



The bridge to possible

# Hyperscale Infrastructure

## Session Introduction

황승희 부사장

시스코코리아 클라우드 인프라스트럭처&소프트웨어 그룹

# 클라우드 데이터센터



Private/Hybrid Cloud with ECO systems

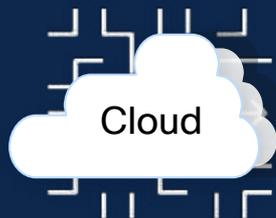
Nexus Switch (N9K)

SDN (ACI)

Compute Platform (UCS, HX)

SAN (MDS)

# Hyperscale Infrastructure Track



Cloud Transformation  
with ECO system



Cost & Operation  
Effective AI  
Infrastructure



Business Continuity  
with Active/Active  
Network

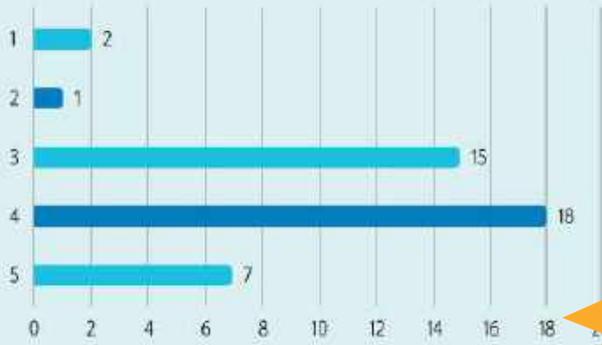
기업의 디지털 전환 가속화

고가용 아키텍처에 의한 비즈니스 연속성

# Cloud Transformation with ECO system

## 빠르고 쉬운 멀티 클라우드 구축의 새로운 접근법

클라우드 도입 및 전환 만족도는 어떠합니까?(5점만점)



출처 : 아이티데일리 2020 클라우드 디지털 트랜스포메이션 세미나 설문 결과

사례를 통한 쉽고 빠른 방안 제시

클라우드 네이티브  
애플리케이션 재정립  
(컨테이너, 클라우드 API)

AI 기반 지능화  
품질 및 비즈니스 관리  
(E2E 서비스 가시성, 선제적 예방)

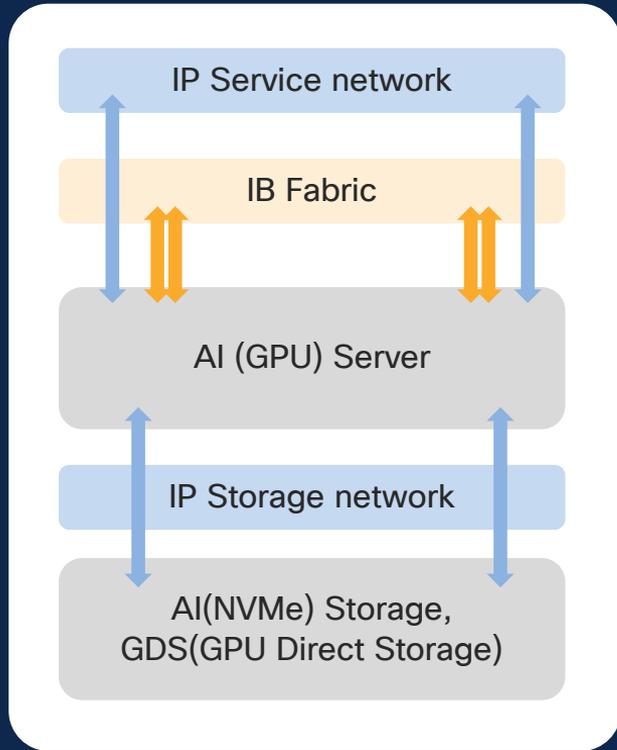
레거시 데이터센터  
인프라 현대화  
(SDDC & Private Cloud)

하이브리드/멀티 클라우드 확대



# Cost & Operation Effective AI Infrastructure

초거대 AI 시대, 네트워크 아키텍처 새판짜기

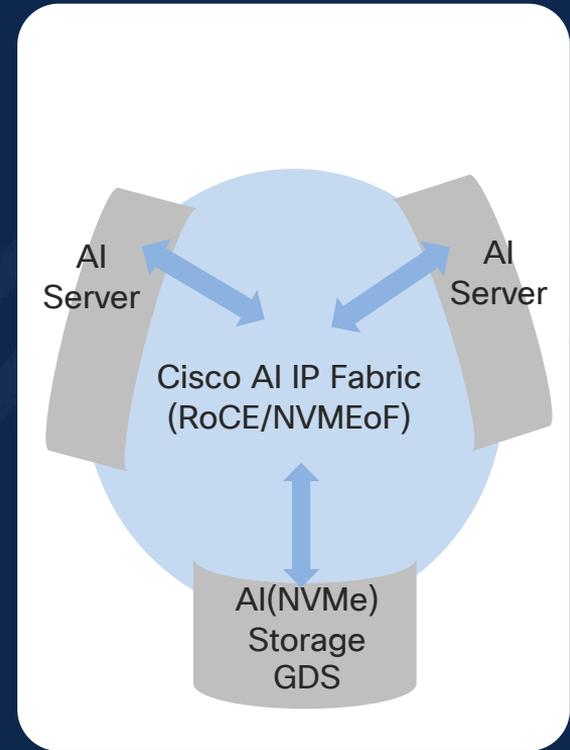


레거시 아키텍처

국내 AI 시장 급성장 :  
1조6천억(38% YoY, IDC)



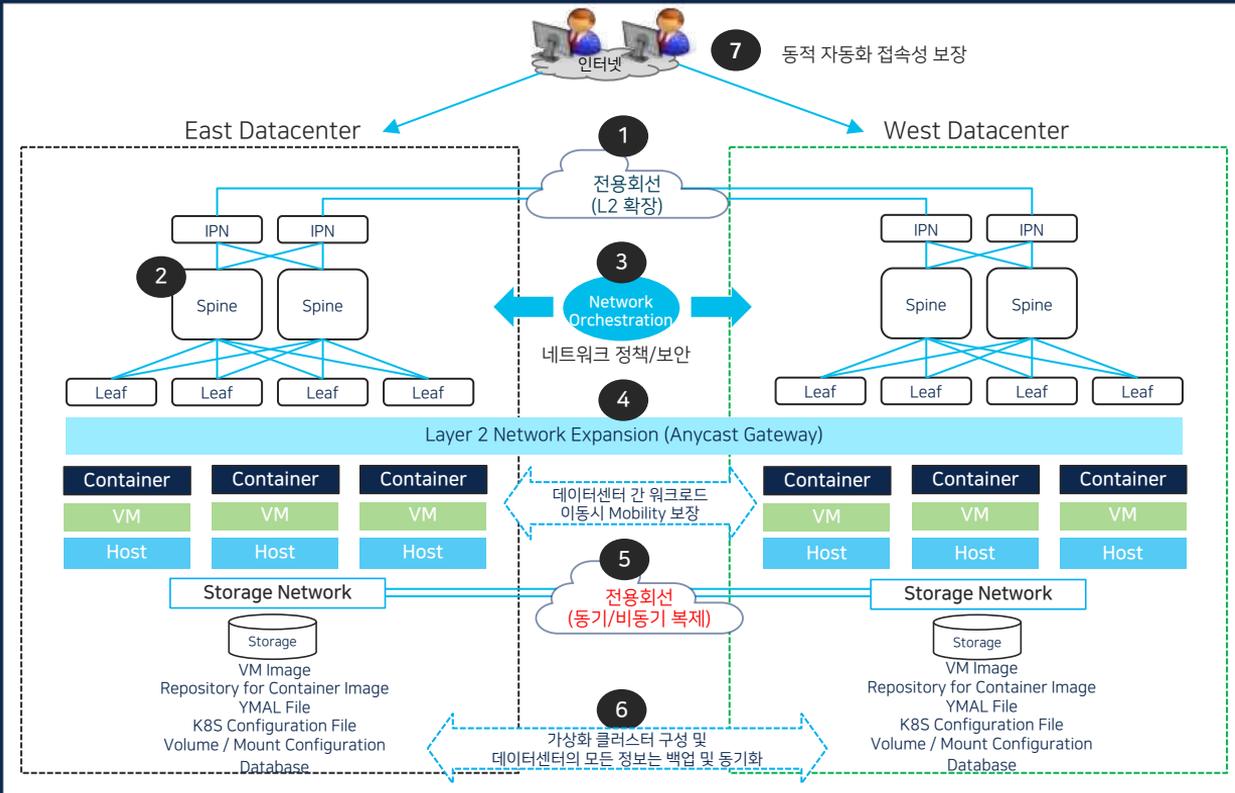
- 투자비용
- AI모니터링
- ESG(저전력)
- 프로세스시간
- 기술지원
- 구성유연성



시스코 혁신 아키텍처

# Business Continuity with Active/Active Network

비즈니스 연속성과 무중단 서비스를 위한 현명한 선택



**AADC 통합관리**

Multi-데이터센터와 Public Cloud까지 Workload와 Network의 통합관리 기능을 제공

**Cloudwave**

- ANML
- Cloud Portal
- Day/ops
- Workload
- VMware
- Container
- Network
- Multi-Cloud
- 통합관리
- Openstack
- AADC
- Hubernetes
- ICG
- Cisco ACI
- Data center

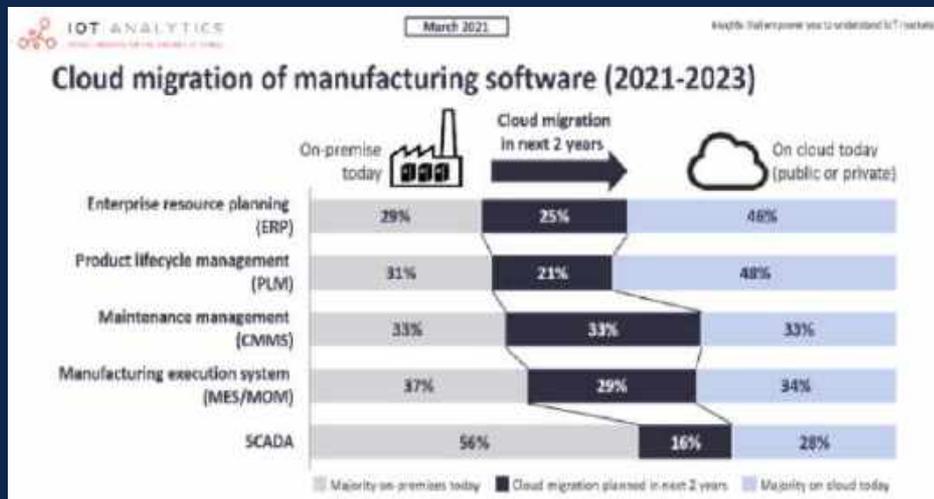
SMN SYSTEMS

# Effective Smart Factory Network

## IT/OT 컨버전스 아키텍처를 통한 스마트팩토리 진화

- 국내 스마트팩토리 시장 성장 11.4% YoY, 15.3B\$/CY26
- \* 제조 자동화, 로봇 및 IoT 연결 등 네트워크 수요 증가
- \* 공정자동화 및 지능화 품질 관리로 고품질 네트워크 요구
- 실시간 트래픽 및 지능화 장애 모니터링, 품질관리 등

- ▶ 멀티 생산라인, 공장의 구성 관리 어려움
- ▶ 개별 CLI 운영 체계, 자동화 구현의 복잡성
- ▶ Security 및 장애격리를 위한 복잡한 ACL
- ▶ 이상징후 감지 및 장애분석의 어려움



중앙 통제적 구성 관리 및 운영 자동화

팩토리 보안 강화 및 구성 유연성

지능화 패브릭 모니터링 & 신속 원인 분석



The bridge to possible

감사합니다.

CISCO *Connect*

#CiscoConnect



The bridge to possible

# 기업의 멀티 클라우드 네트워킹 준비하기

하정환 프로 | Cloud Infrastructure Software Group

이요한 이사 | Cloud Infrastructure Software Group

# Agenda

1



클라우드 송환 (Cloud Repatriation)

Customer Cloud Journeys

2



클라우드 네트워킹 구축

Multi-AZ | Multi-Region

3



클라우드 네트워킹 운영

Day2Ops | AIOps

4



기업의 지속가능성

Sustainability

1



# 클라우드 송환 (Cloud Repatriation)

Customer Cloud Journeys



# 클라우드 송환 (Cloud Repatriation)

Cloud Repatriation : Public Cloud 로 전환된 애플리케이션 및 워크로드를 On-prem infrastructure 혹은 Private Cloud 로 다시 돌려보내는 것



DAVID HEINEMEIER HANSSON

October 19, 2022

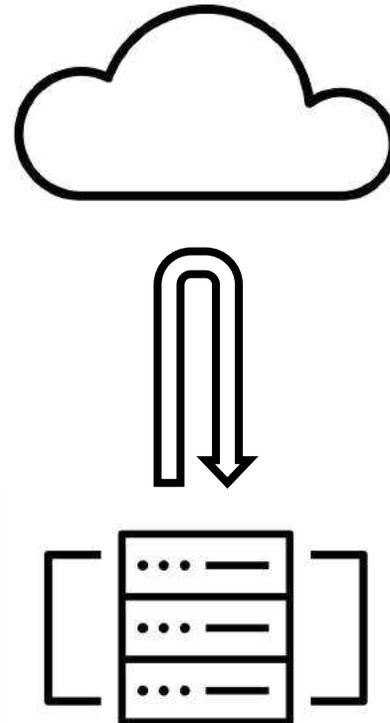
## Why we're leaving the cloud



DAVID HEINEMEIER HANSSON

February 21, 2023

### We stand to save \$7m over five years from our cloud exit



**BANK OF AMERICA** 

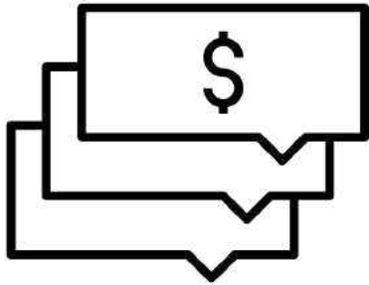
**37signals** 

 **Dropbox**

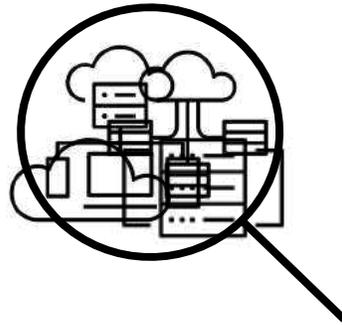
A modern living room with a white desk, a grey sofa, and large windows. In the foreground, there are several cardboard boxes, a roll of bubble wrap, and a roll of packing tape. The text is overlaid on a dark horizontal band across the middle of the image.

고객의 클라우드 송환의 배경은 무엇일까요?

# 클라우드 송환 (Cloud Repatriation)



초과비용  
Cost Overruns



운영복잡성  
Operational  
Complexity



보안  
Security

Top 3 Cloud Repatriation Drivers

53%고객은 예상하지 못한  
클라우드 비용 또는 지출 증가로 영향을 받았습니다.



33%

운영복잡성 경험  
2022

53%

End-to-end 관리의 복잡성  
2023



43%

보안을 위해  
여러 개의 클라우드 사용

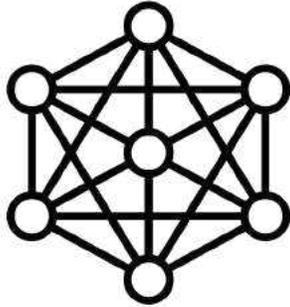
하지만 . . .

53%

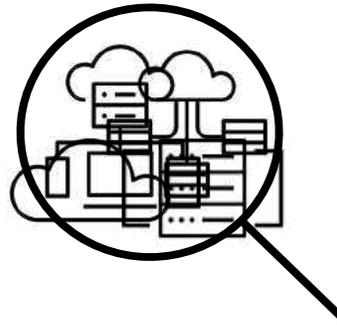
분산 및 하이브리드 워크로드에서 가장 큰  
네트워킹 과제는 **보안**



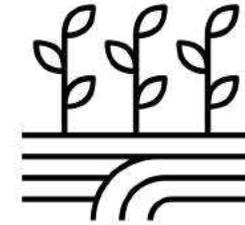
# 클라우드 송환 (Cloud Repatriation)



연결성  
Connectivity



운영복잡성  
Operational  
Complexity



지속가능성  
Sustainability

고객이 마주할 또 다른 고민

# 클라우드 송환의 고객의 과제

클라우드 네트워킹 구축하기  
Multi-AZ Multi-Region

아젠다2



커넥티비티  
Connectivity

- 서비스 배포는 점점 짧아져,
- 빠르고 안정적인 네트워킹 커넥티비티가 요구됨

클라우드 네트워킹 운영  
Day2Ops AIOps

아젠다3



운영 복잡성  
Complexity

- 기업 50%가 2024년까지 AIOps 솔루션을 채택예정
- 인텔리전트 분석은 실시간으로 근본 원인 과 트래픽 파악

지속가능성  
Sustainability

아젠다4



지속가능성  
Sustainability

- 환경, 사회 및 거버넌스 모두 ESG목표를 달성하는 것이 매우 중요
- 장기적으로 회사의 현재 및 미래의 직원, 주주 모두 요구

2



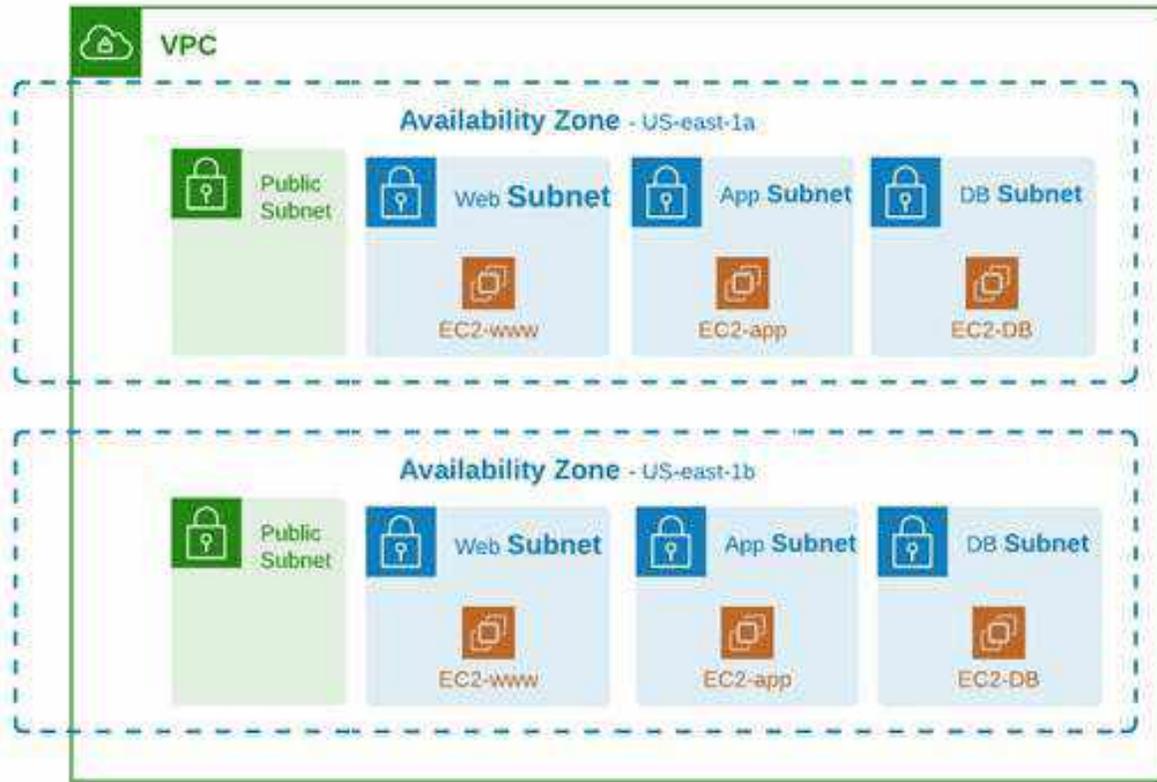
# 클라우드 네트워킹 구축

Multi-AZ | Multi-Region

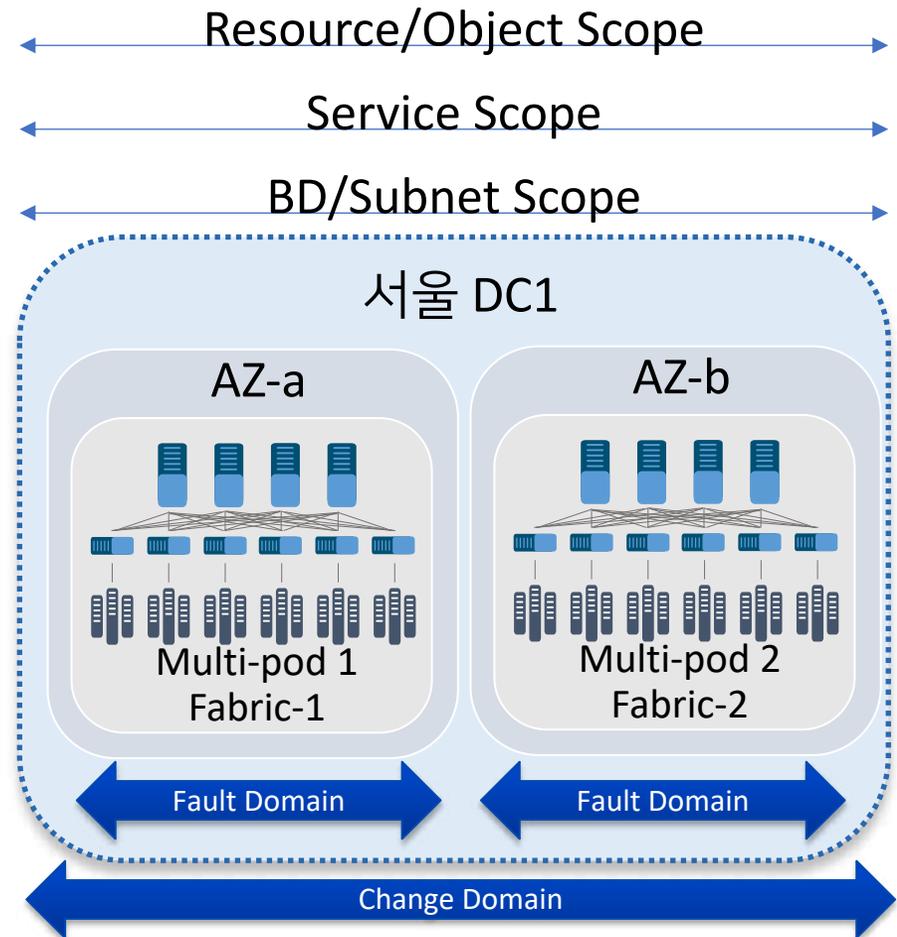


# (온-프레임 클라우드에서도) 가용영역 Multi-AZ 구축하기

퍼블릭 클라우드의  
AZ 네트워킹

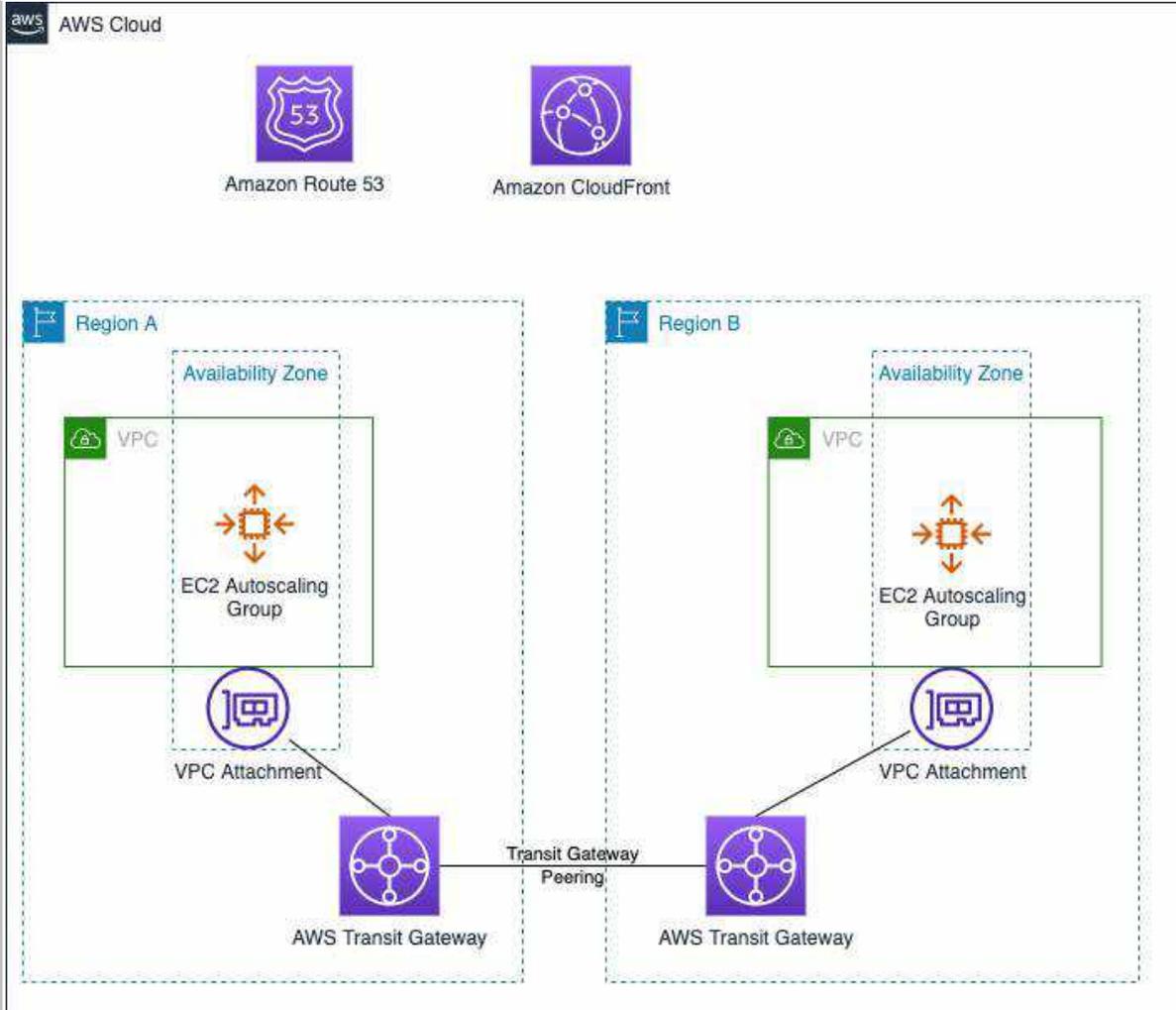


시스코가 제안하는 온-프레임  
클라우드 네트워킹

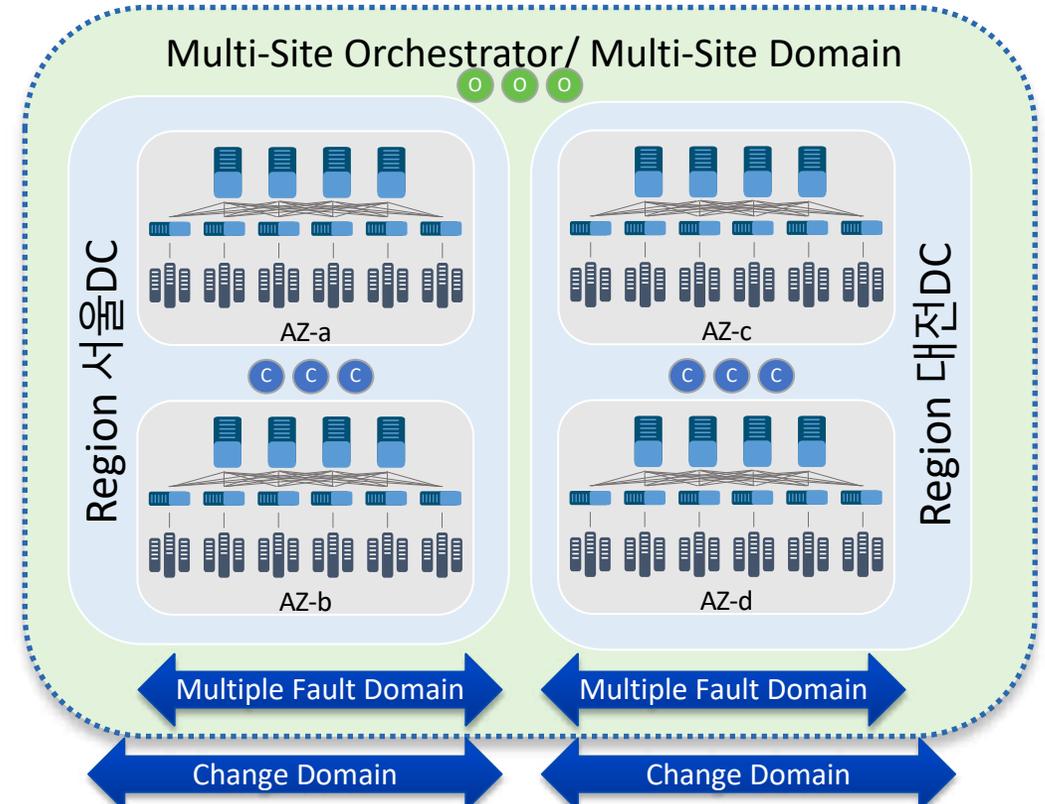
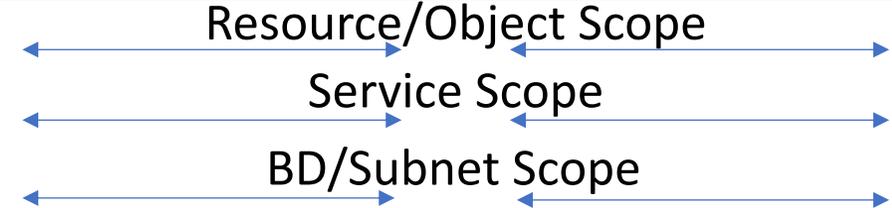


# (온-프레임 클라우드에서도) 가용지역 Multi-Region 구축하기

## 퍼블릭 클라우드의 Multi-Region 네트워킹



## 시스코가 제안하는 온-프레임 클라우드 네트워킹



# One View

데모

## My Sites

Map Table



3



# 클라우드 네트워킹 운영

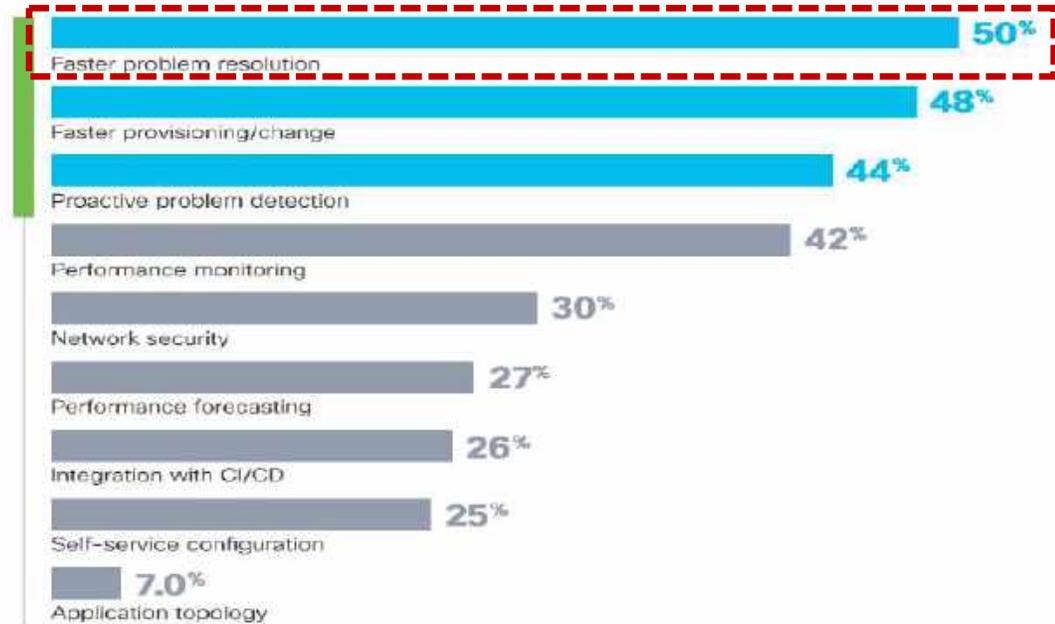
Day2Ops | AIOps



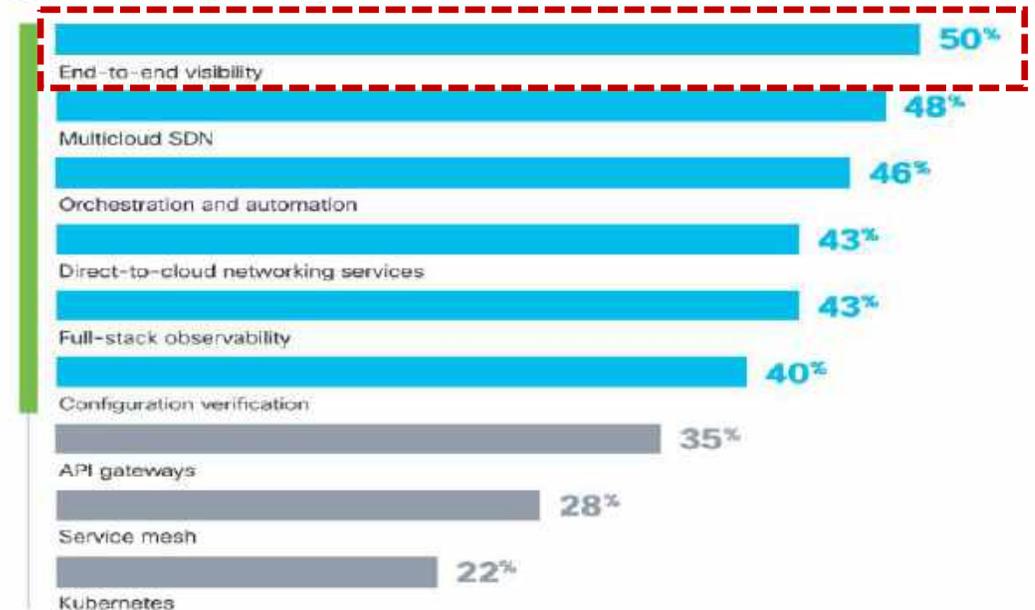
# 2023 Global Networking Trends Report

하이브리드클라우드 네트워킹 환경에서 가장 요구되어 지는 기능 : 빠른장애해결  
분산된 워크로드 관리에 도움을 주는 네트워킹 기술 : End to End 가시성 확보

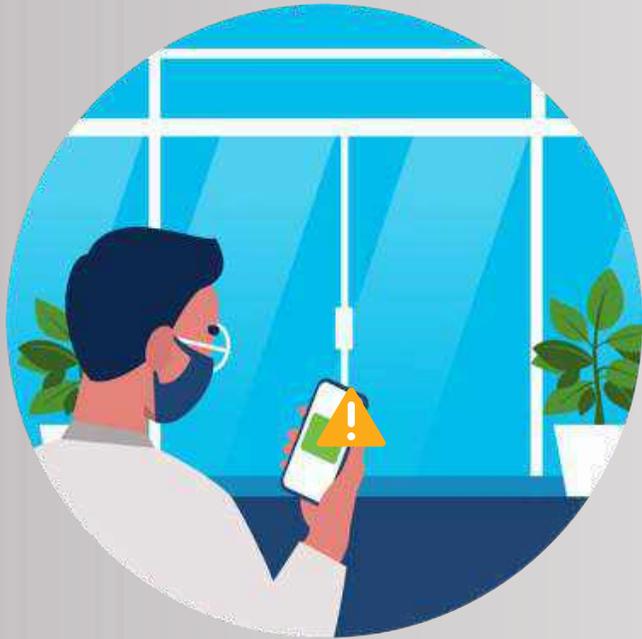
What features or benefits of hybrid cloud networking are most in demand?



Which of the following networking technologies are you considering or do you use to help manage distributed workloads?



# 사전 대응적 AIOps 운영: 고객의 서비스의 가용성 및 가속화



- ✓ 외부 네트워크 또는 다른 사이트의 네트워크와 관련하여 상관관계를 파악합니다.
- ✓ AI 기반의 보증 및 자동화를 사용하여 문제를 해결
- ✓ 성능 저하의 신속한 식별과 복구를 가능하게 합니다.
- ✓ 실시간으로 네트워크를 모니터링 및 분석하여 이상 징후를 식별
- ✓ 근본 원인 분석 및 용량 계획을 제공하고, 문제 해결을 가속화할 수 있는 기능



# “이상징후 알람에 대해, 시스코는 아래와 같이 답변할 수 있습니다.”

## Interface Down

Refresh Actions

### What's wrong?

Port Channel vpc4 configured on Leaf201 and Leaf202 is currently down.

Anomaly Level Warning Status Active

Category: Connectivity Site: San Jose ACI Detection Time: March 29, 2022, 10:00:00 AM

### What triggered this anomaly?

Filter

Interface	Aggregation Type	Active Member Ports	VPC Domain	Switch 1	Switch 2	VPC Neighbor
vpc4	LACP Active	0/2	10	Leaf201	Leaf202	NSK-Switch 1 via CDP

1 Items found Rows per page: 10 < 1 2 ... 10 >

### What's the impact?

24 Endpoints currently have no connectivity.

### How do I fix it?

Recommended Solution (1/2)

Please check that interfaces Eth1/1 and Eth1/2 on Leaf Leaf201 and Leaf Leaf202 respectively are correctly plugged and that the cable is damaged.



What is wrong?

