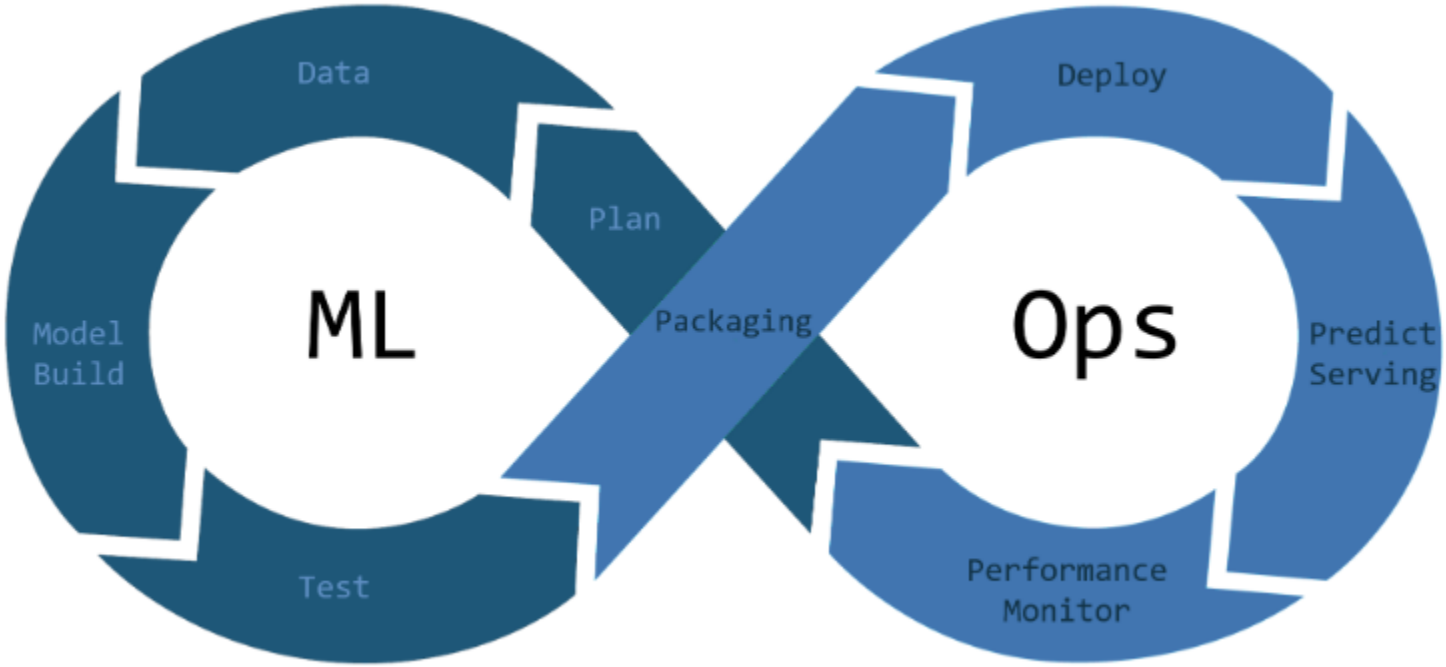
MLOps를 위한 MLflow

연구소 연구2팀
신재영 주임연구원

2023.04.25

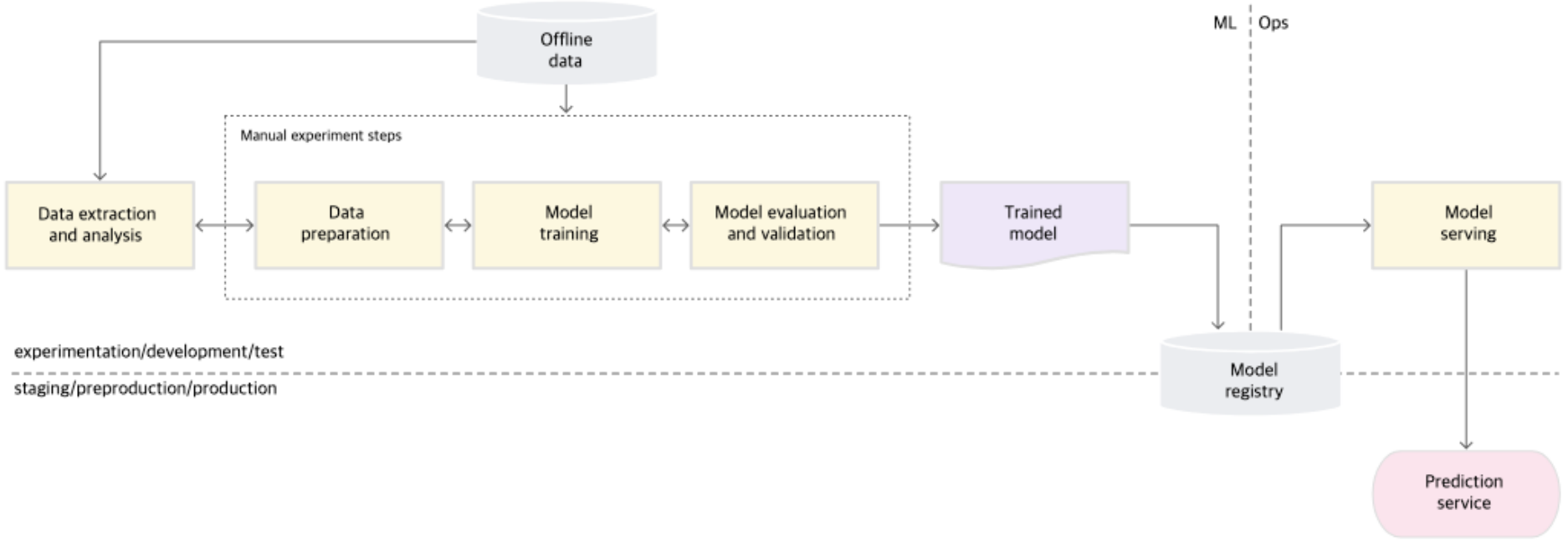
1. MLOps란?

