

DevOps는 알겠는데 MLops는 또 뭐야?



모르는데 어떻게 해요?

모르면 맞아야죠(?)



40분 뒤에 여러분이 모르면
모르면 맞아야죠(?)

제가

8강 8조 최종전
CRASH

MBC
GAME

본 자료는 아래의 자료를 기반으로 만들었습니다.

SCALABLE MACHINE LEARNING

THE STATE OF DATAOPS / MLOPS IN 2018

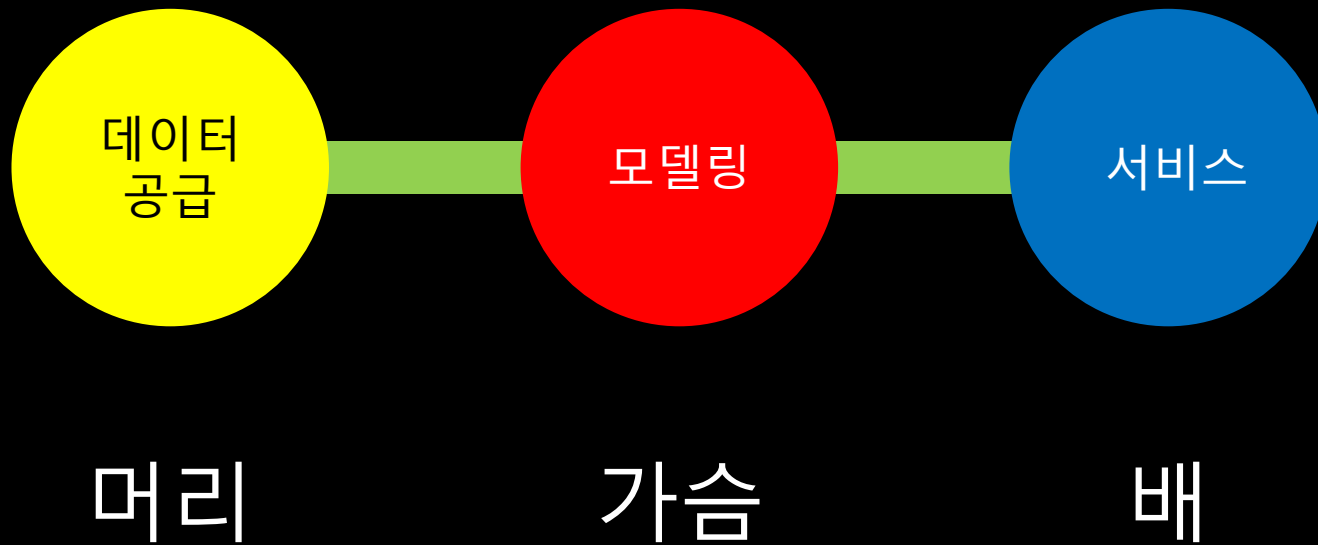
euroscipy 2018

alejandro saucedo

@axsucedo

in/axsucedo

Project with Data science



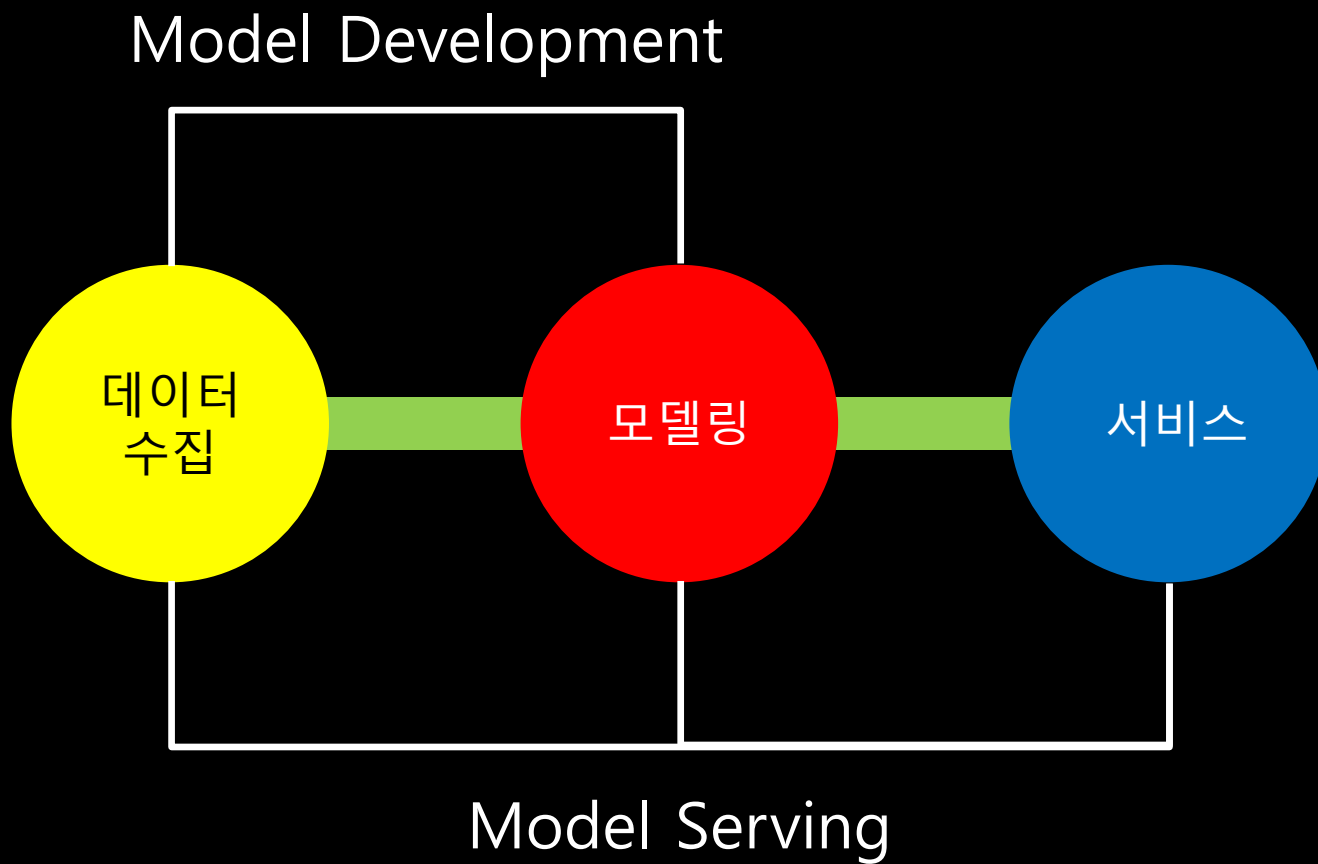
정바
백선생표 만능간장

얼마나
간단한지 보세요



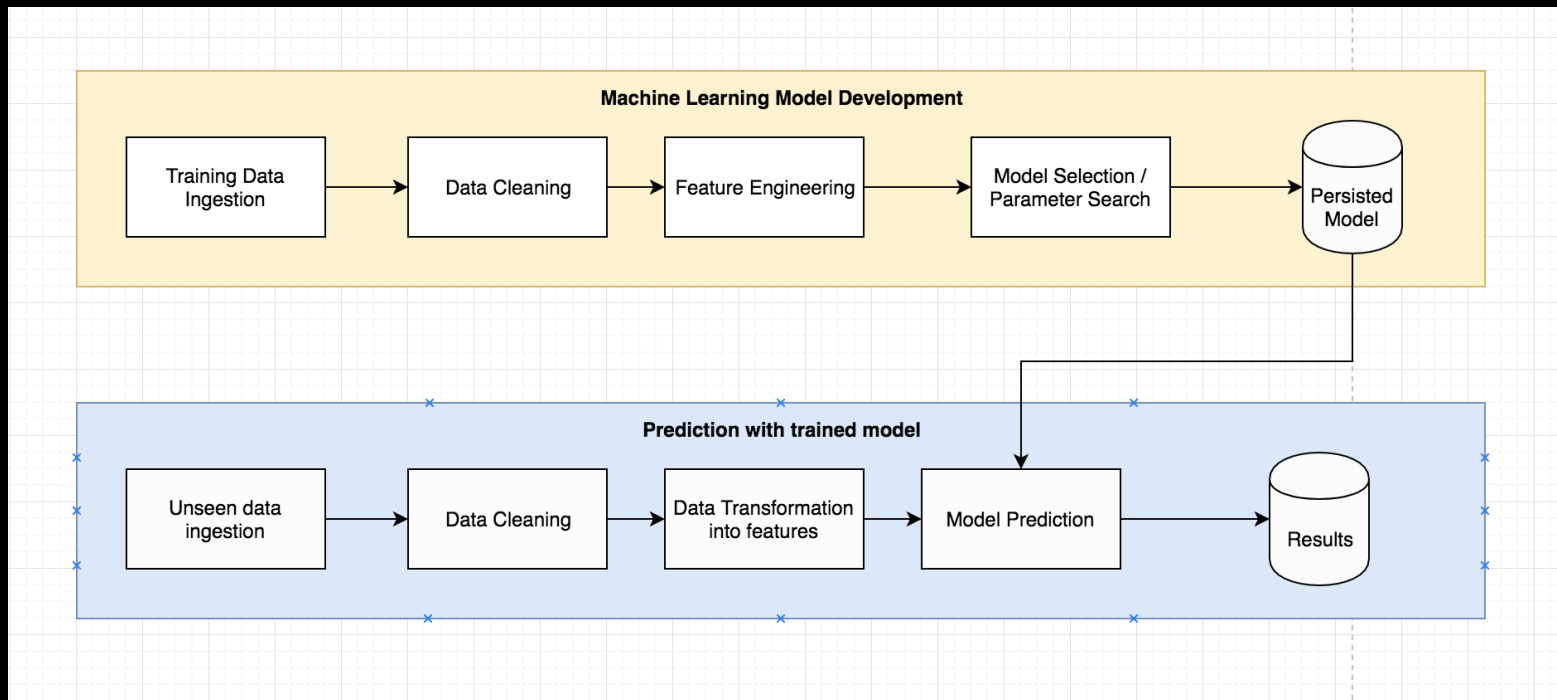
