TEN

회사소개서

Make AI accessible for all.

회사명	주식회사 텐 TEN Inc.
대표이사	오세진
설립일	2020.07.03
주소	서울특별시 강남구 역삼동 테헤란로 146 현익빌딩 1203호
주요 서비스	MLOps 솔루션 'AI Pub' AI 인프라 컨설팅 'RA:X'
직원 수	21명
홈페이지 / SNS	https://ten1010.io/ https://www.youtube.com/@ten1300 https://www.linkedin.com/company/ten1010/

2020년 서비스를 시작하여
각종 사업 선정과 어워드 수상으로 경쟁력과 잠재력을 인정받고 있습니다.

2020

02 AI Pub 서비스로 비즈니스 시작

07 주식회사 텐 법인 설립

2021

03 NVIDIA Inception Program 체결

04 Kingsley Ventures Seed 투자 유치

05 NetApp 공식 파트너쉽 체결

06 벤처기업 인증 획득

06 N&UP 프로그램[1] 선정

07 TIPS[2] 창업성장기술개발사업 선정

08 기업 부설 연구소 설립

08 대한민국 리딩기업대상 K-스타트업대상 'AI 서비스 부문' 수상

11 NVIDIA GTC[3] Top Startup from Korea 선정 및 참가

12 AI 바우처 사업 공급 기업 선정

2022

01 2021 글로벌기업 협업프로그램 글로벌 스케일업 IR 데이 우수상 수상

03 AI 바우처 사업 선정

05 Coaster Auth 오픈소스 프로젝트 공개

07 대한민국 리딩기업대상 K-스타트업대상 'MLOps 플랫폼 부문' 2년 연속 수상

08 창업진흥원 초기 창업 패키지 선정

11 한국평가데이터 기술역량 우수기업 인증 (T4 등급)

12 Pre-series A 라운드 투자 유치 (어센도벤처스, 퀀텀벤처스코리아, 신용보증기금)

2023

01 가상화 리소스 관리, 크기 추천 방법 등 특허 5개 등록 (10-2488614, 10-2488615, 10-2488618, 10-2488619, 10-2488620)

02 제1회 AI 워크샵[4] 개최

07 NetApp Preferred Partnership 체결

08 IBM Partnership 체결

09 Redhat Partnership 체결

11 한국국방과학연구소 딥러닝 연산용 고밀도 GPU기반 클러스터 컴퓨팅 시스템 사업 구축 완료

12 AI Pub GS 인증 획득

12 CIO Review '2023년 가장 유망한 한국 테크기업' 선정

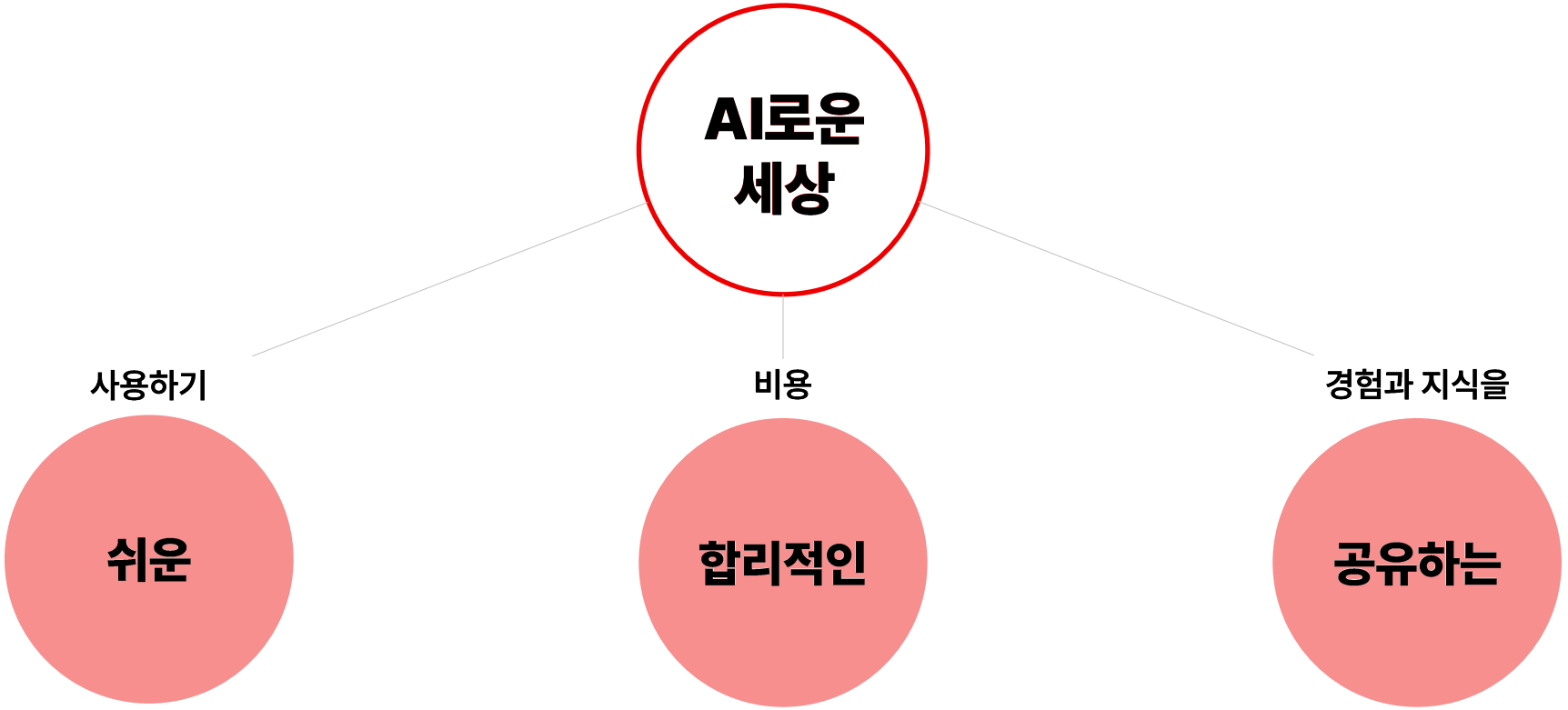
1) N&UP 프로그램 : NVIDIA와 중기부가 협업하여 기업의 글로벌 진출을 지원하는 사업
2) TIPS : Tech Incubator Program for Startup
3) NVIDIA GTC : NVIDIA GPU Technology Conference
4) 제1회 AI 워크샵 : 엔비디아 GPU 기반 AI 인프라 구축과 MLOps 소프트웨어 워크샵

TEN은

세상을 널리 AI롭게 Mission **하기 위하여,**

기술의 장벽을 낮추는 Vision **도구를 만듭니다.**

AI로운 세상에서는, 누구나 AI를 통해 가치를 생산하고 혜택을 받을 수 있습니다. TEN은 AI로움을 위해 보다 합리적인 비용의 사용하기 쉬운 AI 개발 도구를 제공하며, AI에 관한 경험과 지식을 공유할 수 있는 사회를 추구합니다.



워크샵 (23.2)
NVIDIA GPU 기반 AI 인프라 구축과 MLOps 소프트웨어 워크샵

웨비나1 (22.4)
쿠버네티스 환경에서 AI 서비스 및 MLOps 플랫폼의 데이터 관리를 손쉽게 하는 방법

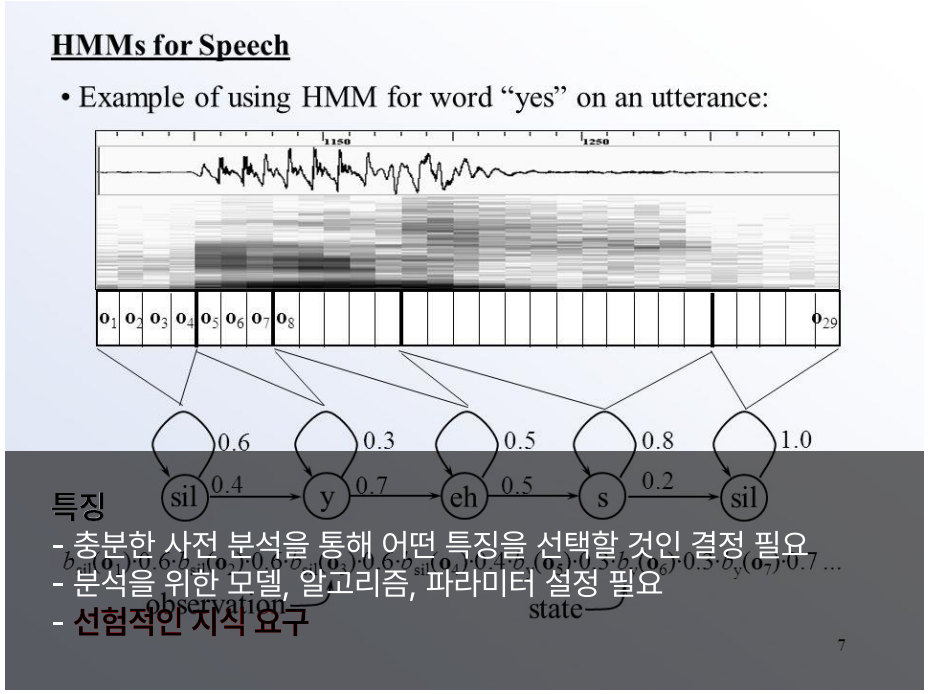
웨비나2 (22.4)
NVIDIA HyperScale AI 인프라 환경과 쿠버네티스 기반 AI 개발 및 운영 환경 구축

딥러닝 시대의 AI,
어떻게 준비하고 계신가요?

지금은 딥러닝의 시대로, 개인부터 기업까지 사회 구성원 모두가 AI를 이야기하고 있습니다.
누구나 자신의 지식을 데이터로 구체화하고, 이를 학습시켜 자동화할 수 있게 될 것입니다.

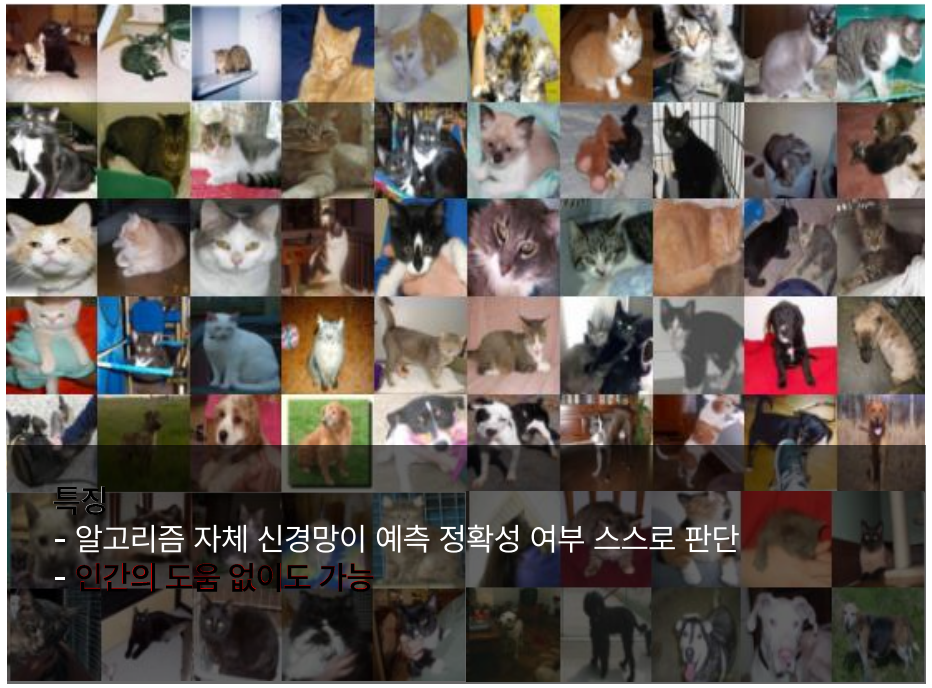
통계적 패턴 인식을 하던 때와는 달리,
물리적 현상을 모델링하기 보다 AI를 만드는 사람의 아이디어가 중요해졌죠.