- Machine Learning(ML) + Operations(Ops)의 합성어로 머신러닝 모델을 안정적이고 효율적으로 배포 및 유지 관리하는 것
- 데이터 관리 및 머신러닝 개발뿐만 아니라 서비스 운영을 통해 안정적으로 서비스를 제공하기 위해 MLOps 개념이 탄생



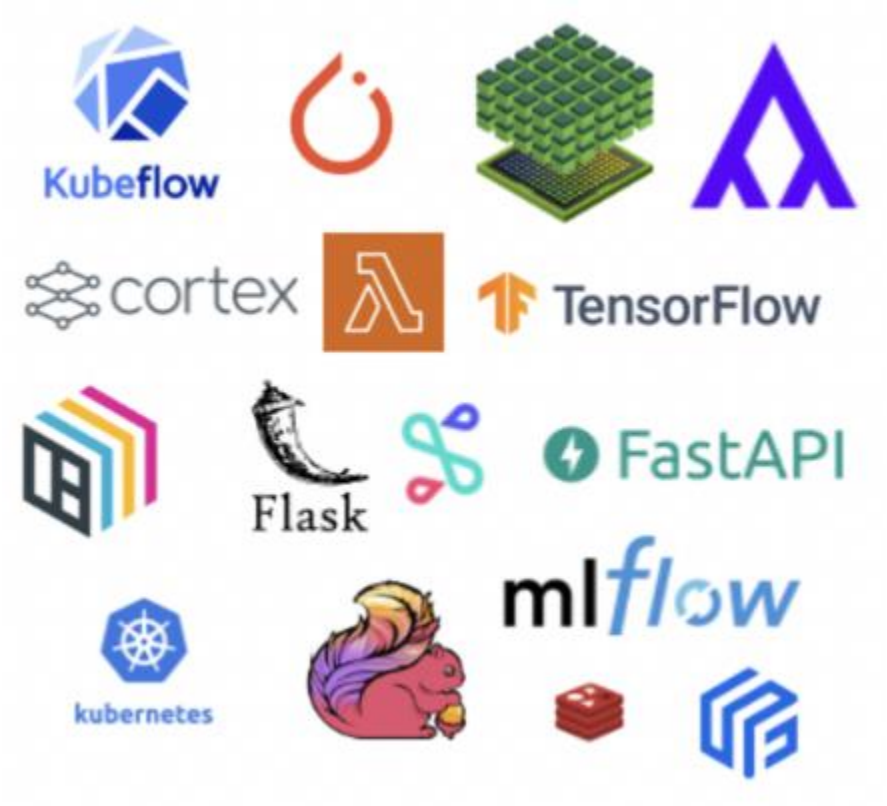
1. MLOps란?

- MLOps는 크게 ML(학습) 단계와 Ops(운영) 단계로 나뉨
- ML 단계 : 데이터 수집, 전처리, 모델 구축, 학습 및 평가 등
- Ops 단계 : 모델 배포, 모니터링, 테스트 등



