NetApp

AI 환경을 위한 End-to-End Data Transference

신 현 호 부장
NetApp Korea, Global Solutions Engineer

© 2019 NetApp, Inc. All rights reserved.

Agenda

I. AI로 부터의 변화

II. Edge - Core - Cloud Journey 그리고 Data Transference

III. Sample Architecture

© 2019 NetApp, Inc. All rights reserved.

AI로 부터의 변화

© 2019 NetApp, Inc. All rights reserved.
NetApp

AI로 인해 세상은 어떻게 바뀌고 있을까?

AI는 거의 모든 산업군에서 강한 영향력을 미치는 세력이 되고 있는 중

40%

디지털 전환(DX) 계획의 40%가 올해 안에 AI 서비스를 사용

50%

엔터프라이즈 인프라의 50%는 2021년까지 AI를 적용 예정

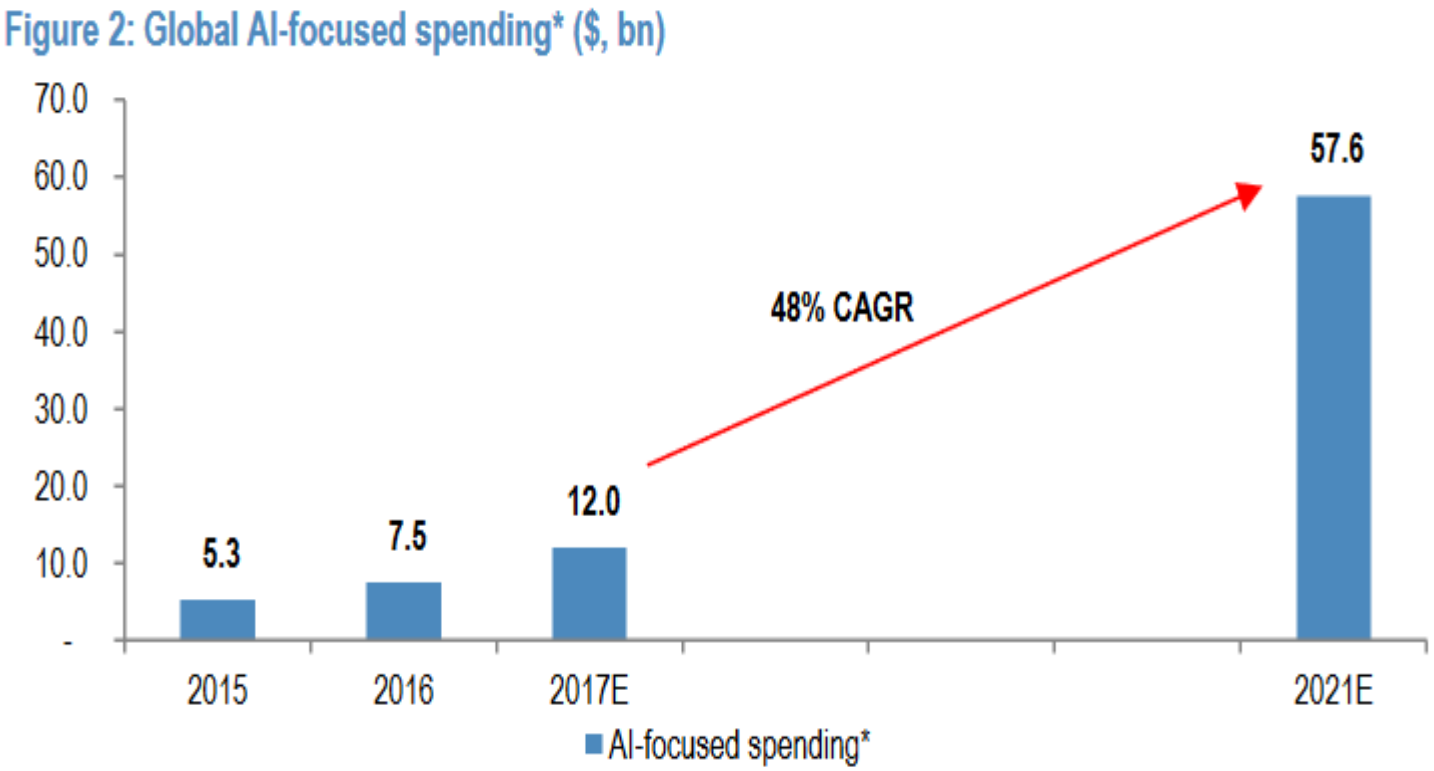
87%

글로벌 비즈니스 리더의 87%는 AI가 3년 안에 고객에게 더 나은 경험을 제공할 것이라고 예상

Source: IDC Storage Workloads, 2018. AI Business, 2018.

