PURE// ACCELERATE DIGITAL KOREA 2021

BREAKTHROUGH

상상을 초월하는 혁신

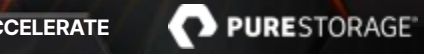
PURESTORAGE®

데이터분석과 AI 환경의 현대화

김태훈 부장
Unstructured Data Specialist
thkim@purestorage.com
Pure Storage Korea

신창호 대표
CEO
chshin@imgr.co.kr
IMGURU

김정묵 운영총괄
COO
jmkim@lablup.com
Lablup



Session Agenda

1. 데이터 분석과 AI 의 요구 환경 변화
2. 데이터 분석 환경의 현대화 - IMGURU
3. AI 환경의 현대화 - Lablup
4. Summary | FlashBlade UFFO - Unified Fast File Object

데이터 분석과 AI 의 요구 환경 변화

데이터 활용이 기업의 미래

현대화된 데이터 활용은 비즈니스 성공의 원동력

TODAY

자유로운 데이터 접근과 활용의 어려움

단일 업무, 관리의 어려움

최신 데이터 및 애플리케이션에는 너무 느림

제한된 데이터 공유 및 재활용 어려움

비효율적인 스토리지 및 컴퓨팅 자원 사용률

FUTURE

UFFO (Unified Fast File and Object) 플랫폼은 디지털 시대 필수 인프라

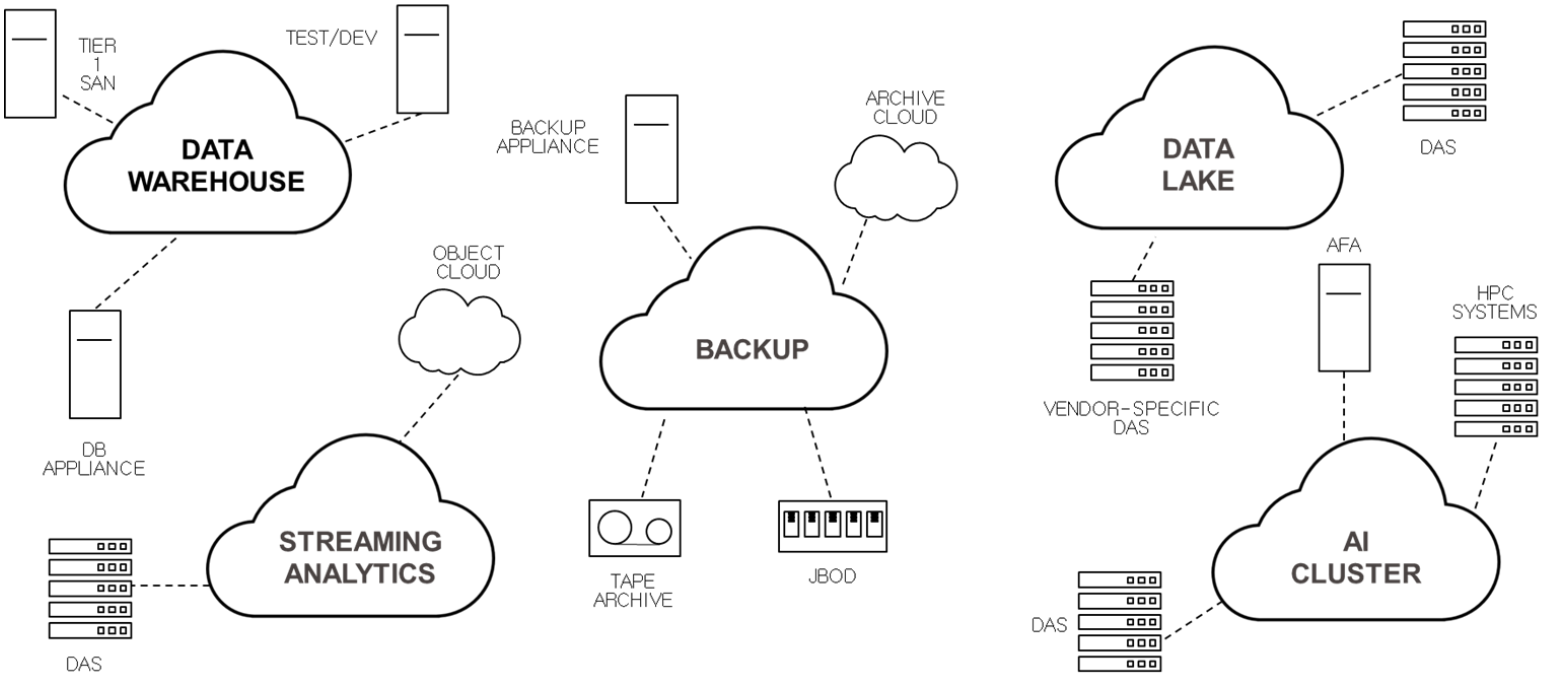
4.8x 수익향상
- 신제품 수익에서 경쟁사를 능가할 가능성이 4.8배 이상 높음