이거 너무 간략한거 아니냐고

그게 실제론 말이죠



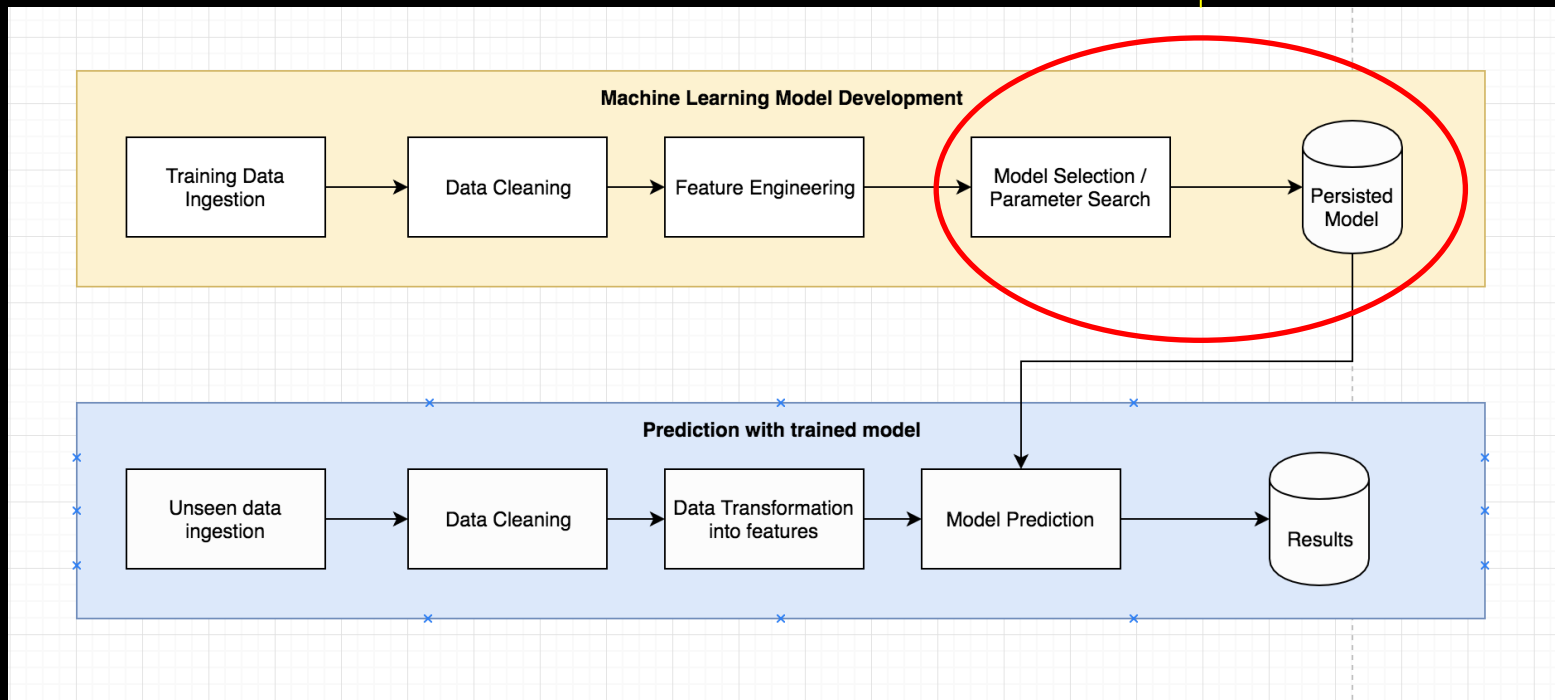
그게 실제론 말이죠

- Model Development
- Model Serving



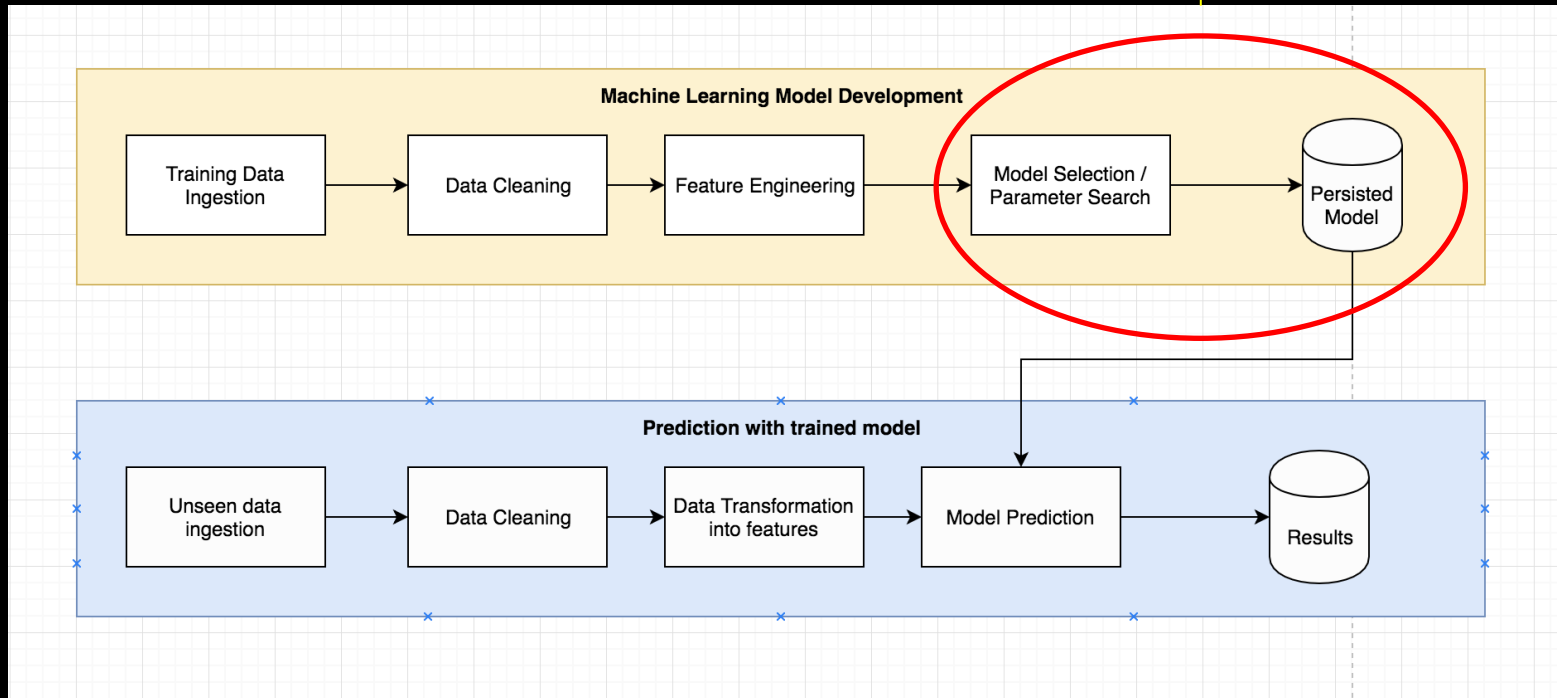
작은 조직의 프로젝트라면

신경써야 할 모델의 수가
그다지 많지 않습니다.