© 2019 NetApp, Inc. All rights reserved.
NetApp

Global AI Focused Spending

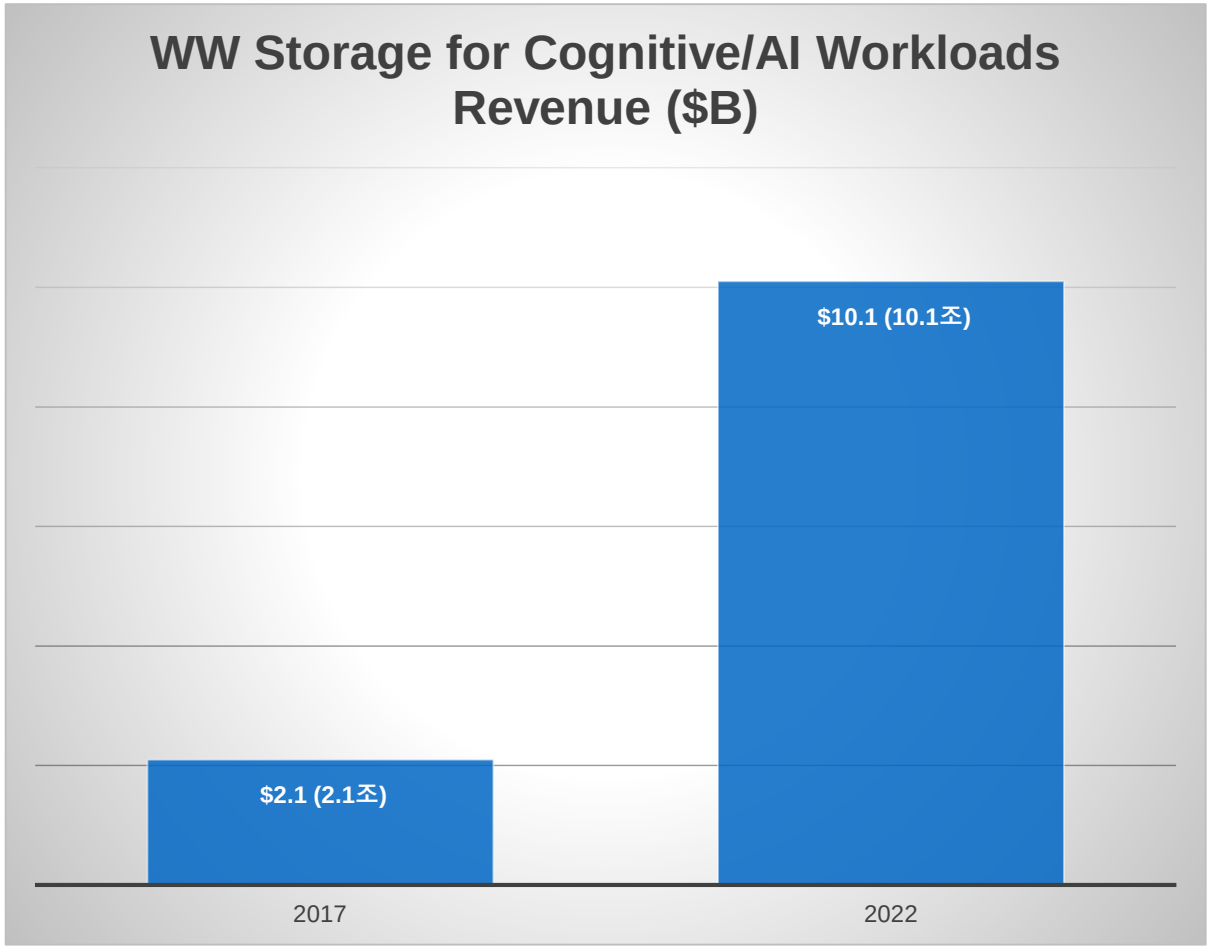


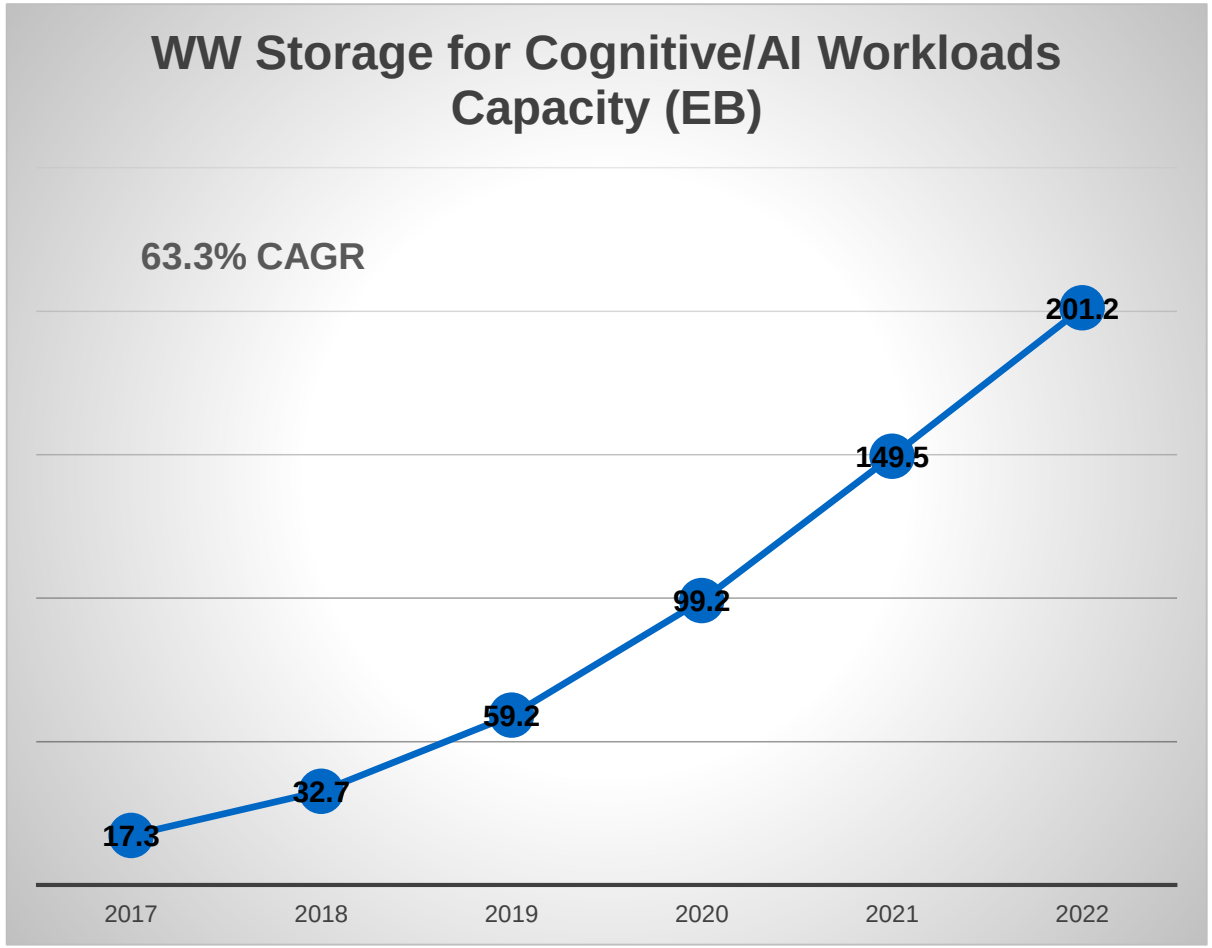
Source: AI-spending estimates from IDC. *Includes AI-focused spending on hardware, software (applications + software platforms), and services (IT consulting & system implementation).