3.2x 고객 만족도 향상
- 사전 분석투자 보고 받을 가능성이 3.2배 이상 높음

2.4x 운영 우수성 향상
- 직원 당 매출 증가 가능성이 2.4배 이상 높음

데이터에서 새로운 인사이트 생성

무질서한 사일로에 갇힘



데이터 분석 요건의 변화

hadoop

DECADE AGO

TensorFlow mxnet Caffe2

Flume docker R python

Cassandra mongoDB Couchbase

Parquet kafka logstash

hadoop Apache Spark elastic

TODAY

DATA IS NOW DIFFERENT

Small to Large Files

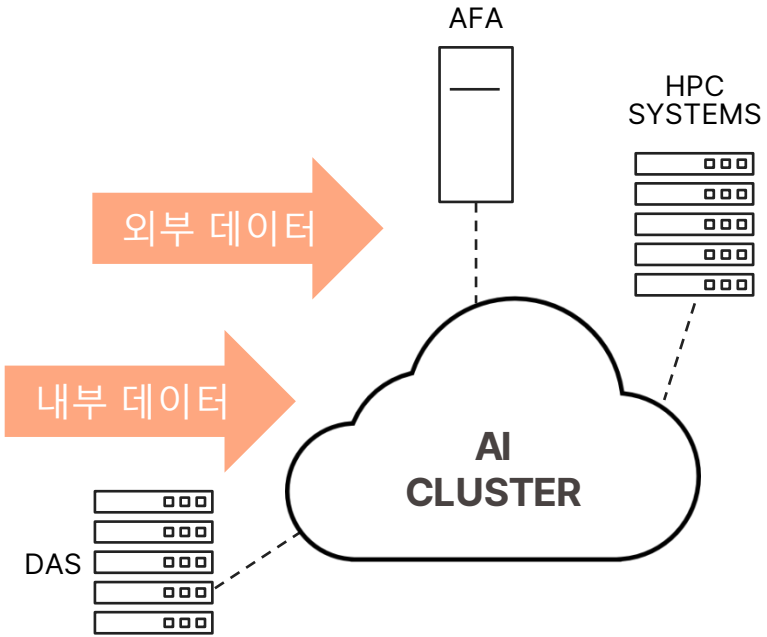
Random to Sequential Access

Real-time or Batched

Apps & Data Evolve Quickly

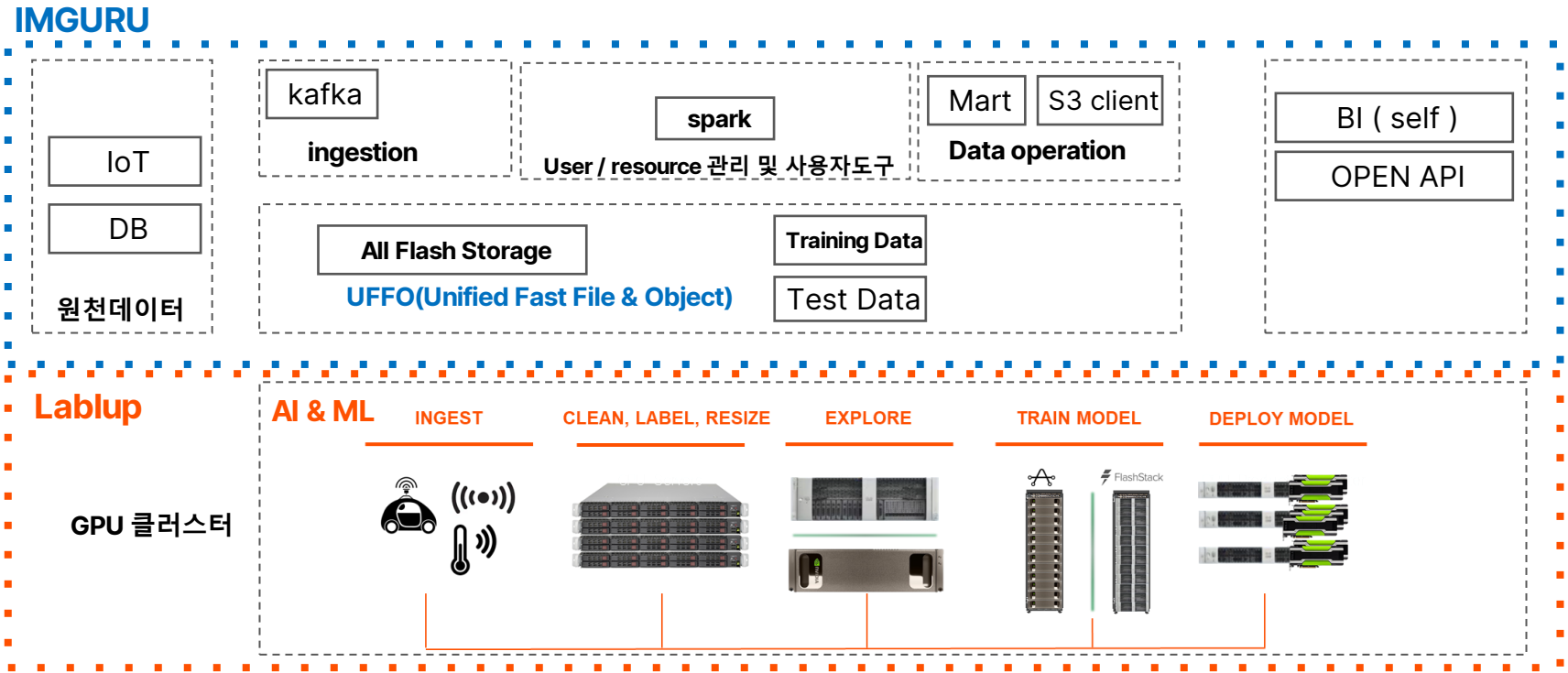
Elastic Infrastructure

AI 요구 환경 변화



- 다수의 개별 AI 프로젝트 수용
- 대규모 리소스가 필요한 AI 프로젝트 수용
- 민첩하고 검증된 아키텍쳐 , 스케일아웃 아키텍쳐
- 내부 생성 데이터, 외부 데이터를 저장하고 공급하는 오픈아키텍쳐

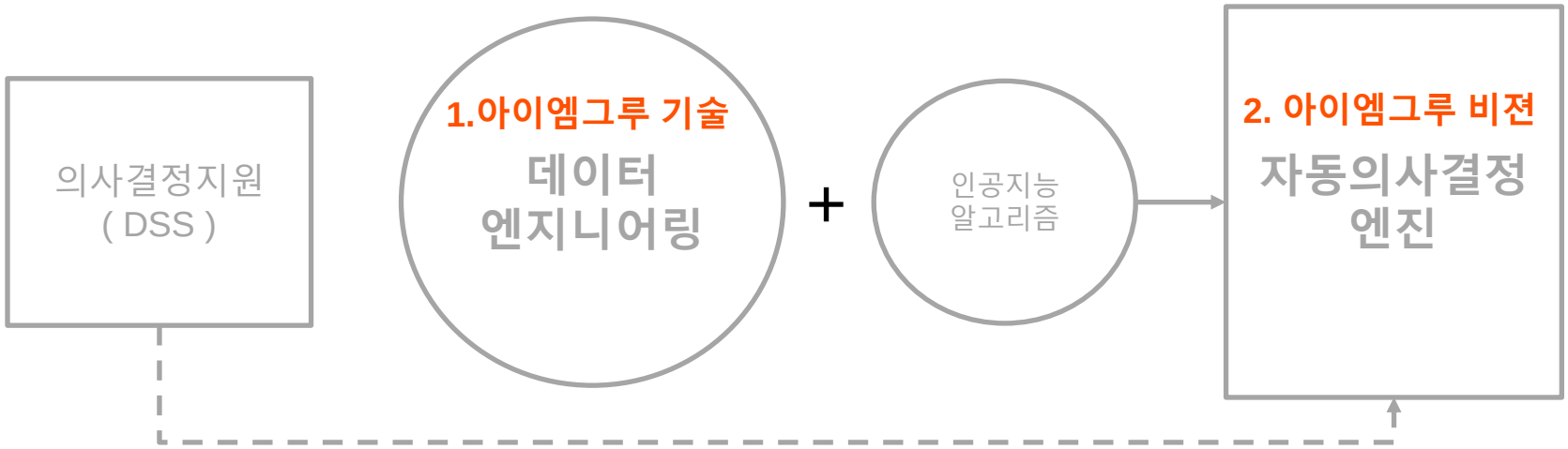
IMGURU & Lablup



아이엠그루
데이터 아키텍처 현대화 (스트림 아키텍처)

