

디지털 트랜스포메이션을 통한  
비즈니스 혁신과 4차 산업혁명 대응을 위한

# GIT 솔루션즈 데이

## AI 환경을 위한 인프라

2018년 10월 17일 / CIA 사업본부 김성태 이사

# 요즘 우리는... 더 이상,

---



## AI, 궁금증을 풀어드립니다.

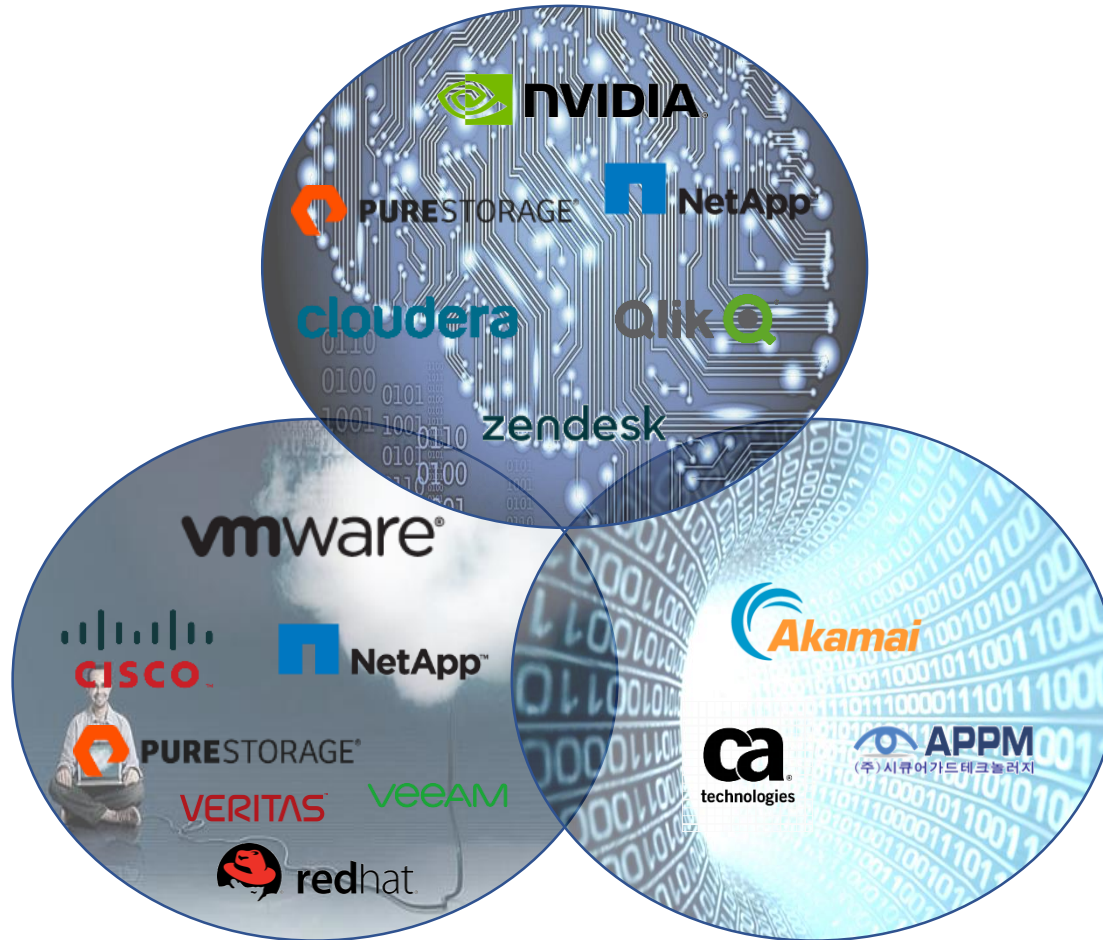


농림축산식품부

농림축산식품부

# 오늘 우리는...

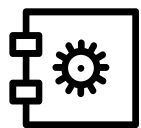
## AI (ML&DL) 및 빅데이터



클라우드 컴퓨팅

정보 보호

# 디지털 트랜스포메이션에 따른 업계 변화



금융 서비스  
18%



오일 및 가스  
15.1%



공공시설  
29%



정부  
17.5%

파괴적  
디지털 혁신으로  
수익 감소 위기에  
놓인 전통 산업

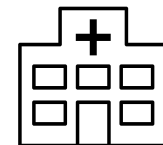
설비  
20%



유통  
>25%



의료  
10.9%



운송  
14.2%



# 데이터를 가장 안전하게 잘 쓰는 나라를 만들겠습니다!

“이제 **데이터**  
가장 잘  
다루는 나라  
돼야.”

대한민국이 **라** 된다

데이터를 가장 안전하게  
잘 쓰는 나라를 만들겠습니다!