통계적 패턴 인식



특징
- 충분한 사전 분석을 통해 어떤 특징을 선택할 것인 결정 필요
- 분석을 위한 모델, 알고리즘, 파라미터 설정 필요
- 선험적인 지식 요구

딥러닝



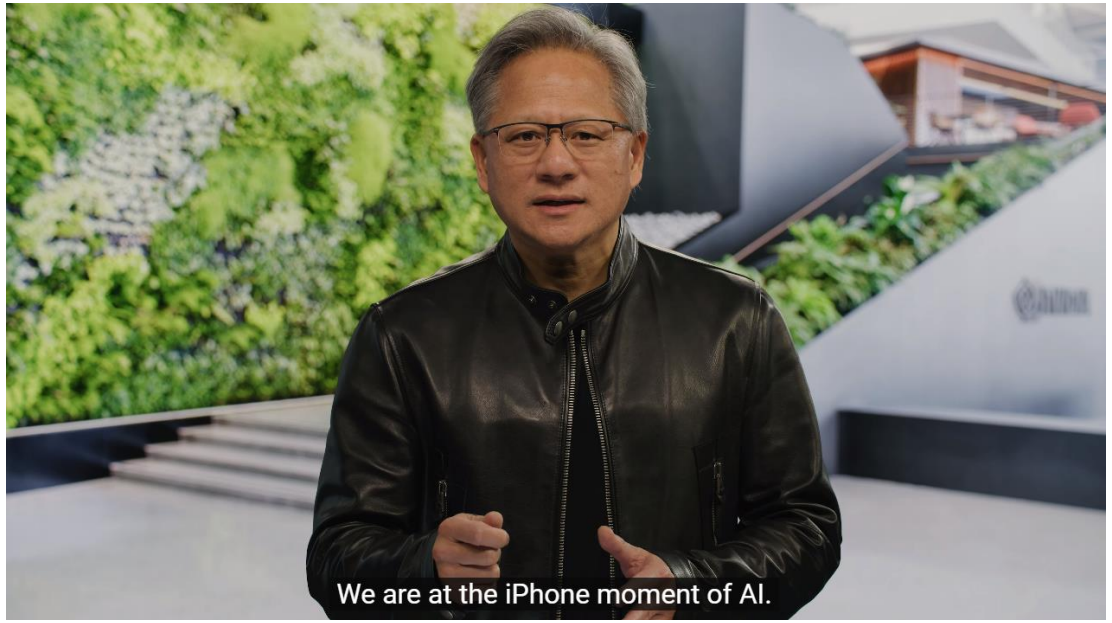
특징
- 알고리즘 자체 신경망이 예측 정확성 여부 스스로 판단
- 인간의 도움 없이도 가능

Building a Cat Detector using Convolutional Neural Networks
— TensorFlow for Hackers (Part III) | by Venelin Valkov | Medium

“우리는 AI의 아이폰 시대 (iPhone moment of AI)에 살고 있습니다.”

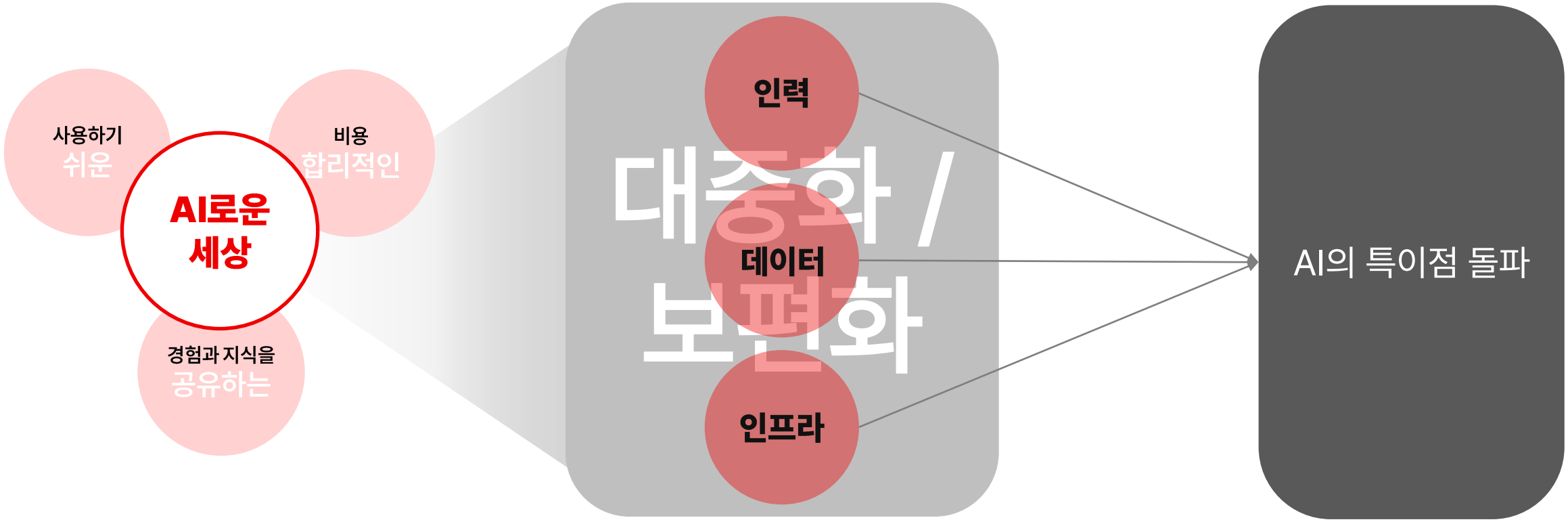
by Jensen Huang

스타트업은 혁신적인 제품과 비즈니스 모델을 구축하기 위해 경쟁하고 있으며, 기존 기업들은 이에 대한 대응책을 모색하고 있습니다.



GTC 2023, Jensen Huang (NVIDIA CEO)

지금 우리는 AI 기술의 특이점을 지나고 있습니다.
그 어느 때보다 AI 관련 인력, 데이터, 인프라의 대중화·보편화가 필요합니다.



특정 지식인이 아닌 모든 이가 AI를 만든다는 인력 보편화의 조건과
자신의 지식을 데이터로 구체화하면 AI를 만들 수 있다는 데이터 보편화의 조건은 점차 해결되고 있지만,
인프라는 철저하게 비용의 문제라 시간과 노력만으로는 해결하기 어렵습니다.

인력

머신러닝 박사 학위가 없어도
AI을 만들 수 있는 시대

데이터

누구나 자신의 지식을
데이터로 구체화하고
이를 학습시킬 수 있음

인프라

인프라 = 비용

딥러닝 출현 후, AI 성능의 패러다임은 컴퓨팅 파워가 주도하고 있습니다.
기존의 통계적 패턴 인식과는 다르게 딥러닝은 컴퓨팅 파워가 성능을 좌우합니다.

즉, 비용을 투자할 수록 성능이 좋아지는 것이죠.
자원 효율화에 대한 필요성은 앞으로도 더 중요해질 것입니다.

통계적 패턴 인식

성능 $\propto$ 컴퓨팅 파워

약한 비례 관계

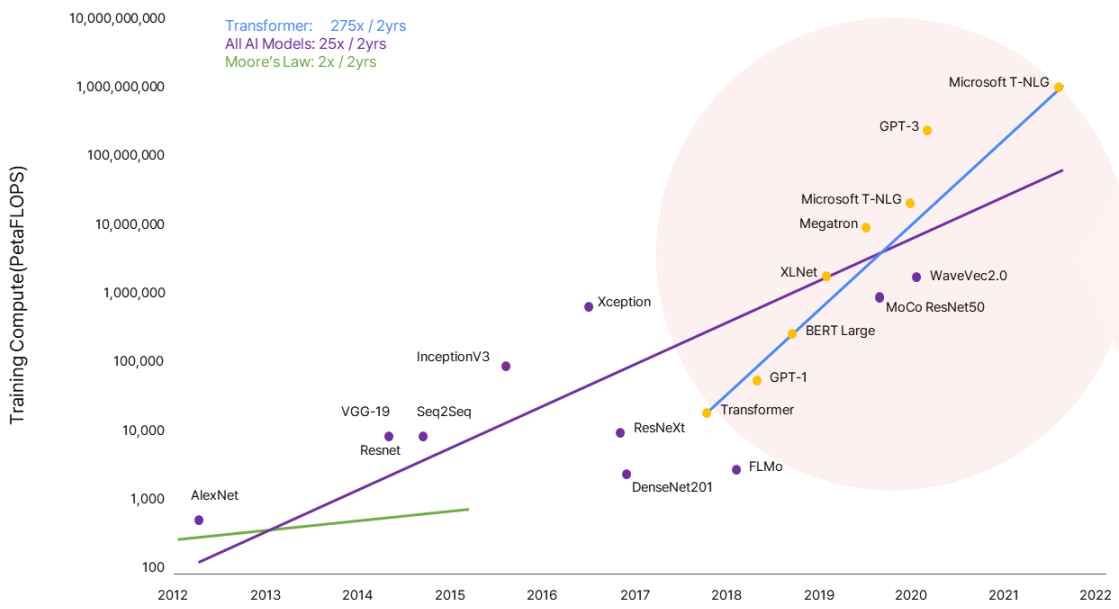
딥러닝

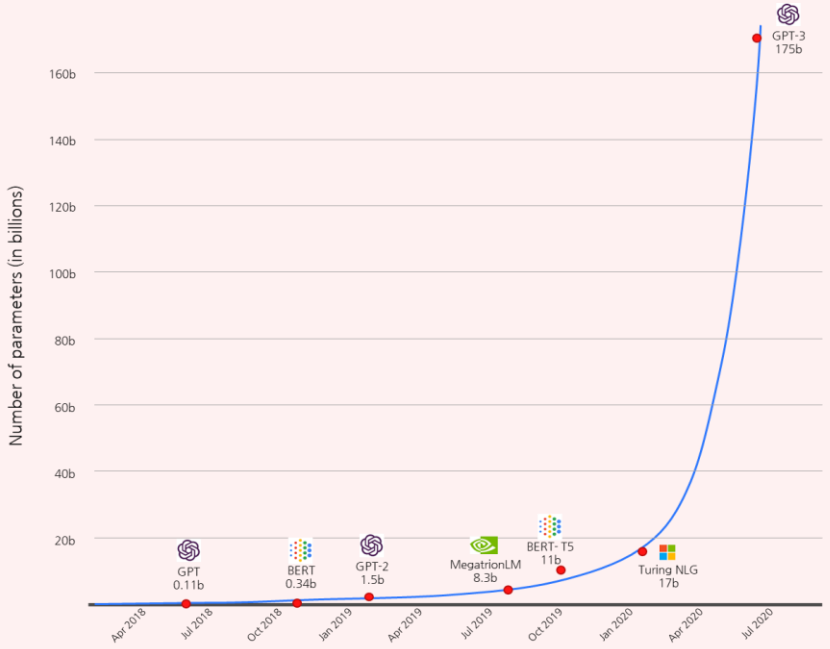
성능 $\propto$ 컴퓨팅 파워

강한 비례 관계

고성능 서버 인프라를 갖추기 위한 노력은 기업의 규모와 업종을 불문합니다. 컴퓨팅 자원 요구량의 증가폭은 선형 스케일에서 더 분명하게 드러납니다. AI의 특이점 도래와 인프라의 중요성을 모두 체감할 수 있죠.

