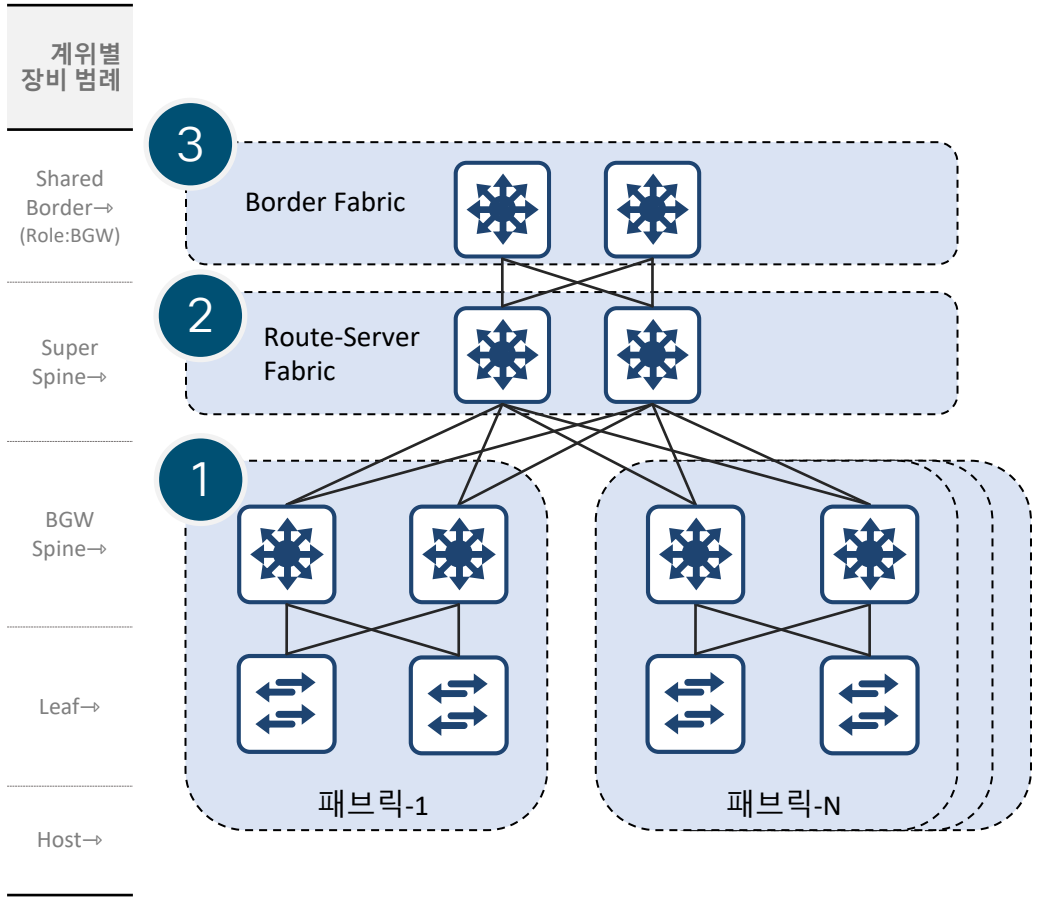
COOP / MP-BGP / LISP 등 Control-Plane 활용 – Control-Plane

- L3환경에서 L2 정보 교환 필요(ARP)
- 호스트 위치 정보 파악

Data Center Networking

VXLAN EVPN을 단계별로 직접 건설하며 이해

건설 계획. 용도에 맞는 적합한 디자인과 투자 규모 선택



1. Spine-Leaf : 하나의 패브릭을 만드는 단위(최초 투자범위)
2. Super Spine : Spine을 집선하는 역할로서, Fabric간의 연결 담당
   2개 이상의 패브릭을 연결할 때 유용
3. Shared Border : 다수 패브릭으로부터 트래픽을 수신하여 외부로 전달

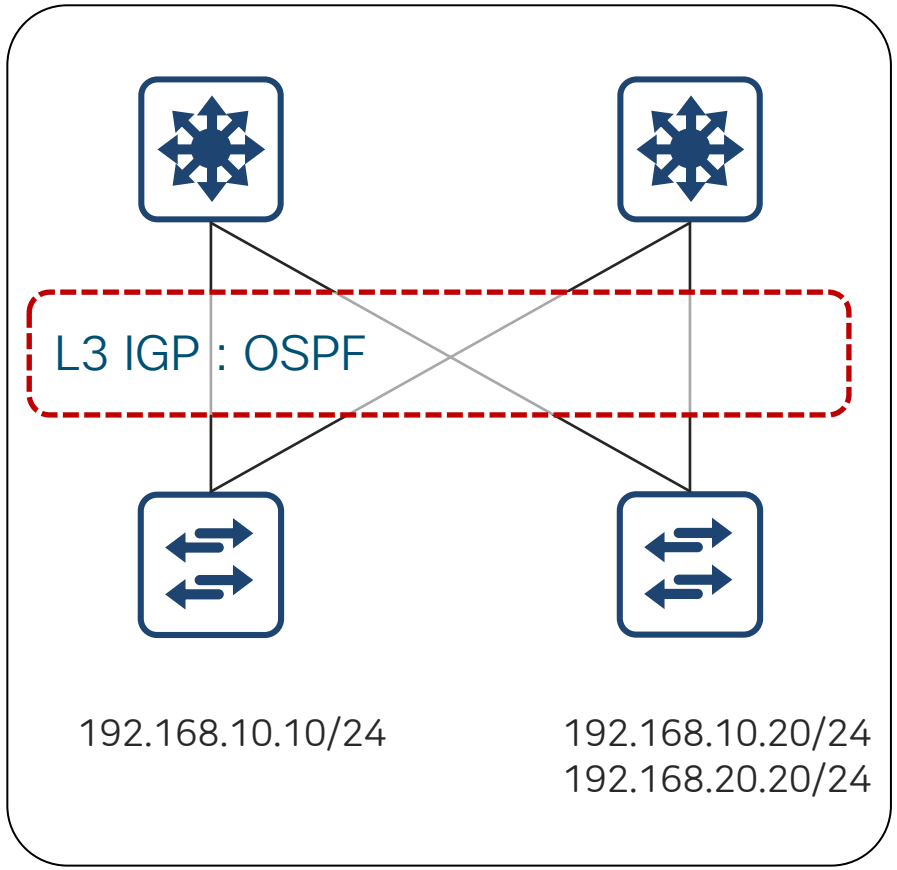
건설 시작. Leaf-Spine의 IGP(OSPF 기준)

계위별
장비 범례

BGW
Spine→

Leaf→

Host→



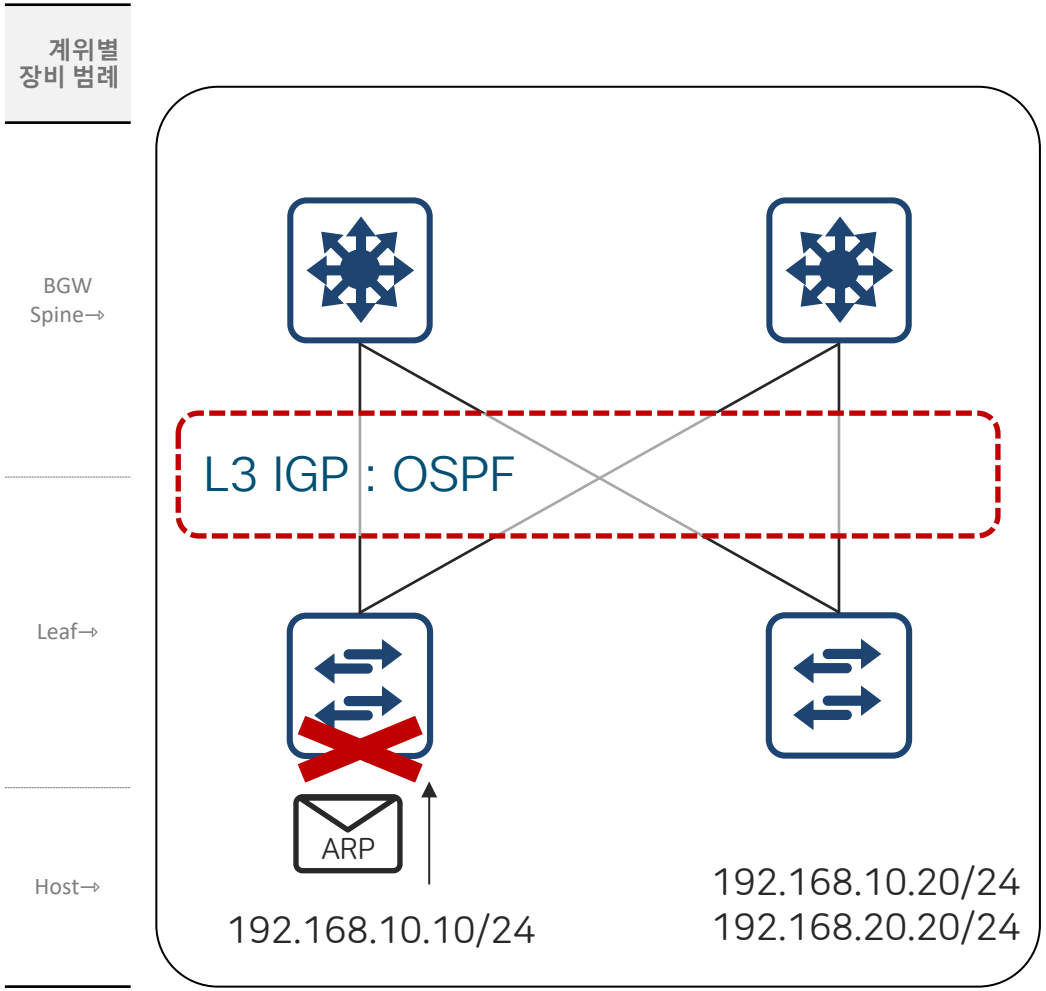
- **Spine과 Leaf 간에는 OSPF를 통한 IGP를 구성**
  > OSPF외 ISIS, EIGRP 등이 가능
  > 운영자가 익숙한 라우팅 프로토콜이 가장 좋은 프로토콜

- **L3 프로토콜을 사용하는 이유**
  > 트래픽의 출발 Leaf에서 도착 Leaf까지 모든 경로를 STP Block없이 Active로 활용하기 위함

```
Leaf 1# show ip route 192.168.20.20
via 1.1.1.1(Next-Hop 1 : Spine1의 인터페이스 주소)
via 1.1.1.2(Next-Hop 2 : Spine2의 인터페이스 주소)
```

Leaf#1의 호스트가 Leaf#2에 수용된 호스트로 통신하기 위한 라우팅 테이블

건설 시작. Leaf-Spine의 IGP(OSPF 기준) – L2 통신 한계



- 하지만 장비간 L3 프로토콜 사용시, 스위치가 다른 경우 ARP 전달 불가
  Leaf #1 → Leaf #2 통신하기 위한 경로 중에 L3가 존재하는 경우

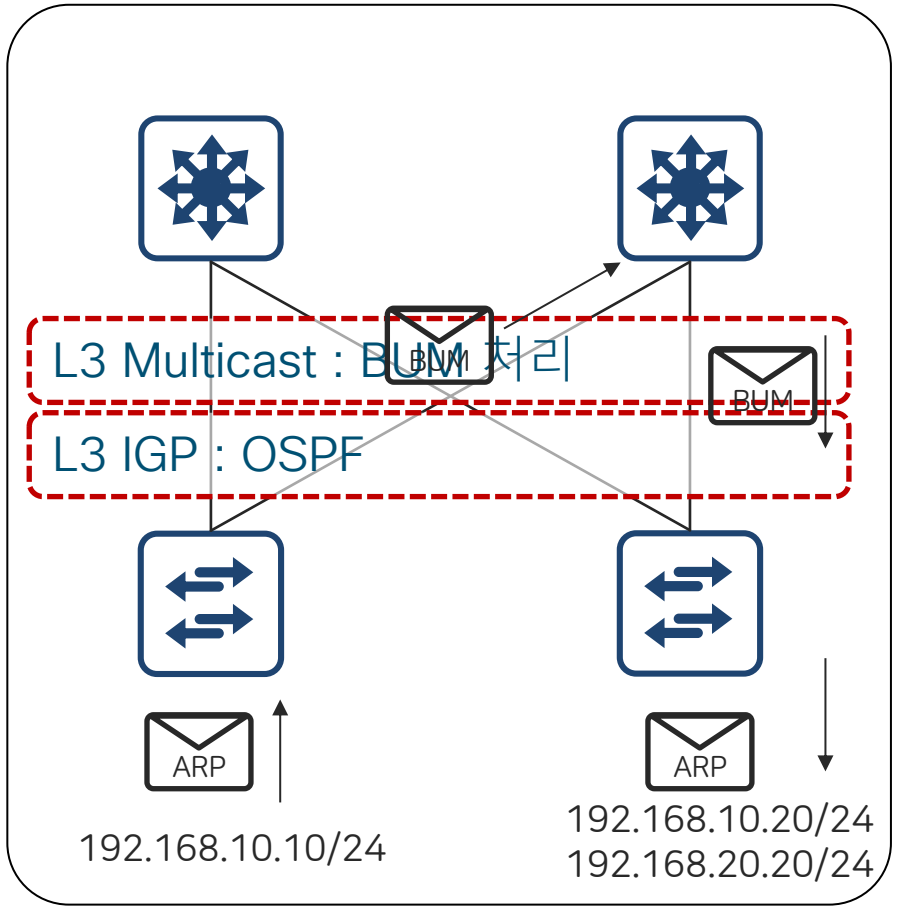
건설 시작. BUM 전달

계위별
장비 범례

BGW
Spine→

Leaf→

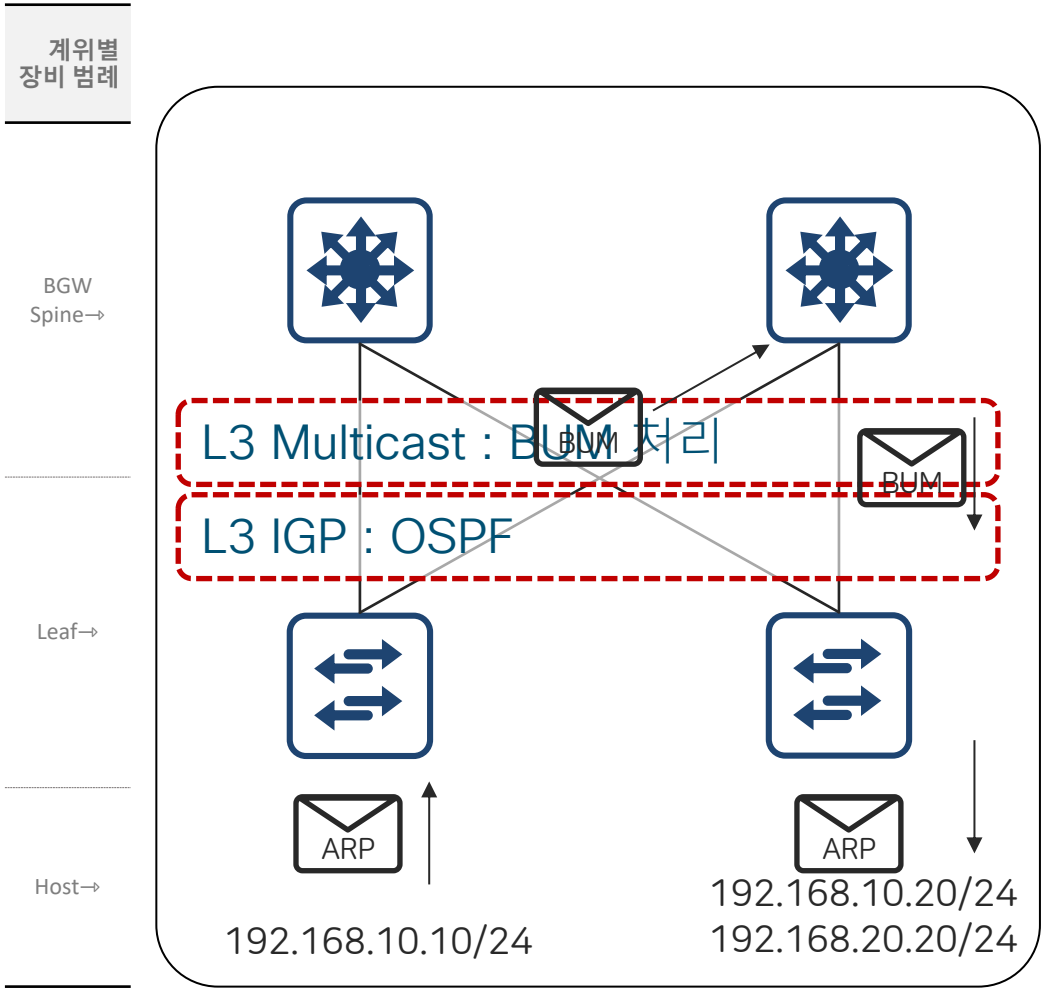
Host→



- VXLAN + Multicast 기술을 통해, BUM 트래픽을 L3통해 전달
- Multicast의 MROUTE TREE > ARP Flooding과 유사

: Multicast RPF Check(Incoming으로 Flooding없음 / Outgoing으로 Floodir

건설 시작. BUM 전달 – Legacy 대비 차별점 없는 통신 방법



- VXLAN + Multicast 기술을 통해, BUM 트래픽을 L3통해 전달
- Multicast의 MROUTE TREE > ARP Flooding과 유사

: Multicast RPF Check(Incoming으로 Flooding없음 / Outgoing으로 Floodir

결국 Legacy와 같이
매번 BUM Flooding이 필요한 비효율적인 방식

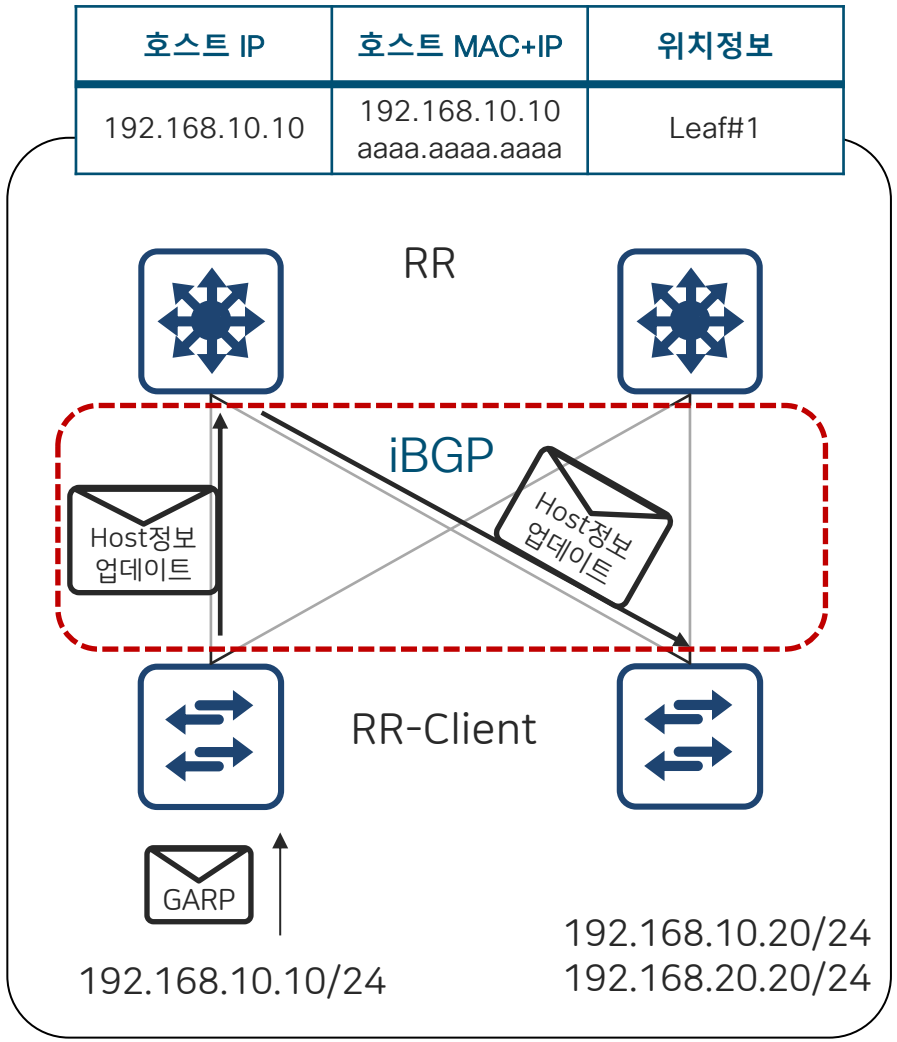
건설 시작. Control-Plane – MP-BGP

계위별 장비 범례

BGW Spine→

Leaf→

Host→



L3 Multicast : BUM 처리

L3 IGP : OSPF

MP-BGP EVPN 기술을 이용한 위치정보 교환 → iBGP Route-reflector

- Host의 위치정보 → 어떤 IP, 어떤 MAC+IP가 어느 Leaf에 수용되었는지?

→ 호스트의 위치정보를 알고 있으므로 불필요한 BUM 제거

Data Center Networking

VXLAN EVPN의 이해 - 아키텍쳐

= MP-BGP Control Plane / VXLAN Data Plane

Data Center Networking

VXLAN EVPN의 이해 - 아키텍쳐

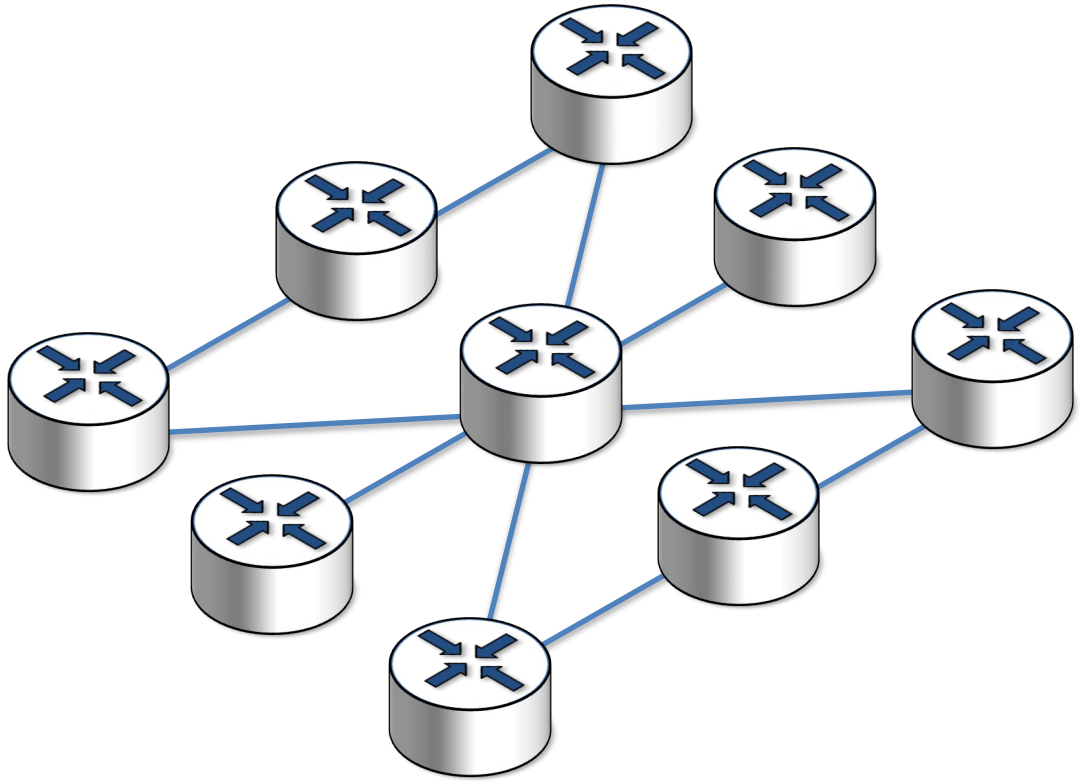
= MP-BGP Control Plane / VXLAN Data Plane