현재 무엇 **what**이 잘못되어 있는지 알람



What triggered this anomaly?

이 이상 징후의 트리거 **trigger**는 무엇입니까?



What is the impact?

영향 **impact**은 무엇입니까? 아무것도 하지 않으면 어떻게 되나요?



How do I fix it?

어떻게 **how** 해야 문제가 해결되는지 권장 사항 및 가능하면 원-클릭 교정

4



# 기업의 지속가능성

Sustainability



# 기업의 지속가능성 (Sustainability)

APJC 기업들 비즈니스 관행에서 지속가능성을 수용하기 시작

The State Of  
Sustainability Among  
Large Firms  
In Asia Pacific<sup>1</sup>,  
Forrester Research,  
10/21

33%

Named a  
sustainability lead

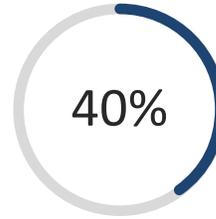
55%

Have greenhouse  
emission targets

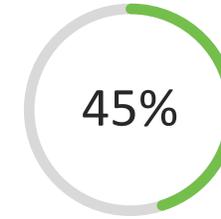
41%

Set a Carbon Neutral  
Date

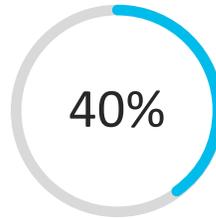
IDC, 2024년까지 지속가능성이 아시아 태평양 기업의 비즈니스  
관행에 포함될 것으로 전망



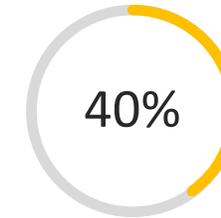
주요 기업들의  
40%는  
지속가능성을  
운영에 포함시킬  
예정



45%의 조직은  
지속가능성을 기술  
구매 결정에 있어  
매우 중요한 고려  
사항으로 간주



주요 기업들의  
40%는  
지속가능성을  
비즈니스 KPI에  
포함시킬 예정



40%의 TOP  
기업들은  
지속가능한 혁신에  
참여

1. Forrester research on 127 large companies in Australia, China, Hong Kong, India, Indonesia, Japan, Malaysia, New Zealand, Singapore, South Korea, Taiwan, and Thailand.

# 지속가능성 (Sustainability) 디자인



## 가시성 (Visibility)

네트워크 에너지 소비 및 비용에 대한 가시성 확보



## 탄소 풋프린트 (Carbon footprint)

장비들의 온실 가스 greenhouse gases (GHG) 의 이해



## 주요 지속 가능성 메트릭 (Key sustainability metrics)

총 전력 사용률 Power Utilization Effectiveness (PUE)



# Emissions

About 80% of your energy this month came from low-carbon sources on average

데모  
베타 버전이며, 현시간은 로드맵!

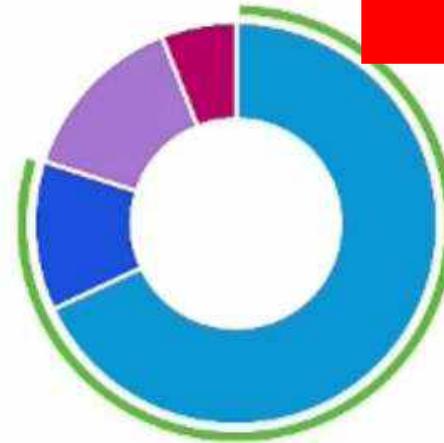
## Emissions this Month

# 194 kg CO<sub>2</sub>e

Estimated monthly carbon dioxide equivalent emissions (based on energy usage alone) from this site only. Emissions are estimates from utility data and third-party services. [See Methodology](#)

**Equivalent of 0.75 car trips**  
of 1070 km each in a gas-powered passenger vehicle. [See More](#)

## Energy Mix



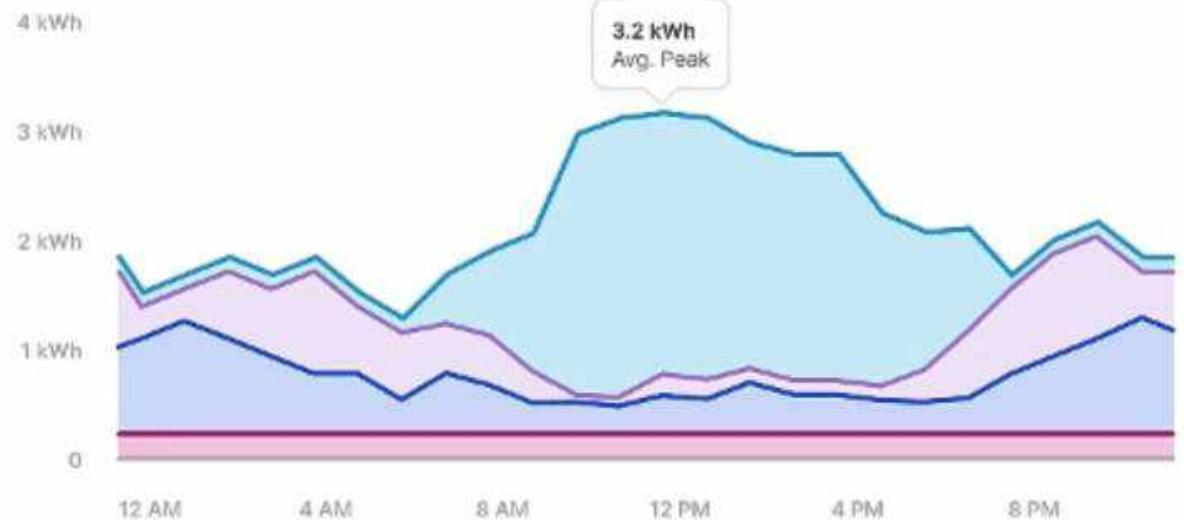
**80%** Average percentage of energy from low-carbon sources

- Low-carbon
  - Solar 68%
  - Wind 12%
- Other sources
  - Natural Gas 14%
  - Coal 6%

## Daily Energy Mix

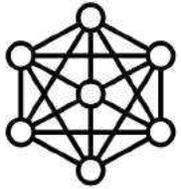
Amount of total energy used by source, over all of the days for November

- Low-carbon
  - Solar 68%
  - Wind 12%
- Other sources
  - Natural Gas 14%
  - Coal 6%



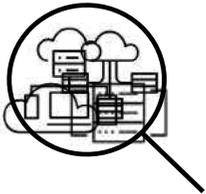
# Summary

## “클라우드의 온프레미스 송환 (Repatriation)을 고민하는 기업의 증가”



### Connectivity

다양하고 분산된 클라우드 환경을 위한 네트워킹 준비 필요



### Operational Complexity

클라우드 환경에서 전통적인 운영방식은 더 이상 통하지 않음



### Sustainability

클라우드 사용유무와 관계없이 기업의 지속가능성은 매우 중요하며, 정량적 지표로써 관리되어야 함



The bridge to possible

감사합니다.

CISCO *Connect*

#CiscoConnect

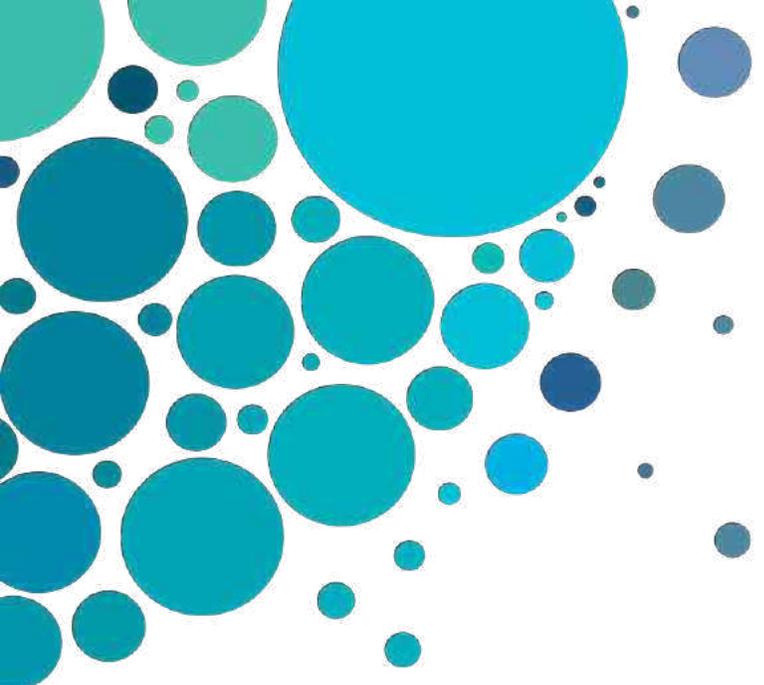
CISCO *Connect*

ALL IN

#CiscoConnect

# 초거대 AI 시대, 네트워크 아키텍처 새판짜기

최수영 상무, 임규현 프로 | 시스코코리아



# Agenda

- AI 마켓 트렌드
- AI 인프라 구조
- 고성능 AI를 위한 네트워크 고려사항
- AI Network를 위한 시스코 네트워킹 제안
- 요약

# AI 마켓 트렌드



# 생성형 AI 활용 사례

Gamma.app 을 이용한 여행 정보 프리젠테이션 생성 (챗대리, 코대리)

**옐로스톤 국립공원, 당신을 몰입시킬 여행의 시작**

여기는 자연이 만든 내셔널 지오그래픽의 땅입니다. 우리의 황해 아래에서 가장 아름다운 국립공원 중 하나를 탐험해보세요.

by Soo Yung Chol



**일정 및 캠핑 정보**

**캠핑**  
옐로스톤 국립공원 내 12개의 캠핑장 중에서 선택 가능합니다. 온수 샤워와 RV용 전기가 제공됩니다.

**일정**  
3일간의 루어르 패키지를 구성해드립니다. 첫째날, 이곳에서 살아 숨쉬는 조류들과 함께하는 자연체험을 떠나갈 것입니다. 둘째날, 화산성 가운이 가득한 숲속에 놀러 오세요. 셋째날, 옐로스톤 역사적인 유산을 탐험해보세요.



**손님의 안전과 주의사항 안내**

- 야생동물과의 만남이 는 조심하세요.**  
만드시 안전을 고려해 동지 하세요. 어떠한 낙상한 세력 신형, 이국적 행태지, 나쁜 것 들한 그것들까지도...
- 행동 지침**  
계발지 안전과 다른 이들이 존중을 위해 공해대에서 벗 을 지켜주세요.
- 기타 정보**  
관광객은 통행료 유한에 대 해 인식하고 이를 추궁해야 합니다.

**Famous Attractions**

Experience the natural beauty of Yellowstone National Park's most iconic attractions, including Old Faithful, Grand Prismatic Spring, and the Yellowstone Grand Canyon. Witness the power of nature with geysers, hot springs, and stunning vistas that will leave you in awe.



**Explore Yellowstone's Natural Wonders with a Three-Day Itinerary**

Discover the best of Yellowstone National Park with this three-day itinerary. From geysers and hot springs to wildlife and stunning vistas, this route will take you on a journey through the park's most iconic attractions. Don't forget your camera!

**Day 2: Wildlife & Scenic Views**

Get an early start to catch the park's wildlife in action. Look for bison, elk, and maybe even a grizzly bear. Then take a scenic drive along the Yellowstone River to see the park's stunning waterfalls and canyons.

1 2 3

**Stay in Comfort at Yellowstone's Best Hotels and Lodges**

Relax and unwind after a day of exploring Yellowstone National Park at one of the many luxurious hotels and lodges in the area. Choose from a range of accommodations, from rustic cabins to elegant resorts, and enjoy world-class amenities and breathtaking views of the park.

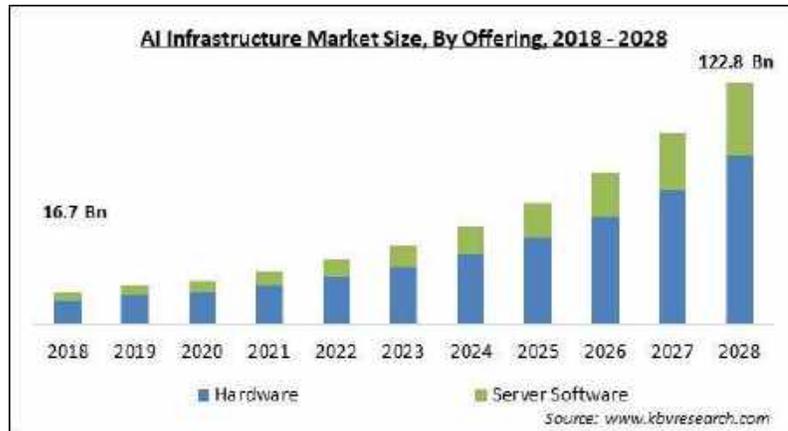


Yellowstone Lodge      Yellowstone Cabin

Source : gamma.app

# 해외 AI 비즈니스 마켓 트렌드

- 글로벌 AI 인프라 시장은 2028년까지 연평균 24.7% 성장 예상
- 주요 동력은 인공지능 기술의 발전과 함께 증가하는 AI 모델의 수요
- Microsoft 365 Copilot 발표 23.03.16
- Google Bard 서비스 출시 23.03.21
- Microsoft GPT-4 기반 New Bing 서비스 출시 23.05.04



www.mk.co.kr > news

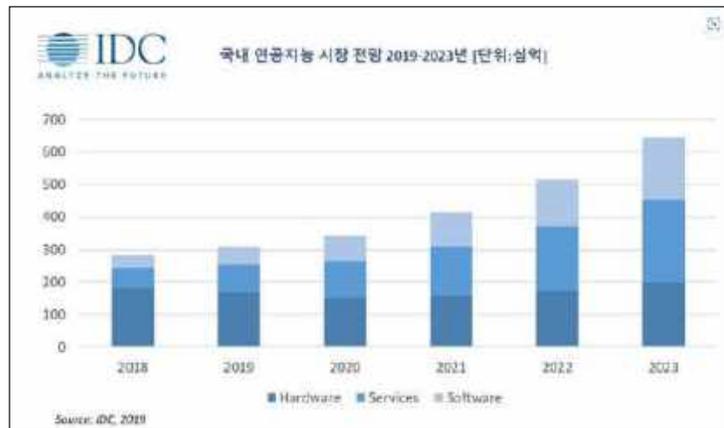
구글, 챗GPT 같은 AI 공개... 람다 기반 바드(Bard) 인공지능 - 매일...  
2023.02.07. 블로그 통해 바드(Bard) 개인 발표 피차이 CEO '우주일' 내에 일반에 공개 API 인동해 수입은 스타트인 활용 가능 딥마인드 등 계열사 공동원 인공지능 개발 구글이 공개한 인공지능 바드 <출처> 구글 > 구글이 만든 챗봇인 챗GPT와 본격.

중양일보 [특] 64면 1단 2023.05.25. 데이비뉴스  
챗GPT 흥에 업은 MS, 윈도에 AI 비서 '코파일럿' 탑재  
앞서 MS는 오피스 계층에 생산 AI를 위한 '마이크로소프트 365 코파일럿' 보안에 생산 AI를 위한 '사쿠라디 코파일럿' 등을 발표했다. 이날 MS는 마이크로소프트 3...  
MS, 윈도 11에 생산 AI 비서 '코파일럿' 넣는다. 뉴스1 2023.05.25. 데이비뉴스

아이뉴스24 PICK 4월 권 데이비뉴스  
엔비디아, 고성능 AI 지원 슈퍼컴퓨터 'DGX GH200' 공개  
엔비디아 DGX 슈퍼컴퓨터는 엔비디아 GH200 그레이스 호퍼 슈퍼칩과 엔비디아 NV링크 스위치 시스템을 탑재했다. 고성능 AI 인의 애플리케이션과 추론 시스템...  
엔비디아 AI 영역 확장하는 슈퍼컴퓨터 'DGX GH200' 발표 델로더 4월 권 1억사골롭(TeraFLOPs) 성능 제공... 엔비디아, DGX GH200... IT비즈뉴스 4월 권

# 국내 AI 비즈니스 시장 트렌드

- 국내 AI 시장은 매년 17% 상승 예상
- AI 전체 비용 중 30%가 하드웨어 인프라 투자
- 네이버, 카카오 등 국내 빅테크 기업들에서 NVIDIA GPU를 구매하여 서비스 개발 중
- 광주 AI 데이터센터 NVIDIA H100 GPU 1000개 도입
- 국내 스타트업들이 대기업과 손잡고 AI 반도체 동맹 결성 (해외 벤더와의 AI 주도권 경쟁)



**AI타임즈**  
광주 AI데이터센터에 국내 최대 규모 슈퍼컴 인프라 들어선다  
광주지능산업협약사업단(단장 김순하)이 오는 10월 오픈 예정인 AI데이터센터에 엔비디아 GPU가 도입된다.

**파이낸셜뉴스**  
네이버클라우드, 엔비디아·아이크로진과 '유전체 분석' 시장 확장한다  
(서울=뉴스1) 김근욱 기자 = 네이버클라우드가 엔비디아(NVIDIA), 유전체 분석 전문...

**조선비즈**  
국내 AI 스타트업 리벨리온 "AI 벤치마크서 엔비디아·힐컴 압도" - 조선비즈  
국내 인공지능 반도체 스타트업 리벨리온은 AI 반도체 기술력 검증 플랫폼(MLPerf)을...

**"한국의 엔비디아 키운다" KT, '리벨리온'과 AI 반도체 동맹**  
국내 AI 반도체 스타트업 리벨리온에 300억원 투자, AI 반도체 사업 본격 진입  
하이퍼스케일 AI 컴퓨팅 통해 외산 GPU 의존도 극복. 국산 AI 반도체 설계·조립  
국내 최초 AI 플랫폼 '위한' 확보. 초대규모 GPU만 조성 후 전용 서버로 구축  
습수 국산 AI 가솔트 클로빌 경쟁, 판로 확대. 국가 AI 반도체 경쟁력 강화에 기여

# AI 인프라 구조



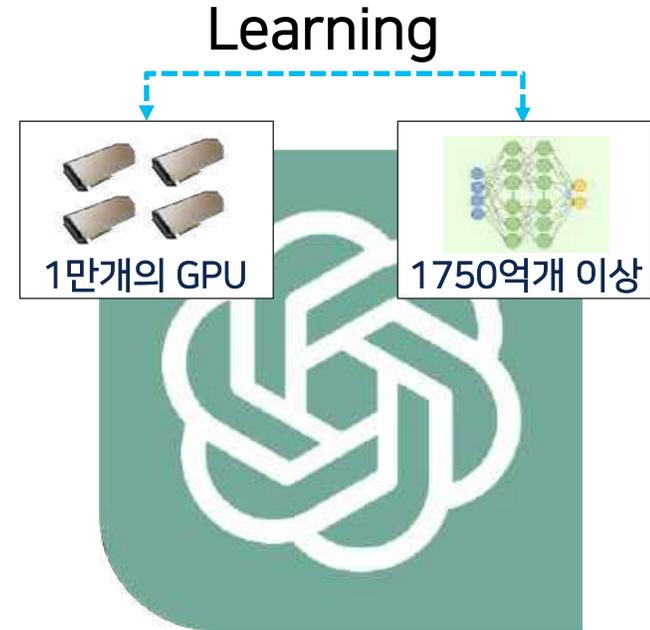
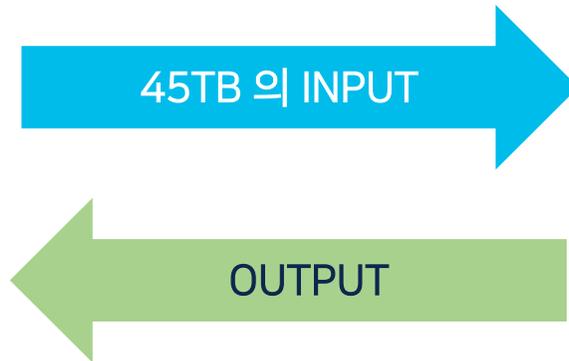
# AI에서 왜 GPU가 핵심인가?

생성형 AI와 같은 ChatGTP-3.5 의 경우 1개 GPU로 약 355년의 러닝시간이 필요  
GPU의 갯수를 늘릴수록 러닝 과정은 획기적으로 줄일 수 있는 구조



**45 TB**

이미 인터넷에 존재하는 정보를 기반으로



**수백 GB 에서 수 TB**

(매개 파라메타 1750억~ 1조개 X 4Byte)

# 고대역폭 네트워크가 요구되는 GPU 성능

NVIDIA V100 이후 수십배 빠른 H100 GPU 출시

여러대의 서버로 GPU 클러스터를 구성하기 위해서는 400Gbps 네트워크가 요구됨

NVIDIA V100 GPU

Network 100Gbps

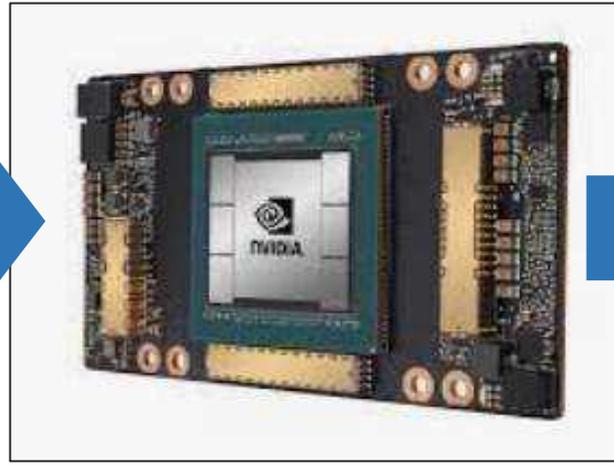


NVLink : 300GB/s

7X

NVIDIA A100 GPU

Network 200Gbps



NVLink : 600GB/s

6X

NVIDIA H100 GPU

Network 400Gbps



NVLink : 900GB/s

대부분의 서버는 25Gbps

현재

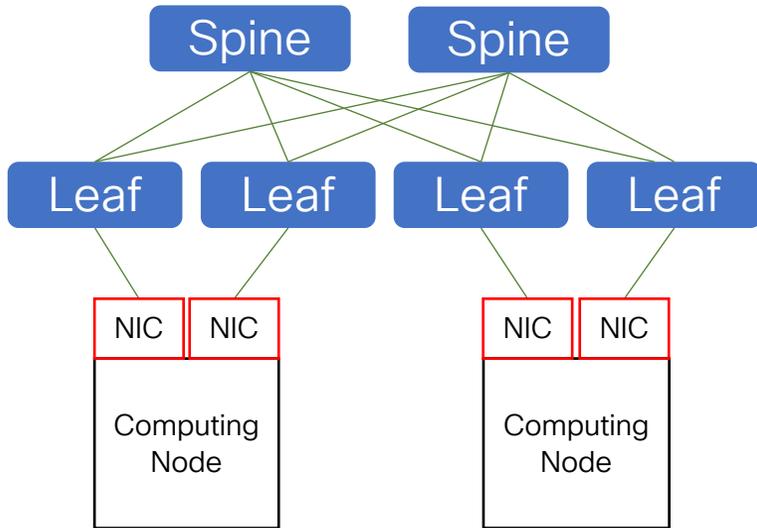
H100이 시장에 주류가 될 수록  
400Gbps 이상의 고대역 네트워크  
수요가 늘어날 것으로 전망

# AI 인프라의 시스템 연결 구조 비교

AI 인프라에는 고성능 GPU간 연결성을 보장하기 위한 네트워크 아키텍처 필요

## 전통적인 서버 연결

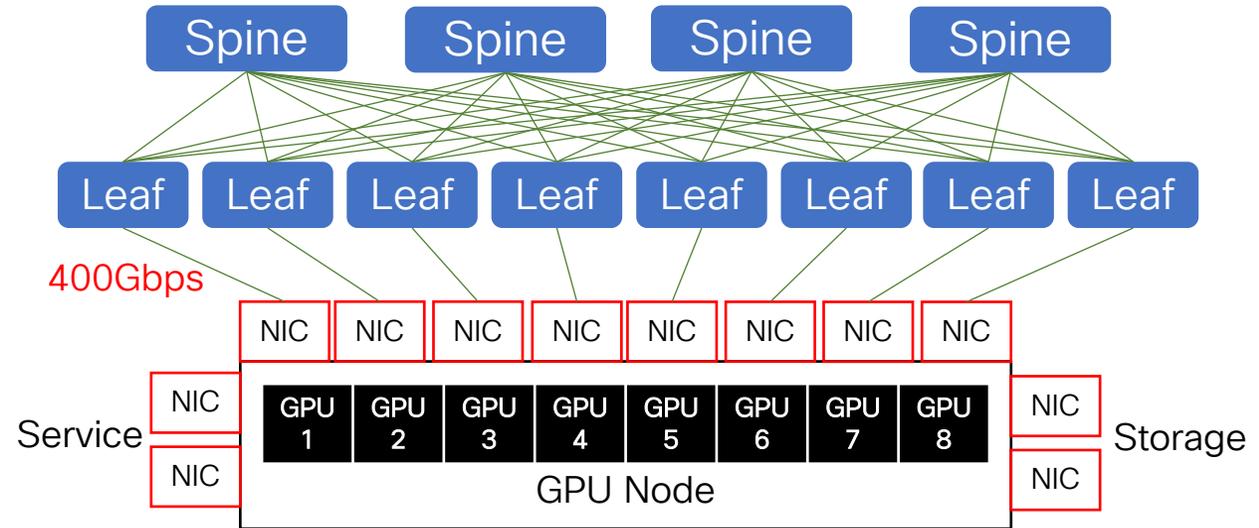
CPU에서 연산된 결과 중심의 네트워크



서버 당 2개의 10G/25G 연결

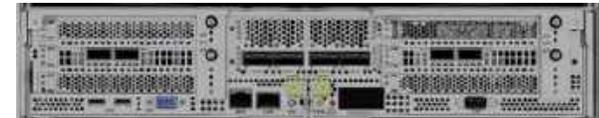
## GPU 서버 연결

GPU 연산속도에 적합한 고성능 네트워크



GPU 당 1개의 400G 연결 (H100 기준)

네트워크 요구사항이  
기하급수적으로 증가

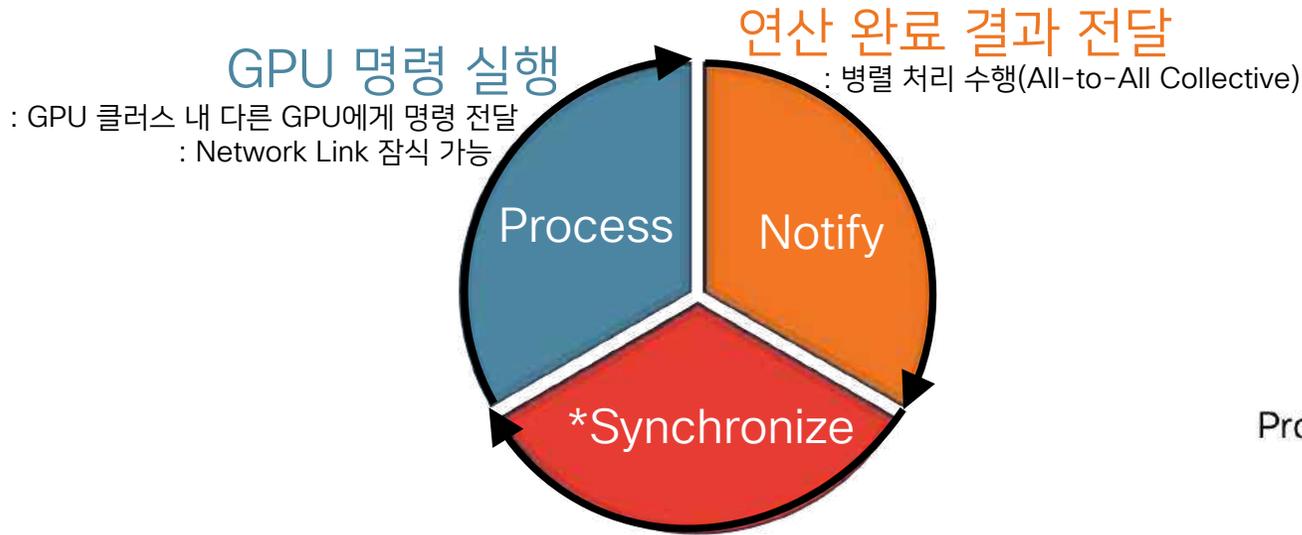


# 고성능 AI를 위한 네트워크 고려사항

# AI Workload Challenge - 트래픽 패턴

전통적 Data Center와 다른 트래픽 패턴

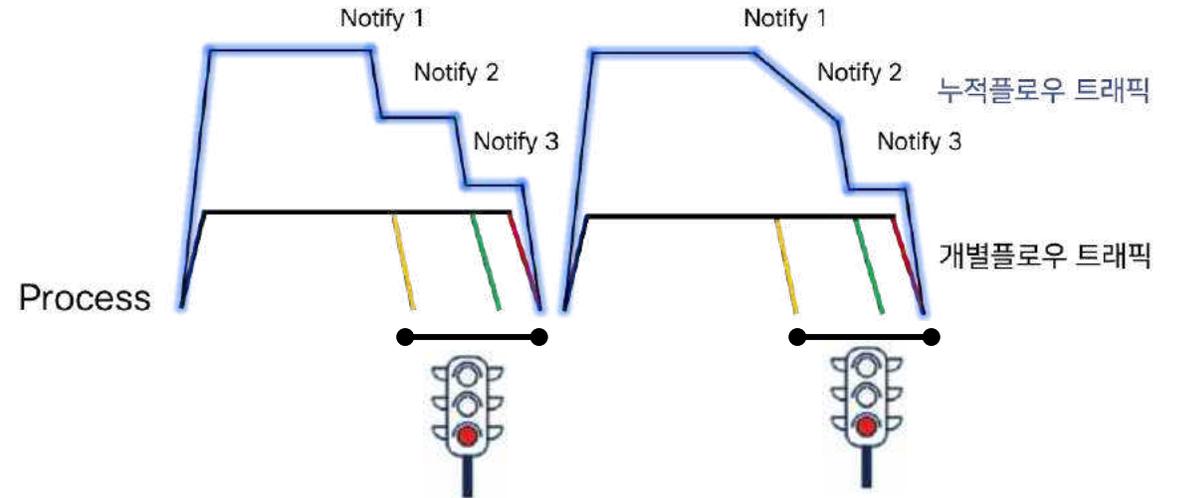
## AI 프로세스에 따른 트래픽 패턴



## 연산 완료 결과 수신 대기 : JCT

: GPU 클러스터 내 모든 GPU로 부터 대기  
: 기존 트래픽 패턴에 없는 AI에서 추가된 패턴

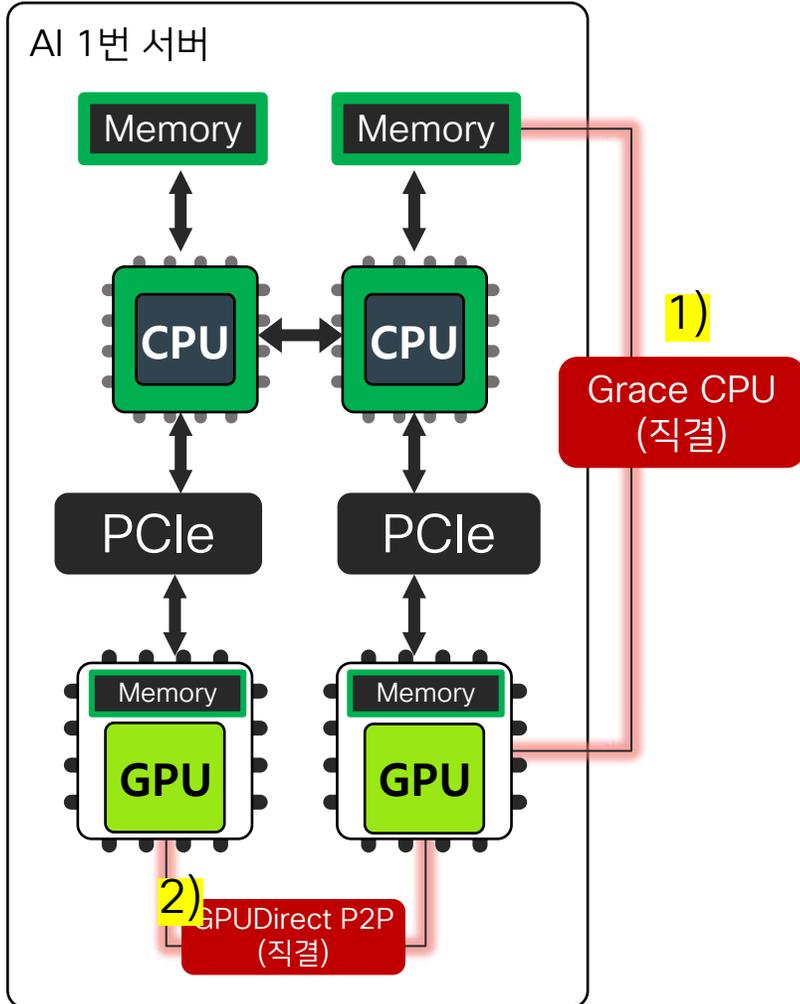
## 고성능 AI 분석을 위한 Lossless/Low Latency 구현 필요



연산 미완료 GPU로부터 Notify 수신대기

연산 미완료 GPU로부터 Notify 수신대기

# AI 서버 내 병목을 줄이기 위한 기술



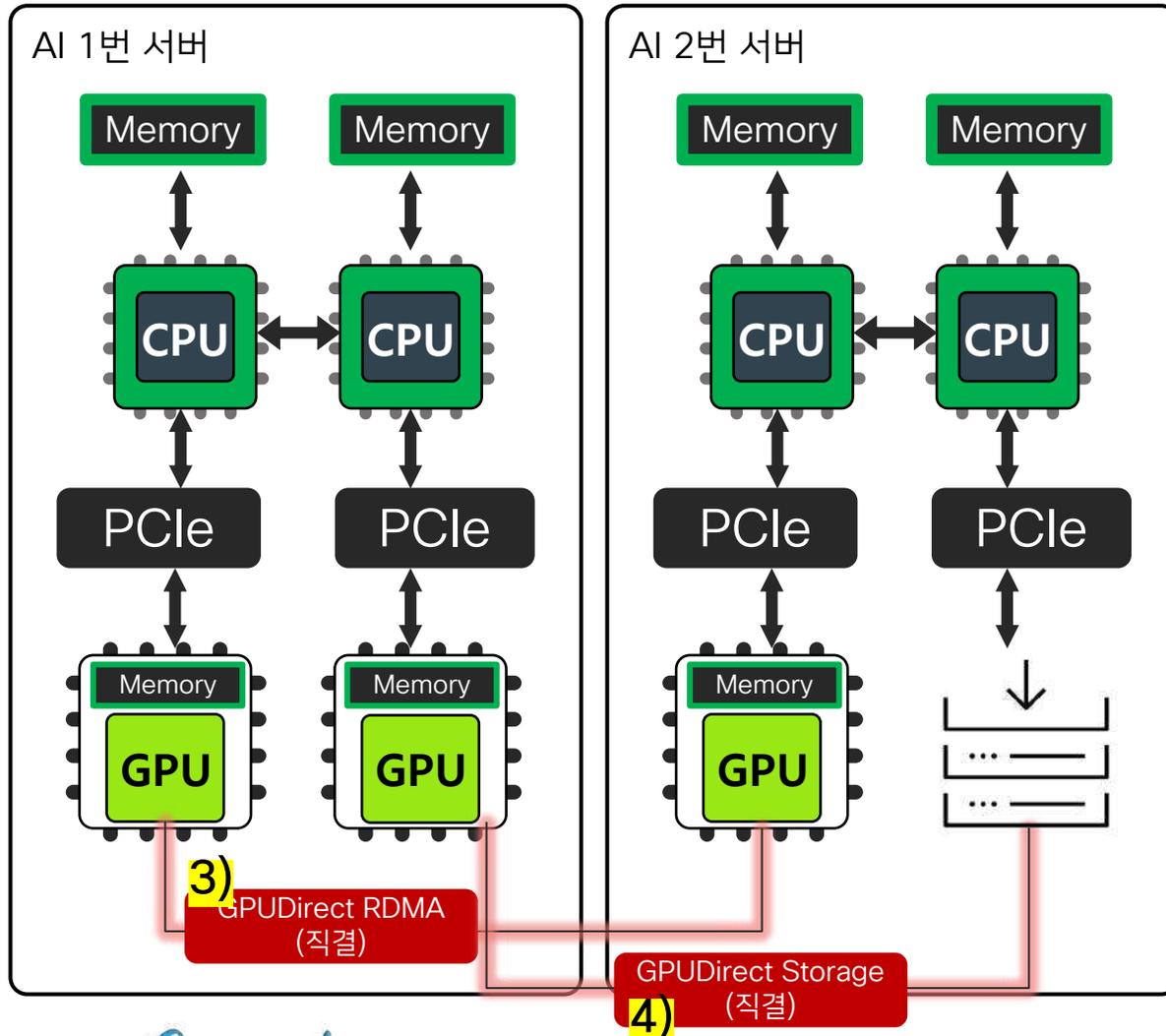
## 1) Grace CPU

- : GPU 내 메모리가 모자랄 경우 Host Memory 참조 필요
- : GPU <-> Host Memory 간 병목제거

## 2) GPUDirect P2P

- : 1개 이상의 N개의 GPU간에 병렬 처리를 통한 딥러닝 필요
- : GPU 0번 <-> GPU 1번 간 병목 제거

# AI 서버 간 병렬처리 시, 병목을 줄이기 위한 기술



## 3) GPUDirect RDMA

: 서버와 서버간의 통신을 통해 딥러닝 분석하는 경우 NIC이 호스트 메모리까지 다녀오지 않고, GPU에 있는 메모리를 바로 읽을 수 있게 사용

## 4) GPUDirect Storage

: Local Storage 또는 원격 스토리지의 데이터를 곧바로 참조

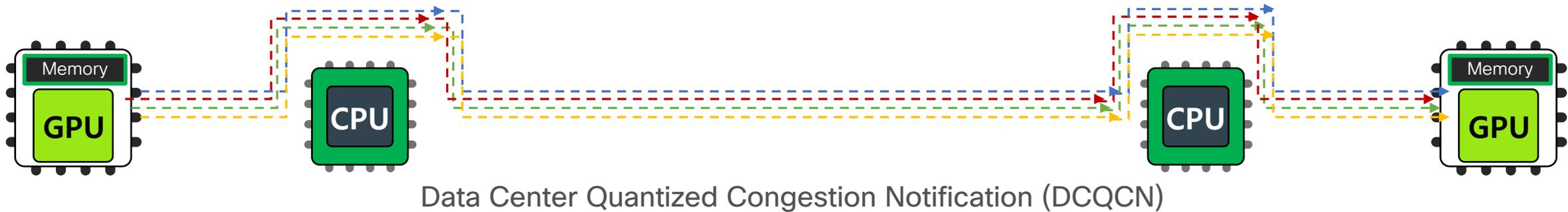


네트워크에서 GPUDirect 통신 시 병목을 제거하며 전송하는 기술

# RoCEv2 (RDMA over Converged Ethernet)

# RoCEv2 와 RDMA (Lossless Network)

핵심 컨셉 : CPU를 미경유 → CPU 보호 및 성능의 극대화



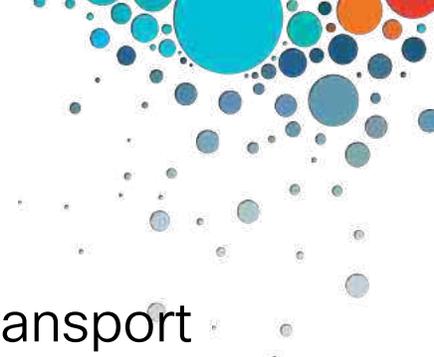
DMA(Direct Memory Access) : 컴퓨터 주변 장치들이 CPU를 거치지 않고, 메모리에 직접 접근할 수 있는 기술