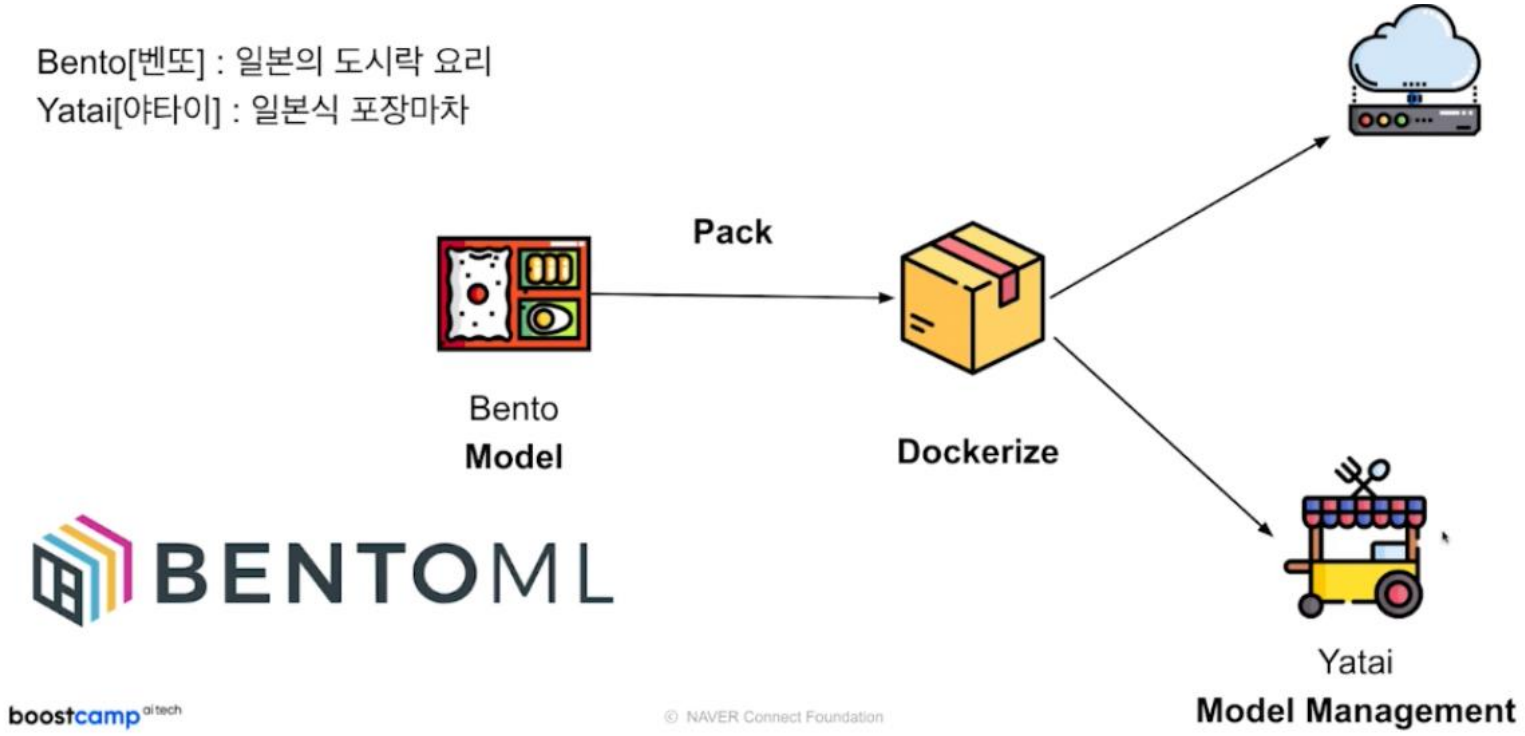
2. MLOps 플랫폼

- 효율적인 머신러닝 라이프사이클 관리와 배포를 위해 많은 플랫폼이 생기고 있음



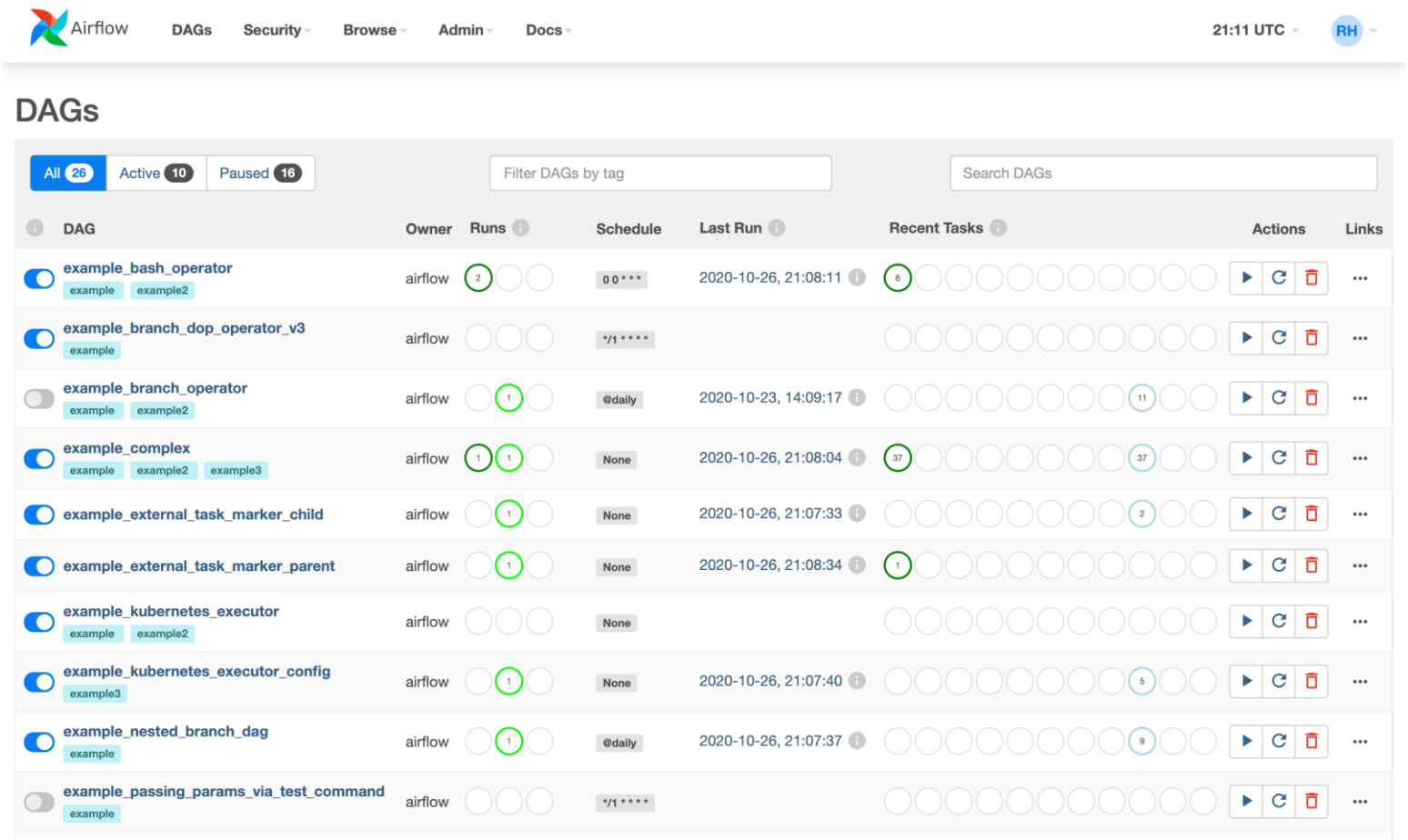
2. MLOps 플랫폼 – BentoML

- 머신러닝 모델을 만들면 해당 모델을 쉽게 배포 및 테스트를 할 수 있게 해주는 오픈소스 라이브러리
- Bento[벤또]는 일본말로 도시락을 말하며, 모델을 도시락에 담아 포장하여 관리 및 배포한다는 의미
- 모델 배포를 위해 모델 Packing을 자동으로 해주며, Docker 이미지를 생성해주는 기능이 있음



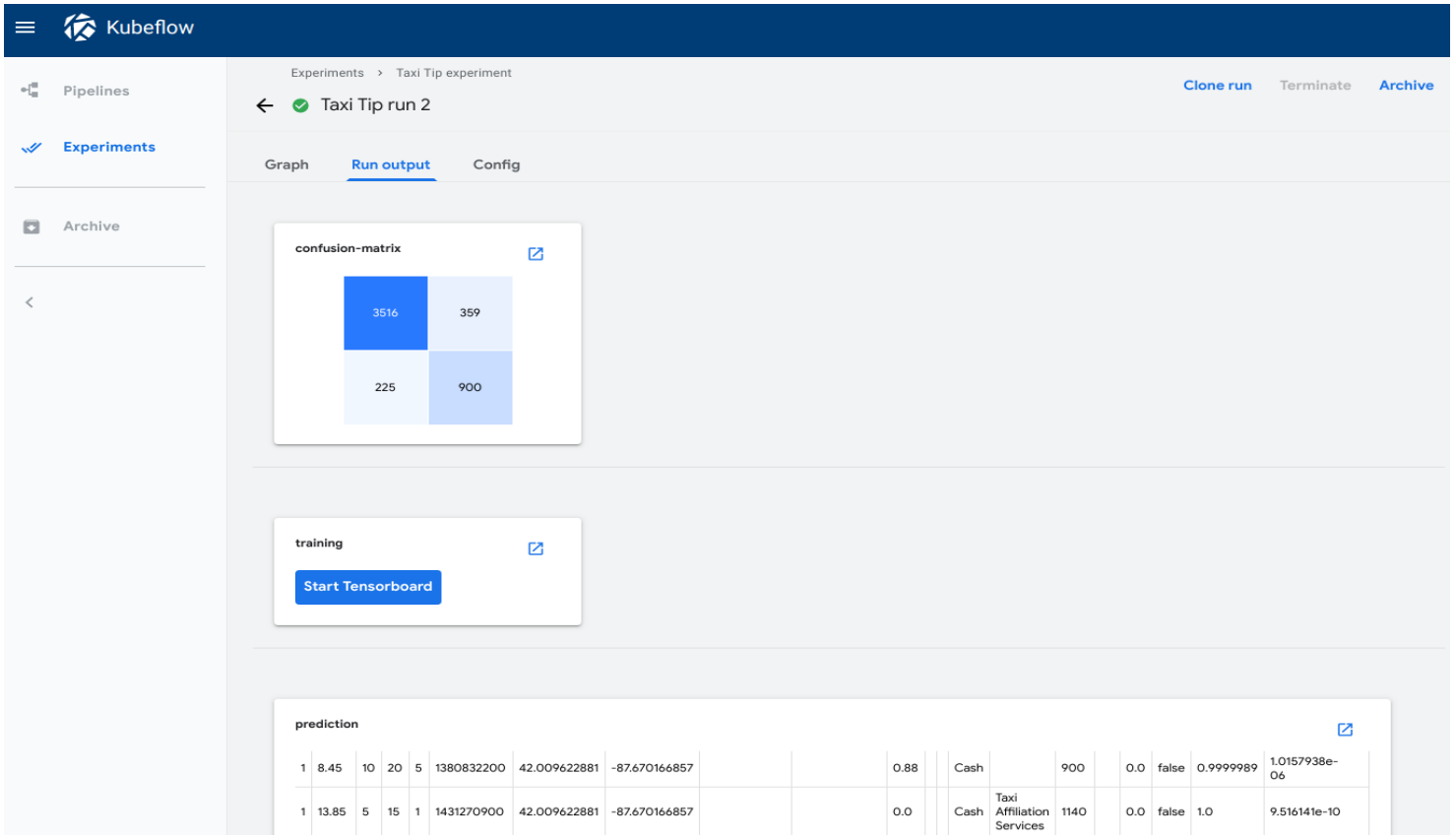
2. MLOps 플랫폼 – Apache Airflow

- 2015년 6월에 Apache가 발표한 데이터 공학 파이프라인을 위한 오픈 소스 워크플로우 관리 플랫폼
- 에어비앤비(Airbnb)에서 점차 복잡해지는 워크플로우 관리를 위해 개발이 시작되었음