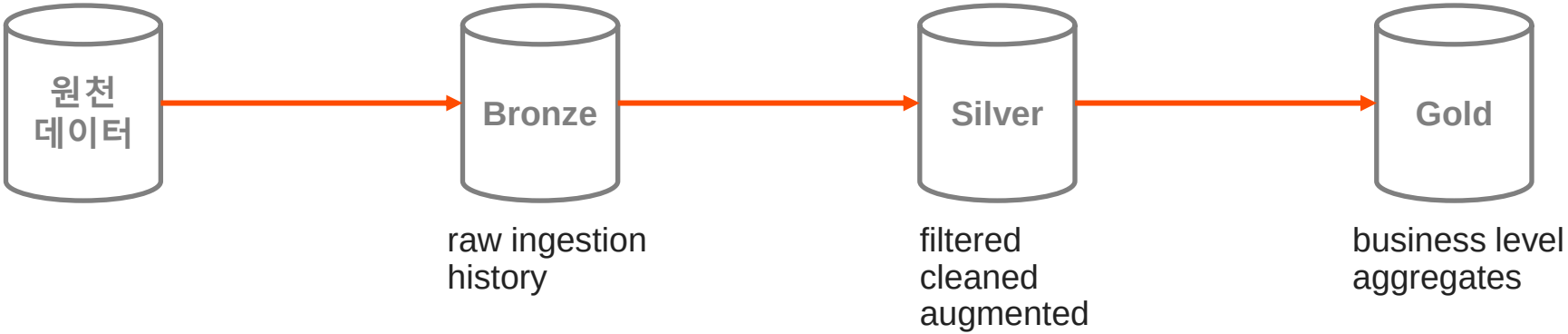
아이엠그루 소개

자동 의사결정 전문 회사 아이엠그루 소개



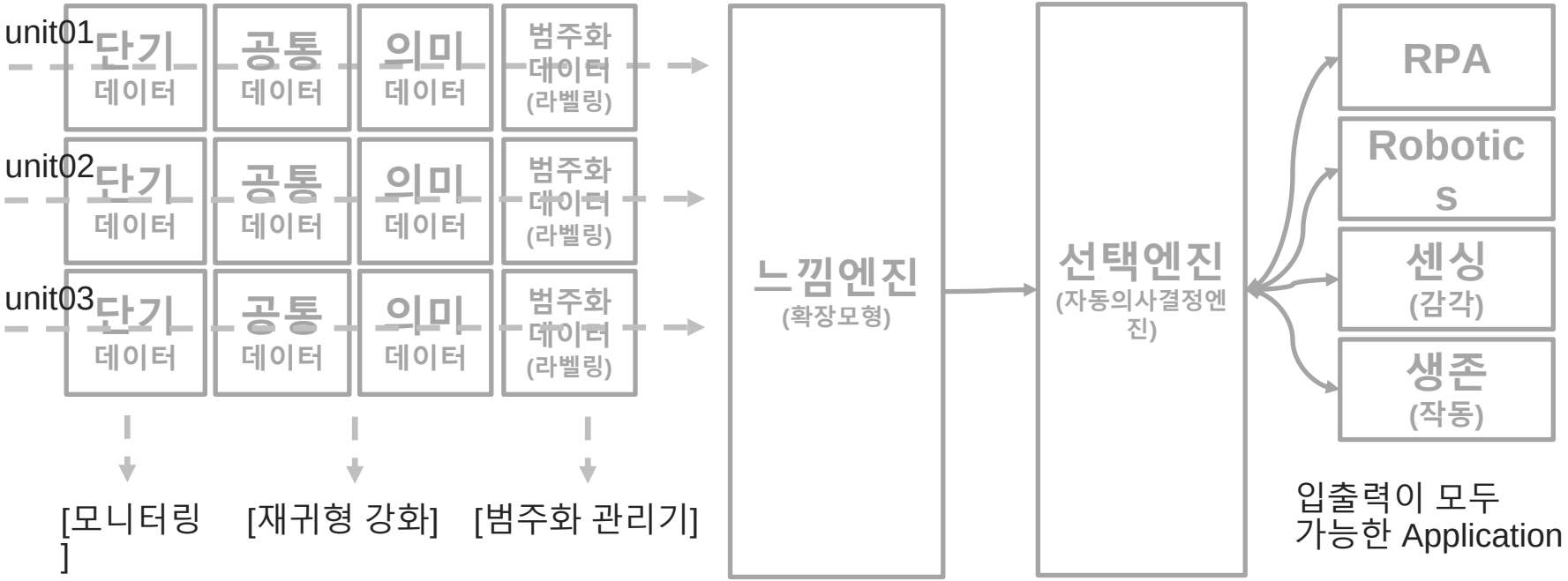
데이터 엔지니어링

데이터 엔지니어링 살펴 보기. 날 것인 데이터를 사용 가치가 있는 형태로 만드는 일을 합니다.



아이엠그루 느낌엔진

아이엠그루의 느낌 엔진의 데이터 구조 이해를 통해 현재 데이터 모델링의 중요성을 확인 합니다



아이엠그루 경험

데이터 기반 의사결정 전문 업체 아이엠그루는 데이터 엔지니어링 기술과 의사결정을 위한 성능 확보 기술을 지향 하며, 이를 위한 솔루션을 보유 또는 리셀링 하고 있습니다

Finance

우리은행
- Build an AI solution(IM:VITA)
- Build a Bigdata Platform

KB국민은행
- Build a Bigdata Platform
- Build an Information Lake
- Build an Untact customer behavior analysis (IM:TORA)

ShinhanCard
- Build a Hyper personalization system

삼성화재
- Build an Insurance service integration system
- Build an Insurance data mart

KB카드
- Build a Hyper personalization system

BIG DIPPER
- Build a Business Platform
- Capital & Technology investment

Public / Wholesales / Manufacturing

emart
- Build an Inventory Management System

ONE store
- DW off road using Data Lake

Zero to Seven
- Website Renewal (DW / BI)

경찰청
- Build an High-tech investigation system (IM:Watson)
- Build a GEO plus(IM:Watson)
- Build a DSS for Profiler (IM:TORA)

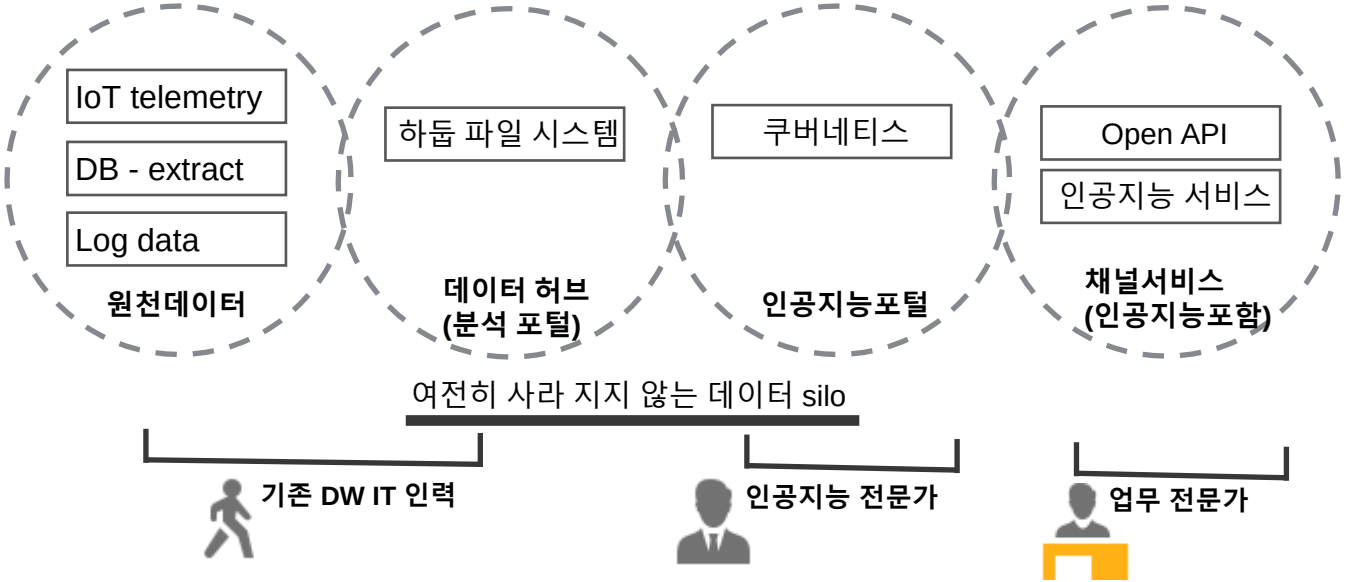
SAMSUNG
- Build Samsung HRI CIC system (IM:Watson)
- Build a GLEX system
- Build a multicampus Advanced CIC