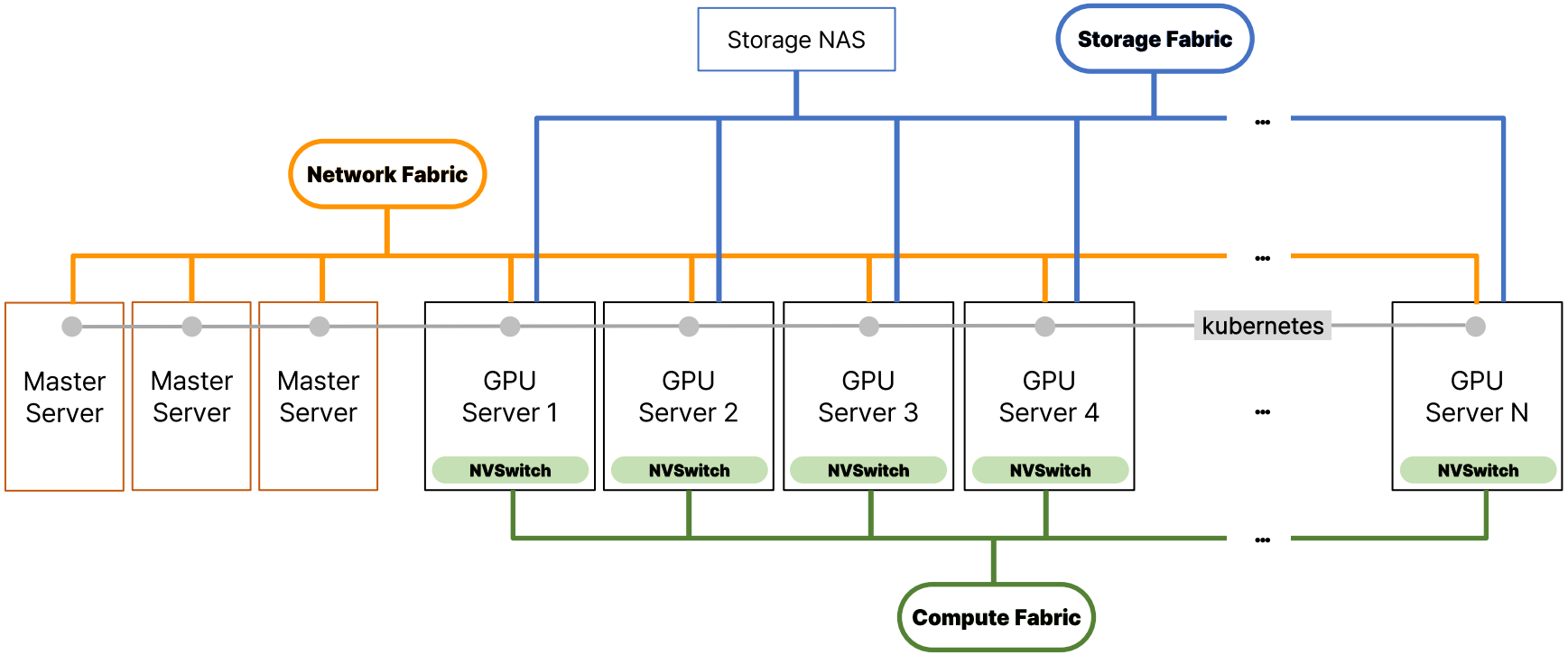
[NVIDIA GTC Keynote, Nov 2021; Log scale]





그러나 인프라 자원을 효율화하는 것은 어려운 일입니다.
AI 모델의 자원 수요가 지속적으로 증가함에 따라, 인프라 구성 역시 복잡해지기 때문입니다.

AI 모델 계산량 증가를 커버하기 위해, 다수의 GPU 머신을 도입할 수밖에 없고
이에 따른 AI 전용 인프라 구성은 더욱 복잡하고 어려워집니다.



인프라 자원의 효율화는
AI 보편화의 핵심입니다.

TEN은,
효율적인 솔루션을 제시합니다.

TEN은 AI를 잘 개발하고 운영할 수 있도록,
MLOps 소프트웨어와 AI 전용 인프라를 서비스하고 있습니다.

SW

Tools

MLOps

HW

Infrastructure

Server, GPU,
Network, Storage

Holistically served by **TEN**

Knowledge

분야전문가 & 데이터

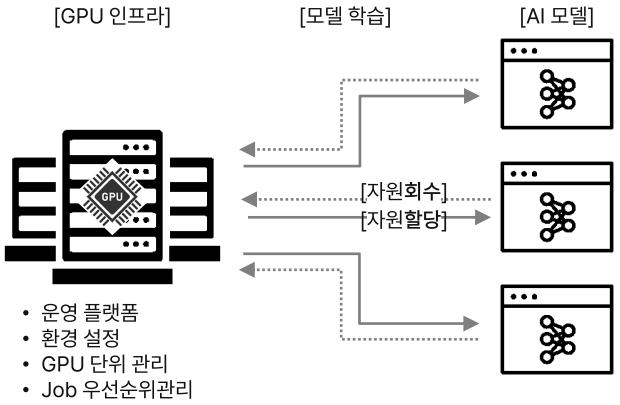
TEN은 AI 모델 학습과 운영 과정에서 직면하는 문제에 대해 인프라 관리 포인트 별로 솔루션을 제시합니다.

AI 모델 학습 과정의 Problem

- 온프레미스 인프라 구축 시
  모델 학습 특성(Resource-hungry)에 최적화된 운영 플랫폼 필요
- 기존 플랫폼 도입 시 서버 단위로 자원을 할당하므로
  GPU 단위의 유휴와 가동률 파악 불가
- 자원 할당 과정에서 서버 설정(OS, Drivers, Libraries등)을
  개발자 별로 변경해야 하는 오버헤드 발생
- 모델 학습 단위의 Job 스케줄링 필요

모델 학습을 위한 솔루션

24 X 7, GPU 자원 가동률 극대화

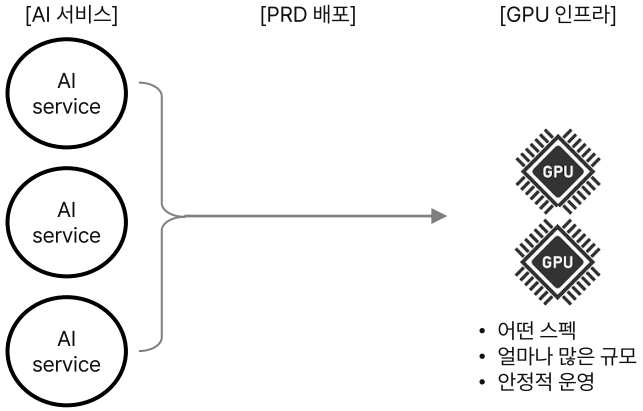


AI 운영 과정의 Problem

- 모델 배포 시 서비스 운영을 위한 GPU 자원 스펙과 규모 산정 불가
- 여러 개의 서비스 운영 시 GPU 자원 내 서비스 간 간섭으로 안정성 확보 불가
- 지속적으로 증가하는 모델의 리소스 사용량에 따른 비용 증가의 문제

모델 운영을 위한 솔루션

최소한의 GPU 사용으로 안정적인 운영 품질 확보



TEN은 컨테이너 플랫폼 COASTER와 MLOps 도구인 AI Pub Dev, AI Pub Ops, AI 인프라 컨설팅 서비스 RA:X를 제공합니다.

컨테이너 플랫폼

COASTER

GUI 기반 AI 개발 / 운영을 위한 MLOps 서비스

AI Pub Dev

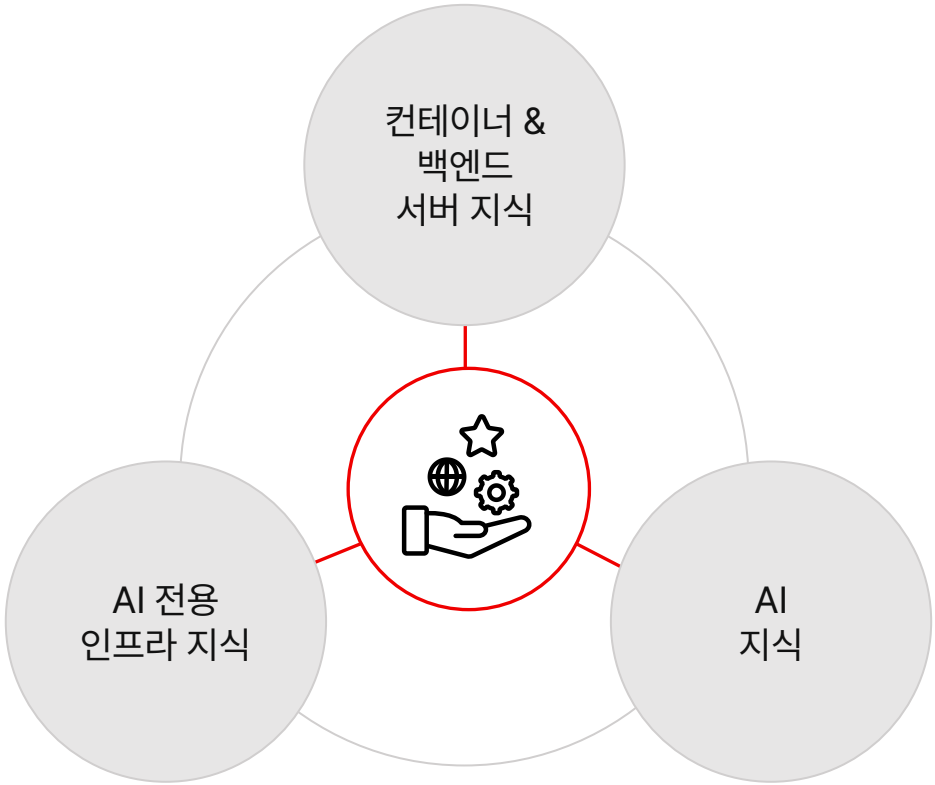
AI Pub Ops

AI 인프라 구축을 위한 컨설팅 서비스

RA:X

TEN's Products

AI 학습과 운영에는 다방면의 전문 지식을 갖춘 전문가가 여럿 필요합니다. TEN은 기업별로 거버넌스를 구성할 수 있도록 지원하는 기능 레벨의 제품을 구성, AI 서비스 운영을 위한 올라운더 전문가를 대체하는 MLOps 도구를 제안합니다.



COASTER

for

Kubernetes에 능숙한 개발자
MLOps 개발자

Command Line Interface

AIPub

for

Kubernetes와 백엔드 개발에
익숙하지 않은 AI 모델러 혹은 AI 서비스 운영자

Graphic User Interface

TEN은 COASTER와 AI Pub으로 AI의 서비스 과정과 고성능 인프라 구성 및 운영까지 모두 고려하는 최적의 솔루션을 제공합니다.

1 Day

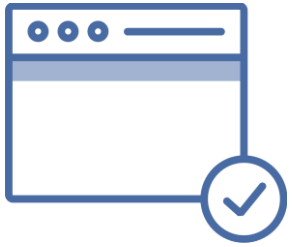
8.6개월이 소요되었던

환경 세팅 - AI 배포까지의 시간 단축

100% util & 1/10 cost

AI 학습단계에는 최고의 가동률로 인프라를 사용하며,

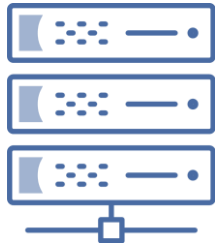
AI 운영단계에는 최소로 인프라를 사용



Off-the-shelf

다양한 분야의 기술 인력을 필요로 하는

복잡한 구축 과정을 생략한 준비된 플랫폼



Tailored

값 비싼 GPU 자원을 최고의 성능으로

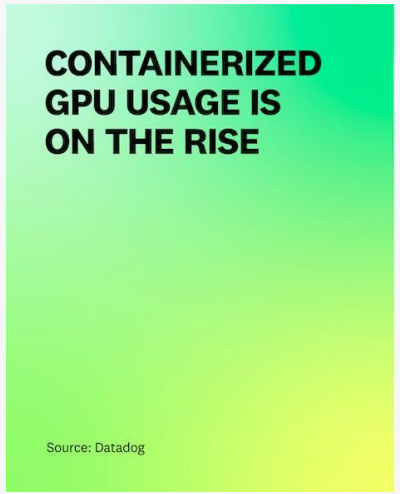
사용할 수 있는 AI 전용 인프라 구성

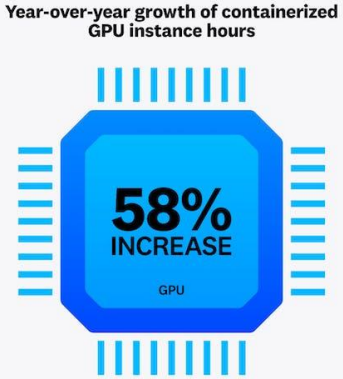
COASTER와 AI Pub은 컨테이너 오케스트레이션의 대세인 Kubernetes를 기반으로 하고 있습니다.

10 INSIGHTS ON
REAL-WORLD CONTAINER USE

FACT 2

Usage of GPU-based compute on Containerized workloads has increased





> We observed a 58 percent year-over-year increase in the compute time used by containerized GPU-based instances (compared to a 25 percent increase in non-containerized GPU-based compute time over the same time period).

10 Insights on real-world container use – Datalog - https://www.datadoghq.com/container-report/

TEN은 많은 고객들의 AI 개발에 대한 목소리에 귀 기울이고, AI Pub을 통해 솔루션을 제시합니다.

" 서버들을 대량으로 구매했는데
이를 필요한 **유관 부서에 나눠주는 관리 오버헤드**가
너무 커서 제대로 운영이 안돼요. "

" 부서 별로 구매한 서버들이 많아지면서
다른 종류의 GPU 서버들을
한번에 관리하기가 어렵고 사용률이 낮습니다. "

" 공과 대학에서 연구지원을 위한 GPU 서버를
구매하였으나 이를 **각각의 연구실에 효과적으로
공유**할 방법이 없어 고민이에요. "

" **여러 사람이 서버를 사용**해야 하는데,
사용자 별 데이터를 관리할 스토리지나 이미지 레지스
트리와 같은 공간을 관리할 방법이 없어요. "

" 이번에 A100 GPU를 구매했는데
MIG 기능을 매번 설정하기가 너무 어렵고
이를 개발자들에게 공유하기도 벅찹니다. "

" 도커 이미지를 기반으로 개발 산출물을 관리하는데요.
각 사용자들이 사용하는 이미지들과 관리자가 팀과
공유해야 할 **이미지들을 일일이 관리**하기 어려워요. "

TEN은 많은 고객들의 AI 운영에 대한 목소리에 귀 기울이고, AI Pub을 통해 솔루션을 제시합니다.

