

SAMSUNG SDS

Foresee

Techtonic 2021

Disrupt

Partner



SDS Cloud 환경에서 Kubeflow를 활용하여 쉽게 구현하는 MLOps

정영일 프로

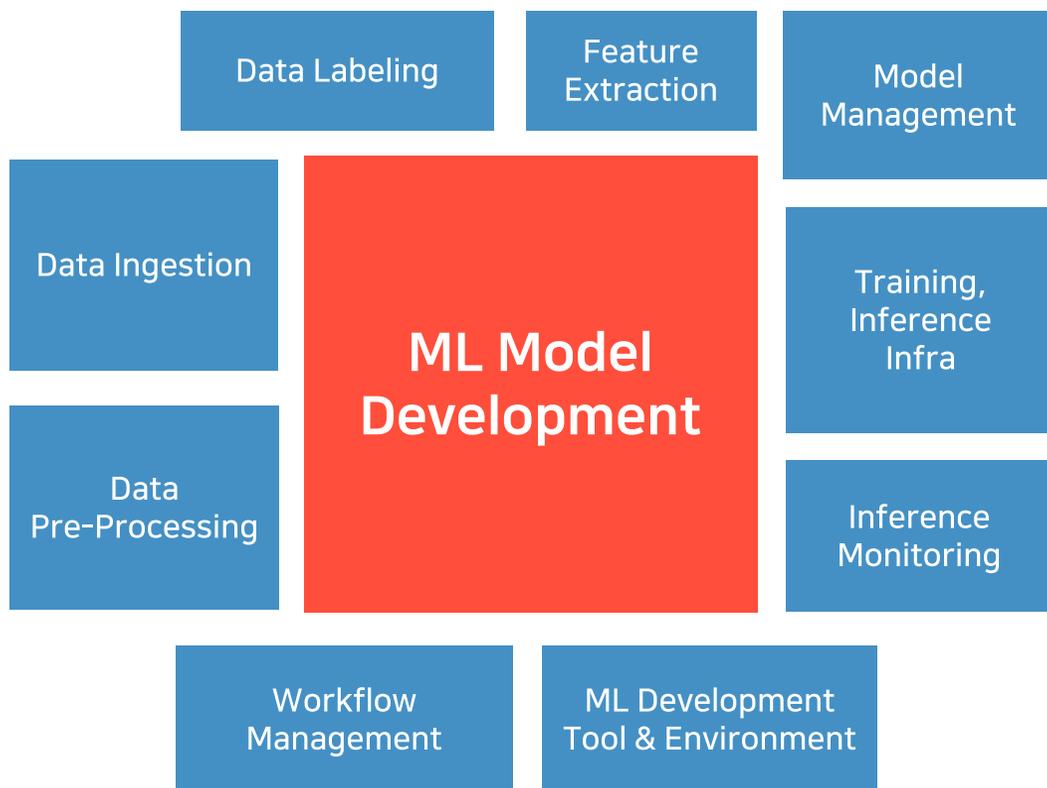
AGENDA

1. Why MLOps?
2. MLOps 오픈소스 Kubeflow
3. SDS Cloud에서 Kubeflow 구성해 보기
4. 향후 계획 : Kubeflow Enterprise Service

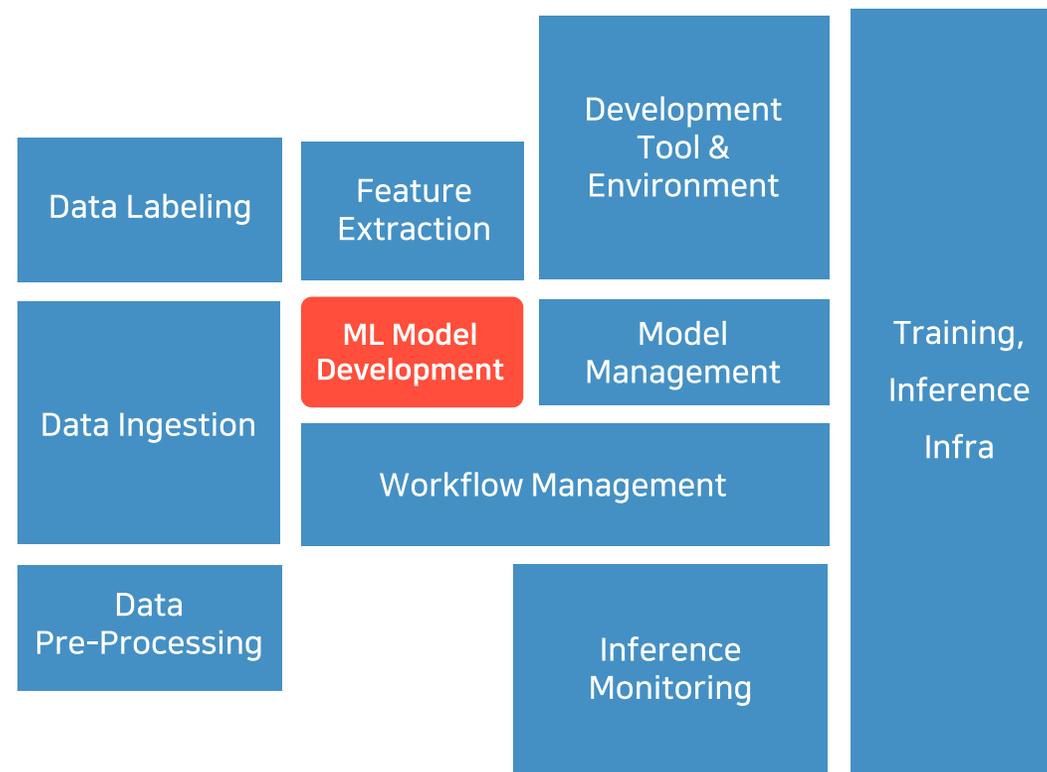
Machine Learning Project 현실