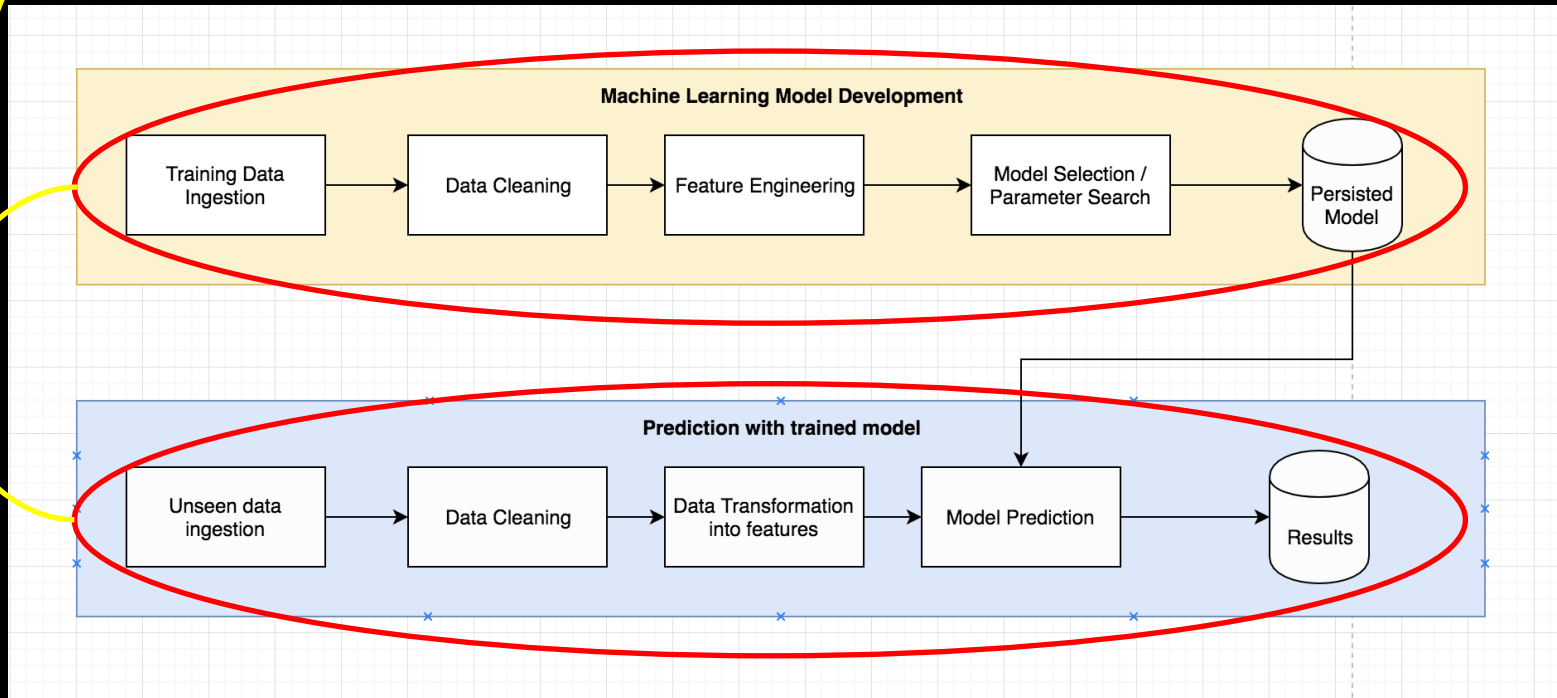
작은 조직의 프로젝트라면

데이터 사이언티스트(모델러)의
수가 적기 때문에 모델링
협업에 특별히 문제가 없습니다.



작은 조직의 프로젝트라면

워크플로우의 수가 적기 때문에
각 단계를 각각 별도의 방법으로
트래킹을 해도 무방하다.