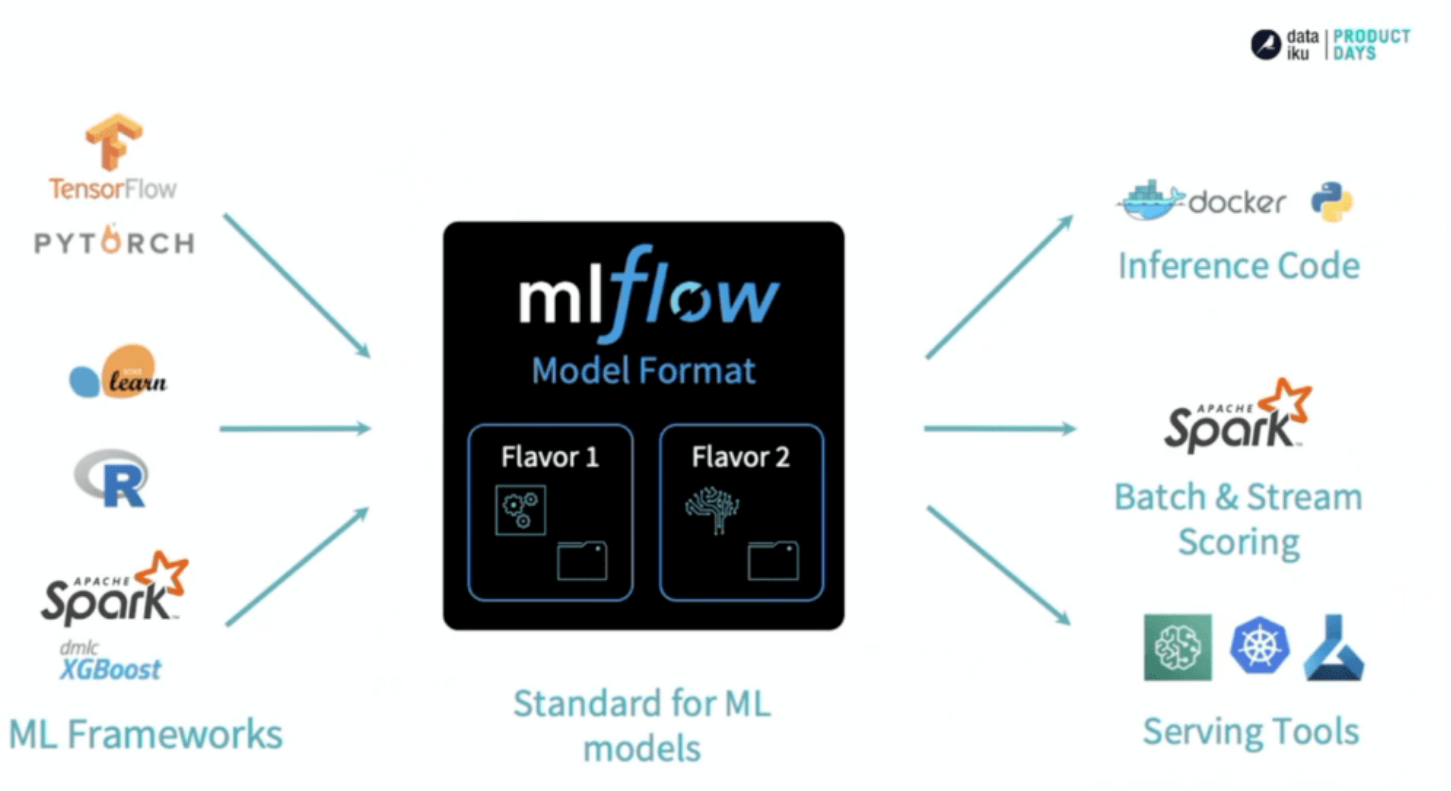
2. MLOps 플랫폼 – Kubeflow

- 머신러닝 워크플로우의 머신러닝 모델 학습부터 배포 단계까지 모든 작업에 필요한 도구와 환경을 쿠버네티스(Kubernetes) 위에서 제공
- 컨테이너를 직접 빌드하거나 커스터마이즈 할 필요 없이 간단히 모델을 배포할 수 있음
- 많은 기업에서 쿠버네티스를 활용하여 Kubeflow를 사용하고 있음



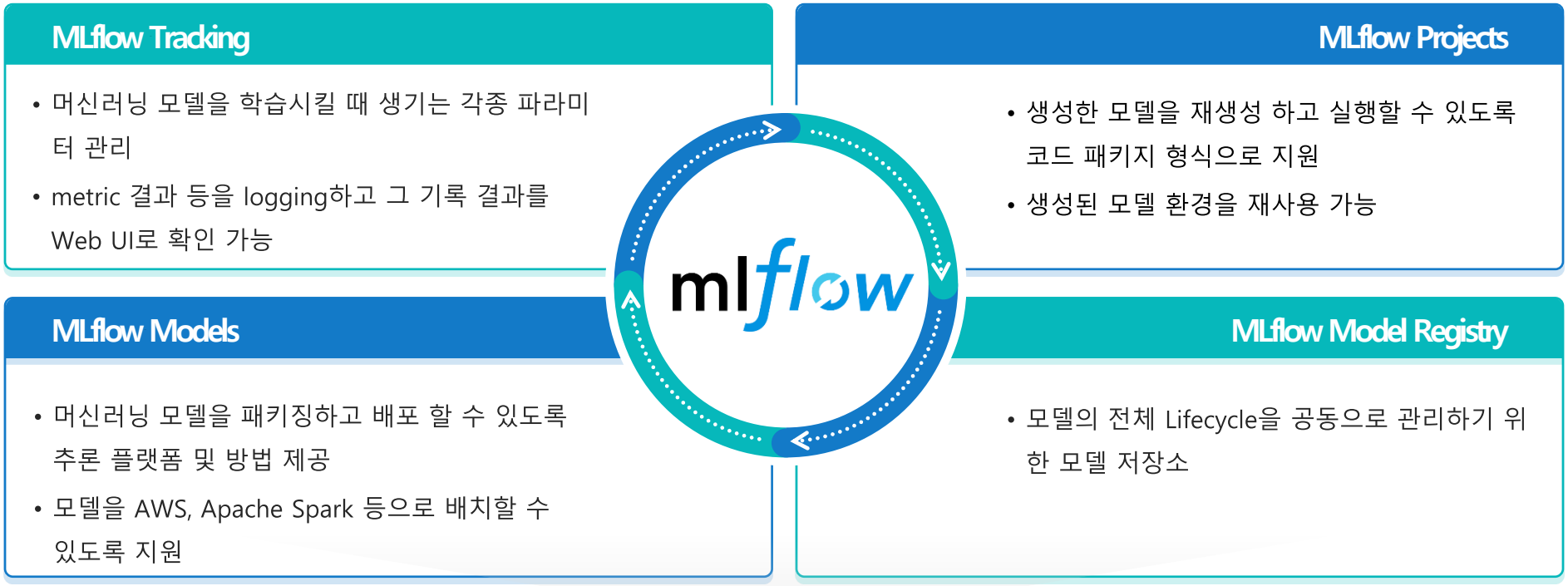
2. MLOps 플랫폼 – MLflow

- Databricks에서 개발한 End-to-End 머신러닝 Lifecycle을 관리하기 위한 오픈 소스 플랫폼
- 머신러닝 모델의 실험을 추적하고 모델을 공유 및 배포할 수 있도록 지원하는 라이브러리