© 2019 NetApp, Inc. All rights reserved.

Storage Revenue and Capacity

Cognitive/AI workloads





Source: IDC WW Storage for Cognitive/AI Workloads Forecast, 2017-2022

© 2019 NetApp, Inc. All rights reserved.

NetApp

세계적으로 급속한 AI 적용

300,000 10년 동안 자동차 사고로 부터 구한 생명

4 Billion 2019년까지 음성 지원 기능을 탑재한 AI device 수

$59.8 Billion 2025년까지 AI software 시장의 성장

1 Billion 2020까지 AI 기능을 탑재한 camera 수

$10 Billion AI를 위한 자본금

$190% 5년간 AI 관련 특허 증가

71% 제조분야의 자동화 전환 잠재력

Fourth 베를린에서 AI startup 수

© 2019 NetApp, Inc. All rights reserved.
NetApp

AI가 제공하는 커다란 이점

경쟁력 유지

- 고객 환경 혁신
- 필요에 따른 예측

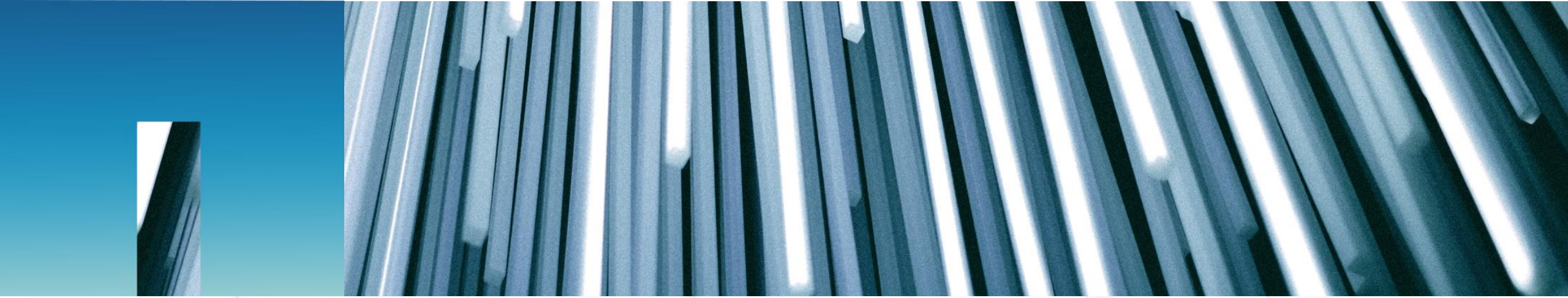
성장의 도모

- 높은 value의 제품과 서비스 혁신
- 새로운 시장 기회 창출

경비 절감

- Process error 감소
- 안전성 향상

© 2019 NetApp, Inc. All rights reserved.
NetApp



AI는 Data에 대한 모든 것

Data is distributed

multiple cloud와 on-premises로 부터 생성되고 사용

Data is dynamic

끊임없이 변화하고 점점 더 클라우드로 이어지는 중

Data is diverse

다양한 형태로부터 수집 : video, audio, images, quantitative, logs etc.

© 2019 NetApp, Inc. All rights reserved

NetApp

Data는 AI에 있어 매우 중요합니다.
하지만 상당한 도전 과제들을 필요로 합니다

(Source IDG research)

51%
Data silos

37%
Technology complexity

35%
Data access

35%
Data preparation

© 2019 NetApp, Inc. All rights reserved
NetApp

Edge - Core - Cloud Journey 그리고 Data Transference

© 2019 NetApp, Inc. All rights reserved.
NetApp

Cloud and Edge Connectivity?