하지만

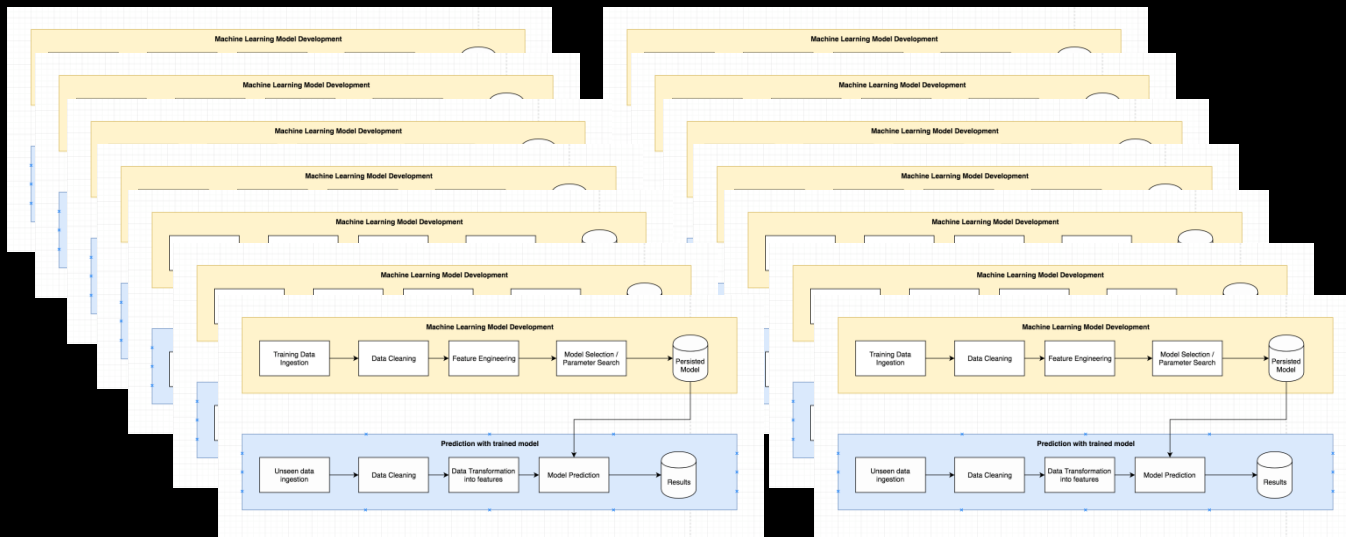


Image : <https://axsauze.github.io/scalable-data-science>

규모가 커지면 이야기가 달라진다

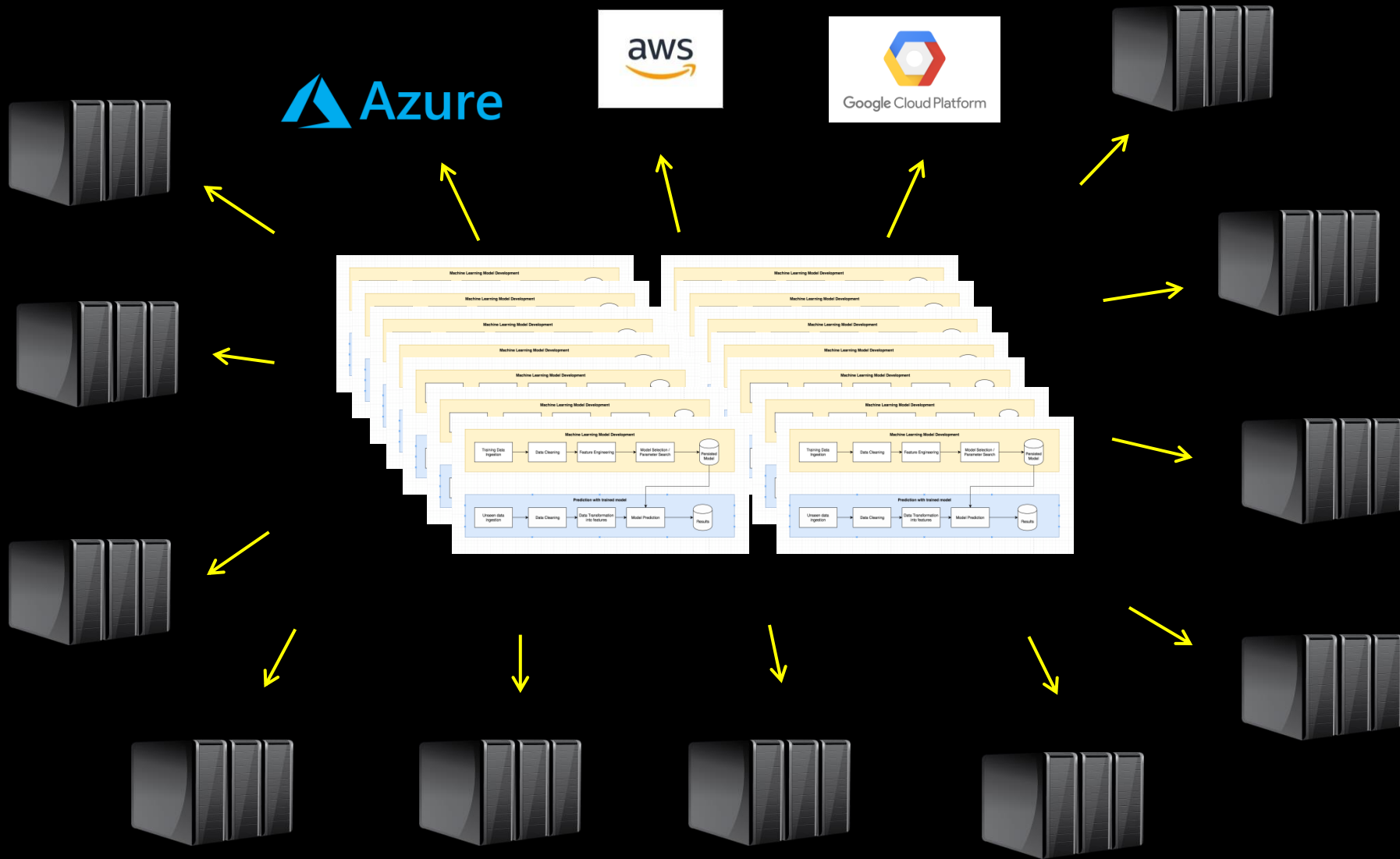
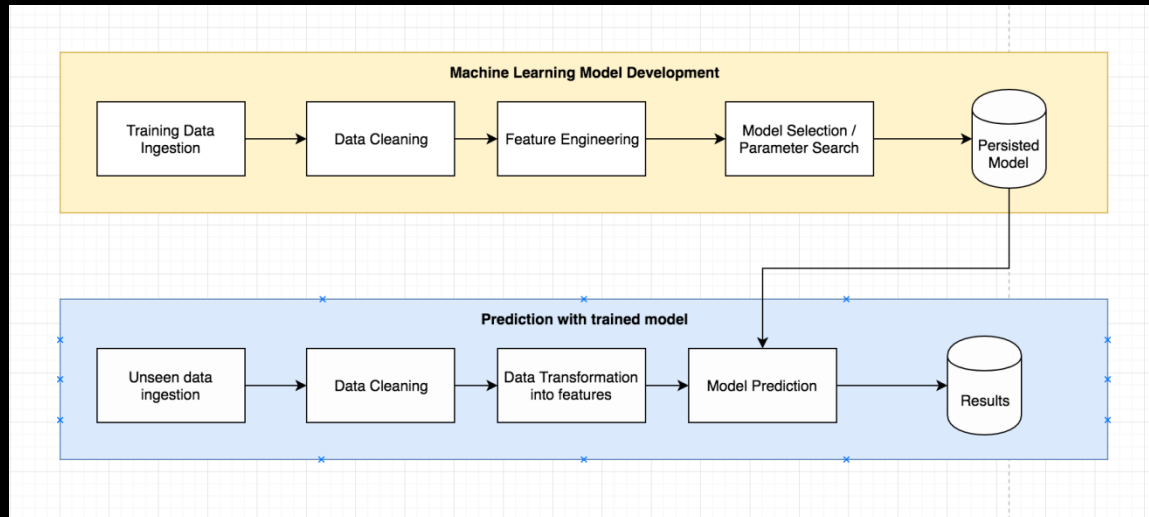