RDMA(Remote Direct Memory Access) : DMA를 네트워크를 통해 원격지(Remote)로 확장하는 기술

\*RoCE(RDMA Over Converged Ethernet) : RDMA 통신을 전달하는 이더넷 기술

DCQcN(Data Center Quantized Congestion Notification) : 위 기술들을 조합해서 만드는, Lossless Network 체계

# RoCEv2 와 RDMA



Protocol	Network Protocol	Network Transport Requirements
	TCP	Lossy
GPUDirect, NVMe	RoCEv2	Lossless
	iWARP (TCP)	Lossy

## (Lossy) TCP/IP 방식의 Flow Control



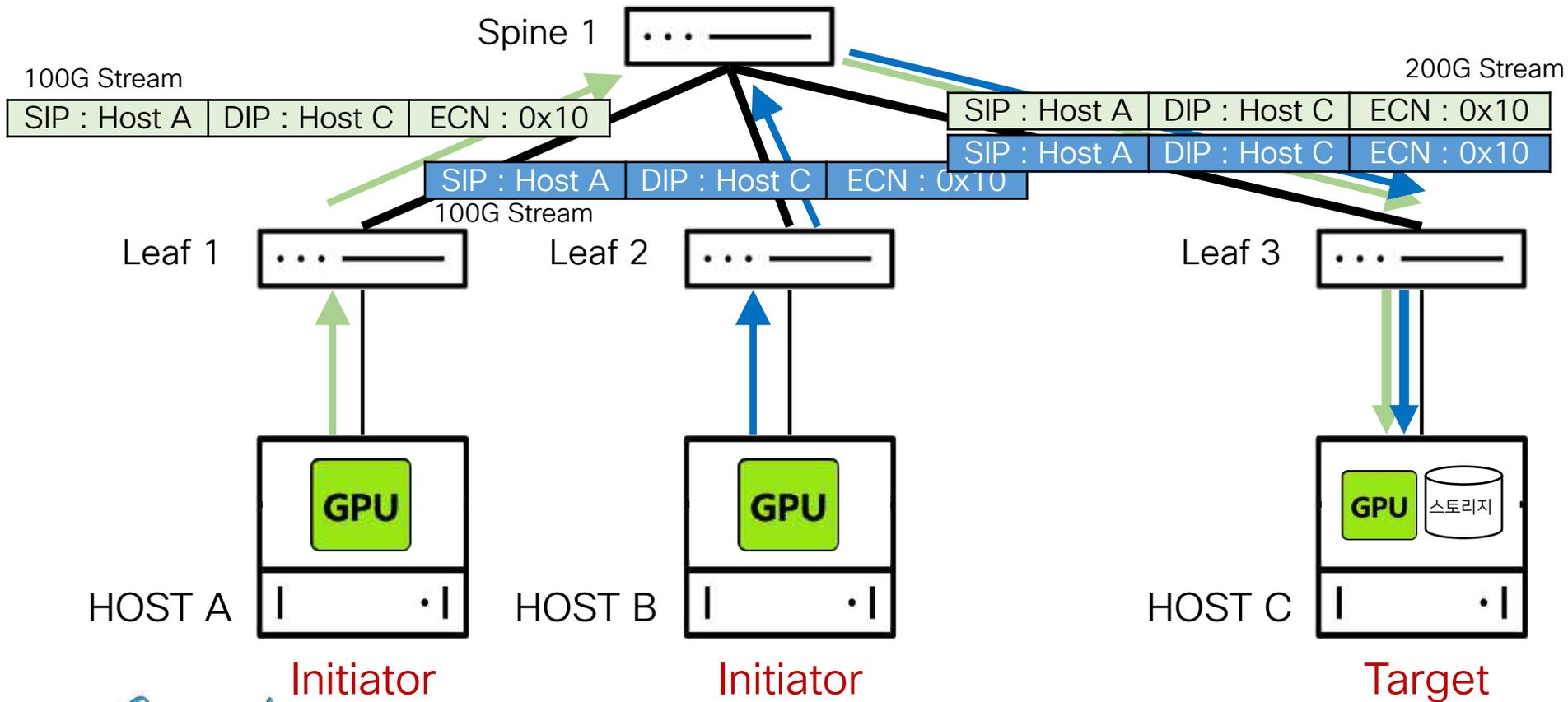
## (Lossless) RoCEv2 방식의 Flow Control

TCP와는 다른 Flow Control 방식 사용  
(뒷장계속)

# DCQCN - RoCEv2 기반의 흐름제어 방식(Lossless)

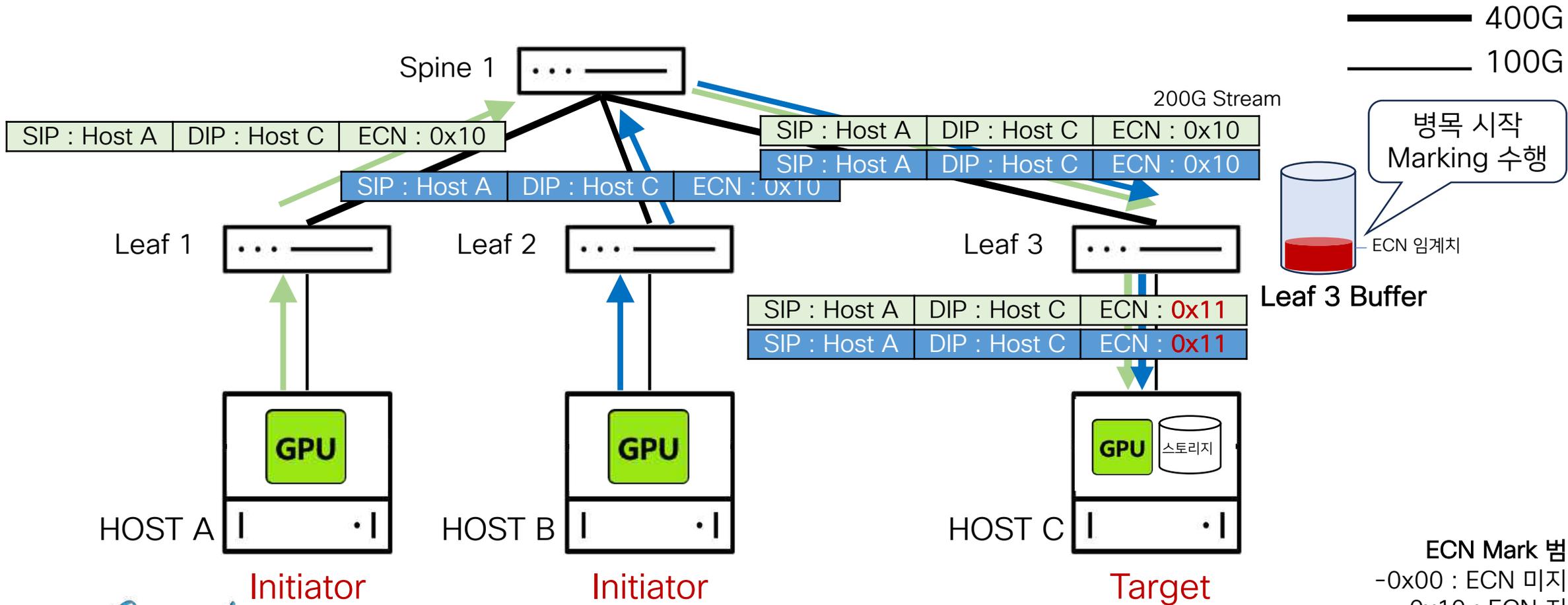
- AI 서버, NVMe 스토리지와 상호 작용을 수행하는 네트워킹
- Flow Control : ECN과 PFC를 이용하는 흐름제어

———— : 400G  
———— : 100G



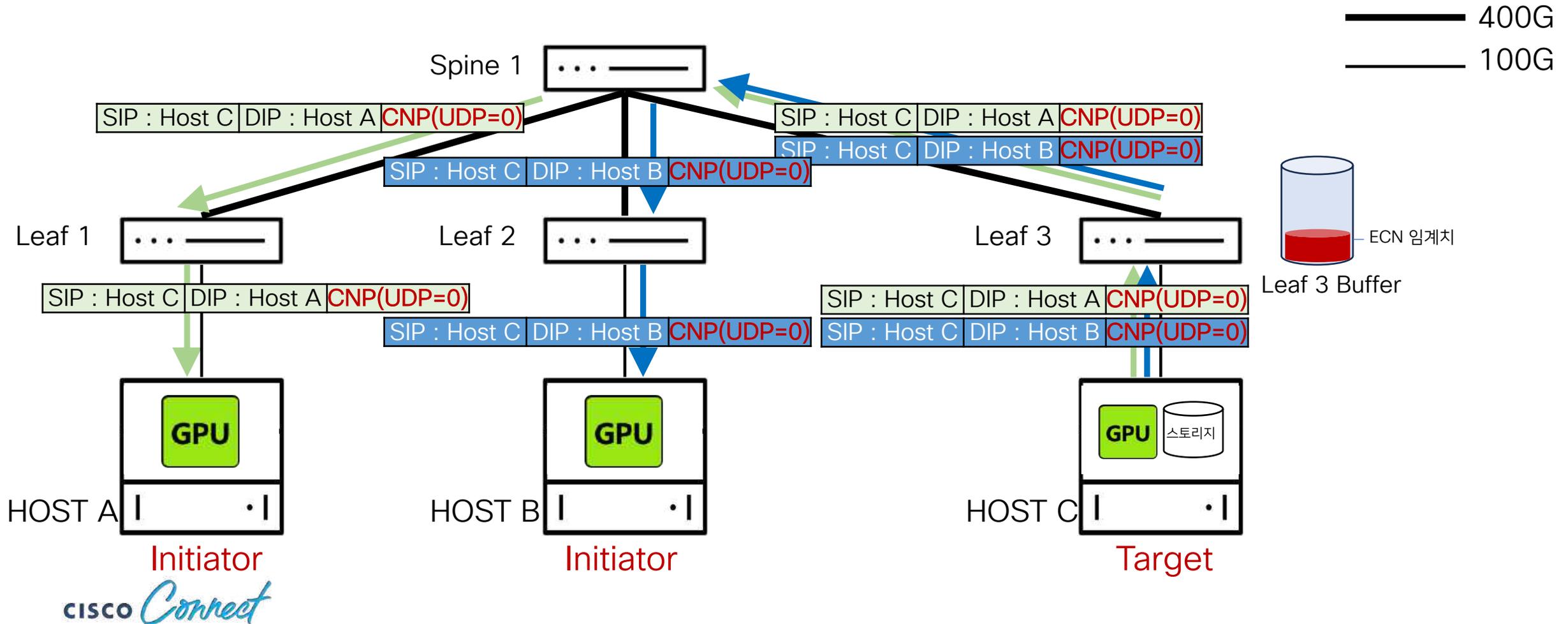
# DCQCN - RoCEv2 기반의 흐름제어 방식(Lossless)

- Leaf 3 : ECN(To Target) Marking : 오는 길에 막히는 구간이 있습니다.



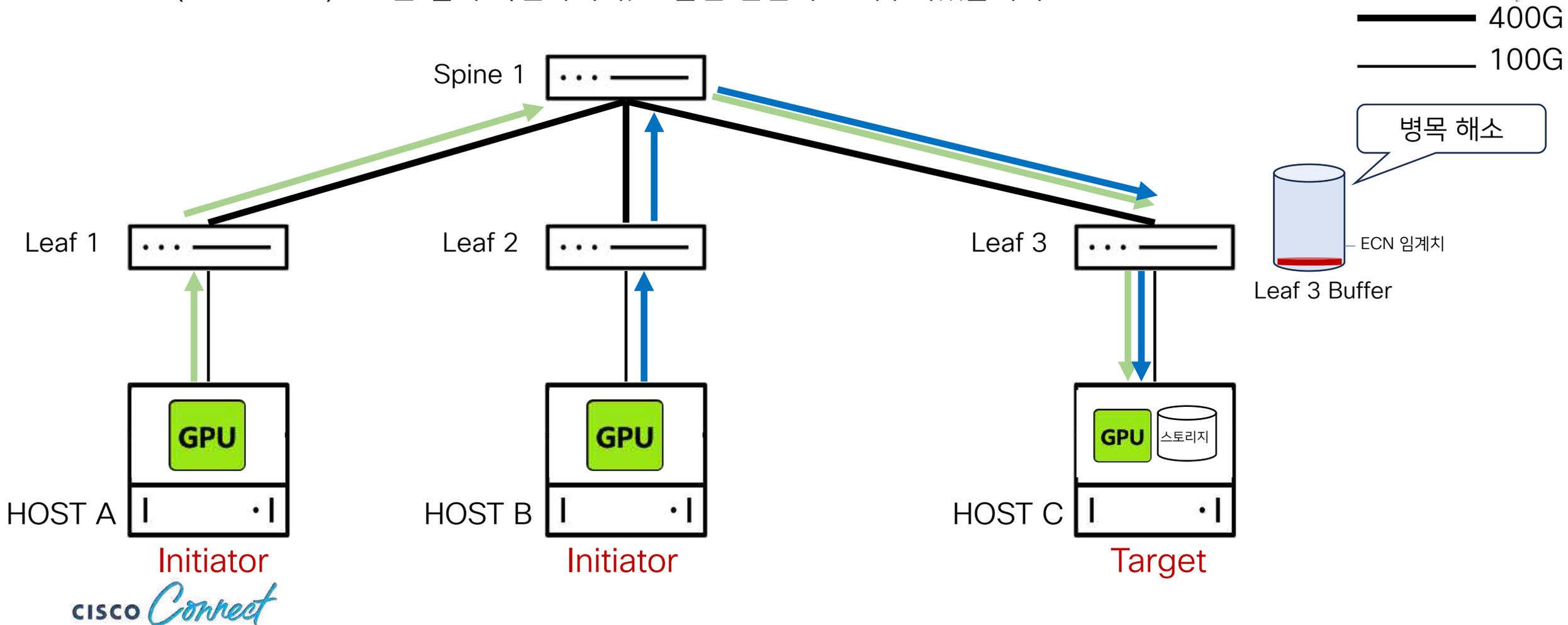
# DCQCN - RoCEv2 기반의 흐름제어 방식(Lossless)

- Leaf 3 : ECN(To Target) Marking : 오는 길에 막히는 구간이 있습니다.
- HOST C → HOST A : CNP(To Initiator) : 오는 길이 막힌다니까, 조금만 천천히 보내주시겠습니까?



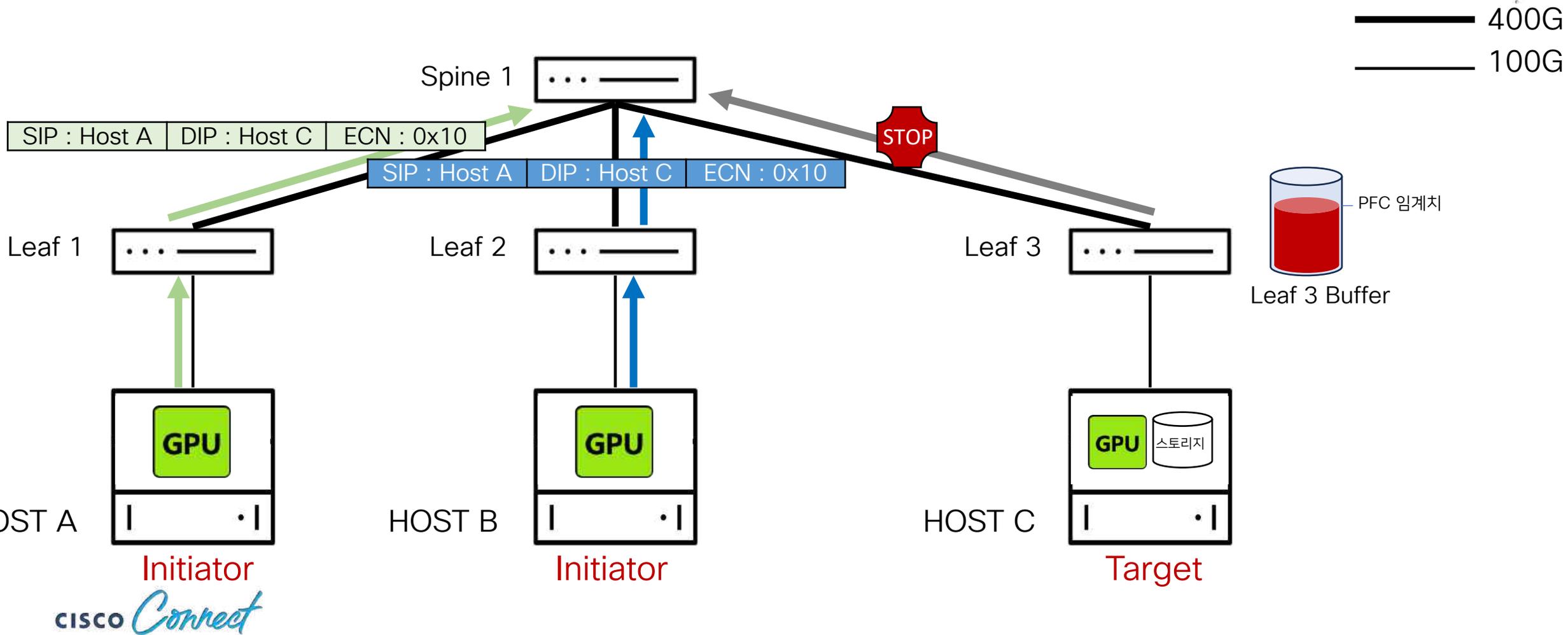
# DCQCN - RoCEv2 기반의 흐름제어 방식 (Lossless)

- ECN(To Target) : 오는 길에 막히는 구간이 있습니다.
- CNP(To Initiator) : 오는 길이 막힌다니까, 조금만 천천히 보내주시겠습니까?



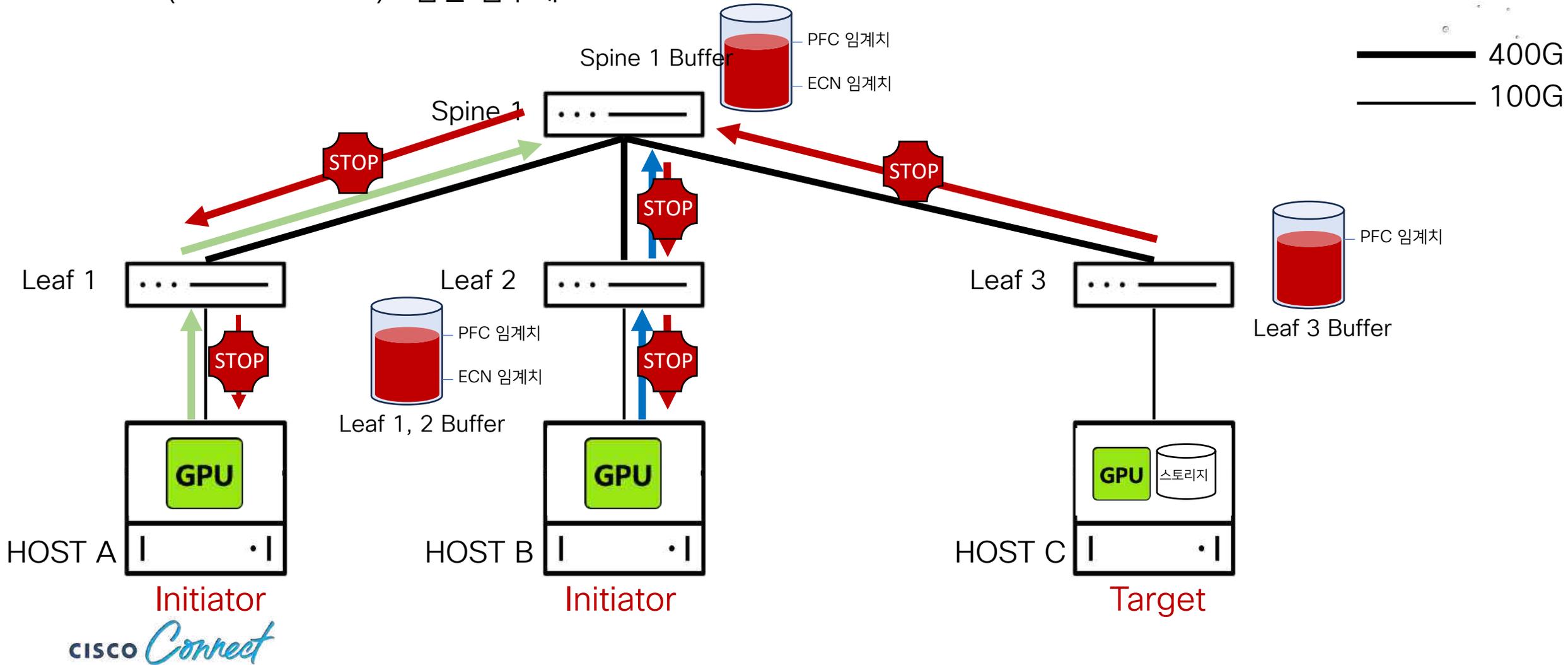
# DCQCN - RoCEv2 기반의 흐름제어 방식 (Lossless)

- PFC(Switch-Switch) : 잠깐 멈추세요



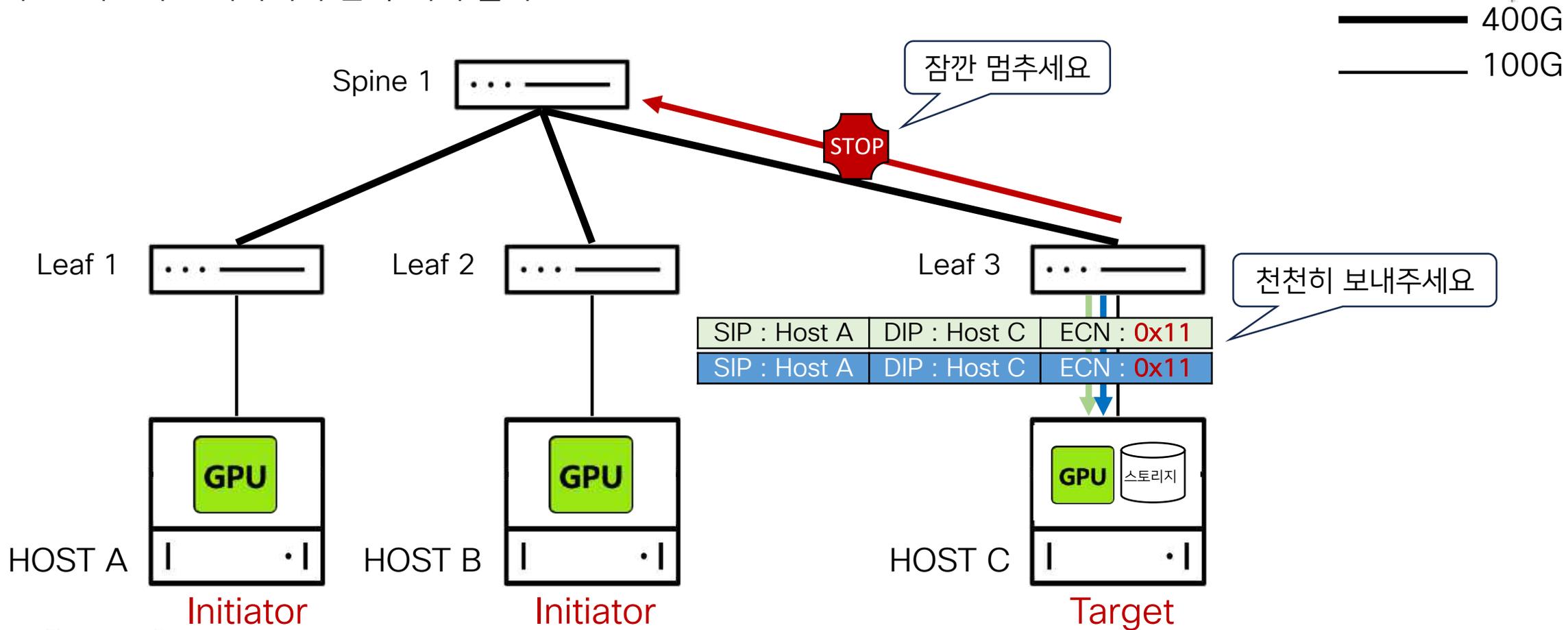
# DCQCN - RoCEv2 기반의 흐름제어 방식 (Lossless)

- PFC(Switch-Switch) : 잠깐 멈추세요



# DCQCN - RoCEv2 기반의 흐름제어 방식 (Lossless)

- AI(GPU) 서버, NVMe 스토리지와 상호 작용을 수행하는 네트워킹
- 서버 or 네트워크 계위에서 단독 제어 불가



# AI 네트워크를 위한 시스코 네트워킹 제안

# AI/ML 어플리케이션을 위한 Cisco 검증 디자인



Products and Services Solutions Support Learn

Explore Cisco

Products & Services / Cloud and Systems Management / Cisco Nexus Dashboard Fabric Controller / White Papers /

## Cisco Validated Design for Data Center Networking Blueprint for AI/ML Applications

Updated: June 6, 2023

Bias-Free Language Contact Cisco

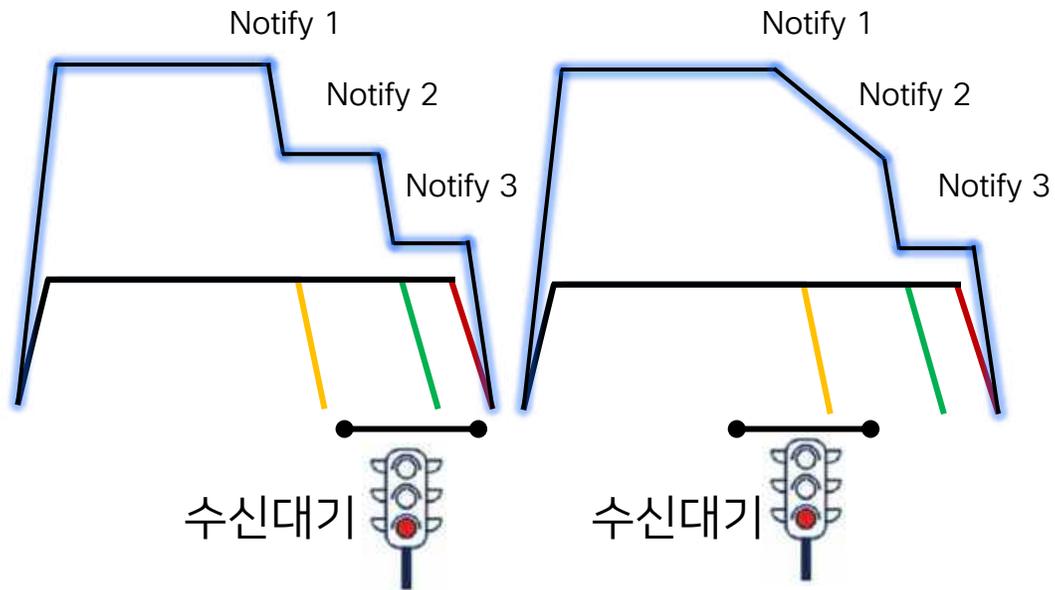
Introduction

Table of Contents

Save Download Print

# Non-Blocking Fabric 검증 디자인

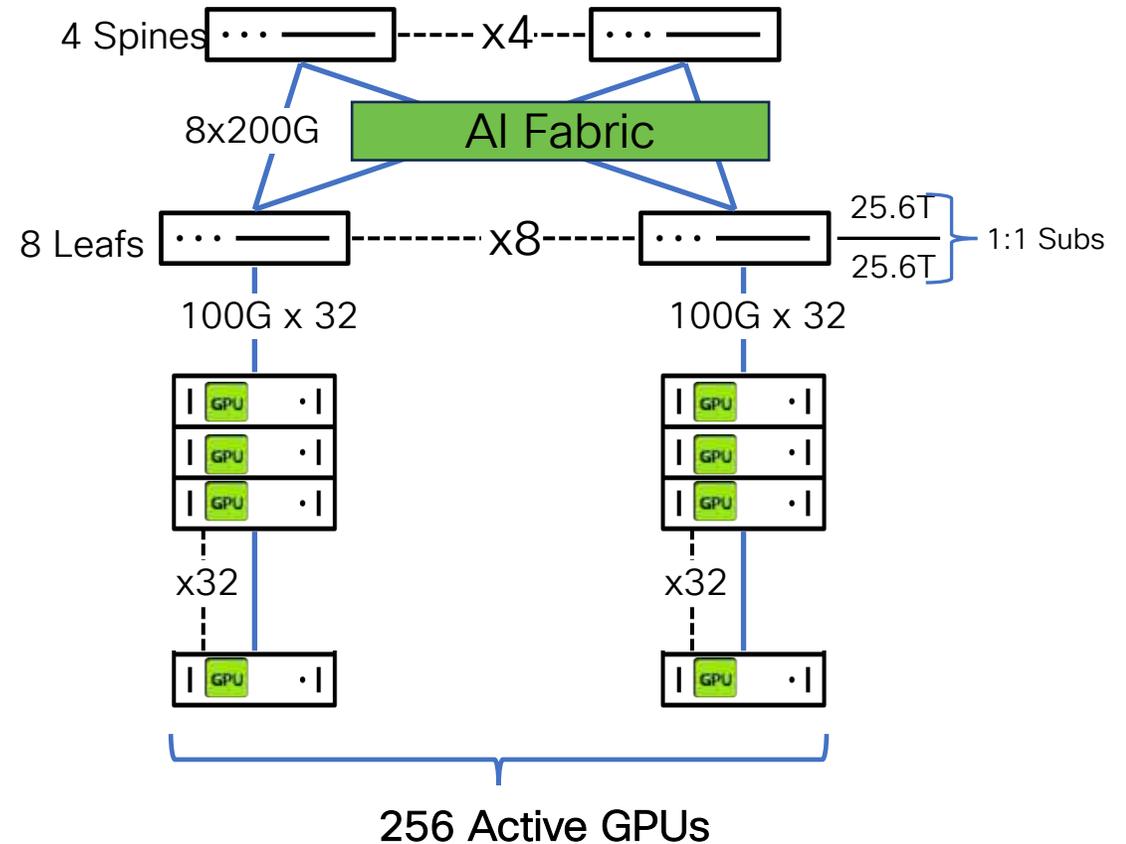
## AI Network 트래픽 패턴



Multi-Job 수행이 요구되는 AI Network는 Lossless 흐름제어 및 Non-Blocking 요구

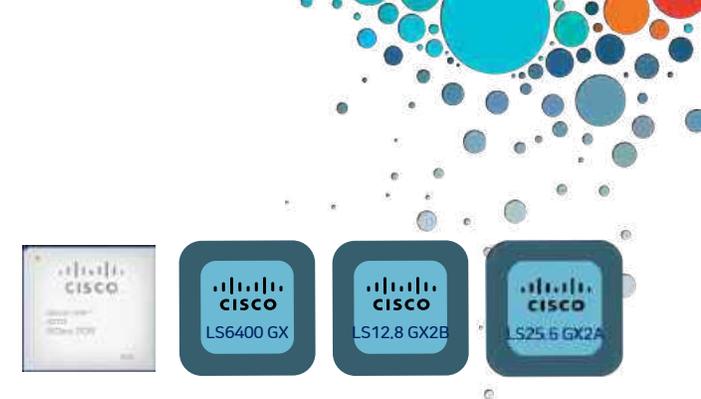
## Non-Blocking Fabric 디자인 예시(1:1 Subs)

\*RoCEv2 기반의 흐름제어

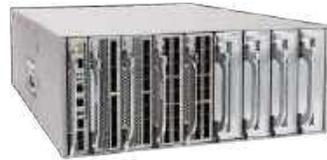


# AI Fabric 디자인을 위한 검증 제품

## Cisco Nexus 9000 Series



Nexus 9400



Nexus 9500 / Nexus 9800



C9364D-GX2A 64p 400G



C9348D-GX2A 48p 400G



C9332D-GX2B 32p 400G



C9316D-GX 16p 400G



**Performance**

**51.2Tbps / 8.35Bpps**

**38.4Tbps / 8.35Bpps**

**25.6Tbps / 4.17Bpps**

**12.8Tbps / 4.3Bpps**

C93600CD-GX  
28p 100G + 8p 400G



C9364C-GX 64p 40/100G



C9332C 32p 40/100G



C9364C 64p 40/100G



**Performance**

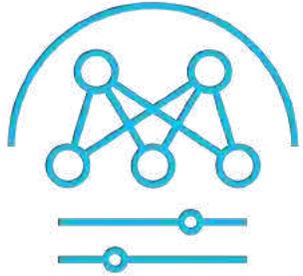
**12Tbps / 4Bpps**

**12.8Tbps / 4.3Bpps**

**6.4Tbps / 4.4Bpps**

**12.84Tbps / 4.3Bpps**

# AI Network Fabric 구축 및 관리 솔루션



**Cisco NDFC**



목적에 맞는 네트워크 패브릭 디자인

AI 네트워크를 위한 표준 설정 배포(장비 간 상호연계)

GPU Learning 중단 보호를 위한 무중단 OS Upgrade 제공

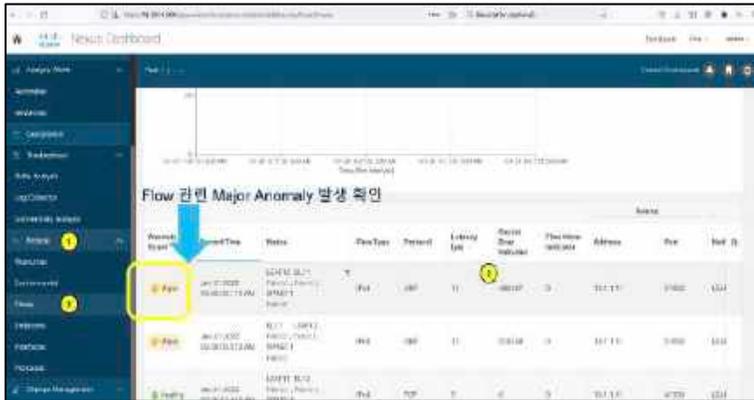
플로우 성능 모니터링을 위한 텔레메트리 서비스 연계

# AI Network 모니터링 제공

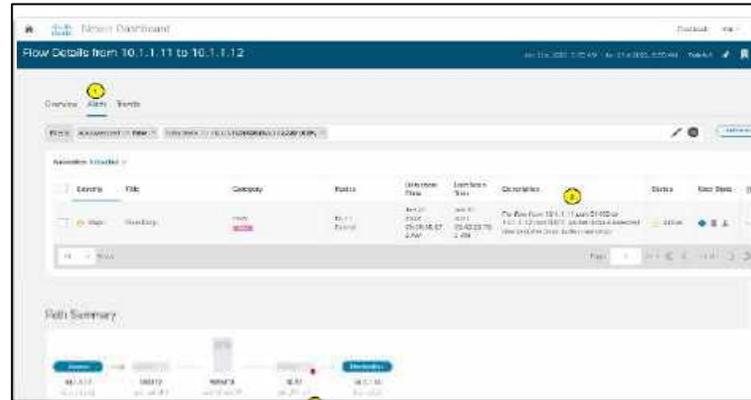


## Cisco ND-Insights

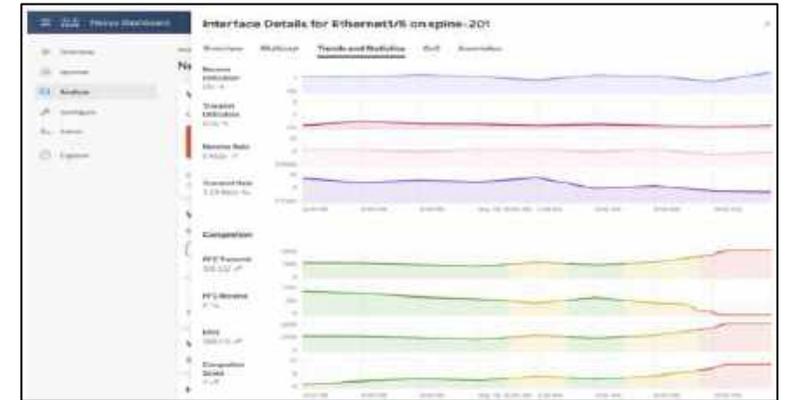
- AI Traffic 개별 트래픽 플로우에 대한 Latency, Drop
- 정확한 이슈 구간 파악
- AI 흐름제어 메트릭 발생 트렌드 감시



Flow 별 이슈 감지(Drop, Latency, Application Type)



이슈 Flow에 대한 Path finder 및 원인 파악



AI Network(RoCEv2) 발생 이슈 메트릭

AI Network에서는 지금까지 모니터링하지 않았던, 새로운 요소에 대한 모니터링이 필요합니다.