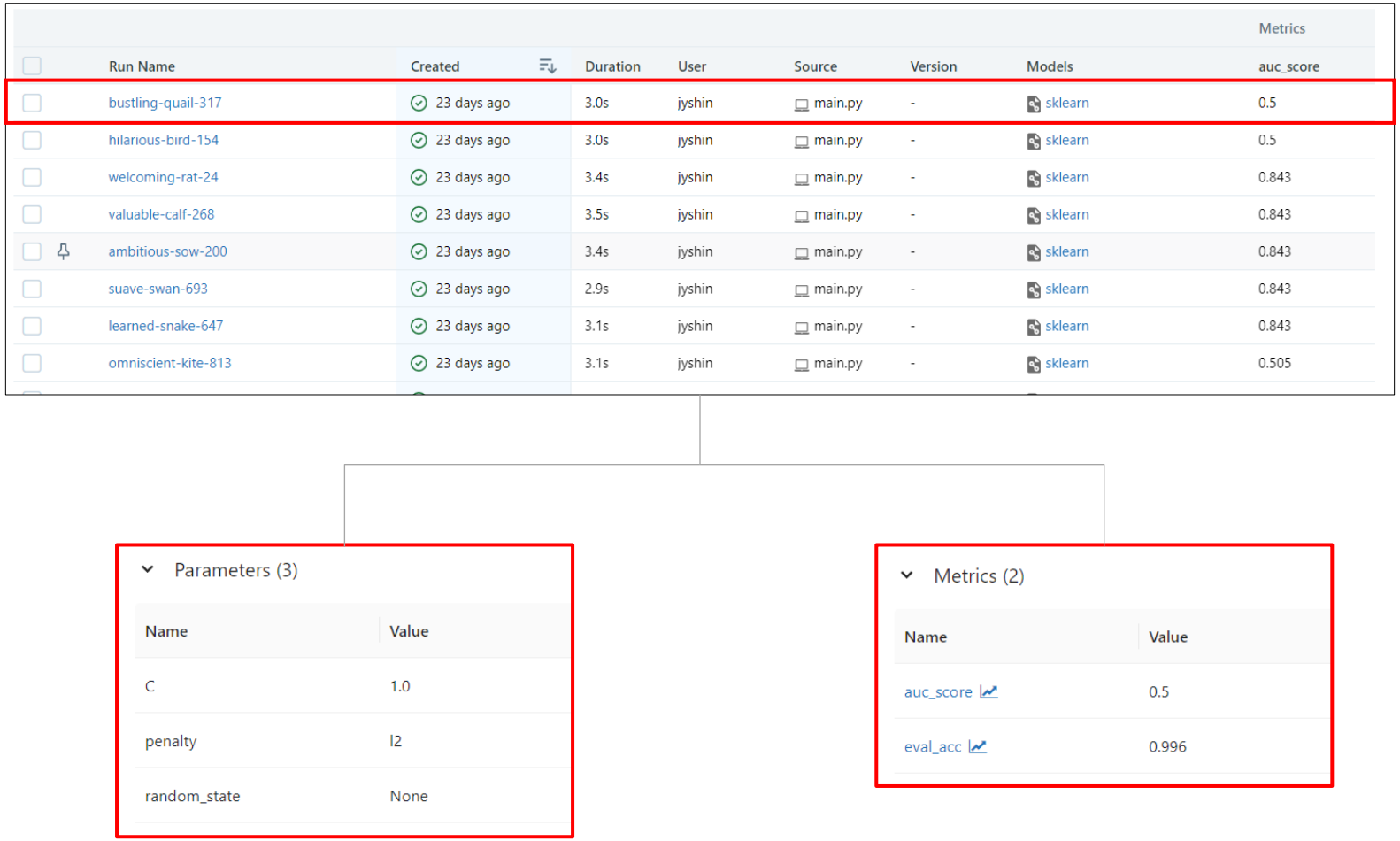
3. MLflow 주요 기능 및 특징

- MLflow의 주요 기능은 총 4가지이며 각각의 특징은 다음과 같음



3. MLflow 주요 기능 및 특징

(1) MLflow Tracking



	Run Name	Created	Duration	User	Source	Version	Models	Metrics: auc_score
	bustling-quail-317	23 days ago	3.0s	jyshin	main.py	-	sklearn	0.5
	hilarious-bird-154	23 days ago	3.0s	jyshin	main.py	-	sklearn	0.5
	welcoming-rat-24	23 days ago	3.4s	jyshin	main.py	-	sklearn	0.843
	valuable-calf-268	23 days ago	3.5s	jyshin	main.py	-	sklearn	0.843
	ambitious-sow-200	23 days ago	3.4s	jyshin	main.py	-	sklearn	0.843
	suave-swan-693	23 days ago	2.9s	jyshin	main.py	-	sklearn	0.843
	learned-snake-647	23 days ago	3.1s	jyshin	main.py	-	sklearn	0.843
	omniscient-kite-813	23 days ago	3.1s	jyshin	main.py	-	sklearn	0.505

Parameters (3)