" 음성 합성 시스템을 도입하였으나 **GPU 서버에서 운영해본 경험이 없어** 서비스하지 못 하고 있어요. "

" AI 서비스를 개발하였으나 **운영 인력과 GPU 서버 운영 노하우**가 없어 안정적인 운영이 어려워요. "

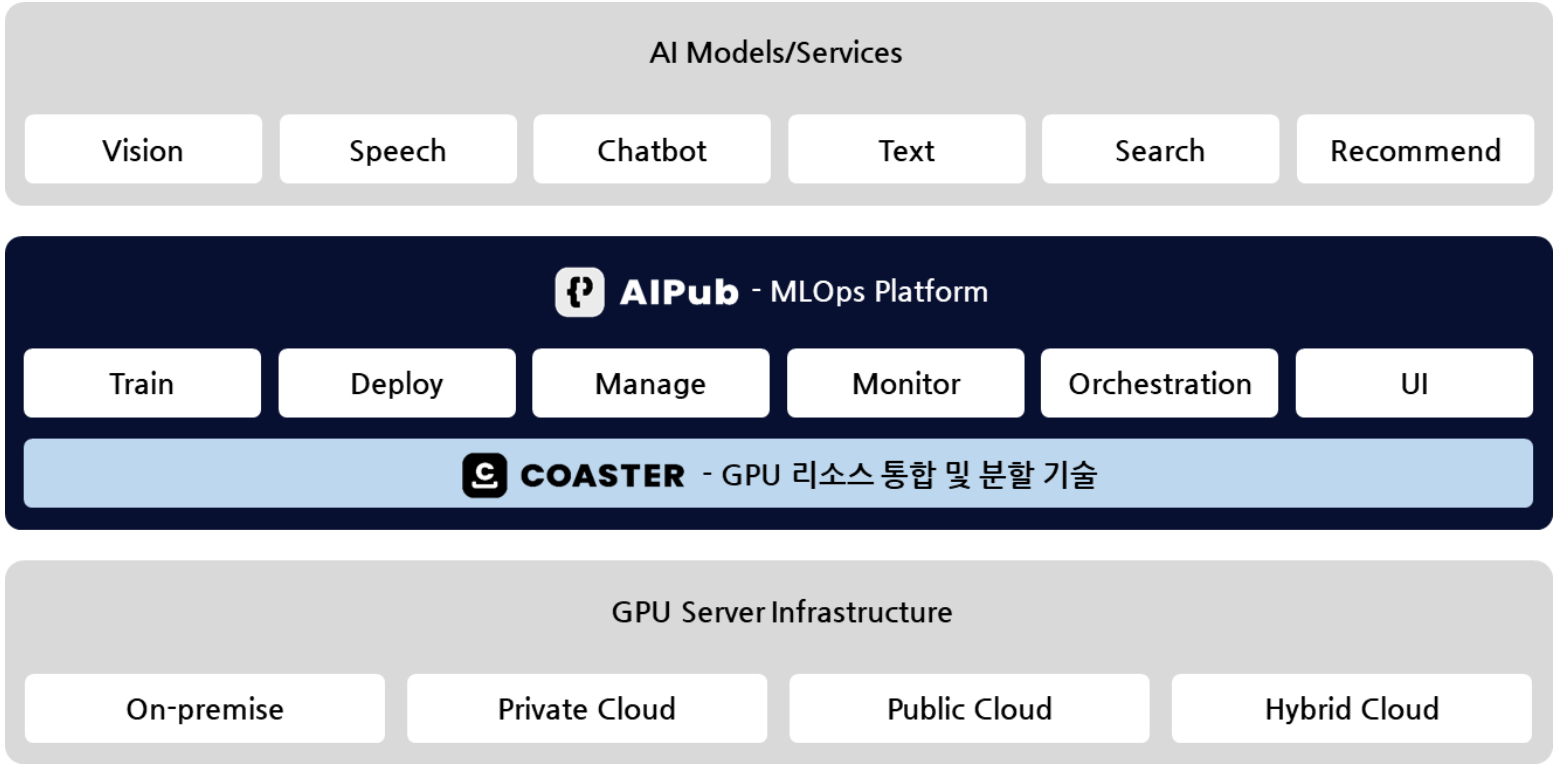
" 서비스를 운영하는 운영자나 **팀 별로 서버 리소스를 구분하고 관리해야 할 필요성**이 있는데 방법이 없습니다. "

" 제품 결함을 검출하는 AI을 개발했는데요. **자동으로 운영하고 관리**하는 소프트웨어는 없을까요? "

" 서비스의 종류도 다양하고 각각의 서비스 버전에 따라 **연동된 도커 이미지도 다르다**보니 이를 관리할 방법이 없어요. "

" **서비스 별로 트래픽에 대한 모니터링**이나 응답이 실패하는 경우를 확인하고 관리할 필요성이 있는데 좋은 방법이 없을까요? "

AI Pub은 AI Lifecycle의 운영과 인프라 활용 문제 해결에 집중하는 완전 관리형 솔루션입니다. 다양한 인프라 환경을 기반으로 어떤 AI 모델에도 서비스할 수 있도록 고안되었습니다.



AI 전문 밴더들과 파트너십을 통해
AI 전용 서비스를 올인원으로 제공하고 있습니다.

GPU Partnership

NVIDIA

HPC Storage technical support Partnership

IBM

Preferred Partners

NetApp®

Security Partnership

SUSE

TEN

Kubernetes Partnership

Red Hat

TEN Solution Partnership

NVIDIA NPN Partner

BN I&C

NVIDIA NPN Partner

Uclick

BOKYUNG

DDN Storage 총판

Open Network SYSTEM

NVIDIA NPN Partner

iCRAFT

PLATFORMIT

NVIDIA NPN Partner

MAYMUST

NVIDIA NPN Partner

MKIS CORE

MAVERICK INFORMATION SYSTEMS CORE

SPK

DWcts

에스넷시스템

INTECH & Company

NVIDIA NPN Partner

GOODUS

JPLANIT

다양한 분야의 기업 및 학교에서 TEN의 서비스를 이용하고 있습니다.

[Client] 기업

국방과학연구소 AGENCY FOR DEFENSE DEVELOPMENT

세종특별자치시

LG전자

KB국민카드

한화시스템

maum.ai

AI Inventech

basgenbio

JPLANT

EJN

[Client] 교육

연세대학교 YONSEI UNIVERSITY

서강대학교 SOGANG UNIVERSITY

한양대학교 HANYANG UNIVERSITY

성균관대학교 SUNGKYUNKWAN UNIVERSITY

[PoC]

kt cloud

HYUNDAI MOBIS

SAMSUNG DISPLAY

사람들이 Pub처럼 편하게 모여 AI를 개발 및 운영할 수 있는,
대중적인 AI 배포 도구를 생각했습니다.

AI Pub
AI Public
AI Publish

COASTER

Kubernetes를 확장하여 기능을 강화한 컨테이너 플랫폼

TEN

COASTER는 Kubernetes를 확장하여 GPU 인프라 운영과 사용자 관리 기능을 강화한 컨테이너 플랫폼입니다.

COASTER

주요 서비스	서비스 상세
GPU 자원의 분할 사용	GPU 1개의 Utilization과 Memory를 100개 블록으로 나누어 활용
GPU 자원의 조회와 할당	Kubernetes의 확장 명령어로 클러스터 전체의 컴퓨팅 자원 조회
User 권한 관리 – Group	리소스 접근 권한을 사용자 그룹단위로 설정 및 관리
스케줄러 대기열 관리	작업 대기열 상의 우선 순위 변경

COASTER 주요 기능 1.

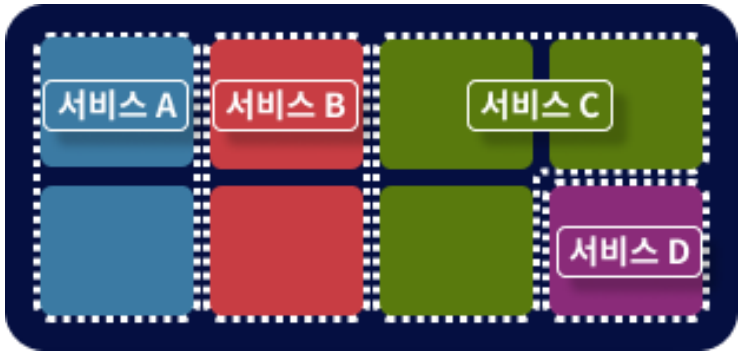
GPU 자원의 분할 사용 기능을 지원합니다.

GPU를 Block 단위로 컨테이너에 할당하게 되면, GPU 1개에 여러 컨테이너를 띄울 수 있을 뿐만 아니라
각 컨테이너 간 리소스 사용량 침해를 막을 수 있어 안정성을 담보해 줍니다.

Kubernetes Native : GPU 할당

GPU 1개 단위로만 컨테이너 할당 가능하여
1개 GPU에 여러 개의 컨테이너를 띄울 수 없음

Coaster Extended : GPU 할당



GPU 1개의 utilization와 memory를
1% 단위로 분할하여 100개 블록으로 나누어 활용 가능

COASTER 주요 기능 2.

GPU 자원의 조회와 할당 기능을 제공합니다.

Kubernetes의 확장 명령어로 클러스터 전체의 컴퓨팅 자원을 조회하고 적절한 타입의 GPU에 필요한 Block 수량을 컨테이너에 할당할 수 있습니다.
각 서버의 GPU 타입을 알 수 없고 서버 별로만 자원 조회가 가능하던 기본 Kubernetes와 다른 사용성을 경험하실 수 있습니다.

Kubernetes Native : GPU 자원 조회

```
Capacity:
  attachable-volumes-aws-ebs:  39
  cpu:                         4
  ephemeral-storage:           83873772Ki
  hugepages-1Gi:               0
  hugepages-2Mi:               0
  memory:                      16093900Ki
  nvidia.com/gpu:              1
```

접속한 노드의 자원 상태 정보만 확인 가능

Coaster Extended : GPU 자원 조회

```
[root@aws-master-01 ~]# kubectl get cr -A
NAME              RESOURCE_NAME                  TOTAL          FREE
block-tesla-t4    ten1010.io/block-tesla-t4      400            130
cpu               cpu                            16000m         9200m
gpu-tesla-t4      ten1010.io/gpu-tesla-t4        4              2
memory            memory                         1600017238Ki   1283834Ki
[root@aws-master-01 ~]#
```

확장 명령어를 사용해 클러스터 전체의 컴퓨팅 자원 조회

Kubernetes Native : GPU 자원 할당

```
apiVersion: v1
kind: Pod
metadata:
  name: gpu-pod
spec:
  containers:
    - name: cuda-container
      image: nvcr.io/nvidia/cuda:9.0-devel
      resources:
        limits:
          nvidia.com/gpu: 2 # requesting 2 GPUs
```

GPU 단위로 할당

Coaster Extended : GPU 자원 할당

```
apiVersion: v1
kind: Pod
metadata:
  name: block-pod
spec:
  containers:
    - name: tensorflow
      image: tensorflow/tensorflow
      resources:
        requests:
          cpu: 2000m
          memory: 4Gi
        limits:
          ten1010.io/block-tesla-t4: 70
```

GPU를 분할하여 할당

COASTER 주요 기능 3.

그룹 단위로 유저 권한을 관리할 수 있습니다.

정책을 공유할 유저와 네임 스페이스, 공유 저장소, 이미지 레지스트리, 서버 노드 등 리소스 접근 권한을
하나의 그룹으로 묶어 한 번에 관리할 수 있습니다.