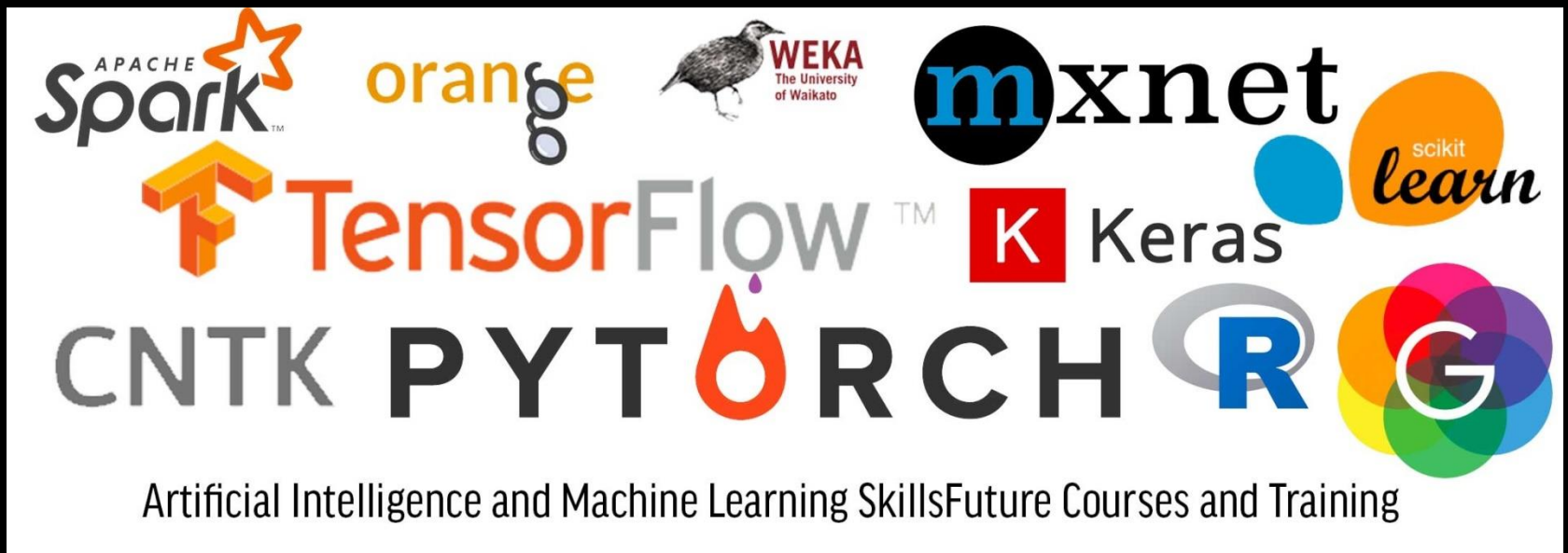
Image : <https://axsauze.github.io/scalable-data-science>

데이터 플로우의 복잡성 증가

- 많은 데이터 워크플로우의 수
- 표준화된 추적없이 데이터가 수정됨
- 규모가 커질수록 플로우와 스케줄링의 복잡성을 관리하는 것은 불가능해짐

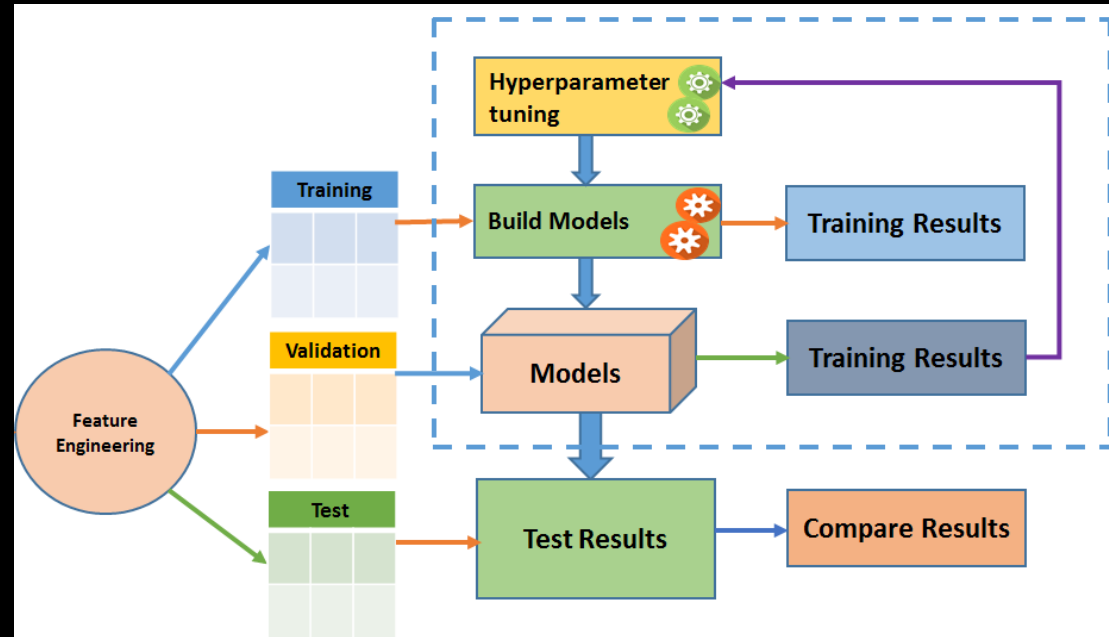


각자 선호하는 툴은 다 다릅니다.



까다로워지는 모델 서빙

- 각자 다른 버전의 모델이
각자 다른 환경에서 실행
- 점점 더 모델 배포와 버전
되돌리는 것이 복잡해짐



따라서 까다로워지는 track back

우리같은 엔지니어는
굳이 논문같은거 안보고도
모델이 잘못됐다는 걸
알 수 있습니다.

VS

데이터 파이프라인에
벌래같은 거 키우니까
데이터가 잘못 들어와서
모델이 뻥나는 거잖아



DATA engineer

그래서 전 뭐하는데요?