LG전자
- Build an LG.com Search Platform (2010 ~ 2019)
- Build a VOC analysis platform (IM:TORA)

1. Search
2. 자연어 처리
3. DW / BI
4. Data Lake
5. 초개인화
6. 용의자 추천

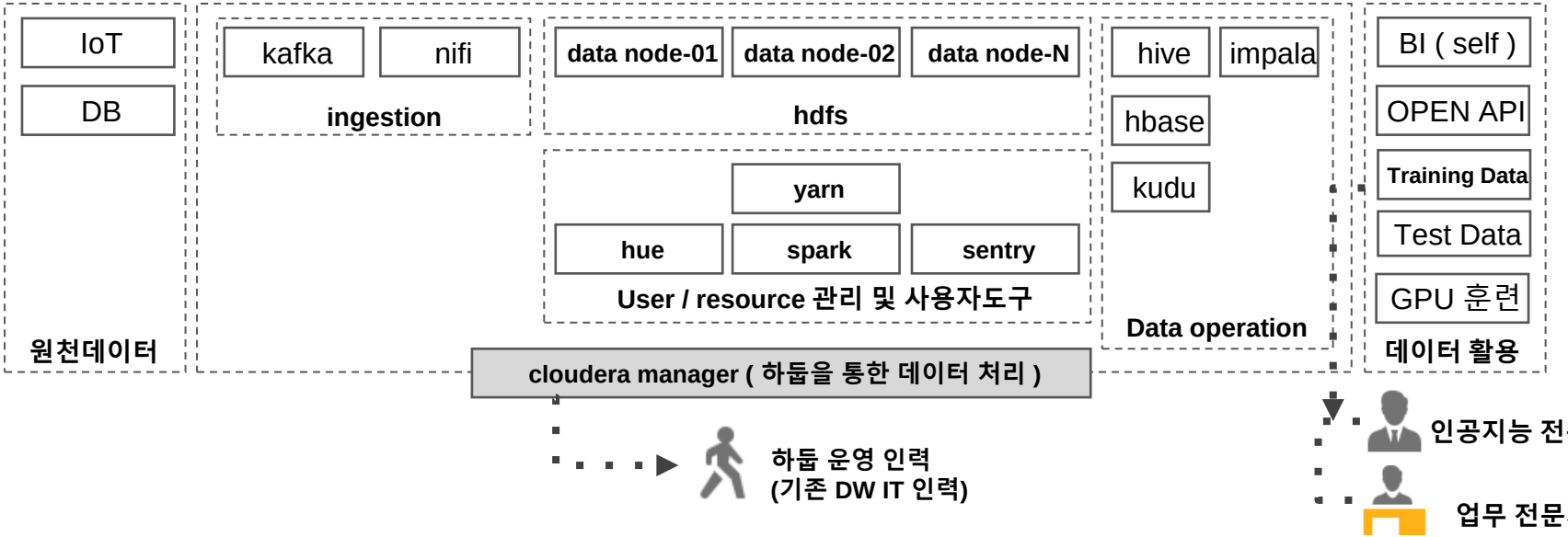
데이터 분석 시스템의 현재

이미 데이터 분석 시스템이란 이름으로 고객들이 보유한 데이터 분석 인프라는 이렇게 구성 됩니다



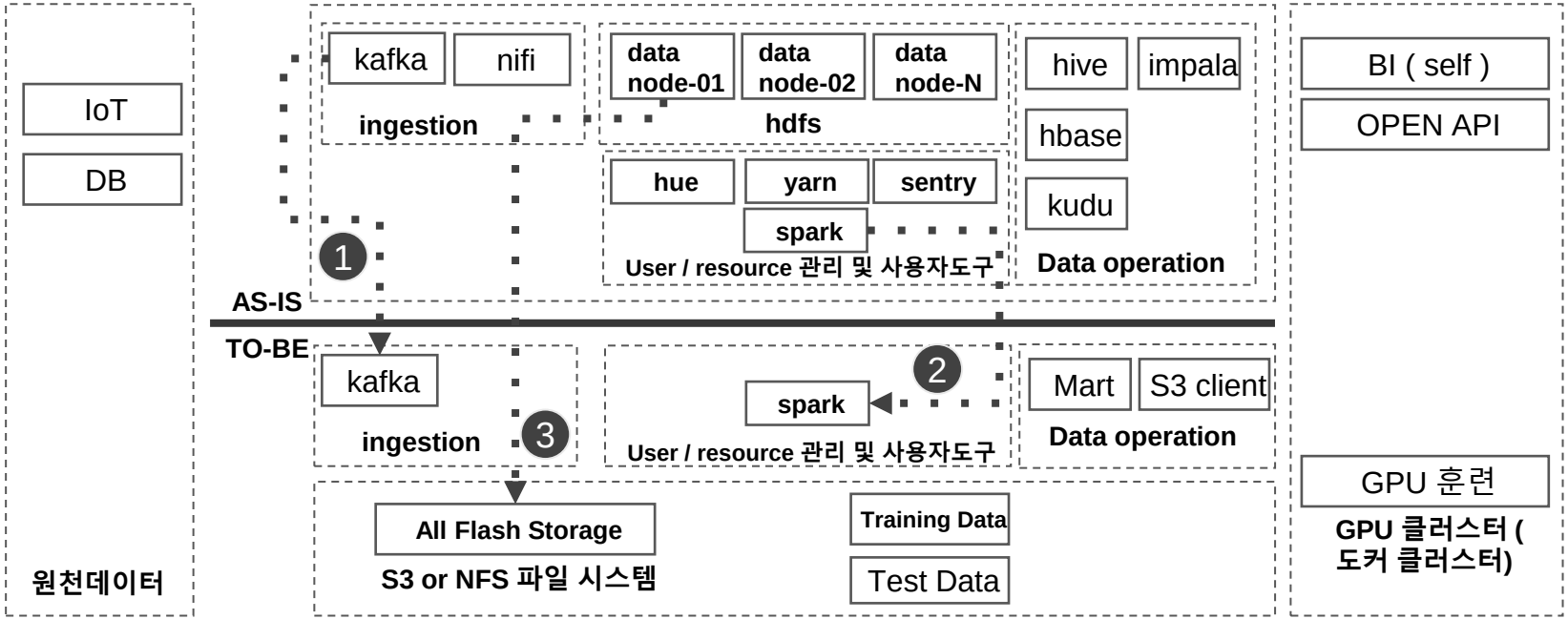
HDFS 상세 예제

상세 시스템 구성도로 사용의 실제를 보고 넥스트 고객 시스템의 방향을 구상 합니다



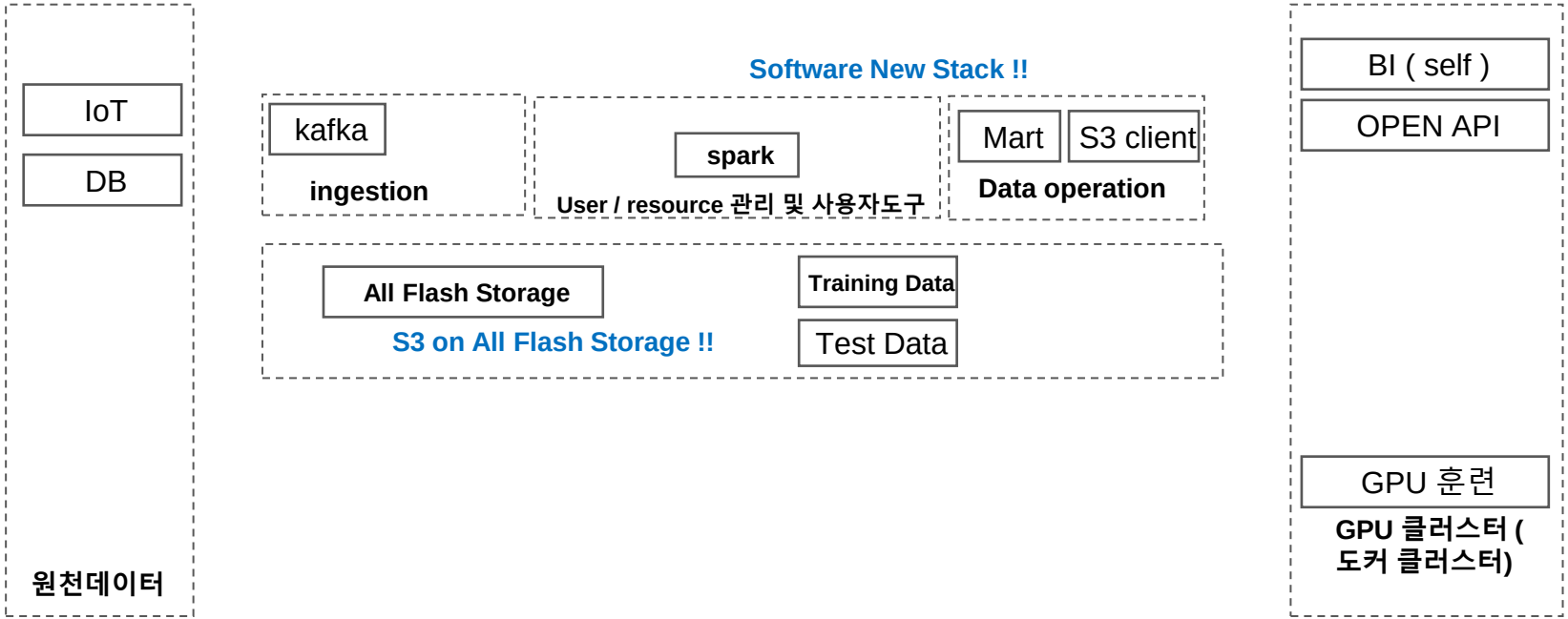
요구 사항 해결을 위한 새로운 아키텍처

나열된 문제점들을 왜, 어떻게 해결 하였으면 좋을지 고객들의 바람을 정리 합니다



스트림 아키텍처에 스토리지의 중요성

이제 까지 이야기한 많은 데이터 모델이나 알고리즘은 모두 (1) 성능확보 (2) 지속, 확장성 (3) 관리용의성 을 갖어야 합니다. 그래서 퓨어 스토리지가 필요합니다



Backend.AI : Hyperscale AI Infra w/ Ease

김정묵
COO
jmkim@lablup.com
Lablup Inc.

Lablup Inc : Leadership

Jeongkyu Shin — Co-founder / CEO

- 포스텍 물리학과 박사 / 컴퓨터공학과 복수 전공
- 복잡계 뇌과학 및 기계학습 기반의 Agent-based model 전공
- ML/DL Google Developer Expert
- 기계학습 파트 및 시스템 설계 담당

Joongi Kim — Co-founder / CTO

- KAIST 전산학과 박사
- 80Gbps 성능의 GPU 기반 네트워크 가속 및 부하분산 기술 최초 개발
- 분산처리 시스템 설계 및 개발 담당
- Best Student Paper in ACM EuroSys Conference