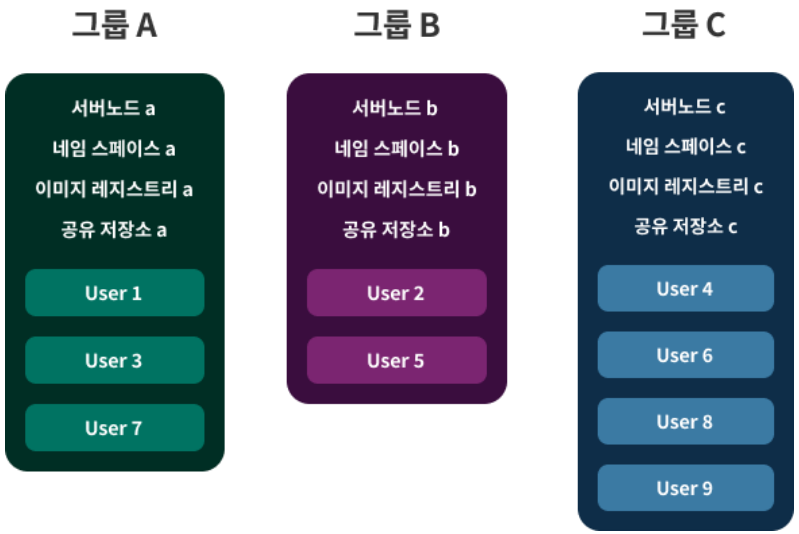
Kubernetes Native : 권한 관리

유저 생성 시 유저별로 모든 리소스 접근 권한을
따로 관리해야 하여 관리가 복잡하고 어려움

Coaster Extended : 관한 관리

하나의 Group에 정책을 공유할 유저와
네임 스페이스, 공유 저장소, 이미지 레지스트리, 서버 노드 등
리소스 접근 권한을 묶어 한 번에 관리 가능

coaster의 오픈소스 프로젝트 : Resource-group-controller



COASTER 권한 관리 예시

그룹 A :
User 1, 2, 3 에게만
노드 a / 네임 스페이스 a / 이미지 레지스트리 a / 저장소 a 접근 권한 부여

그룹 B :
User 2, 5 에게만
노드 b / 네임 스페이스 b / 이미지 레지스트리 b / 저장소 b 접근 권한 부여

그룹 C :
User 4, 6, 8, 9 에게만
노드 c / 네임 스페이스 c / 이미지 레지스트리 c / 저장소 c 접근 권한 부여

COASTER 주요 기능 4.

스케줄러로 Queue에 있는 작업들의 우선순위를 자유롭게 변경할 수 있습니다.

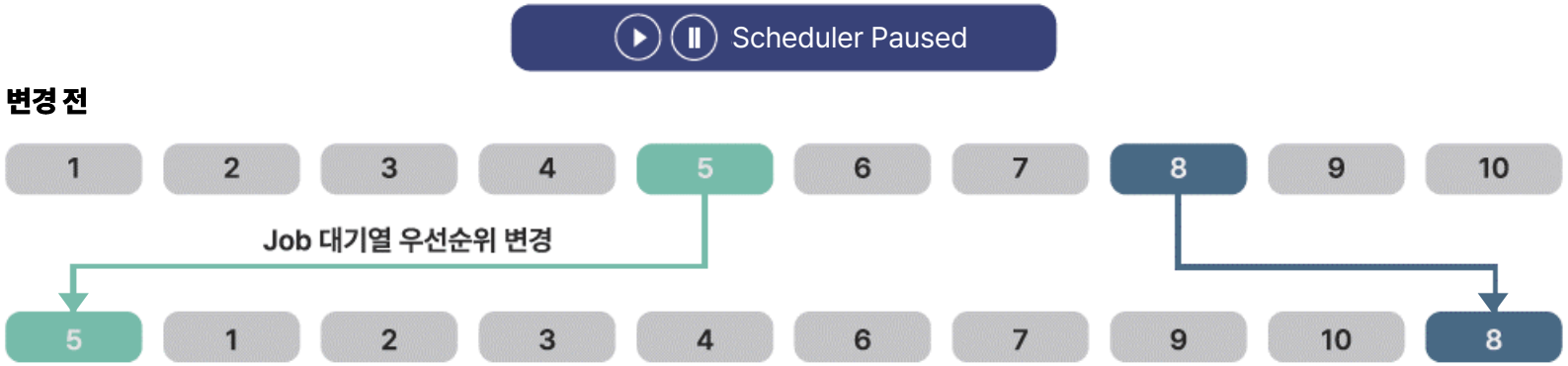
Kubernetes Native : 스케줄러

Kubernetes의 기본 스케줄러는
FIFO(First In First Out) 방식으로 대기열 관리

Coaster Extended : 스케줄러

Queue에 있는 작업들의 우선순위를
자유롭게 변경 가능

COASTER 스케줄러 예시



변경 후 : 변경 전 대기열 5번째, 8번째 작업을 각각 대기열 1번째(기존 5번째 작업), 10번째(기존 8번)로 변경 완료

COASTER의 기능을 DEMO 영상으로 확인해 보세요.

AIPub Dev

COASTER를 코어로 하여 AI 개발을 지원하는 완전 관리형 서비스

AI Pub Dev는 COASTER를 코어로 하여 AI 개발 업무를 지원합니다.
모델 학습, 리소스 및 워크로드 등에 관한 완전 관리형 서비스를 제공합니다.

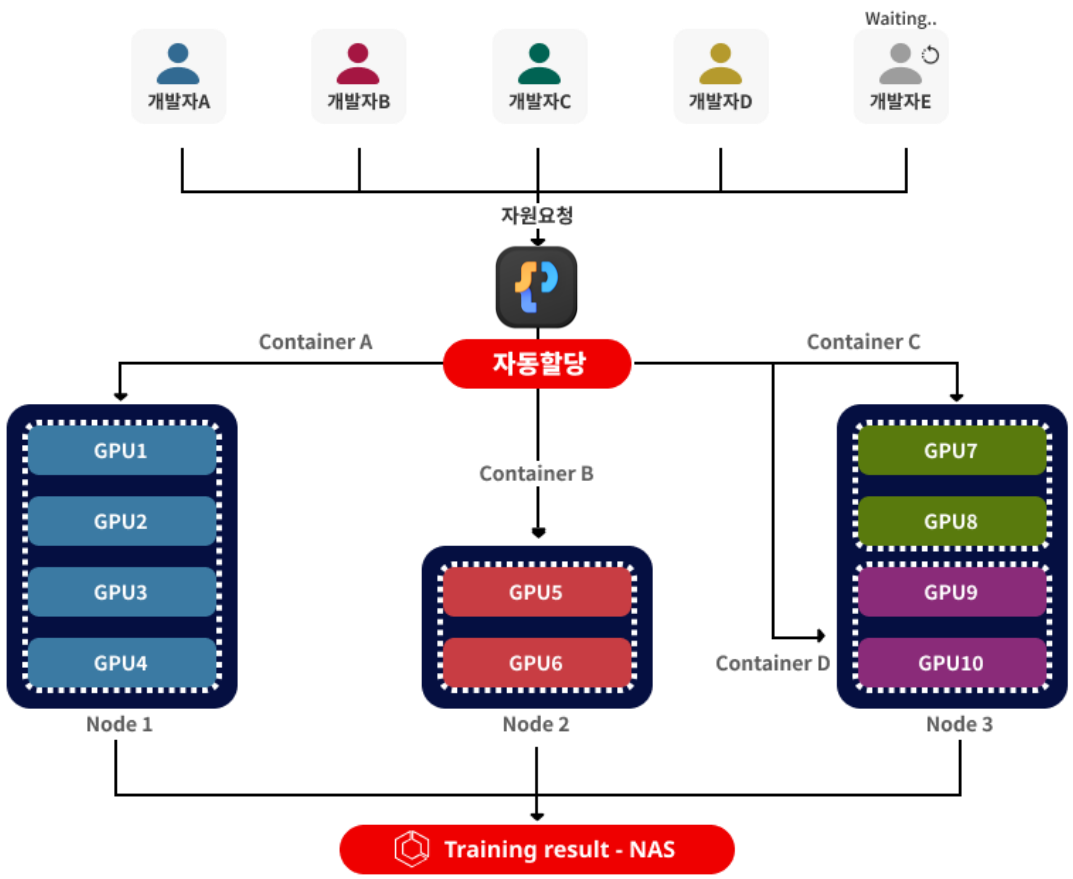
AI Pub Dev

주요 서비스	서비스 상세
워크스페이스 생성	사용자의 개발 환경을 이미지의 형태로 관리
	개발 이미지 기반 워크스페이스 생성
	주피터노트북 및 텐서보드 연동
모델 학습	AI 학습 별로 필요한 자원을 자동으로 할당하여 작업 수행
	GPU 리소스와 CPU 리소스 신청 가능
리소스 관리	사용자 계정 별 리소스 사용 제한 설정
	유휴 리소스 회수
	노드 별 워크스페이스 관리
	노드 별 MIG 설정
	전체 인프라 모니터링
워크로드 관리	스케줄러 멈춤/재개 기능
	대기열 관리 및 우선순위 조절 기능
사용내역 관리	사용자 계정별 리소스 사용 내역 관리
	사용내역 다운로드 기능

AIPub Dev 주요 기능 1.

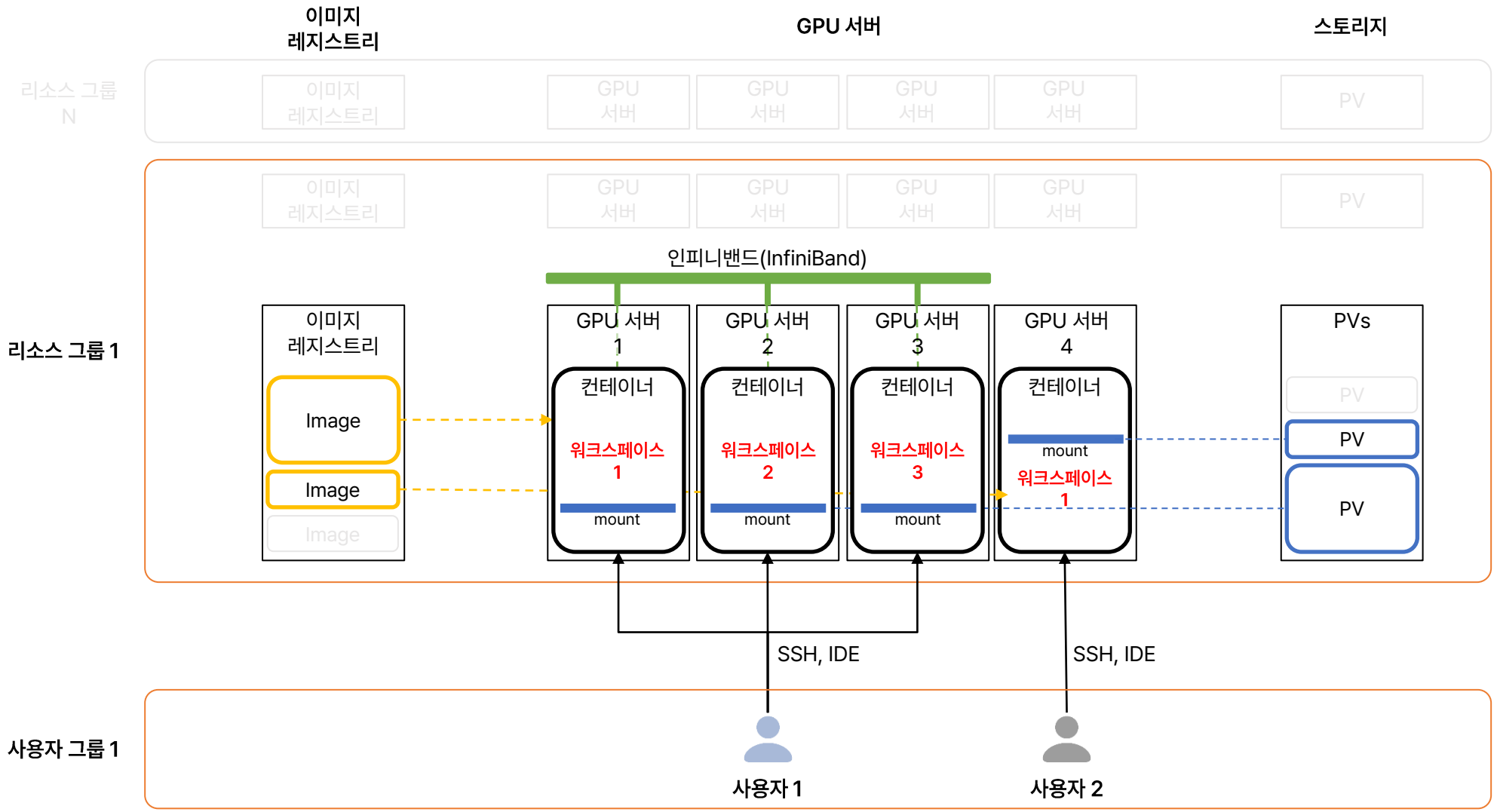
팀 혹은 사용자 단위로 AI 인프라를 할당할 수 있습니다.
팀 개발자들의 자원 사용량 및 팀 별 사용량을 집계할 수 있습니다.