# Summary

## AI Network 디자인 고려사항

- GPU 성능 발전에 따라, AI 서버는 400G Non-Blocking으로 구성
- AI Network은 RoCEv2 기반의 Lossless Network 기술 적용
- AI Network 품질 모니터링과 ECN, PFC를 관리방안 마련



The bridge to possible

감사합니다.

CISCO *Connect*

#CiscoConnect

CISCO *Connect*

ALL IN

#CiscoConnect

CISCO *Connect*

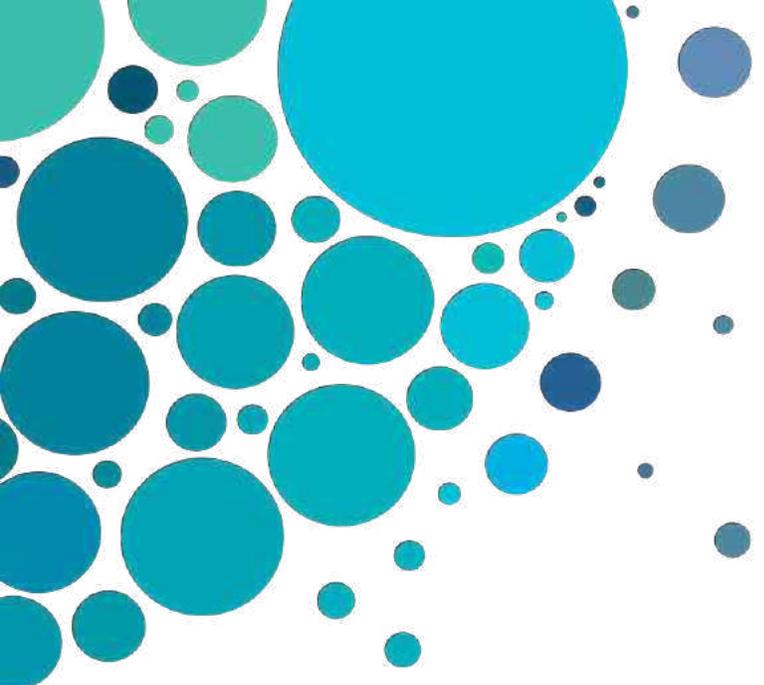
ALL IN

#CiscoConnect

# 전환기의 IT 인프라, 어떻게 운영해야 하나?

진화하는 IT 인프라 운영 고도화를 위한 시스코의 기술지원서비스

윤희진 이사 | Service Sales Specialist, Cisco Korea



# Agenda

- 기술지원서비스 세대별 방향성
- 기술지원서비스 진화의 배경
- IT인프라 운영을 위한 핵심요소
- 시스코가 제시하는 솔루션 (예시)
- 요약

# 세대별 기술지원서비스의 변화

1세대



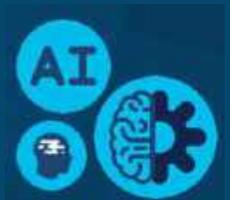
SMARTnet / SNTC

2세대



Solution Support

3세대



Success Tracks

Now!

하드웨어 장비 이슈 해결

연계된 이기종 제품 이슈 해결

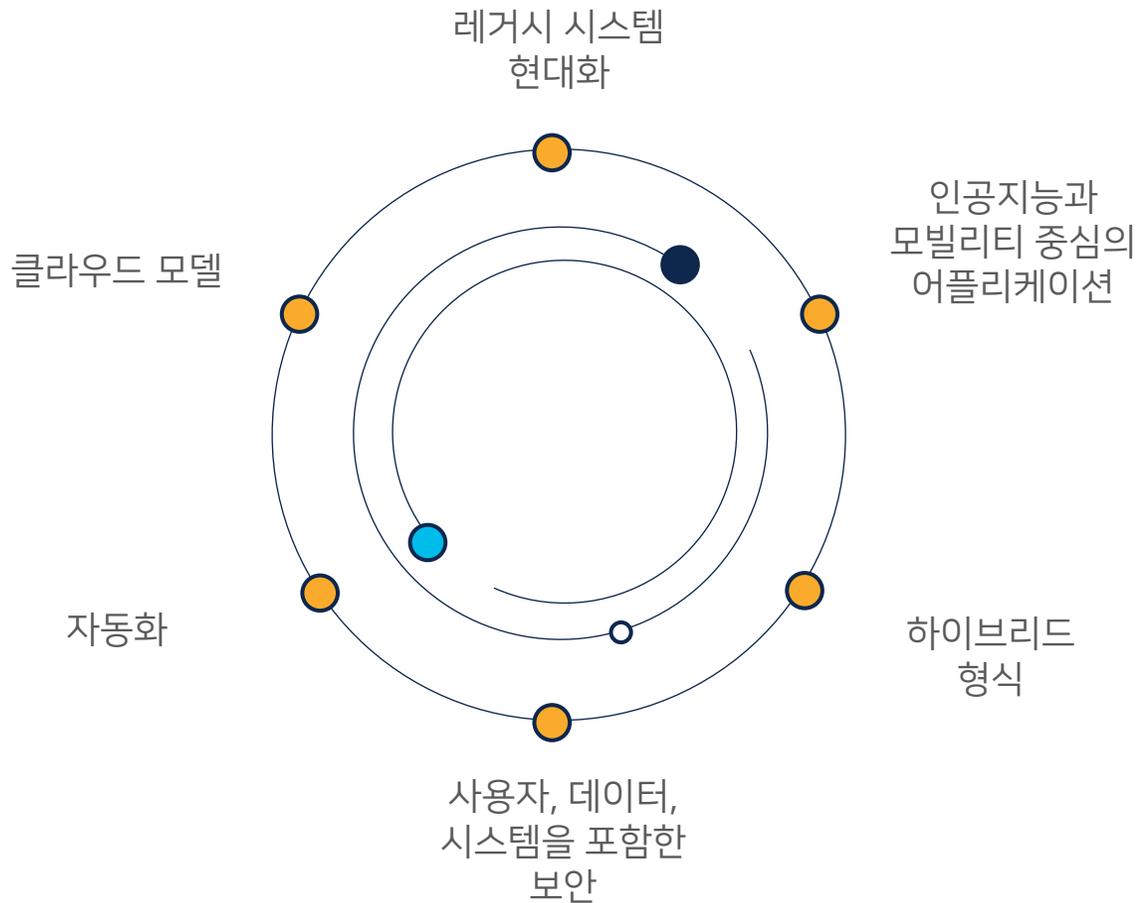
하드웨어 장비 이슈 해결

인공지능 및 기능 활용

연계된 이기종 제품 이슈 해결

하드웨어 장비 이슈 해결

# 현실적으로 '매우' 복잡한 디지털 트랜스포메이션



"2030년까지 IT 기술역량 부족에 의한,

8,500

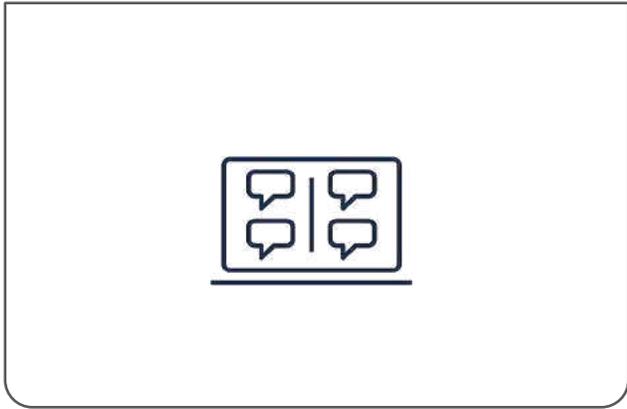
만개의 일자리 부족현상이 일어날 것이며,  
이는 연간 8조 5천억 달러의 매출손실로 이어질 것."

*Fortune April 2022*

# IT 환경은 복잡성 확대에 대한 민첩한 대응을 요구

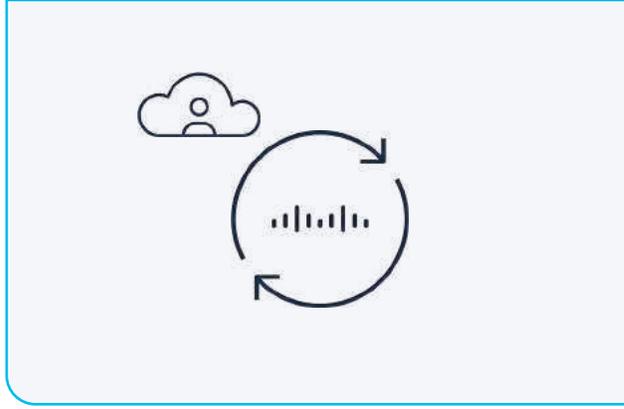
PAST

수동적 대응에서

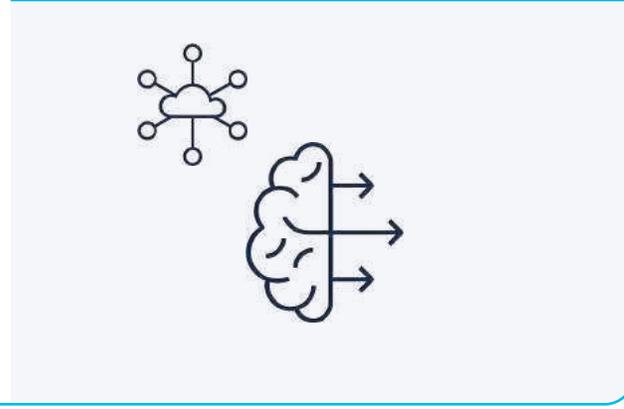


PRESENT

능동적이고



예측 가능한 대응



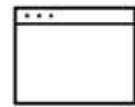
문제 발생후



인력



문제해결



도입 및 운영  
전반 활용



자동화  
+ 인사이트



학습역량

# 시스코가 제시하는 솔루션

# IT 솔루션 도입에서 활용 최적화 시점까지의 여정

## 기술지원서비스 항목

### ✓ 사후 기술지원



- 시스코와 연계된 타 벤더사 제품을 포함한 기술지원

### + 기술 전문가 가이드



- 1:N 고객 웨비나
- 1:1 맞춤형 전문가 코칭
- 전세계 고객사 모임 커뮤니티

### + 인사이트 & 분석



- 자산관리 자동화
- 고객 운영환경 안정성 관리
- AI/ML 운영환경 최적화

### + 맞춤형 학습



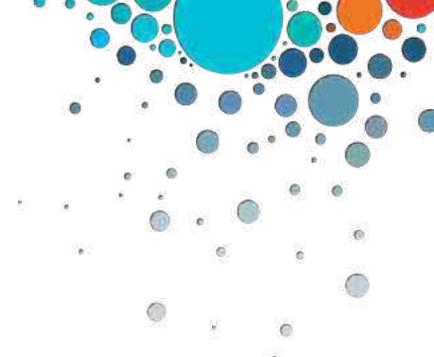
- 관련된 모든 이러닝 콘텐츠
- 리모트 랩 지원

## 통합된 하나의 포털에서 관리 및 사용 (CX Cloud)

### 기대효과

- **더욱 빨라진 사후 기술지원 서비스** - 간소화된 절차에 의한 빠른 사후 기술지원 서비스
- **실시간 자산관리** - 자동화를 통해 보유하고 있는 네트워크 인프라 자산에 대한 완벽한 관리
- **장애 사전 예방** - 운영환경에 대한 선제적 보안 취약성 조치 및 잠재적 위험에 대한 사전 예방
- **운영능력 고도화** - 전문가 가이드 및 맞춤형 학습 콘텐츠를 통한 운영능력 고도화

# 고객사 IT 운영환경에 대한 서비스 적용 예시

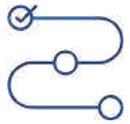


시나리오	
네트워크 인프라 자산에 대한 인사이트	운영중인 장비중 도래하는 LDoS* 정보를 한번에 파악하여 마이그레이션 플랜 제공 <small>* Last Day of Support</small>
네트워크 인프라 자산에 대한 보안 조치	운영 서비스중 특정 "A" 기능을 쓰는 경우 심각한 보안 취약점에 노출, 즉시 해당 자산을 파악하고, 잠재적인 위협에 노출된 자산까지 한번에 파악
잠재적인 네트워크 단절 위험을 사전 예방	운영 서비스중 Crash가 발생할 수 있는 장비들을 사전에 파악
환경에 적합한 OS 선정	서비스중인 운영 환경에 맞는 안정적인 OS의 선정

# 요약



# IT 인프라 운영 고도화의 필요성과 방향성



소프트웨어 기반 산업화에 의해 복잡해진 IT 운영환경 고도화의 필요성



기존의 수동적 대응에서 능동적이고, 예측가능한 대응을 할 수 있는 역량



시스코의 방향성은 IT 관리 간소화, 최적화, 그리고 미래 혁신 역량에 집중



The bridge to possible

감사합니다.

CISCO *Connect*

ALL IN

#CiscoConnect

# Cisco ACI 기반 AADC 사례와 통합관리 방안

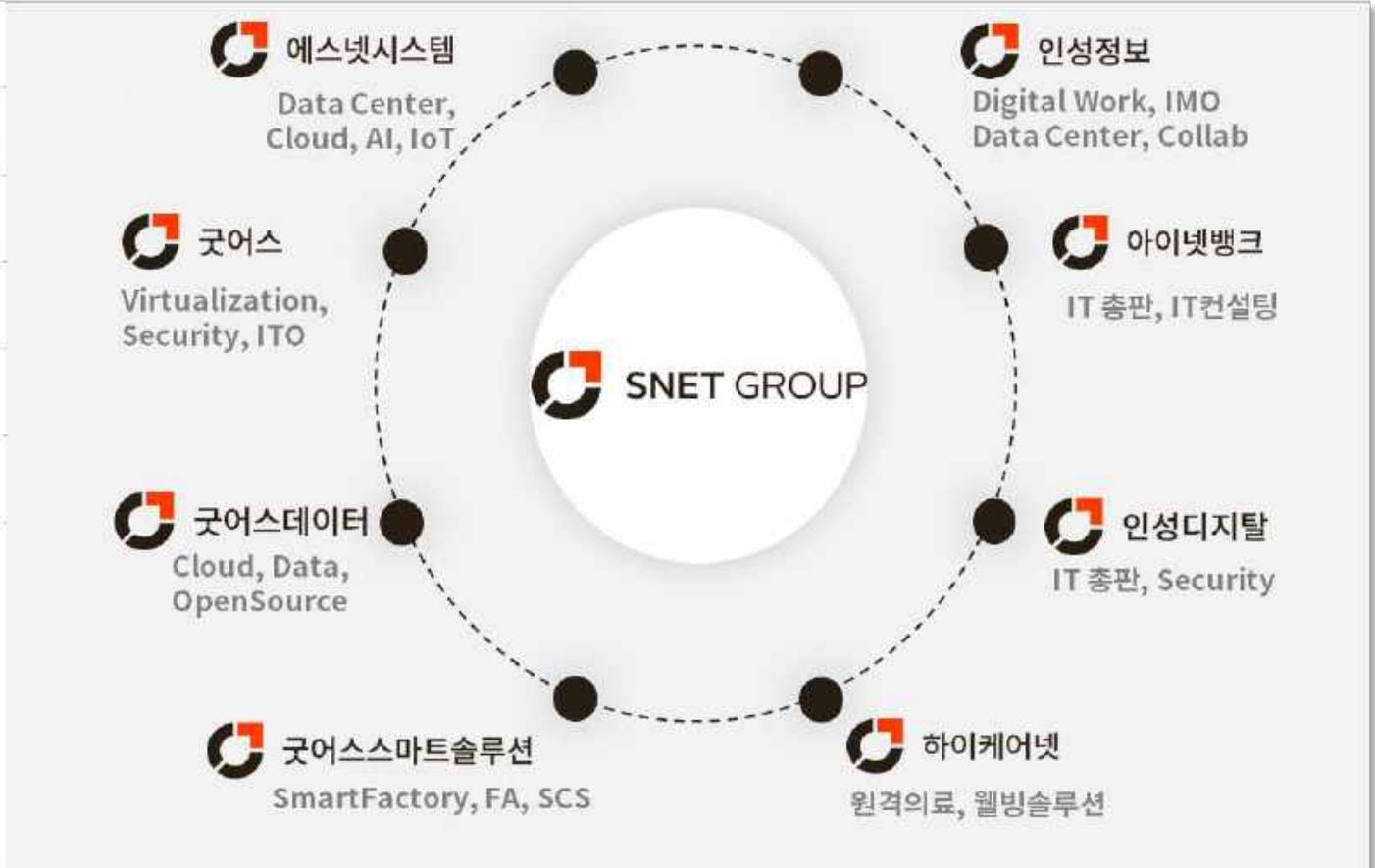
(최근 데이터센터 동향과 Self Service Portal)

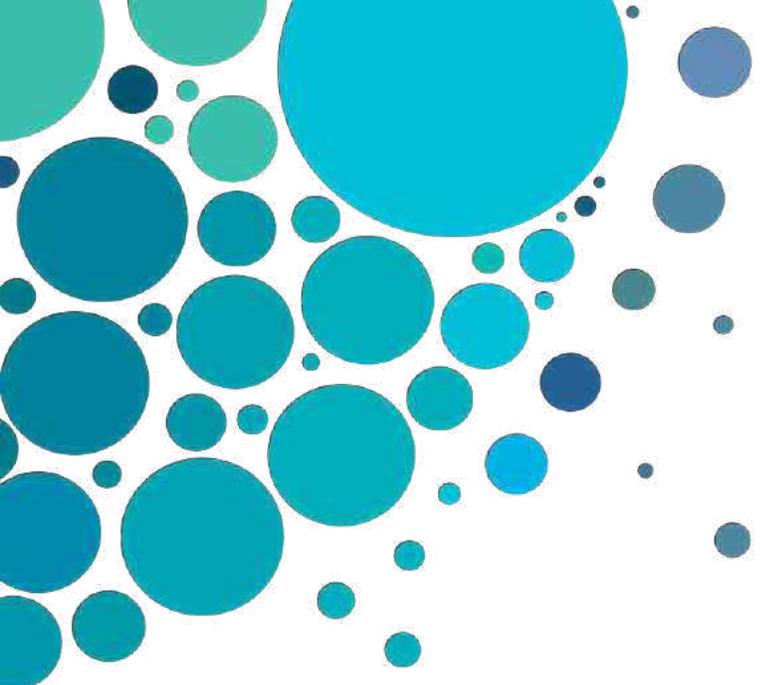
하진철 상무 | 에스넷시스템(주)

# Snet Group 소개

20년 이상 국내 정보통신 분야를 선도해 온 에스넷그룹은 2020년 10월 인성정보와의 결합을 통해 에스넷시스템, 굿어스, 굿어스데이터, 굿어스스마트솔루션, 인성정보, 인성디지털, 아이넷뱅크, 하이케어넷을 계열사로 구성하여 **ICT 전문 역량을 융합하고 이를 바탕으로 각 분야별 사업을 영위**하고 있습니다.

회사명	에스넷그룹
설립일	1999년 2월 8일
상장일	2000년 4월 27일
회장	박호대
임직원 수	1,269명
매출액	7,040억 (2022년 연결기준)
주요 비즈니스	ICT 컨설팅/설계/구축 서비스, 운영/관제 서비스, IT 라이프사이클 서비스를 제공 영역: Cloud/DC, IT Infra, Smart Factory, AI/BigData/IoT, Smart Workplace, Healthcare, Distribution





# Agenda

- SDN기술에 대한 이해
- 데이터센터 Trend
- 인프라 고도화 단계
- 통합관리 및 AADC Demo
- Summary

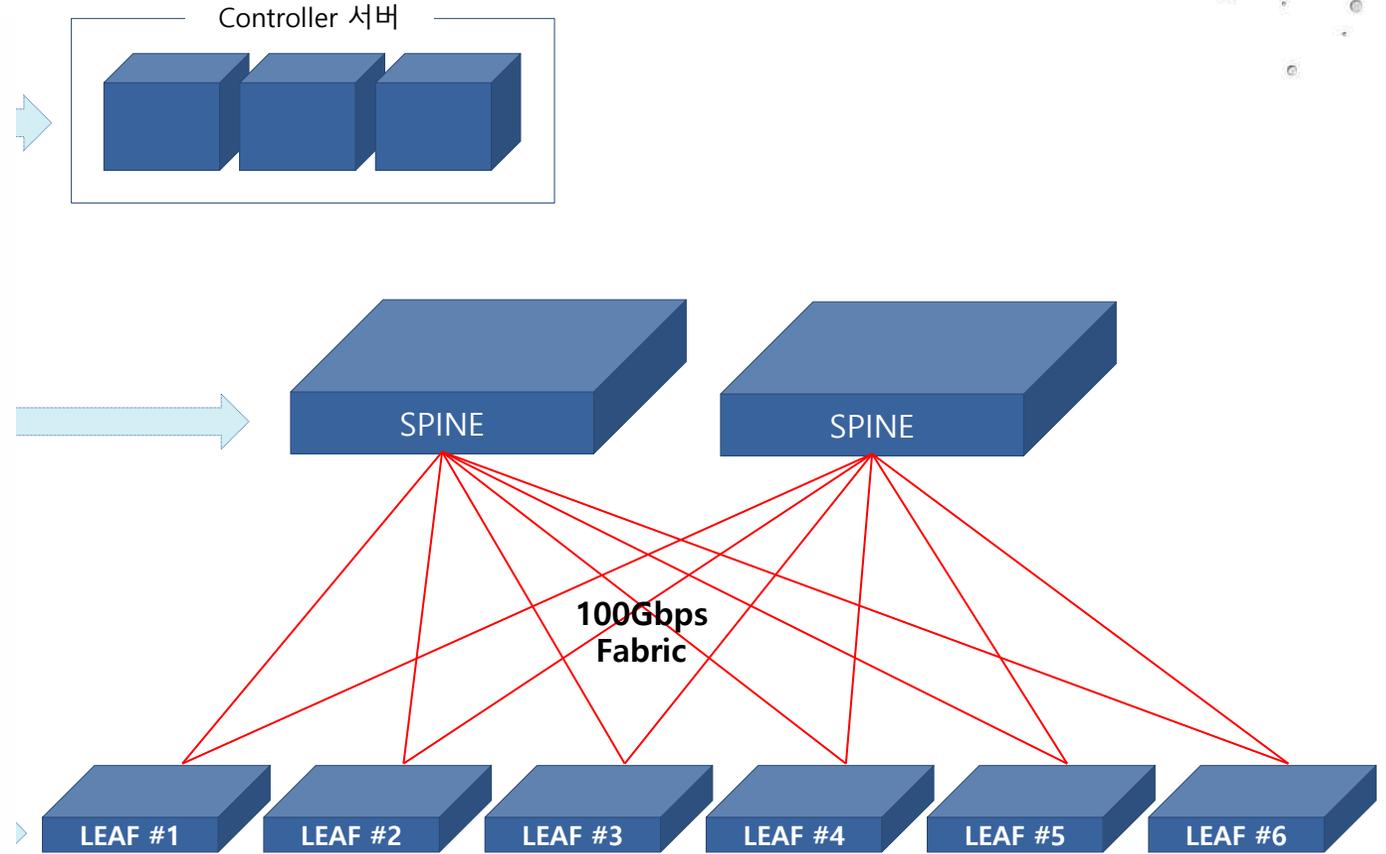
# SDN 기술에 대한 이해

SDN이란 무엇이며 어떻게 동작하는가

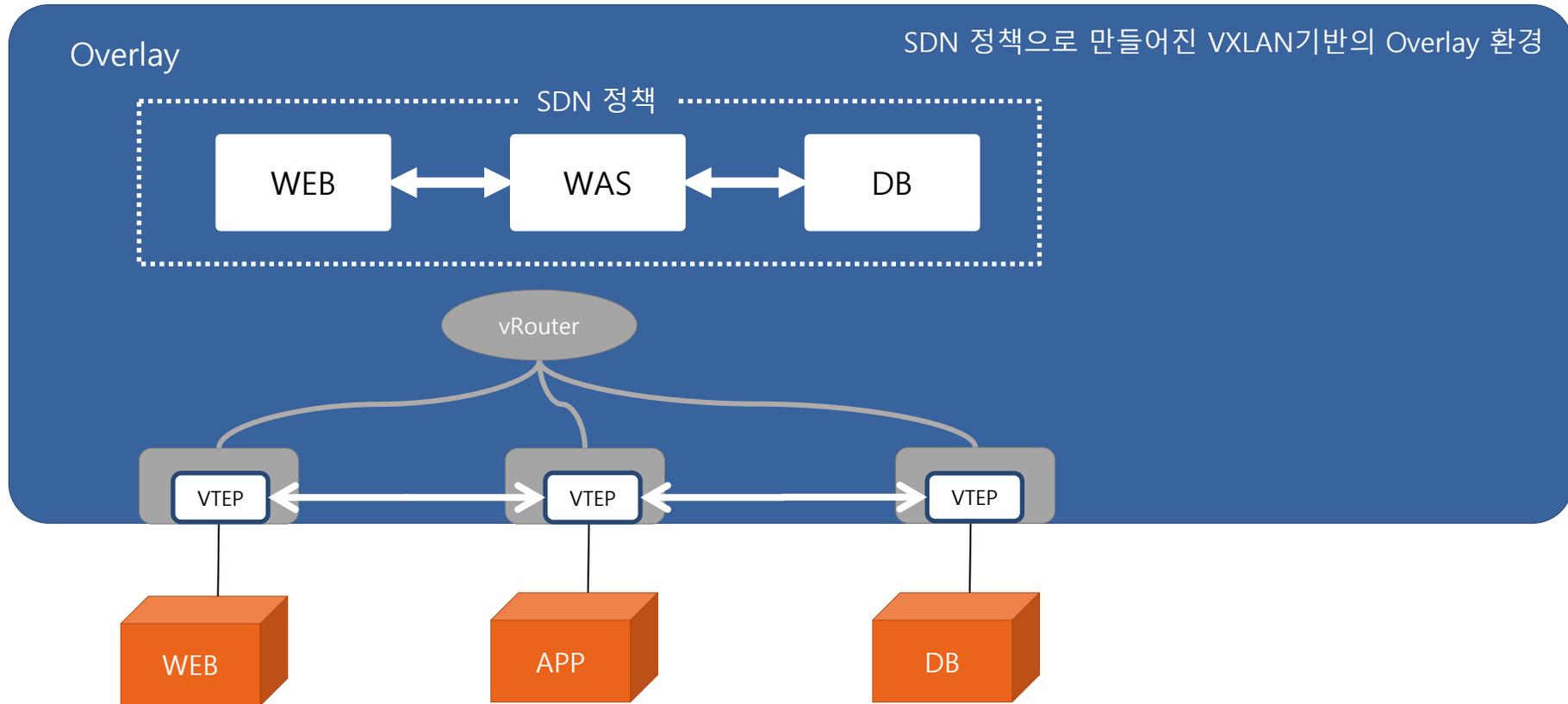
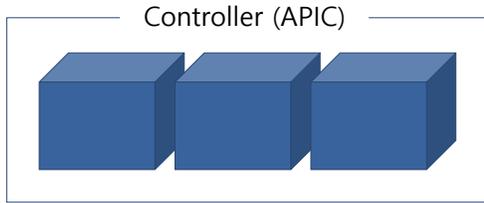


# 데이터센터에서 사용되는 SDN Fabric - Underlay

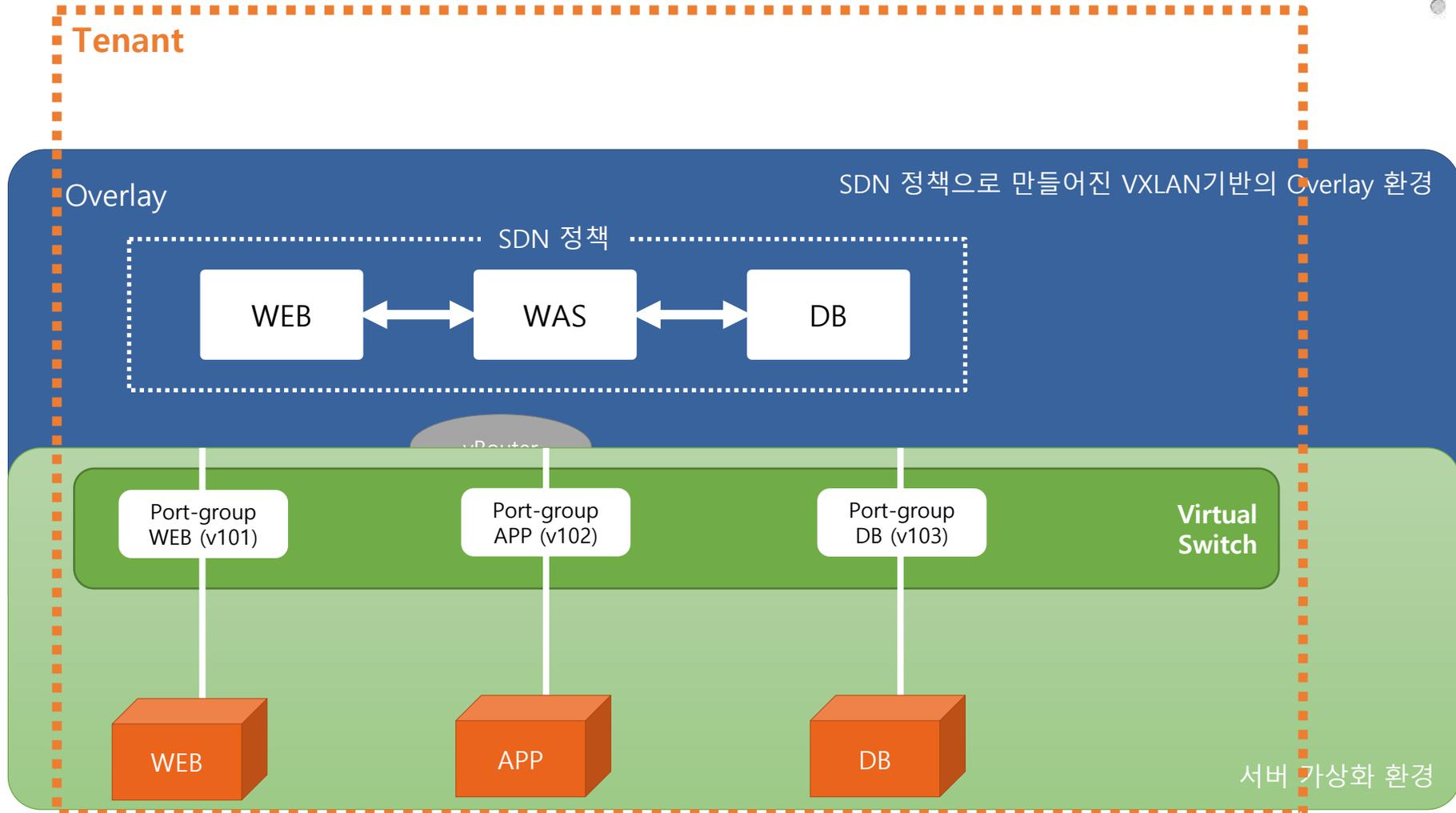
Software Define 기술은 용도에 따라 다양하게 사용되지만, 특히 데이터센터에서는 SDN Fabric 개념이 사용됩니다. Underlay는 모두 자동화 되며, 케이블 연결만으로 자동 인식됩니다.



# 데이터센터에서 사용되는 SDN Fabric - Overlay



# 데이터센터에서 사용되는 SDN Fabric - Overlay



# 참고 - 서버 가상화 구성

서버 가상화 환경에서 구성 원리

VM	
CPU	4 Core
MEM	16 GB
DISK	200 GB
NIC	Port-Group

VM	
CPU	8 Core
MEM	32 GB
DISK	400 GB
NIC	Port-Group

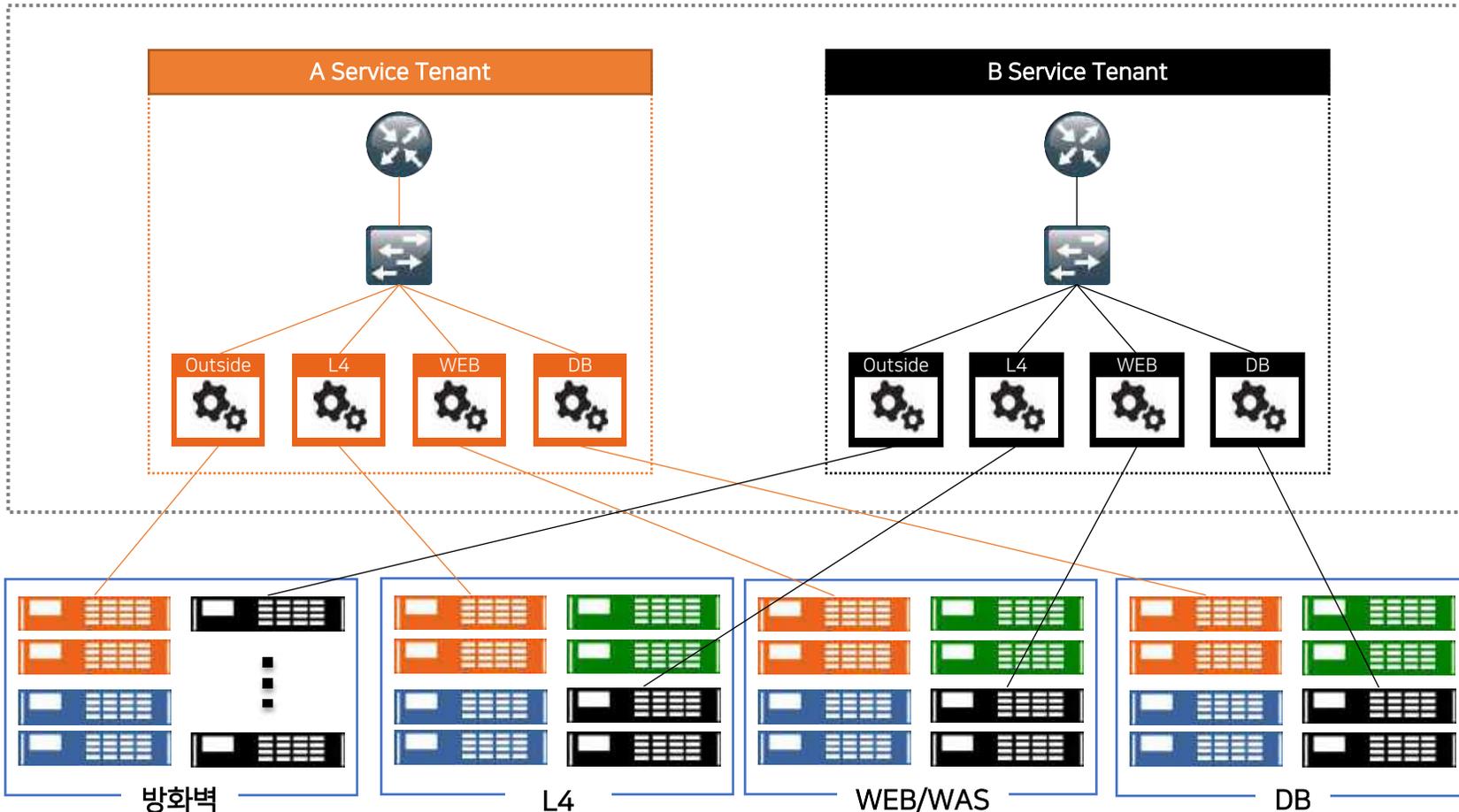
전체 32 Core  
사용 4Core + 8Core  
잔여 20Core

CPU 32 Core	Memroy 256 GB	DISK 1 TB	NIC
----------------	------------------	--------------	-----

# 데이터센터에서 사용되는 SDN Fabric - Overlay

SDN Overlay 환경은 마치 서버 가상화 환경에서 VM이 서버의 Resource를 공유하여 사용하는 것과 유사합니다.

Network Resource인 방화벽과 L4, Web/WAS 및 DB를 마치 VM의 CPU와 Memory, DISK를 가져다 사용하는 것처럼 활용합니다.