위치정보제공

데이터전달

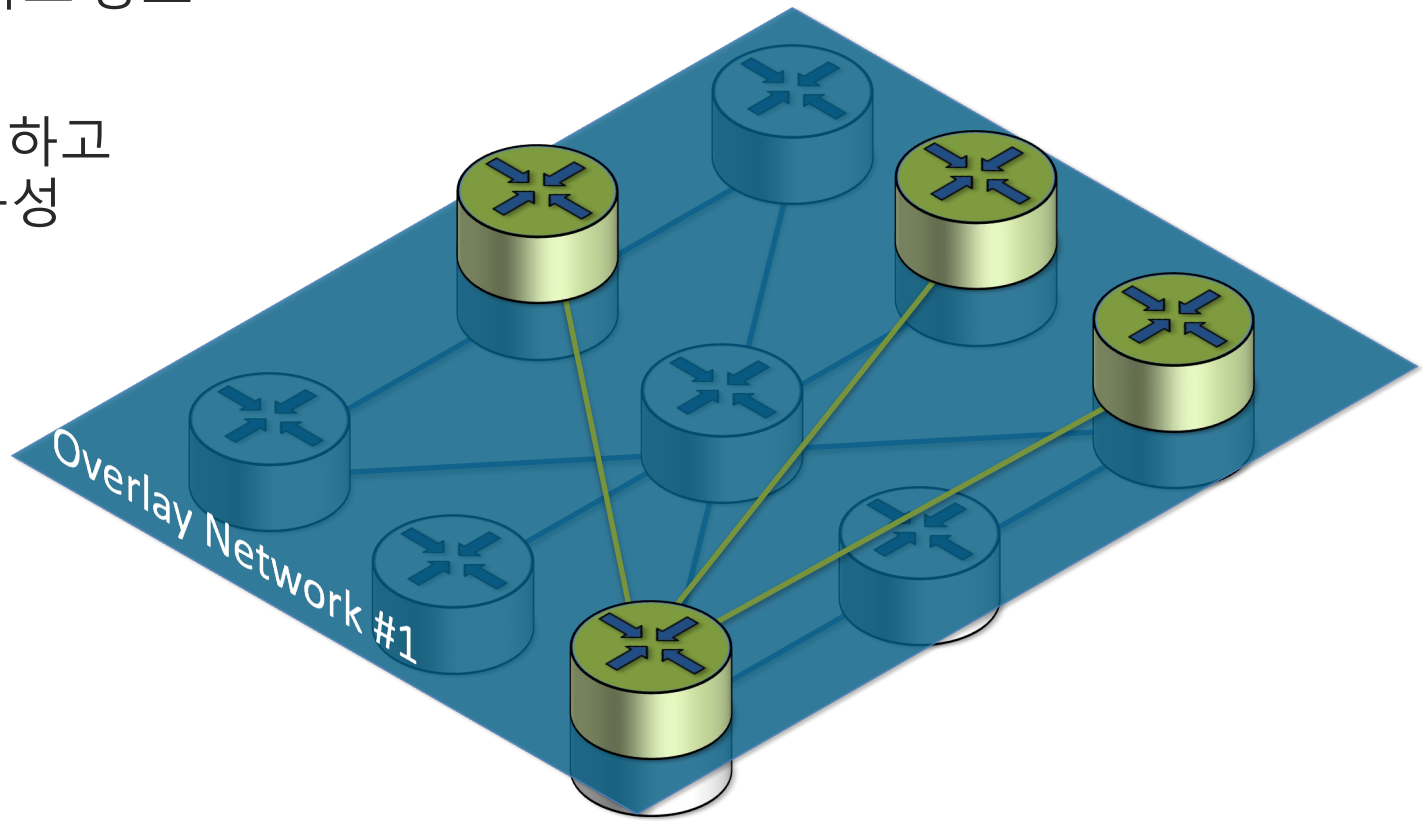
Overlay 네트워크 개념

- 네트워크 구축과 동일하게 물리적 스위치와 스위치 간 링크를 연결하고 상호 통신 가능하게 연결



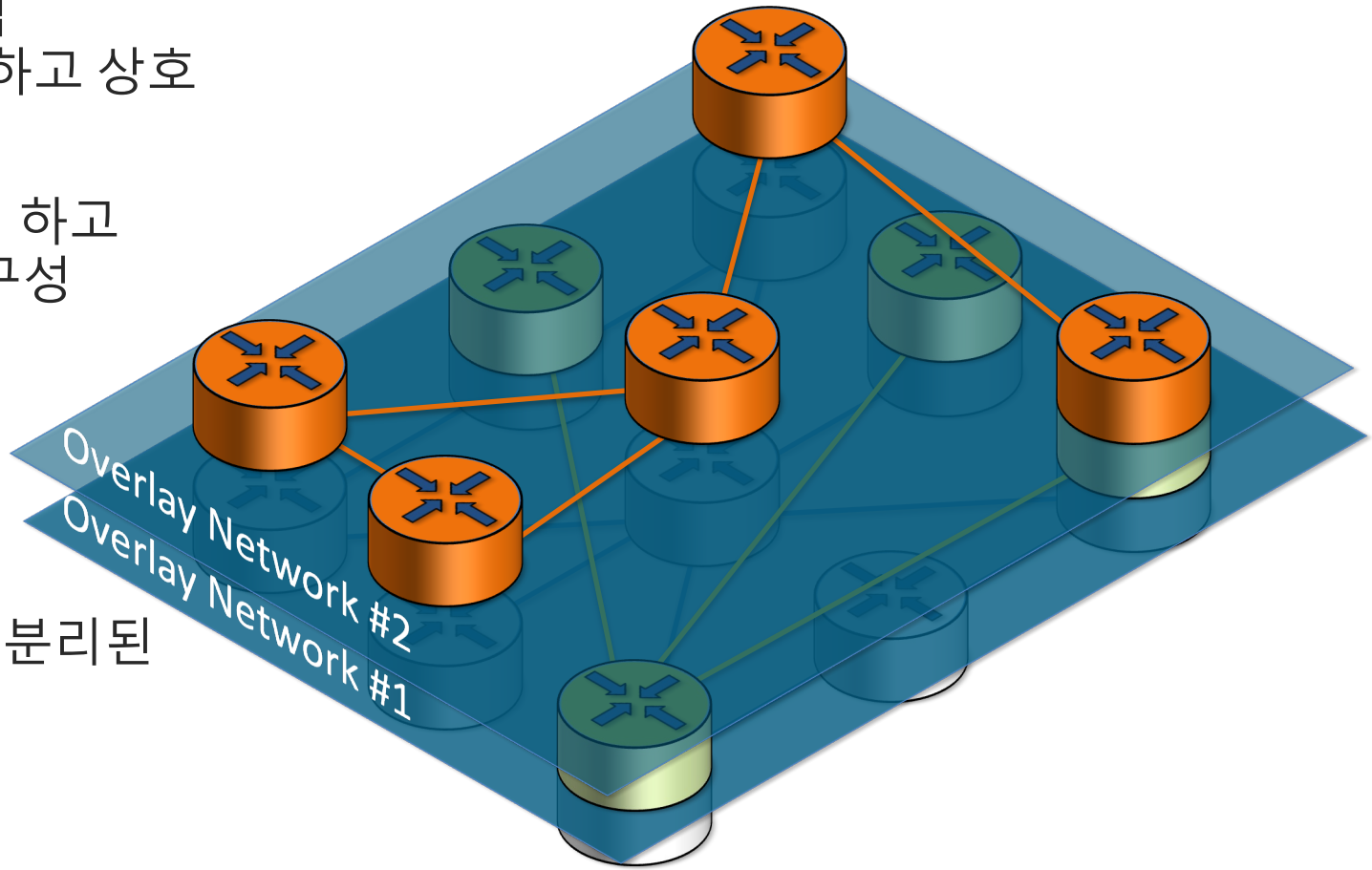
Overlay 네트워크 개념

- 네트워크 구축과 동일하게 물리적 스위치와 스위치 간 링크를 연결하고 상호 통신 가능하게 연결
- 논리적 네트워크인 ”Overlay” 구성 하고 그 장비들은 독자적인 토폴로지 구성
- Underlay는 Overlay에서 생성한 Data의 전달만 담당

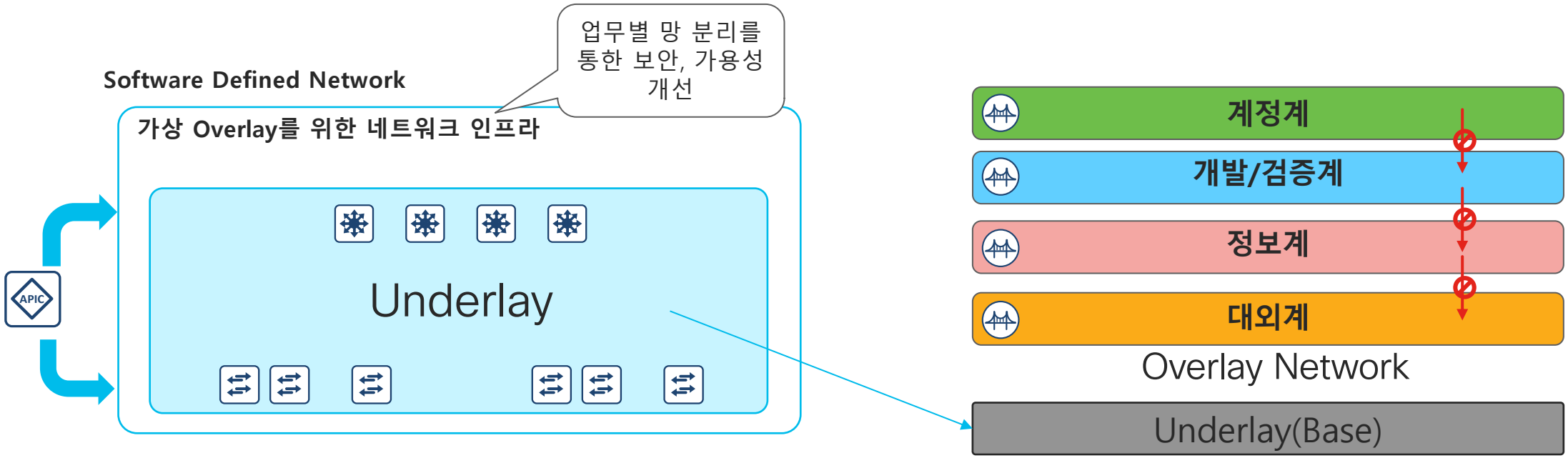


Overlay 네트워크 개념

- 네트워크 구축과 동일하게 물리적 스위치와 스위치 간 링크를 연결하고 상호 통신 가능하게 연결
- 논리적 네트워크인 ”Overlay” 구성 하고 그 장비들은 독자적인 토폴로지 구성
- Underlay는 Overlay에서 생성한 Data의 전달만 담당
- 여러개의 “Overlay”를 생성 할 수 있으며 이는 서로 다른 그룹 (고객, Tenant)를 위한 논리적으로 분리된 네트워크 구조를 제공



Overlay 네트워크 이해



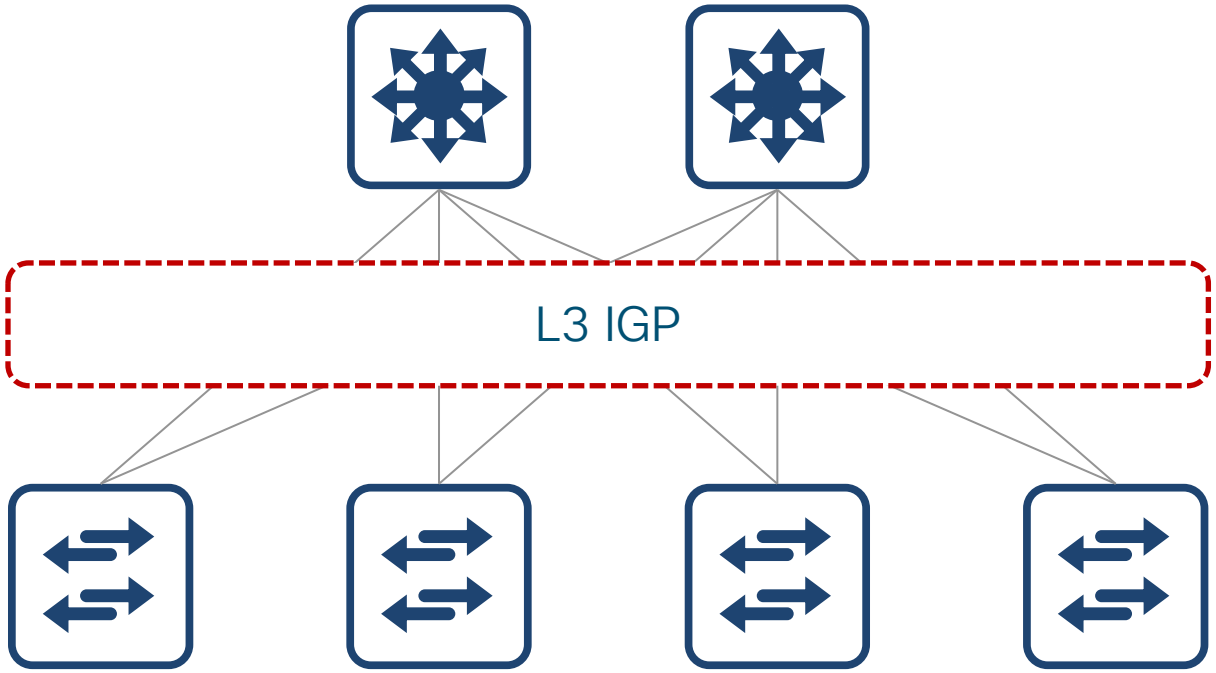
네트워크를 나누기 위한 방법론

VXLAN EVPN을 구성하는 기반 기술

범례	용도	효과	비고
VLAN	Subnet 구분	-	
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장	L2 VNID, L3 VNID
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding	pim sparse mode
IGP	L3 Underlay	STP 제거, Data-Plane	주로 OSPF 활용
VRF	망분리	동일 인프라에 가상 네트워크 구축	MAC-VRF /MAC-IP-VRF
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거	확장네트워크 연동활용

IGP 기반의 L3 Underlay

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거



L3 Underlay를 사용하는 이유

- 포트자원 낭비없이 최대한 활용
  - ✓ Failover/Failback에 따른 Forwarding Delay 제거
  - ✓ Ether-Channel없이, ECMP 기반 로드발란싱 구현

VXLAN의 이해

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거

VXLAN?

- ✓ VXLAN 은 Network Overlay 기술
- ✓ VXLAN 은 IP 기반의 Routed Network 위에 Layer-2 & Layer-3 overlay networks을 구성하는 기술
- ✓ VXLAN 은 MAC in UDP encapsulation 사용 (UDP destination port 4789)

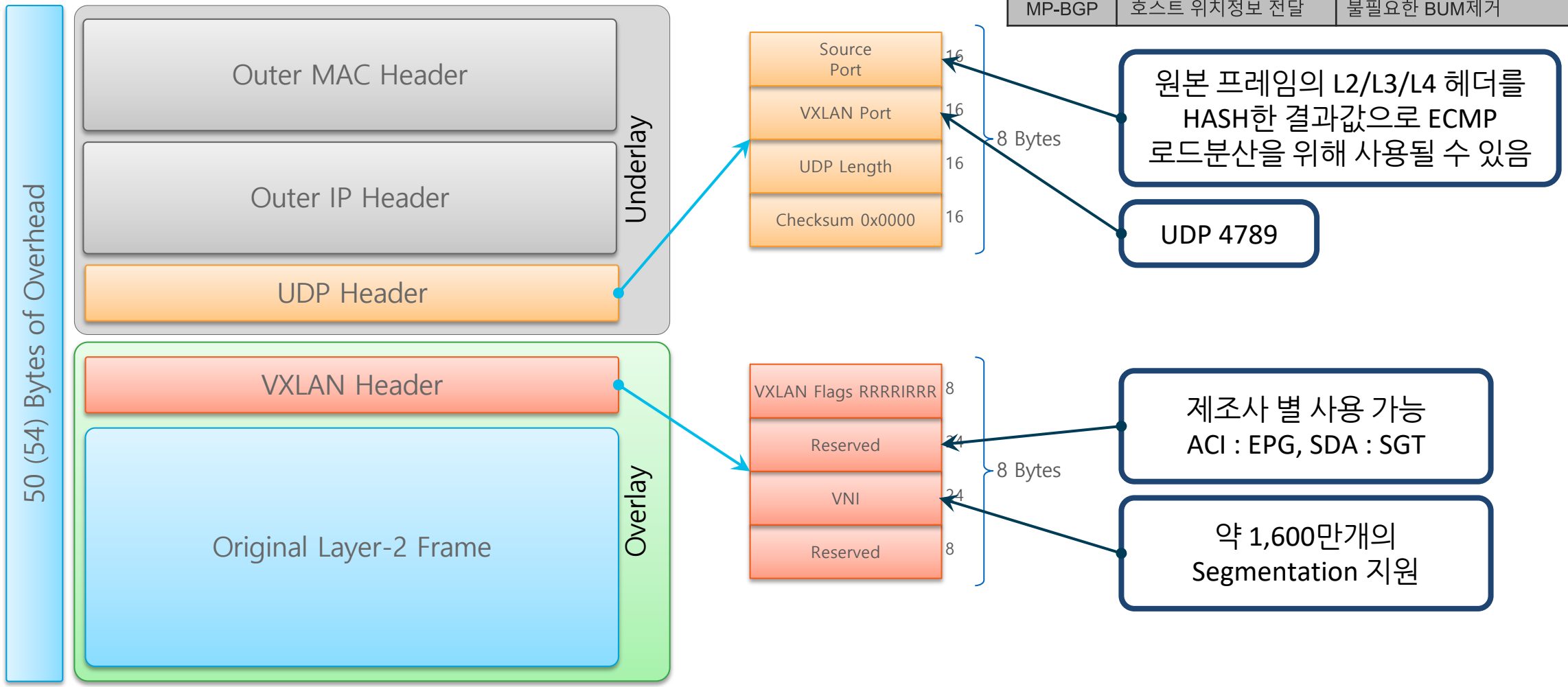
Why VXLAN?

- ✓ 표준 기반의 오버레이 기술
- ✓ ECMP 기반의 L3 환경을 그대로 유지/활용
- ✓ 멀티 테넌시를 위한 세그멘테이션 지원
- ✓ 주소공간(Name-Space)을 16M 으로 확장
- ✓ 물리 및 가상 네트워크 인프라를 통합
- ✓ SDN을 위한 기반 기술

VXLAN 는 네트워크에
세그멘테이션(Segmentation), IP 이동성(Mobility)와 확장성(Scale)을 제공

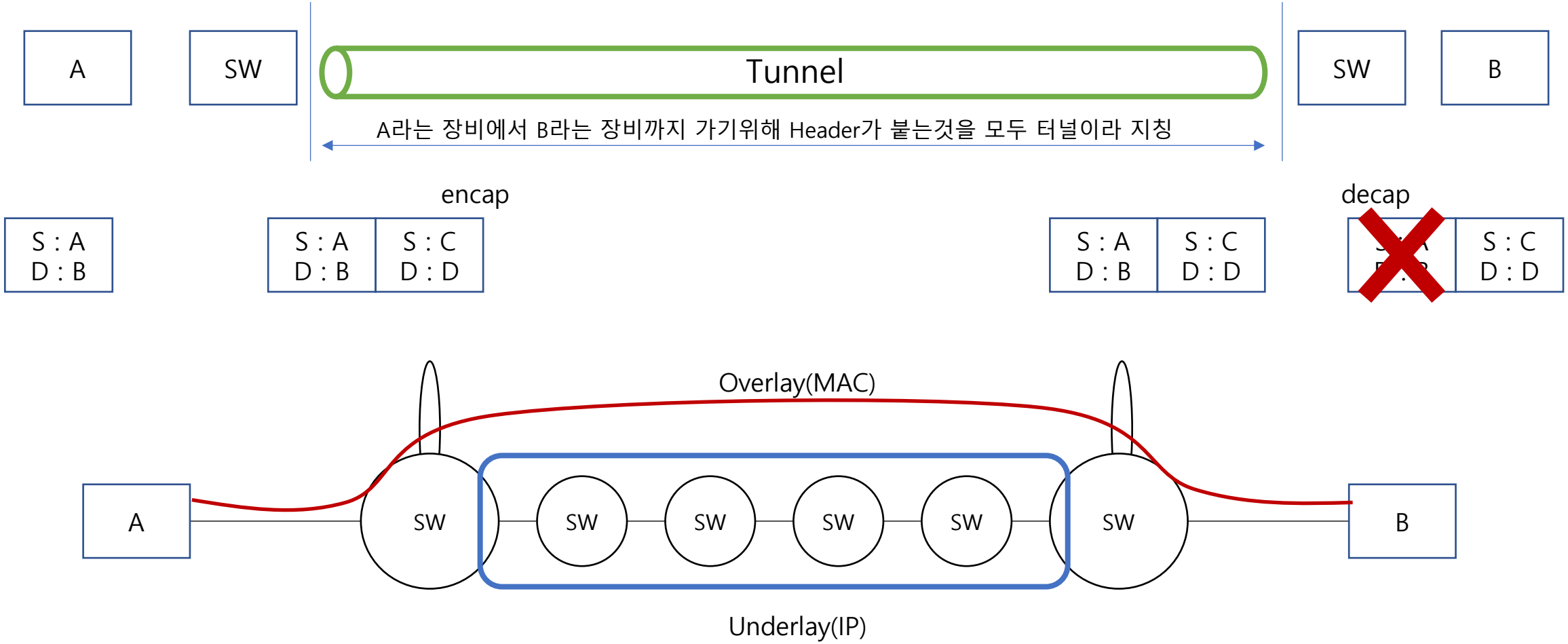
VXLAN의 이해

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거



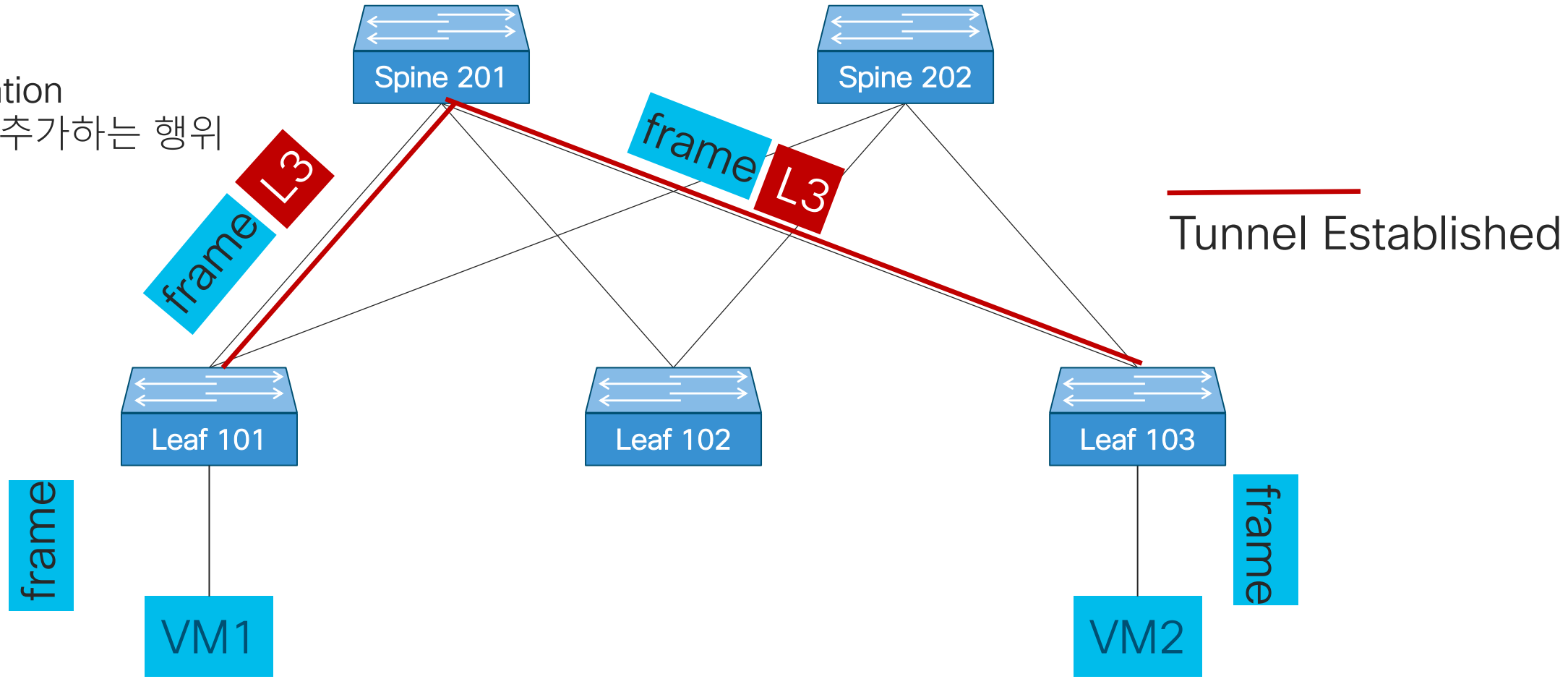
Tunnel의 이해

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거



Tunnel의 이해

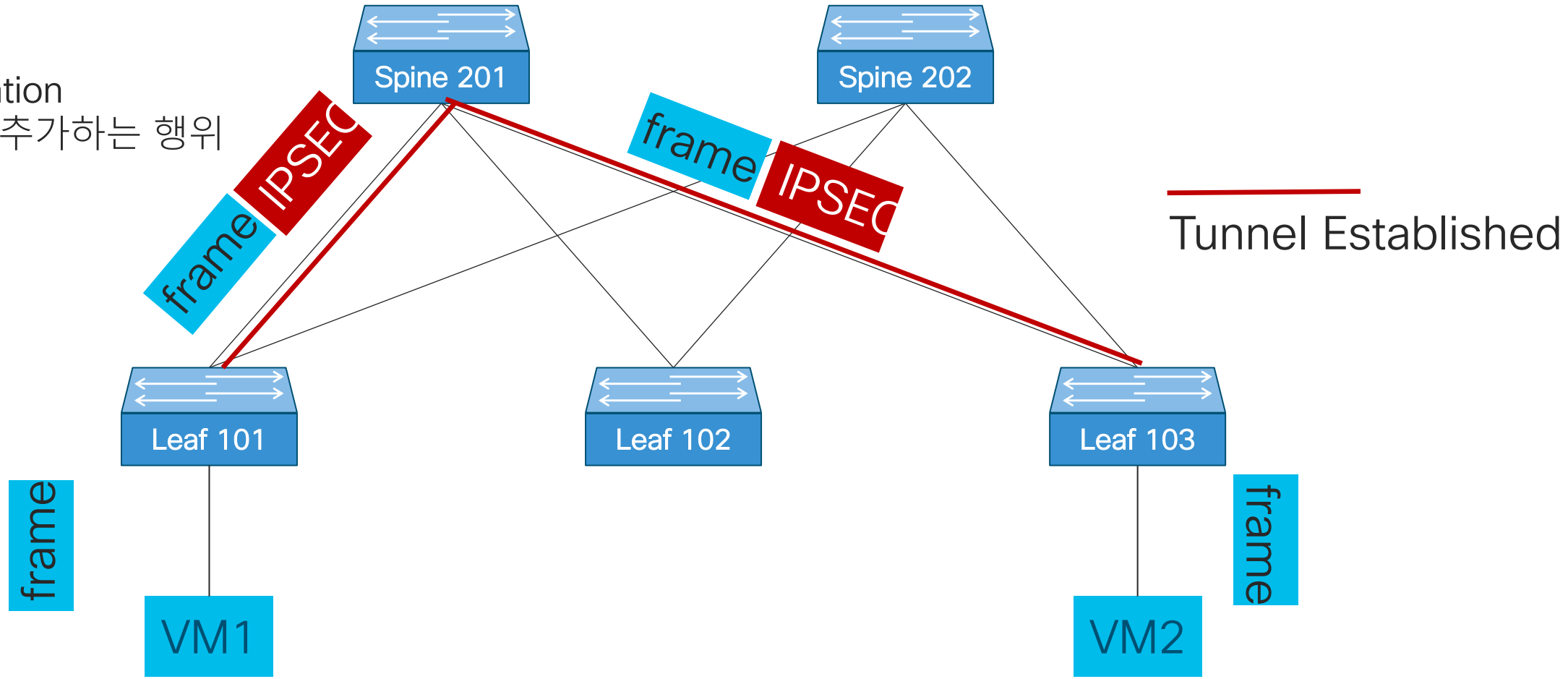
Tunnel
= Encapsulation
= header를 추가하는 행위



일반적인 Tunnel은 Session이 한번 생기면, Session이 Fin발생 전까지 변동 없음
= 생성된 Tunnel의 Traffic이 특정 경로를 지속적으로 점유