AI 개발용 인프라를 구축하여 중앙 관리할 때에도 자원 할당 및 관리가 가능하고, 각 단위 별 사용량을 측정할 수 있습니다.



AIPub Dev 주요 기능 2.

소속된 그룹에서 접근할 수 있는 리소스들을 통해 워크스페이스를 생성할 수 있습니다.



AI Pub Dev의 기능을 DEMO 영상으로 확인해 보세요.

AIPub Ops

COASTER를 코어로 하여 AI 운영을 지원하는 완전 관리형 서비스

AI Pub Ops는 AI 운영에 필요한 기술들로 구성되어 있습니다.
COASTER와 Kubernetes를 바탕으로, Service Mesh 기능과 Application을 위한 기능을 포함합니다.

Application	학습한 AI 모델을 안정적으로 운영할 수 있도록 서비스화
Service Mesh	서비스 검색, 로드 밸런싱, 시간 초과 및 재시도를 제공하며 관리자가 클러스터의 보안을 관리하고 성능을 모니터링
Kubernetes	GPU가 분할된 기준으로 전체 클러스터 자원을 조회, 할당, 관리
Coaster	GPU를 분할하여 컨테이너에 할당하여 서비스에 필요한 만큼 효율적으로 자원을 사용

AI Pub Ops는 COASTER를 코어로 하여 AI 운영 업무를 지원합니다.
서비스 생성 및 관리, 리소스 등에 관한 완전 관리형 서비스를 제공합니다.

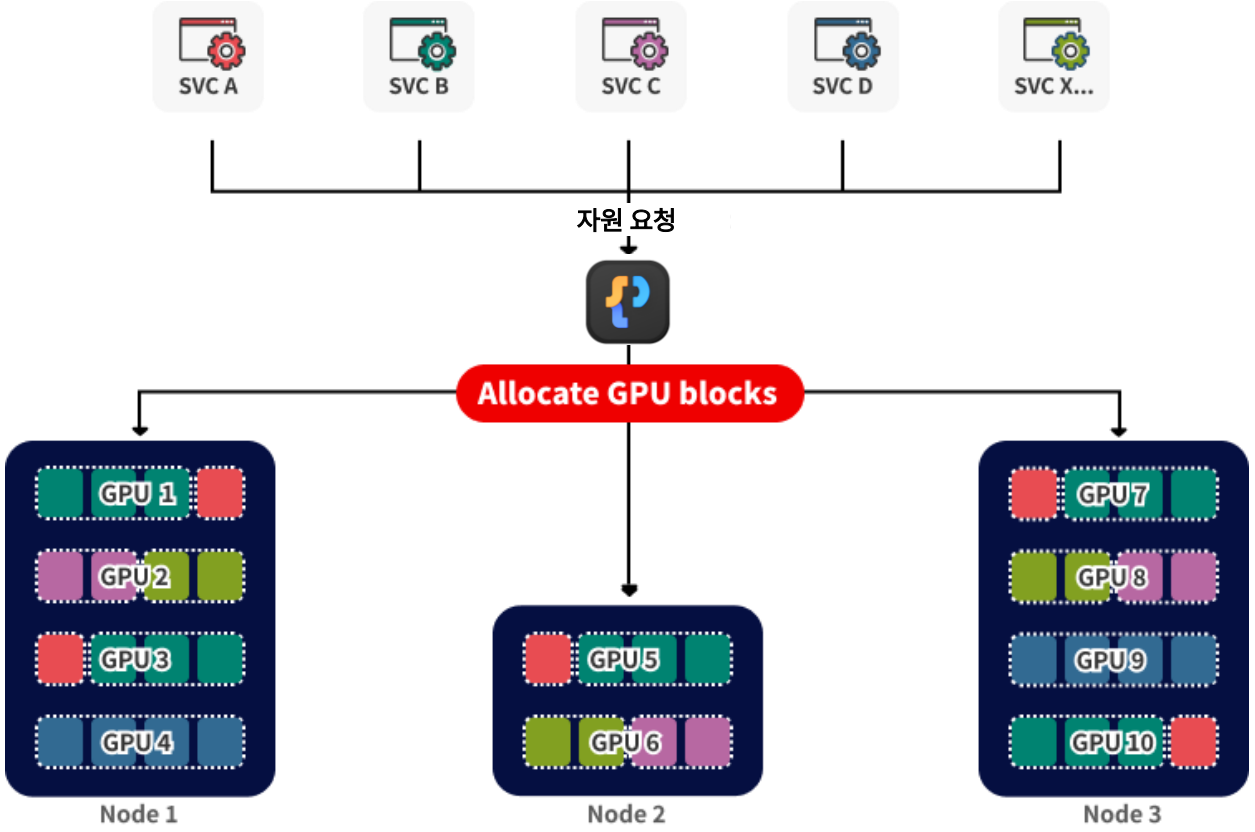
AI Pub Ops

주요 서비스	서비스 상세
서비스 생성 및 업데이트	UI를 통한 서비스 생성 / 중지 / 삭제 및 배포 가능
	UI를 통한 무중단 서비스 업데이트
	버전 관리 및 서비스 롤백 기능
서비스 모니터링	서비스 목록과 서비스 상세를 통한 운영 상태 모니터링
	서비스 장애 시 알림 및 로그 확인을 통한 트러블 슈팅
리소스 그룹 관리	관리자가 리소스 그룹을 생성 및 사용자 권한 설정 기능
	리소스 그룹 편집 기능
리소스 관리	서비스 별 GPU 블록 단위 할당 가능
	GPU 블록 및 서버의 실시간 가동률 모니터링
사용내역 관리	서비스 별 리소스 사용내역 관리
	사용내역 다운로드 기능

AIPub Ops 주요 기능 1.

사용자의 서비스를 이미지의 형태로 관리합니다.
GPU 1개를 100분할 하여 최소 단위로 운영해 비용 절감에 기여합니다.

비개발자도 서비스를 생성하거나 중지, 삭제할 수 있으며, 업데이트 및 롤백도 가능합니다.



AIPub Ops 주요 기능 2.

퍼블릭 클라우드의 서비스 운영 및 관리 기능을 제공하여 온프레미스 서버에서도 AI 서비스 운영을 돕는 기능을 지원합니다.

High Availability
이중화를 통해
Single point of failure 제거

Rolling Update
서비스 중단 없이
상시 업데이트 가능

Load Balancing
바쁘지 않은 서버로
서비스 요청을 자동 분배

Scale-out
서비스 요청에 따라
서버 수를 자동으로 늘려 트래픽 처리

Fail over 대응
서비스 Fail over 시 탐지 기능 및
서비스를 새로 띄워 안정성 확보

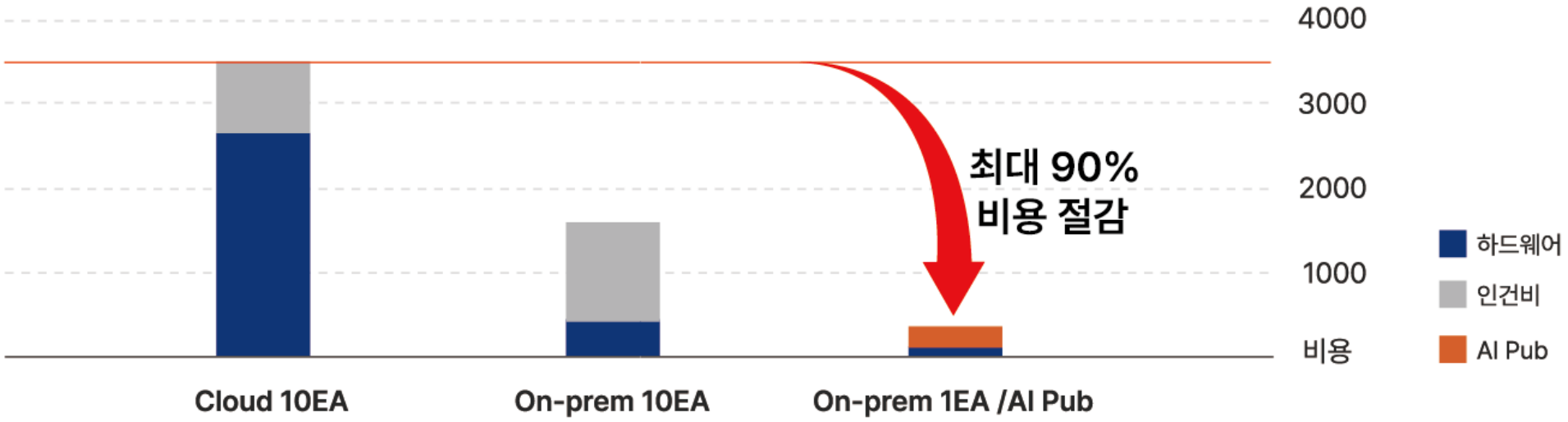
이상징후 알림
중요 운영 이벤트 발생 시
카톡, 슬랙으로 실시간 알림

AI Pub Ops를 활용하여 운영 비용을 최대 90%까지 절감할 수 있습니다.

퍼블릭 클라우드에서 AI 서비스를 운영하는 대신, 같은 급의 서버를 구축 후 AI Pub을 도입하여 비용을 절감할 수 있습니다.
GPU 분할 기능을 통해 서버 자원의 1/10만 활용하여도 서비스 운영이 가능하고,
서비스 운영을 위한 기능들을 직접 개발하거나 유지 보수하지 않아도 되기 때문에 인건비도 절약할 수 있습니다.

[고객사의 AI 서비스 50개 5년 운영을 위한 인프라 구축 비용]

(실제 사례: T4 4EA 장착한 서버 10EA, 중급 개발자 4명 관리 → T4 4EA 장착한 서버 1EA, 관리 인력 없이 AI Pub으로 대체)



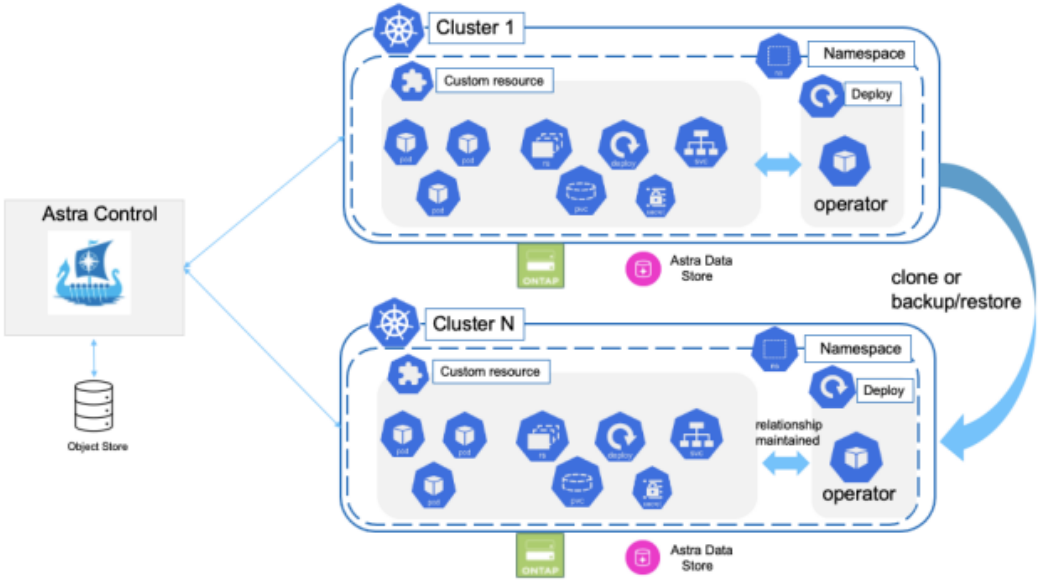
AI 서비스 운영의 어려움 해소를 위해
TEN에서 운영 대행 서비스를 지원합니다.

오랜 기간 쌓은 AI 운영 노하우를 기반으로 서비스를 대신 운영해드립니다.
고객은 지속적인 서비스 안정화를 위해 시간과 비용을 투자하지 않아도 되며, AI 가치 생산에만 집중할 수 있습니다.