Jonghyun Park — Co-founder / Research Director

- 포스텍 물리학과 박사
- DNA-단백질 바인딩 속성 분석 빅데이터 연구
- 연구 지원 시스템 설계, 개발 및 연구 담당

Jeongmook Kim — CO O

- 포스텍 화학공학과 석사
- SK Innovation Platform Technology R&D Center
- 시스템 운영 및 파이프라인 담당

Lablup Inc : Tech

- 검증
  - NVIDIA DGX-Ready Software partner (Asia 최초, 유일)

- 기술
  - 자체 컨테이너 오케스트레이터
  - 드라이버 레벨의 GPU 분할 가상화 기술
  - 하이브리드 클라우드 플랫폼 기술
  - 연산 노드간 직결 보안 계층 기술

- 특허
  - 컨테이너 기반의 GPU 가상화 방법 및 시스템 / KR 10-2018-0169620
  - 사용자가 요청한 다수개의 라이브러리를 탑재한 세션 컨테이너 제공 방법 / KR 10-2019-0049711

- 수상
  - D.Camp 주최 D.Day 우승 (2016.4)
  - NVIDIA AI Conference Inception Award 우승 (2018.11)
  - 소프트웨어산업 발전 유공자 대통령표창 / 신정규 대표 (2018.12)
  - 대한민국 인터넷 대상 기술혁신부문 과기정통부장관상 (2019.12)
  - 공개SW 산업발전 유공자 과기정통부장관 표창 (2019.12)
  - 과기정통부 주최 대한민국 SW 품질대상 우수상 (2020.11)

Lablup Inc : Customers

Enterprises

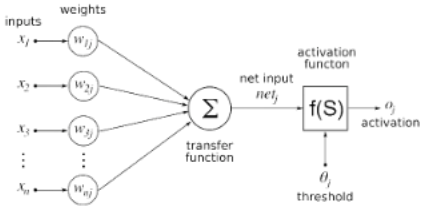

Public

Universities




Situation : Hyperscale AI

Deep Learning = 수십억 개 이상의 매개변수를 갖는 행렬 계산의 반복



$$\begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 58 & \\ & \end{bmatrix}$$