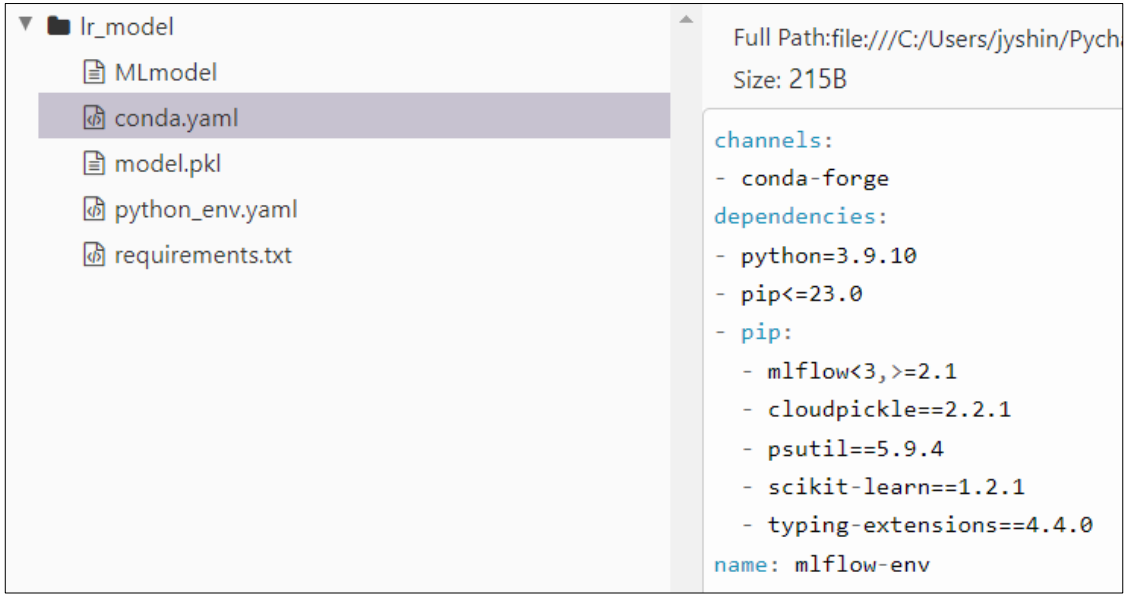
Name	Value
C	1.0
penalty	l2
random_state	None

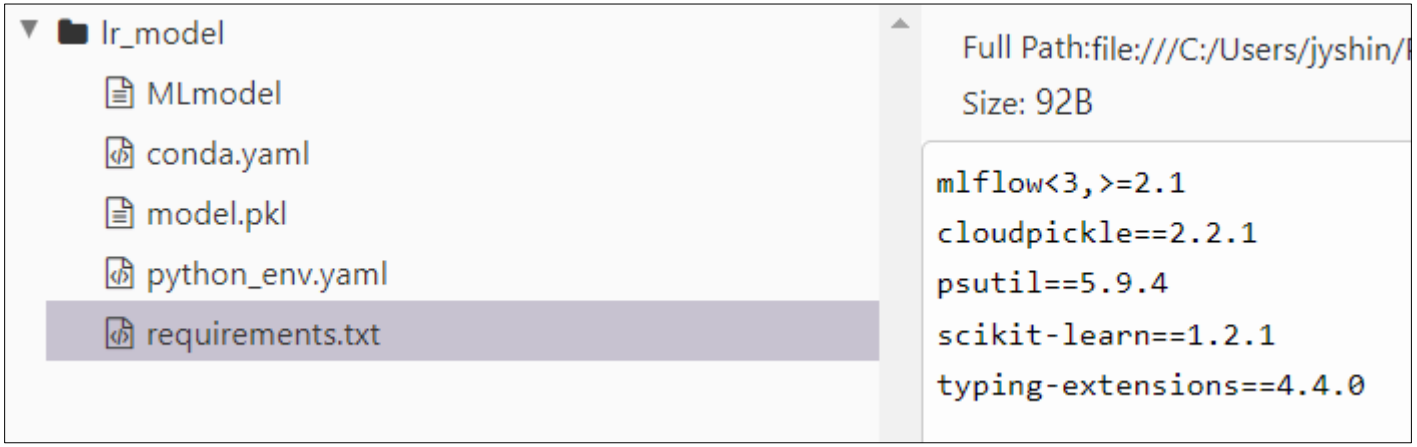
Metrics (2)