“산업화 시대  
석유가 성장  
기반이었다면, 4차  
산업혁명시대,  
**미래 산업의  
원유가 바로  
데이터,**”

# 데이터가 불러올 엄청난 기회

비즈니스

KT그룹, 5G와 인공지능에 5년간 23조 투자

NewsRoom Exclusive 경제

삼성전자, 인공지능(AI) 분야에 총력전  
美 실리콘밸리, 英 캠브리지 대학, 美 뉴욕에 6번째 AI센터  
은행 하반기 채용 "공정성 지키자" AI 도입 등 야심차

2019년

이용할

트랜스포

진화하는 자율주행차 카메라...딥러닝으로 패턴까지 예측

머니투데이방송 권순우 기자 | 2018.09.10 11:21

AI로 암 진단하는 시대 왔다

파이낸셜뉴스 | 입력 : 2018.09.08 15:29 | 수정 : 2018.09.08 15:29

조달청, 정부물품관리에 인공지능 기술 활용

파이낸셜뉴스 | 입력 : 2018.09.27 13:49 | 수정 : 2018.09.27 13:49



# AI 플랫폼 고려사항

## 개발자 생산성



“ 지금 당장 시작해야만 해,  
현업에서는 어제 결과를  
전달하고 싶어하거든

## 확장 성능



“ 나는 가장 강력한  
GPU가 필요해

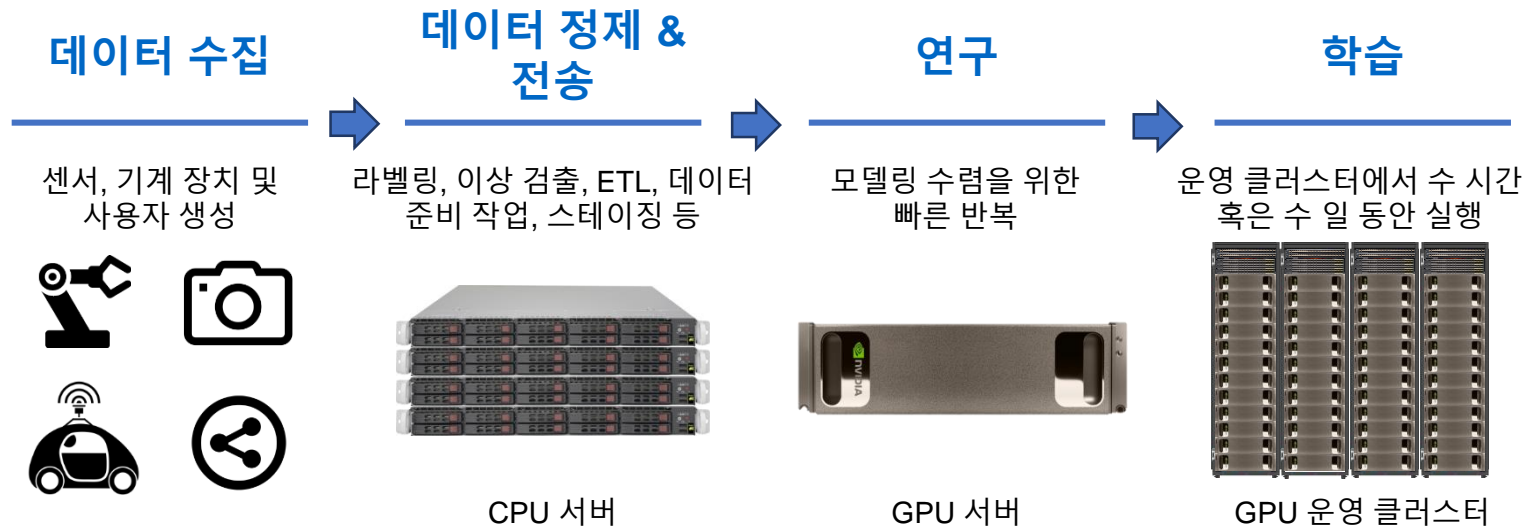
## 총소유비용



“ 예산이 제한되어 있어,  
가능하다면 가장 적은  
비용으로 가능했으면 해

# AI, ML 및 DL 환경의 워크로드 분석

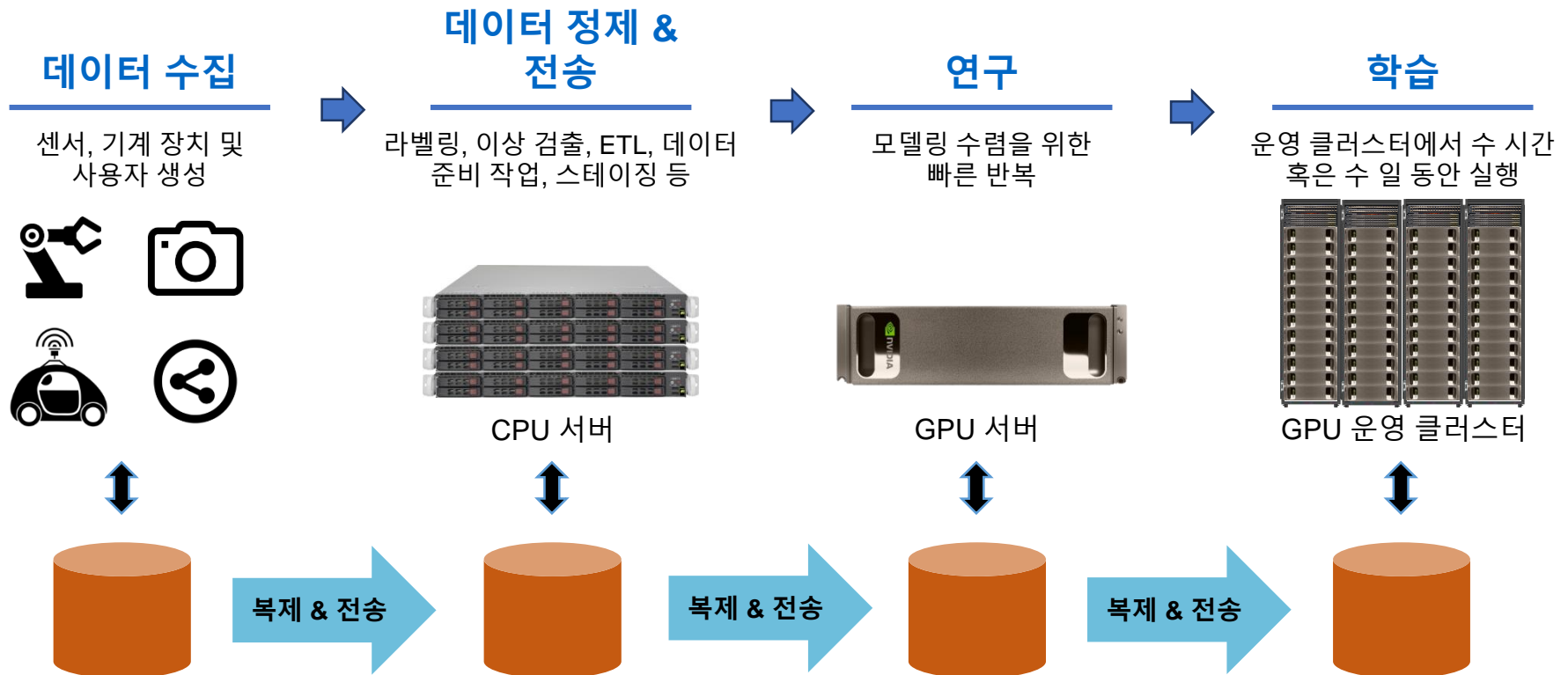
AI 인프라 환경을 위한 데이터 수집에서 학습 단계에 이르기 까지 단계별 I/O 워크로드 속성





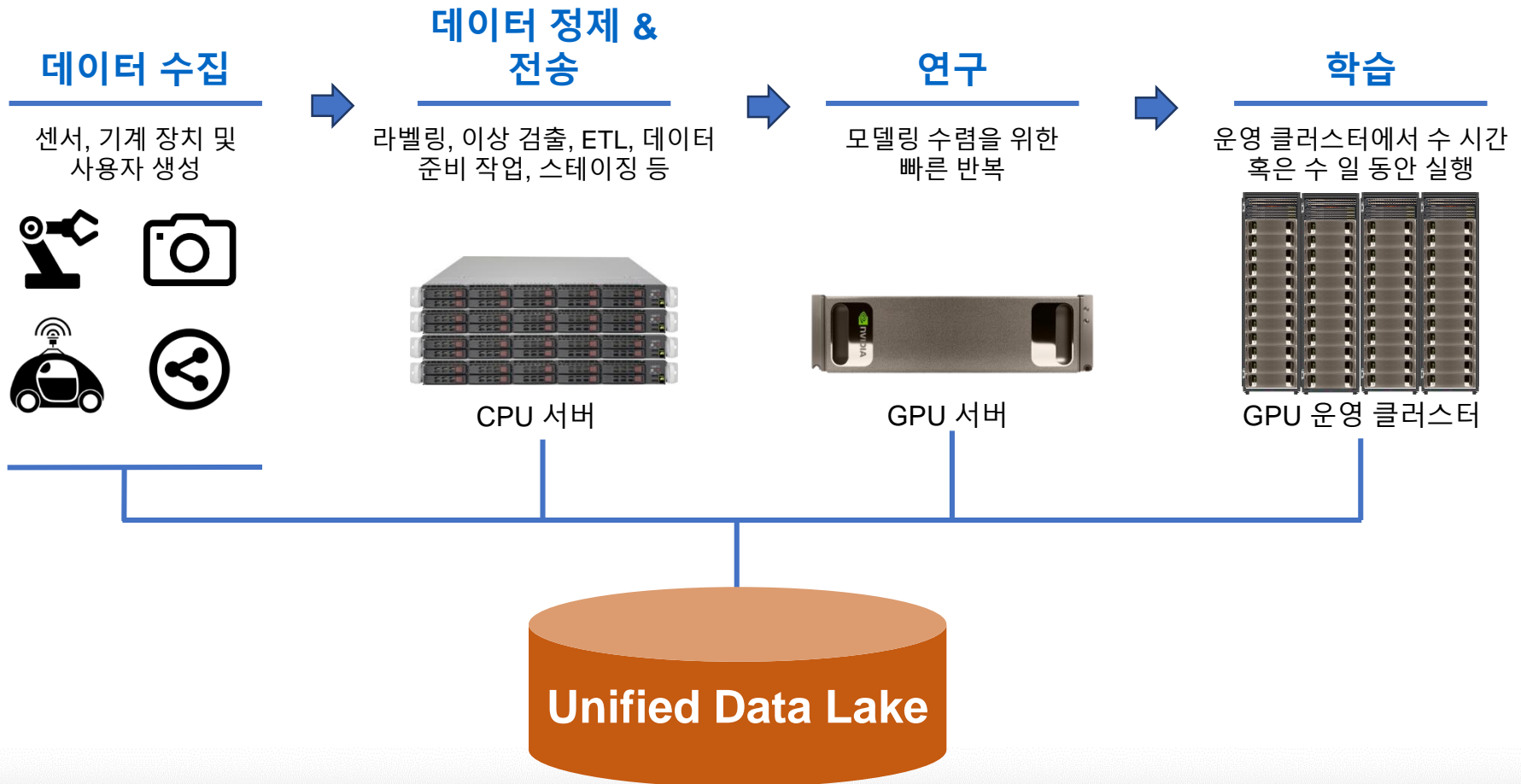
# AI, ML 및 DL 환경의 워크로드 상세 흐름 #복잡?