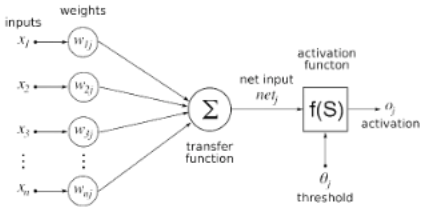
$$1 \cdot 7 + 2 \cdot 9 + 3 \cdot 11 = 58$$

Reference: NVIDIA 2017 "A NET COMPUTING ERA"

계산량 산정: GOPS * bandwidth

Situation : Hyperscale AI

Deep Learning = 수십억 개 이상의 매개변수를 갖는 행렬 계산의 반복

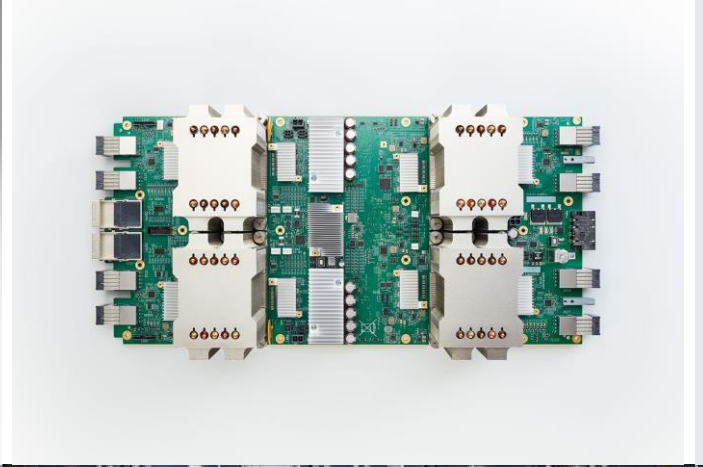


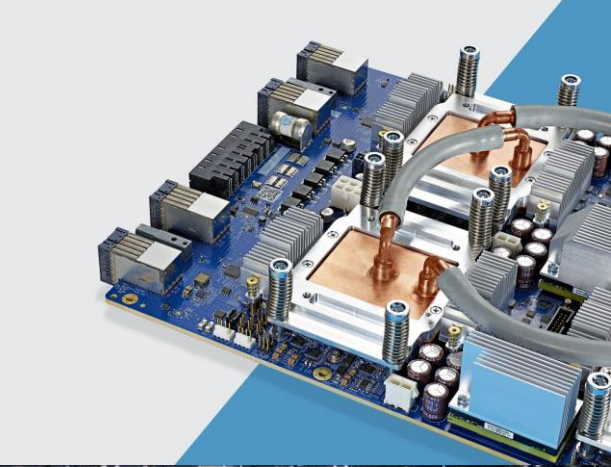
$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix} = \begin{bmatrix} 58 & \\ & \end{bmatrix}$$

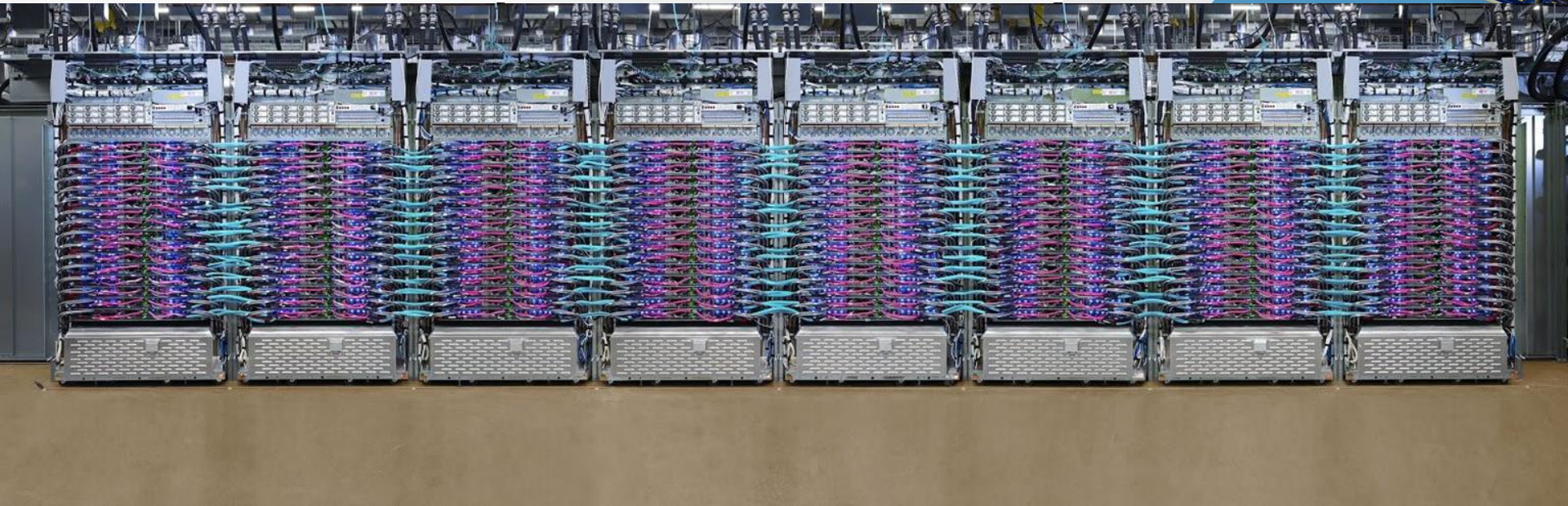
$$1 \cdot 7 + 2 \cdot 9 + 3 \cdot 11 = 58$$

Reference: NVIDIA 2017 "A NET COMPUTING ERA"