Name	Value
auc_score	0.5
eval_acc	0.996

3. MLflow 주요 기능 및 특징

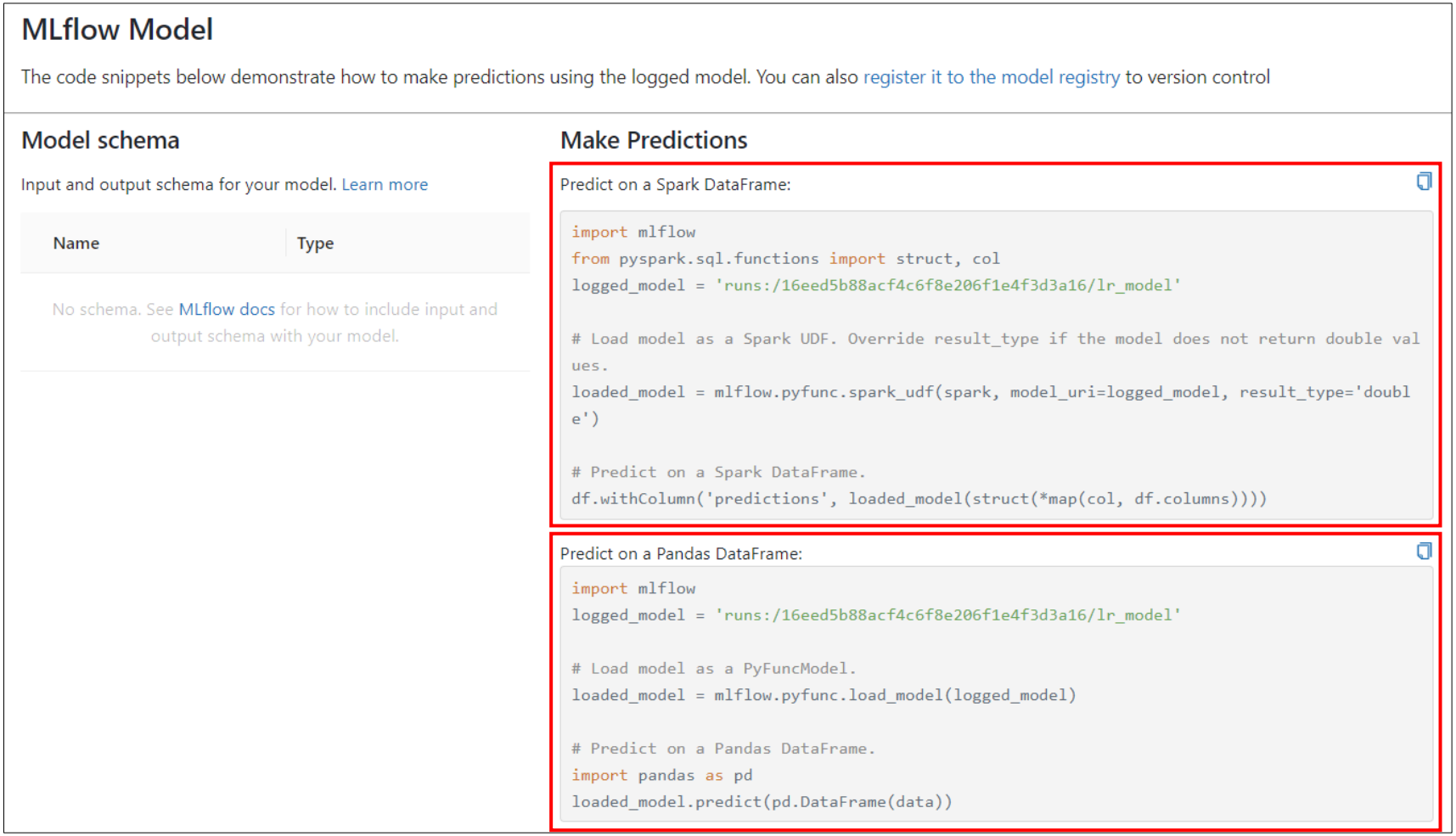
(2) MLflow Projects





3. MLflow 주요 기능 및 특징

(3) MLflow Models



3. MLflow 주요 기능 및 특징

(4) MLflow Model Registry

	Run Name	Created	Duration	User	Source	Version	Models	Metrics: auc_score
☐	bustling-quail-317	23 days ago	3.0s	jyshin	main.py	-	sklearn	0.5
☐	hilarious-bird-154	23 days ago	3.0s	jyshin	main.py	-	sklearn	0.5
☐	welcoming-rat-24	23 days ago	3.4s	jyshin	main.py	-	sklearn	0.843
☐	valuable-calf-268	23 days ago	3.5s	jyshin	main.py	-	sklearn	0.843
☐	ambitious-sow-200	23 days ago	3.4s	jyshin	main.py	-	sklearn	0.843
☐	suave-swan-693	23 days ago	2.9s	jyshin	main.py	-	sklearn	0.843
☐	learned-snake-647	23 days ago	3.1s	jyshin	main.py	-	sklearn	0.843
☐	omniscient-kite-813	23 days ago	3.1s	jyshin	main.py	-	sklearn	0.505

Registered Models

Share and manage machine learning models. Learn more

Create Model

Search by model names or tags | Search | Clear

Name	Latest Version	Staging	Production	Last Modified	Tags
lr_model	Version 1	-	-	2023-04-20 11:15:07	-