서비스 제공을 위한 ML Project에서는 데이터 처리부터 모델 개발/운영까지 여러 구성요소들을 고려하여 구축 필요

ML Project에 대한 인식



현실에서의 ML Project



Challenges in Machine Learning System

개발/학습/튜닝/배포 등
환경 구성의 어려움

모델 서비스 적용과
협업의 어려움



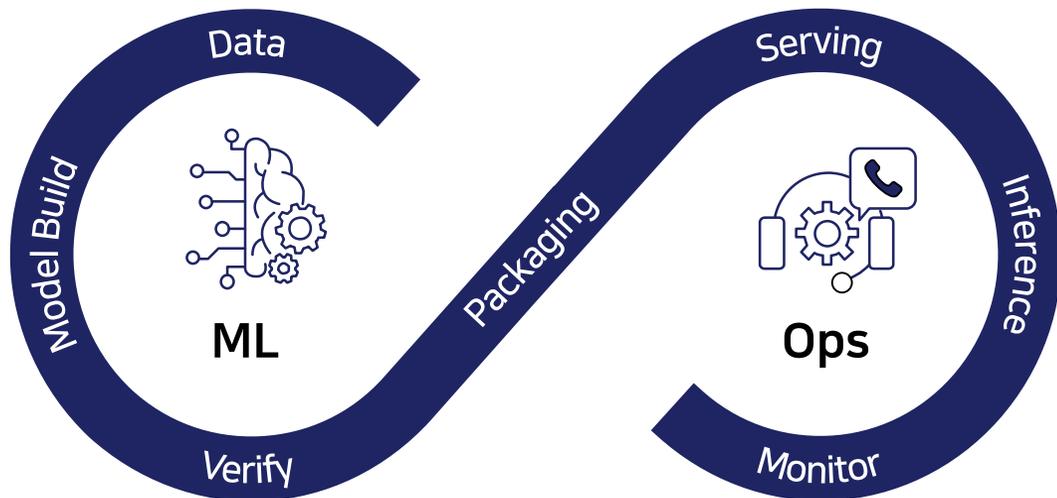
안정적이고 탄력적인
인프라 구성의 어려움

추론 모델 성능 관리와
ML Lifecycle 관리의 어려움

MLOps (ML + DevOps) 란?