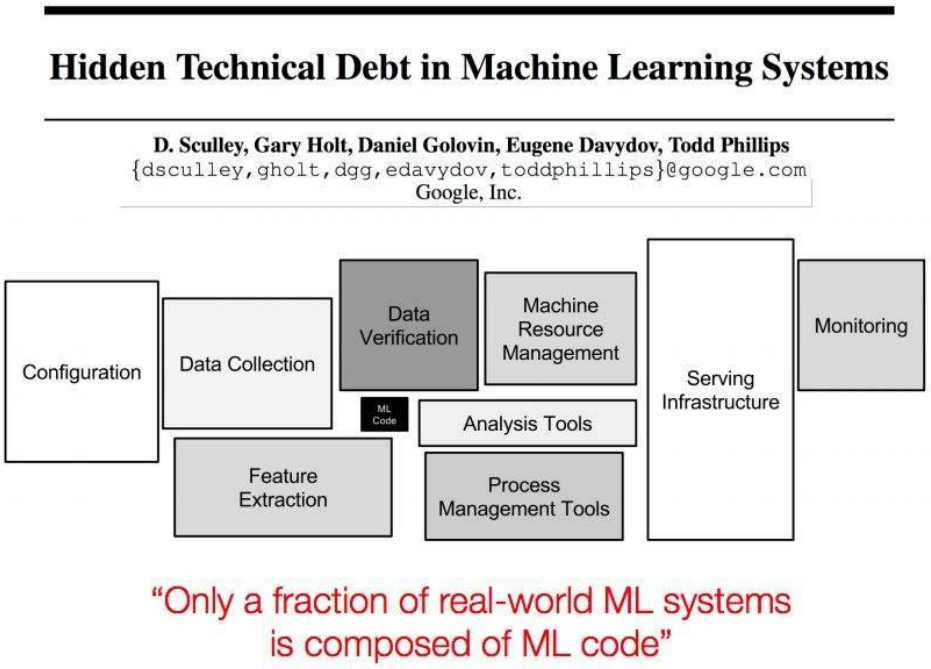
계산량 산정: GOPS * bandwidth







Challenges : AI Infra Mngmt



- 플랫폼 구축의 기술적 난이도 상승
  - 다른 분야와의 차이점: **빠른 변화 사이클** (하드웨어 및 소프트웨어 플랫폼, AI 프레임웍, 모델 등)
  - 구축이 끝나는 시점에서는 이미 구세대 플랫폼이 됨
  - 이로 인한 관리의 어려움
- 비자동화된 부분이 상당히 많음
  - 설정, 데이터 수집, 검증, 특징 추출, 리소스 관리, 인프라 관리 등
  - 대규모 데이터 관리

Lablup's Answer : Backend.AI

AI 프레임워크용 엔터프라이즈 클러스터 백엔드

차별점 :

- 분할 GPU 자원 할당 및 공유
- 쉬운 GUI 와 강력한 CLI / SDK 제공
- 훌륭한 관리 편의성과 높은 사용률

Lablup's Answer : Backend.AI

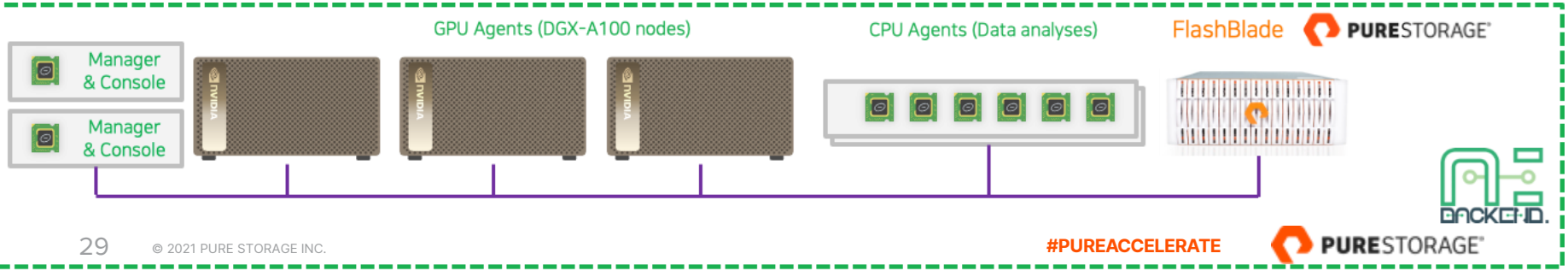
AI 프레임워크용 엔터프라이즈 클러스터 백엔드

차별점 :

- 분할 GPU 자원 할당 및 공유
- 쉬운 GUI 와 강력한 CLI / SDK 제공
- 훌륭한 관리 편의성과 높은 사용률

Case Study : 해외A (대단위 개방형 AI 연구개발 플랫폼)

- Multi-region, Multi-organization 사용자용 머신러닝 최적화 클러스터 팜 구축
- 구성
  - 도입 규모 : A100 GPU 200대, 추가 고성능 CPU nodes (데이터 분석용)
  - **Backend.AI Reservoir** 추가 도입으로 완전 Air-gapped 환경 구축
- 고객 혜택
  - 대단위 팜 구성 설계 제공 및 SLA 극대화를 위한 **고가용성** 구성
  - 멀티노드 분산 훈련 지원 기능 및 GPU간 네트워크 기반의 대규모 초고속 딥러닝 훈련
  - Backend.AI Reservoir를 통해 **PyPI 및 Ubuntu 저장소를 완전 폐쇄망 내에서 자유롭게 사용 지원**
  - 기관 내외부 동시 서비스 시 시스템/데이터 보안을 위한 격리 도메인 구성



Unique Differentiation : Why?

성능

- VM 및 Kubernetes 기반 솔루션들 대비 동일 하드웨어로 고성능 달성
- AI, ML, HPC, 수치해석 등 연구 개발에 최적화
- 고성능 컴퓨팅에 특화된 다양한 배치, 자원 할당 및 병목 제거 구현

편의성

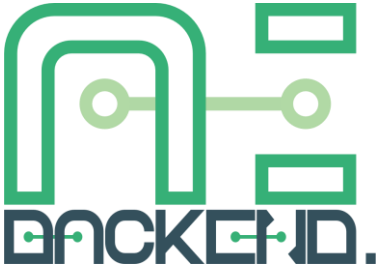
- ML / HPC 전문가들이 직접 만든 플랫폼
- 일관된 플랫폼 GUI (웹 및 데스크탑 앱) / 관리 동작 및 스크립팅을 위한 CLI
- 시스템 관리 컨트롤 패널을 통한 상세한 관리자 제어 기능

확장성

- 완전 문서화된 API 및 SDK (Python, Node.js) 제공
- 다양한 GPU 및 머신러닝 가속 H/W 지원[1]
- On-Prem에서 Public Cloud, Hybrid 클라우드까지 쉬운 확장

비용 절감

- GPU 분할 가상화(Fractional GPU™) 를 통한 고가 GPU의 활용성 증대 및 고가용성 달성
- 더 적은 하드웨어로 동일한 성능, 동일한 하드웨어로 더 높은 성능 제공
- 강력한 장애 대응 (연속성 Fail Over, 쉬운 장애 원인 분석 및 로그 API / 로그 솔루션 통합)



AI / ML / HPC 를

R&D 부터
Business Service,
AI Service 추론 및 제공까지

하나의 일관된 플랫폼으로 통해

효과적으로 관리

#PUREACCELERATE

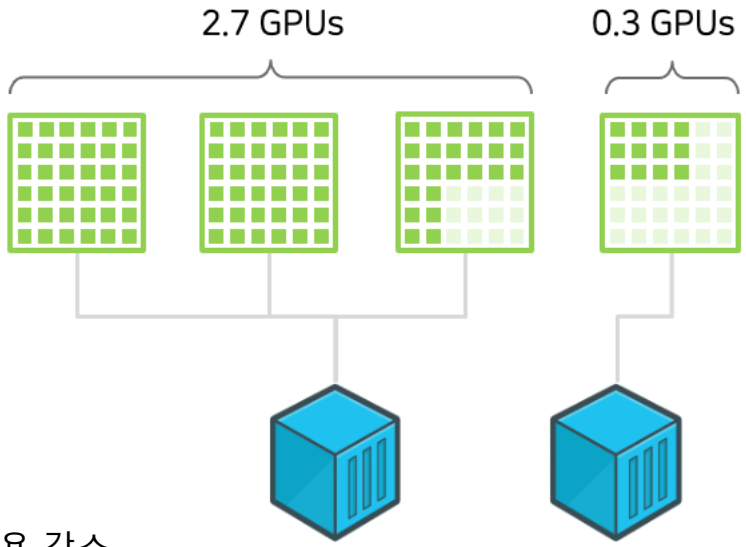
PURESTORAGE®

Backend.AI : GPU 가상화 (fractional GPU™) 기반의 유연한 자원 관리

- **컨테이너 기반 GPU** 스케일링
  - 컨테이너별로 CUDA SMP 및 GPU RAM을 나눠줌
    - ✓ 예) 2.7 GPUs, 0.3 GPUs를 각 컨테이너에 할당
  - 단일 GPU 공유 : 교육 및 추론 워크로드에 적합
  - 다중 GPU 할당 : 모델 훈련 등 대규모 워크로드에 적합
  - 자체 개발한 CUDA 가상화 계층으로 구현 [등록 특허]