1 | 10 / page

4. MLflow Model Pipeline 구현

개발 환경

- python 3.9.1
- mlflow 2.1.1
- numpy 1.23.5
- pandas 1.5.3
- scikit-learn 1.2.1

MLflow 튜토리얼

- https://github.com/jaeyeongs/mlflow_example

5. MLflow 필요성 및 활용 방안

1

간편한 사용성

2

반복적인 실험에 적합

3

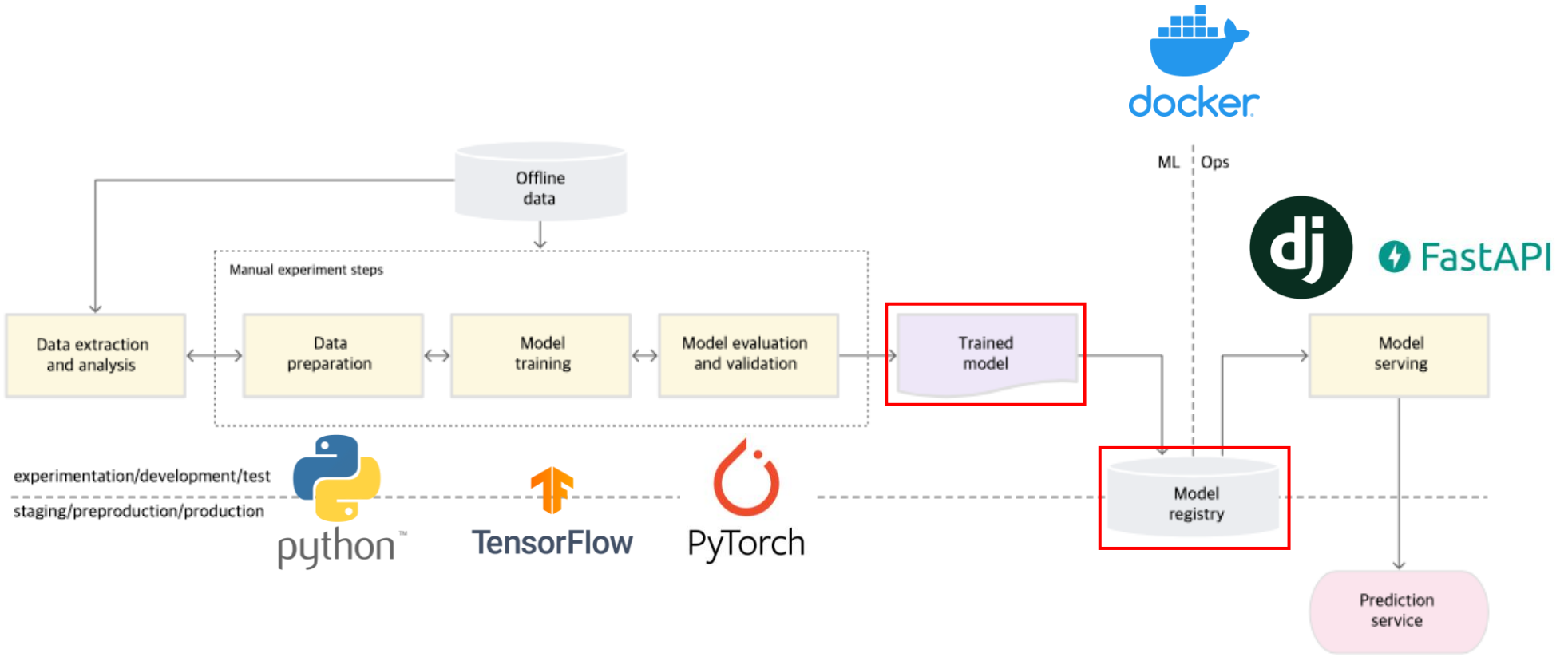
모델 관리 용이

4

모델 정보를 UI로
쉽게 보여줌

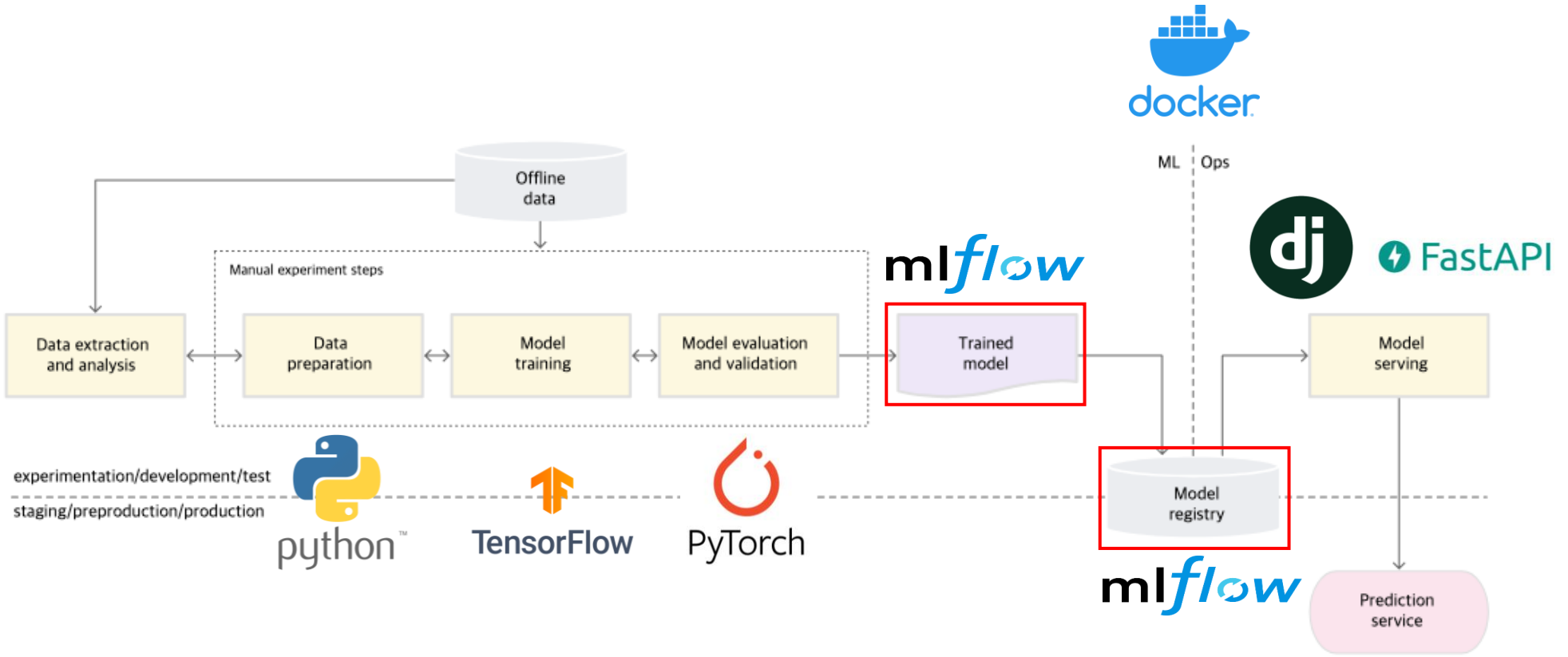
5. MLflow 필요성 및 활용 방안

- 현재 연구소에서는 Model 구축 및 관리를 위한 플랫폼이 없음
- Model 훈련 정보 및 성능 측정과 같이 반복적인 실험을 할 수 있는 환경이 필요



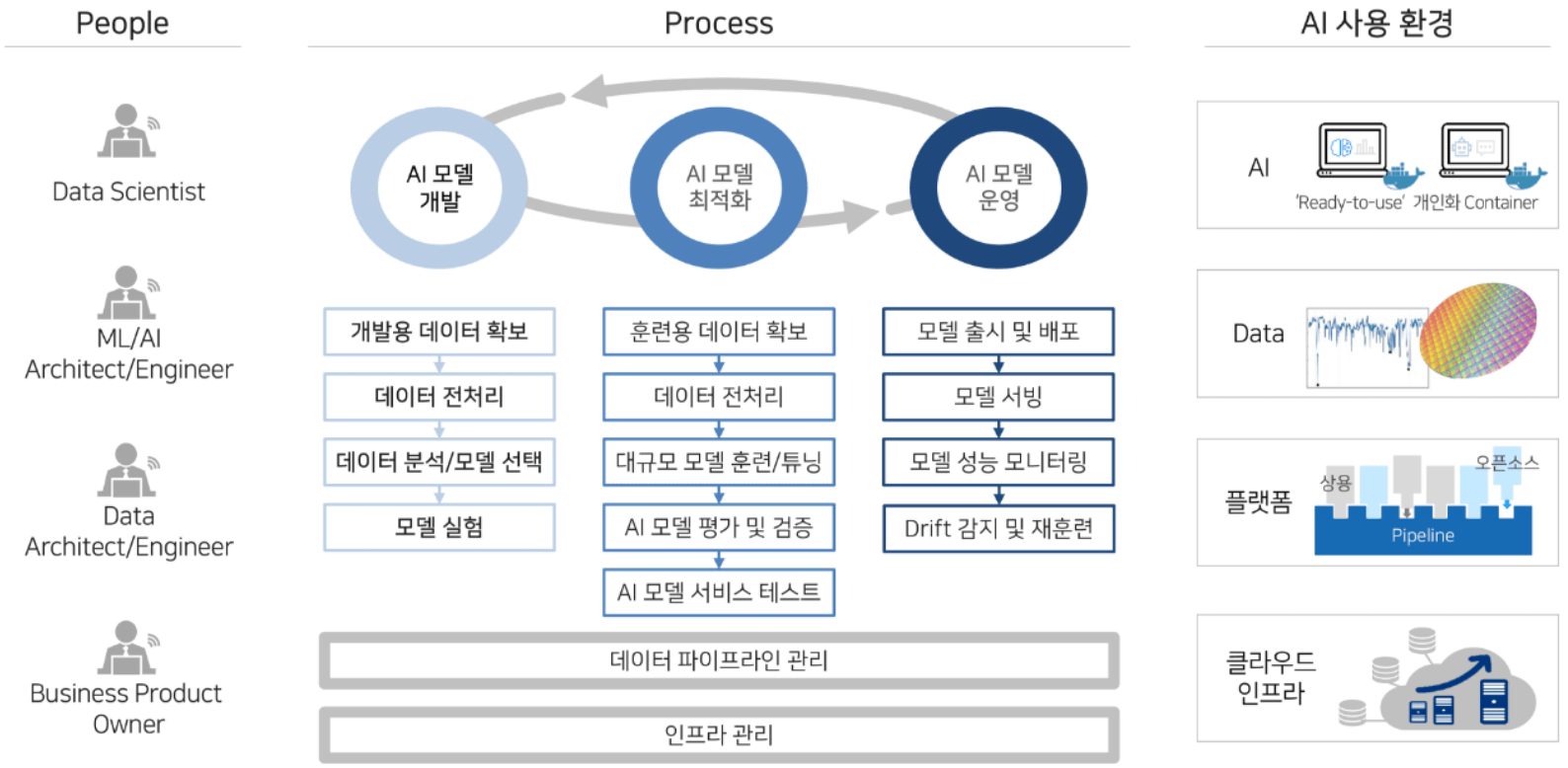
5. MLflow 필요성 및 활용 방안

- MLflow를 통해 구축한 모델을 관리 및 저장하여 모델 유지 보수 가능
- MLflow를 활용하여 서비스 배포를 더욱 효율적으로 할 수 있음



5. MLflow 필요성 및 활용 방안

- 최근 다양하고 많은 모델들이 생성되고 있는 추세에서 모델을 최적화하는 Task의 중요성이 높아지고 있음
- MLflow와 같은 MLOps 플랫폼을 사용함으로써 체계적인 모델 운영 관리가 가능함
- Spotify는 2019년 12월 파이프라인 도입을 통해 Data Scientist 의 생산성이 700% 향상되었다고 밝힌 바 있음



Q&A