Edge

Core

Cloud

aws

Microsoft Azure

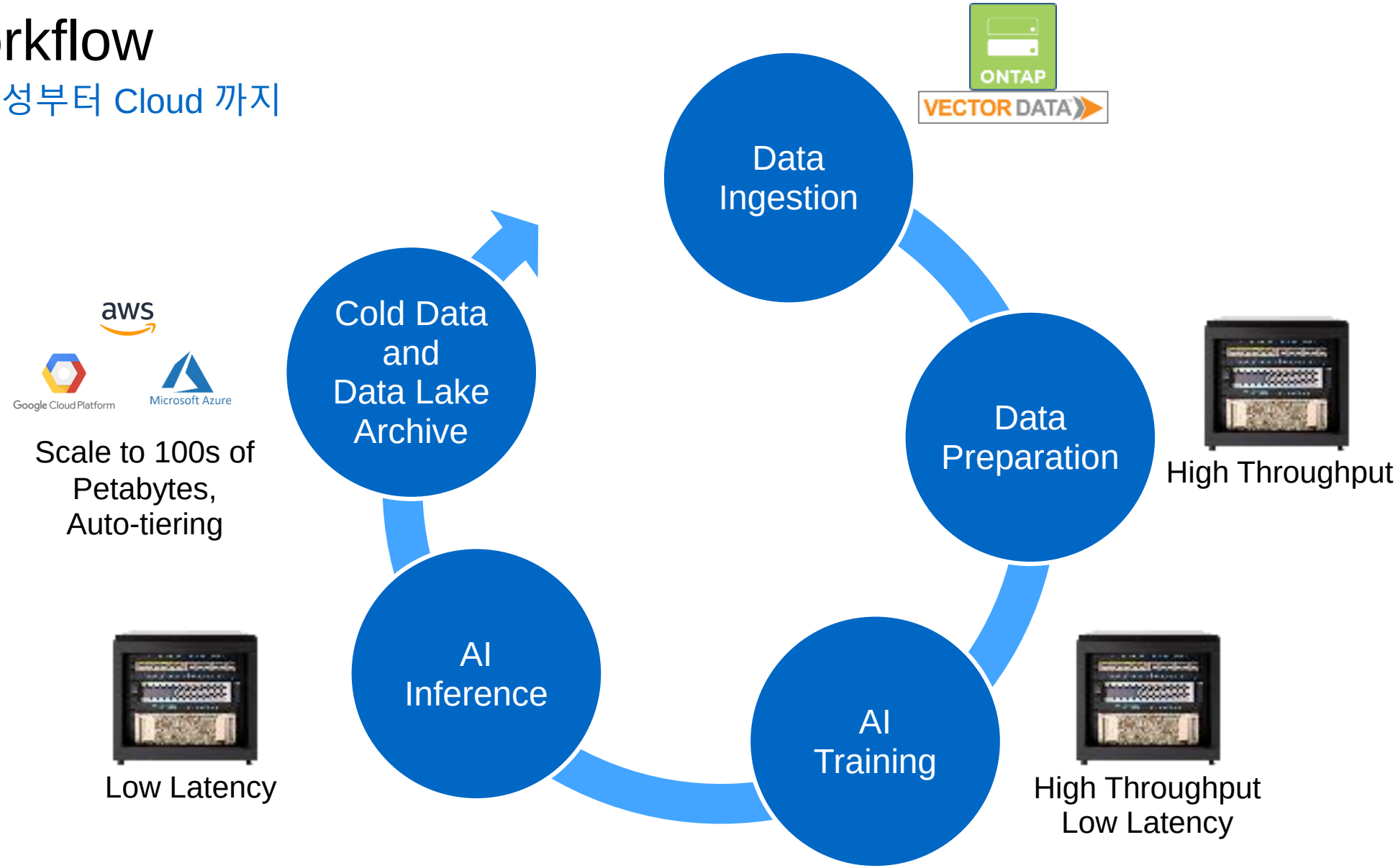
Google Cloud Platform

© 2019 NetApp, Inc. All rights reserved.