ML engineer



DATA scientist

어....그러니까 음....

전부다?!

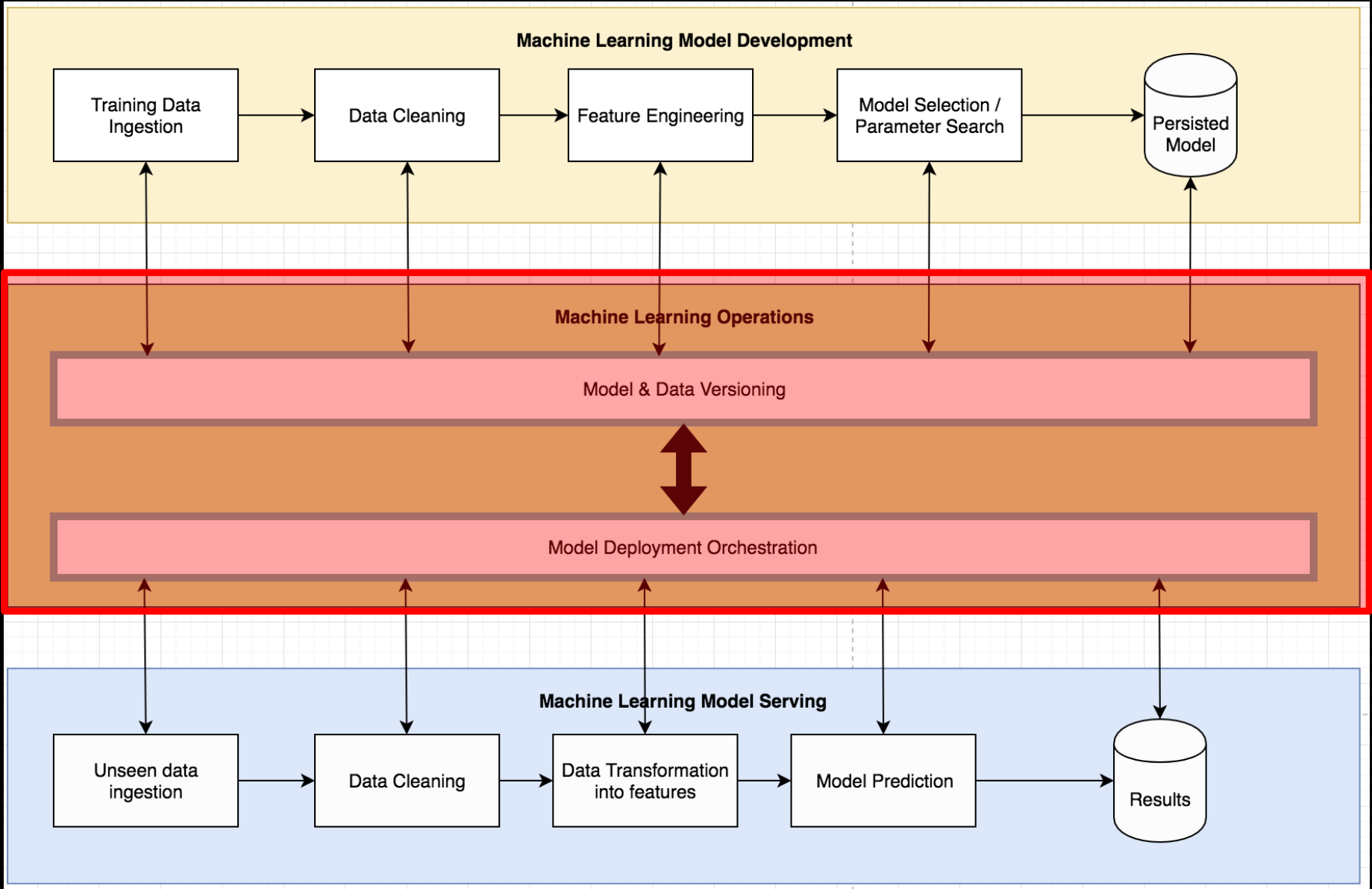


굳이... 할 필요가

ML engineer 는
데이터 파이프라인, 모델의 개발
그리고 서빙까지 신경써야 합니다.

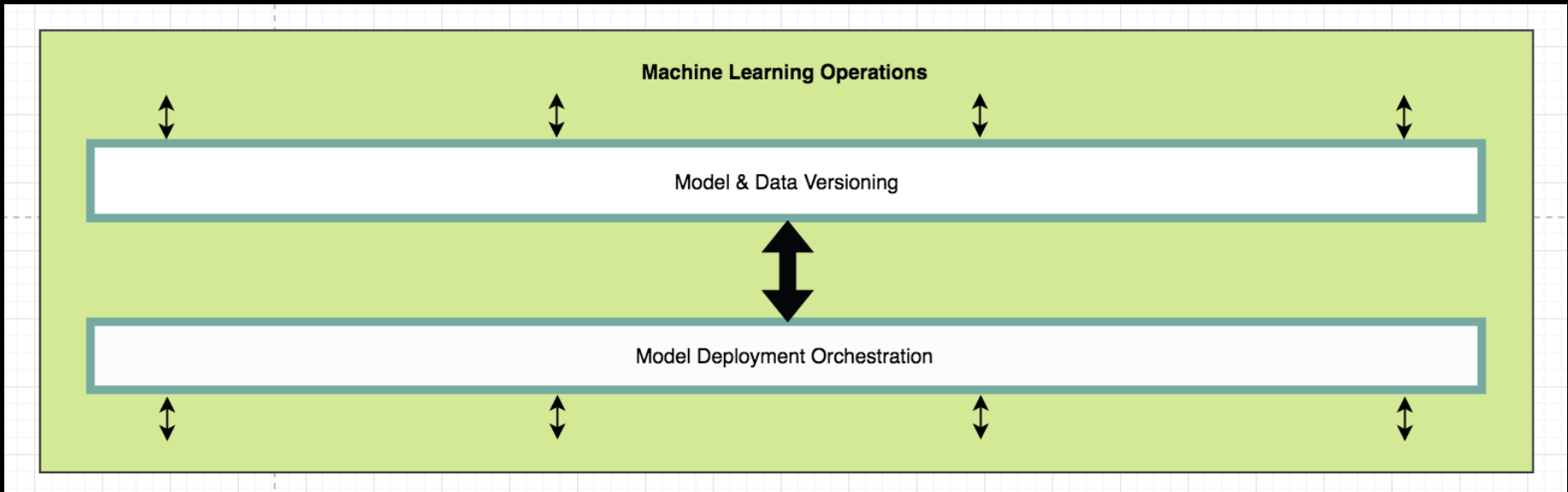
ML-OPS





결국 해결해야하는 것은 두가지입니다.