ML Project가 실제 운영까지 이어지지 못하는 데서 시작하여, ML 모델 개발과 운영에서 반복되는 문제를 해결하고자 함

MLOps 개념



모델 개발 가속화와 운영 유지 관리 목표로 여러 단계를 유기적으로 연결하고 지속적으로 개선해 나가는 체계

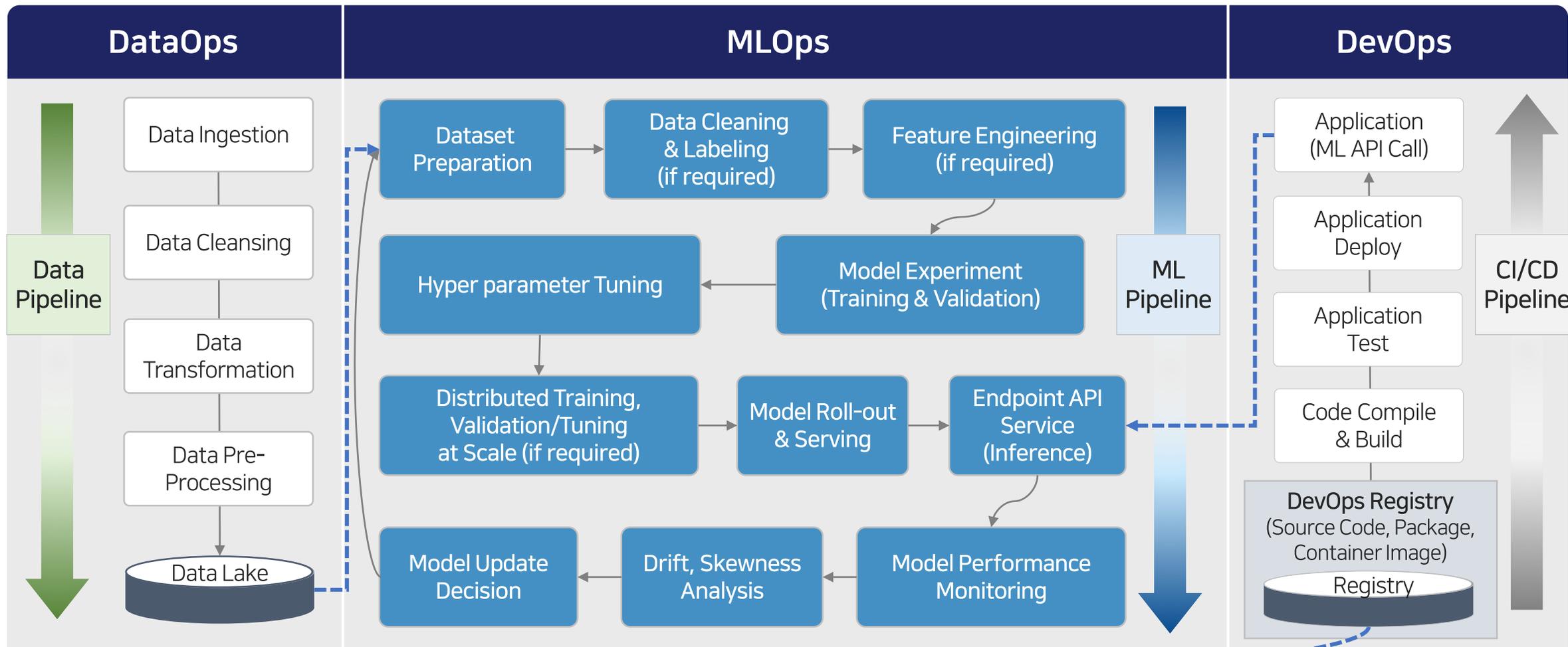
주요 동향

- ✓ **오픈소스**
 - Kubernetes 기반 ML Workflow를 지원하는 Kubeflow, 모델 라이프사이클 관리를 지원하는 MLflow 등
- ✓ **상용 플랫폼**
 - AWS SageMaker, MS Azure ML 등 클라우드 서비스 기업
 - 오픈소스 기반 차별화로 Startup, 상용플랫폼 다수 등장