손쉬운 NW 생성/삭제

가상 Network인 Tenant는 마치 VM을 생성하는 것처럼 손쉽게 생성하고, 삭제 가능

장애 시 빠른 복구

VM처럼 Snapshot 기능을 제공하여 작업 전 Snapshot을 구성하고, 즉시 복구 가능

VM 및 Container와의 연동

가상머신을 관리하는 Platform 또는 컨테이너를 관리하는 Kubernetes와 SDN이 연동되어 각 환경에서 설정하는 Workload가 Network와 손쉽게 연동

GUI 기반의 통합관리

모든 설정을 GUI기반으로 설정하여 손쉽게 관리

Day2 Operation (운영) 관리 기능 제공

가상화 된 환경을 손쉽게 모니터링 하고, 트래픽에 발생하는 이슈를 실시간으로 확인하며, AI/ML 기능을 이용하여 운영/관리에 편의성 제공

# 데이터센터 Trend

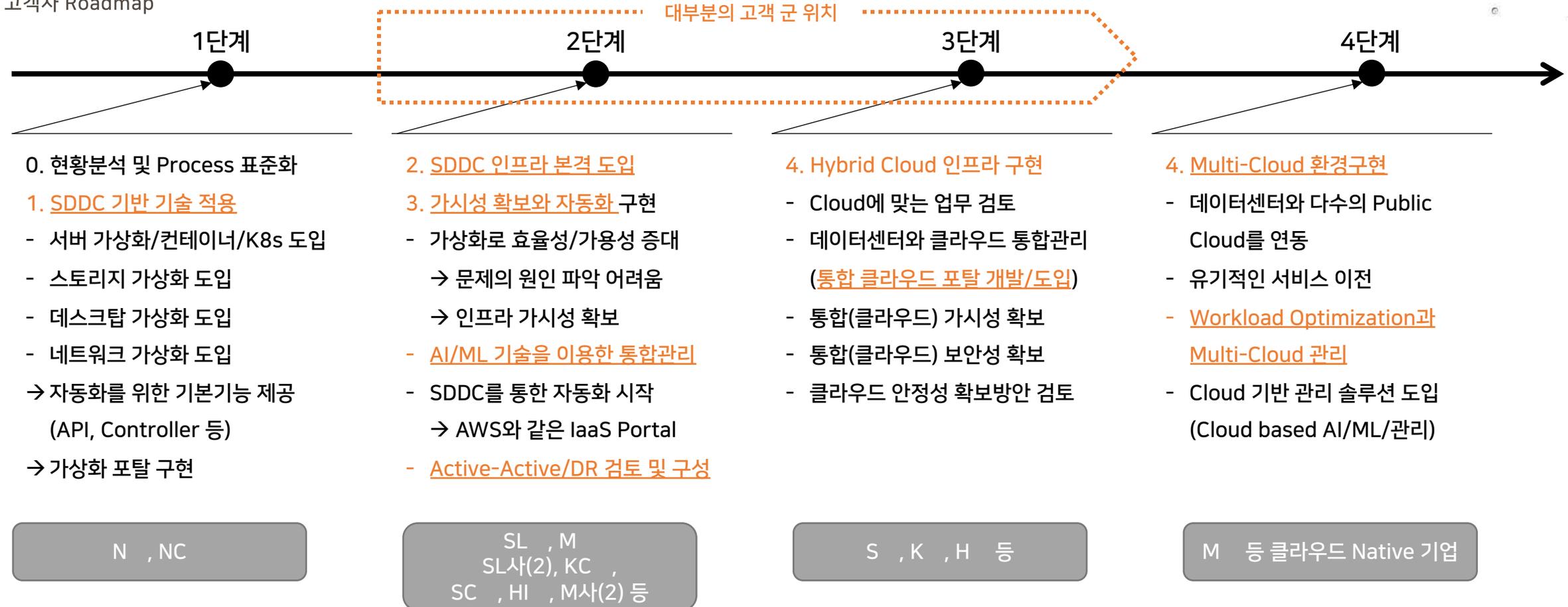
실제 고객 도입 사례를 통해 본 동향



# 데이터센터 시장 동향

인프라 구축 및 관리 기술의 변화에 따라, IT 자원의 효율적인 이용 뿐 아니라 AI/ML 기술을 이용한 현황 분석 및 예측과 자동화된 인프라 관리 방식을 도입  
고객의 Private 데이터센터 뿐 아니라 Public Cloud 사업자와의 연동으로 더 효율적인 곳에 인프라를 배치하는 단계로 진화 중

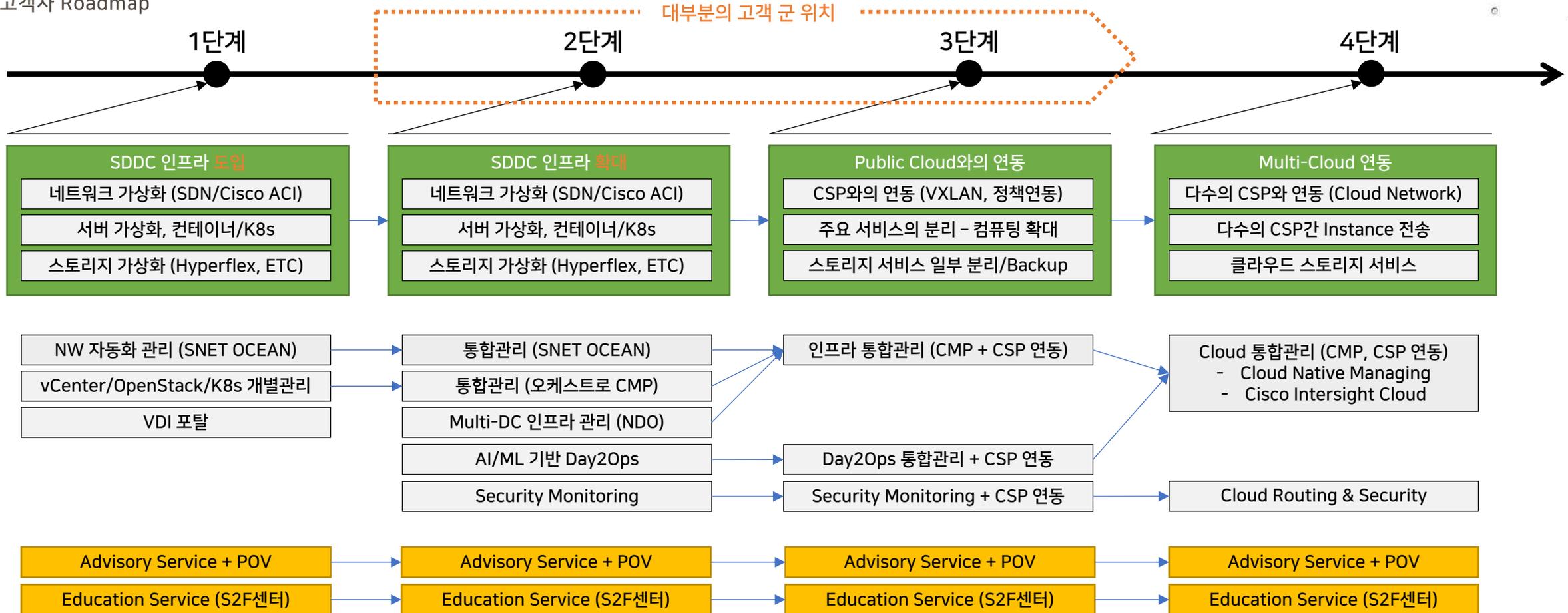
고객사 Roadmap



# 데이터센터 기술 동향

SDDC 및 Hybrid Cloud 환경과 더 나아가서 Multi-Cloud 환경으로 인프라를 고도화 하기 위해 주요 검토 내용 (일반적으로 3년~5년 PLAN) 인프라의 효율적 운영을 위해서는 컴퓨팅 자원 관리 뿐 아니라 모든 것을 연결하는 네트워크와 클라우드까지 관리영역이 확장되어야 합니다.

고객사 Roadmap



# 인프라 고도화 단계

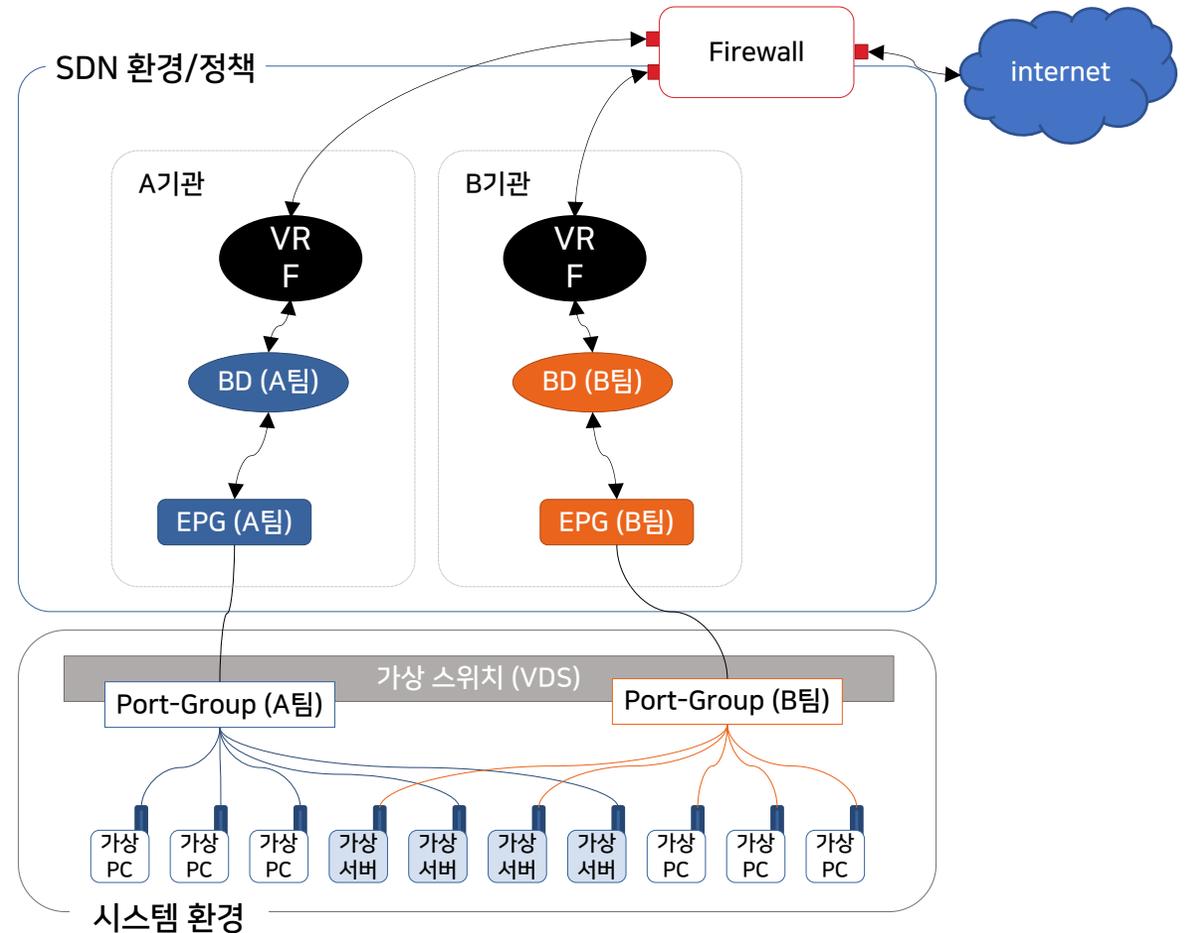
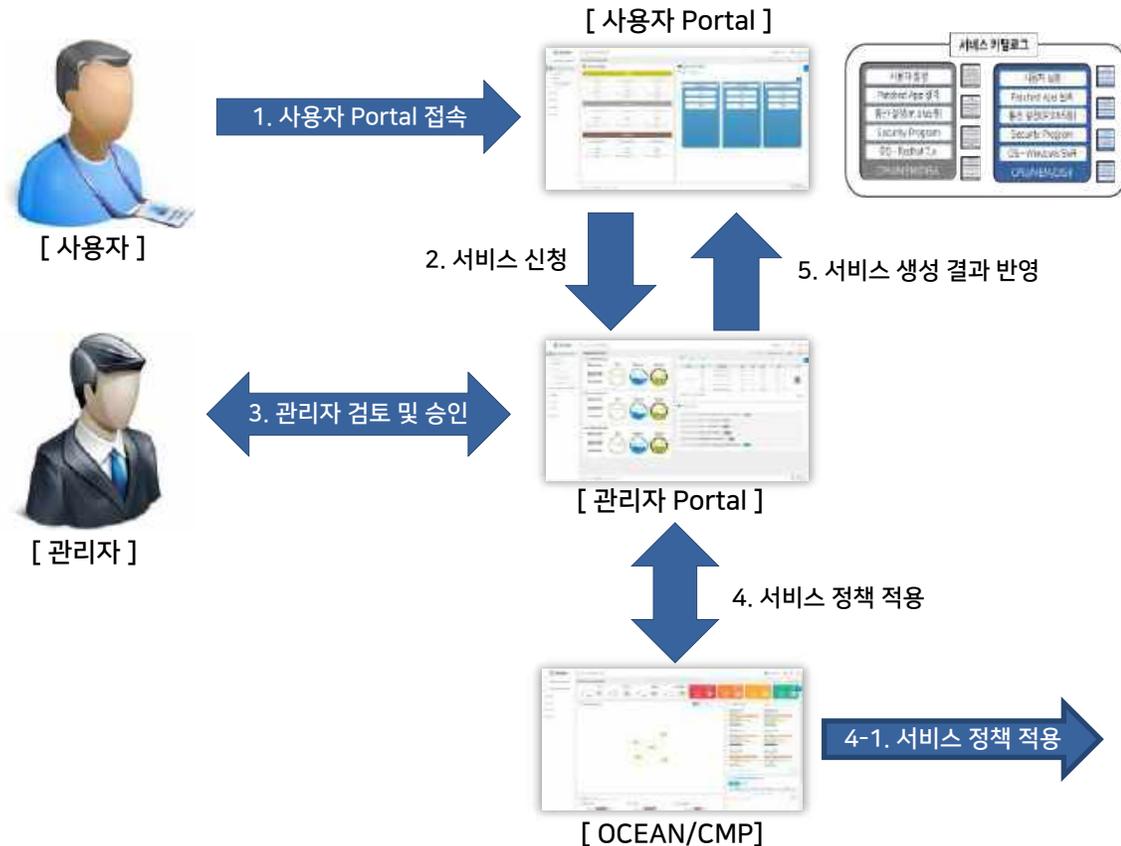
구축 단계별 고려사항



# 1단계 - SDDC 시범 도입 (기반 기술 검증)

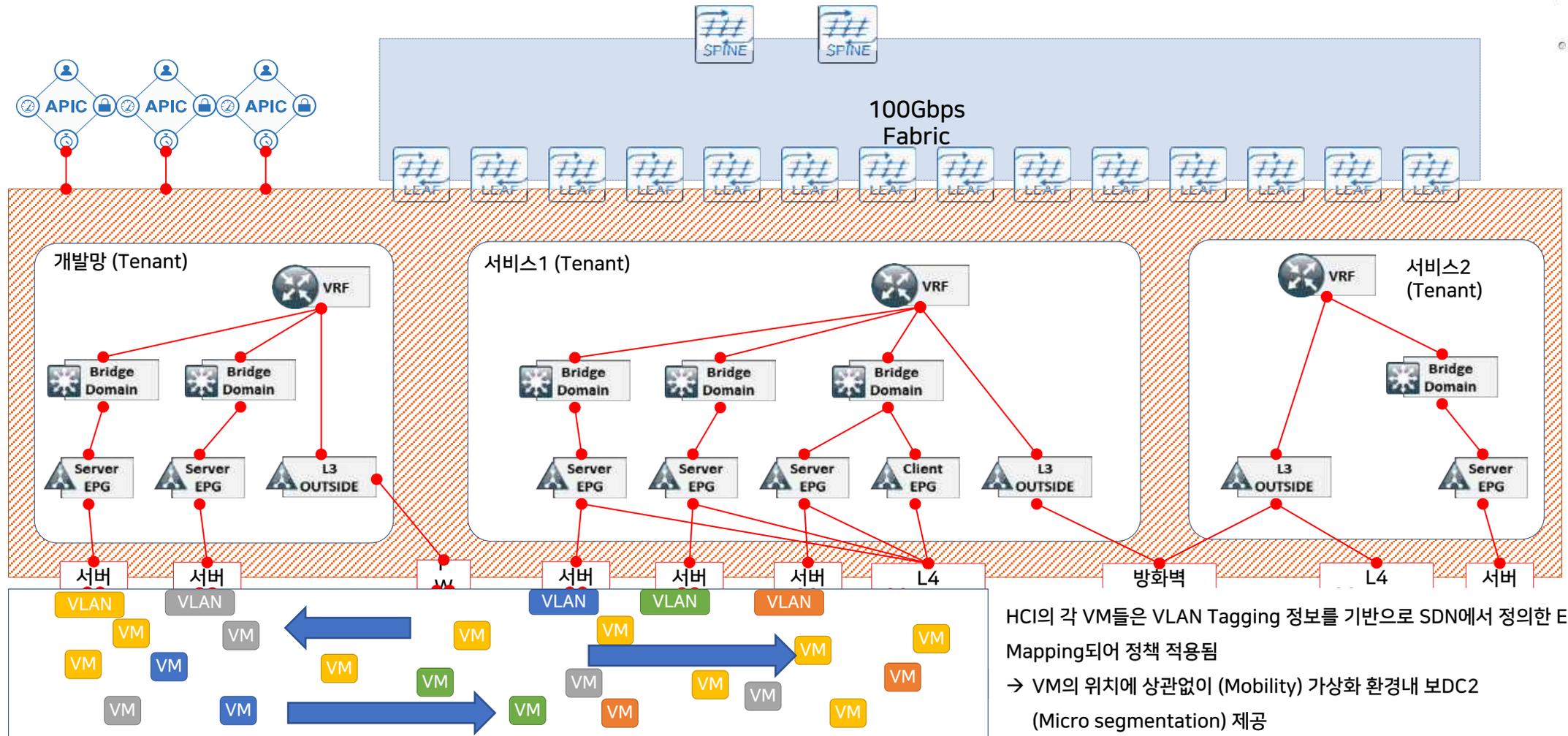
SDDC를 위한 기반 기술인 인프라 가상화 기술을 활용하여 Scale-out 기반의 최적의 인프라 구축 환경 구현이 필요합니다.  
SDDC Portal을 통해 가상 PC 또는 서버 생성시 자동으로 SDN과 연동되어 인프라 자동 확장 및 보안성을 제공합니다.

SDDC 포탈을 이용한 관리기능 구현



# 1단계 - SDDC 시범 도입 사례

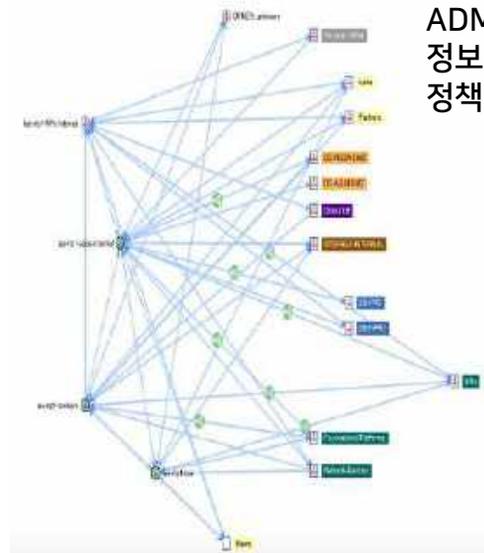
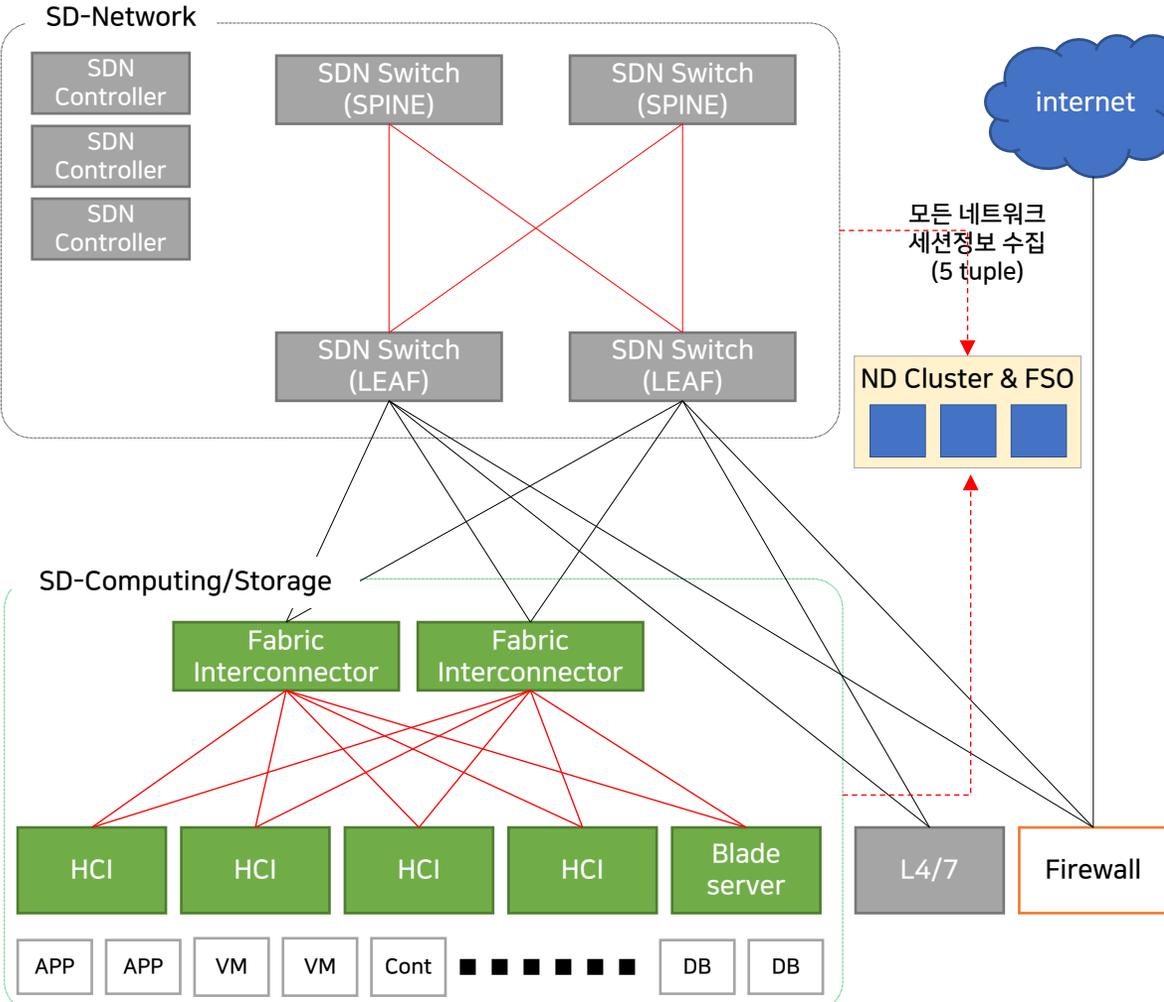
가상 인프라 환경을 통해 물리적인 인프라는 공유하지만 논리적으로 완벽하게 분리된 Software Define 환경으로 운영됩니다.  
서비스의 요구에 맞게 인프라를 손쉽게 확대하고 재구성 할 수 있습니다.



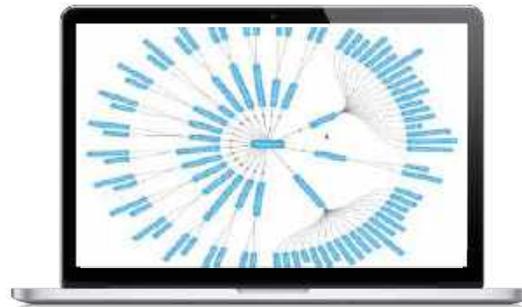
HCI의 각 VM들은 VLAN Tagging 정보를 기반으로 SDN에서 정의한 EPG에 Mapping되어 정책 적용됨  
→ VM의 위치에 상관없이 (Mobility) 가상화 환경내 보DC2 (Micro segmentation) 제공

# 2단계 - SDDC 본격 확장 및 가시성 확보

가상화 환경이 구성되면서 고도화 될수록 인프라의 가시성 확보에 어려움을 겪게 되며, 이를 해결하기 위한 전용 솔루션 확보가 요구가 발생하게 됩니다. Day2 Operation (운영환경)에 있어서, 가시성은 서비스의 효율적인 운영 뿐 아니라 안정성 향상을 위해서도 중요합니다.



ADM(Application Dependency Mapping) 정보를 기반으로 상관관계를 분석하고 이를 통해 정책의 설정 및 위반 여부 확인/통보

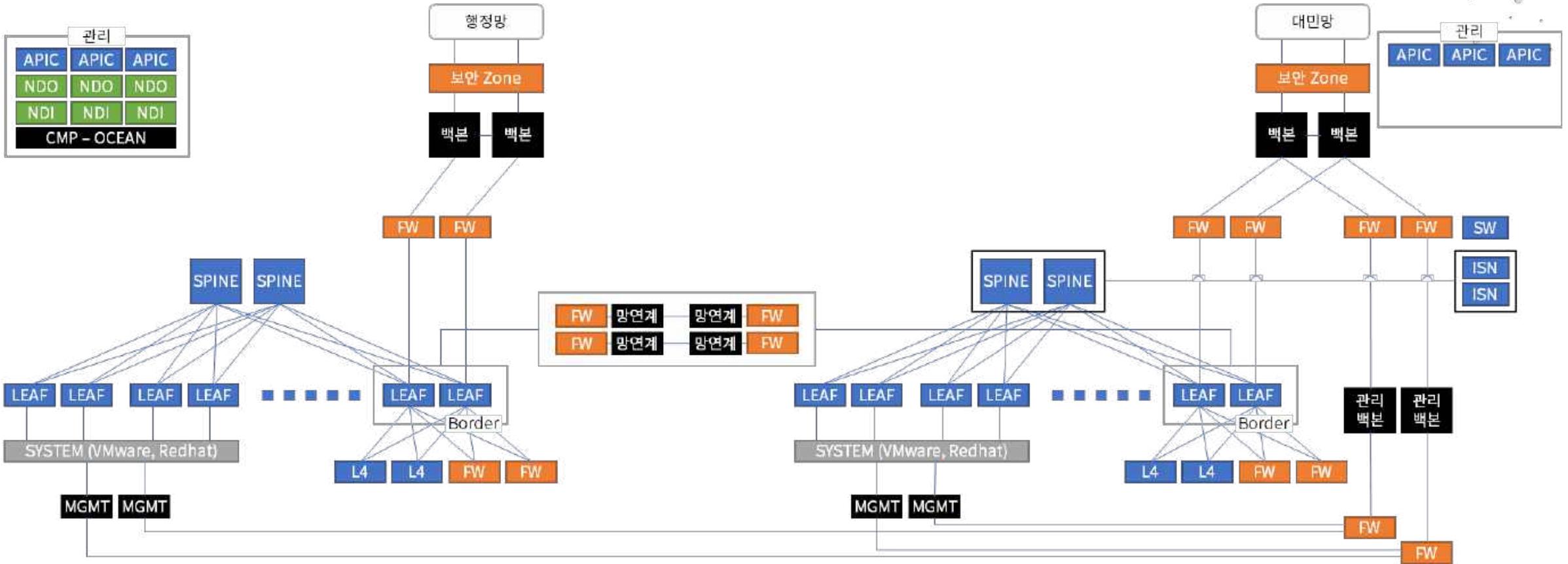


상호 연결된 관계 뿐 아니라 해당 Application 간 통신 상세 정보 확인이 가능하며 특히 과거 특정시간에 발생한 서비스 이슈의 원인을 분석할 수 있음

# 2단계 - DR 또는 A-A 데이터센터 (공공 데이터센터 사례)

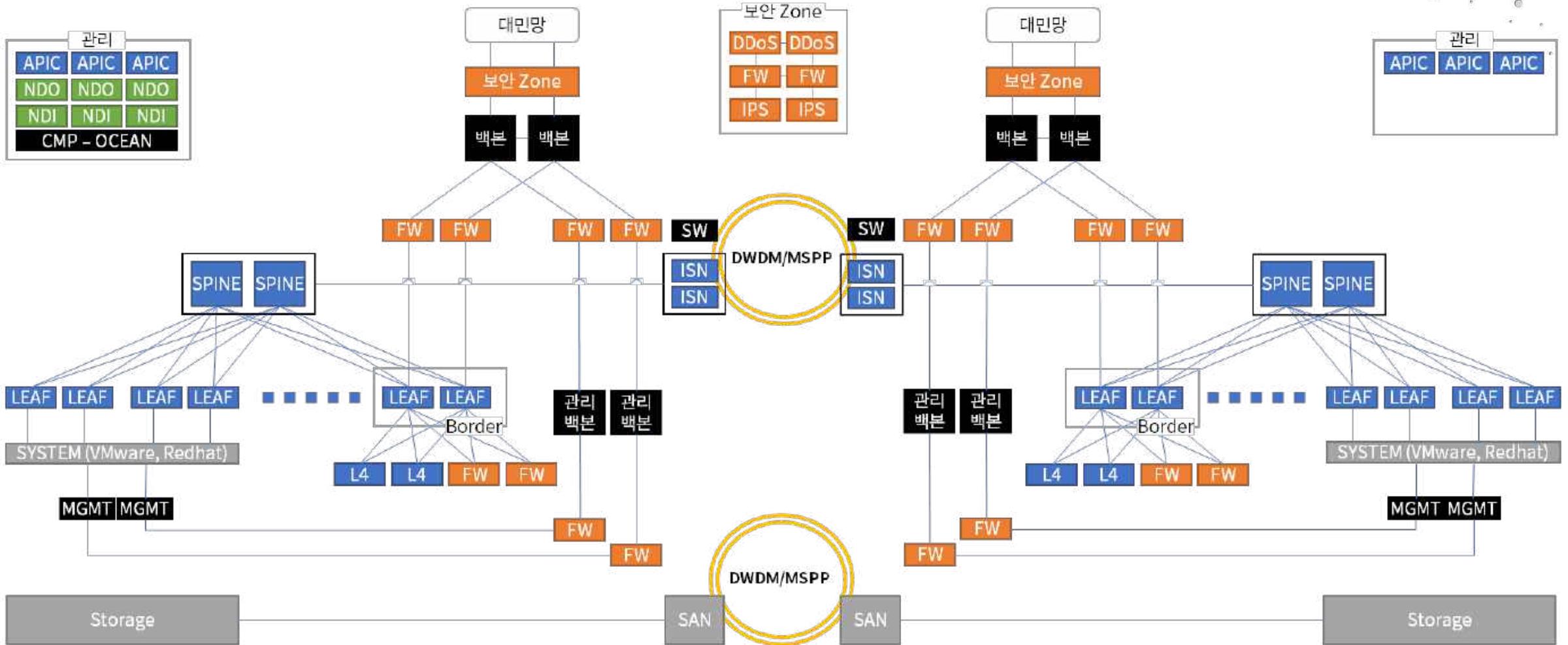
공공 클라우드 데이터센터 설계 Reference

행정망과 대민망을 위한 인프라는 별도 분리하고, 망연계 솔루션을 이용하여 연동합니다.



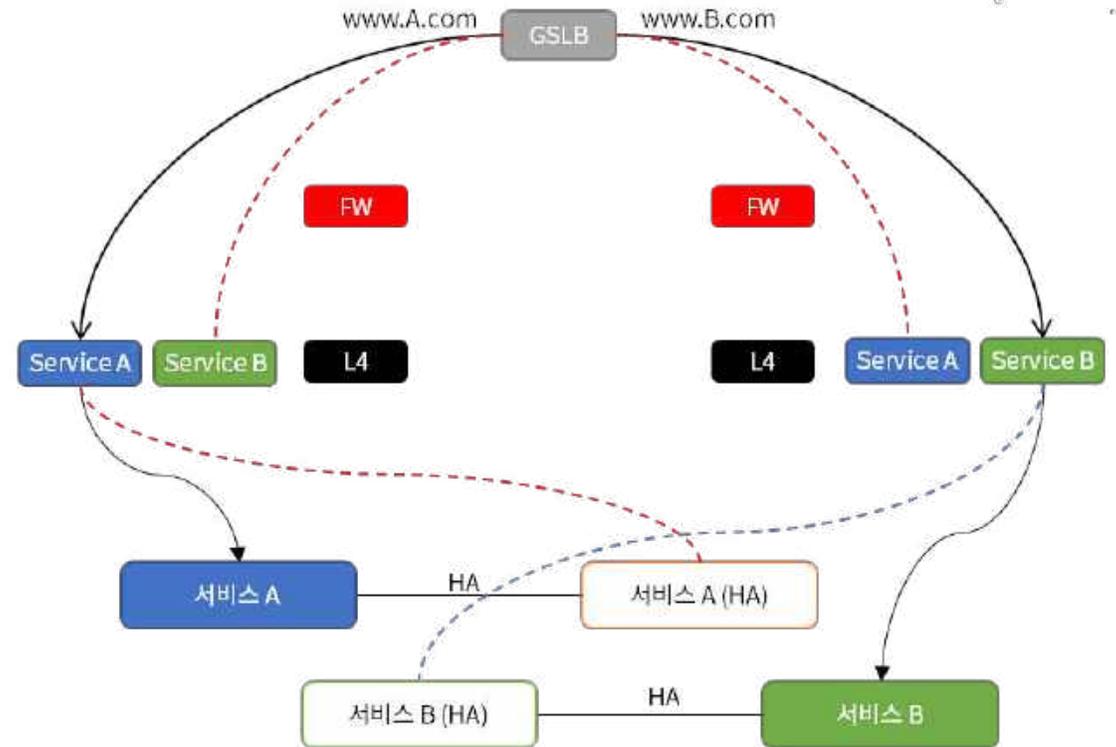
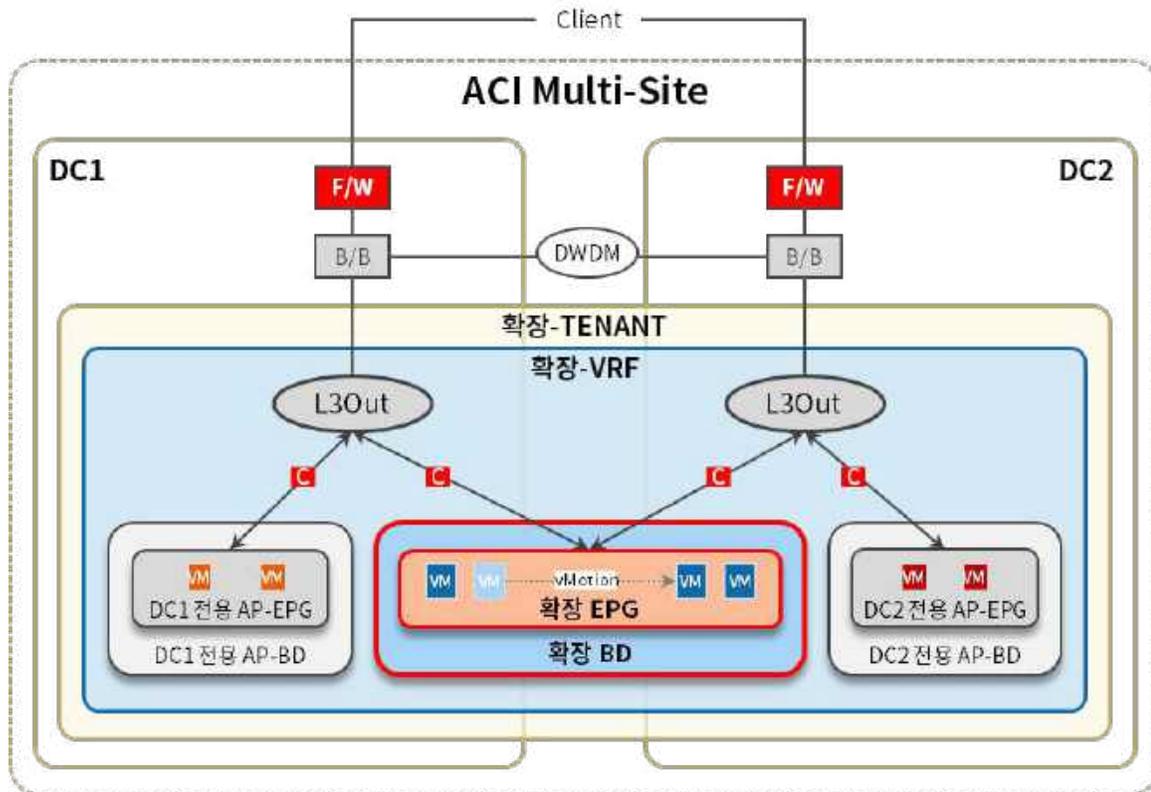
# 2단계 - DR 또는 A-A 데이터센터 (DR 또는 A-A 데이터센터)

2개의 데이터센터를 서로간 연동하기 위해서는 데이터 전송을 위해 전용의 DWDM 또는 MSPP 회선으로 구성합니다.  
VxLAN 기술을 이용해 2개의 데이터센터를 서비스 별로 L2 확장합니다.