AI 인프라 환경의 단계별 데이터 저장소 별도 구성 시 문제점



# AI, ML 및 DL 환경의 워크로드 상세 흐름

AI 인프라 환경을 위한 데이터 수집에서 학습 단계에 이르기 까지 단일 데이터 파이프라인 구성 필요



# AI 플랫폼 도입 시 고려사항

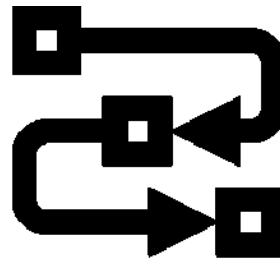
디자인, 배치 및 운영 상 병목 지점에 대한 해결 방안 필요

## 추정에 의한 디자인



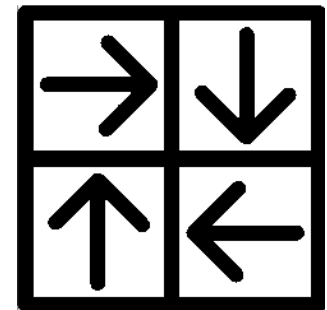
확장 시 예측 가능한  
성능을 제공할 수 있는  
아키텍처 여부 확인

## 배치의 복잡성



컴퓨팅, 스토리지, 네트워크  
및 소프트웨어의 구매, 설치  
및 문제 해결 방식

## 다중 지원 접점



구축된 스택 내 다중  
계층에 걸친 다수 벤더간  
논쟁

# 향상된 AI 인프라 구축을 위한 파트너십



# PureStorage AIRI (AI-Ready Infrastructure)



## AIRI

## 업계 최초의

완벽한 AI 환경을 위한 인프라스트럭처

### 하드웨어

**NVIDIA® DGX-1™** | 4x DGX-1 Systems | 4 PFLOPS of DL Performance

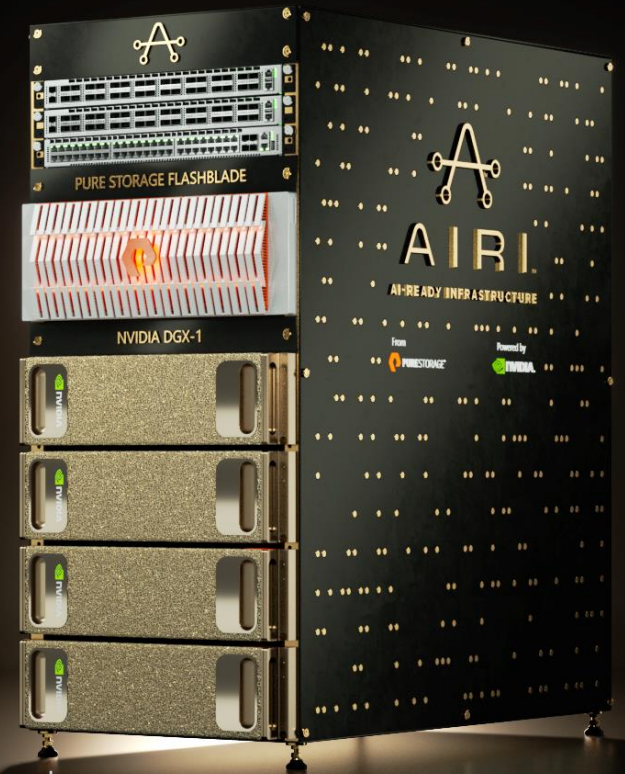
**PURE FLASHBLADE™** | 15x 17TB Blades | 1.5M IOPS

**CISCO OR ARISTA** | 2x 100Gb Ethernet Switches with RDMA

### 소프트웨어

**NVIDIA GPU CLOUD DEEP LEARNING STACK** | NVIDIA Optimized Frameworks

**AIRI SCALING TOOLKIT** | Multi-node Training Made Simple



# NetApp ONTAP AI

- 하드웨어
  - NVIDIA DGX-1 supercomputers
  - NetApp AFF A800 cloud-connected all flash
  - Cisco Nexus 3232C 100GbE switches
- 소프트웨어
  - NVIDIA GPU Cloud Deep Learning Stack
  - NetApp ONTAP 9
  - **Trident, dynamic storage provisioner**
- 지원
  - Fast track technology partner collaboration