AI가 확산되고 대규모 운영 환경에서 구축 사례가 증가함에 따라 MLOps를 지원하는 플랫폼 다수 등장

MLOps와 DataOps, DevOps의 관계

MLOps는 Data 측면에서 DataOps와 연계, DevOps와는 Registry 도구와 Application 연계 활용

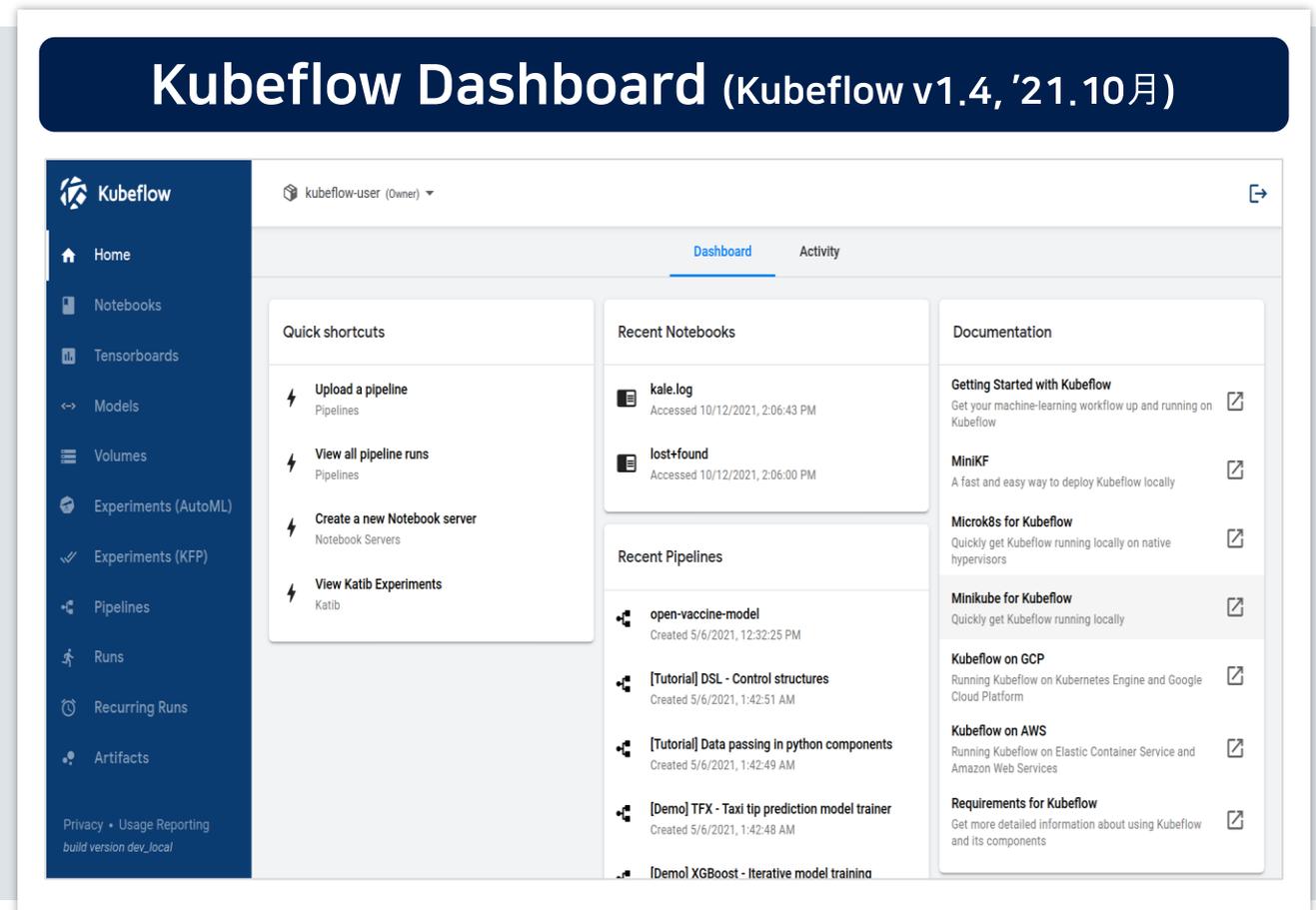


MLOps를 지원하는 Kubeflow

모델개발, 학습, 배포 및 추론 과정의 복잡한 ML Workflow를 Kubernetes 환경에서 **Simple, Portable, Scalable**하게 지원하는 Cloud Native 기반 오픈소스 플랫폼

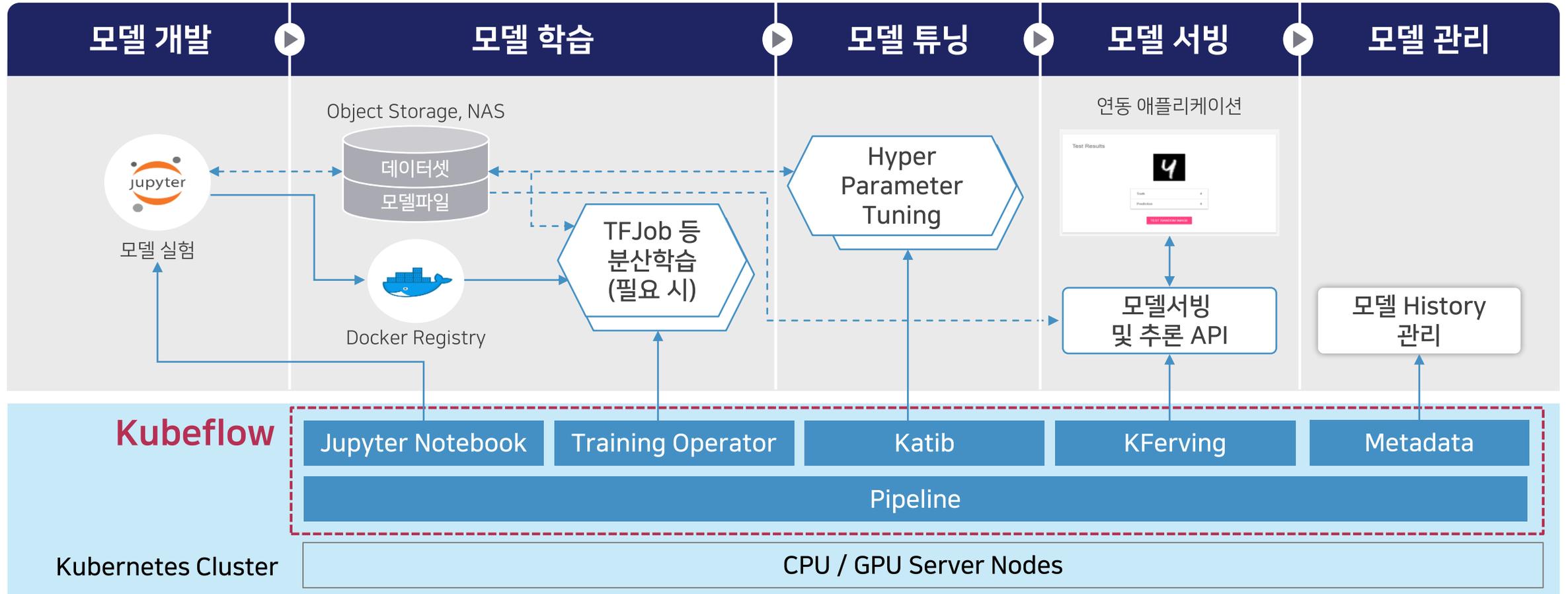
“Kubeflow = Kubernetes + ML flow”