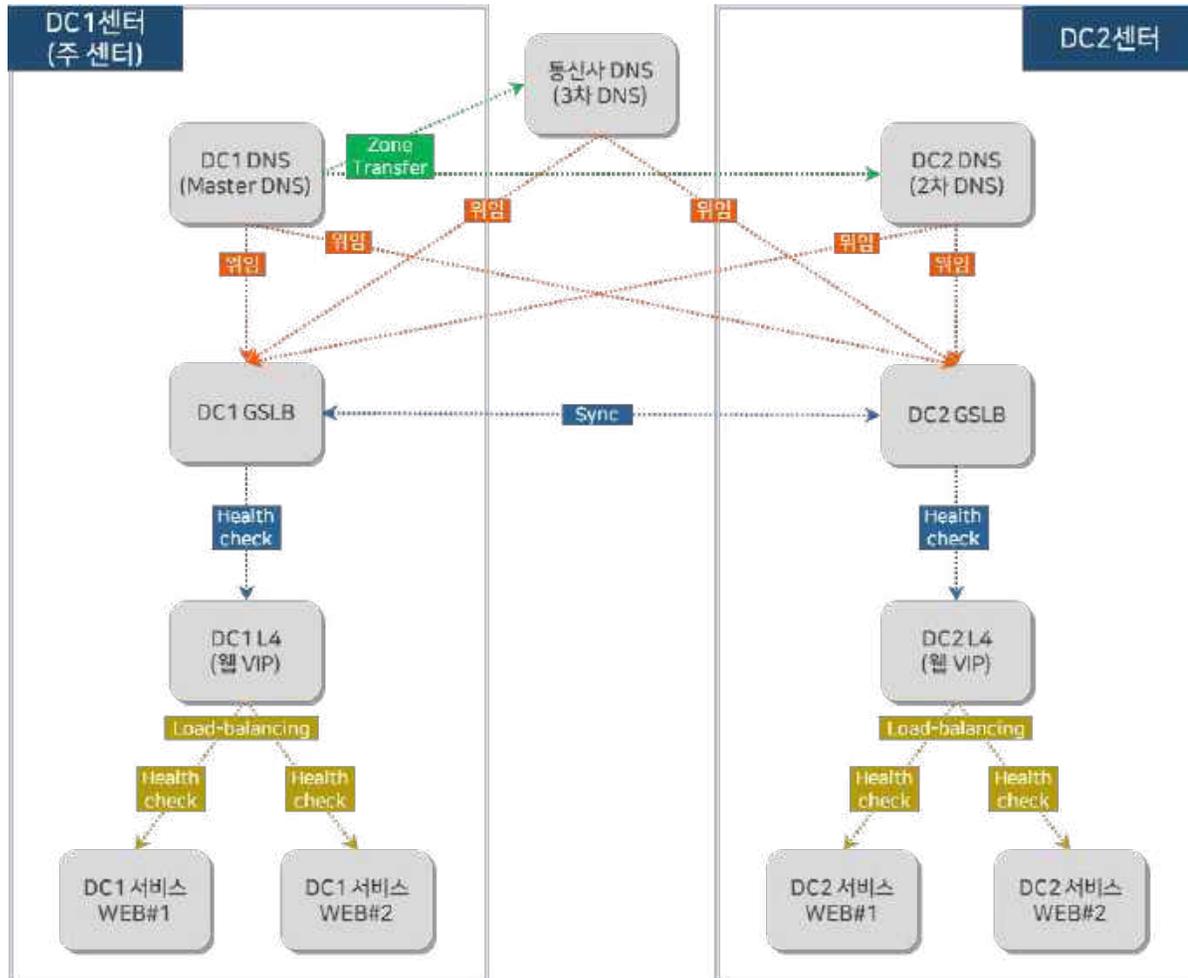
# 2단계 - DR 또는 A-A 데이터센터 (논리적 구성)

논리적으로, 2개의 데이터센터는 데이터센터 자체 서비스와 DC간 통신 서비스로 분리하여 구성하며, 상호간 GSLB 구성을 통해 상호간 서비스를 이중화 합니다.



# 2단계 - DR 또는 A-A 데이터센터 (DNS 및 GSLB 구성)

DNS와 GSLB 구성은 각 서비스의 안정성을 위해 부하분산 처리 합니다.

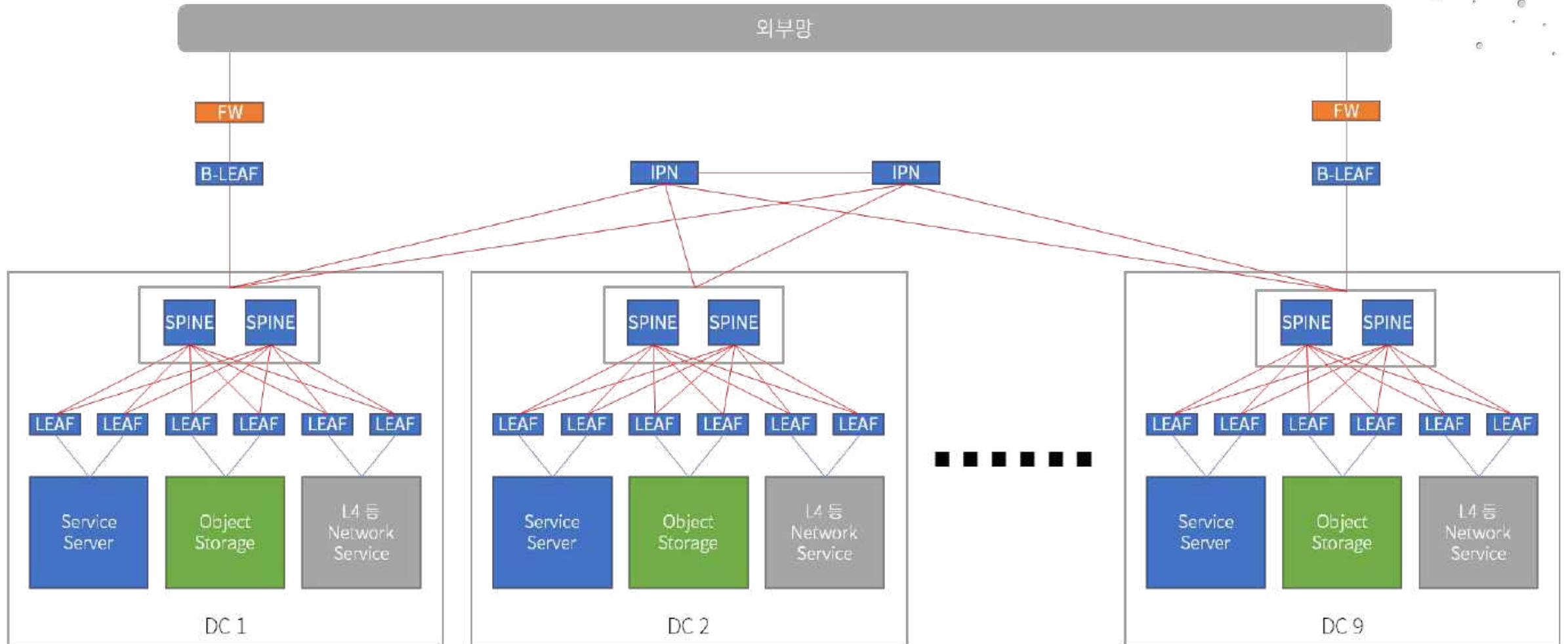


## 구성 세부 내역

- 네트워크 구성
  - DC1센터 서비스 공인 IP 인터넷으로 전파
  - DC2센터 서비스 공인 IP 인터넷으로 전파
  - DC1센터 서비스 공인 IP를 DC1센터 서비스 L4 VIP로 NAT 처리
  - DC2센터 서비스 공인 IP를 DC2센터 서비스 L4 VIP로 NAT 처리
- DNS 구성
  - DC1센터 DNS를 마스터 DNS로 설정
  - Master DNS와 2차(DC2센터), 3차(통신사) DNS 간 Zone Transfer 설정
  - 서비스 DNS 네임쿼리 GSLB 서버에 위임 처리 설정
- GSLB 구성
  - DC1센터 GSLB에 DC1/DC2센터 서비스 공인 IP 등록 및 TTL 값 조정
  - DC2센터 GSLB에 DC1/DC2센터 서비스 공인 IP 등록 및 TTL 값 조정
  - TTL 값은 센터 재해 시 Failover 시간을 보장 함
  - 단, 짧은 Failover을 위해 TTL 값을 짧게 설정 할 경우 DNS와 GSLB가 그만큼 많은 네임쿼리 요청을 처리해야 함
  - DC1/DC2센터 L4/L7 VIP에 대한 Health check 설정
- L4/L7 구성
  - DC1센터 L4/L7은 DC1센터 Web 서버 Load-balancing 설정
  - DC2센터 L4/L7은 DC2센터 Web 서버 Load-balancing 설정

# 2단계 - 통합 관리를 위한 인프라 확장

데이터센터의 통합관리 목적으로 동일 지역에 존재하는 다수의 데이터센터의 네트워크를 확장하여 하나의 네트워크 환경으로 관리/운영

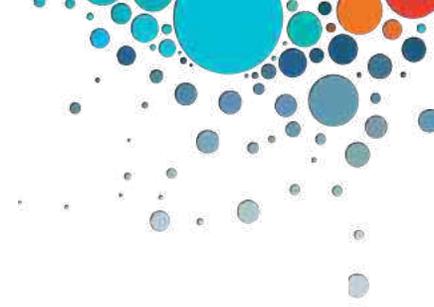


# 통합관리

Cloudwave 솔루션 Demo



# Cloudwave 솔루션 소개 및 특징



**1. 멀티 클러스터 통합 관리**

**2. Active-Active 데이터센터**

**3. A.I/ML 기반 Day2 Operation 지원**

# 구현 환경

Active-Active 데이터센터  
Multi-Cluster 환경

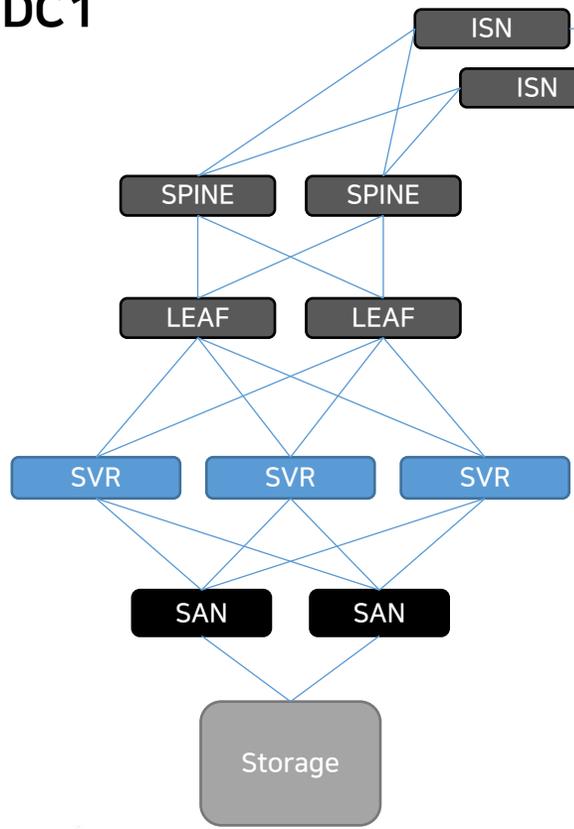
- Day2 Operation
- Cloudwave
  - CMP OCEAN
  - 심포니
  - Nexus Dashboard

- Public Cloud
- GSLB
  - VPN



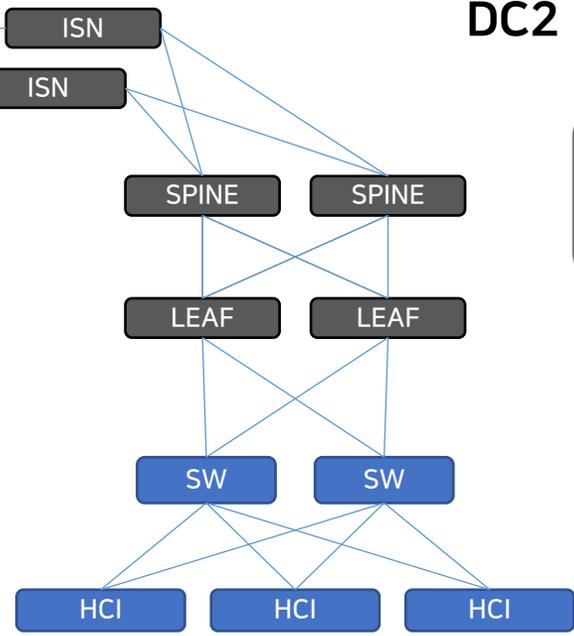
DC1

- MGMT
- SDN 컨트롤러
  - vCenter

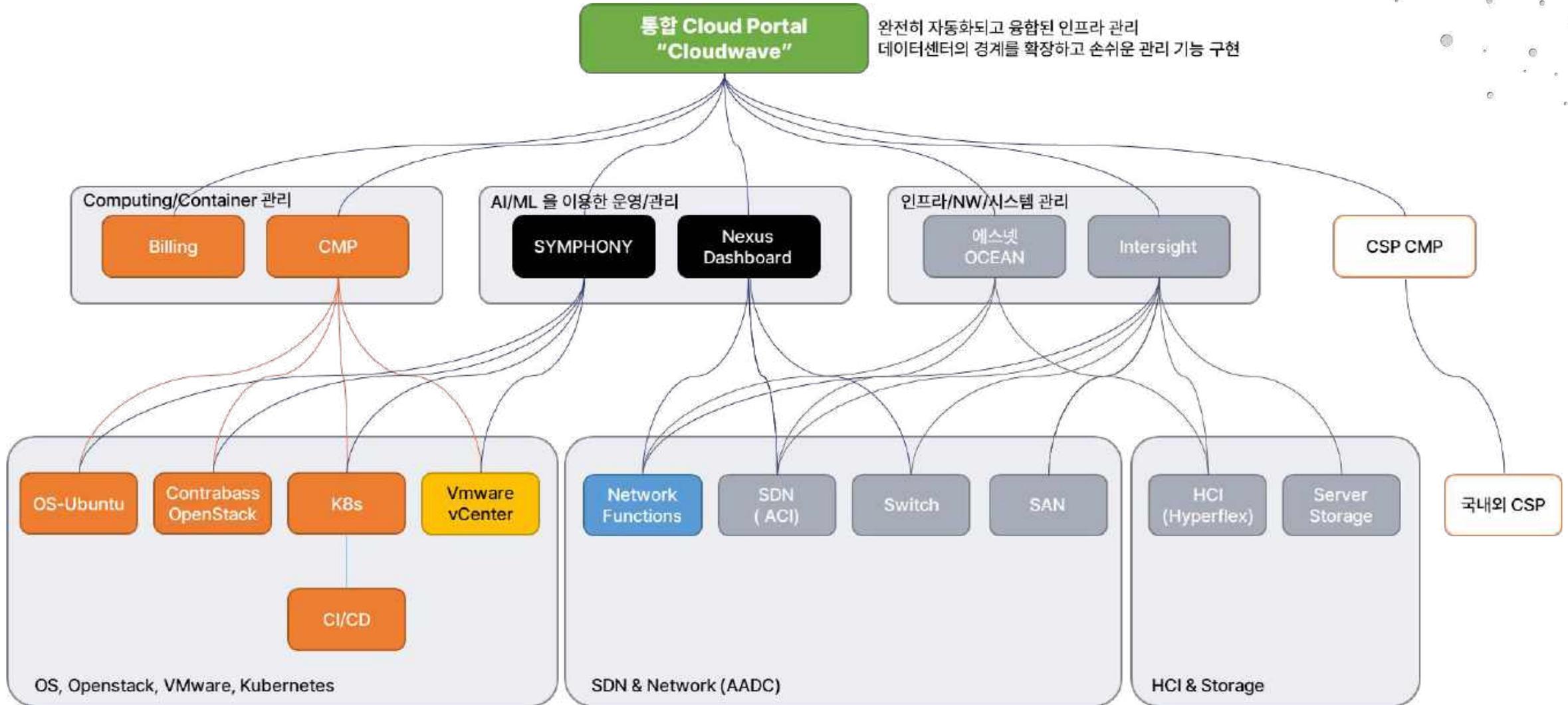


DC2

- MGMT
- SDN 컨트롤러
  - vCenter



# 통합 Cloud Architecture



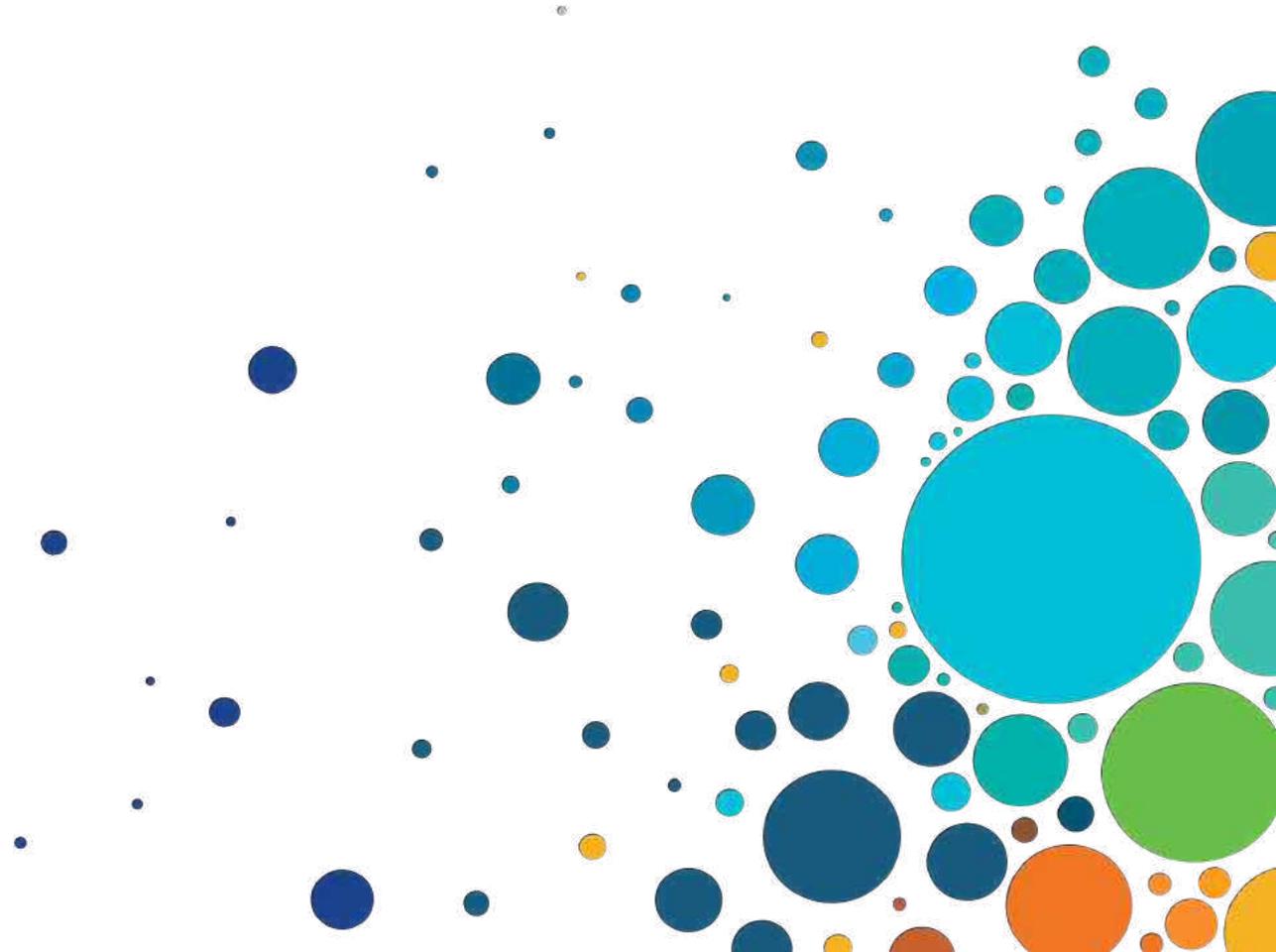
# 사용자 포탈

Self-Service Portal Demo (2분30초)



# 관리자 포탈

통합관리 Demo (4분15초)

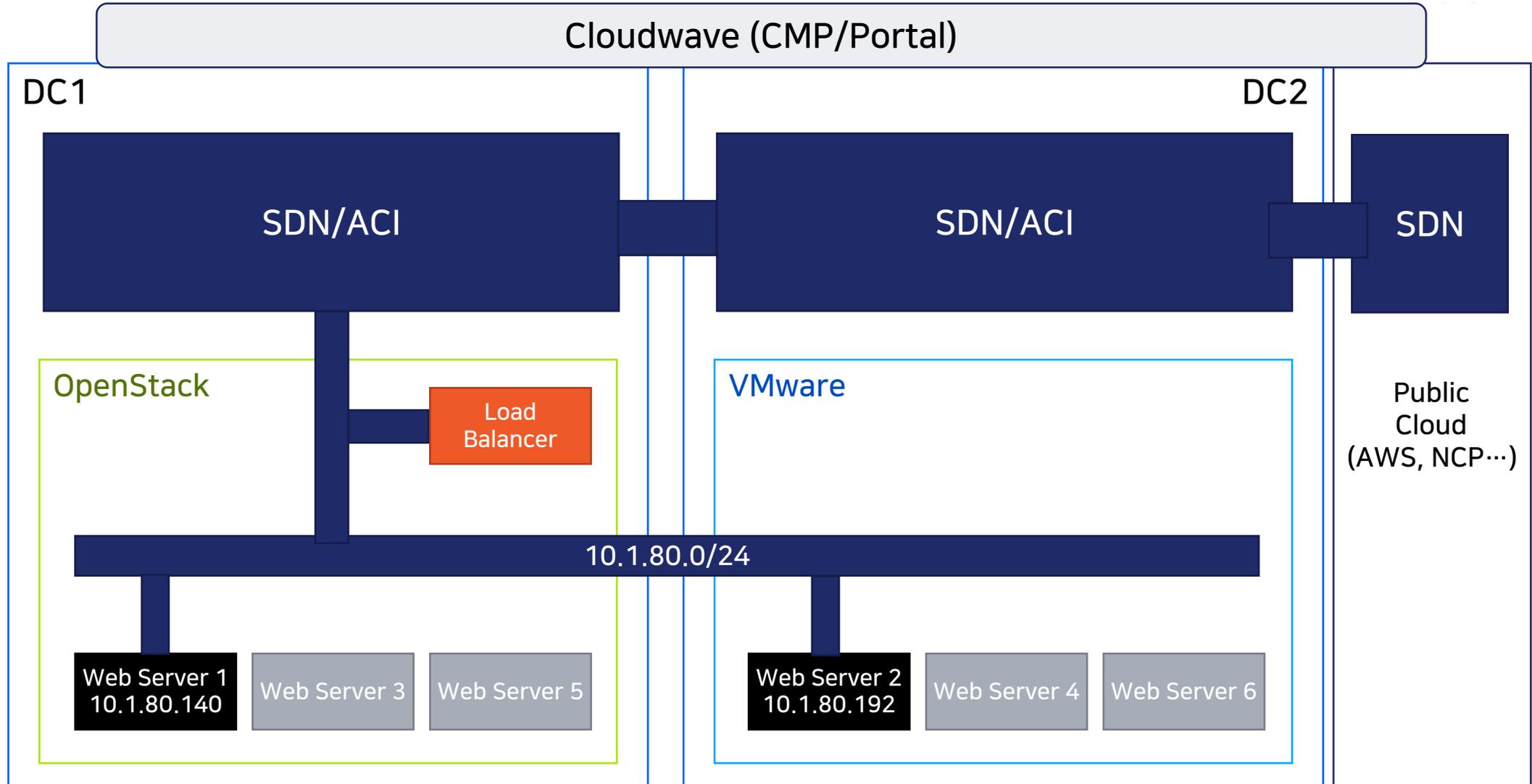


# AADC 구성 Demo

Multi DC간 L4를 이용한 Load Balancer 구현 (2분56초)



# 통합 네트워크 관리



# 통합 네트워크 관리 - L4 스위치 Octavia 생성

CONTRABASS OKESTRO CMP

로드밸런서 > 네트워크 > 로드밸런서

상태	로드밸런서명	설명	IP	운영 상태	관리 상태	작업
🔍	snet-octavia		10.1.80.37	오프라인	🔍	⋮

로드밸런서 등록

가상머신 추가

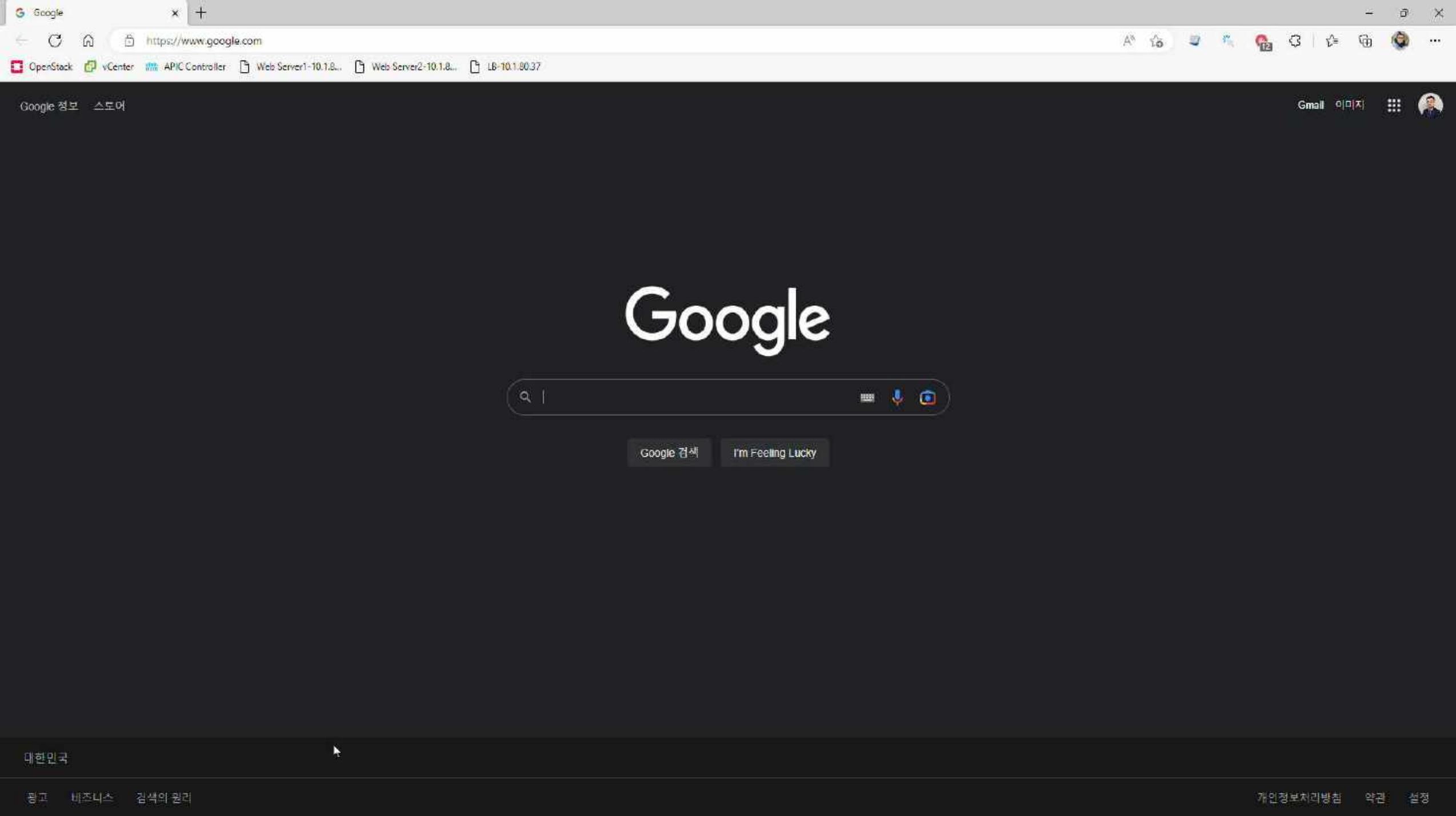
로드밸런서를 등록합니다.  
포시 항목(\*)은 필수 입력 항목입니다.

가상머신명	서브넷	IP주소	포트	삭제
webserve_r1	snet-sub-net	10.1.80.140	80	🗑️
webserve_r2	snet-sub-net	10.1.80.192	80	🗑️

전체 - 🔍 검색하기

가상머신명	서브넷	IP 주소	추가
가상머신 정보가 없습니다.			

이전 다음



# Google

Google 검색 I'm Feeling Lucky

# Summary

지난 20여 년간 다양한 사업에서 경험한 **Network과 시스템에 대한 에스넷시스템의 Know-how와 역량**을 집중하여 고객의 클라우드 운영 환경을 **단일 플랫폼으로 관리**할 수 있도록 개발했습니다.

VMware와 Openstack, Kubernetes환경이 구성된 Private Cloud 환경 뿐만 아니라, **AWS나 NCP와 같은 CSP와의 연동, 즉 멀티 하이브리드 클라우드 환경**을 하나의 CMP에서 통합 관리 가능하며, 시장에서 가장 많이 사용되고 있는 SDN 솔루션과 **통합되어 네트워크와 컴퓨팅을 동시에 배포하고 자동화 하는 솔루션은 금번 개발 제품이 유일합니다.**

SDN 기술을 이용하면 2개 이상의 Multi Datacenter를 **Active-Active로 운영** 가능하며, AWS와 같은 **Public Cloud와 연동기능을 통해 손쉽게 일관된 정책을 구현**하실 수 있습니다.

**AI / ML 기술을 이용**하여 Network에 대한 가시성과 인프라 운영에 대한 효율성을 제공하고, Computing에 대한 적정자원관리와 HPA (Horizontal Pod Auto-scaling) 파라미터 추천 기술을 이용하여 **서비스에 대한 선제적 대응**이 가능합니다.



The bridge to possible

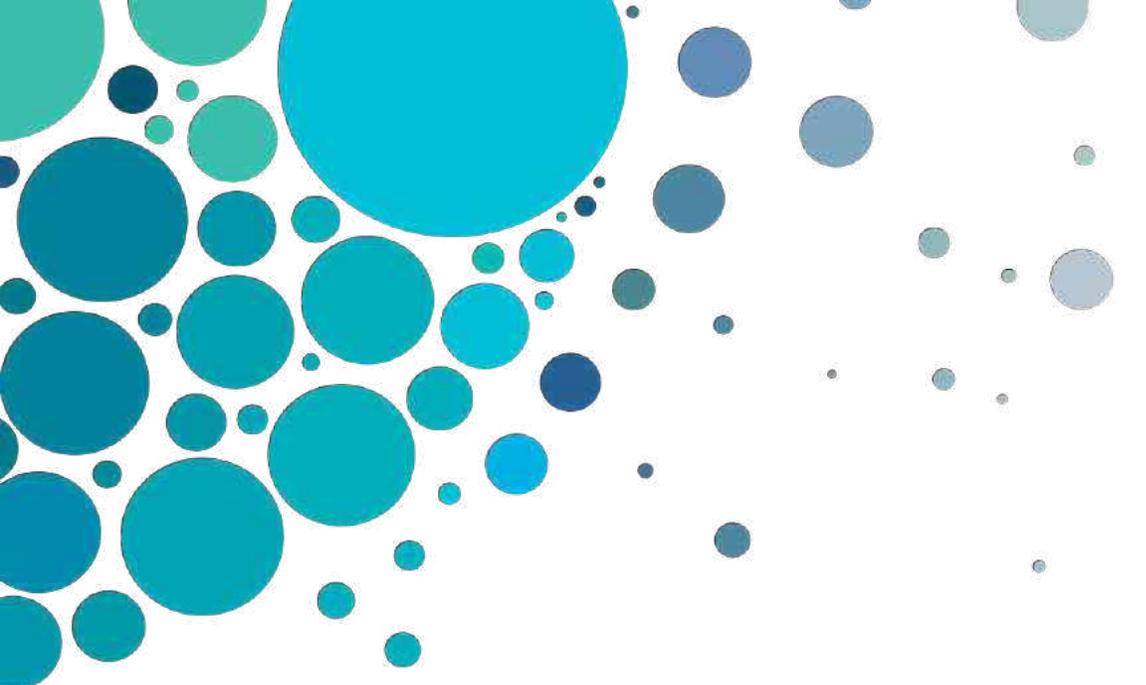
감사합니다.



The bridge to possible

# IT/OT 컨버전스 아키텍처를 통한 스마트팩토리의 진화

이정표 프로 | 시스코코리아, 산업용 네트워크 세일즈그룹 프로



# Agenda

- 1 Why**  
왜 IT/OT 컨버전스가 필요한가?
- 2 What**  
어떤 것을 IT/OT 컨버전스라고 말하는가?
- 3 How**  
이를 위해 어떤 역량이 필요한가?
- 4 Summary**  
이것만 기억해주시기 바랍니다

# “OT”를 어떻게 바라보고 계십니까?

Operational Technologies

설비

센서/모터

공장

자동화

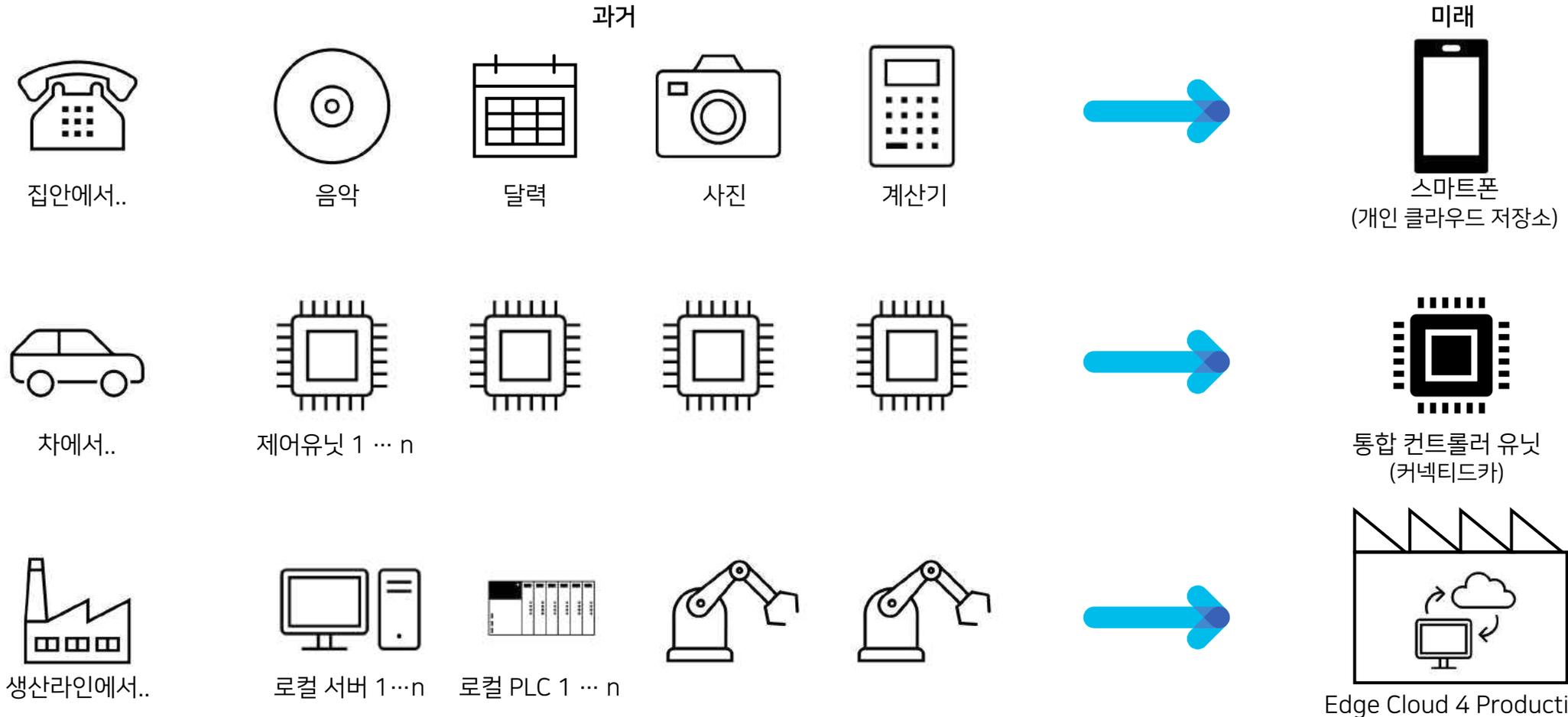
디지털트윈

OT 데이터



# “Data” 처리장치는 빠르게 통합되고 있습니다

Data 처리는 로컬 하드웨어 기반에서, 통합 엣지로, 그리고 Edge-Cloud로 전환중에 있습니다



# OT의 IT화는 이미 시작되었습니다

Audi의 예시는 미래의 제조환경이 하드웨어 제어 환경에서, 소프트웨어 제어 환경으로 변화할 것임을 시사합니다

Audi production today...