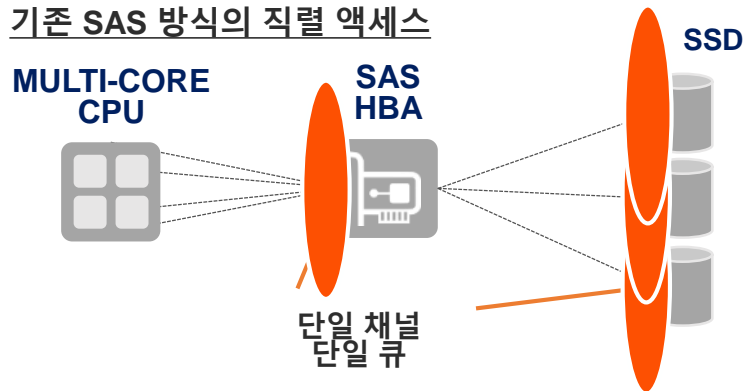


POWERED BY  
 **NVIDIA**

# AIRI & ONTAP AI의 공통점

## 1. 최신의 NVMe 스토리지 탑재 (PureStorage FlashBlade & NetApp AFF A800)

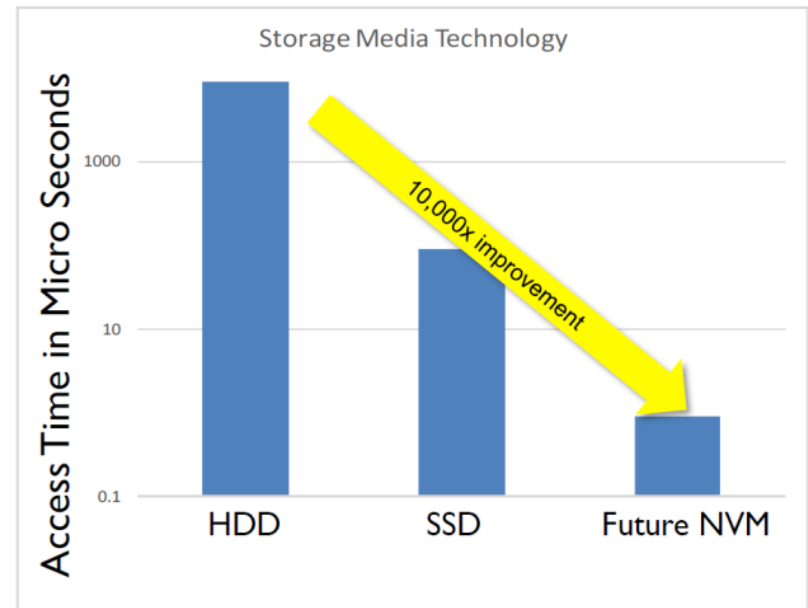
기존 SAS 방식의 직렬 액세스



새로운 NVMe 방식의 병렬 액세스



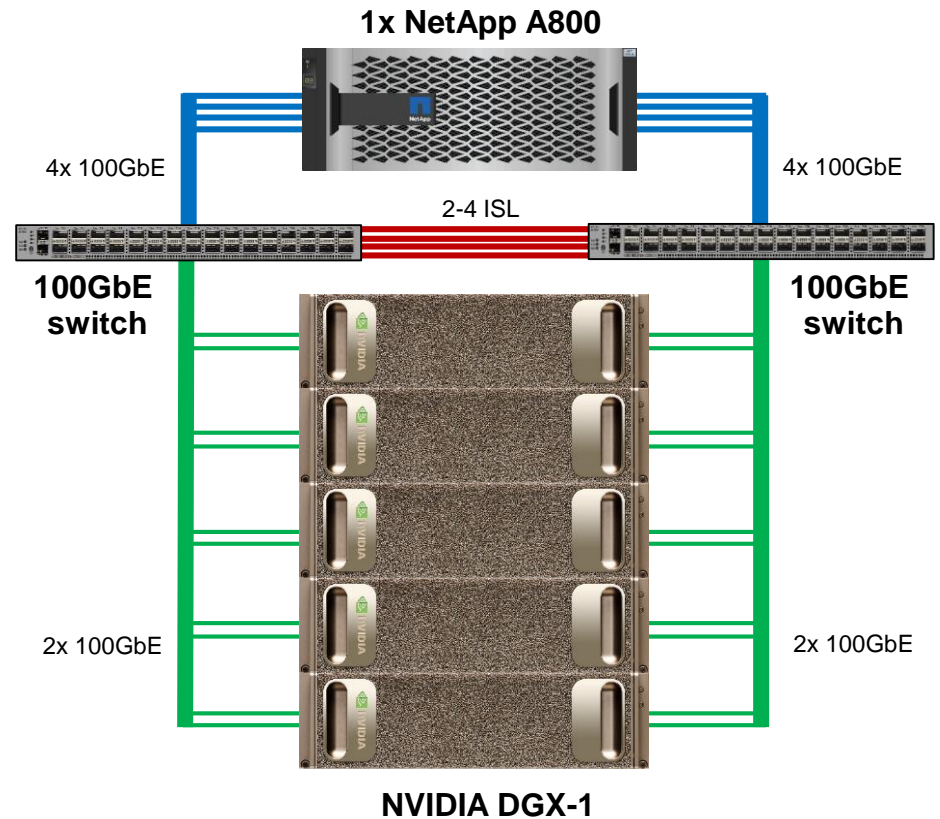
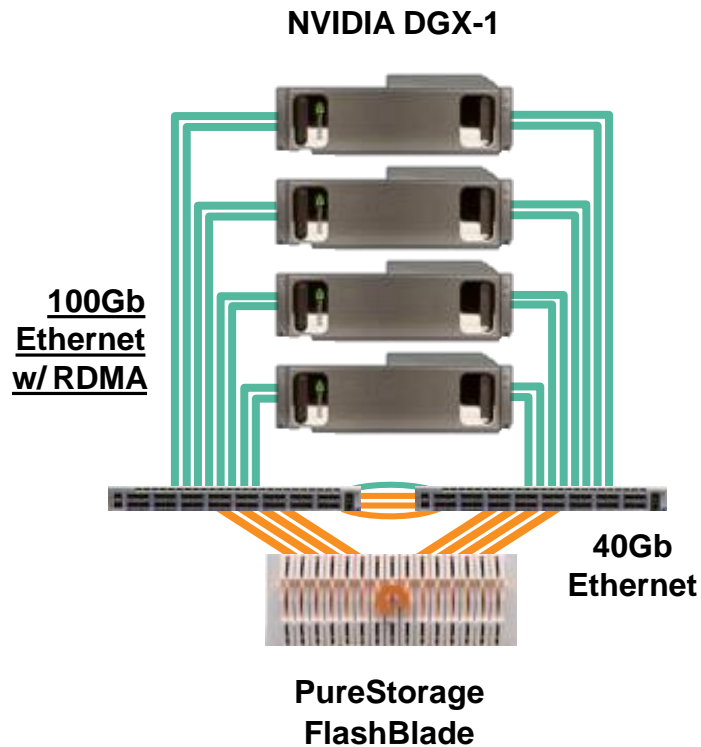
## 고성능 업무 환경 개선 → NVMe