Tunnel의 이해

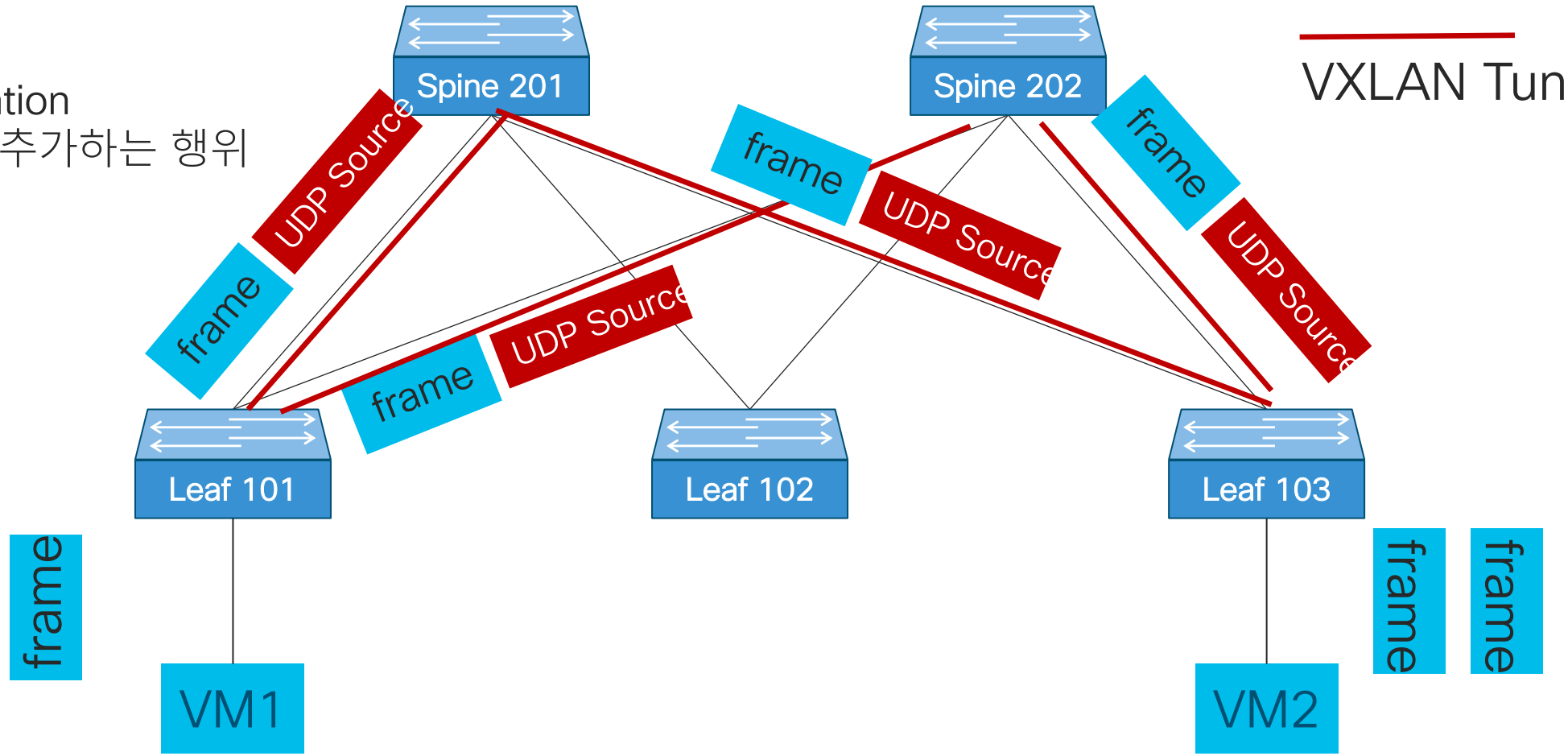
Tunnel
= Encapsulation
= header를 추가하는 행위



일반적인 Tunnel은 Session이 한번 생기면, Session이 Fin발생 전까지 변동 없음
= 생성된 Tunnel의 Traffic이 특정 경로를 지속적으로 점유

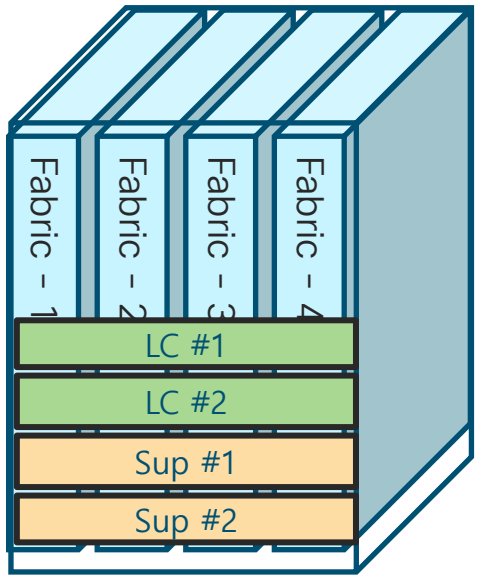
VXLAN Tunnel의 이해

Tunnel
= Encapsulation
= header를 추가하는 행위

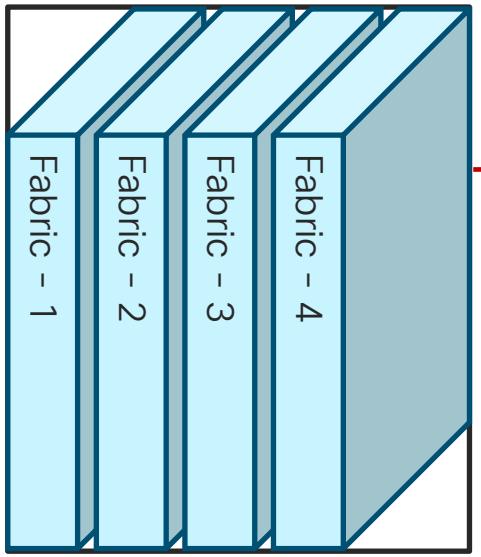


VXLAN Tunnel은 UDP의 Source Port를 기반으로 VTEP간 Load Balancing

Fabric구축의 목적 = 거대한 모듈러 스위치를 만드는 것

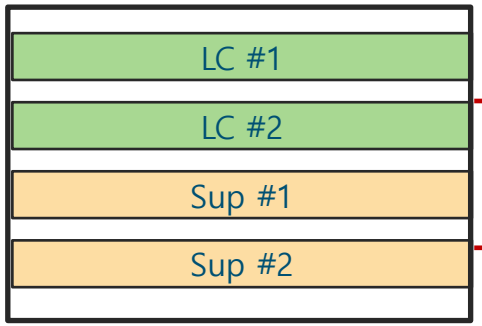


모듈러 스위치



Chassis(Back)

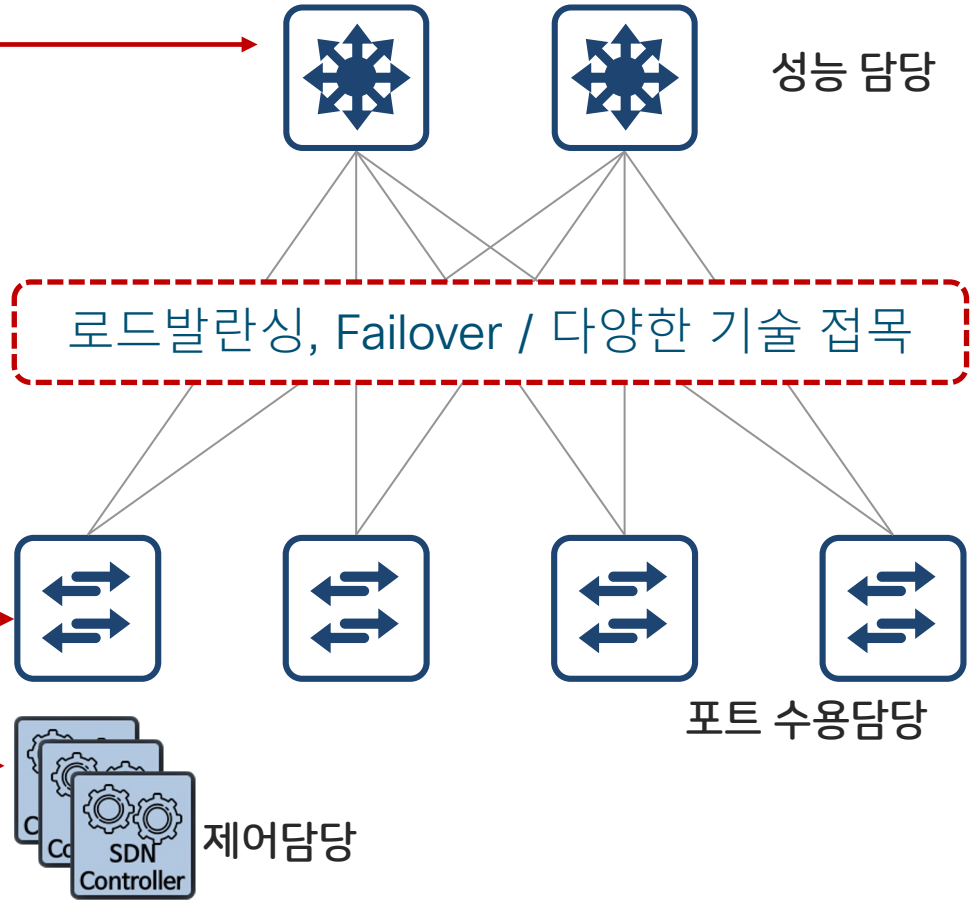
Fabric Module → Spine



Chassis(Front)

Line Card→ Leaf

Supervisor → Controller

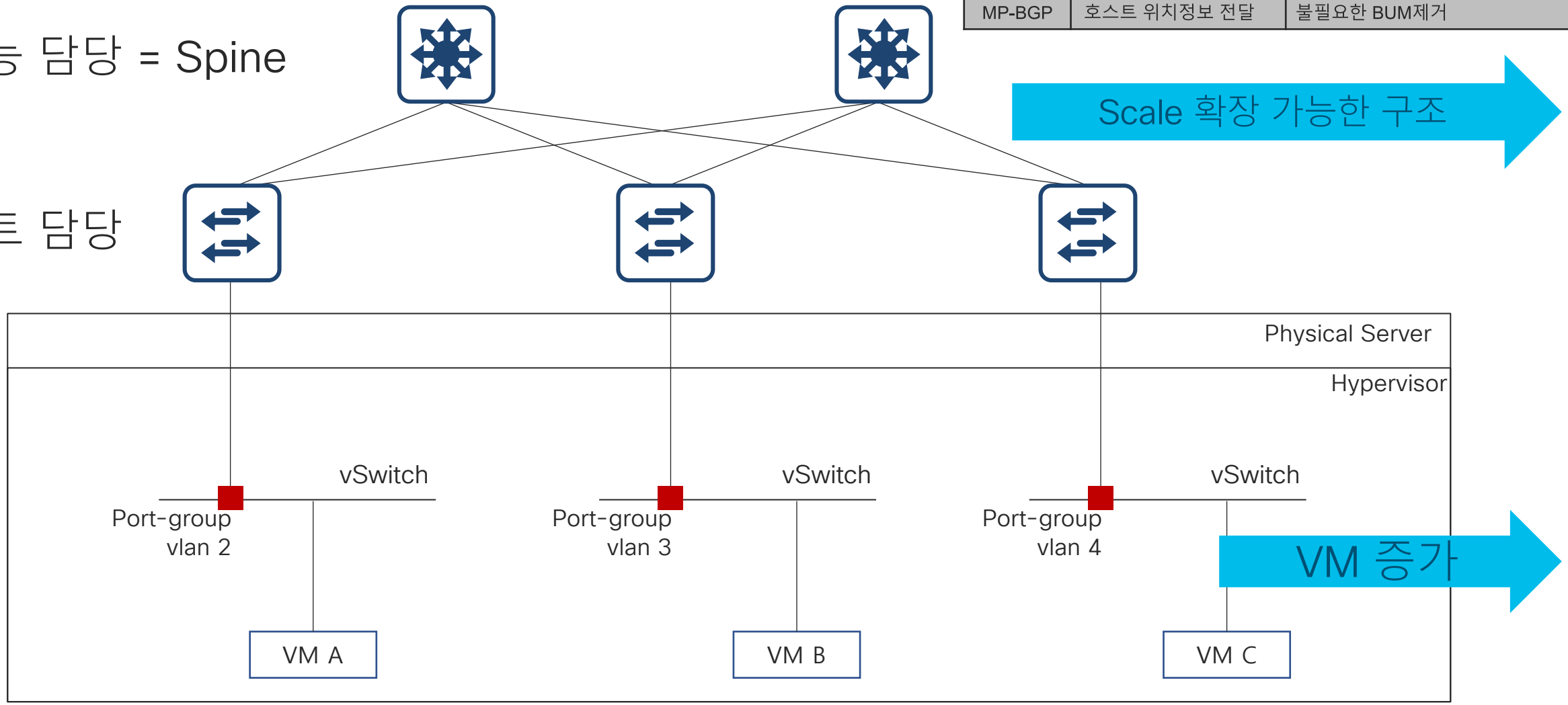


SPINE/LEAF = 거대한 모듈러 스위치

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거

성능 담당 = Spine

포트 담당



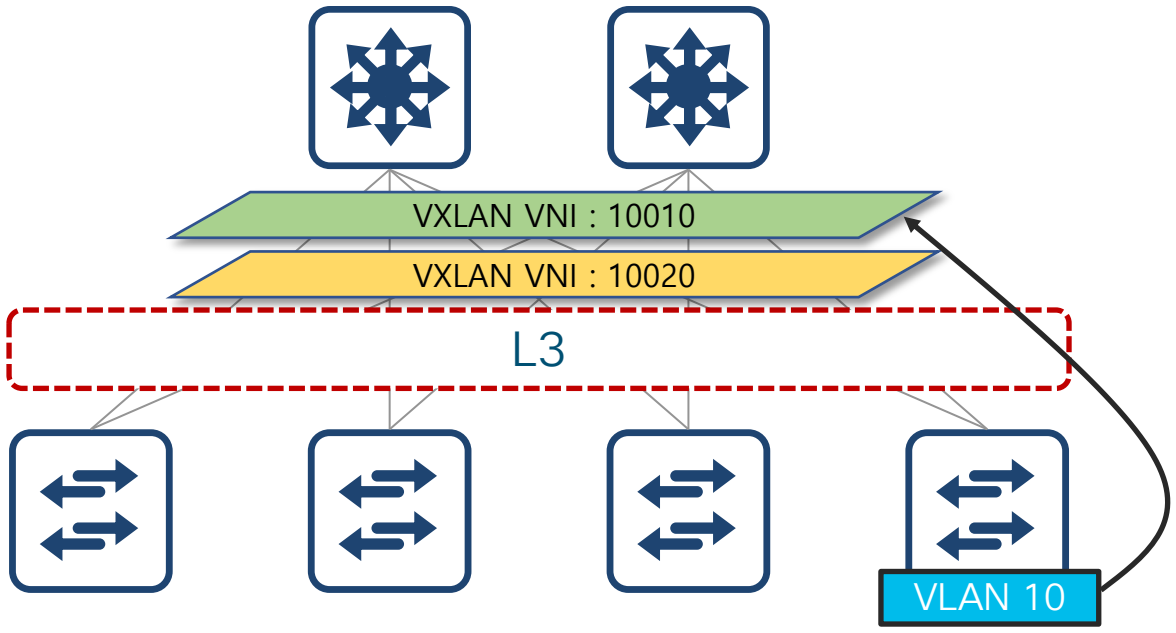
Data Center Networking

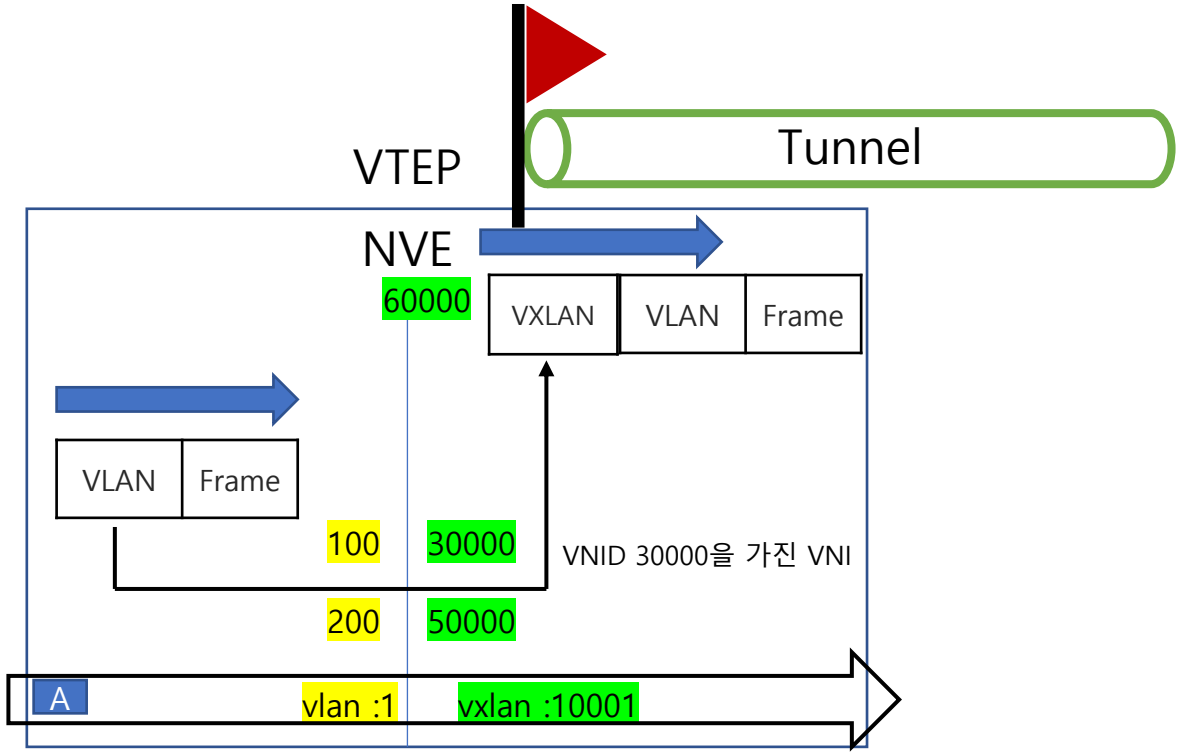
VXLAN EVPN의 이해 – Protocol Behavior

= MP-BGP Control Plane / VXLAN Data Plane

BUM 처리를 위한 Multicast 활용 - VLAN 별 VXLAN ID 매핑

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거

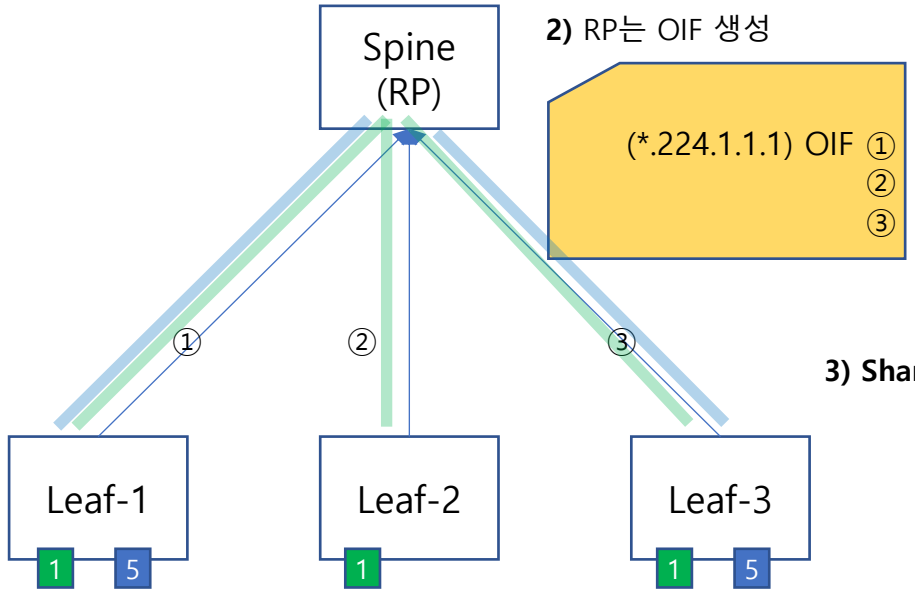




VLAN은 L3 Network를 통과할 수 없음 > VXLAN 맵핑을 통해 해결

BUM 처리를 위한 Multicast 활용

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거

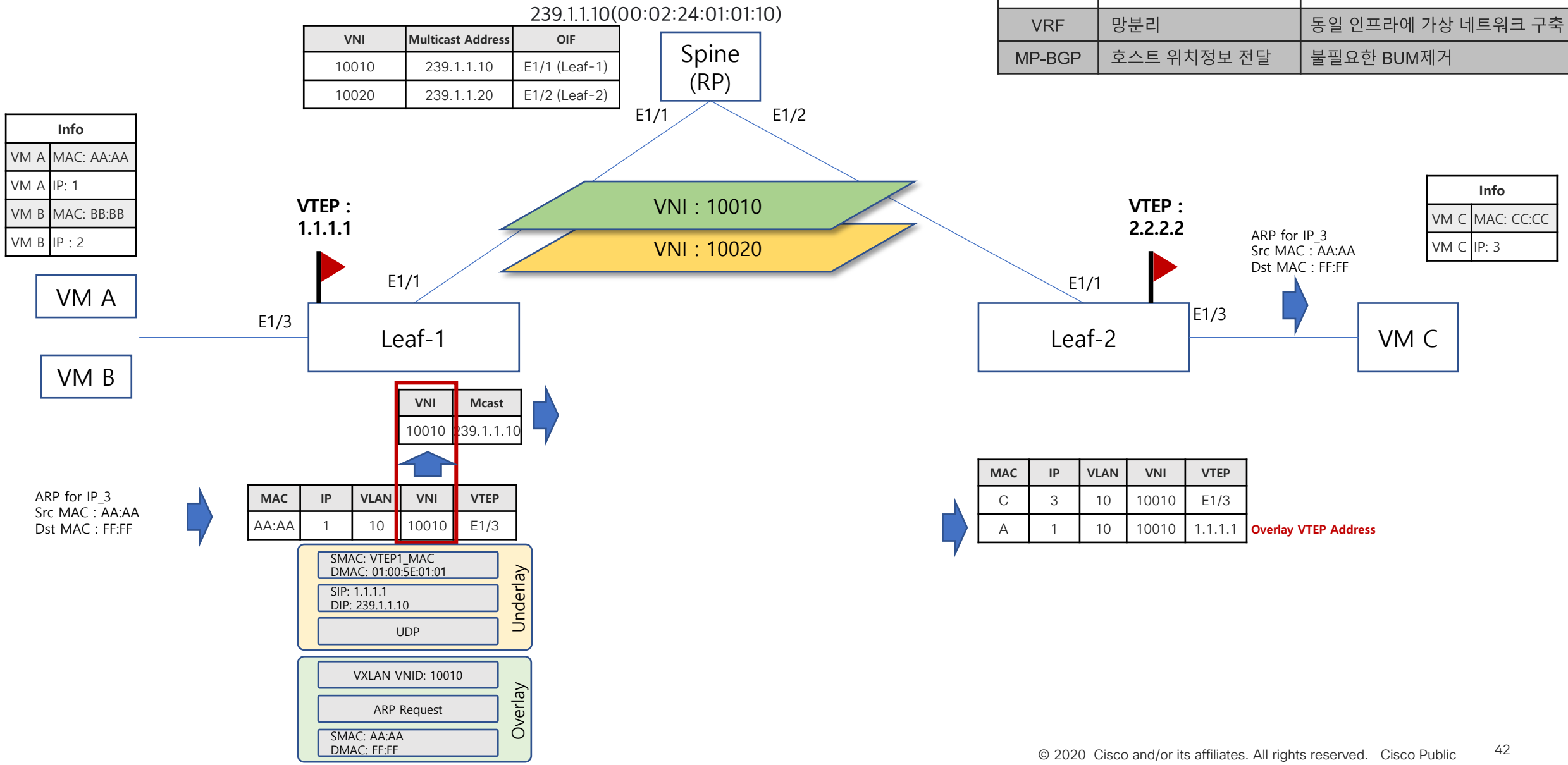


4) Multicast는 Tree에서 특성상
자기가 받은 쪽이 아닌 다른 OIF로 트래픽을 던지기 때문에
Leaf-3이 던진 Multicast는 Leaf 1,2번으로 Multicast 전달

✓ **VNID 마다 구분이 되는 멀티캐스트 주소가 생성되고, 이 VNID의 멀티캐스트 주소로 RP에게 join**

BUM 처리를 위한 Multicast 활용

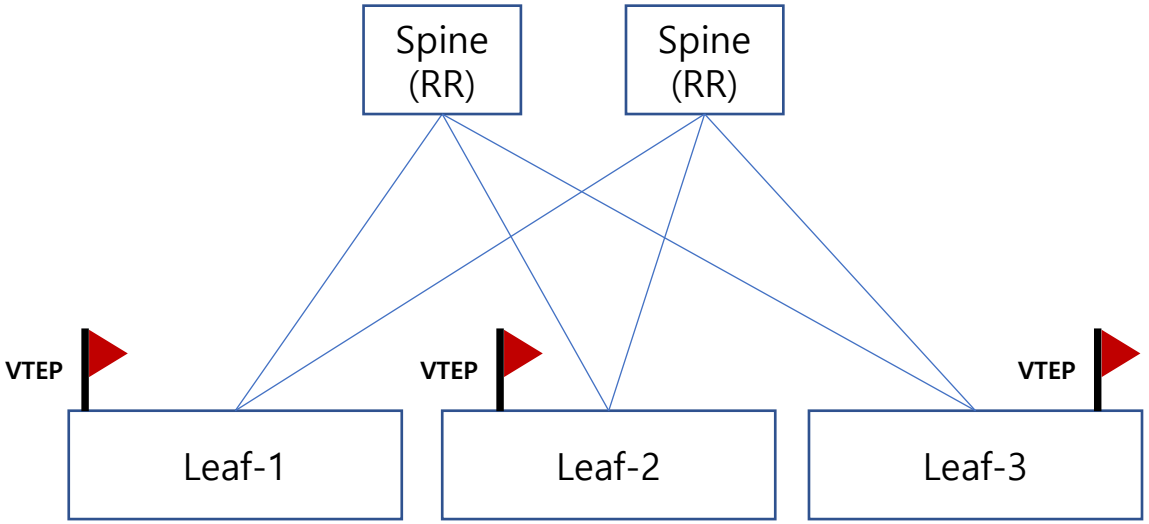
범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거



MP-BGP기반의 Control-Plane

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거

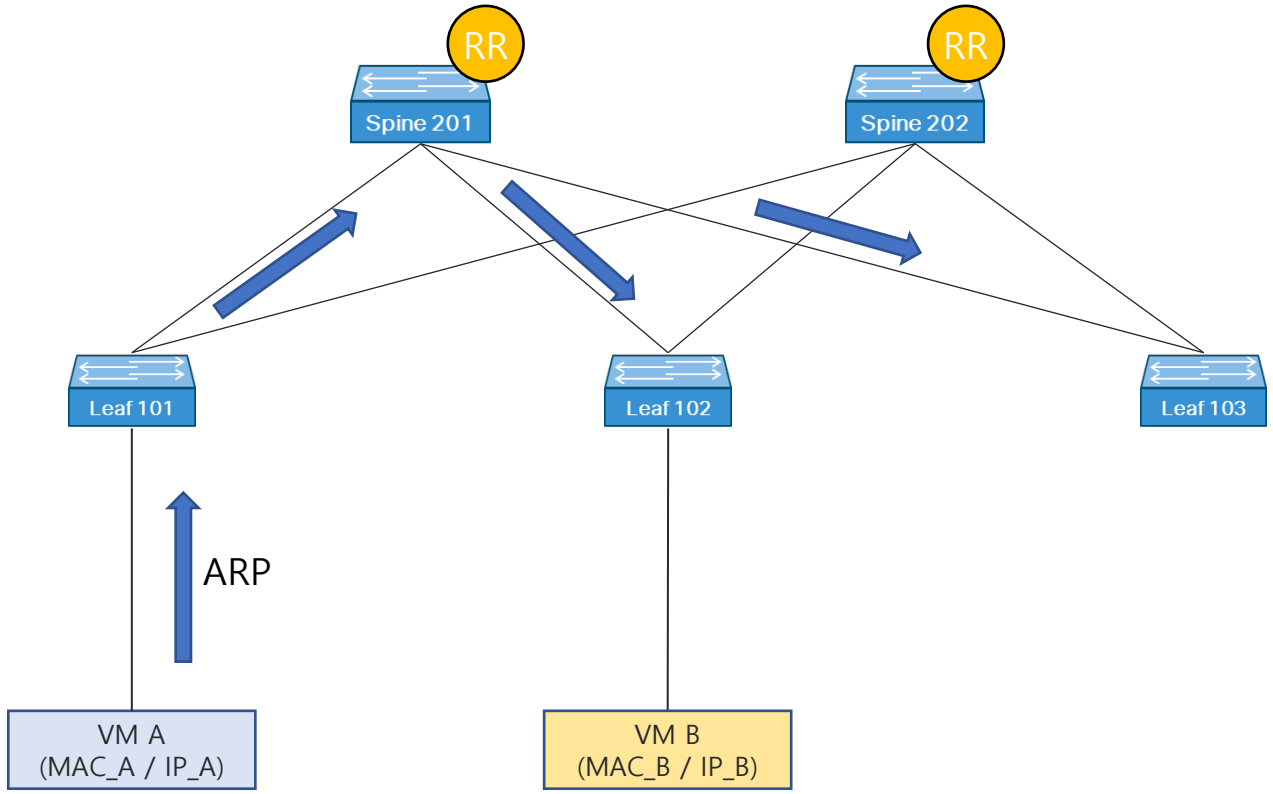
- 호스트 Route와 Underlay protocol과의 완전한 분리
- Route Reflector는 확장/최적화를 목적으로 사용
- 호스트/서브넷 경로 및 fabric 외부 경로 reachability 를 위해 Leaf 노드에서 MP-BGP 사용