NetApp

AI Workflow

Data의 생성부터 Cloud 까지



© 2019 NetApp, Inc. All rights reserved.
NetApp

NetApp provides solutions for every stage of the data pipeline

Edge

Software Defined Storage

Core

Reference architectures

NetApp

NVIDIA

CISCO

Cloud

Microsoft Azure

aws

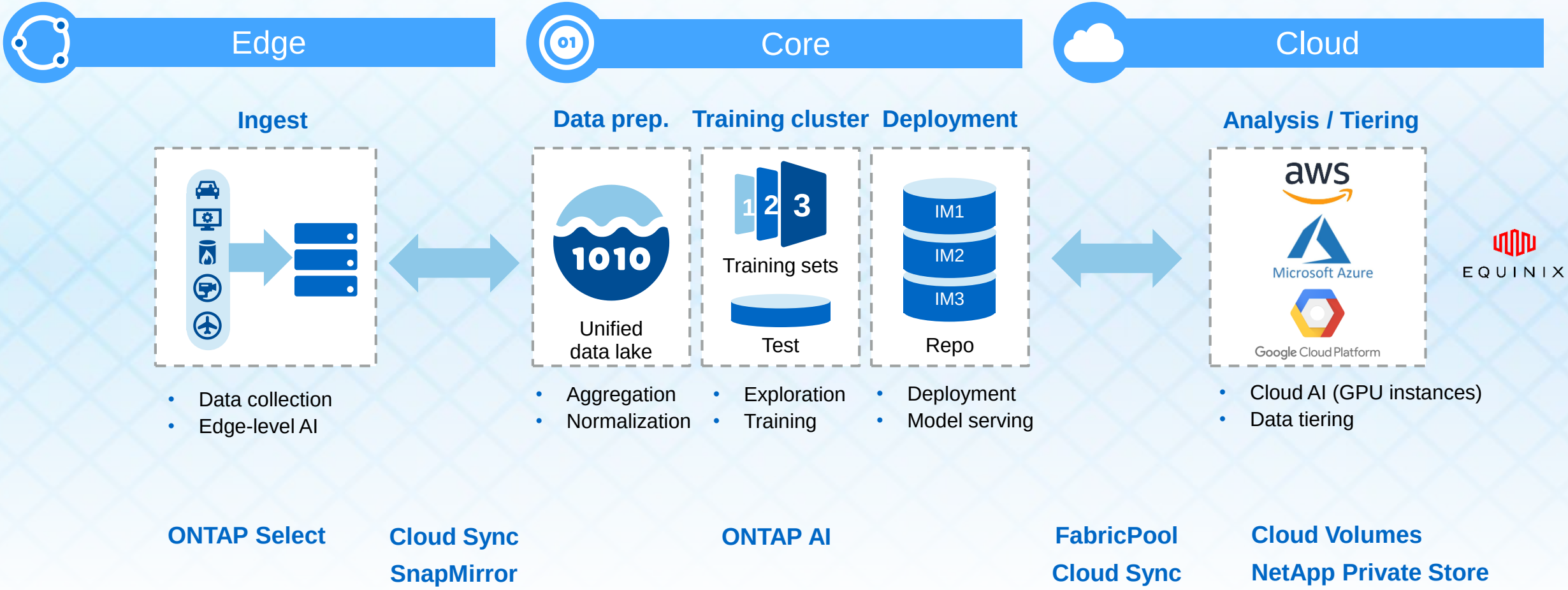
Google Cloud

© 2019 NetApp, Inc. All rights reserved

NetApp

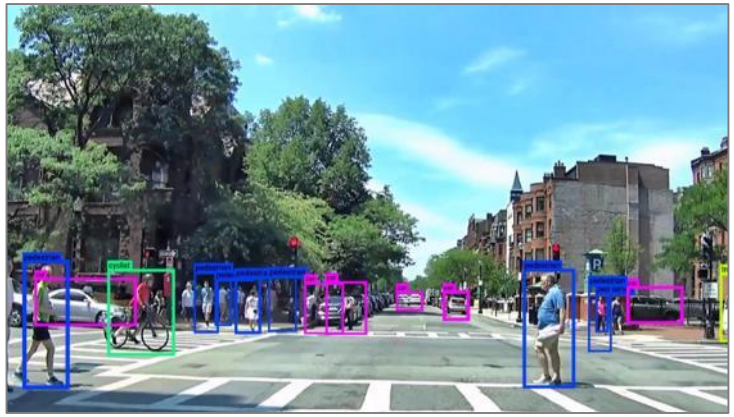
Edge to Core to Cloud

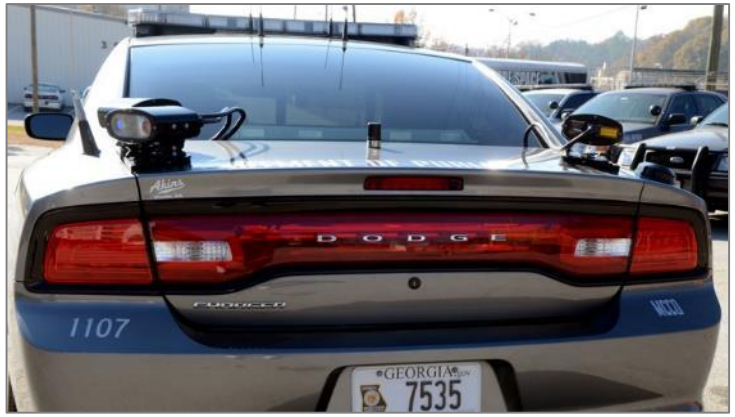
Seamless data management



© 2019 NetApp, Inc. All rights reserved.
NetApp

Edge





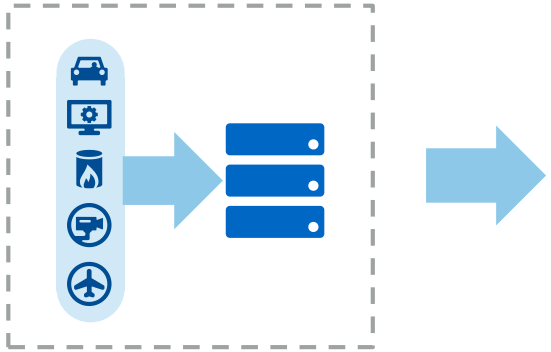
© 2019 NetApp, Inc. All rights reserved.
NetApp

Deep Learning Edge Architecture

Edge 단계에서의 'Intelligence'는 실시간 처리를 가능하게 함과 동시에 cloud로 data 전송량을 감소

Edge

Ingest



Requirements

- 수많은 source로부터의 data를 통합
- 효율적인 data storage
- 효율적인 복제
- Open data ingestion

NetApp Differentiation

- Unified
- Smart Replication
- Efficient
- Scalable
- Upgradeable

AFF

9 ONTAP

ONTAP

© 2019 NetApp, Inc. All rights reserved.

NetApp

Software Defined Storage for the Enterprise

ONTAP Select

USE CASES

- Edge aggregation
- Mobile data acquisition system
- On-premises Cloud edge
- Detached edge replication
- Remote offices

Enterprise-class data 절감 및 보호

유연한 설치

하이브리드 클라우드 간 중단 없는 복제 및 이동

Cloud와 유사한 신속성

© 2019 NetApp, Inc. All rights reserved.
NetApp

Vector Data

Vault Edge

- 견고한 small form factor data 관리 솔루션
- Vector Data 다수의 sensor로 부터 data를 통합
  - edge device의 계층적 표시
- ONTAP Select® (NetApp's SDS offering)
  - *Data 중복제거 및 보호*
- Storage 복제 솔루션을 이용해 WAN을 통한 data 전송
  - Source에서 destination volume으로 빠른 data 복제
  - 중앙 시스템으로 효율적인 data backup
- Applications
  - DoD, Military, Transport, Aerospace, Oil & Gas …

Specifications:

CPU:	Intel Quad-Core Processor
Memory:	32GB RAM
Storage:	RAID Controller with Battery Backup Up to 8 x 2.5" SSD
Dimensions:	10.79in x 11.38in x 6.89in
Weight:	16.4 lbs (empty), 17.5 lbs (loaded)
Temperature Ranges:	+5°C to +55°C operational; -40°C to +70°C storage
Operating shock:	40Gs; 11ms
Random vibration:	10Hz to 2000Hz
Humidity:	Up to 95% RH Non-condensing

© 2019 NetApp, Inc. All rights reserved.