Audi production tomorrow:  
a smart factory not a maintenance nightmare

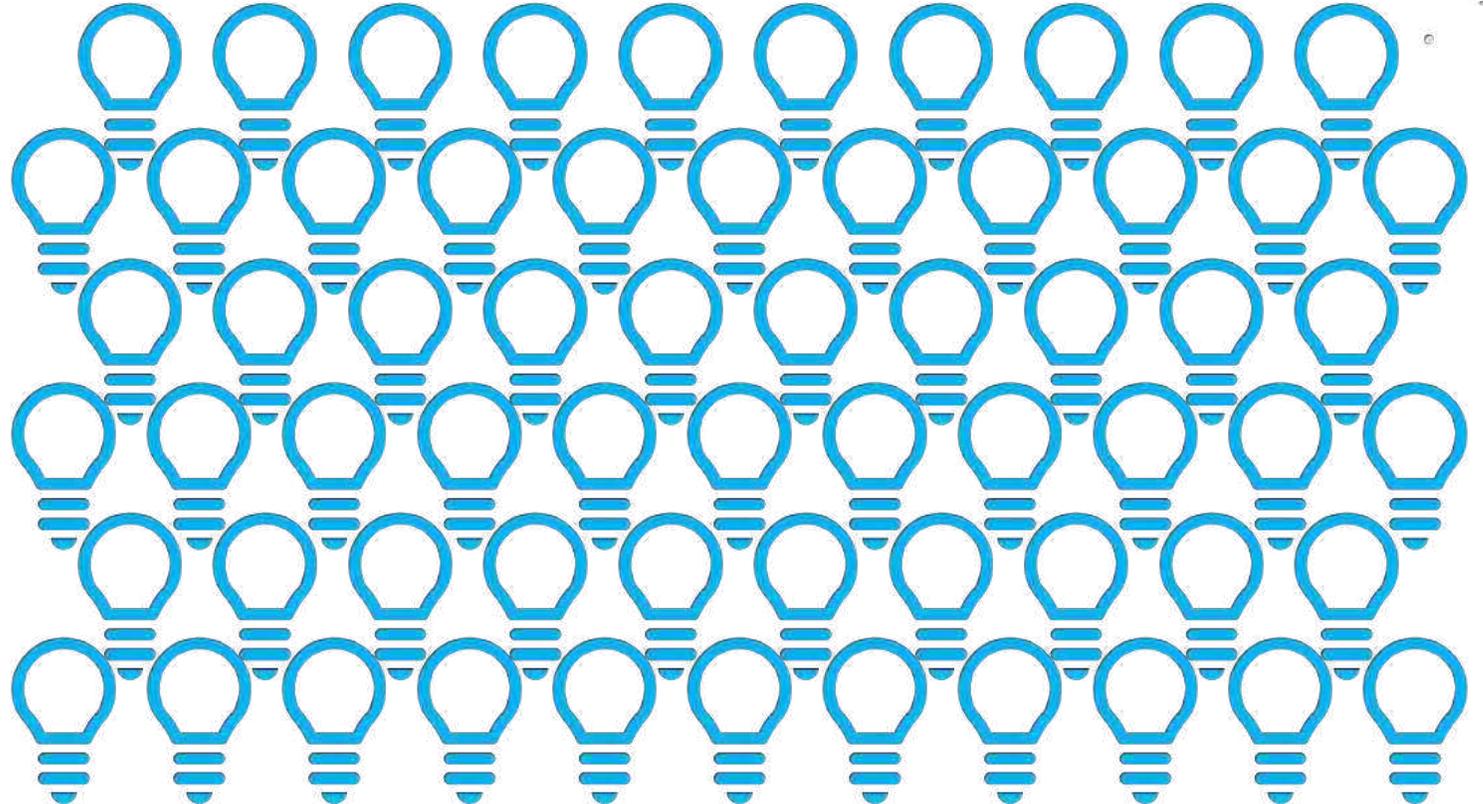


# Imagine “N” times of update



하나의  
스마트전구

VS

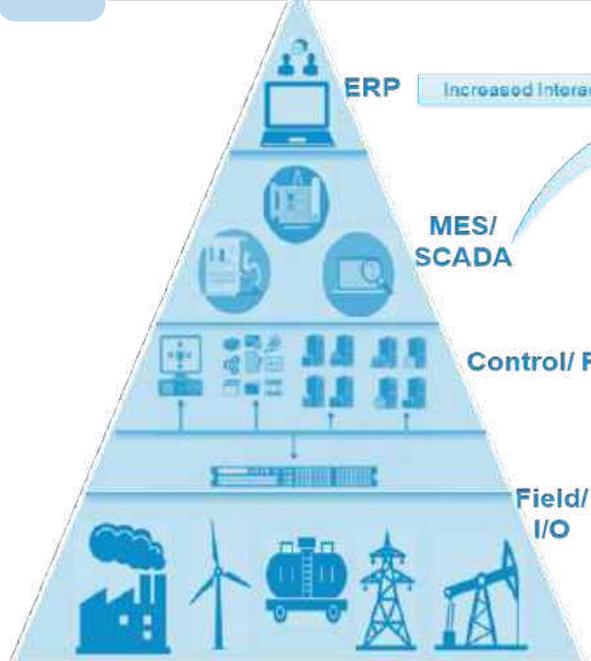


1,000개의 다양한 제조사에서 구입한  $10^5$  개의 스마트 전구

# IT와 OT의 연결 “Simple” 해야 합니다

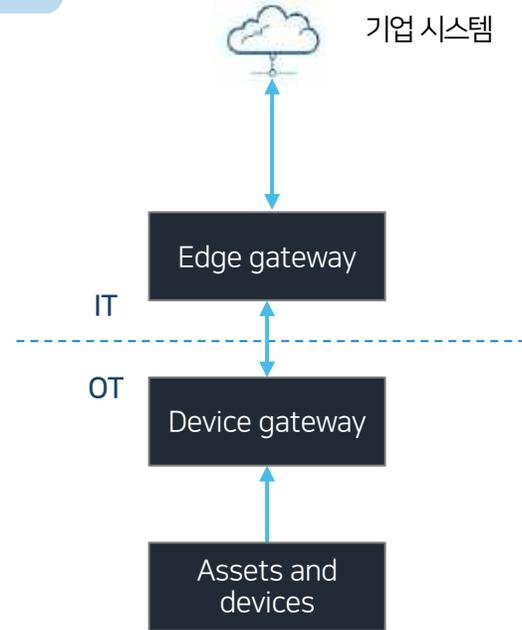
현재 OT의 복잡성은 매우 높아진 상태입니다. Frost & Sullivan 보고서에 따르면, 미래 제조 환경은 단순화, 연결성, 개방형으로 전환될 것이며, 과거 독자사양 및 이기종 스택은 사라지고 데이터 표준화, 분석 성능 및 보안이 고려된 아키텍처로 전환될 것입니다.

## Industry 3.0 과거 / Hierarchical



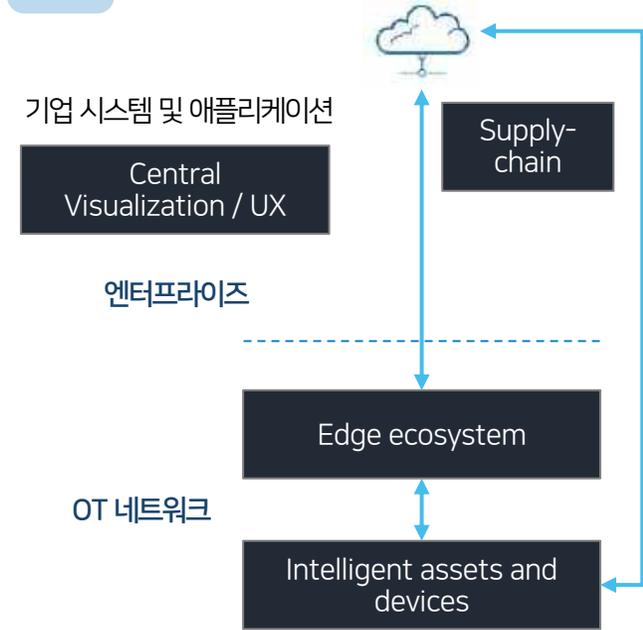
독자 사양 | 이기종 산업용 프로토콜  
| 지연 발생 | 실시간 분석 기능 없음

## Industry 4.0 현재 / Edge-to-cloud



표준화 프로토콜 | 오픈 아키텍처 | 프로토콜 해석 | 지연  
최소화를 위해 Edge에서 진단 분석 | 실시간 진단 분석

## Industry 5.0 미래, 2035+ / Asset-to-cloud



인텔리전스 내장 자산/IIoT | Edge (또는) Cloud간  
안전한 보안 연결 | 실시간 진단 분석 | Edge ecosystem

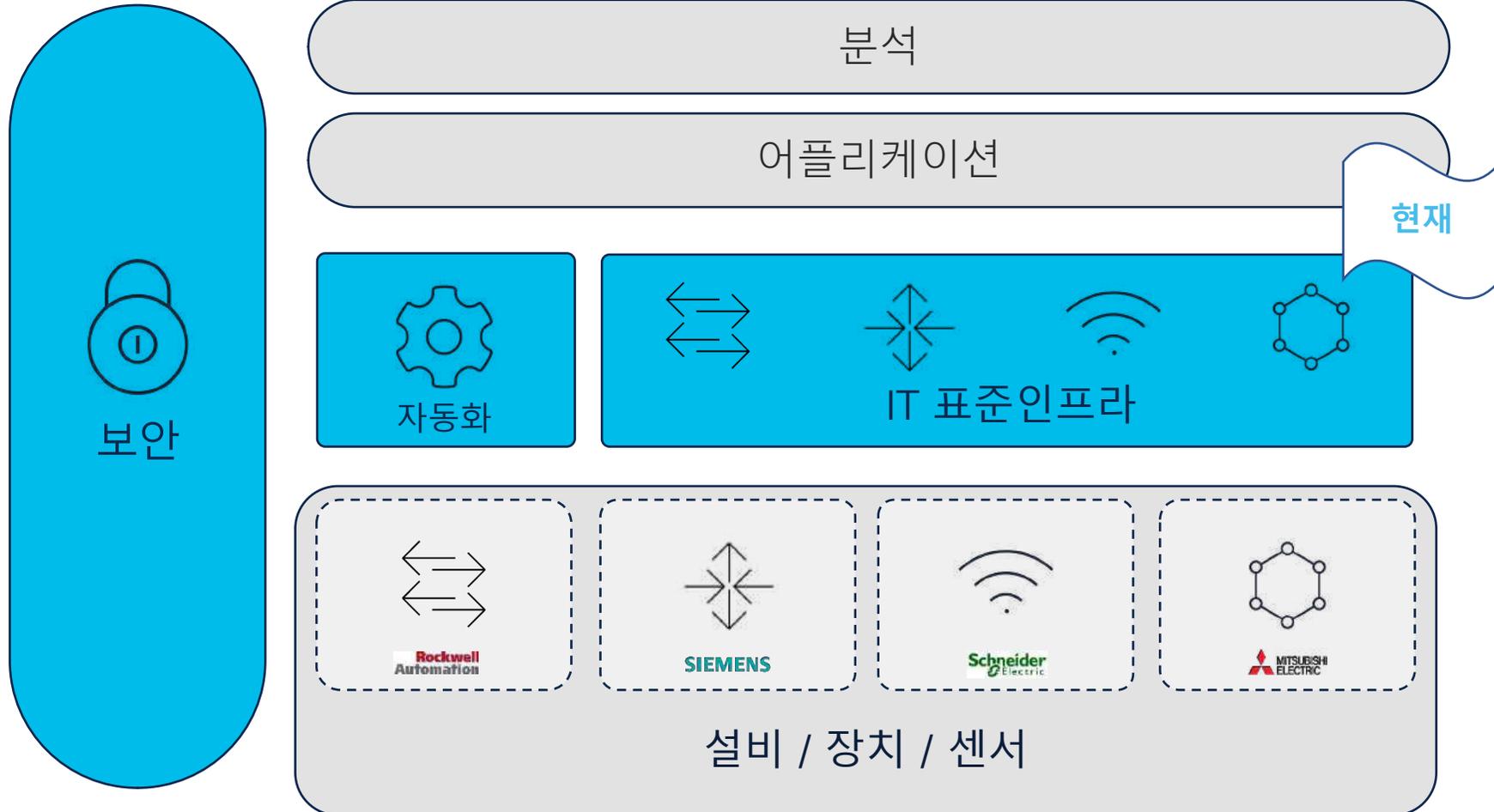
# What

어떤 것을 IT/OT 컨버전스라고 말하는가?



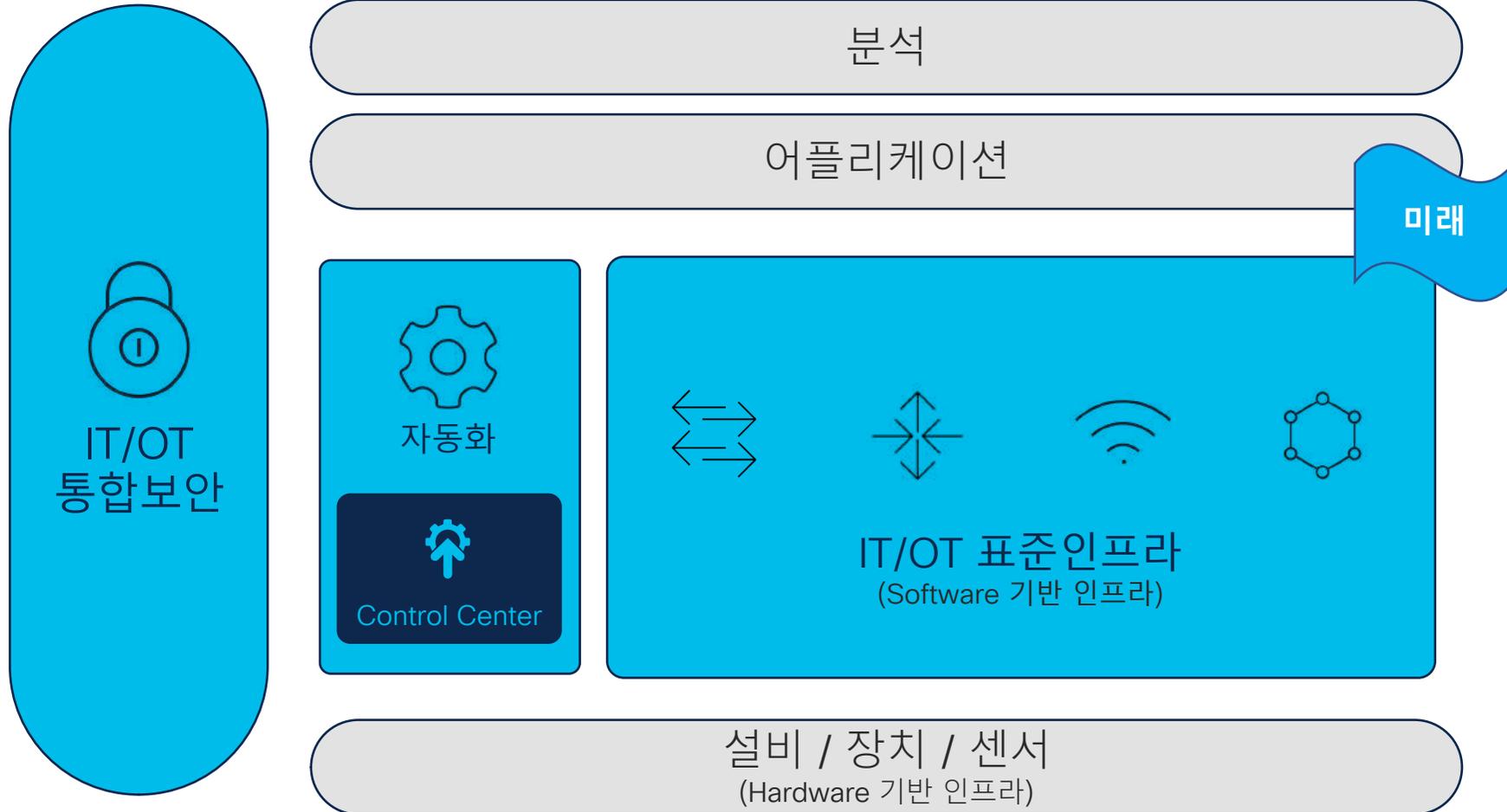
# 제조 IT/OT 인프라의 현재와 미래

현재의 제조 OT 인프라는 각 전문 영역의 설비, 시스템, 레거시 기술로 인해 Silo된 형태로 운영중에 있습니다

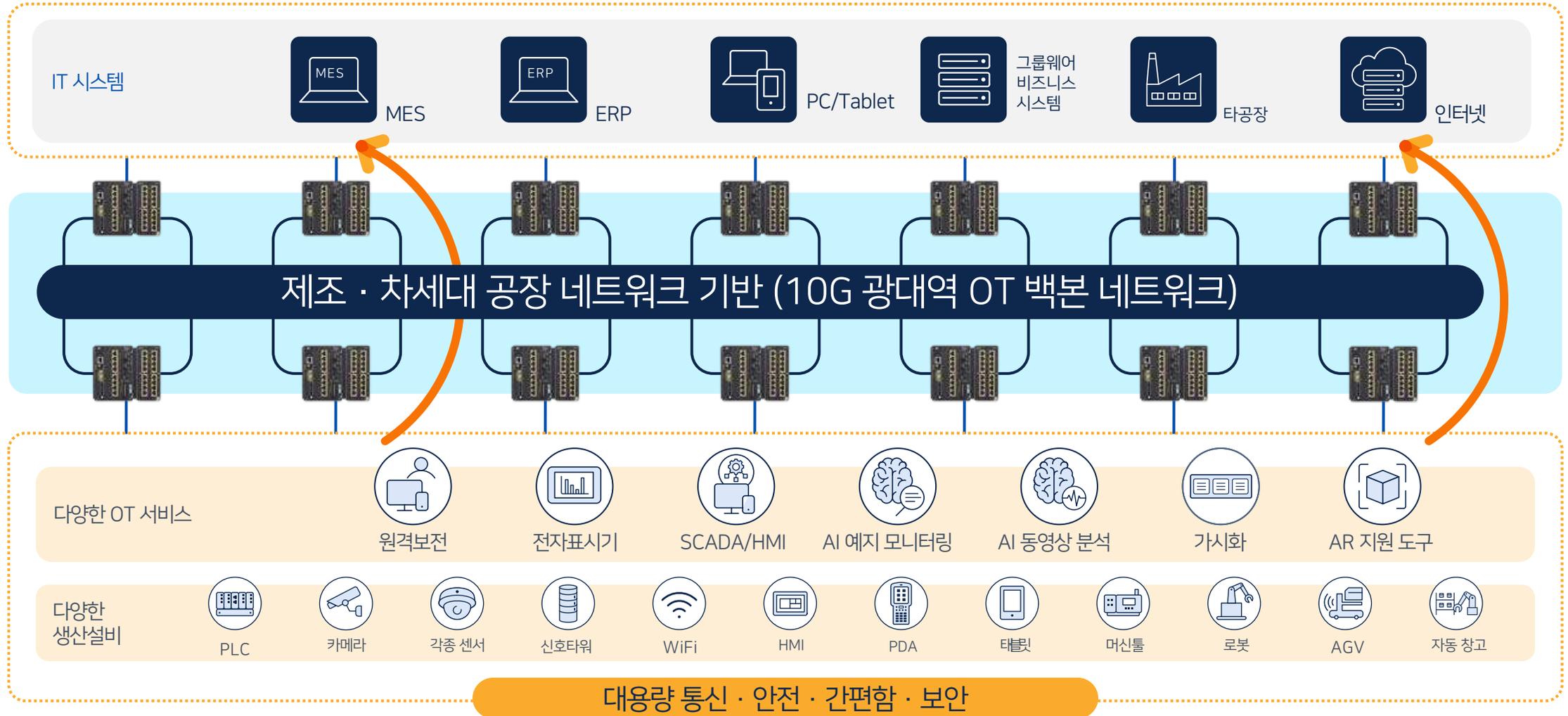


# 제조 IT/OT 인프라의 현재와 미래

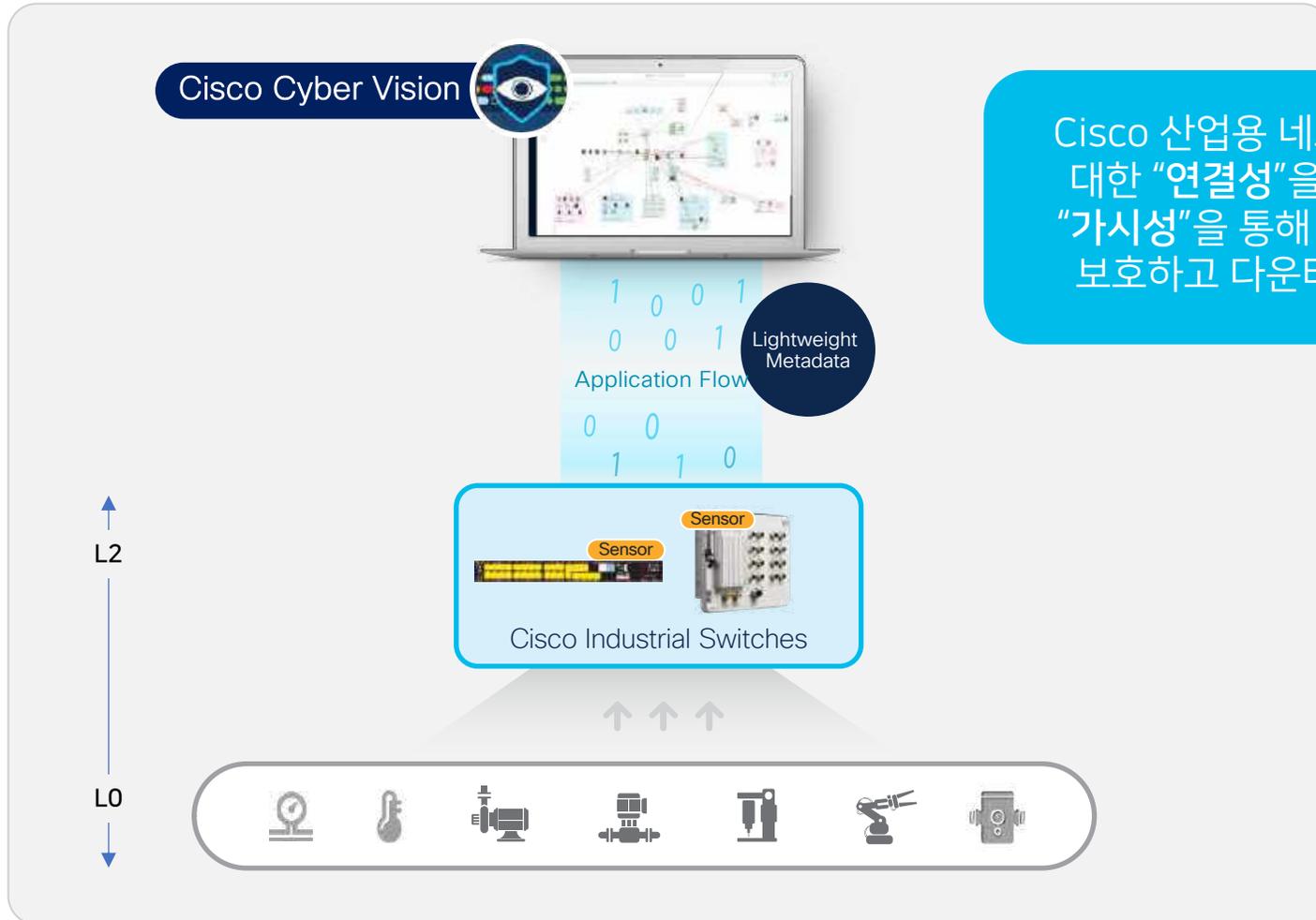
미래의 제조 인프라는 IT와 OT가 단순하면서도 완전하게 소프트웨어 기반으로 통합된 아키텍처로 운영될 것입니다



# [적용 사례] IT/OT 컨버지드 구현을 위해 10G망으로 전환



# [적용 사례] 컨버지드 인프라를 통해 연결성 및 가시성을 동시에



Cisco 산업용 네트워크는 중요 자산에 대한 "연결성"을 제공할 뿐만 아니라 "가시성"을 통해 사이버 위협으로부터 보호하고 다운타임을 감소시킵니다



# How

이를 위해 어떤 역량이 필요한가?



# [1] 레퍼런스 아키텍처를 통한 검증된 설계 도입

Cisco Validated Design

시스코의 오랜 학습을 통해 사전 검증된 “표준화 설계모델”을 통해 “소프트웨어 기반” OT 인프라로 전환 가속화

## CVD 아키텍처 도입 기대효과

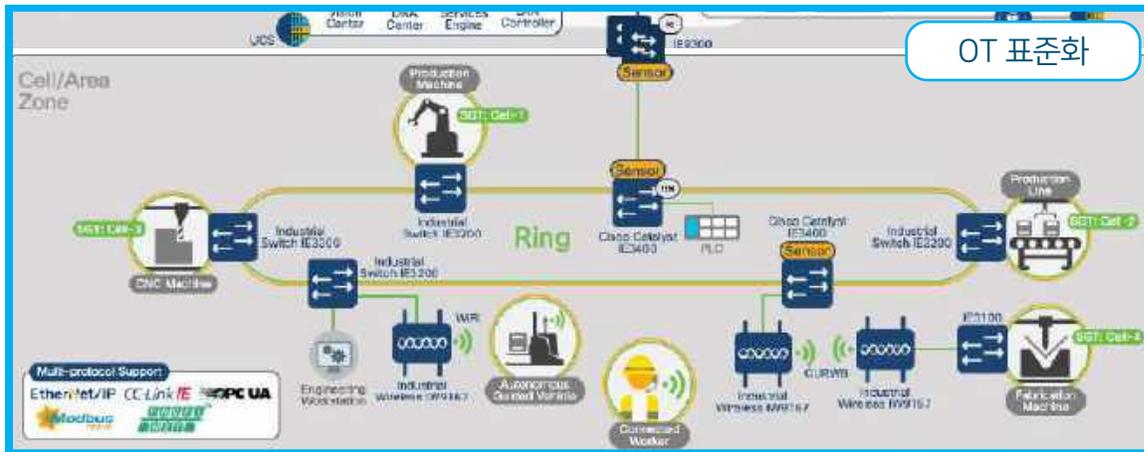
- 사전 검증된 레퍼런스 아키텍처에 따른 도입 가속화
- 설비 네트워크 표준화(L0-L2)를 통한 중복투자 축소
- 코드형 인프라(Infrastructure-as-a-Code)를 통한 자동화

## 증가하는 OT 인프라 요구사항 만족

- OT 세그멘테이션(Zone & Cell)
- OT 가시성 및 단말관리
- OT 기술변화 수용 및 다양한 OT 벤더 시스템 통합

## 새로운 디지털 기술 도입 준비

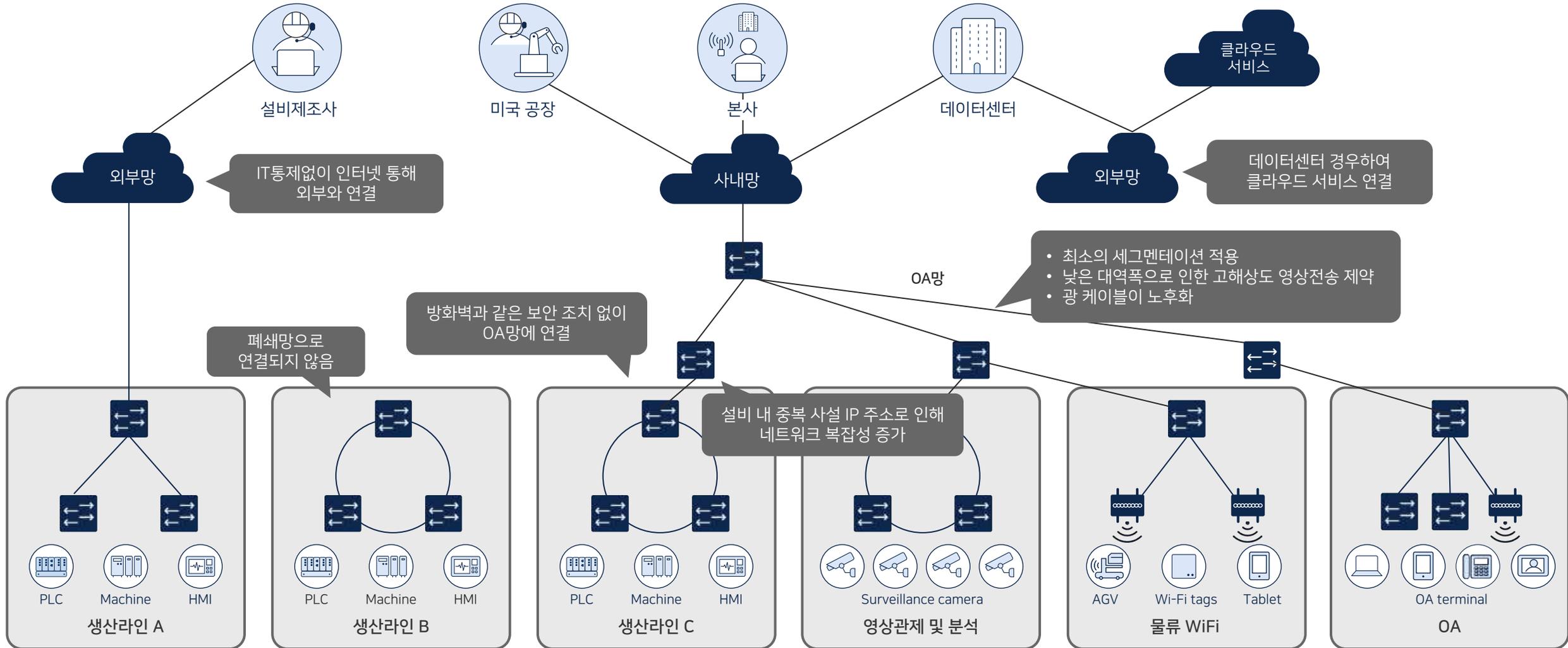
- 팩토리-to-클라우드 서비스 지원
- 글로벌 운영화에 따른 원격 전문가 지원 및 설비연결
- 가상화 어플리케이션을 엣지 단에 구현



주요 시스템 공급사와 협력

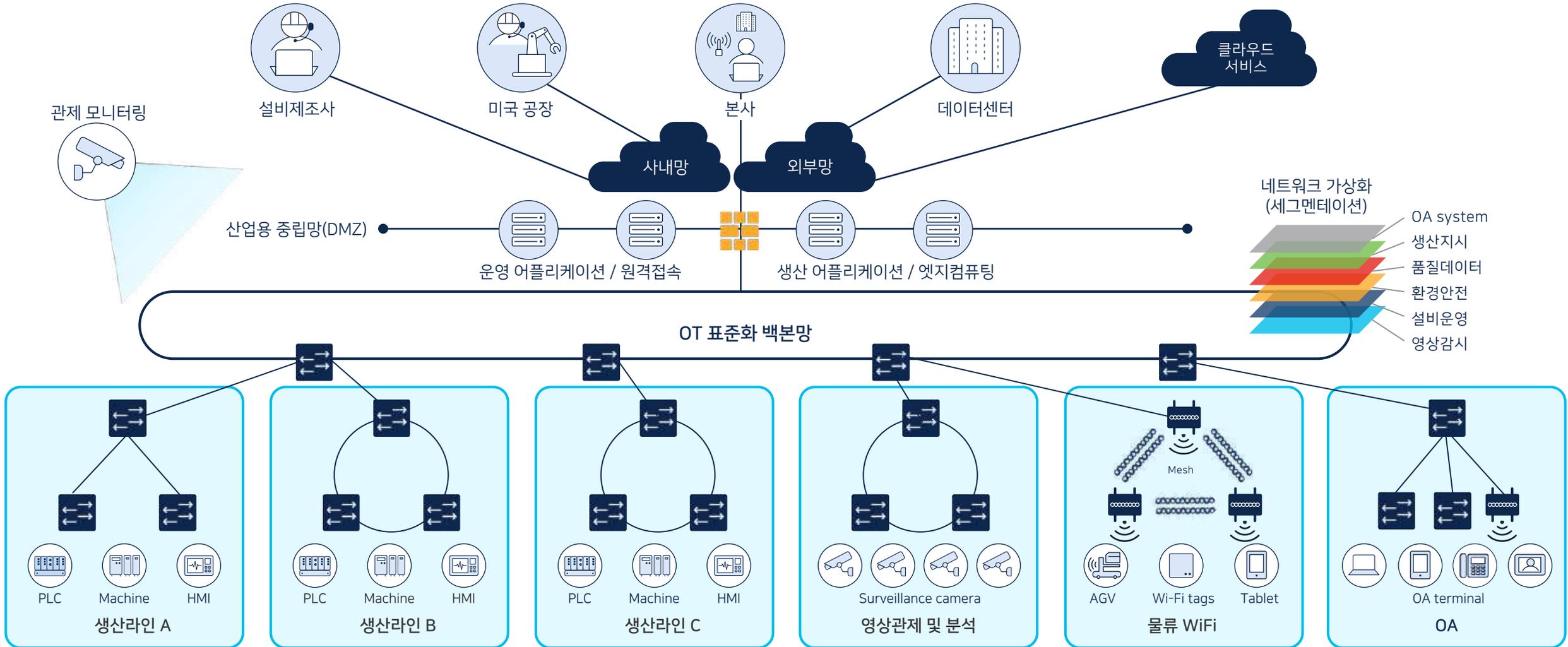
# As-Is 제조 레거시 인프라

현재 공장의 모습 예시



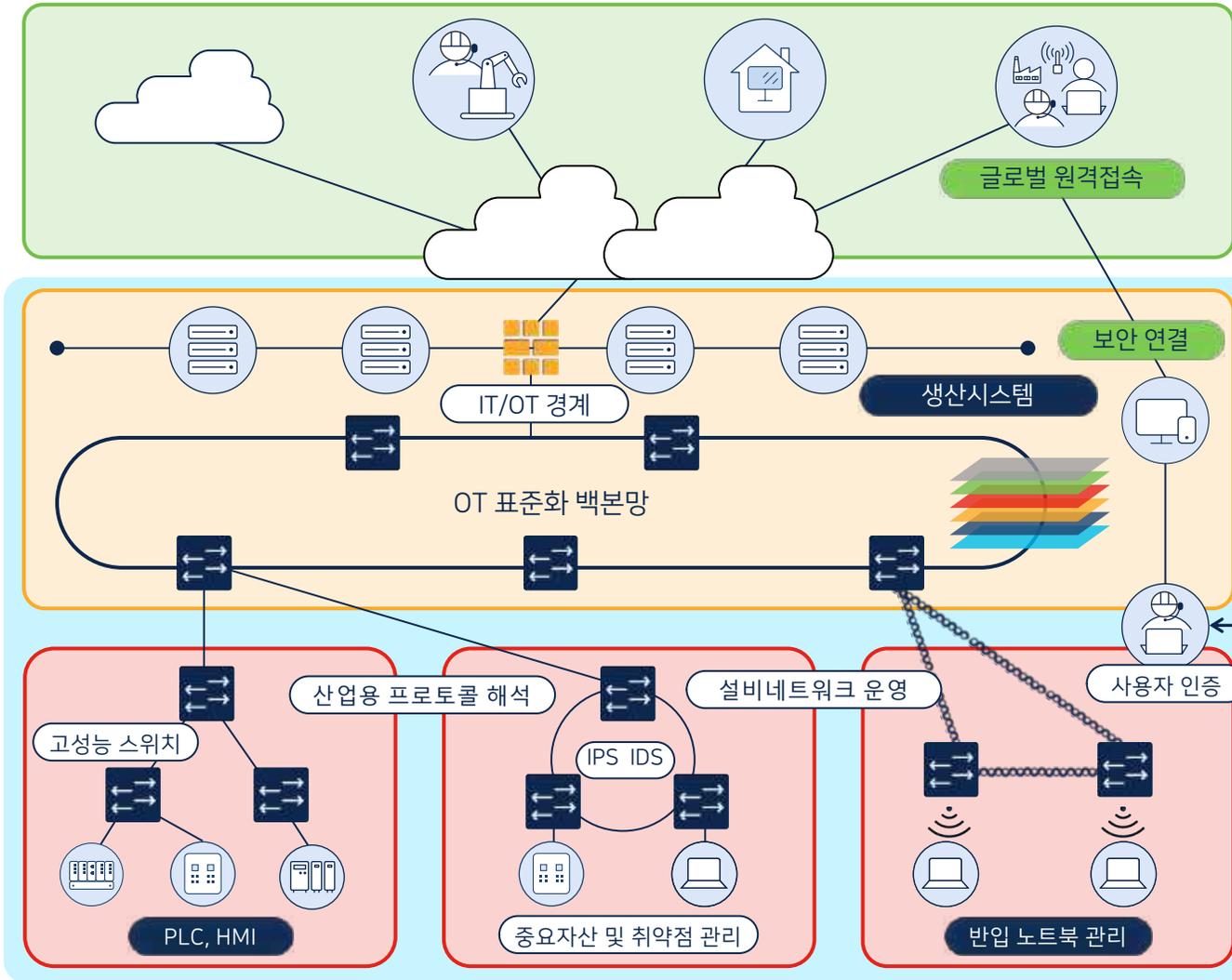
# To-Be 제조 IT/OT 컨버지드 인프라

향후 공장의 모습 예시



# To-Be 제조 통합 보안 아키텍처

향후 공장의 모습 예시



## 1 자산 식별, 취약점 탐지 및 NW 고도화



### DNA: 통합 네트워크 운영

- IT/OT, 무선 등 모든 네트워크 통합 운영
- NMS 관점에서 시각화
- 네트워크 장애 감지



### Cyber Vision: ICS 관리 및 취약성 시각화

- 산업용 프로토콜 이해
- 자산 및 ICS 취약성 식별
- 네트워크에 장비 추가 및 삭제 기록

## 2 네트워크 세그멘테이션



### Firepower: IT 네트워크 및 생산 장비 네트워크 경계 방어

- 차세대방화벽/IPS 엔진/멀웨어 차단
- 허용된 통신만 안전하게 연결
- IT/OT 공유 데이터 서버
- 글로벌 원격전문가 접속 허용



### ISE: 접근권한 및 제어

- 사용자 및 장치 액세스 관리
- 외부 벤더의 비인가 단말의 네트워크 접속 통제



### IE 스위치: 관리형 스위치로 표준 구현 보안 기능으로 방어

- 포트 비활성화
- 논리적 분리(목적지 제어)
- 케이블 제거 감지

# [2] IT-OT 통합 운영이 가능한 기술을 선택해야 합니다

IT 조직에서 사용하는 기술을 OT 환경에서 사용해야 합니다



**'기업용 환경'에 맞게 설계됨**

포트 집적도, 대역폭, PoE 전력량

**'산업환경'에 맞게 설계됨**

온도, 습도, 진동, 충격, 전기 서지 등에 대한 산업 인증

**OT 사용자 용이성**

SD-CARD를 통한 빠른 제품교체 방법, I/O 알람 릴레이 등

**산업용 특징 및 기능**

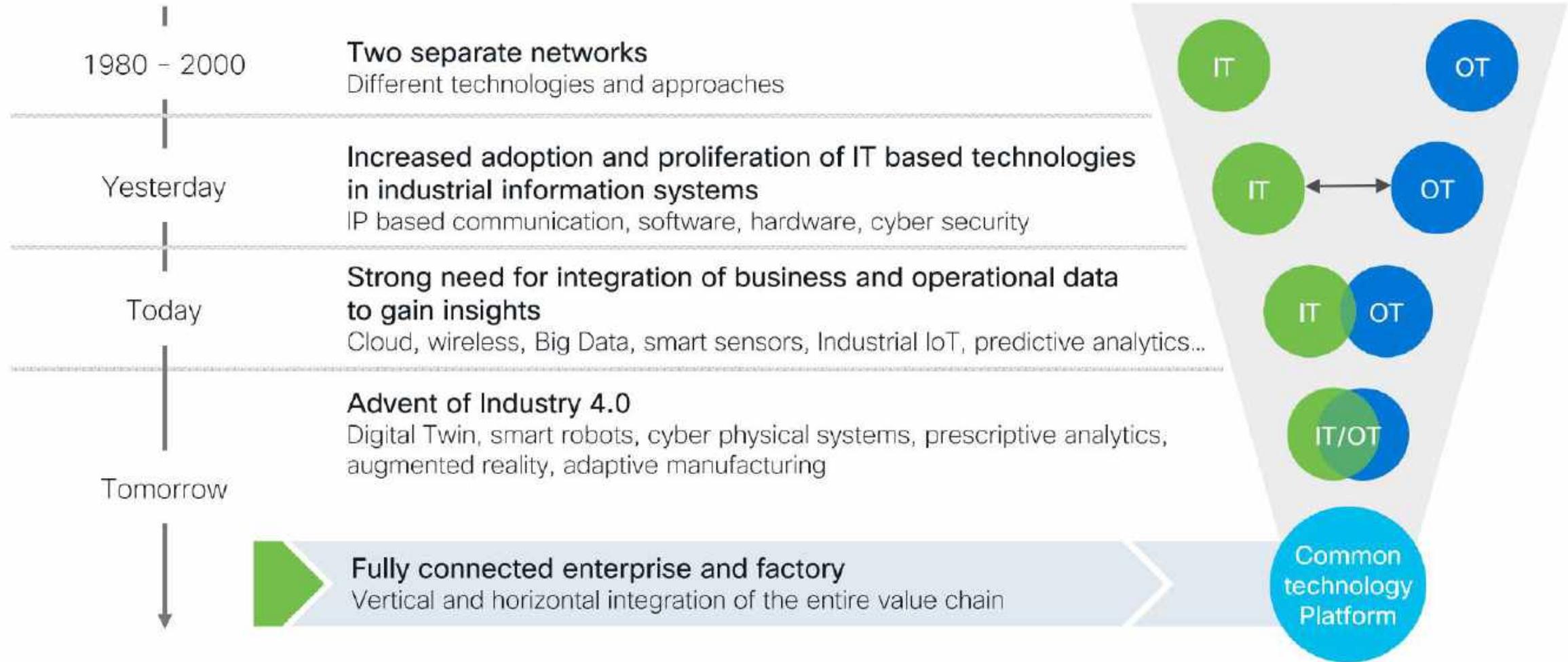
PROFINET, Ethernet/IP, Modbus TCP, MRP, DLR과 같은 프로토콜 지원

운영시스템 (Cisco IOS XE)

네트워크 자동화 관리도구 (Cisco DNA Center)

# [3] IT와 OT 조직의 협업이 KEY

20년동안 분리되어 왔던 IT와 OT 조직에서, 이제는 공통의 기술 플랫폼을 통해 협업이 필요합니다



# Summary

이것만 기억해주시기 바랍니다



# OT 시장 패러다임의 변화

If I need a specific function,  
I buy a **box** that provides it

# NEW OT 시장 패러다임의 변화

If I need a specific function,  
I buy **software** that provides it  
(that runs virtualization/containerized on general purpose hardware)

# 요약

- ✓ 시스코는 15년 이상 필드에서 검증된 Cisco Validated Design(CVD) 설계 방법론을 통해 고객사에 맞는 맞춤형 아키텍처를 제시합니다
- ✓ 시스코는 소프트웨어 기반 인프라로 전환하기 위해 필요한 IT/OT 컨버전스 표준화 설계 및 글로벌 구축 사례를 보유한 유일한 공급사입니다
- ✓ IT/OT 컨버전스가 실현되기 위해서는 1) 통합 설계모델, 2) 통합 운영기술, 3) IT/OT 조직의 협업이 필요합니다

시스코의 IT/OT 설계 컨설팅 필요한 경우 “설문지”에 요청해주세요



The bridge to possible

감사합니다.