- ✓ Google, Arrikto, Cisco, Bloomberg, MS 등 오픈소스 개발참여
- ✓ 솔루션 업체 (VMWare, Nvidia 등) 및 주요 CSP (Amazon, Google, MS)사는 자사 AI 플랫폼과 Kubeflow 연계 상품 제공
- ✓ ML 오픈소스 생태계에 영역별 최적 오픈소스 조합
 - 다양한 ML Framework 및 개발언어/툴/환경 제공
 - 개발된 모델의 빠른 추론 서비스 구성 및 Workflow 자동화 기능
- ✓ Kubernetes 기반 운영 안정성, 유연성 제공



Kubeflow 핵심 구성요소

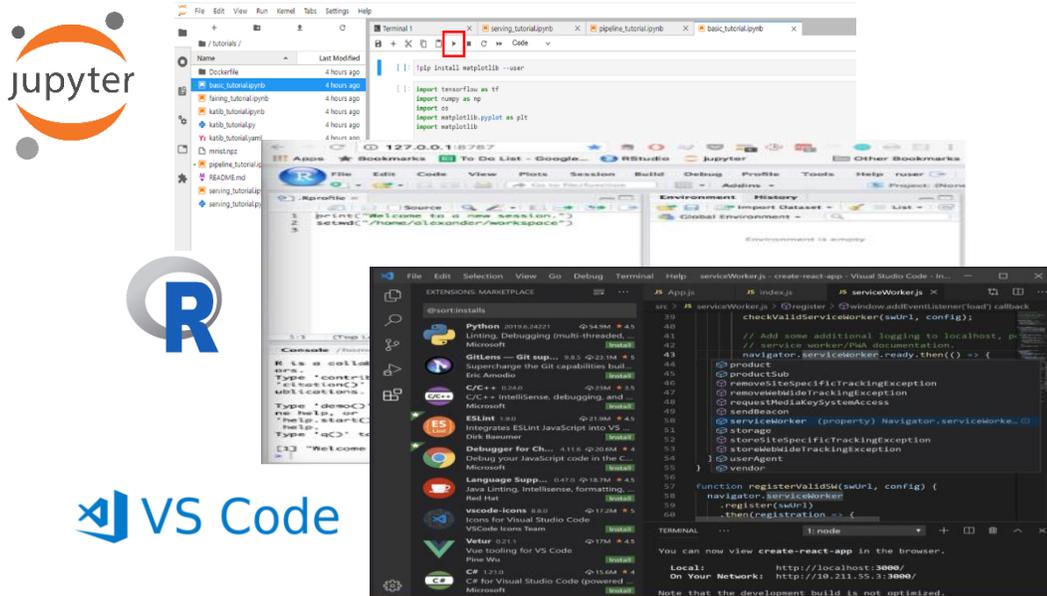
Kubernetes 기반으로 모델 개발, 학습, 튜닝, 서빙, 관리 등 ML Workflow 단계별 유용한 툴과 파이프라인 기능 제공



Kubeflow 주요 기능 (1/2)