MAC, IP, Subnet을 공유할 수 있는 MP-BGP

MP-BGP기반 Learning & Distribution

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거



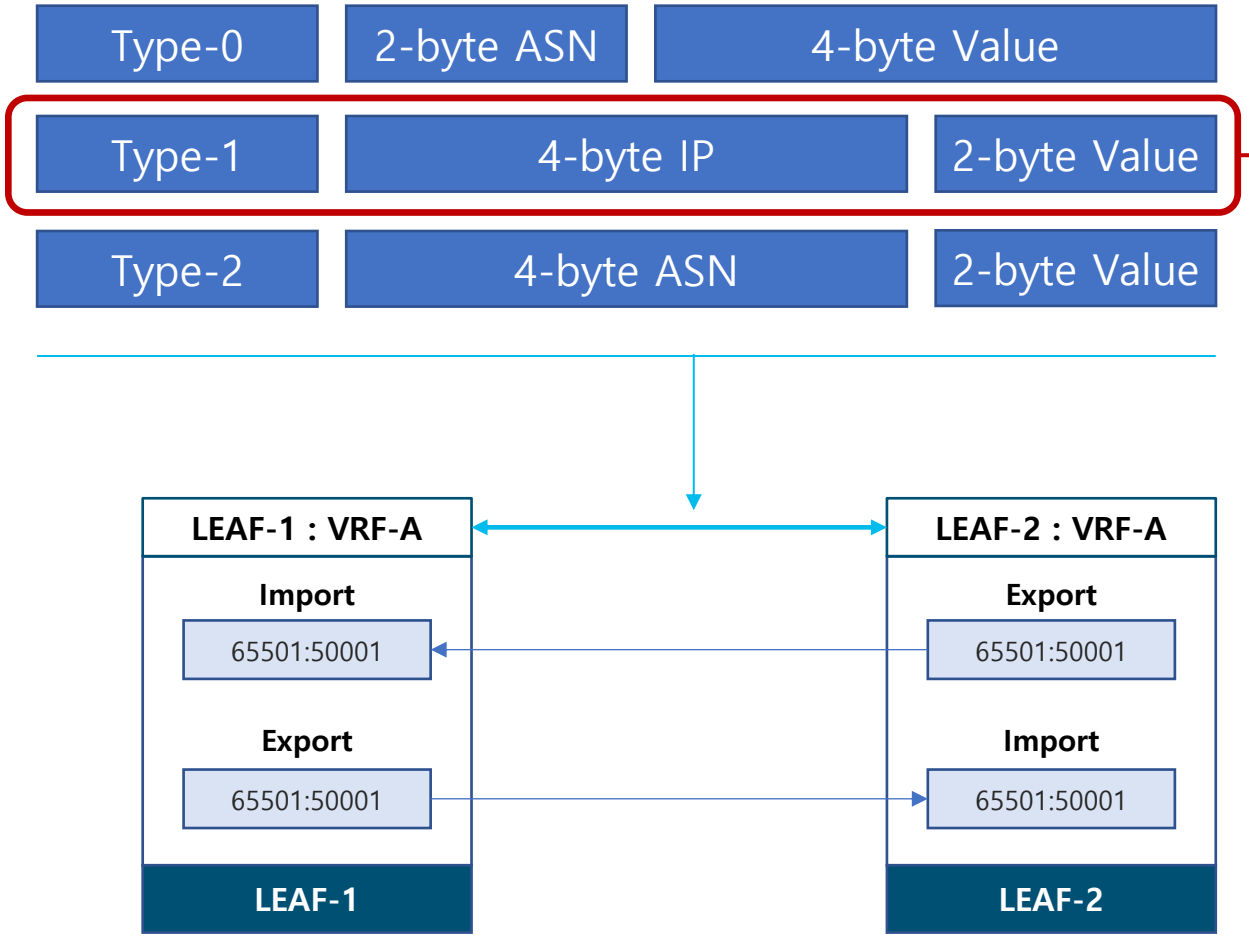
ARP Suppression & Learning

BGP Control-Plane은 host정보를 LEAF가 모두 소유 하여 Leaf가 대신응답
→ ARP Suppression
→ EVPN 기본동작

Table에 없으면 Multicast를 통한 BUM 전달

VXLAN
– Route Distinguisher Format & Type

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거



RD = Auto 로 선언 시 Type-1 사용

- VTEP IP : VRF ID == Ex) **1.1.1.1 : 3**

```
DC1-LEAF-101# show vrf
 VRF-Name                           VRF-ID State    Reason
 black-hole                              3 Up       --
 common:Accordion-VRF                   16 Up       --
 common:L3Out                           21 Up       --
 common:mykube-vrf                       6 Up       --
 common:OCP-VRF                          5 Up       --
 CommonInfra:A                          18 Up       --
 HX-iSCSI:A                             10 Up       --
 management                              2 Up       --
 mgmt:inb                               14 Up       --
 Nicepayments:NICE_TESTVRF               7 Up       --
 overlay-1                               4 Up       --
 SD-WAN-SiteA:A                         12 Up       --
 SDA:SDA                                 8
 ServiceGraph_Test:A                    13
 T1-Hyundai_Dep_Group:VRF-01            11
```

RT = Auto

- ASN : VNI == 동일 BGP AS : 50010([ex]vlan10)

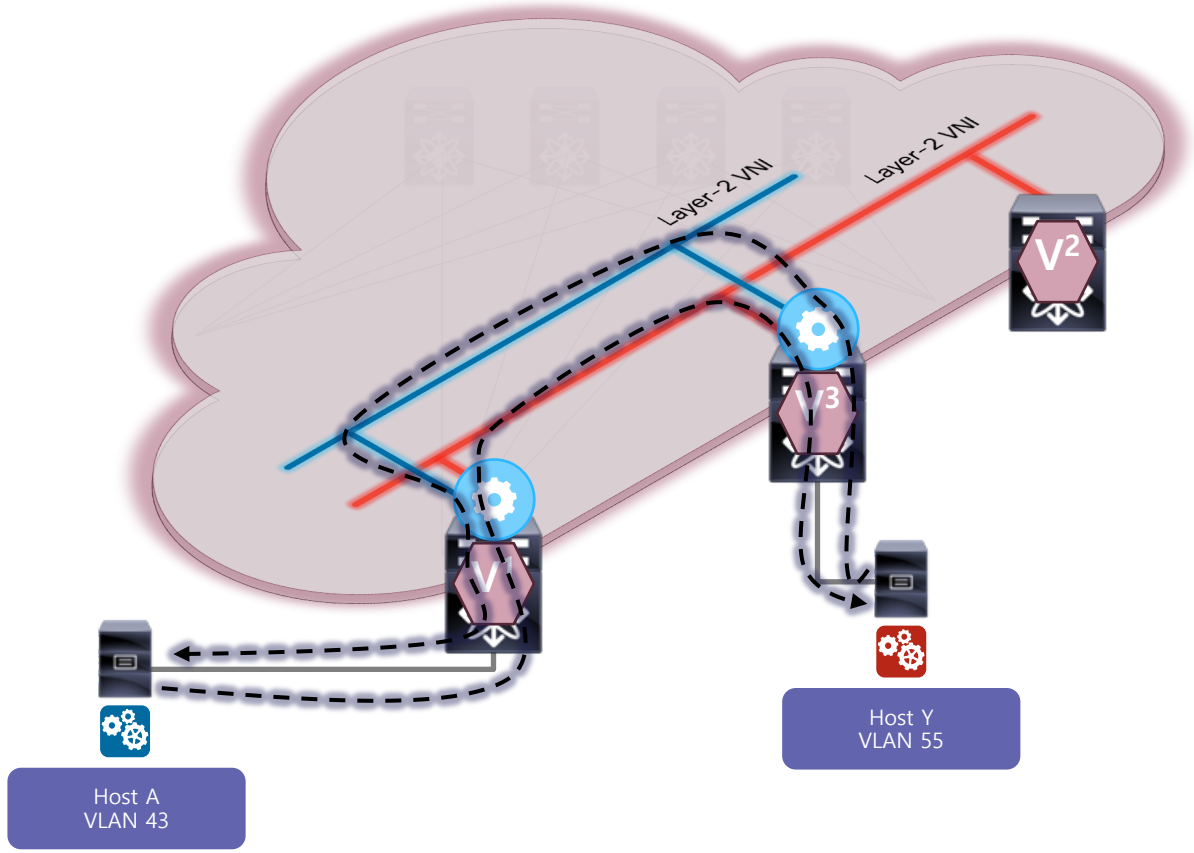
- BGP AS: EBGP의 경우 BGP AS가 달라서 Auto 사용 불가
- VNI : VNI로 ID를 만듬

- ASN : VNI == Ex) **65501 : 50010**

>> 동일 IBGP 65501내에서 50010을 가진 스위치 간 경로 수신/송신

VXLAN – Asymmetric IRB

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거



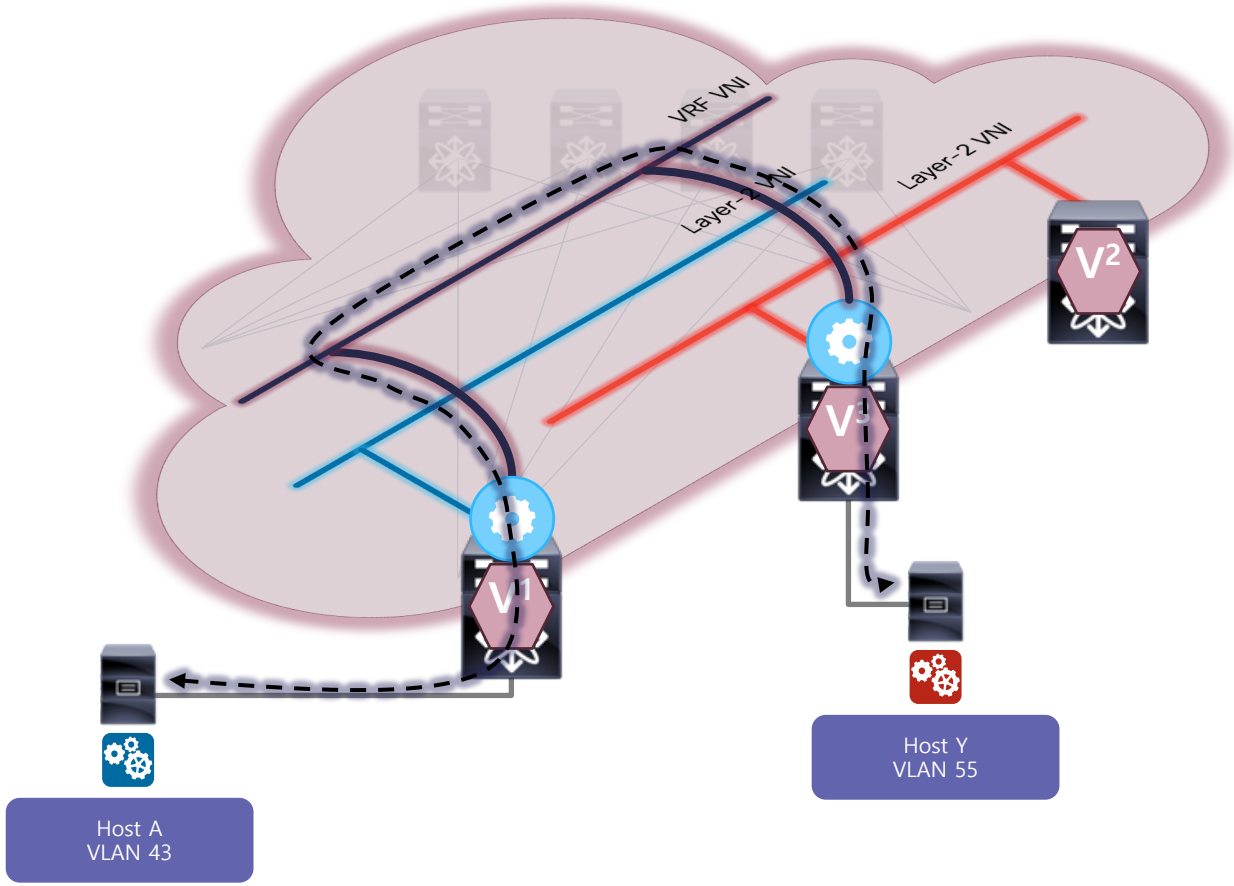
- **Asymmetric**
  - Inter-VLAN Routing과 유사
  - Source와 Destination VNI는 라우팅 해야하는 모든 Leaf에 있어야만 함
  - 라우팅 이후의 트래픽은 Bridged 트래픽과 Share
  - 분산형 라우팅 방식에 적합하지 않음

  Host A 의 VLAN/VNI “blue” 에서 “red”로 가는 트래픽은 V^1 에서 라우팅

  Host Y 의 VLAN/VNI “red” 에서 “blue”로 가는 트래픽은 V^3 에서 라우팅

VXLAN – Symmetric IRB

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거



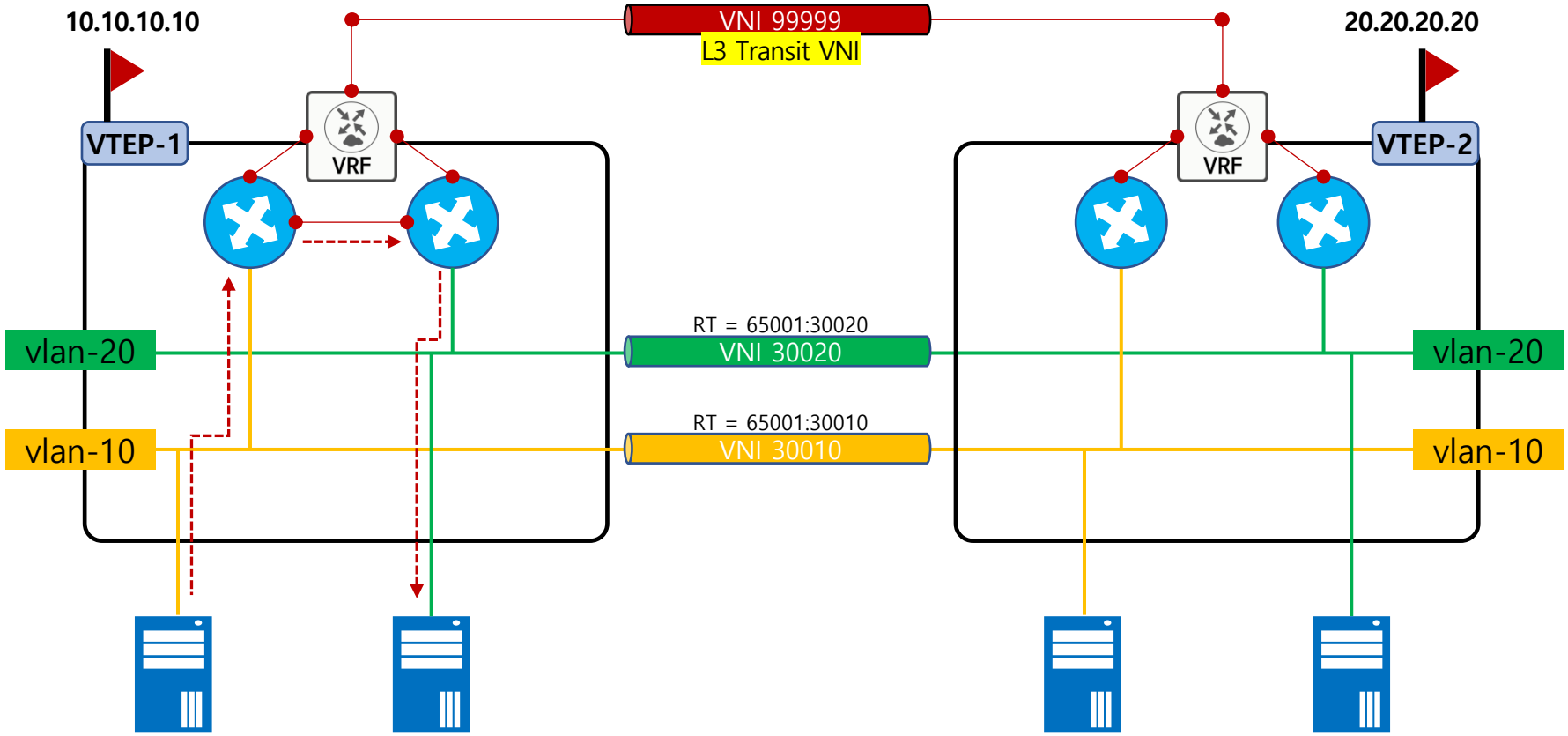
- **Symmetric**
  - Source와 Destination VNI가 어디 있던지 관계 없음
  - 라우팅 이후의 트래픽은 Bridged 트래픽과 별도로 분리
  - 라우팅 트래픽을 위한 별도의 VNI (Per VRF)

    Host A 의 VLAN “blue” 에서 “red”로 트래픽을 보내기 위해 V^1 에서 “purple” VNI로 전송

    Host Y 의 VLAN “red” 에서 “blue”로 트래픽을 보내기 위해 V^3 에서 “purple” VNI로 전송
  - **Cisco VXLAN/EVPN 에서 사용**

VXLAN – Symmetric IRB

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거



L3 Transit VNI = VRF

>> L3 전용 65001 : 9999

L2 VNI = MAC-VRF

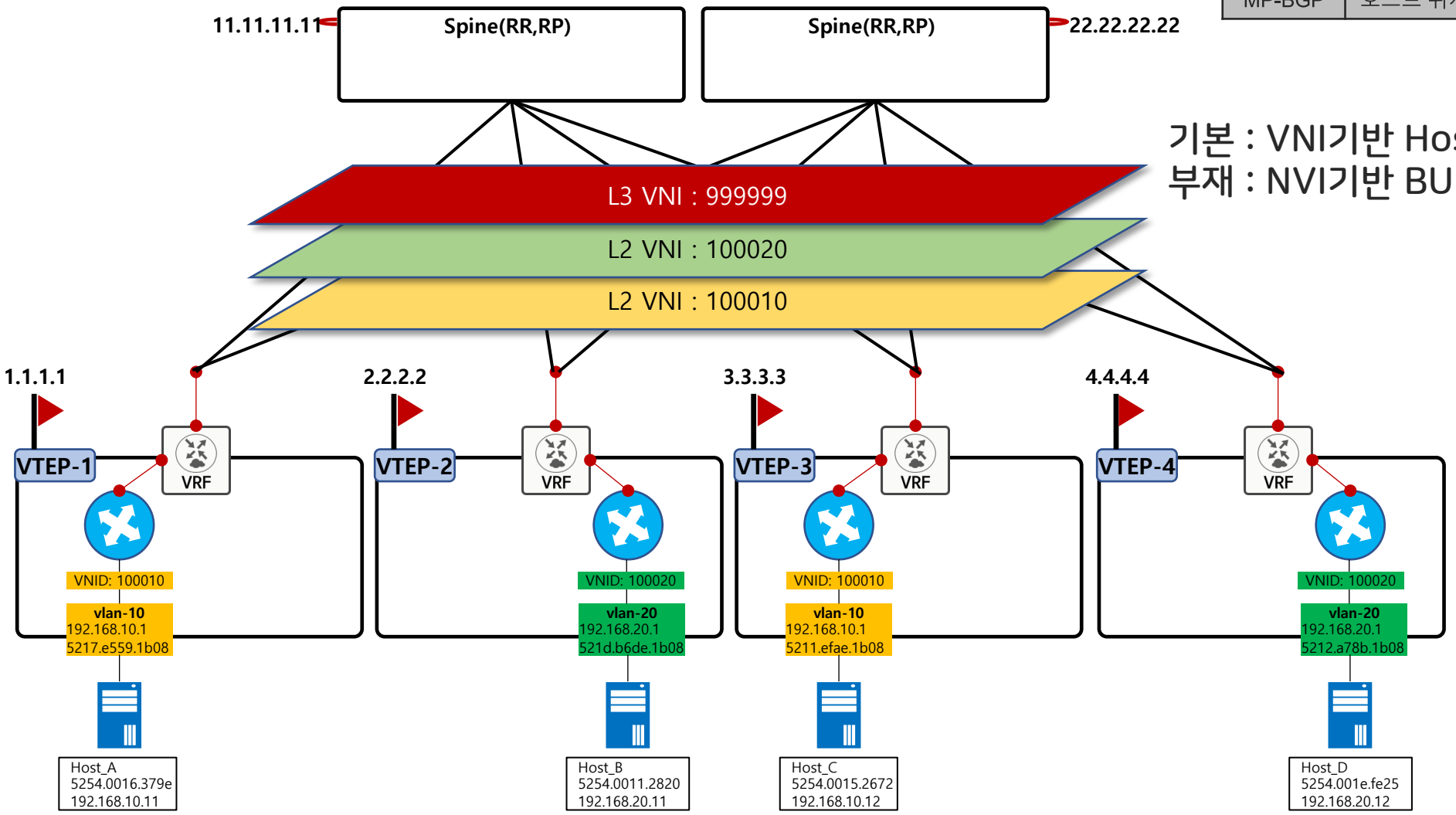
>> L2 전용 65001 : 32767+VLAN

NDFC
12.1.1 (여기서 시작)
L2 L3 VNI 합쳐서 1500개
12.1.2(고객 업그레이드 필요)
합쳐서 2500

NDO
VRF 40 개

VXLAN – Symmetric IRB

범례	용도	효과
VXLAN	더욱 세밀한 Subnet구분	더욱 세밀한 Segmentation과 L2확장
IGP	L3 Underlay	BUM 제거, STP 제거, Data-Plane
L3 Multicast	BUM 전달	지정된 대상에게만 BUM Forwarding
VRF	망분리	동일 인프라에 가상 네트워크 구축
MP-BGP	호스트 위치정보 전달	불필요한 BUM제거



기본 : VNI기반 Host Table 참조 - MP-BGP
부재 : NVI기반 BUM Forwarding - Multicast

Cisco CISG

감사합니다.

CISG

심플한 SDN 운영과 구축 – NDFC소개

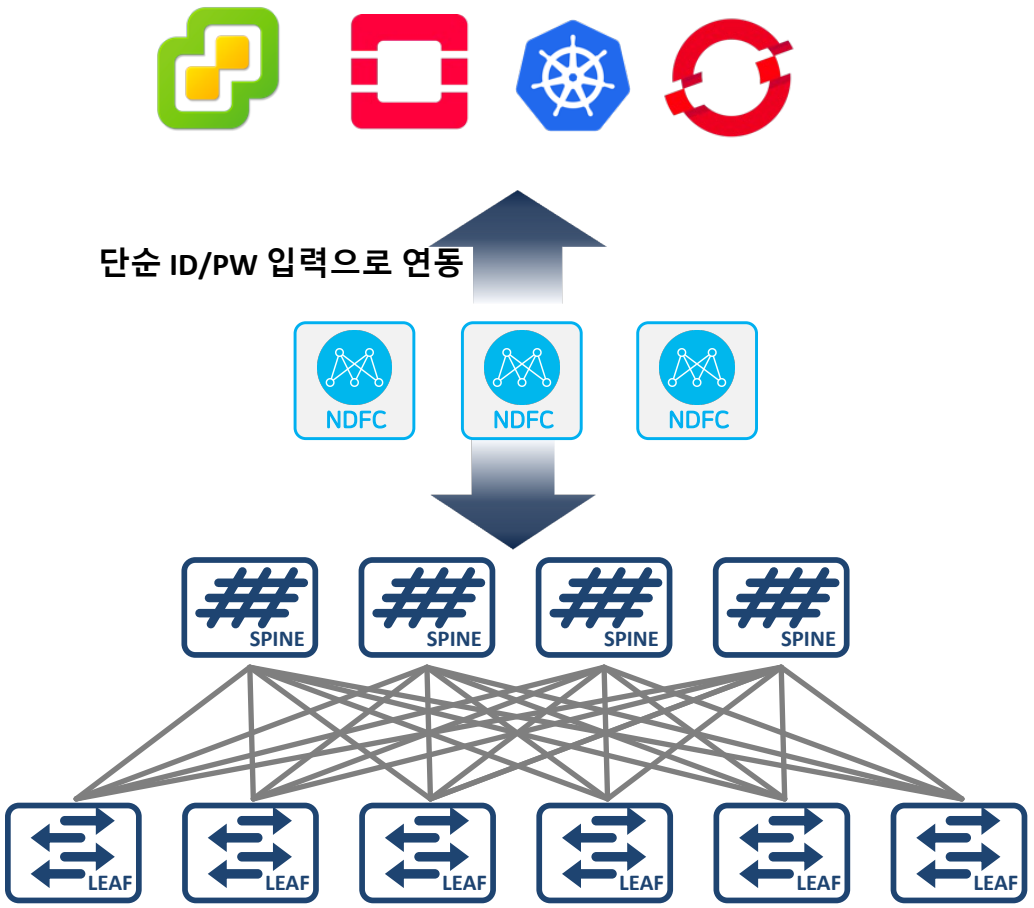
Cisco Systems Korea

NDFC 소개

- Nexus Dashboard Fabric Controller

NDFC(Fabric Controller) 기반의 네트워크 운영과 서비스 연계

유연한 방식의 Controller



- **중앙 집중형 관리 모델**
  - 트래픽에 개입 하지 않는 정책형 컨트롤러
  - 타 시스템 연계를 위한 개방형 모델 (REST API, Plug-ins)

- **변화 된 네트워크 구조**
  - Scale Out, 가용성 향상을 위한 Spine-Leaf
  - East-West 트래픽을 고려한 40G/100G/400G 의 하드웨어 기반의 고 대역폭 처리
  - 하드웨어 기반의 분산형 제어부 및 오버레이 처리로 네트워크 안정성 향상
  - 기존 환경을 포함하고 VM, Container, Openstack 등 새로운 서비스 가시성 지원

내부 구성원들이 느낄 수 있는 인프라 투자가 필요합니다.