# AIRI & ONTAP AI의 공통점

## 2. 100GbE Ethernet Network Switch 구성



# AIRI & ONTAP AI의 공통점

## 3. GPU 서버의 Local SSD Caching Area 사용 권장

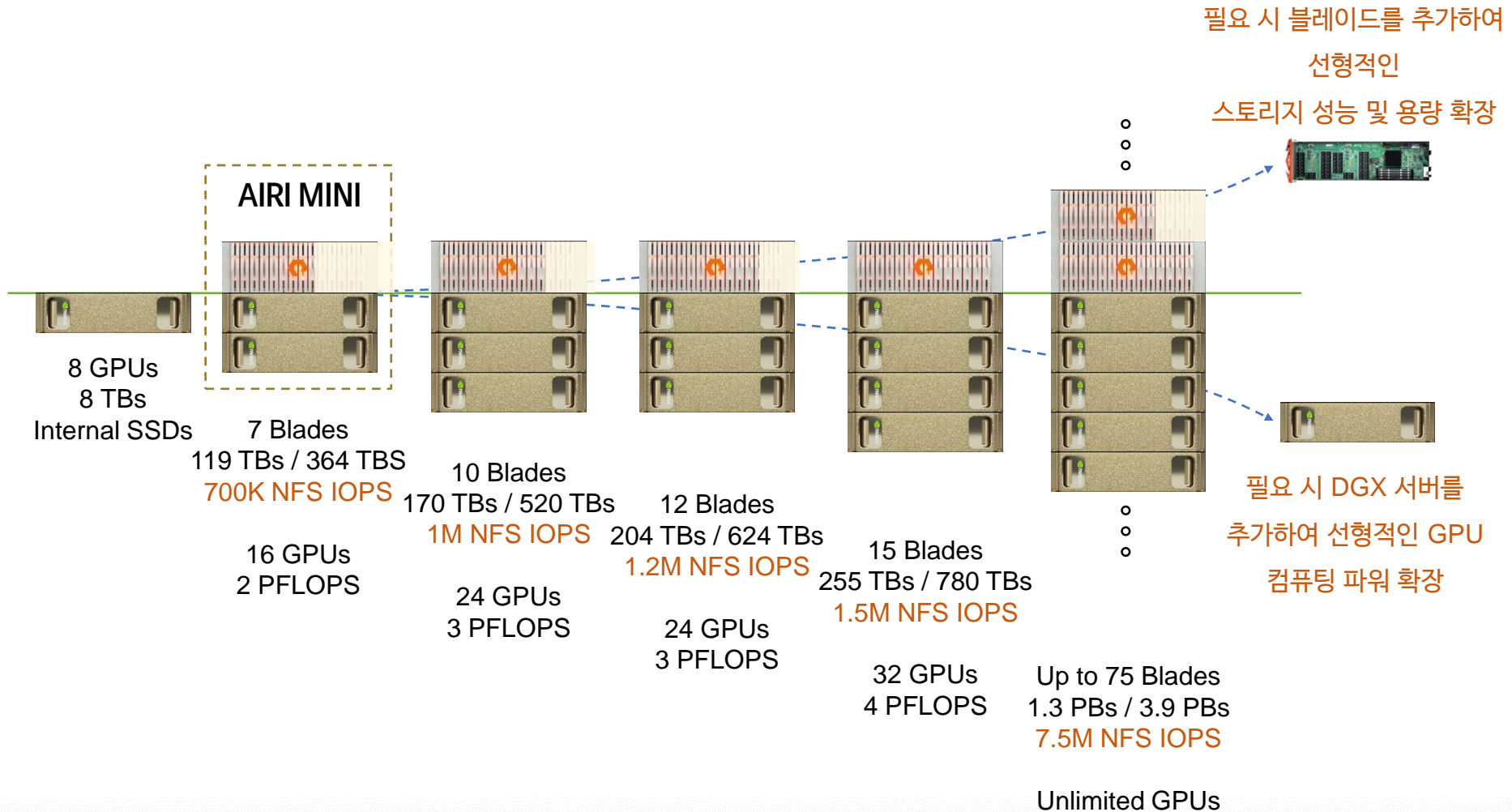
### 3.5. Configuring NFS Mount and Cache

The DGX-1 includes four SSDs in a RAID 0 configuration. These SSDs are intended for application caching, so you must set up your own NFS drives for long term data storage. The following instructions describe how to mount the NFS onto the DGX-1, and how to cache the NFS using the DGX-1 SSDs for improved performance.