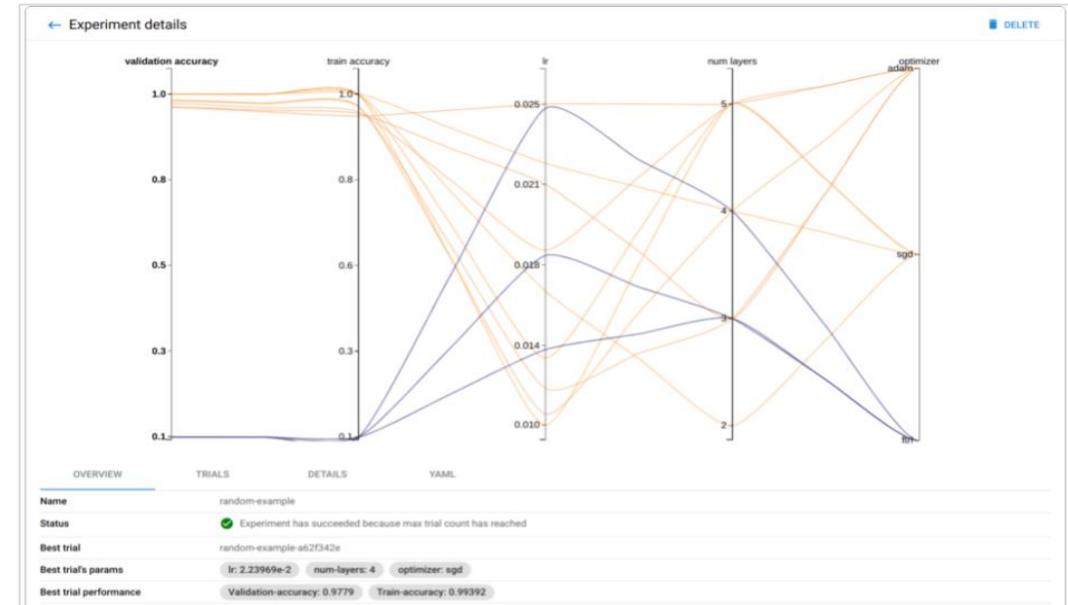
다양한 ML 개발 및 학습 환경 제공

- ML Framework를 포함한 Jupyter Lab 환경 제공
 - Rstudio, Visual Studio Code 등 여러 IDE 환경 제공
 - 분산 학습을 쉽게 Kubernetes 환경에서 컨테이너로 실행 가능한 Operator 지원 (TFJob, PyTorchJob, MPIJob 등)



모델 튜닝 (Katib)

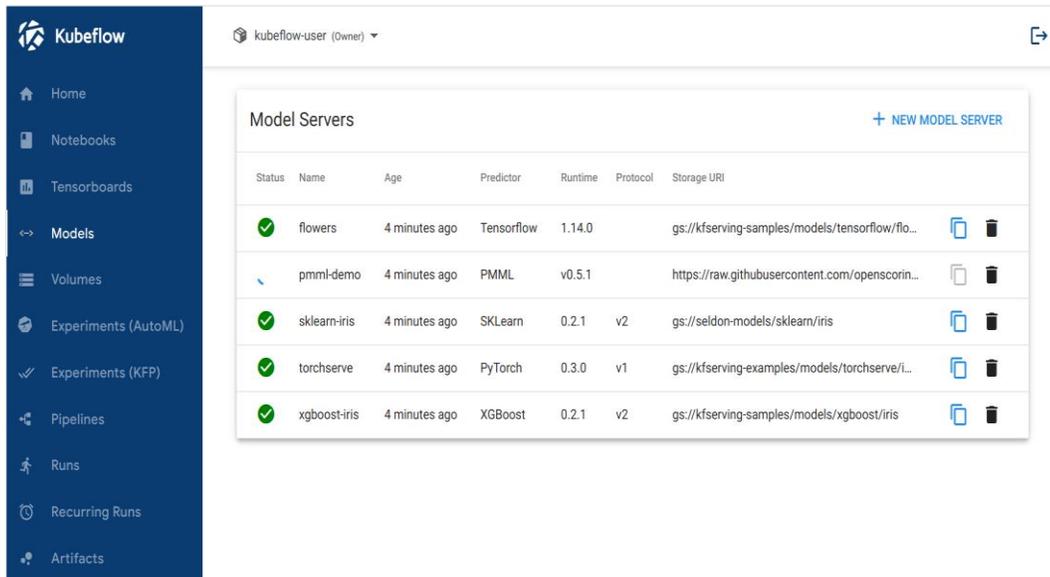
- 하이퍼파라미터 최적화 기능 지원
 - 목표 (accuracy 등), 탐색범위 (Learning rate, Layer수 등), 탐색 알고리즘 (random, grid 탐색 등), 최대 시도 등 설정
 - ML 코드내 파라미터를 yaml 변수와 연동하여 구동



Kubeflow 주요 기능 (2/2)

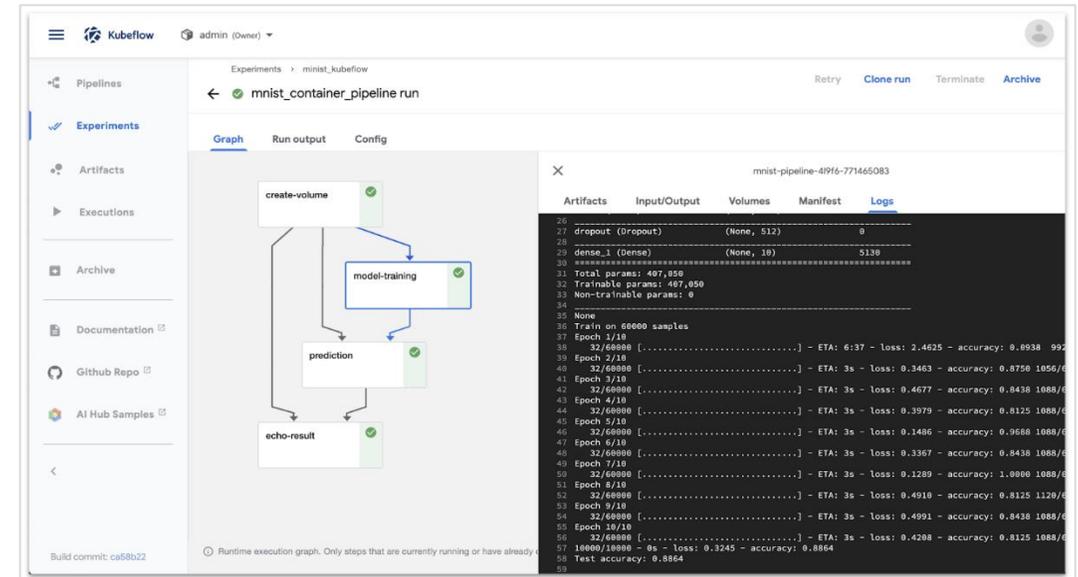
모델 서빙 (KFServing)