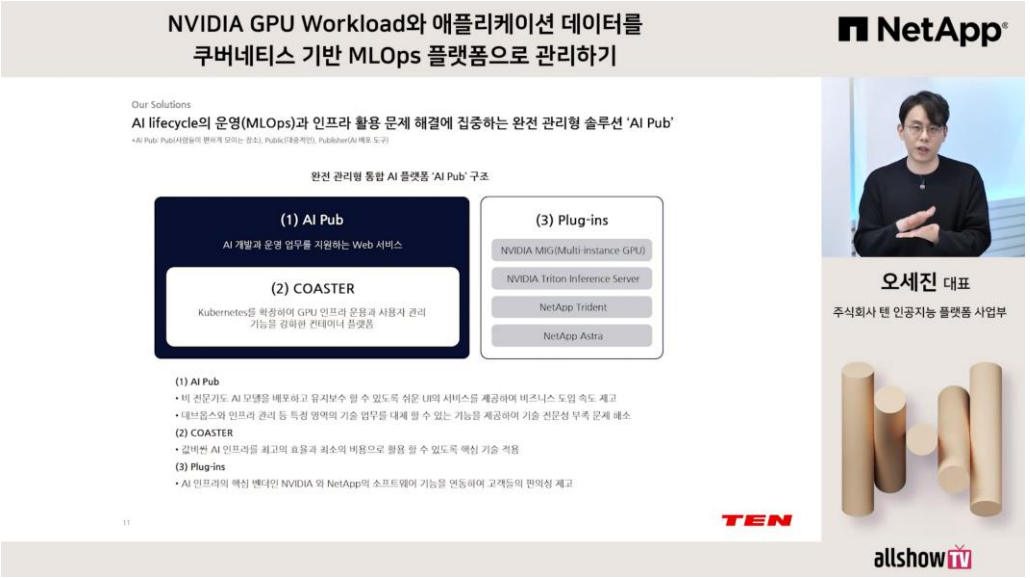
AI Pub Ops

AI Operation by
TEN

AI Pub Ops는 NetApp Astra와 연동되어 있어, 여러 마이크로 서비스로 구성된 어플리케이션의 데이터 관리와 백업, 마이그레이션과 롤백을 지원합니다.



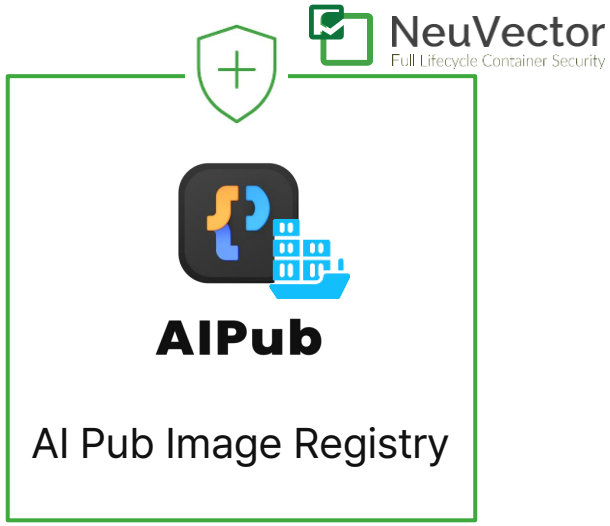
COASTER X NetApp



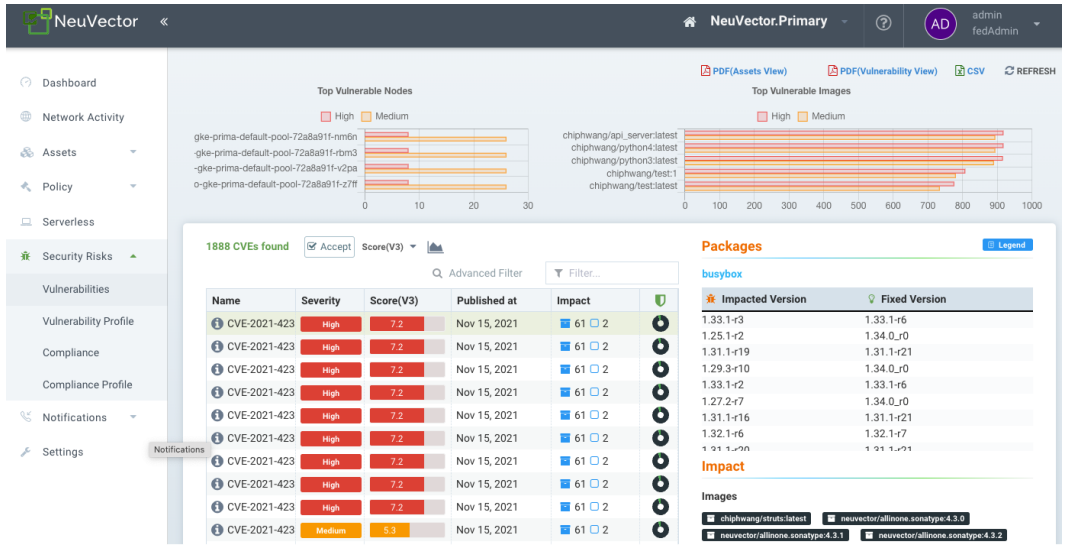
[웨비나]
'쿠버네티스 환경에서 AI 서비스 및 MLOps 플랫폼의 데이터 관리를 손쉽게 하는 방법'

AI Pub Ops는 NeuVector의 스캐닝 기능을 사용하여 다양한 경로로 유입되는 도커 이미지에 대한 보안 필터링이 가능합니다.

컨테이너 이미지, 클러스터를 구성하는 노드, 동작 중인 컨테이너의 취약점을 검사할 수 있습니다.



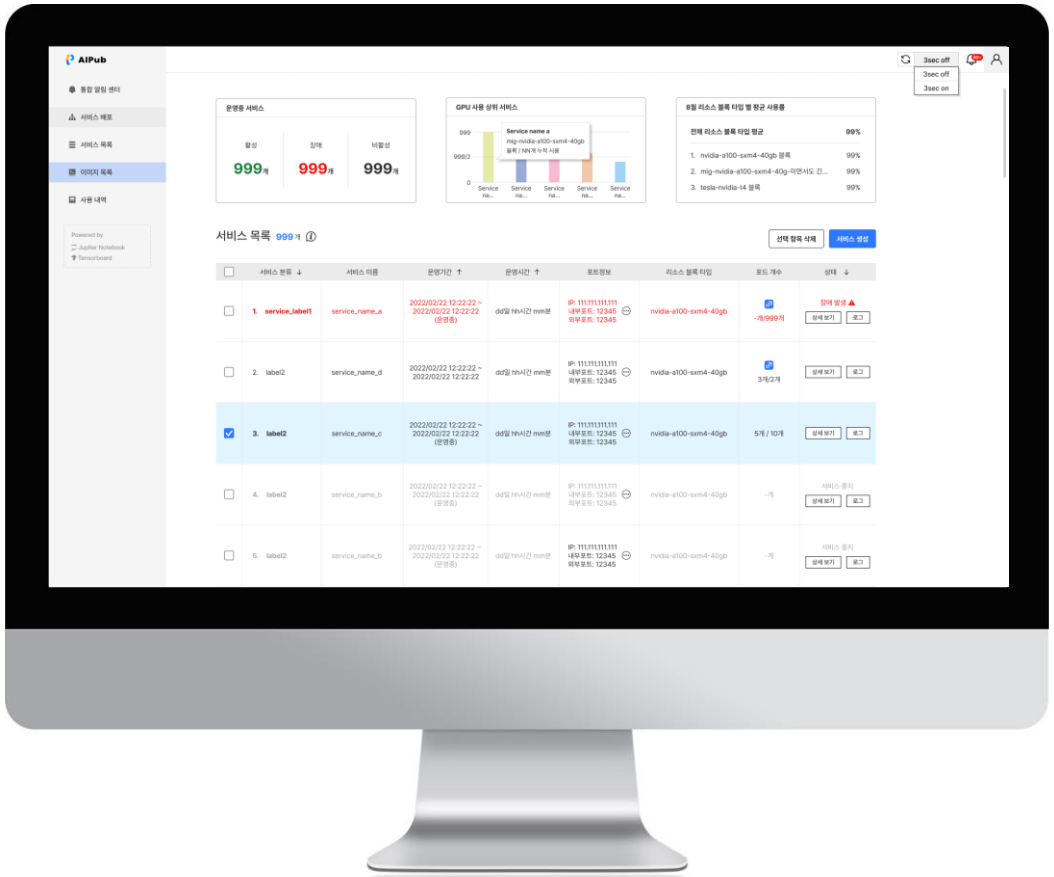
COASTER X SUSE

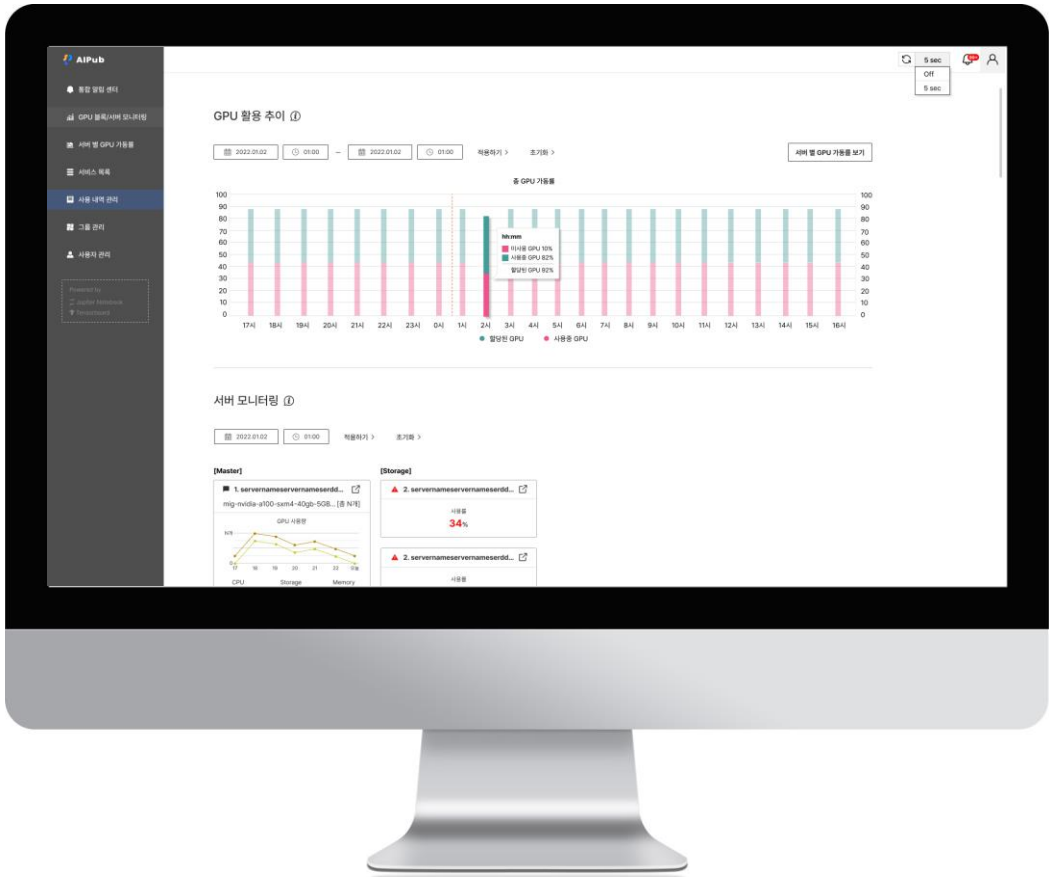


CVE를 기준으로 스캐닝	CVE는 다양한 취약점을 통일해 고유한 넘버링을 부여한 것으로 장치, 시스템, 프로그램 해킹에 악용될 수 있는 컴퓨터 보안 취약성 및 시스템 결함에 코드를 부여하고 관리