Start with Better Switch – Nexus 9K

ASIC

Datacenter Scale ASIC

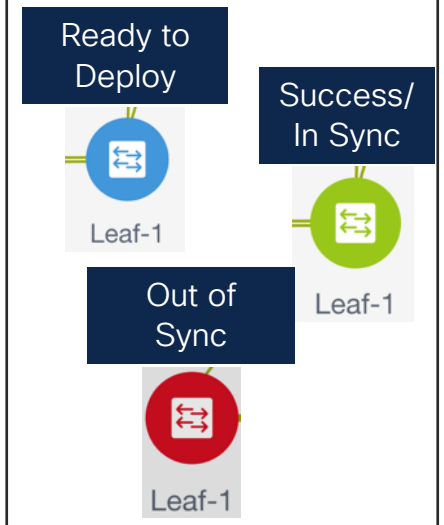
Nexus 9000 시리즈

- Cisco 자체 설계 ASIC
- Software에 따라 컨트롤러 기반의 ACI 혹은 단독형 OS로 사용
- ASIC 기반 VxLAN 오버레이 지원
- ASIC 기반 기반 다양한 Telemetry 지원
- ASIC 기반 암호화 지원 (MACSec)
- 1/10/25/40/50/100/400G 등 용도에 따른 다양한 속도 제공
- ASIC 기반 RoCE, DCBX 등 Datacenter 네트워크 기술 지원

NDFC(Fabric Controller) 기반의 네트워크 운영과 서비스 연계

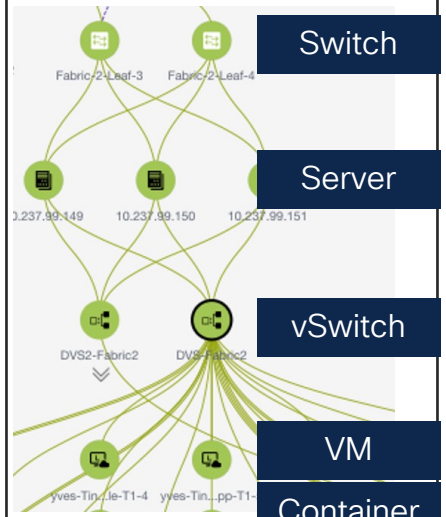
유연한 방식의 Controller

실시간 변경관리



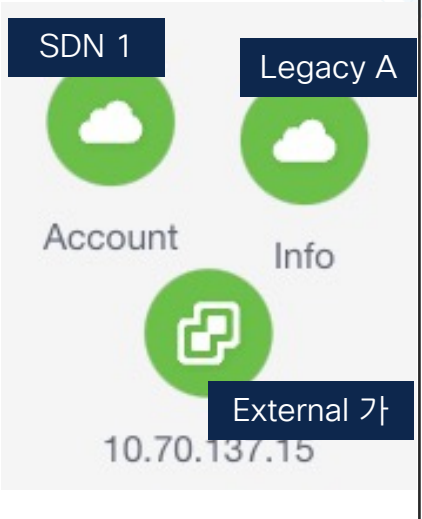
실시간 Compliance Check

서버 가상화 가시성



Openstack Kubernetes VMWare

Legacy~Classic 단일 관리 및 자동화

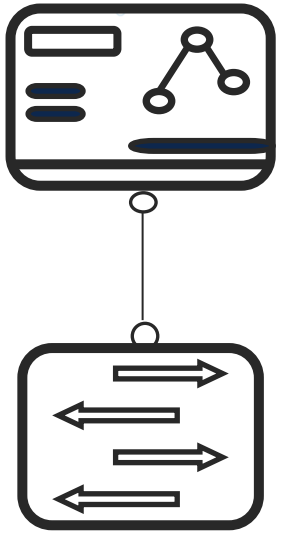


업무 영역별 Fabric 중앙관리
(SDN, Legacy, External)

API, 자동화 중앙포인트

PipeLine

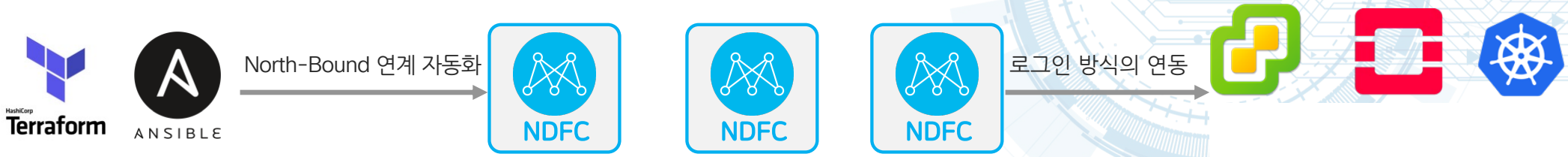
Telemetry연계



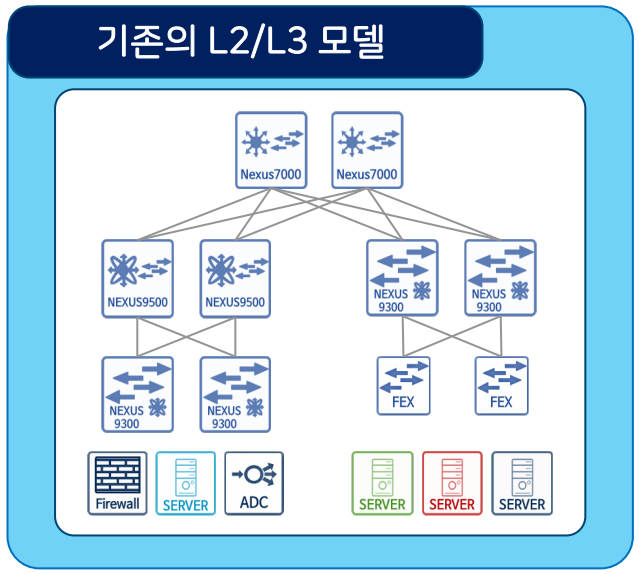
SW Telemetry
HW Telemetry

NDFC(Fabric Controller) 기반의 네트워크 운영과 서비스 연계

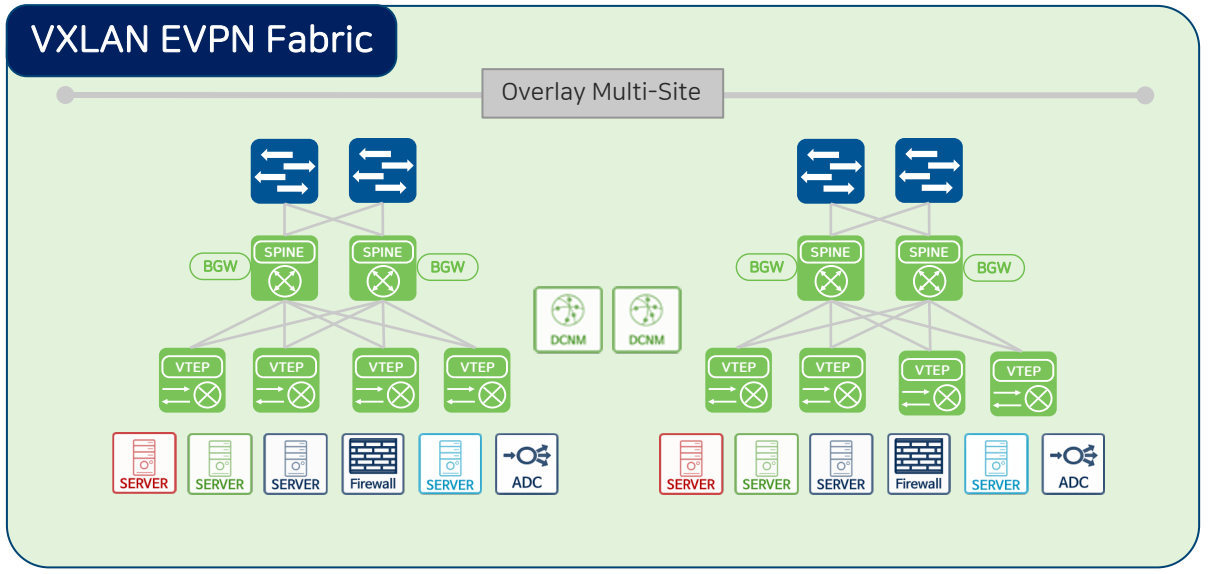
유연한 방식의 Controller



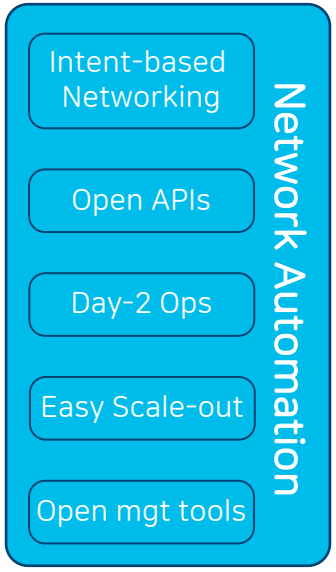
Brown, Greenfield → 단일 컨트롤러에서 모두 수용



기존 운영 네트워크

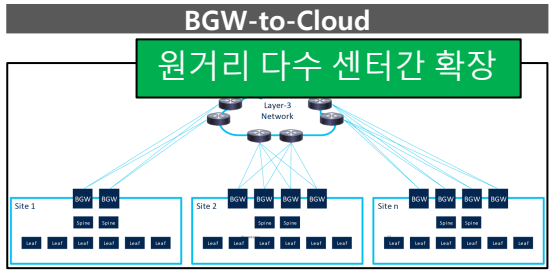


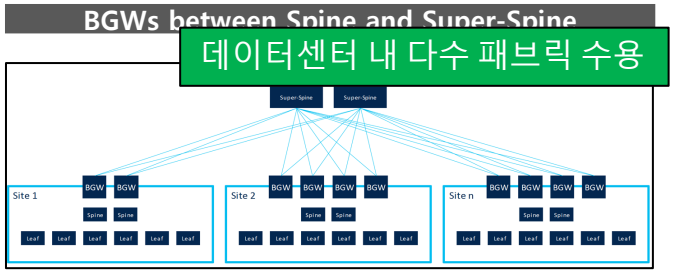
SDN Trunky, Programmable network fabric

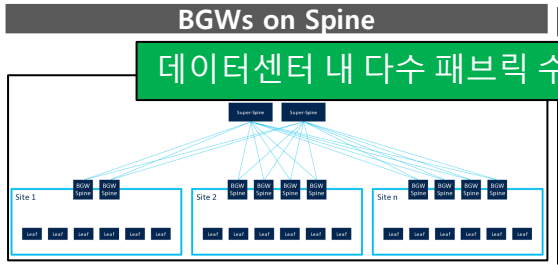


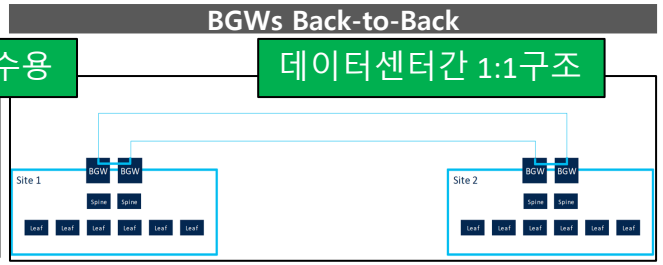
원하는 디자인에 대한 모든 구성 템플릿 기반 자동화

설계사상, 투자규모, 확장을 고려할 때는 다양한 네트워크 토폴로지가 고려됩니다. 시스코 NDFC는 서비스에 맞는 유연한 토폴로지를 효율적으로 운영가능하게 합니다.









1. East West 간 트래픽 성능 확장

- 스파인 스위치를 늘려 리프 스위치 간의 백플레인 성능(속도) 향상
  ex)*샤시형 스위치의 패브릭 확장과 같은 개념*
- 스파인 <-> 리프 스위치 연결 포트에 대해 40G, 100G, 400G의 속도를 지원

2. 서버 등 서비스 연결 포트 확장

- 포트 확장이 필요할 때 신규 리프 스위치를 확장하여 수용 포트 수 확장
  → ex) *샤시형 스위치의 인터페이스 모듈 확장과 같은 개념*
- 리프 스위치를 패브릭에 연결하면 컨트롤러에서 자동으로 인식하게 되며, 위자드 방식의 메뉴를 통해 손쉽게 리프스위치를 연결하게 됨

3. 스파인 <-> 리프 간 논-블로킹 패브릭

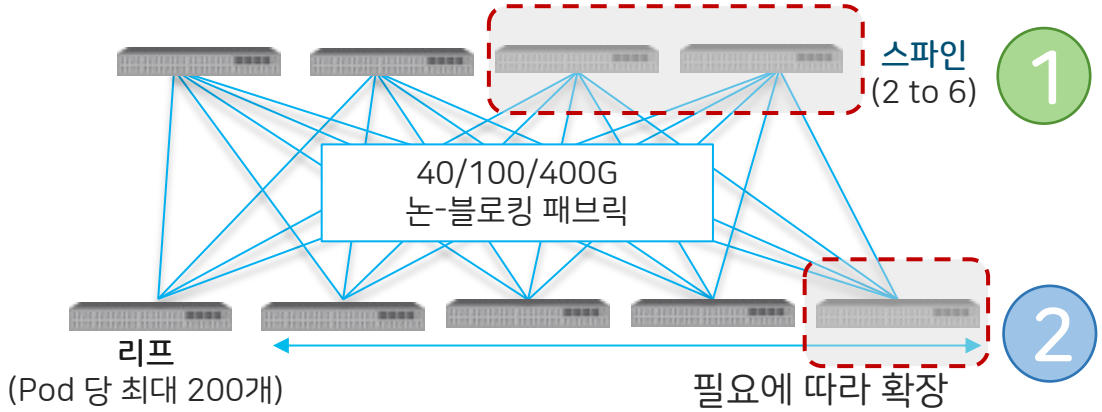
- 스파인 <-> 리프 간 루프방지를 위한 STP로 블로킹이 생기는 구간없이 모든 링크에 대해 100% 활용하며 ECMP(Equal Cost Multi Path) 부하분산 지원
  → ex) *샤시 내에서는 STP, 블로킹이 없는 것과 같은 개념*

단계적 투자방식을 고려할 수 있는 아키텍쳐이며,
영역별로 필요시 구매할 수 있는 유연한 아키텍쳐 입니다.

단계적 투자를 위한 SPINE-LEAF 아키텍쳐 – 가용성 및 확장성

SDN 최초 도입시, 운영에 대한 적응이나 투자금액, 안정성 검증과 같이 다양한 요소가 필요합니다.

컨트롤러
(1,3 또는 이상)



1. East West 간 트래픽 성능 확장

- 스파인 스위치를 늘려 리프 스위치 간의 백플레인 성능(속도) 향상
  ex)*샤시형 스위치의 패브릭 확장과 같은 개념*
- 스파인 <-> 리프 스위치 연결 포트에 대해 40G, 100G, 400G의 속도를 지원

2. 서버 등 서비스 연결 포트 확장

- 포트 확장이 필요할 때 신규 리프 스위치를 확장하여 수용 포트 수 확장
  → ex) *샤시형 스위치의 인터페이스 모듈 확장과 같은 개념*
- 리프 스위치를 패브릭에 연결하면 컨트롤러에서 자동으로 인식하게 되며, 위자드 방식의 메뉴를 통해 손쉽게 리프스위치를 연결하게 됨

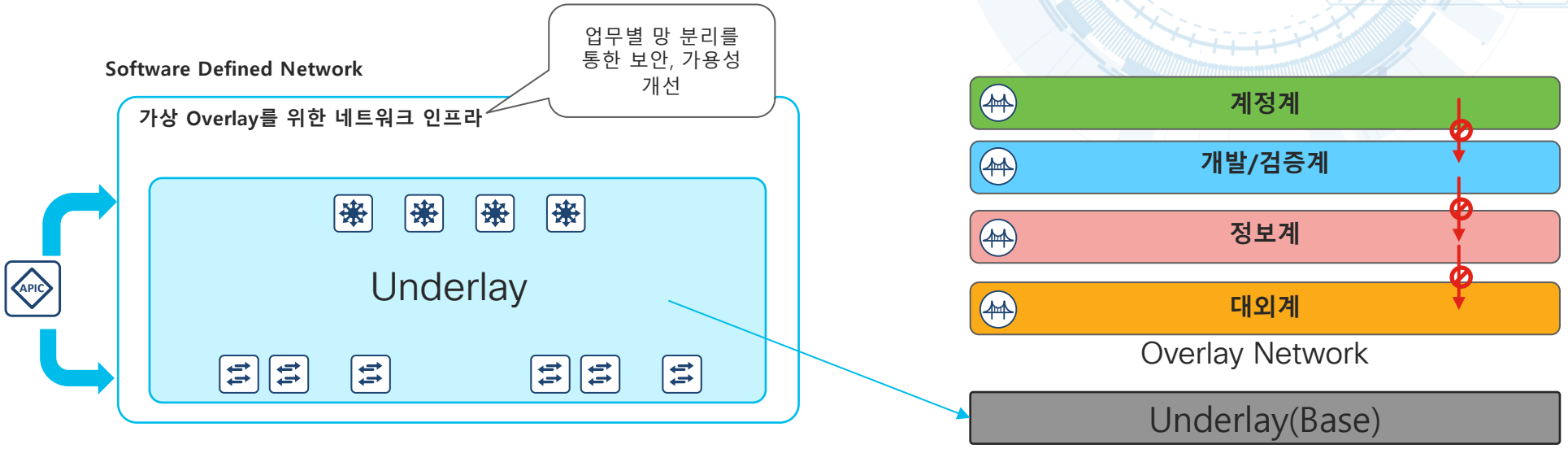
3. 스파인 <-> 리프 간 논-블로킹 패브릭

- 스파인 <-> 리프 간 루프방지를 위한 STP로 블로킹이 생기는 구간없이 모든 링크에 대해 100% 활용하며 ECMP(Equal Cost Multi Path) 부하분산 지원
  → ex) *샤시 내에서는 STP, 블로킹이 없는 것과 같은 개념*

단계적 투자방식을 고려할 수 있는 아키텍쳐이며,
영역별로 필요시 구매할 수 있는 유연한 아키텍쳐 입니다.

가장 손쉬운 Overlay 네트워크 – 단일 하드웨어 내 논리적 망분리

신규 서비스를 수용할 때마다, 신규 스위치 또는 네트워크를 생성해야 하는 과투자에 대해 해결할 필요가 있습니다.



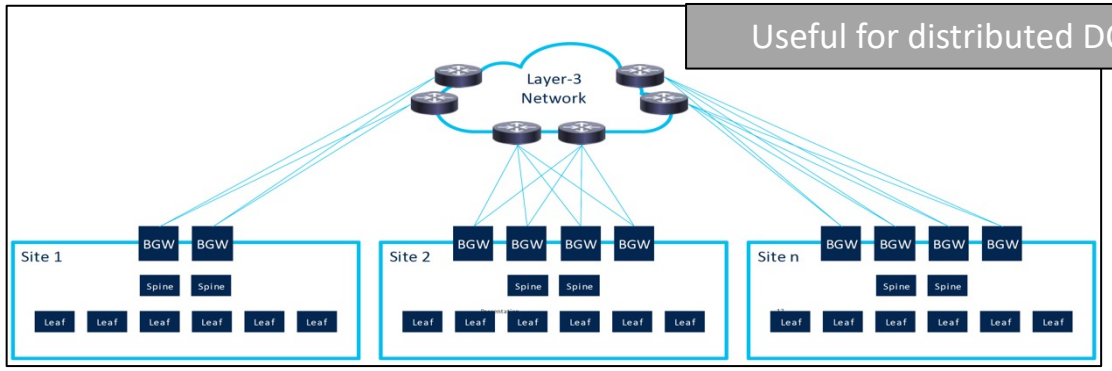
단일 하드웨어 꾸러미 내에서, 망분리가 필요한 계열사/서비스 등을 ISMS 보안요건을 만족하면서,
망분리를 수행하기 때문에, 신규 서비스/기존 서비스에 관계없이 네트워크를 분리할 수 있습니다.

Benefit 1. Supporting Various Fabric Types

According to customer situations

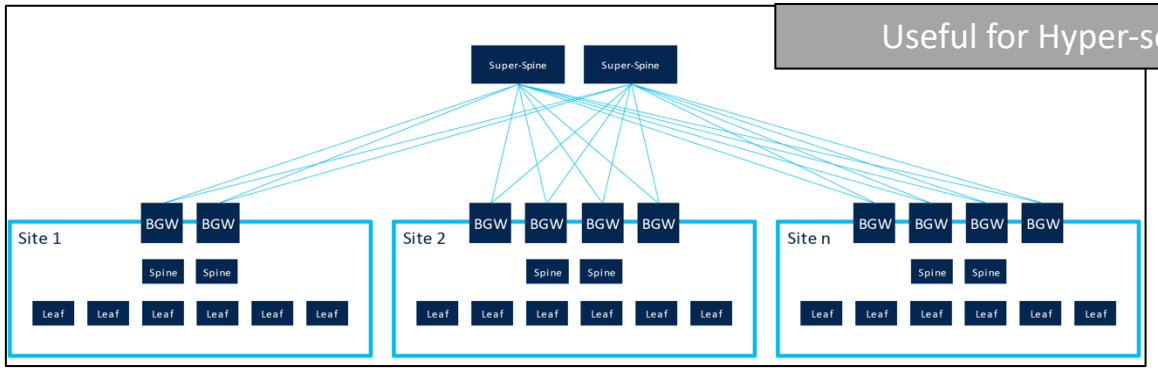
BGW-to-Cloud

Useful for distributed DC



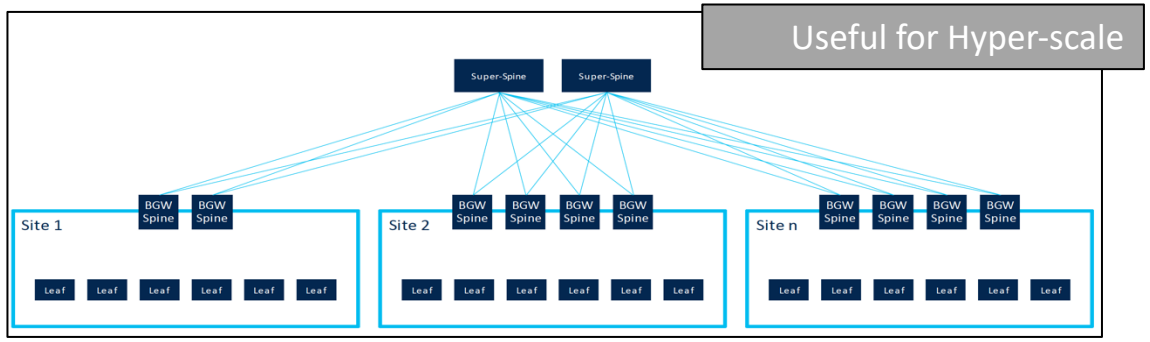
BGWs between Spine and Super-Spine

Useful for Hyper-scale



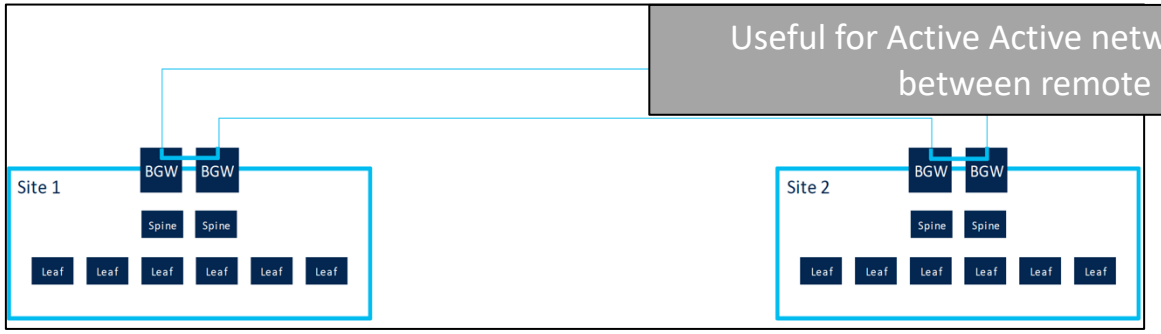
BGWs on Spine

Useful for Hyper-scale

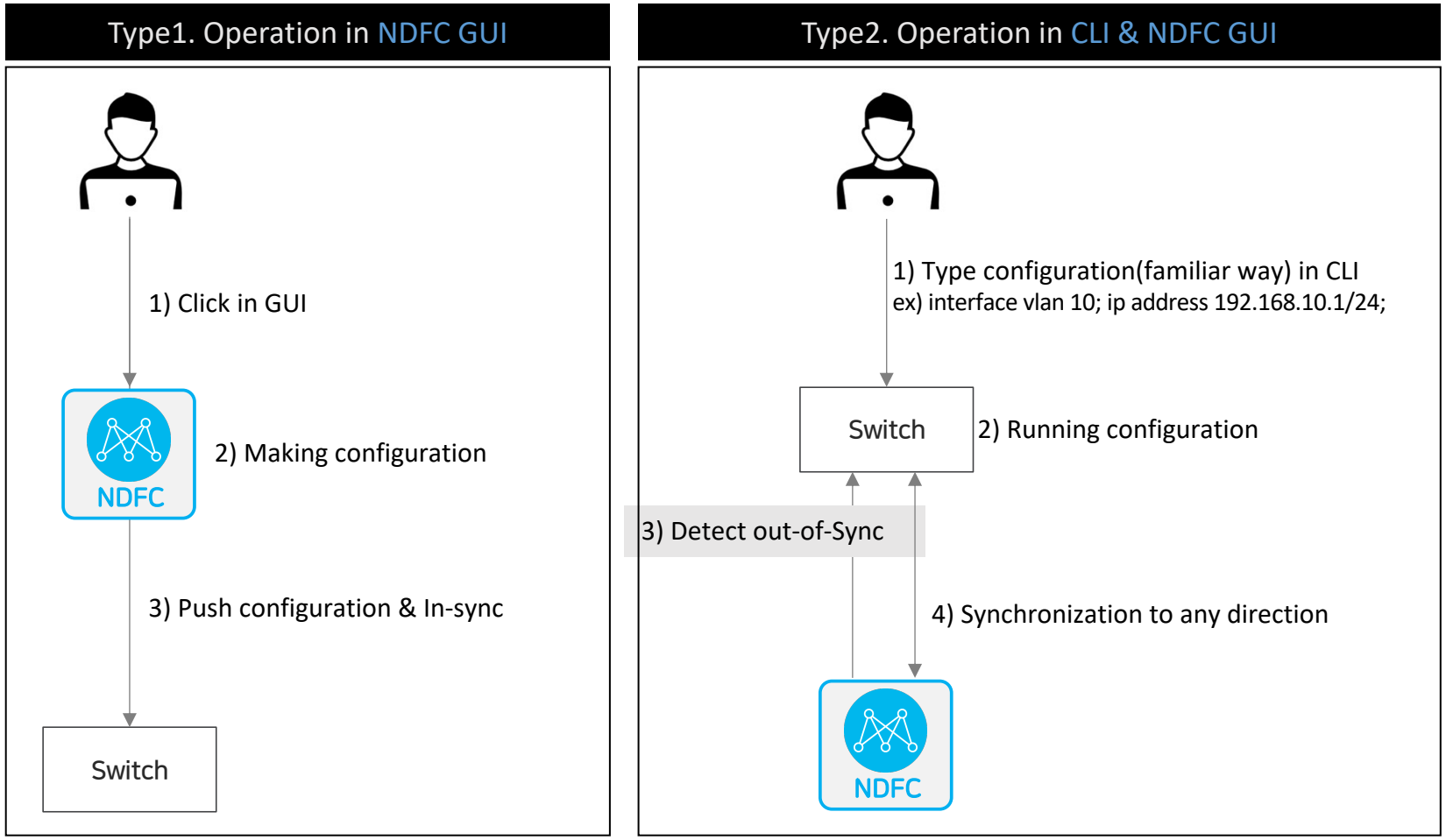


BGWs Back-to-Back

Useful for Active Active network between remote DCs



Benefit 2. Mixed Operation with CLI & GUI



Operation of CLI & GUI

→ Comfortable operation

Benefit 3. Template based Configuration

NDFC-Template based

Kit

Template-based configuration deployment
- built-in *Template for key settings
- Create and deploy a new template directly for settings that are not in the NDFC