- **NVIDIA 플랫폼 통합**
  - Validated NVIDIA DGX-Ready Software

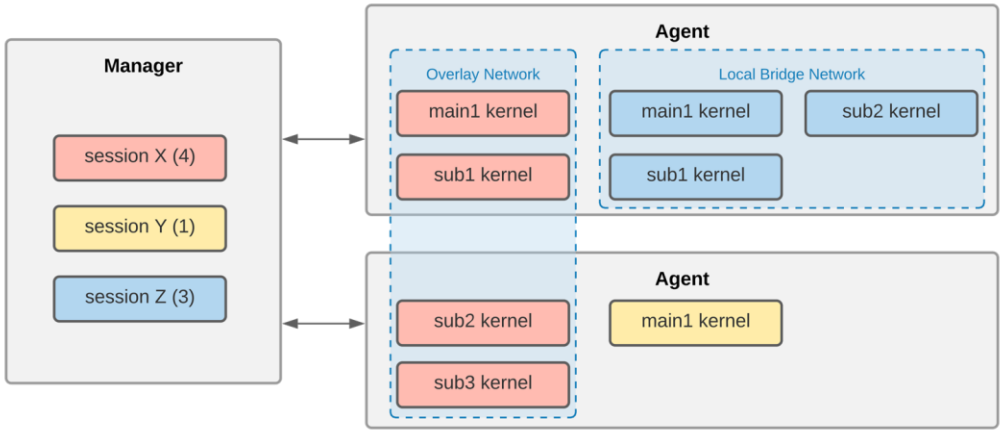
- 고객 혜택
  - 고가의 하드웨어인 GPU 의 사용률 향상을 통한 구입 비용 감소
  - 고가의 훈련용 GPU를 분할하여 추론 GPU로 운영
    - ✓ 노후 장비 활용성 최대화 (훈련 → 추론용으로 전환)



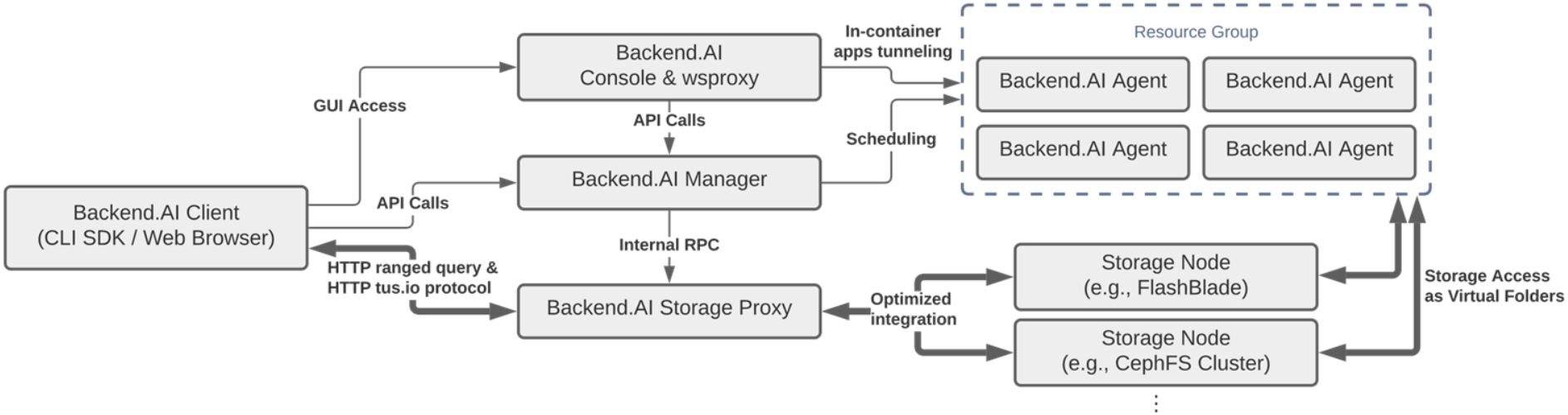
GPU 부분할당 예시

Backend.AI : 클러스터 세션 기반 멀티 노드 분산 처리

- 멀티 노드 워크로드 지원
  - 여러 노드를 필요로 하는 워크로드를 위한 인스턴트 노드 그룹 (클러스터 번들) 생성
    - ✓ 분산처리 딥러닝 훈련 및 빅 데이터 분석 플랫폼용
  - 독자적인 네트워크 주소 할당 및 자체 환경 변수 노출
  - 클러스터 번들 (19.09~), 세션 번들 (20.09~) 지원



Backend.AI : 스토리지 프록시 설계 w/FlashBlade



FlashBlade integration

- RapidFile Tools : 고속 스캔 및 사용 통계
- Purity REST API : 파일 시스템 당 성능 지표 측정
  (IOPS, I/O latency, I/O throughput)

Unified Fast File and Object Storage(UFFO)

- 업계 최초 고성능 파일 & 오브젝트
- 모든 비정형 데이터 워크로드 통합
  자유로운 접근성(S3/NFS/SMB)
- 유연한 성능 확장
- 관리 및 운영 효율성
- TCO, ROI 절감
- 완벽한 데이터 보호(Safe Mode)
- 클라우드로의 쉬운 연계

AI 레퍼런스 아키텍처

고성능 통합 파일/오브젝트

실시간 데이터 분석

FLASHBLADE

업계 최초 데이터 허브 전용 고성능 데이터 서비스 플랫폼

블레이드 기술

탄력적인 데이터 처리 및 강력한 성능의 스토리지 유닛

Purity FB 3

분산 소프트웨어 기반의 무제한 성능/용량 확장

스케일-아웃 패브릭

소프트웨어 정의 패브릭을 통한 비즈니스 규모에 따른 선형적 확장

Thank You!

Contact us to continue the conversation

김태훈 부장
Unstructured Data Specialist
thkim@purestorage.com
Pure Storage Korea

신창호 대표
CEO
chshin@imgr.co.kr
IMGURU

김정묵 운영총괄
COO
jmkim@lablup.com
Lablup

공식 웹사이트	www.purestorage.com/kr
공식 유튜브	www.youtube.com/c/PureStoragekr
공식 페이스북	www.facebook.com/purestoragekorea
네이버 블로그	blog.naver.com/purestorage_korea