CISCO *Connect*

#CiscoConnect

# 기업의 IT 챌린지를 해결할 PURE의 핵심 데이터 솔루션 전략

강신우 이사

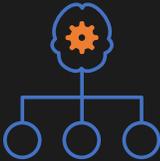
Sr. System Engineer | PureStorage Korea

# 2023년 IT 핫 키워드

## 키워드



ESG, Sustainability



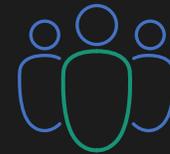
Chat GPT



무중단 인프라



## 현업의 과제



데이터 센터의 상면/전력/냉각을 어떻게 줄일 수 있을까요?



우리 회사에 맞는 최적의 AI 인프라는 어떻게 설계해야 하나요?



서비스 다운타임 발생  
= 퇴근은 없다, 주말도 없다

# 2023년 IT 핫 키워드

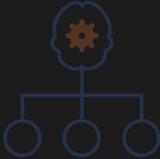
키워드



ESG, Sustainability



기후 변화



Chat GPT



무중단 인프라

# 2023년 IT 핫 키워드

## 키워드



ESG, Sustainability



Chat GPT



무중단 인프라



# 2023년 IT 핫 키워드

## 키워드



ESG, Sustainability



Chat GPT



무중단 인프라



## 데이터센터

- 탄소 배출
- 비용 관리

# 2%

데이터센터는 전세계 에너지의 2%를 사용하며, 2025년까지 최대 20%에 이를 것으로 예상됩니다.<sup>1</sup>

## 올해 수도권 전력공급신청 중 65%는 '데이터센터'... '전력난' 우려

발행일: 2023-05-24 16:00 | 지면: 2023-05-25 | 1면

<표>서울·인천·경기 전기사용예정통지에 따른 전력공급방안 검토결과 현황 (단위: 건)

구분	공급가능	공급불가	검토중	합계
수량	18	96	437	551

구분	데이터센터	데이터센터 외	합계
수량	358	193	551

자료: 한국전력공사, 홍경민 의원실

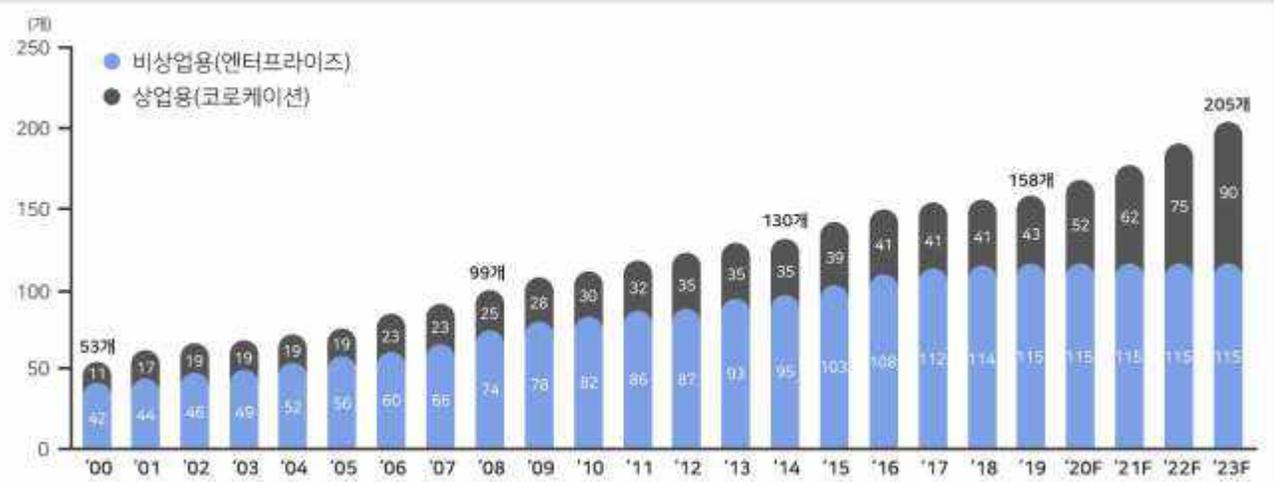
올해 수도권에서 전력공급을 요청한 사용자 중 65%가 데이터센터인 것으로 나타났다. 한국전력공사는 올해 수도권에서 접수받은 전력공급 신청 551건 중 96건에 대해서는 이미 전력공급이 어렵다고 통보했다. 가뜰이나 전력이 부족한 수도권에 데이터센터까지 몰리면서 계동포화는 물론 양질의 일자리 창출도 어렵게 만들고 있다는 지적이다.

경제 > 전자/IT

## 온실가스 배출 주범 지목된 '데이터센터'...탄소중립 실현 가능할까

차 박주환 기자 | 승인 2021.07.31 11:54 | 댓글 0

국내 데이터센터 개수



# 2023년 IT 핫 키워드

## 키워드



ESG, Sustainability



Chat GPT



무중단 인프라



## 데이터센터

상면 비용: 1RU 당 월 30,000원

전력 비용: 한전 1kWh 당 70원

냉각 비용: 3412 BTU/hr 당 월 79,000원

250 PiB를 5년간 운영 하는 비용

= "150억"

인프라 규모에 따른 천문학적 운영 비용 발생

# 지속가능한 IT를 목표로 제작된 제품

## 1. 고집적 단일 디스크 용량

24TB

48TB

75TB

300TB

2026년도 로드맵

## 2. 업계 최고 데이터 절감율 보장

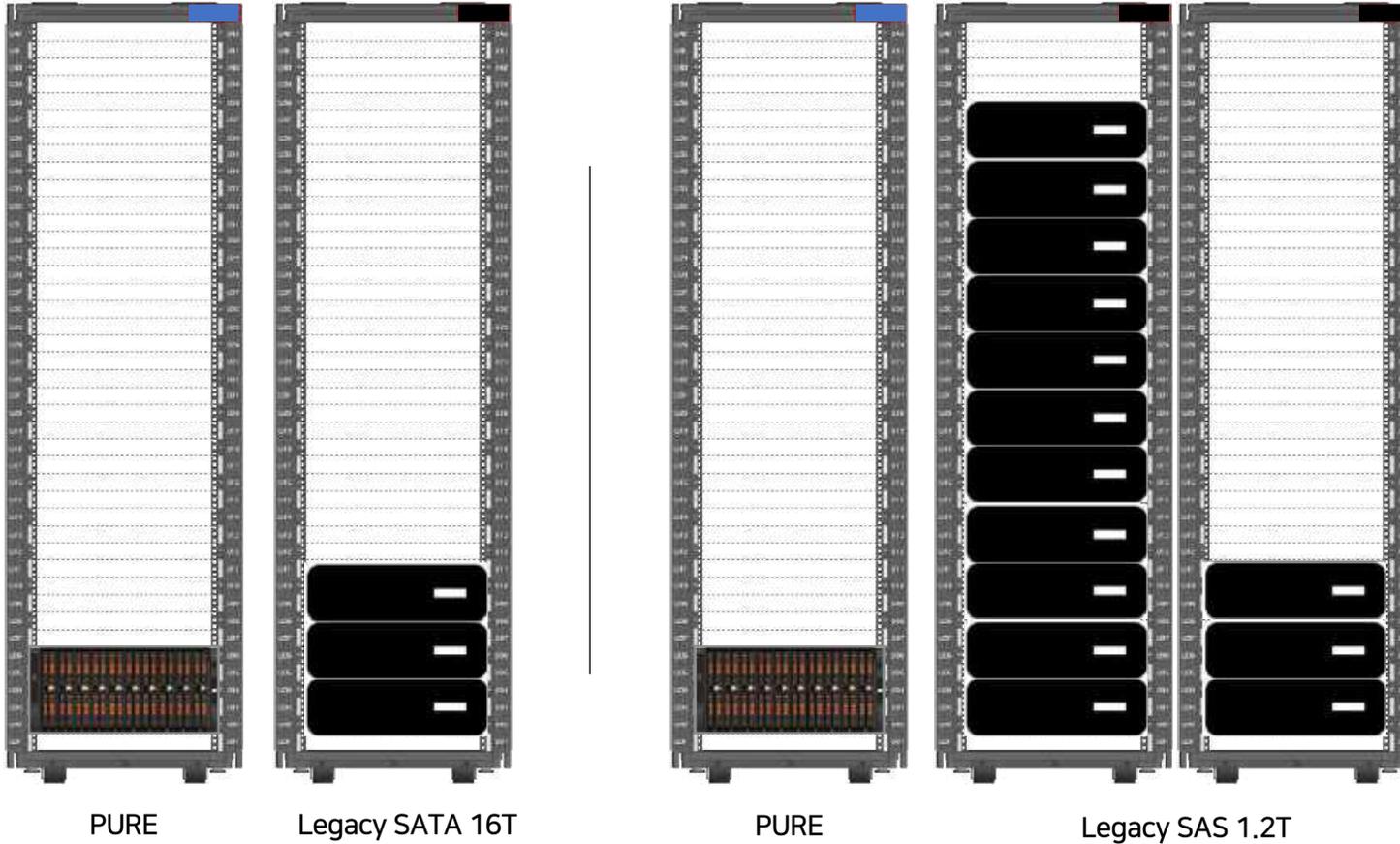


- 15.9:1 실 절감율
- VDI OS 영역
- 일관된 성능



# 퓨어스토리지의 상면 절감 효과

1PB 기준 랙상면 구성 예시이며, 디스크 구성과 비교하면 10배 이상 상면 공간을 절약할 수 있습니다.



	PURE	SATA (16T)	SAS (1.2T)
데이터 절감율	2:1	1:1	1:1
랙 상면	5U	10U (2배)	48U (9.6배)
전력량 (Peak)	3,160 W	4,630 W (1.5배)	21,288 W (대비 6.7배)
발열량 (Peak)	11,314 BTU/hr	17,672 BTU/hr (1.6배)	73,072 BTU/hr (6.4배)

# 실 고객 사례

유럽 금융 고객사에서 퓨어스토리지를 통해 기존 약 1,200개의 랙을 단 75개 랙으로 운영하고 있습니다.



## < Customer Case Study >

- 유럽 최대의 투자 은행 사례
- 기존 스토리지 설치 공간을 1,200개 → 75개 랙으로 절감 (약 94% 절감)
- 상면 절감 → 에너지 절감으로 직결
- 해당 고객사는 UN에서 발표한 Net-Zero Banking Alliance 를 달성에 집중

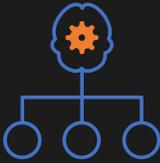
<https://blog.purestorage.com/perspectives/leading-investment-bank-enhances-digital-services-with-modern-storage/>

# 2023년 IT 핫 키워드

## 키워드



ESG, Sustainability



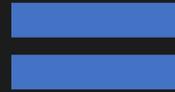
Chat GPT



무중단 인프라

## AI 인프라 구축

- 컴퓨트 → GPU
- 네트워크 → IB or Eth
- 스토리지 → PURE!!!



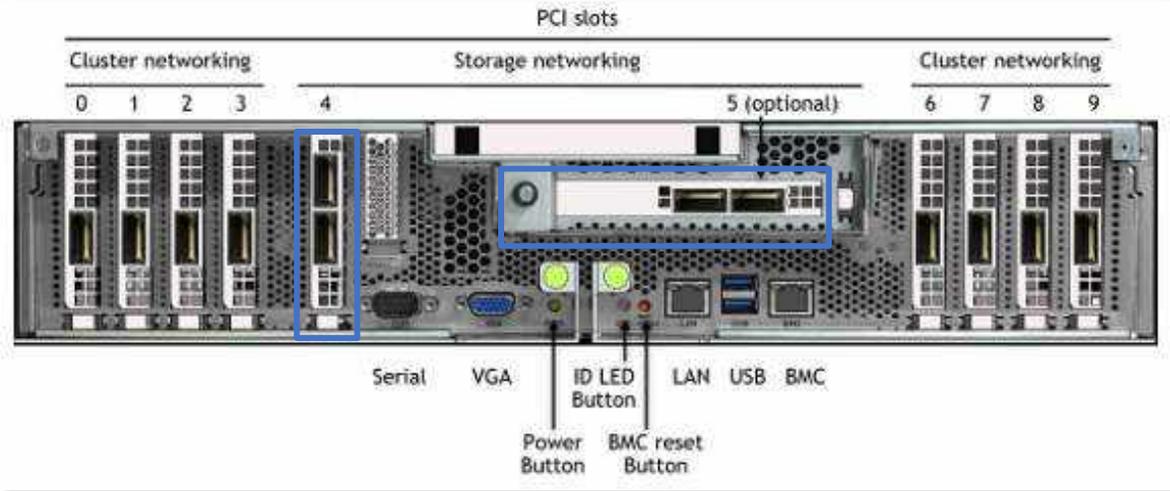
# Infiniband vs Open Ethernet?

인피니밴드가 최고입니다만?!



# Infiniband vs Open Ethernet?

1. NVIDIA 권고가 IB 입니다?! → 스토리지 연결을 위한 기본 권고 구성은 이더넷입니다.



DGX A100 포트 구성 및 데이터 시트 가이드

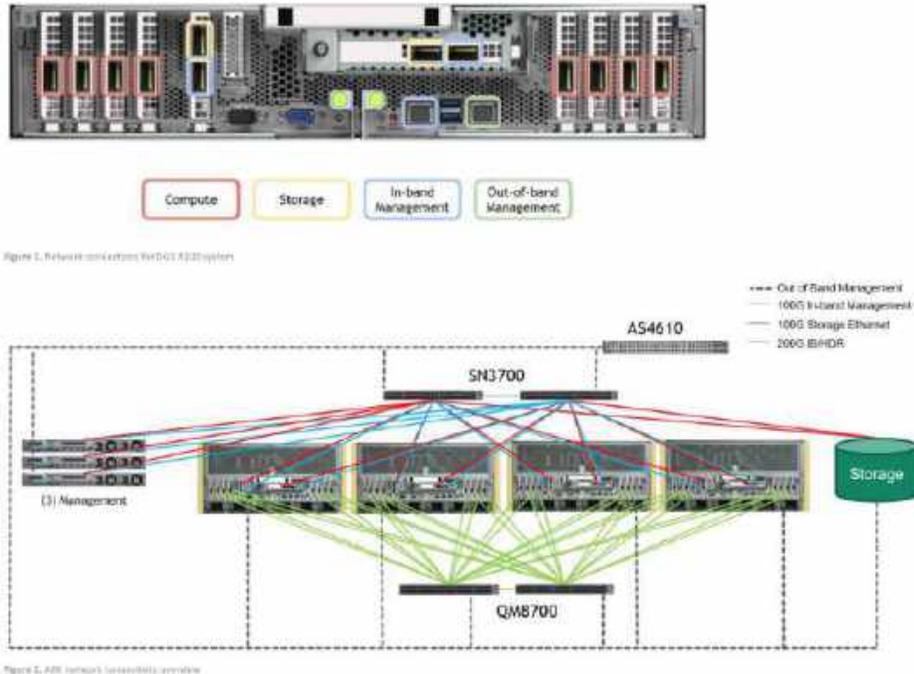
## Unmatched Data Center Scalability with Mellanox

With the fastest I/O architecture of any DGX system, NVIDIA DGX A100 is the foundational building block for large AI clusters like NVIDIA DGX SuperPOD™, the enterprise blueprint for scalable AI infrastructure. DGX A100 features eight single-port Mellanox ConnectX-6 VPI HDR InfiniBand adapters for clustering and 1 dual-port ConnectX-6 VPI Ethernet adapter for storage

- NVIDIA NVLink 구성을 위한 200Gb Infiniband 구성
- DGX와 Storage 연결을 위한 100G Ethernet 구성

# Infiniband vs Open Ethernet?

2. 컴퓨팅과 스토리지를 단일 네트워크로 구성?! → 개별 네트워크 구성으로 성능 보장 및 장애 도메인 분리

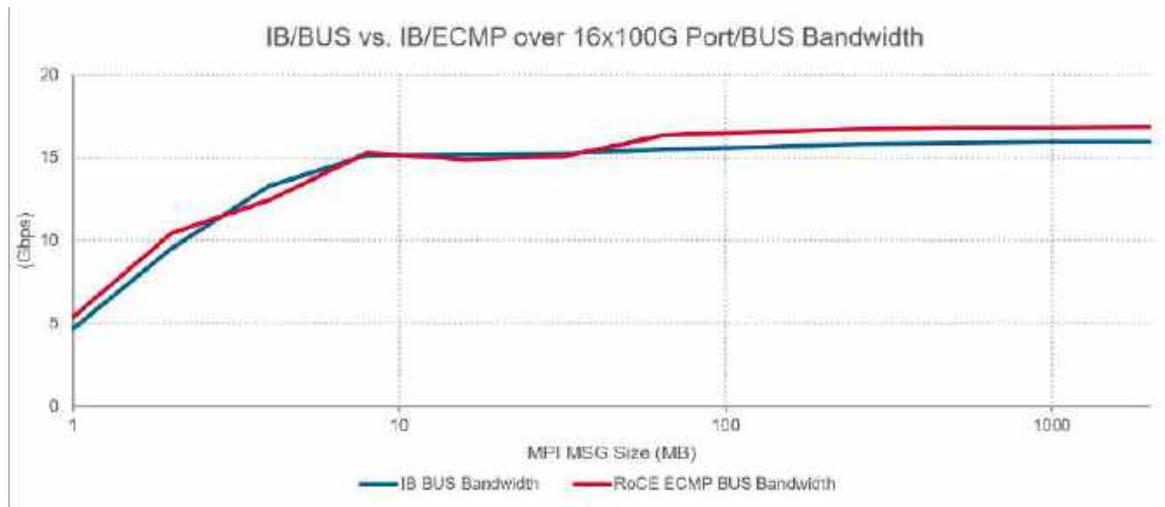


- Disaggregated 구성
- 스토리지 연결과 노드간 통신 네트워크 분리
- 네트워크 장애 시 영향도 최소화

<https://www.broadcom.com/blog/deploying-ai-ml-training-clusters-with-ip-ethernet>

# Infiniband vs Open Ethernet?

3. IB가 성능이 훨씬 빠릅니까?! → RoCE vs IB 성능은 동등한 수준<sup>1)</sup>



- 고성능 AI/ML Training 클러스터에 적합
- 고대역폭 / 종단간 혼잡제어 / 로드 밸런싱 / 패브릭 관리
- 비용 효율적 구성 및 오픈 하드웨어 기반 구성
- 분산 학습 모델에 따라 적합한 프로토콜 선택 필요
- NFS 기반의 오픈 프로토콜을 고성능으로 적용
- 스토리지 병목을 제거할 수 있는 아키텍처 고려 필요

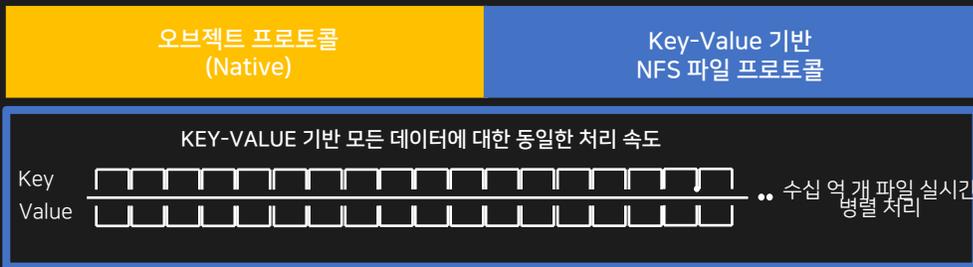
NCCL 테스트 처리량 비교: 이더넷 vs IB

<https://www.broadcom.com/blog/deploying-ai-ml-training-clusters-with-ip-ethernet>

# FlashBlade//S – Key Value 기반 고성능 설계

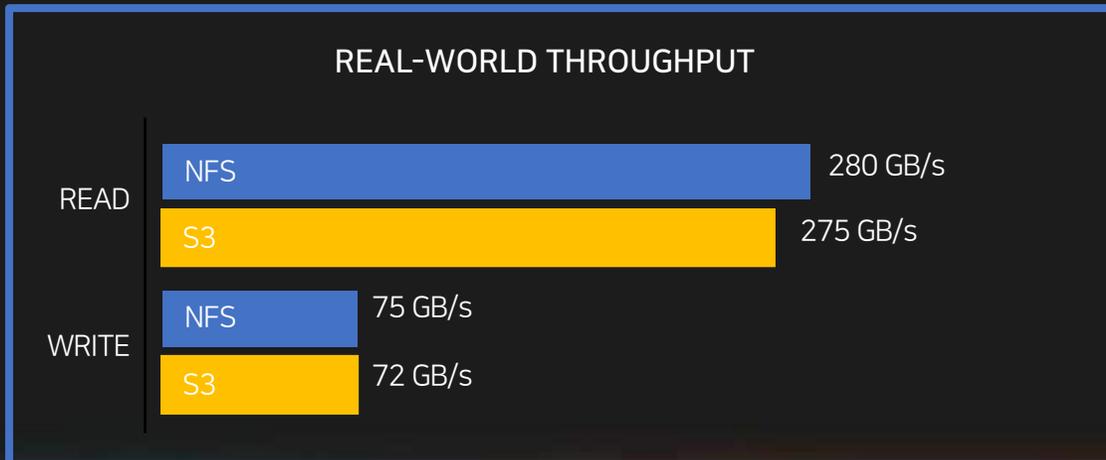
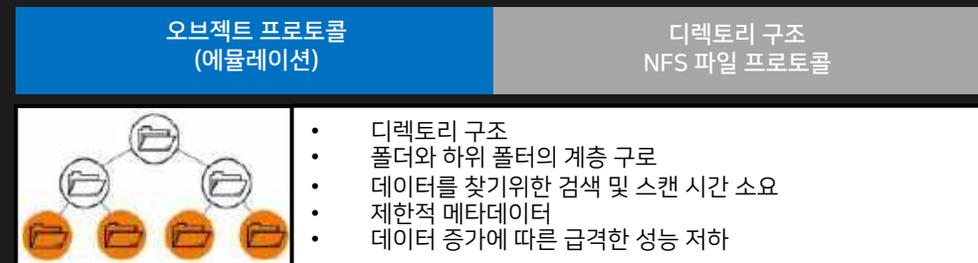
디렉토리 기반 재래식 파일에서의 고질적인 메타데이터 처리 문제 해결을 통한 획기적인 성능 개선

## FlashBlade UFFO 아키텍처



VS

## 재래식 파일 아키텍처



## FlashBlade Unified Fast File & Fast Object

- 네이티브 멀티 프로토콜
- 일관된 고성능
- 풀 샷시 구성 시 최대 280 GB/s 읽기 성능
- 풀 샷시 구성 시 최대 75 GB/s 쓰기 성능
- 블레이드 확장에 따른 선형적 성능 향상

3:1 Compressible data 기준

# 2023년 IT 핫 키워드

## 키워드



ESG, Sustainability



Chat GPT



무중단 인프라

## 시스템 장애



- 물리적 장애

- 논리적 장애

# 데이터 관점에서의 장애 유형

## 물리적 장애 [Physical Failure]

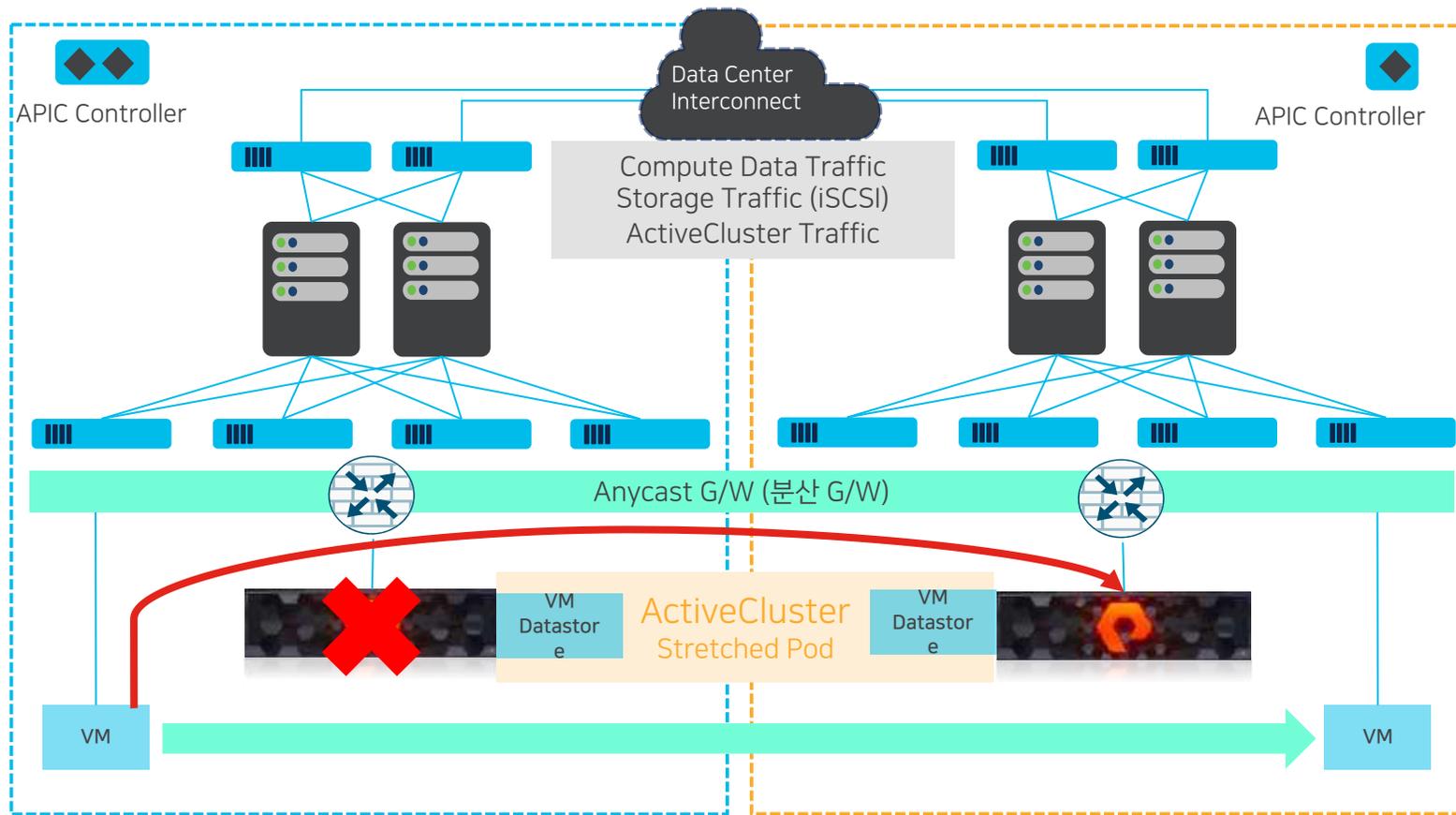
: 하드웨어 또는 전원, 데이터센터와 같은 물리적인 장비의 문제로 발생하는 장애

- 자연재해 - 지진, 폭우, 지진, 화재
- 하드웨어 노후화로 인한 실패
- 전기 / 네트워크 / 스토리지 장애

→ 하드웨어 이중화를 통한 무중단 운영

# Cisco Validated Designs (CVD) with Pure Storage

다양한 구성의 이미 검증된 CVD(Cisco Validated Design) 문서 제공



Multi-Pod 구성으로 L2 Active-Active 구성

Pure Storage Active Cluster 구성

Site 장애 발생시...

- 1) Site A에 Storage 장애
- 2) Site A에 VM은 Site B의 스토리지 통신 가능
- 3) Site를 넘어야 하기 때문에 성능 저하는 발생
- 4) VM을 Site B로 vmotion하여 해소

네트워크 Side : Active - Active 통신  
스토리지 Side : ActiveCluster로 실시간 복제로  
A-A DC 구현

# 데이터 관점에서의 장애 유형

## 논리적 장애 [Logical Failure]

: 애플리케이션 또는 운영체제 상의  
데이터 삭제 및 손상으로 발생하는 장애

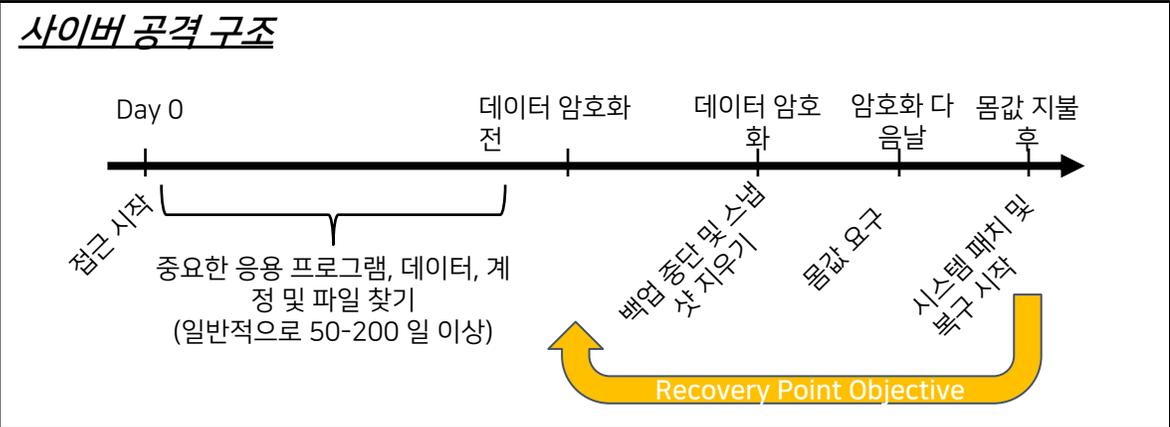
- Human Fault
- 운영자 / DBA 에 의한 작업 실수
- 사이버 공격으로 인한 데이터 손실

→ 백업을 통한 데이터 복구

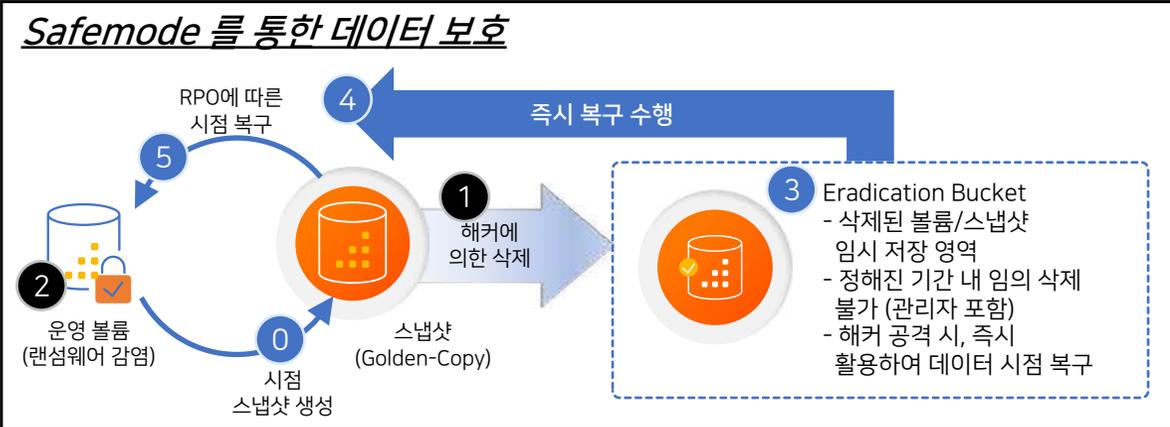


# 스냅샷을 통한 논리 장애 복구 및 랜섬웨어 보호

## 사이버 공격 구조 및 SafeMode 를 통한 복구 방안



- Safemode 기능
- 임의 삭제된 볼륨에 대한 임시 보호 기능
- 별도 휴지통 공간에서 정해진 타이머 기간동안 삭제 불가
- 사이버 공격 / 사용자 실수에 대한 보호 기능 활용
- 퓨어 서포트를 통해 PIN-CODE 확인 후 임의 삭제 가능



### 실제 인도의 MSP 기업 침해 및 Safemode 활성화 사례

**Status :** Ransomware attack occurred on Jun 14th.

2022-06-14 02:59:54 | 1949506 | customer | pureuser | purepgroup destroy DARCL

2022-06-14 03:00:00 | 1949507 | customer | pureuser | purepgroup destroy JS-Furnishings

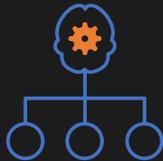
2022-06-14 03:00:06 | 1949512 | customer | pureuser | purepgroup destroy RISHI

# 2023년 IT 핫 키워드 정리 및 맺음말

## 키워드



ESG, Sustainability



Chat GPT



무중단 인프라



## Pure's 솔루션



- 업계 최고의 데이터 절감
- 업계 최고의 고집적 SSD



- 이더넷을 통한 오픈 AI
- Key-Value 기반 초고속 성능



- CVD 기반 ActiveCluster
- 랜섬웨어 보호 및 즉시 복구

# 감사합니다.

Contact US

홈페이지 : [www.purestorage.com/kr](http://www.purestorage.com/kr)

페이스북 : [www.facebook.com/purestoragekorea](http://www.facebook.com/purestoragekorea)

블로그 : [www.blog.naver.com/purestorage\\_korea](http://www.blog.naver.com/purestorage_korea)

유튜브 : [www.youtube.com/@PureStoragekr](http://www.youtube.com/@PureStoragekr)