*Template: Definition of code or script for pre-setting

ACI-Object based

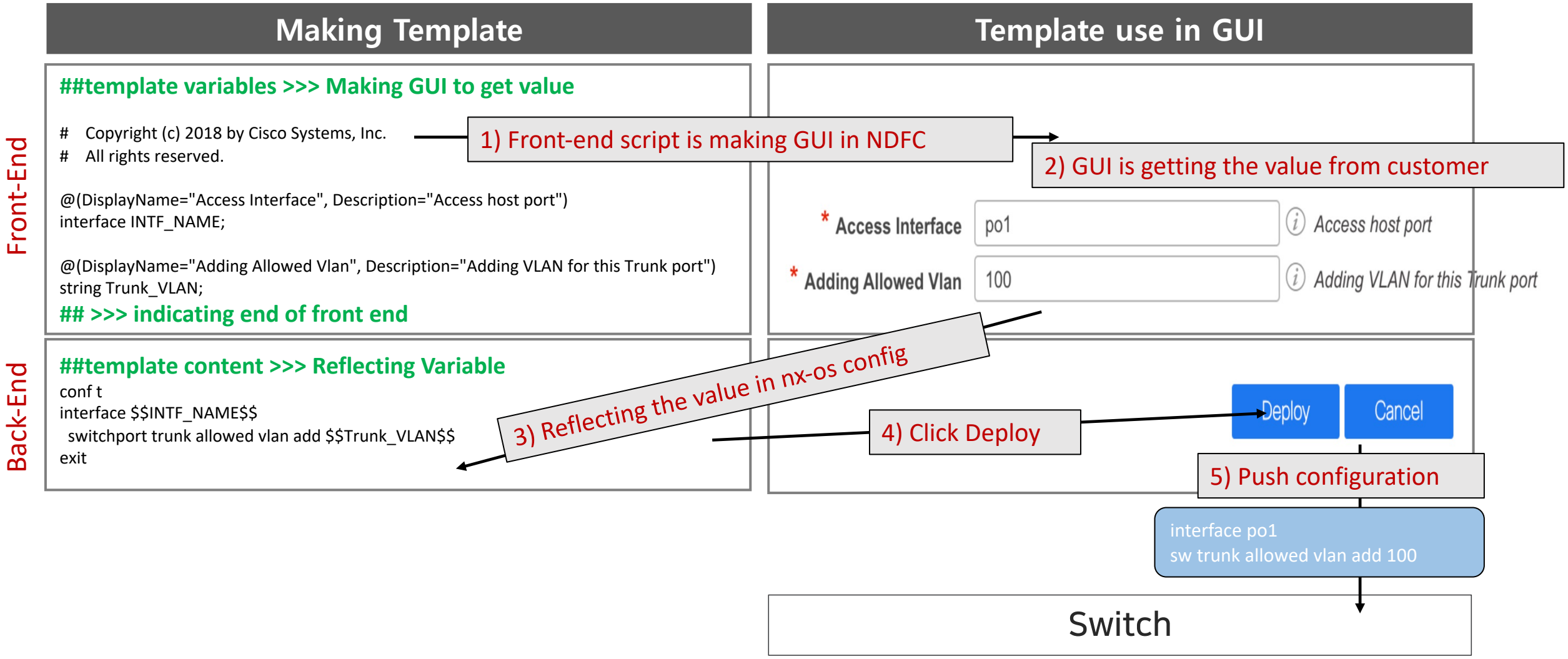
Car

Object-Modeling-based configuration deployment
- Object-modeling for all settings, no need to create configuration.

*Fully built-in to APIC

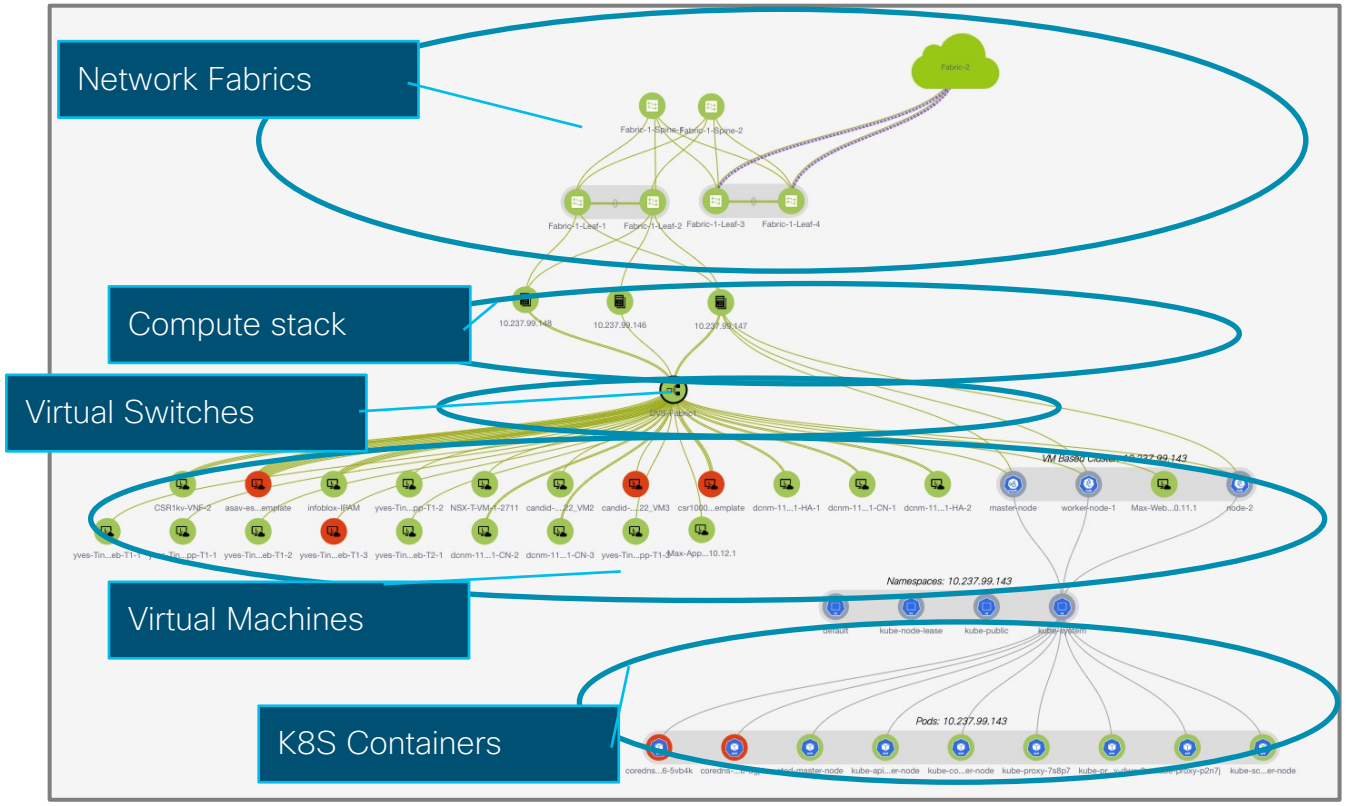
Benefit 3. Template Based Configuration (It is Customizable)

What is template? You can make it.



Benefit 4. Visualizer for Virtual Infrastructure.

From Vmware, Openstack to K8S



Virtual Infrastructure Manager

> From network to all hypervisoor

Vmware / Openstack / Kubernetes(important)

*Vmware : Standard vSwitch, VDS

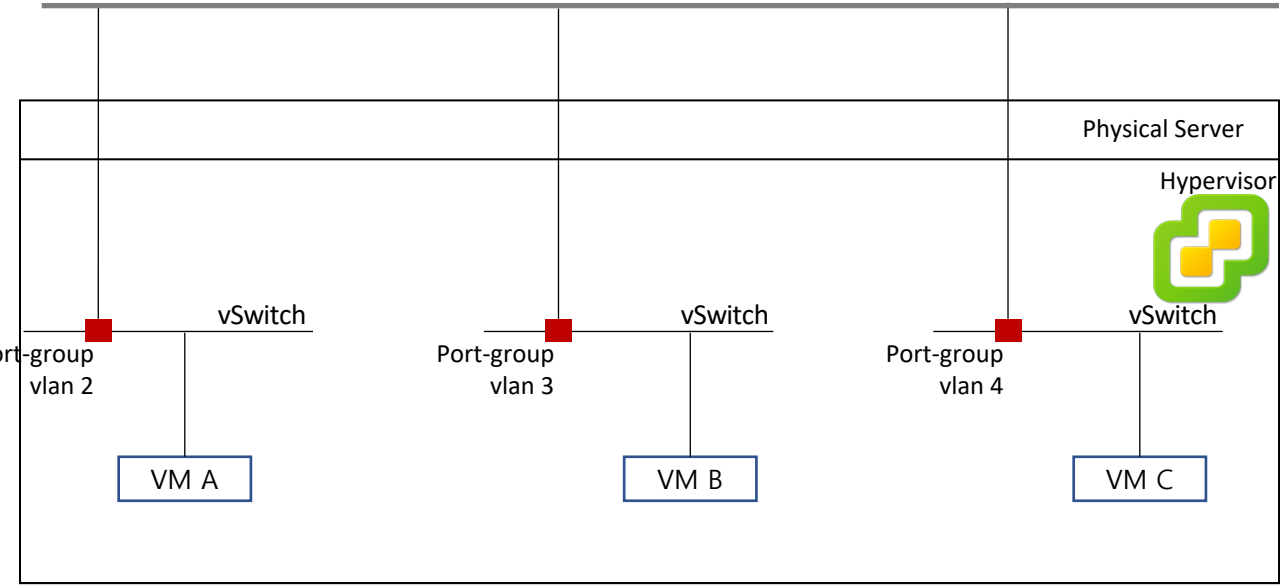
**Kubernetes : not install additional CNI

Benefit 4. Visualizer for Virtual Infrastructure.

Why Visulizer is more effective with K8S?

Leaf Switch

```
Leaf-1# show ip arp
Address         Age       MAC Address     Interface       Flags
99.99.99.11     00:03:57  3890.a5c5.6e03  Vlan2
99.99.99.12     00:03:57  3890.a5c4.c01b  Vlan3
99.99.99.13     00:03:57  3890.a5c3.c12d  Vlan4
```



Can see all the information from Hypervisor

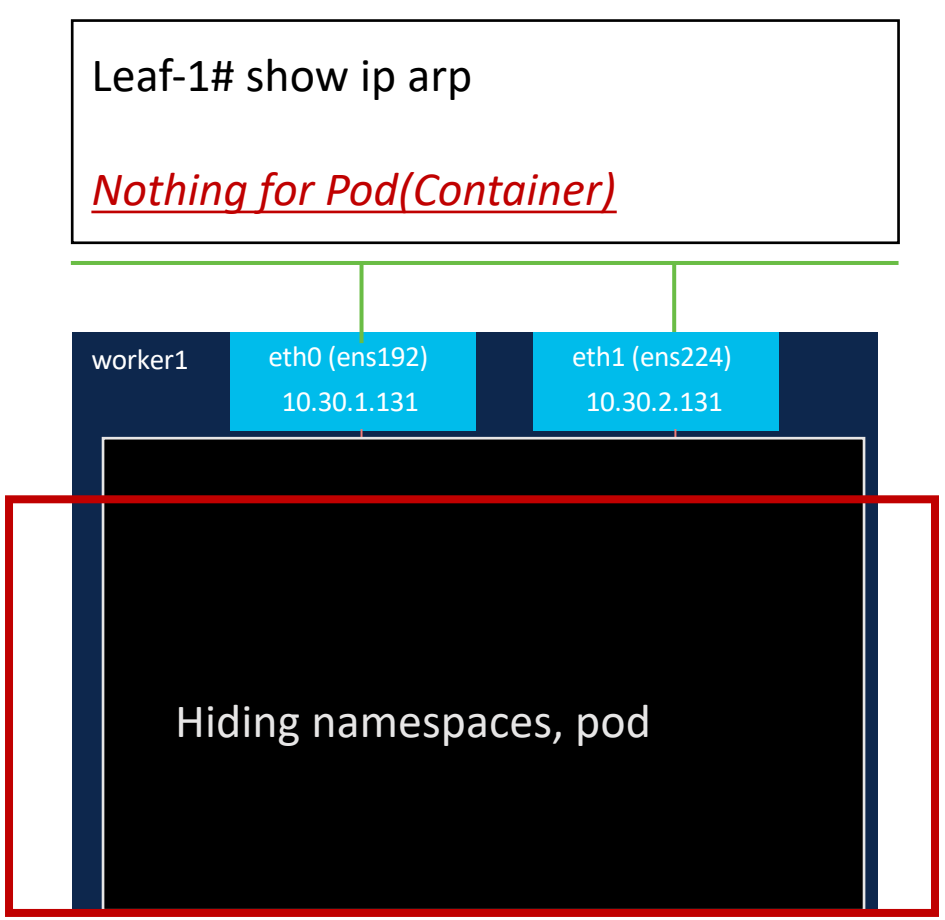
> Easy to operate this kind of environment.

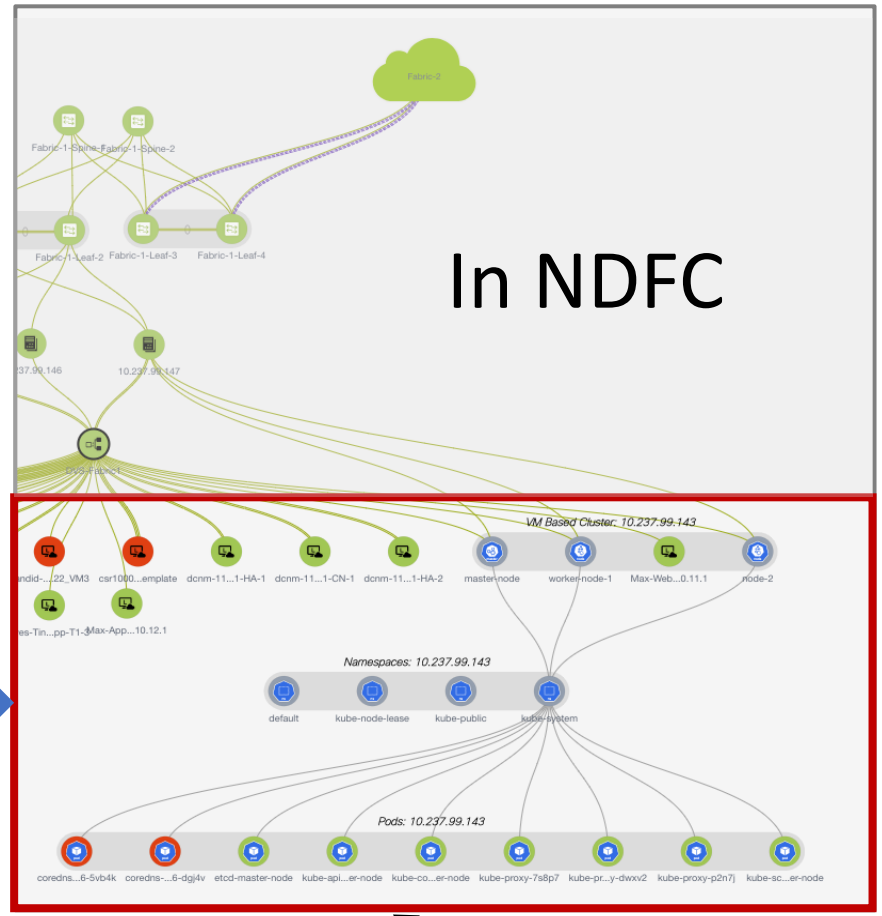
Benefit 4. Visualizer for Virtual Infrastructure.

Why Visulizer is more effective with K8S?

Leaf-1# show ip arp

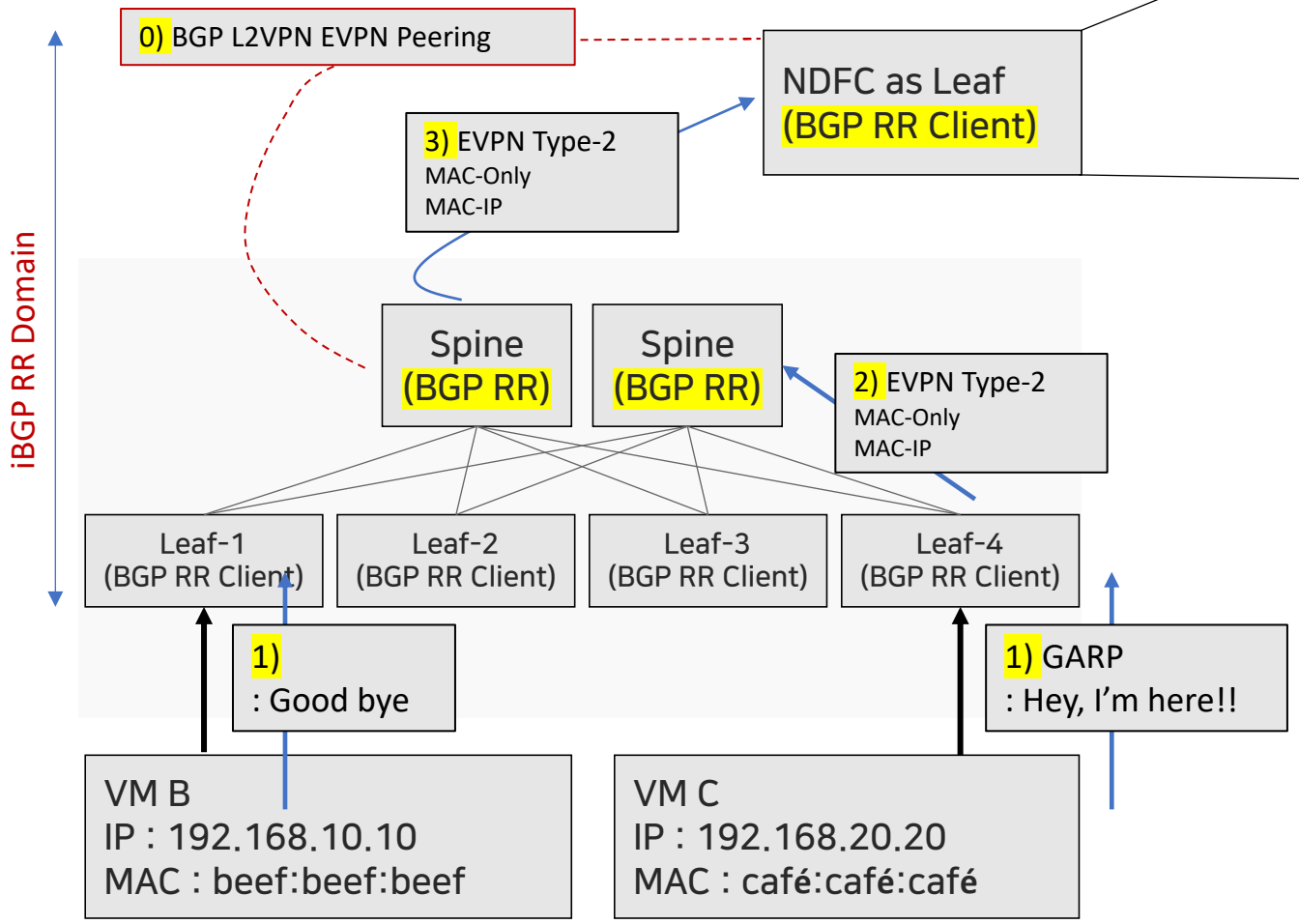
Nothing for Pod(Container)





In NDFC, we can see all the information they are hiding

Benefit 5. EPL(End Point Locator)



Endpoint Locator

IP	MAC	Location	Life
192.168.10.10	beef:beef:beef	Eth-1(Leaf-1)	DELETE
192.168.20.20	café:café:cafe	Eth-1(Leaf-4)	ADD

0) NDFC is also RR Client in iBGP(Fabric)

1) VM sends GARP to leaf

2) Leaf is always updating EVPN Type-2
with MAC-Only / MAC-IP
including RD(Which Leaf:VRF)

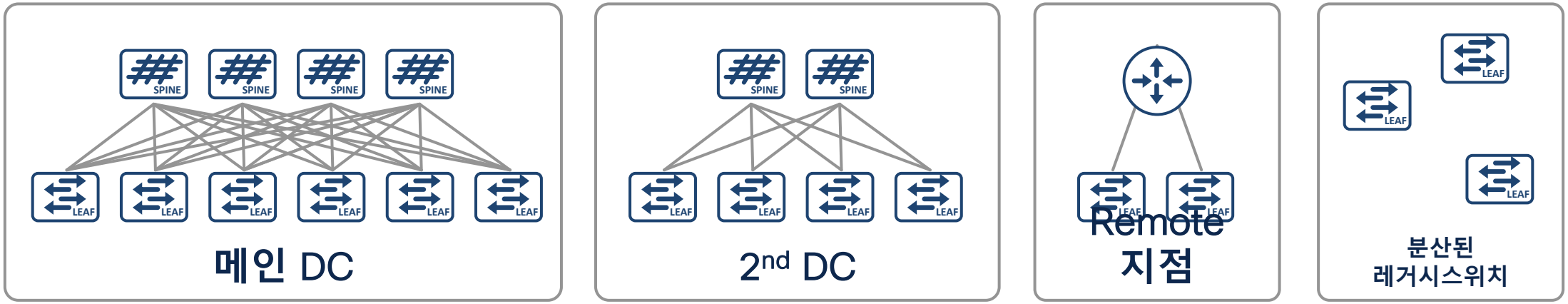
3) Spine is reflecting EVPN Type-2 to NDFC

4) Customer can see VM location and status in Real time in NDFC.

다양한 구조의 네트워크 확장 모델 지원 – SDN Anywhere

데이터 센터 확장과 업무 단일화에 필요한 모든 영역에 대한 운영 솔루션을 제공합니다.

단일 NDFC Controller : 추가 컨트롤러 불필요

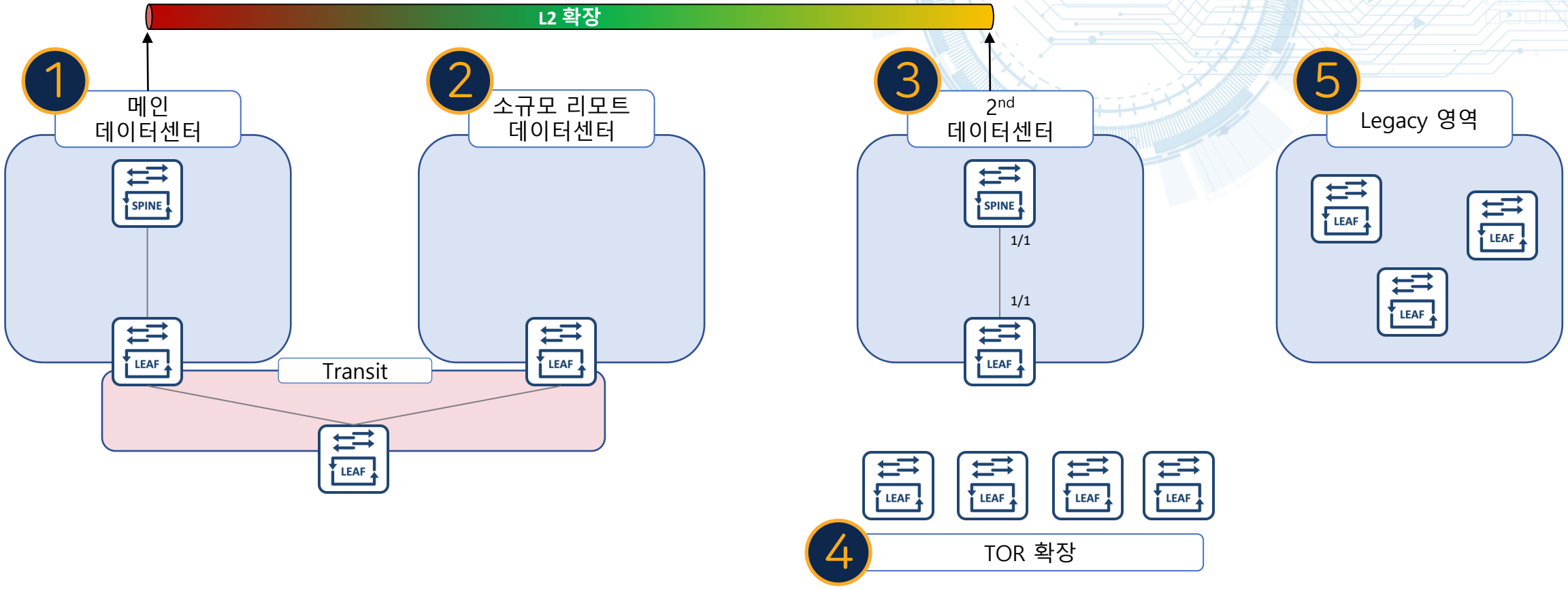


네트워크 확장 가능 / 분리 가능

물리적으로 떨어진 센터 간 동일 Subnet 확장/관리가 가능하며,
규모/위치/용도에 따른 모든 유형의 Network 구조를 지원합니다.

다양한 구조의 네트워크 확장 모델 지원 – SDN Anywhere

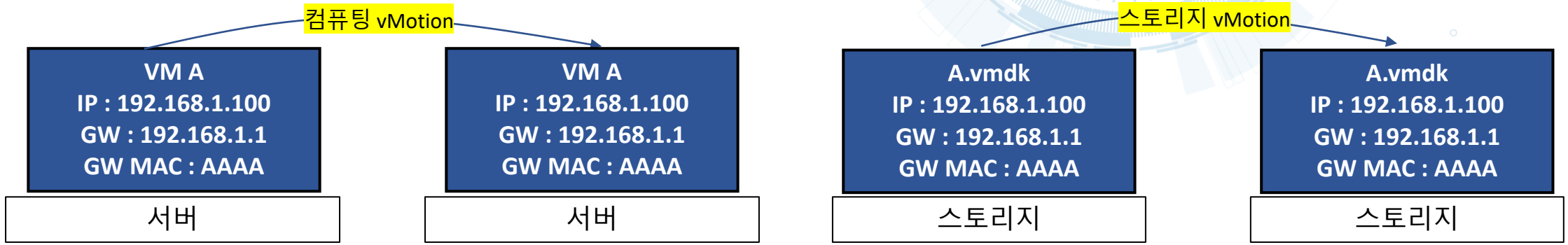
데이터 센터 확장과 업무 단일화에 필요한 모든 영역에 대한 운영 솔루션을 제공합니다.