ONTAP Select Validated Partners

Ruggedized Platforms

PacStar

VECTOR DATA

CUBIC | DTECH LABS

GERMANE SYSTEMS™

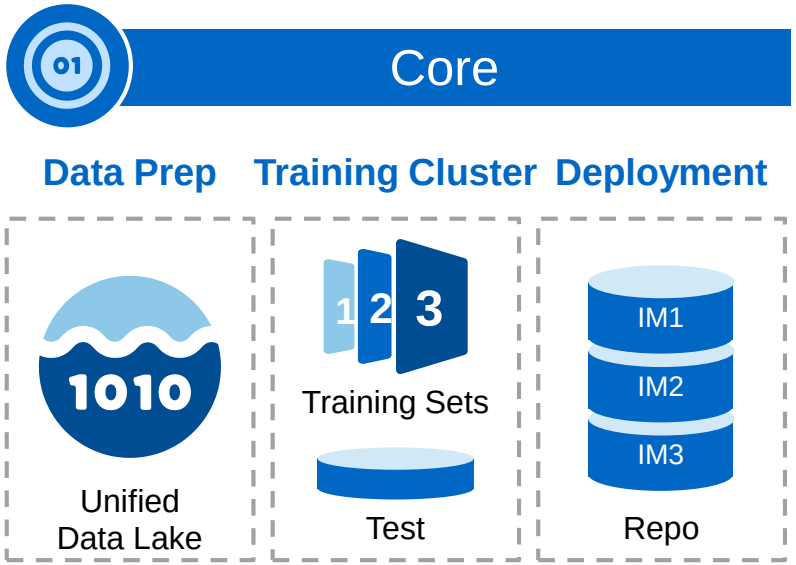
KLAS TELECOM

© 2019 NetApp, Inc. All rights reserved.

NetApp

Deep Learning Core Architecture

학습 정보를 중앙화 하고 insight를 Edge로 전달



Requirements

- 데이터 통합 및 처리 site
- Multiple data sources (NAS and SAN)
- 독립적인 확장: compute and storage
- random 및 sequential IO를 위한 높은 성능

NetApp Differentiation

- Unified data lake
- Performance
- Scalability
- Cost optimization
- Server/compute separation with CI
- Choice of where to perform core functions

AFF CI 9 ONTAP ONTAP

© 2019 NetApp, Inc. All rights reserved.
NetApp

NetApp ONTAP AI

Simplify, accelerate, and integrate your data pipeline for deep learning with NetApp and NVIDIA

Proven architecture

- NVIDIA DGX-1 supercomputer와 NetApp cloud-connected all-flash storage

Simple to deploy

- 디자인 복잡성과 추측성을 제거

Build an integrated data pipeline

- edge에서 core, cloud까지 데이터의 흐름을 유연하게 관리

Delivers performance and scalability

- 작게 시작해서 필요에 따라 확장

© 2019 NetApp, Inc. All rights reserved.

NetApp AI Family

Start small, scale big

HALF RACK

1x DGX-1
1x AFF A220
1x 5624Q

HALF RACK

2x DGX-1
1x AFF A300
2x 5624Q

HALF RACK

4x DGX-1
1x AFF A700
2x 5624Q

FULL RACK

9x DGX-1
1x AFF A800
2x 3232C

FULL RACK

3x DGX-2
1x AFF A800
2x 3232C

FLEXPOD AI

1x UCS 480 ML M5
1x UCS 5108
1x AFF A800
2x 9336C
2x 6454 FI

© 2019 NetApp, Inc. All rights reserved.
NetApp

Cloud Services for Deep Learning

서로 다른 데이터를 통합하고 다양한 데이터 서비스를 제공

Cloud

Analyze

Direct Connect

Express Route

Microsoft Azure

amazon webservices™

Google Cloud Platform

EQUINIX

Requirements

- Production 전반에 걸쳐 서로 다른 소스에서 클라우드 딥러닝 툴셋으로 정보를 통합
- active & archive data 접근
- multi-cloud에 걸친 data 이동
- Data 보호 및 보안

NetApp Differentiation

- Unified data
- Cost optimization
- Multi-cloud access from NPS
- Built in backups and archive
- Move data wherever it's needed
- In-place performance with cloud

AFF

ONTAP

NPS

Cloud Sync

Cloud Volumes

© 2019 NetApp, Inc. All rights reserved.

NetApp

All of the Biggest Clouds Choose NetApp

Data services that move data, persist data, orchestrate and integrate data

Google Cloud Platform

Azure

aws

© 2019 NetApp, Inc. All rights reserved.

NetApp

NVIDIA AI inference server using NKS(NetApp K8S Service)

NVIDIA NGC portal leveraging branded NKS for AI Inferencing

NVIDIA branded NetApp Kubernetes Service and tightly integrated it into their NGC platform for AI workloads.

This is a demo of deploying a simple TensorRT inferencing application using NKS as a backend.