- ✔ 유연하고 확장 가능한 모델 배포 및 추론 지원
 - 다양한 ML Framework 모델 서버를 활용한 서빙 지원
 - 개발된 모델의 서비스화를 위한 Endpoint API 생성
 - Inference 시, A/B Test와 Auto-Scaling 지원



파이프라인 (Pipeline)

- ✔ ML Workflow 생성과 실행 관리
 - 각 Step을 Graph로 표현하여 실행 모니터링 및 공유 가능
 - Pipeline UI 혹은 Python SDK를 활용하여 Workflow 구성
 - Workflow 변경 시, 각 Step을 재사용하거나 교체 가능



SCP Kubeflow Service 소개

SDS는 SCP (Samsung Cloud Platform)에서 사용자가 직접 구성 가능한 Kubeflow 서비스 제공 (<https://cloud.samsungsds.com>)

SCP (Samsung Cloud Platform) 란?

- ✓ 총 10개 상품 영역 31개 기본 클라우드 상품 제공
- ✓ 사용자가 SDS 운영자 개입 없이 **자원의 생성부터 모니터링, 빌링까지** 전체 사용 과정을 **셀프로** 수행할 수 있는 서비스
- ✓ Enterprise 고객 IT시스템을 **클라우드**로 신규 구축하거나, 기존 시스템을 **클라우드**로 전환 시 활용