도입 시, 단계적 확장 가능한 시나리오

손쉬운 네트워크 확장 – IP 모빌리티(vMotion 예시)

vMotion 및 백업의 제약조건과 특성



- vMotion과 백업은 원본 VM을 그대로 복제 하는 방식으로서, VM의 IP 변경이 없습니다.
- 이에 따라, BackEnd Code의 config파일을 수정해야합니다.
- 하드코드된 어플리케이션의 경우 IP변경이 불가합니다.

→ 이러한 특성 때문에, RPO/RTO = 0 을 목표로 구성하기 위해서는, 센터 내/외를 불문하고 확장된 공통된 IP대역 / 게이트웨이 IP / 게이트웨이 MAC이 필요합니다.

어플리케이션이 곧 회사의 비즈니스 입니다.
vMotion과 백업 이외에도 네트워크의 기술제약이 비즈니스의 걸림돌이 되지 않도록,
유연한 설계를 위해 센터간 네트워크를 확장하는 기술을 고려해 봐야합니다.

Enhanced ISSU(Fixed) : 무중단 서비스 운영

- ✓ 단일 Sup 엔진(CPU)상에서 컨테이너 기술을 활용하여 무중단 OS 업그레이드를 수행
- ✓ Zero Packet loss 기반의 무중단 OS 업그레이드 (데이터 서비스 중단 없음, 포트 리셋 없음)
- ✓ 지원 플랫폼: Nexus 9300-EX (7.0(3)I7(3)이상), Nexus 9300-FX (9.3(5)이상)

  주1) Enhanced ISSU 적용을 위해서는 OS boot mode를 컨테이너(LXC) 모드로 전환 필요(스위치 reload 필요). vPC로 구성된 경우 Peer 스위치도 동일 Boot Mode 적용 필요

  주2) Target 버전으로 업그레이드시 Host Linux Kernel 버전의 업그레이드가 필요한 경우 Standard ISSU와 비슷한 수준의 Control Plane의 지연 발생

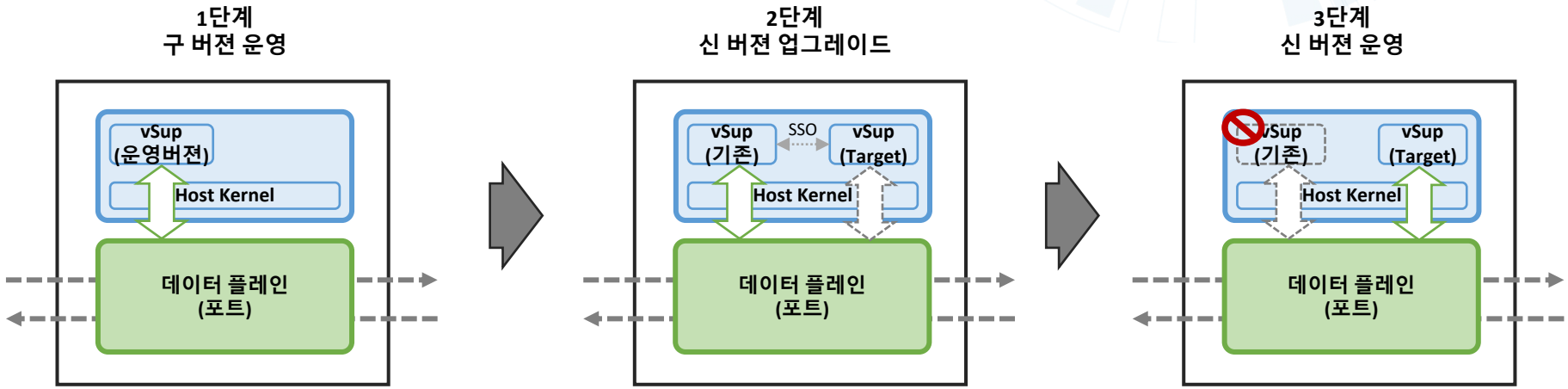


그림: Nexus 9300 스위치의 Enhanced ISSU의 동작

적용 케이스

- ✓ 무중단 OS 업그레이드 (Zero Packet Loss)
  -> 서버가 직접 연결되는 ToR(box형) 스위치에서 구현되고 포트 재설정이 필요 없기 때문에 NIC teaming 상태 유지 (서비스 영향도 최소화)

감사합니다.


CISCO

차세대 모니터링 방식

Nexus Telemetry 기반의 모니터링 기법

- ND-Insights

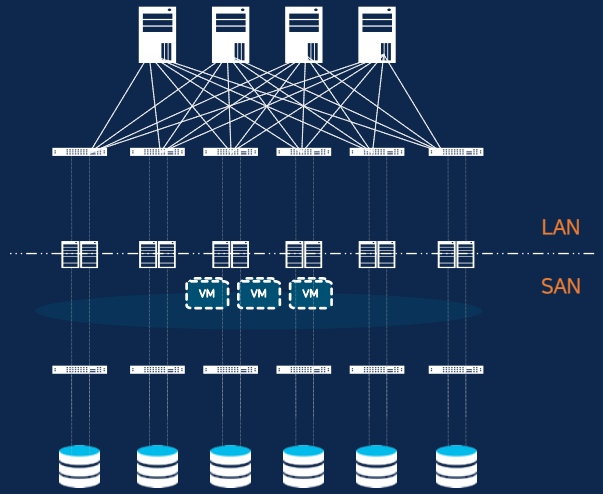
Cisco Systems Korea

2022. 11

ND-Insights : SDN에 적합한 모니터링 툴

Powering automation
Unified agile platform

Cisco Nexus Dashboard



네트워크 센서 기반의
SW/HW 텔레메트리 정보 전달

광범위 데이터 수집
및 데이터 모델링

ML 기반의
텔레메트리 상관 관계 분석

AI 기법을 활용한
이상행위 탐지 및 예방

ND Insight
Nexus

Nexus Dashboard
Insights

- 소프트웨어 텔레메트리 - 제어부 프로토콜 상태 정보, 장비 환경 정보, 다양한 카운트 정보 등
- 하드웨어 텔레메트리 - 시스코 CloudScale ASIC / ASIC NPU 에 내제된 텔레메트리 기능

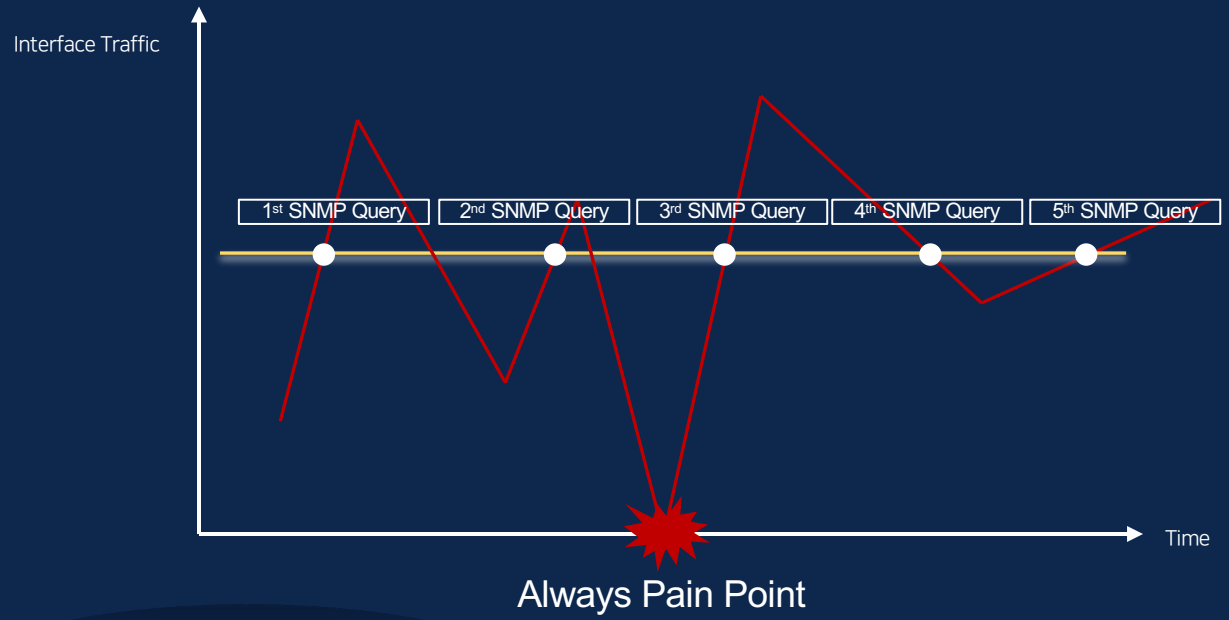
네트워크 텔레메트리
- 기존방식의 한계

네트워크 텔레메트리

기존 네트워크 텔레메트리 방식의 한계와 문제점

Powering automation
Unified agile platform

Cisco Nexus Dashboard



SNMP 모니터링의 한계

- Pull 기반 메커니즘
- Polling 주기 조정 등 수동 개입 필요
- 요청하는 특정 수치 메트릭에 제한한 모니터링

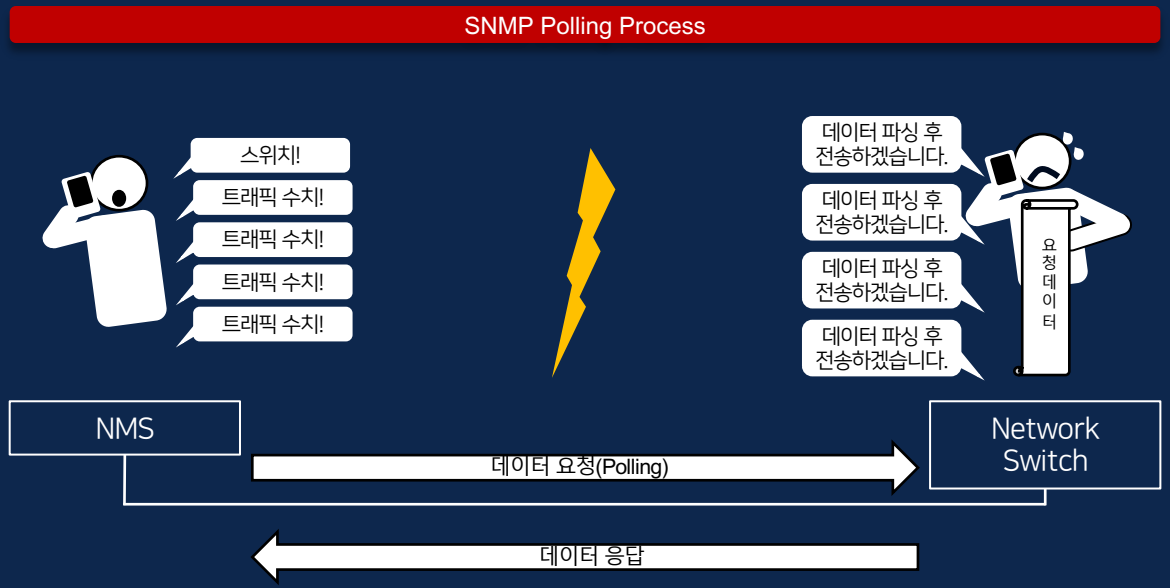
- 더 짧은 Polling 주기 조정 필요
- 단일 스위치 내에서 다수의 메트릭을 확인하기위해 다수의 Polling 필요

네트워크 텔레메트리

기존 네트워크 텔레메트리 방식의 한계와 문제점

Powering automation
Unified agile platform

Cisco Nexus Dashboard



SNMP 모니터링의 한계

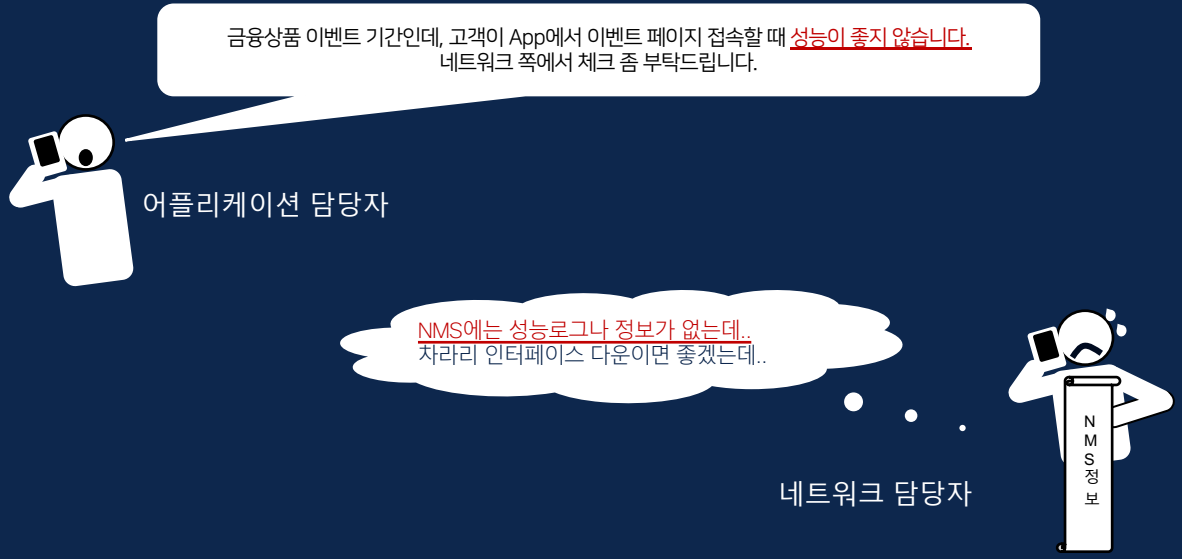
- 잦은 Request & Response 반복
- CPU Job 증가
- 더 많은 SNMP 요청
  → CPU 부하
- 더 많은 장비일 경우
  → 네트워크 내 트래픽 폭주
- 모니터링 수집실패 야기

네트워크 텔레메트리

기존 네트워크 텔레메트리 방식의 한계와 문제점

Powering automation
Unified agile platform

Cisco Nexus Dashboard



서비스 중심 분석 부족

- 기존 방식은 주로 Control-Plane (개별스위치)을 모니터링하는 방식
- 서비스에 대한 가시성 보다는, 개별 스위치의 상태 모니터링

- “서비스 = 비즈니스”이나 가시성 부족

네트워크 텔레메트리

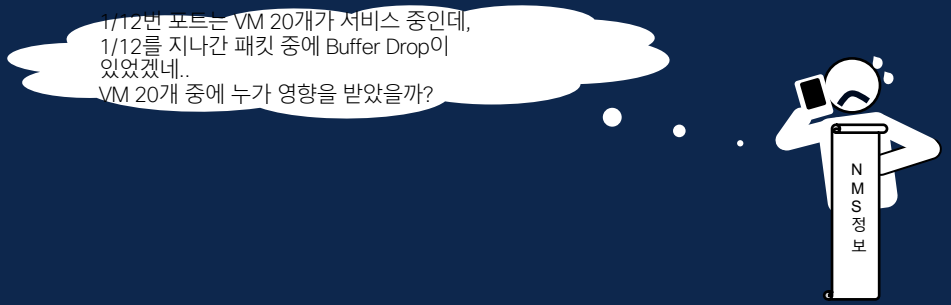
기존 네트워크 텔레메트리 방식의 한계와 문제점

Powering automation
Unified agile platform

Cisco Nexus Dashboard

%TAHUSD-SLOTX-4-BUFFER_THRESHOLD_EXCEEDED:
Module 1 Instance 12 Pool-group buffer 80 percent threshold is exceeded!

>>1/12 포트에서 80%이상의 버퍼를 사용하여 임계치가 초과 되었습니다.



서비스 중심 트러블슈팅 부족

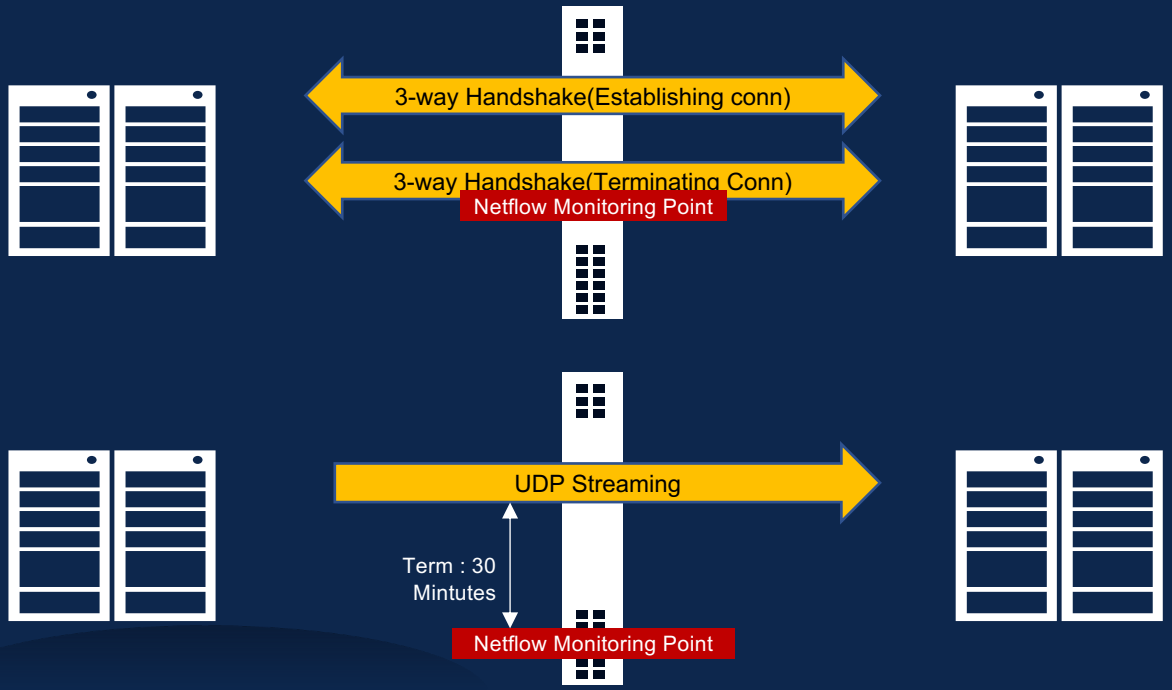
+ 기존 방식은 주로 Control-Plane (개별스위치)을 모니터링하는 방식
+ 서비스에 대한 가시성 보다는, 개별 스위치의 상태 모니터링
+ “서비스 = 비즈니스”이나 가시성 부족

네트워크 텔레메트리

기존 네트워크 텔레메트리 방식의 한계와 문제점

Powering automation
Unified agile platform

Cisco Nexus Dashboard



Netflow/sFlow 모니터링의 한계

+ TCP는 Session FIN 이후 모니터링가능
+ UDP는 일정 주기 이후 모니터링가능

+ 실시간 Flow 모니터링 불가

+ 하드웨어 연계 모니터링 불가

Ex) Flow 성능정보 일부 수집 가능하나, 하드웨어와의 연관성 체크 불가

Software/Hardware Telemetry를 결합한 인프라 운영 솔루션

- ND-Insights

Cisco Nexus Dashboard

Simple to Automate, Simple to Consume

Powering automation
Unified agile platform

Insights

Insights
- FT / FT Event / MicroBurst 탐지 / 상관관계
- Operation 지원(운영권고, 변경관리)

SAN controller

SAN SDN Controller(FC-SAN 기반)
- FC 기반의 SAN Switch Controller
- Flow 성능 수집 / 표준기반 스토리지 연동

Orchestrator

Multi-Site Orchestration
- Between On-Prem ACI Site
- Cloud ACI와 On-Prem ACI Site/NDFC Site

Fabric controller

NX-OS 기반 SDN Controller
- Configuration 자동화(VXLAN EVPN 등)
- Image, Performance, Topology 관리

Data broker

Nexus as Brokering Network Switch
- TAB Network Controller
- Packet Modify / GRE Termination

Fabric discovery

Fabric Discovery(이기종 장비 관리)
- Catalyst / 타사 Switch(제약)

Cisco Data Center 솔루션의 Hypervisor

Private cloud

Public cloud

Cisco Nexus Dashboard

Simple to Automate, Simple to Consume

Telemetry의 결합 필요

Powering automation
Unified agile platform

Cisco Nexus Dashboard

Device Control-Plane 상태 정보

Software telemetry

Device Data-Plane 상태 정보

5-Tuple
Interface/Queue 정보
Buffer Drop 지표
ACL, Policer Drop
Latency, Burst Threshold 초과

LAN

Hardware telemetry

Software Telemetry와 Hardware Telemetry 정보가 연계된 모니터링 제공

네트워크 텔레메트리

Data-Plane 모니터링의 중요성과 Telemetry 기능의 중요성

Powering automation
Unified agile platform

Cisco Nexus Dashboard

스위치는 물리적으로 1대이지만,
2종류의 Plane으로 분리되어 운영됩니다.

Control-Plane

SNMP

PIM

OSPF

BGP

Console Port

MGMT Port

eth1/1

Data Port

eth1/2

Data Port

서비스 = Business

Data-Plane

Control-Plane

- 스위치 포트 업 / 다운 상태 정보
- OSPF Neighbor 다운
- Configuration

Data-Plane

- Application간 Latency
- Application간 Drop
- Application간 Re-Transmission

스트리밍 텔레메트리 방식 제공

Data-Plane 모니터링의 중요성과 Telemetry 기능의 중요성

Powering automation
Unified agile platform
Cisco Nexus Dashboard

스위치는 물리적으로 1대이지만,
2종류의 Plane으로 분리되어 운영됩니다.