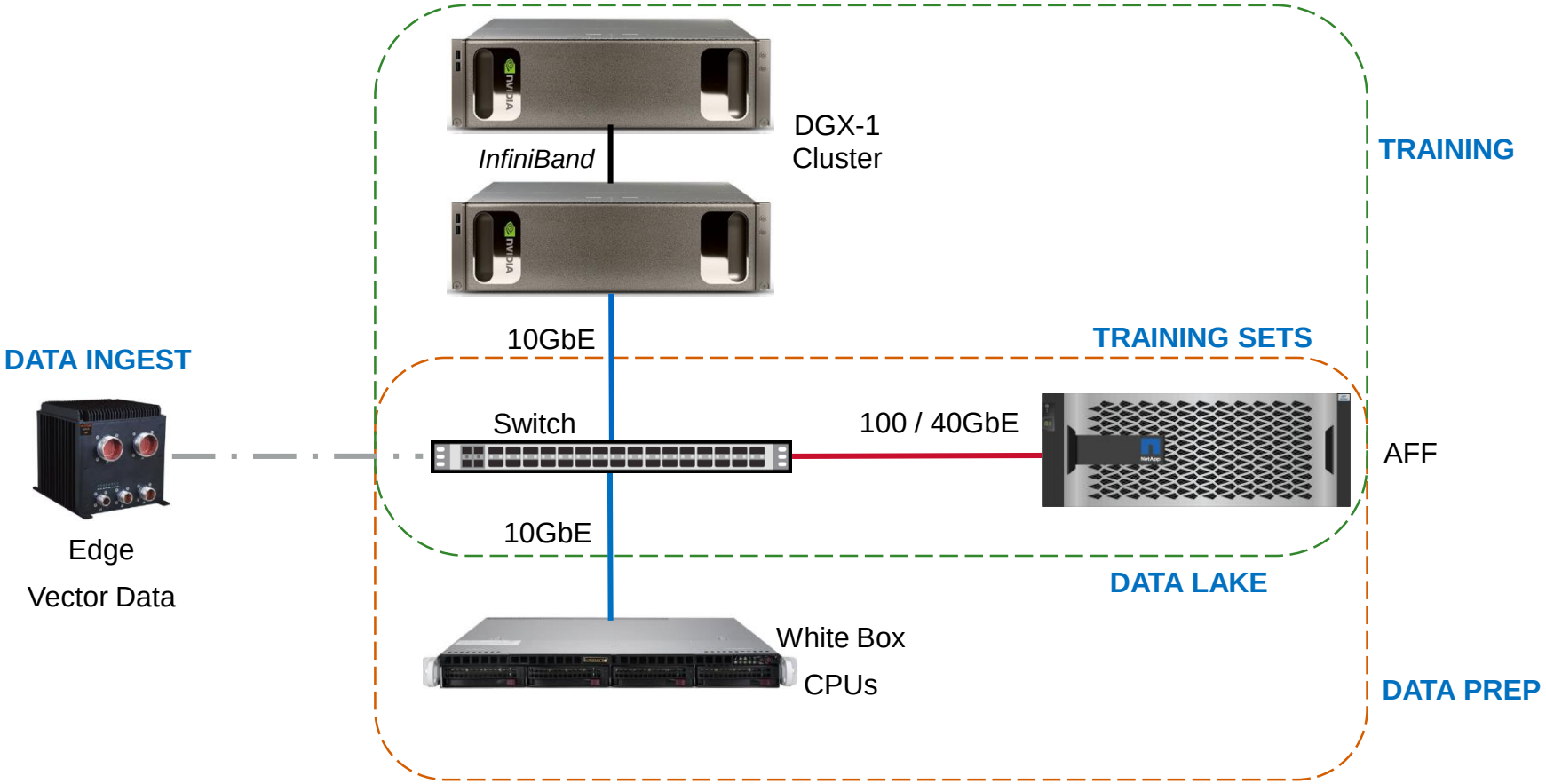
© 2019 NetApp, Inc. All rights reserved.
NetApp

Sample Architectures

© 2019 NetApp, Inc. All rights reserved.