Make sure your DGX-1 is set up in Base OS mode, that you have an NFS server with one or more exports with data to be accessed by the DGX-1, and that there is network access between the DGX-1 and the NFS server.

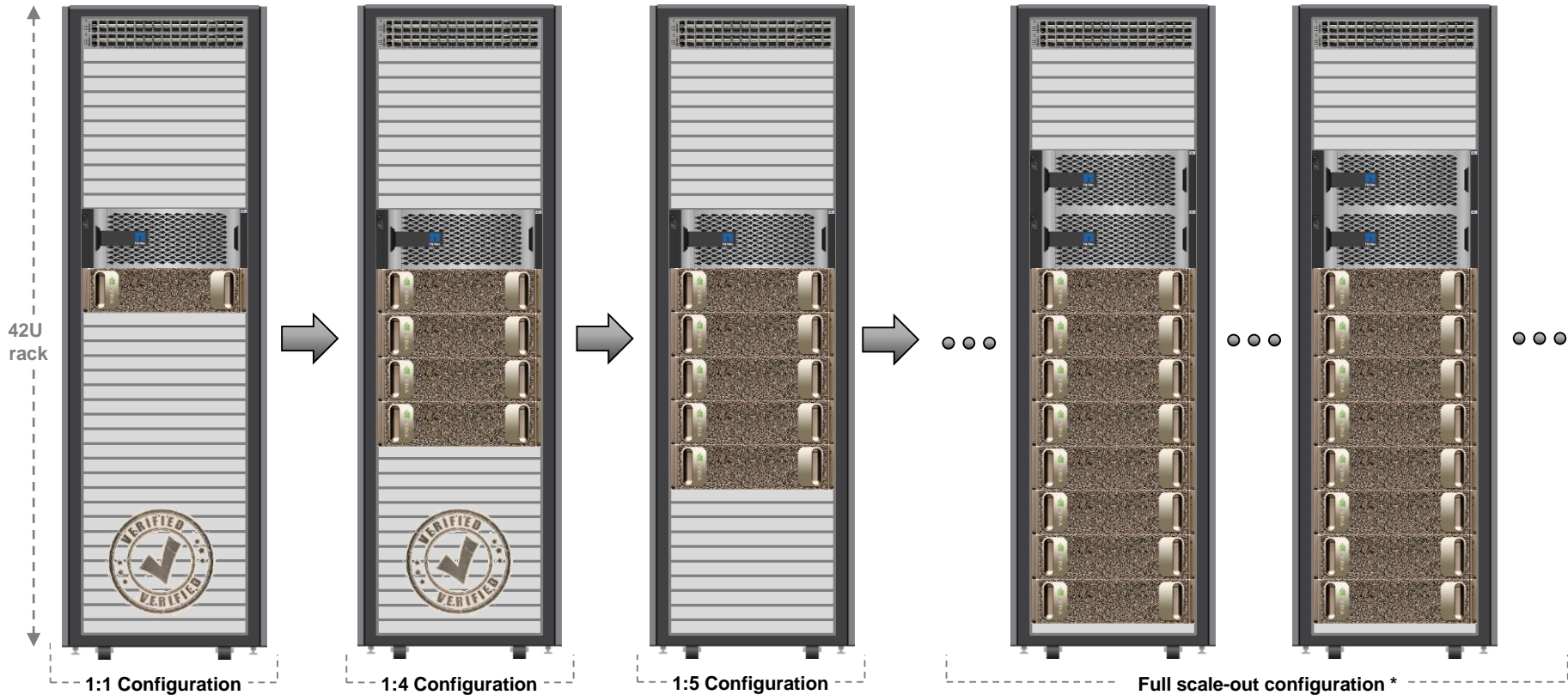
# AIRI & ONTAP AI의 공통점

## 4. 확장성 – Scale-Out 구성 (PureStorage AIRI)



# AIRI & ONTAP AI의 공통점

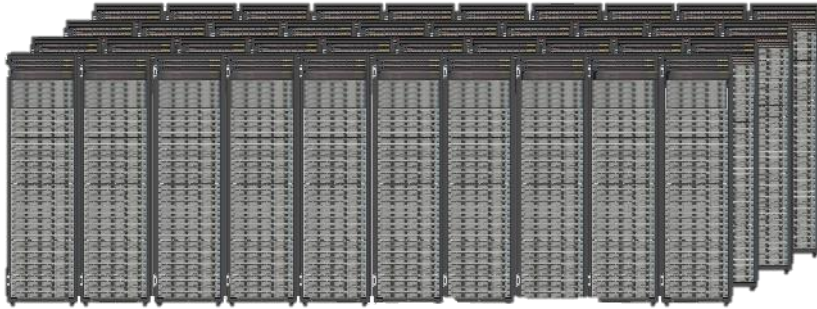
## 4. 확장성 – Scale-Out 구성 (NetApp ONTAP AI)