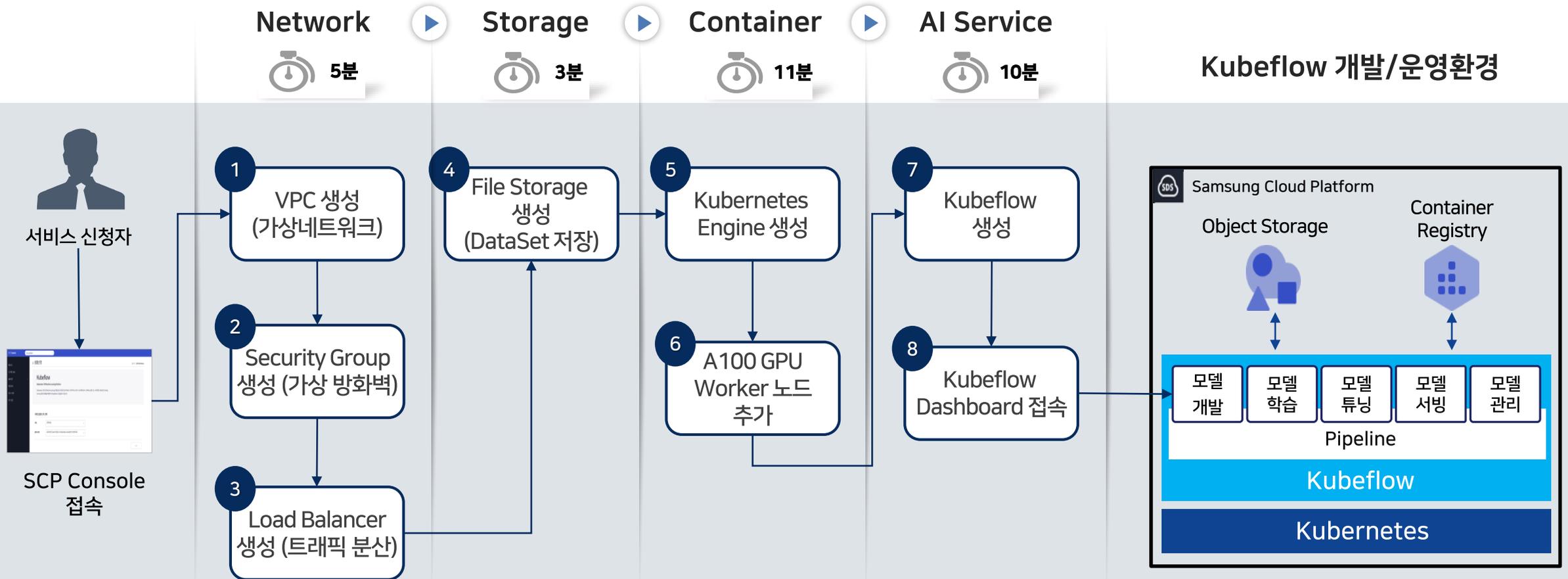
SCP Kubeflow Service 개요

- ✓ Kubeflow Mini와 Enterprise ('22년 개선) 상품으로 구성 : **Kubeflow 확장하여 스케줄링, 모니터링, 로깅 기능 탑재**
- ✓ Kubernetes Engine에서 **Kubernetes Cluster** 생성 후, **Kubeflow**를 해당 Cluster에 배포
- ✓ Kubernetes Cluster 생성 시 최소 CPU 노드 3대 이상 필요, **GPU 노드 (V100, A100)**를 확장하여 구성

Kubeflow Service 구성 방법

SCP 클라우드 상품을 활용하여 Kubeflow에 필요한 인프라 및 SW자원을 수분 이내 Self-Service 구성

※ MLOps 개발/운영환경 제공 리드타임 단축 : 최소 2주 이상 → 수십분 이내



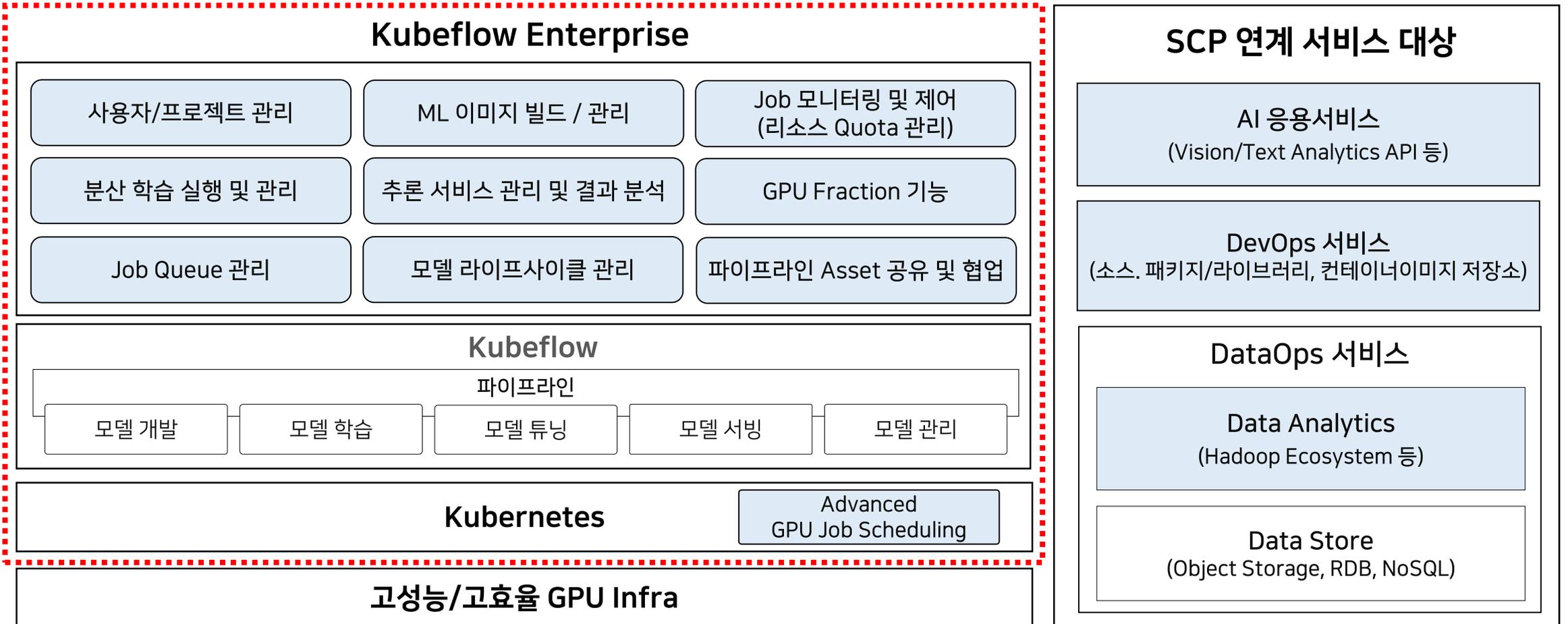
SCP Kubeflow Demo



Kubeflow Enterprise Service

'22년 Kubeflow 엔진 위에 SDS 기술력을 더하여 Enterprise 수준의 MLOps 플랫폼을 클라우드 서비스로 제공

※ 클라우드 환경을 선호하는 고성능 GPU 인프라와 MLOps 플랫폼을 필요로 하는 전문 Data Scientist를 보유한 Enterprise 고객 대상



Thank you

SAMSUNG SDS