Sample AI/DL Training Cluster

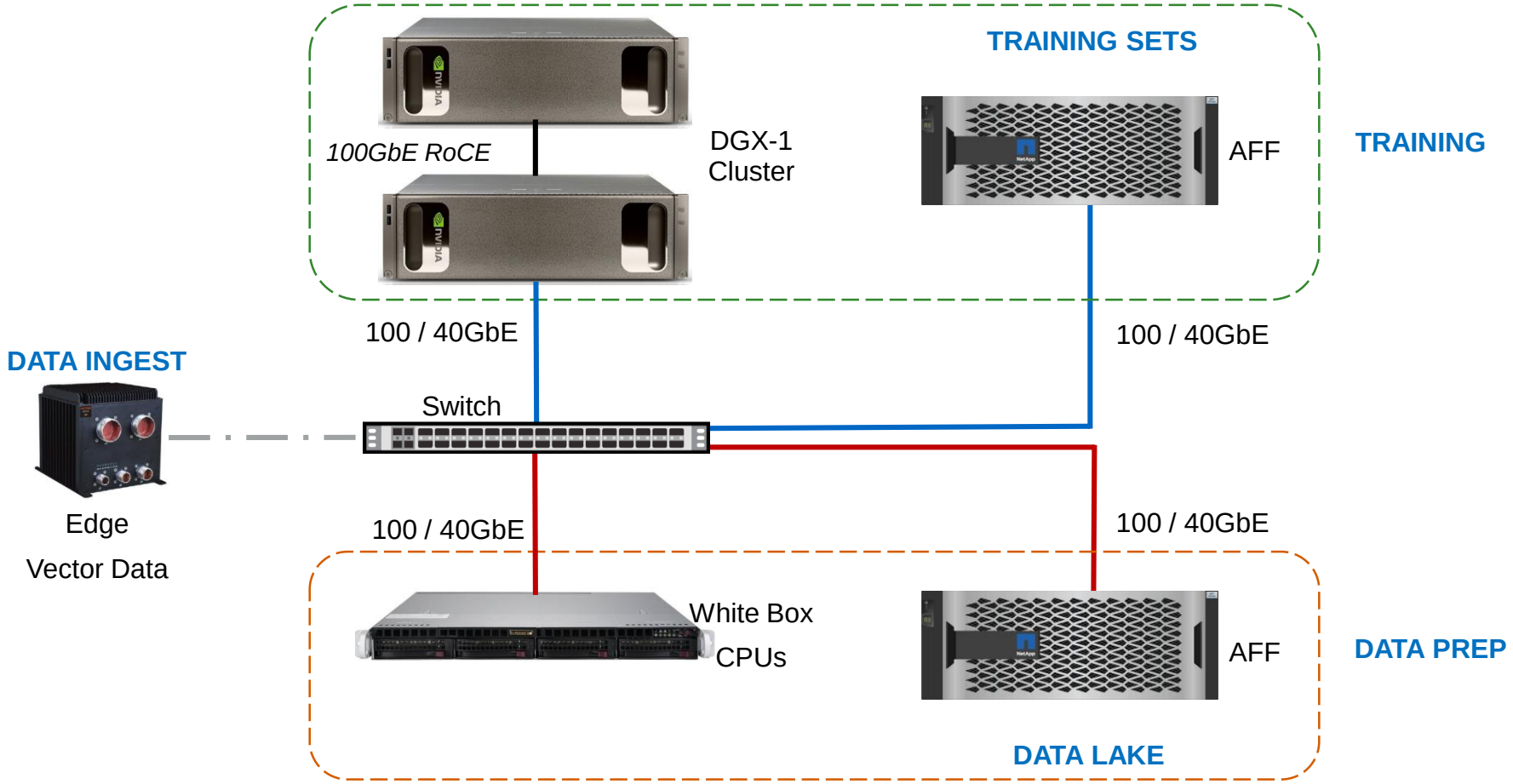
In-Place AI/DL with Data Lake



© 2019 NetApp, Inc. All rights reserved.

Sample AI/DL Training Cluster

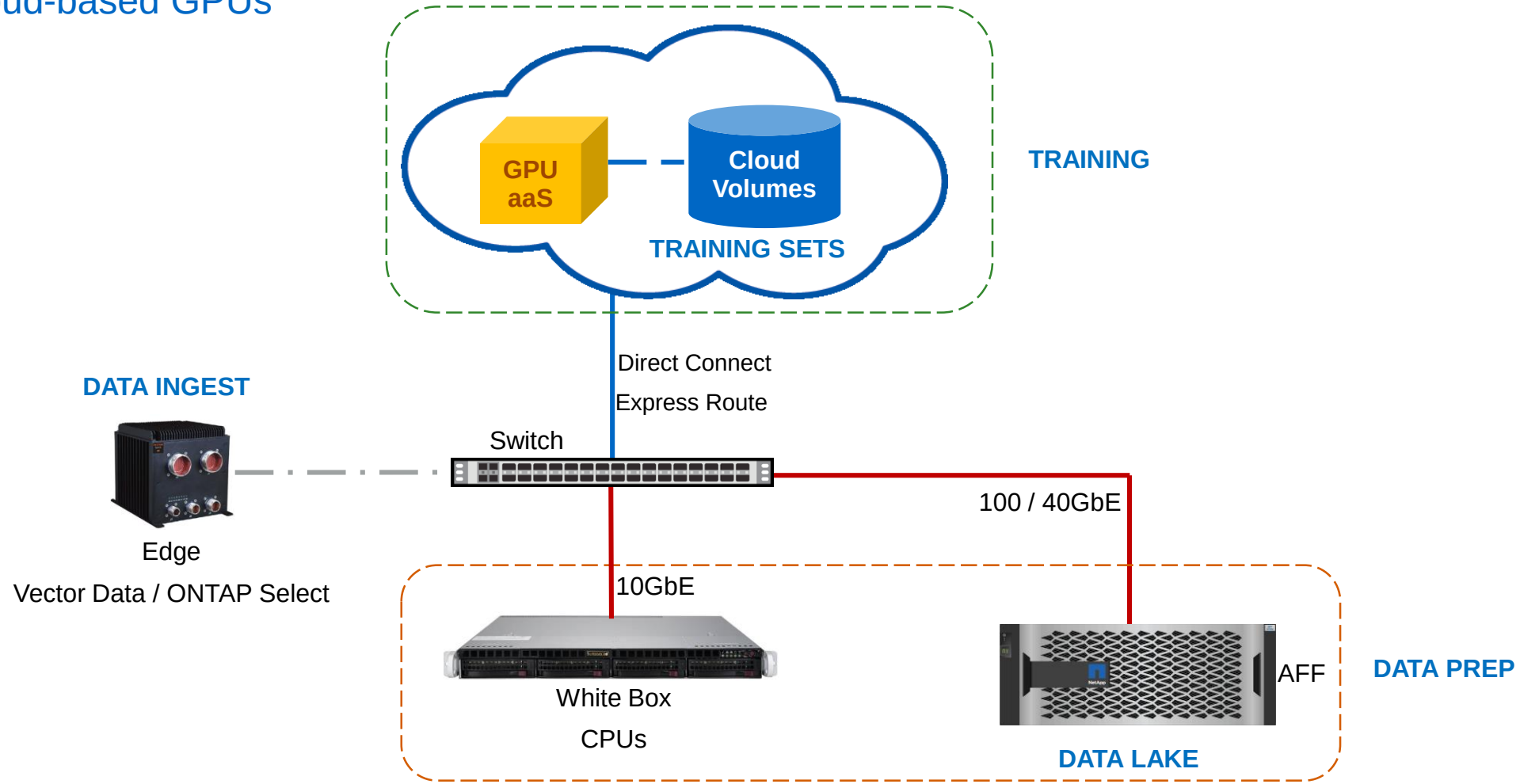
Nvidia DGX-1 100GbE RoCE Cluster, 100 / 40GbE NFS Storage



© 2019 NetApp, Inc. All rights reserved.

Sample Cloud AI/DL Training Cluster

Using Cloud-based GPUs



© 2019 NetApp, Inc. All rights reserved.

Why NetApp for AI?

Smart

- 스마트한 데이터 이동
- 데이터 파이프라인의 단순화
- 원활한 클라우드 통합
- 프로페셔널 서비스

Powerful

- 업계에서 가장 완벽한 클라우드 데이터 서비스를 제공
- Unified data platform
- Future-proof infrastructure

Trusted

- Data protection and compliance designed-in
- Trusted partner
- AI ecosystem에서 최고의 파트너쉽

© 2019 NetApp, Inc. All rights reserved
NetApp

감사합니다

© 2019 NetApp, Inc. All rights reserved
NetApp