Control-Plane

- 스위치 포트 업 / 다운 상태 정보
- OSPF Neighbor 다운
- Configuration

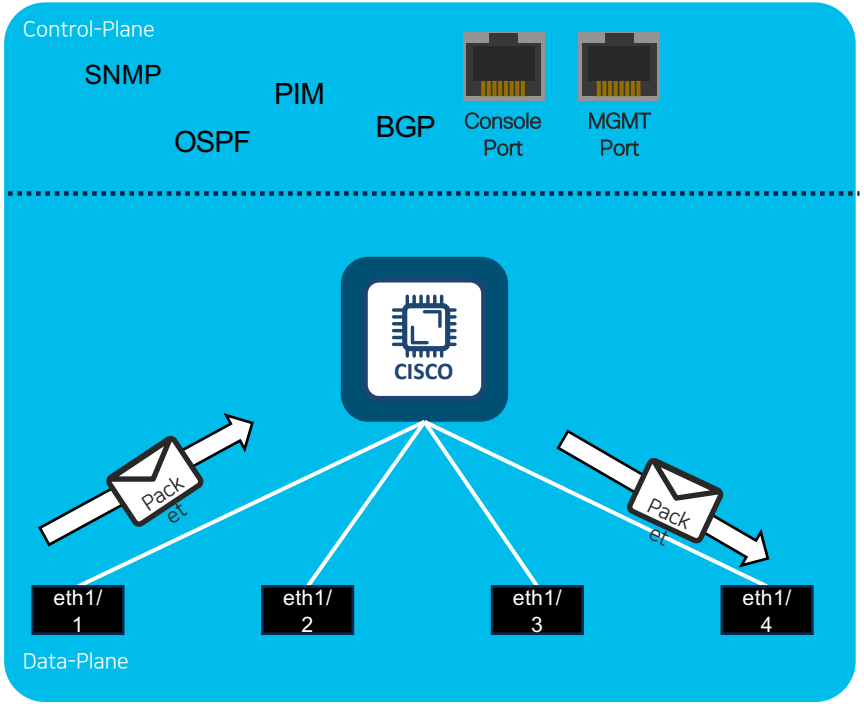
우리가 이미 모니터링중인 영역 : 스위치의 상태

Data-Plane

- Application간 Latency
- Application간 Drop
- Application간 Re-Transmission

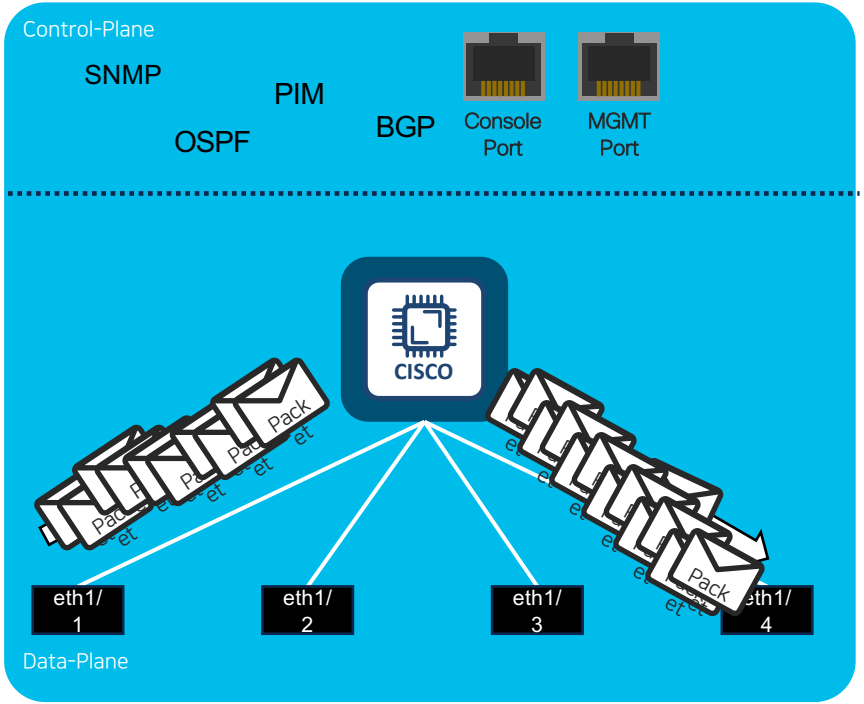
우리가 모니터링 하지 않는 영역 : 어플리케이션 상태

데이터 플레인 모니터링



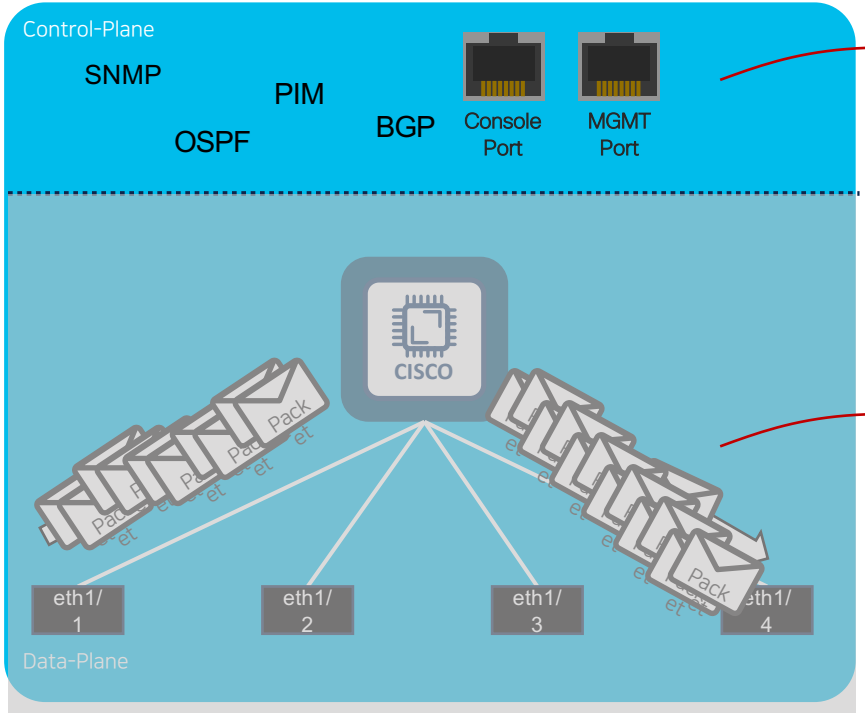
정상

데이터 플레인 모니터링



Burst 발생

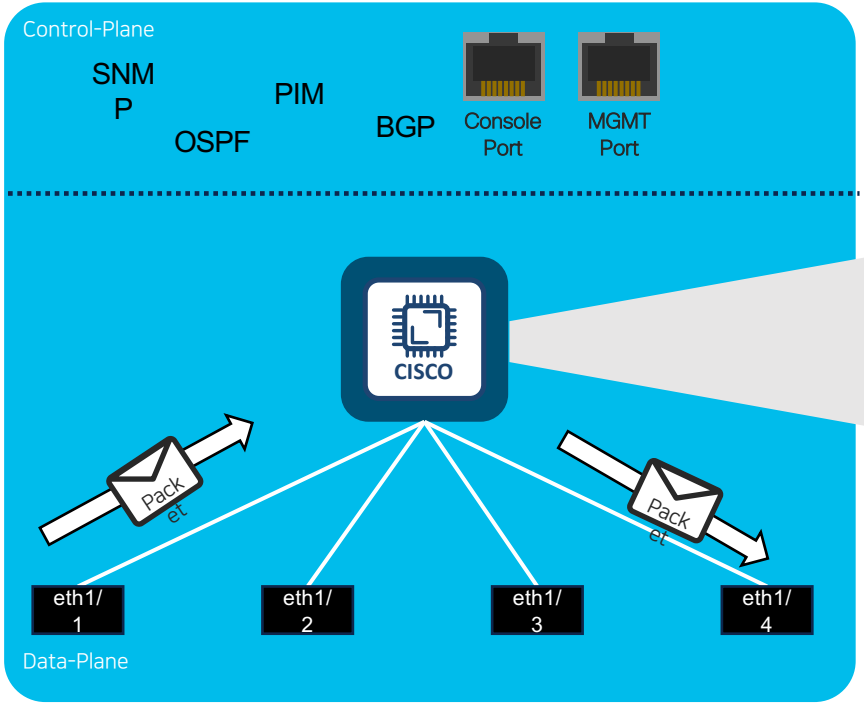
데이터 플레인 모니터링



NMS 기반 모니터링 영역

Burst 발생

데이터 플레인 모니터링 – 패킷 플로우

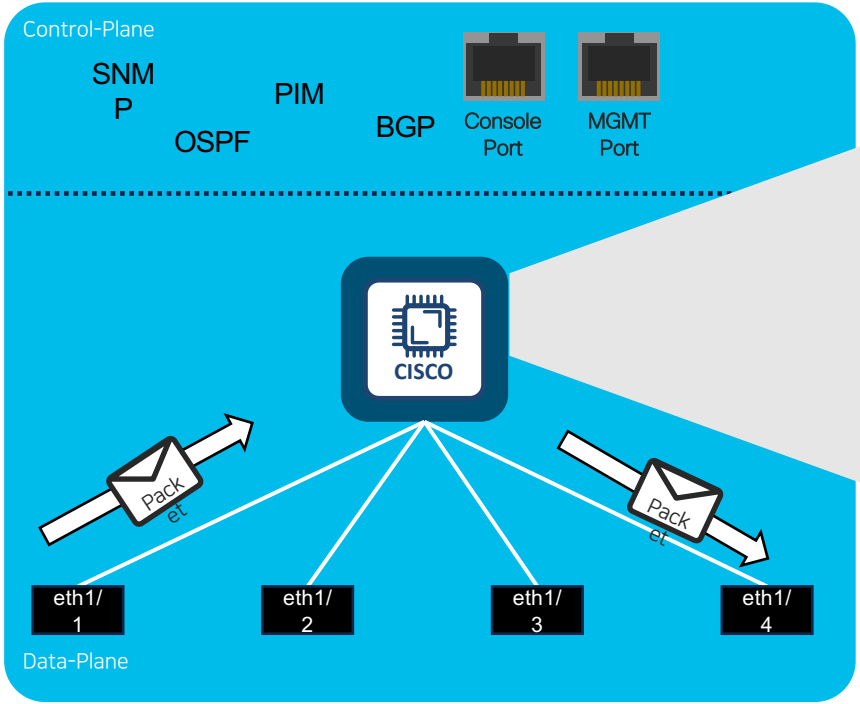


5-Tuple(통신 플로우)
플로우 테이블

출발IP	도착IP	출발Port	도착Port	프로토콜	레이턴시	시간
VM 111	VM 222	65321	443	TCP	7	17:12
VM 333	VM 444	64321	443	TCP	4	17:13

\+ 전체 장비의 상관관계 분석

데이터 플레인 모니터링 – 패킷 플로우

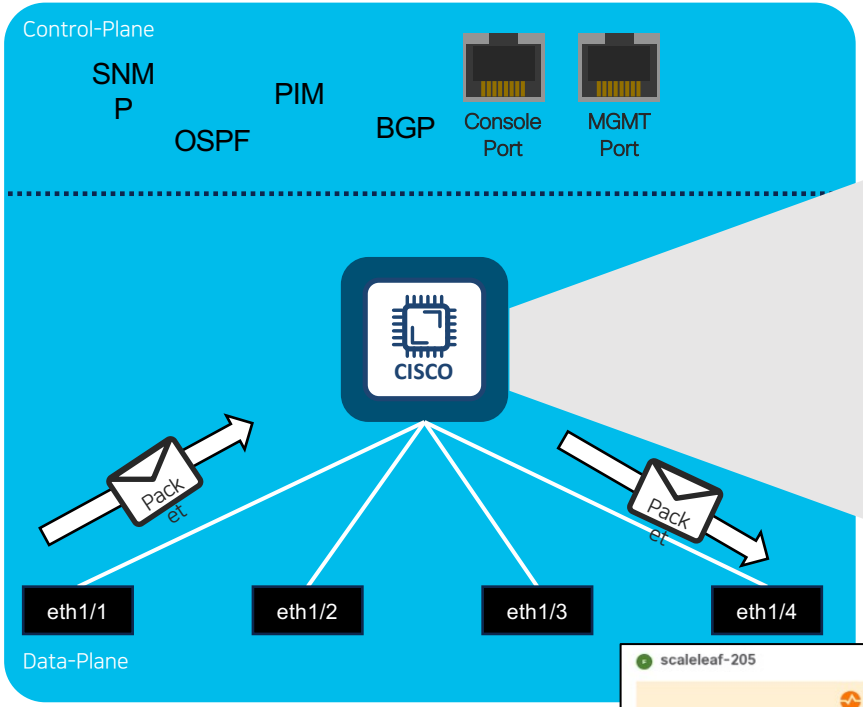


스위치(플로우가 지나가는 하드웨어)	
출발포트	도착포트
1/1	1/4

5-Tuple(통신 플로우)
플로우 테이블

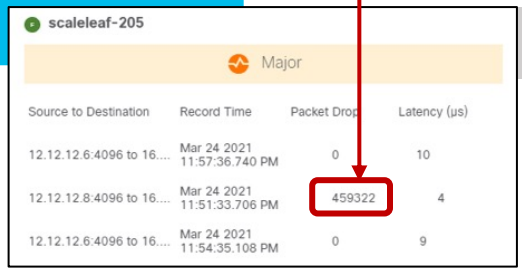
출발IP	도착IP	출발Port	도착Port	프로토콜	레이턴시	시간
VM 111	VM 222	65321	443	TCP	7	17:12
VM 333	VM 444	64321	443	TCP	4	17:13

데이터 플레인 모니터링 – 패킷 플로우 상의 이벤트 결합

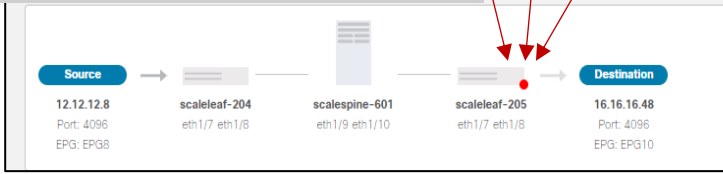


스위치(플로우가 지나가는 하드웨어)	
출발포트	도착포트
1/1	1/4

5-Tuple(통신 플로우) 플로우 테이블							플로우 테이블 이벤트		
출발IP	도착IP	출발Port	도착Port	프로토콜	레이턴시	시간	버퍼드랍	버스트드랍	리트렌스미션
VM 111	VM 222	65321	443	TCP	7	17:12			
VM 333	VM 444	64321	443	TCP	4	17:13			



기준 이벤트로 원인 분석



Cisco Nexus Dashboard

Simple to Automate, Simple to Consume

Telemetry의 결합 필요

Powering automation
Unified agile platform

Cisco Nexus Dashboard

Device Control-Plane 상태 정보

Software telemetry

Device Data-Plane 상태 정보

5-Tuple
Interface/Queue 정보
Buffer Drop 지표
ACL, Policer Drop
Latency, Burst Threshold 초과

LAN

Hardware telemetry

Software Telemetry와 Hardware Telemetry 정보가 연계된 모니터링 제공

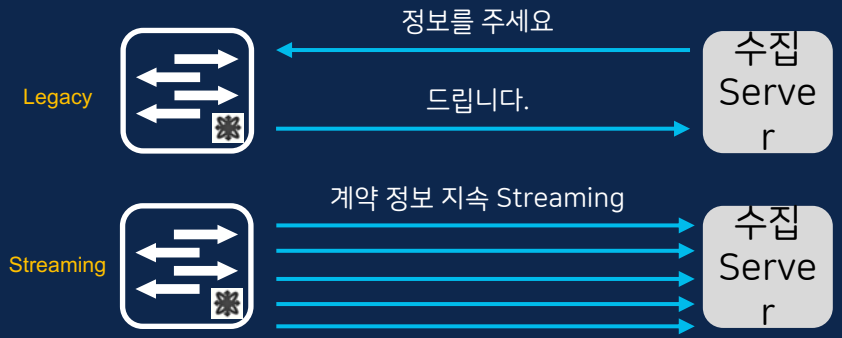
Cisco Nexus Dashboard

기존 방식의 비효율성을 제거하는 Streaming Telemetry의 등장

Powering automation
Unified agile platform

Cisco Nexus Dashboard

스트리밍 텔레메트리를 통한
Poll 방식 텔레메트리의 비효율성 제거



기대효과

통일된 구조화된 데이터 전송 : 수집의 효율성

시간 단위 / 변경 단위 수집 : 정보의 특성에 따른 수집

CPU 부하 감소와 스케일의 보장 : 장비부하 감소

Data-Plane(HW ASIC Telemetry) 텔레메트리 활용
어플리케이션 성능 모니터링 가능

기대효과

어플리케이션 성능 모니터링 가능

Per Packet 수준의 성능 모니터링

(5-Tuple, Timeline, Micro-Burst, Buffer/Queue Drop 등)

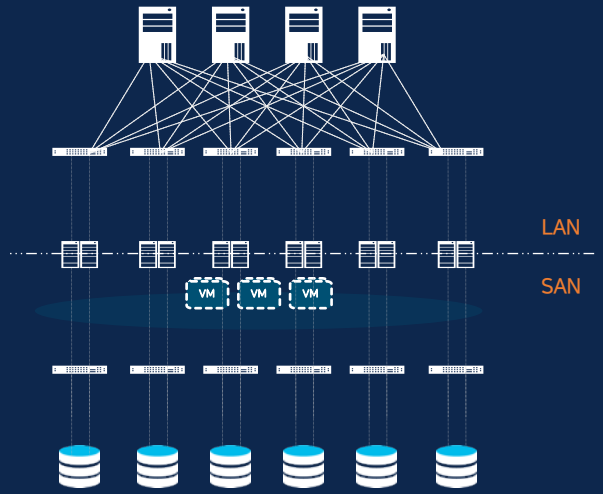
Data-Plane과 Control-Plane 상태 정보와의 연계

Cisco Nexus Dashboard Insights

Simple to Automate, Simple to Consume

Powering automation
Unified agile platform

Cisco Nexus Dashboard



네트워크 센서 기반의
SW/HW 텔레메트리 정보 전달

광범위 데이터 수집
및 데이터 모델링

ML 기반의
텔레메트리 상관 관계 분석

AI 기법을 활용한
이상행위 탐지 및 예방

ND Insight
Nexus

Nexus Dashboard
Insights

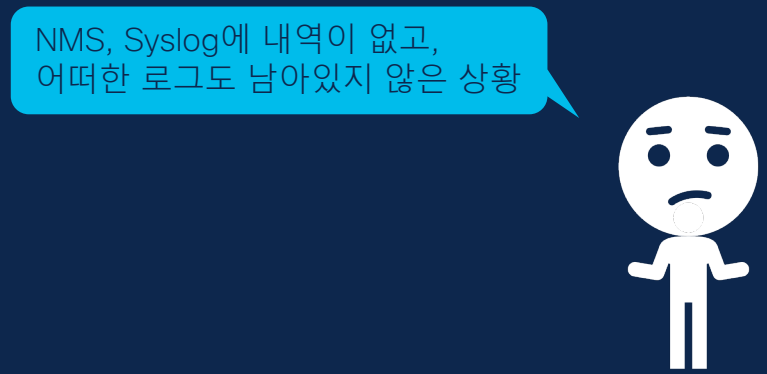
- 소프트웨어 텔레메트리 - 제어부 프로토콜 상태 정보, 장비 환경 정보, 다양한 카운트 정보 등
- 하드웨어 텔레메트리 - 시스코 CloudScale ASIC / ASIC NPU 에 내제된 텔레메트리 기능

Use Case 1. Packet Drop

Insight
Nexus Dashboard

Scenario

- 어제 이 시간에 특정 서비스 속도가 느리고, 간헐적으로 끊김
- 속도가 느렸지만, 서비스는 가능했던 상황
- 네트워크 관점의 분석 필요

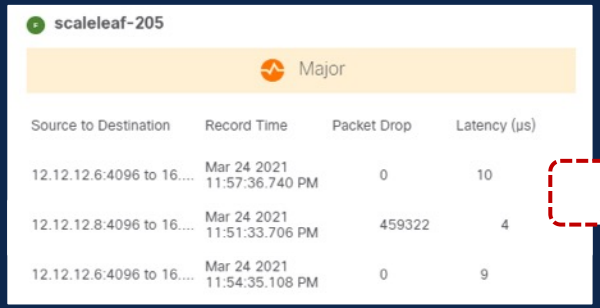


Insight

Nexus Dashboard

Use Case 1. Packet Drop

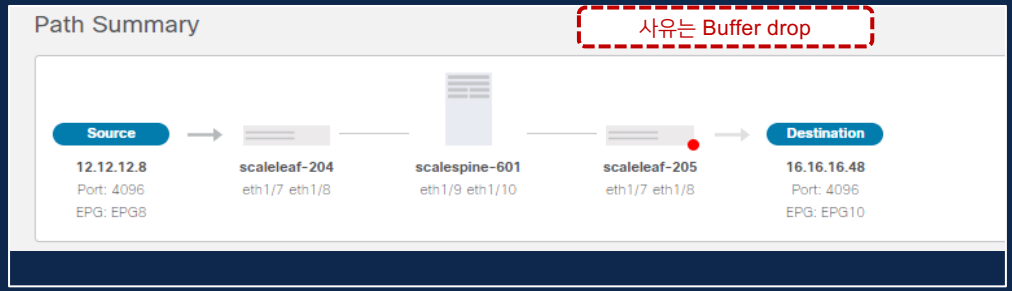
- 과거의 인프라의 상태 관리
- Per Packet 수준의 어플리케이션 상태 정보 확인
- 단일 스위치가 아닌, 연관된 모든 스위치를 기준으로 확인



밤 11:51분에 Packet Drop이 45만개 이상 발생한 통신 기록 확인

패킷 드랍을 기준으로 분석시작

다른 스위치보다, Drop이 더욱 발생한 스위치 발견

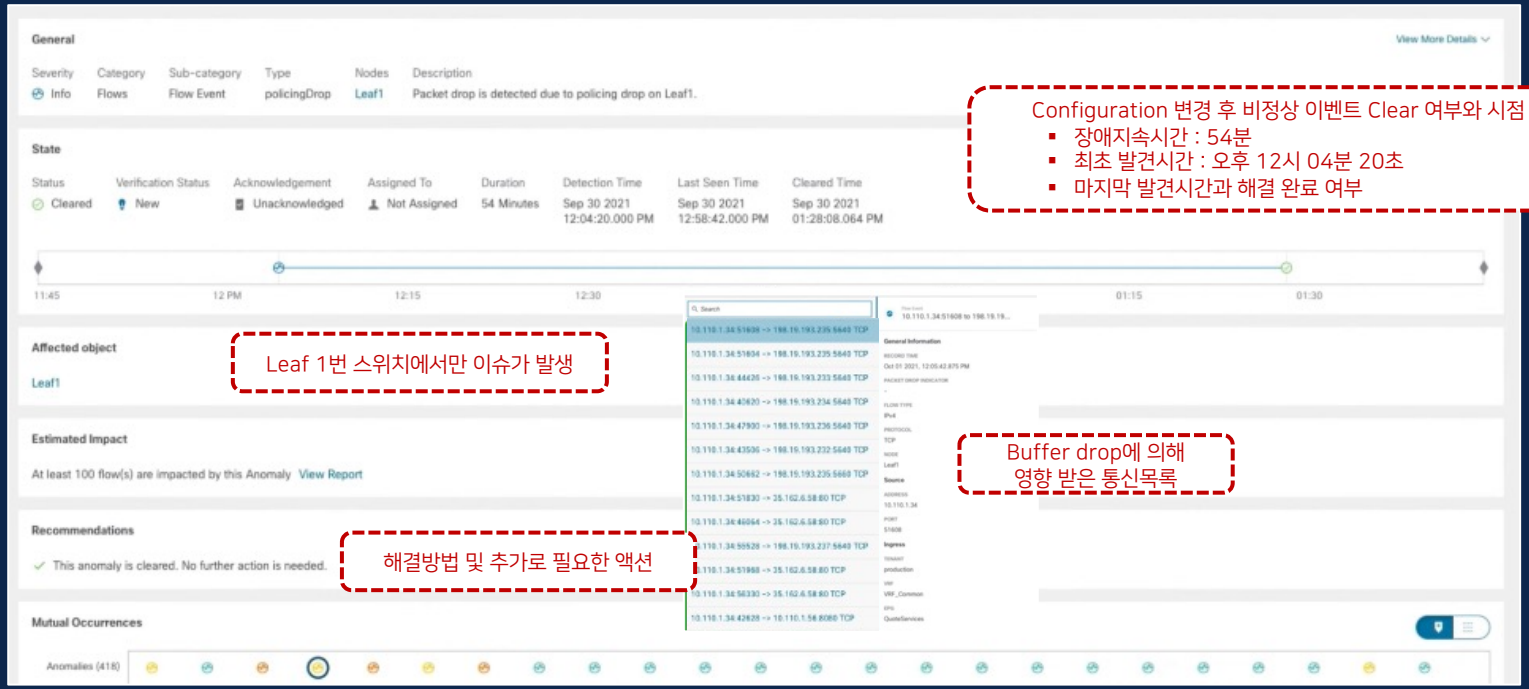


Insight

Nexus Dashboard

Use Case 1. Packet Drop

네트워크 상태 분석



Configuration 변경 후 비정상 이벤트 Clear 여부와 시점 확인

- 장애지속시간 : 54분
- 최초 발견시간 : 오후 12시 04분 20초
- 마지막 발견시간과 해결 완료 여부

Leaf 1번 스위치에서만 이슈가 발생

Buffer drop에 의해 영향 받은 통신목록

해결방법 및 추가로 필요한 액션

Use Case 2. Micro-Burst

Insight
Nexus Dashboard

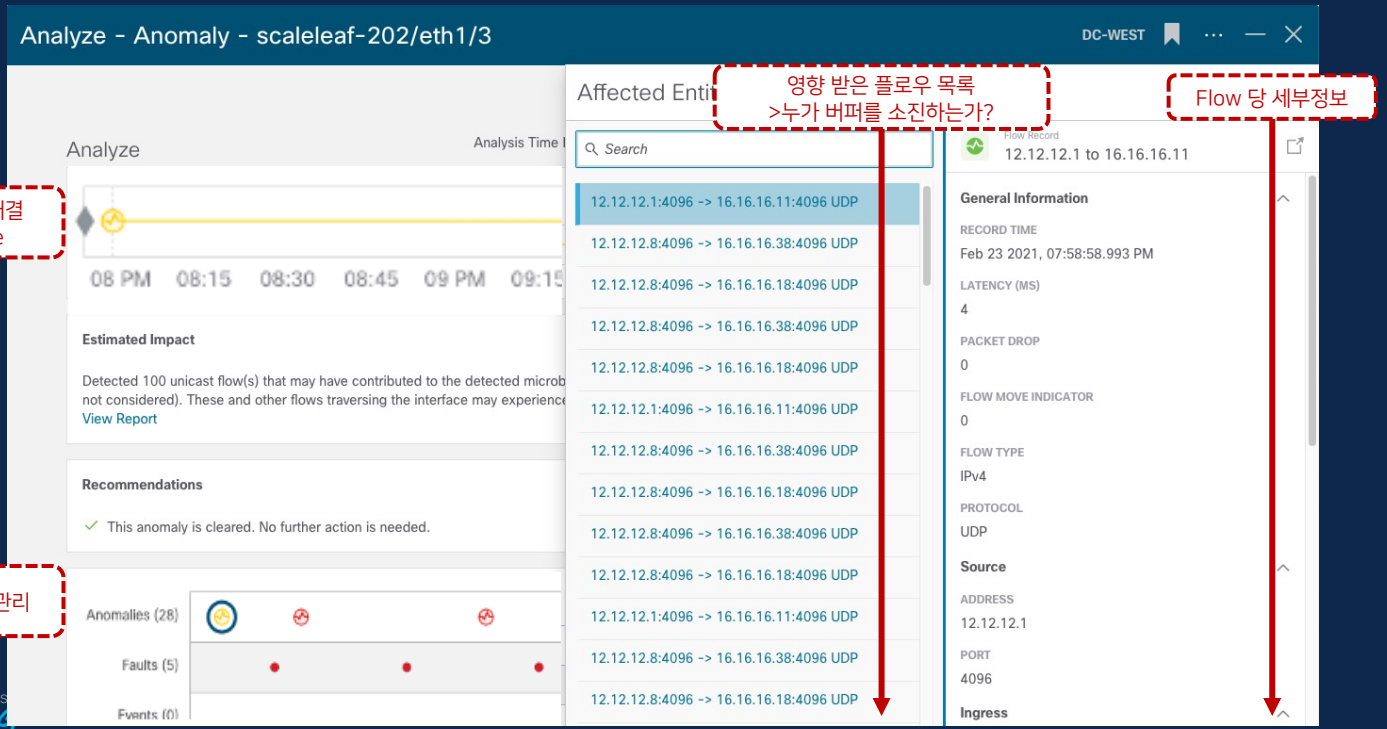
Scenario

- Microburst에 의한 순간 트래픽 폭주로, 3번 포트에 Buffer 초과 Syslog가 NMS에 기록되었음
- Burst에 의해 영향 받은 명확한 플로우 확인 필요

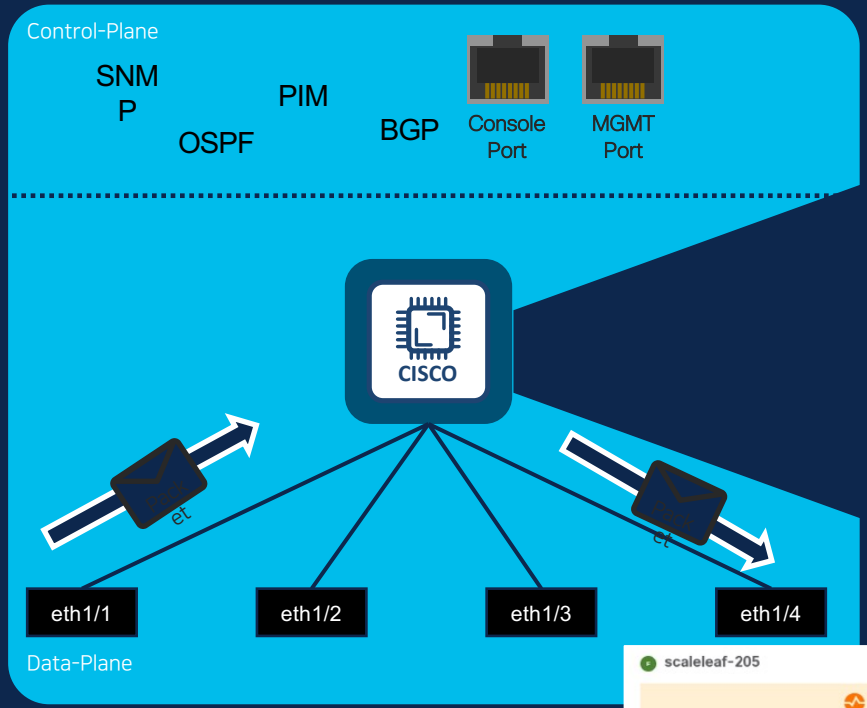
Syslog :::
%TAHUSD-SLOTX-4-BUFFER_THRESHOLD_EXCEEDED: Module X Instance Y Pool-group buffer Z percent threshold is exceeded!

Use Case 2. Micro-Burst

- Microburst 문제 지속/해결 TimeLine >> Microburst 이슈 해결 여부 확인
- Burst 이력 관리 >> Burst 발생 시간과 패턴 분석
- Burst에 의해 영향 받은 Flow 관리 >> 예상하지 못했던 영향 Flow 확인

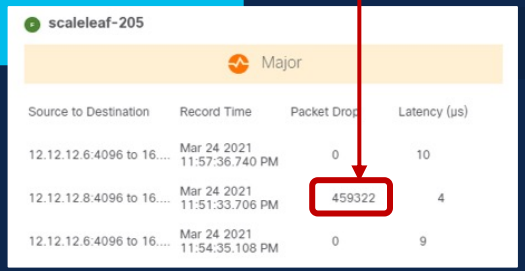


데이터 플레인 모니터링 – 패킷 플로우 상의 이벤트 결합



스위치(플로우가 지나가는 하드웨어)	
출발포트	도착포트
1/1	1/4

5-Tuple(통신 플로우) 플로우 테이블							플로우 테이블 이벤트		
출발IP	도착IP	출발Port	도착Port	프로토콜	레이턴시	시간	버퍼드랍	버스트드랍	리트렌스미션
VM 111	VM 222	65321	443	TCP	7	17:12			
VM 333	VM 444	64321	443	TCP	4	17:13			



기준 이벤트로 원인 분석

The bridge to possible