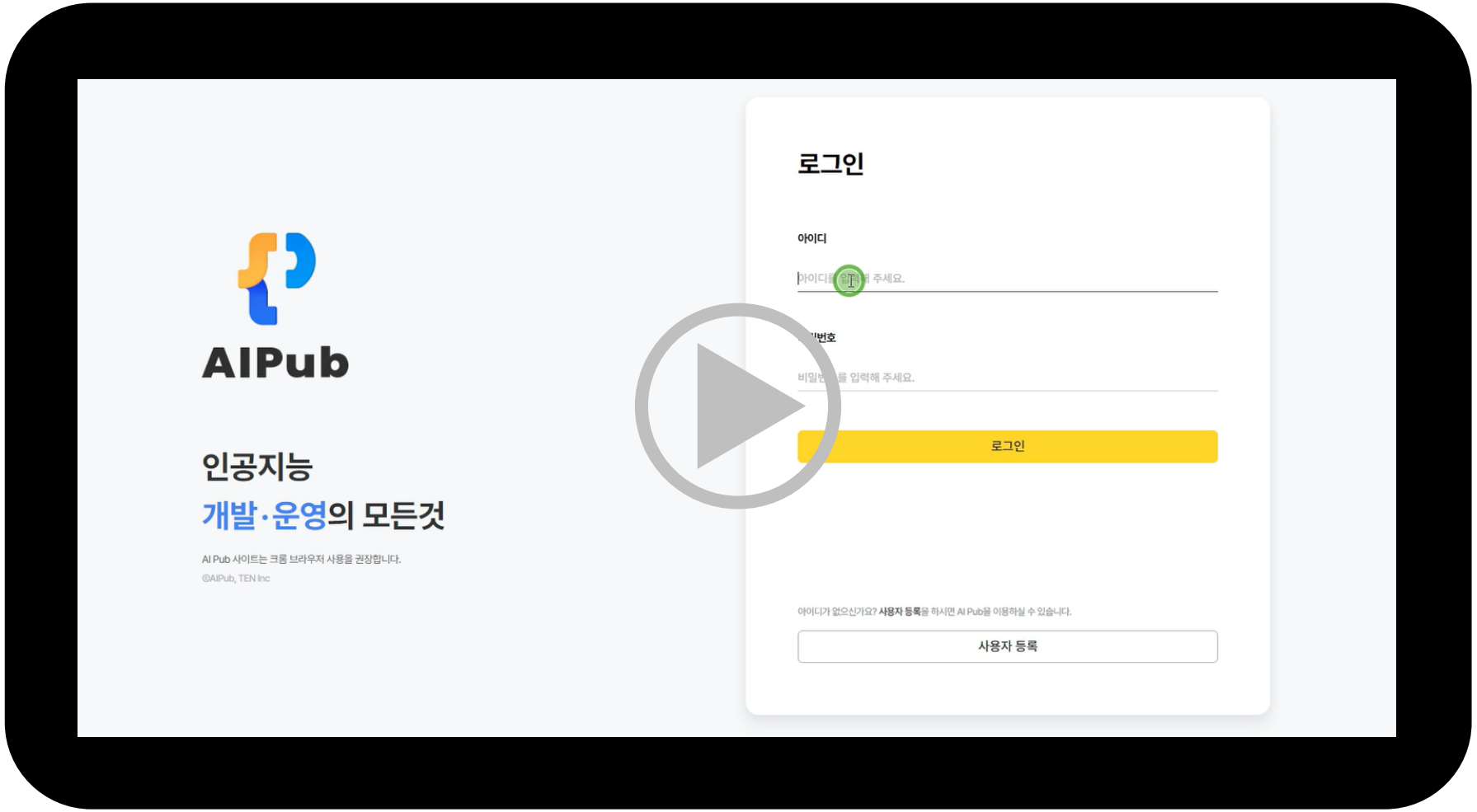
AI Pub Ops는 AI 자원 관리자와 서비스 운영자를 위한 모니터링 서비스를 제공합니다.

서비스와 시스템의 상태를 탐지하고 리스크를 관리할 수 있습니다.





AI Pub Ops의 기능을 DEMO 영상으로 확인해 보세요.



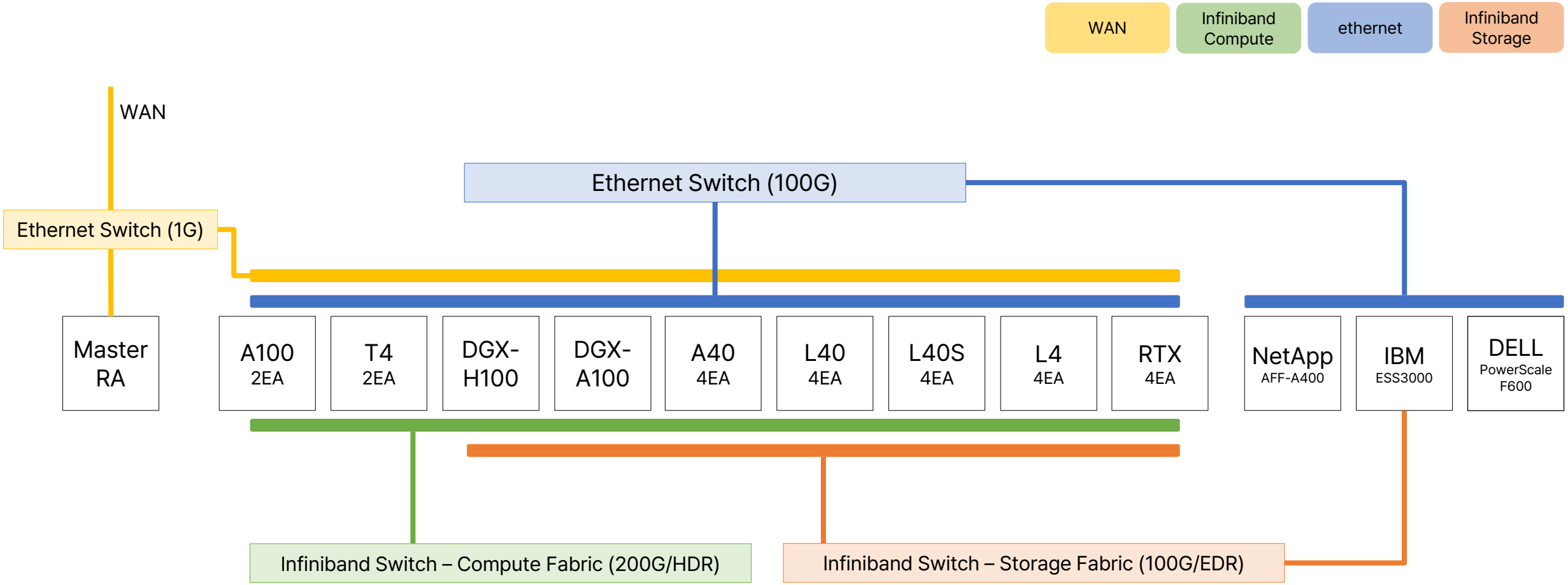
RA:X

AI Reference Architecture of TEN

TEN의 노하우를 더한 테스트 기반의 AI 인프라 구축 컨설팅 서비스

하드웨어를 예산에 맞춰 각각 구입하도록 제안하지 않습니다.
인프라 전체의 가치를 볼 수 있게 하겠습니다.

TEN이 직접 구축하고 운영하는 Reference Architecture을 소개합니다.



RA:X 를 기반으로 한 인프라 구성 컨설팅은 아래의 프로세스로 진행됩니다.

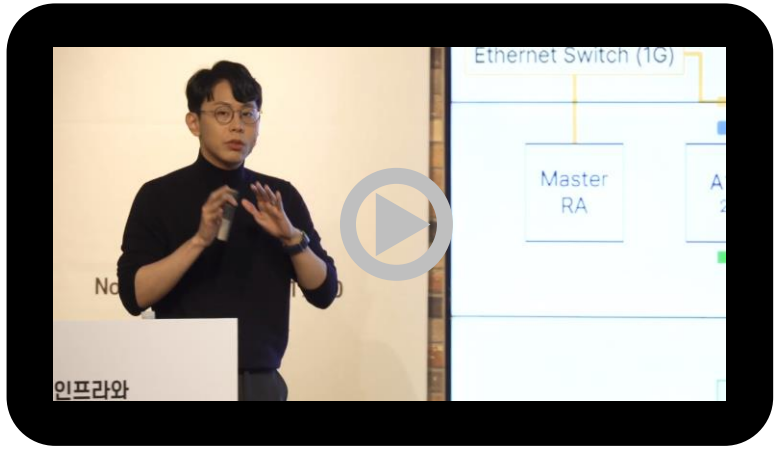
Kick-off (1일) > **Discover** (3-5일) > **Analyze** (5일) > **Report** (1일)

[설문 영역]

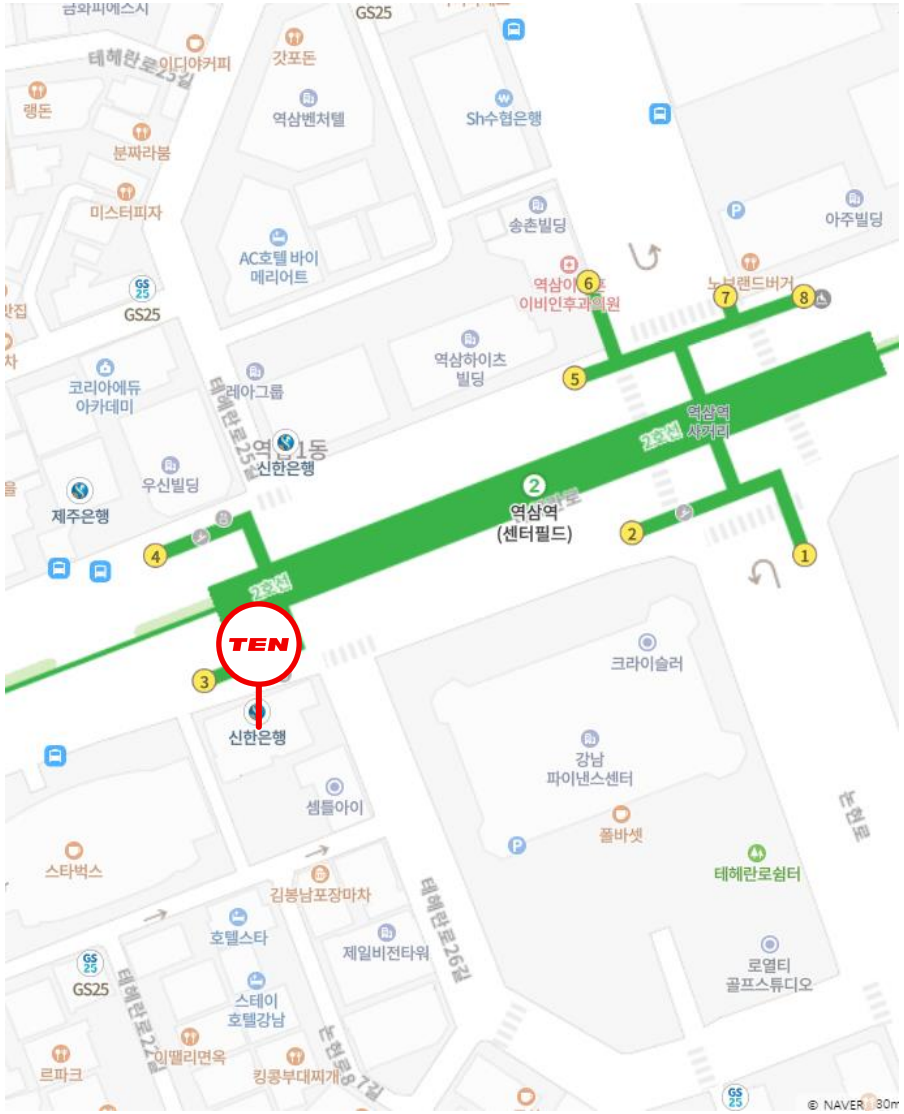
- 비즈니스 전략
- 데이터 사이언스
- 거버넌스
- 애플리케이션 등
- 데이터 엔지니어링
- 인프라 요구사항
- 보안

TEN은 AI 인프라 구성의 어려움과 해결 방안에 대한 세미나, 워크샵을 개최합니다.
인프라에 대한 고민에 공감하고, 해결을 돕기 위한 TEN의 목소리를 다양한 곳에 전합니다.

2023.11
AI 인프라와 NVIDIA 인증 솔루션
통합 구성 워크샵
Opening Session by TEN



2023.11
AI 인프라와 NVIDIA 인증 솔루션
통합 구성 워크샵
Session C, Closing Session by TEN



MORSE ME
WHEN YOU NEED
HELP WITH AI.

경쟁력 있는 AI를 위한 파트너가 필요하신가요?

고효율 솔루션과 신뢰도 높은 파트너십을 갖춘 주식회사 텐과의 협업으로
이전과 다른 AI 개발·운영을 경험하실 수 있습니다.

A 서울특별시 강남구 역삼동 테헤란로 146 현익빌딩 1203호

P +82-2-6956-1071

M helloten@ten1010.io

H https://ten1010.io/

AI로운
세상을 향해.

Make AI
accessible for all

THANK YOU

Copyright © 2024 Ten Inc. All rights reserved