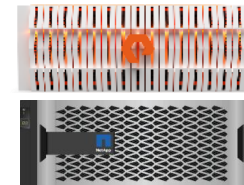
\* Based on 35kW racks

# AIRI & ONTAP AI의 공통점

## 5. 집적도 (상면, 전력, 냉각 비용 획기적으로 절감)



**NVIDIA® DGX-1™**  
Delivering performance  
of 40 racks of CPUs



**PURE  
FLASHBLADE™ &  
NetApp AFF A800**  
Delivering performance  
of 10 racks of Disk



# 해외 AI 인프라 구축 사례 – Zenuity



Make it real.

## 2021년 까지 자동 주행 자동차 현실화

Volvo와 Autoliv의 합작 법인인 Zenuity는 2021년까지 실제 생산 자동차를 위한 자동 주행 소프트웨어 출시를 목표로 하고 있다. 그들은 딥러닝 인프라를 구축하기 위해서 NVIDIA DGX-1 서버와 PureStorage FlashBlade 시스템을 선택하여 AI 계획을 촉진시키고 있다.

# 해외 AI 인프라 구축 사례 - PAIGE



PAIGE

“ 데이터는 AI 혁명을 이끄는 연료입니다. 전세계에서 가장 큰 **종양병리학** 기록보관소에 액세스하면서, 우리는 대용량의 데이터가 의학적으로 인증된 AI 애플리케이션으로 빠르게 전환될 수 있도록 가장 향상된 딥 러닝 인프라를 필요로 했습니다..

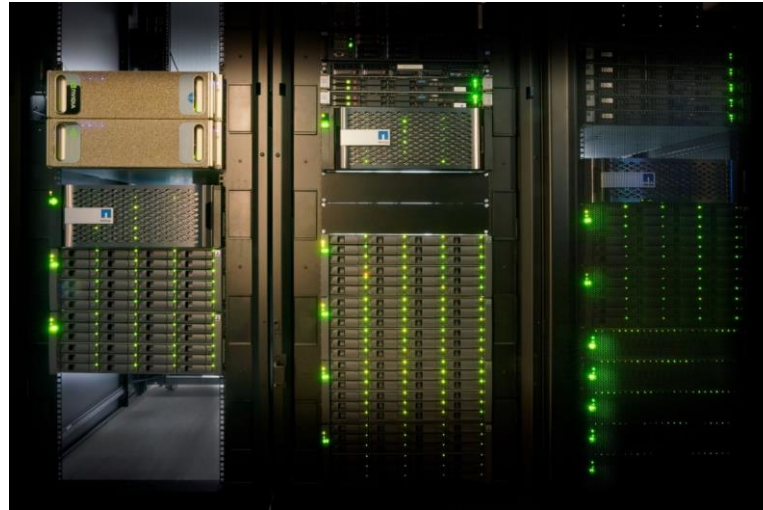
DGX-1과 FlashBlade의 강력한 조합은 **의학 업계**에 AI를 촉진시키겠다는 우리의 임무를 촉진시켰습니다. AIRI는 우리의 AI 인프라를 더욱 배가 시켜줄 수 있는 동일한 핵심 기술로 설계되었으며, 다른 기업을 위해서도 AIRI와 함께 AI 계획을 시도할 때 무엇이 가능한지를 알게 되어 매우 흥분됩니다.

Dr. Thomas Fuchs  
Founder, Chief Science Officer  
Twitter @ThomasFuchsAI



# 해외 AI 인프라 구축 사례 – Cambridge Consultants

- Cambridge Consultants에 의한 Digital Greenhouse
  - 전세계 가장 향상된 AI 연구 시설 중에 하나
  - 최첨단 딥 러닝 기술을 연구하고 개발하기 위한 연구소
  - 매년 400개 이상의 프로젝트 수행
  - 전세계 900명 이상의 직원 근무



Cambridge  
Consultants

“딥 러닝은 우리가 근무하고 있는 거의 모든 시장에 혁명을 일으키고 있다. NVIDIA DGX 슈퍼컴퓨터와 NetApp 올-플래시 스토리지가 함께 구성된 NetApp ONTAP는 딥 러닝을 위한 데이터 파이프라인을 단순화하고 촉진시키고 있다.”

Monty Barlow, Head of Artificial Intelligence, Cambridge Consultants

# 결론 – AI 인프라를 위한 이상적인 데이터 플랫폼

## AI (ML&DL) 및 빅데이터

#1

내부 RAM 혹은 SSD에 저장된 수준과 유사한 DL 학습 성능을 제공할 것

We provide

#2

“모든 데이터를 위해” 초고속, 대용량 스토리지를 이용하여 전체 AI 데이터 파이프라인을 위한 타협하지 않는 성능과 단순함을 제공할 것



클라우드 컴퓨팅

정보 보호

디지털 트랜스포메이션을 통한  
비즈니스 혁신과 4차 산업혁명 대응을 위한

# GIT 솔루션즈 데이

감사합니다.

---