- Reproducibility
- Orchestration

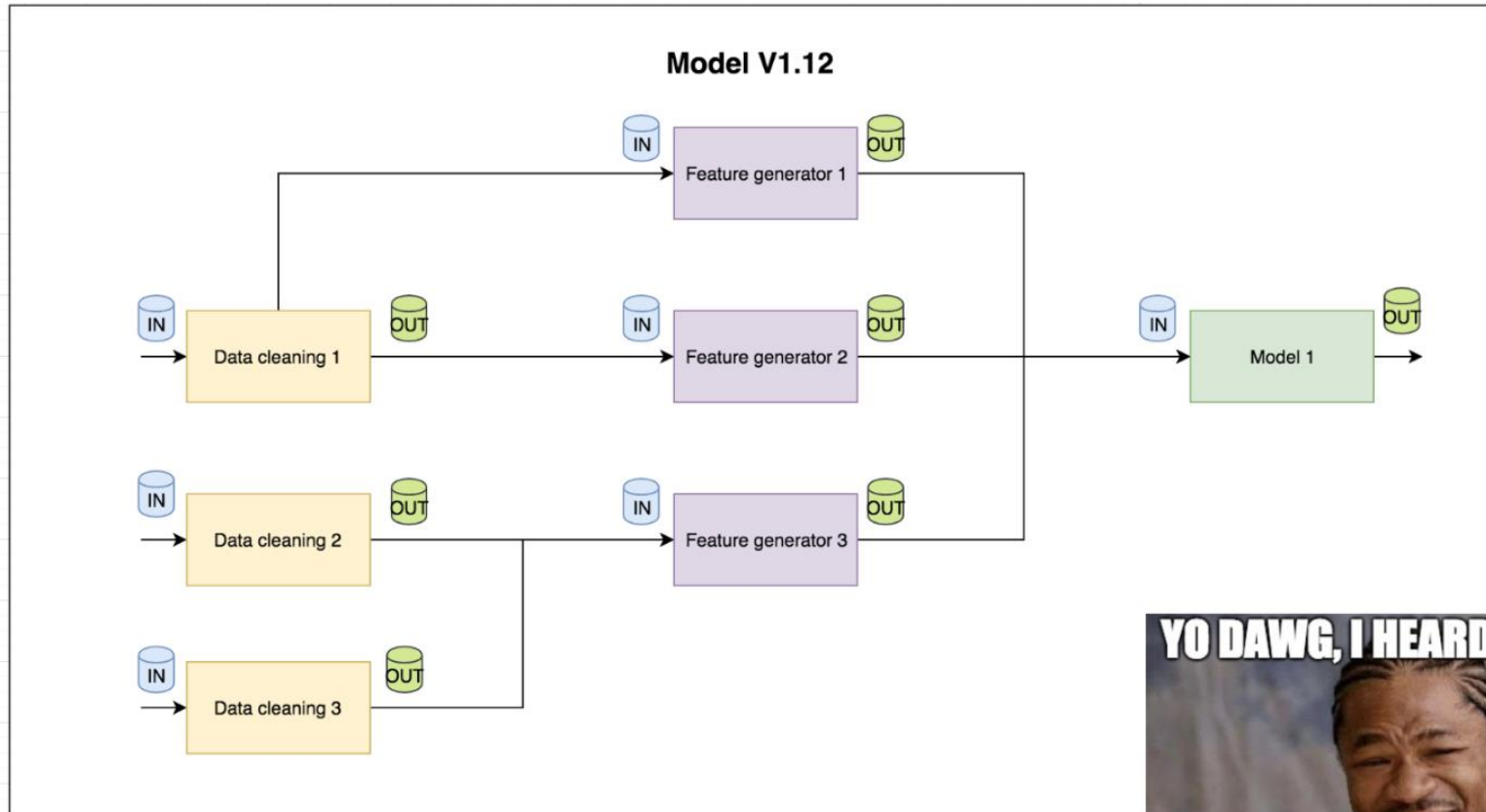


Reproducibility

를 위해서는 이것들을 해결해야 합니다.

- 버그나 에러가 났을때 trace가 가능해야 함
- 결과로 이어지는 단계의 투명성
- 컴포넌트 모듈화 -> 재사용성
- 여러 라이브러리와 프레임워크를 지원하는 추상화
- 버전 롤백될때 재현성이 유지되며 견고해야 함

We can abstract our entire pipeline and data flows



어떻게 하면 이 문제를 해결할까?

The screenshot shows the DVC website homepage. At the top is a navigation bar with the DVC logo, links for FEATURES, DOC, CHAT, GITHUB, and SUPPORT, and a 'Get Started' button. The main content area has a light blue background. On the left, it says 'Open-source Version Control System for Machine Learning Projects' with a 'Download (Windows)' button and a 'Watch video' button. On the right, there are four terminal snippets showing DVC commands: `$ dvc add images`, `$ dvc run -d images -o model.p cnn.py`, `$ dvc remote add myrepo s3://mybucket`, and `$ dvc push`. Below this is a 'Learn more' link. At the bottom, it says 'DVC tracks ML models and data sets' and provides a brief description of DVC's purpose.

DVC

FEATURES DOC CHAT GITHUB SUPPORT [Get Started](#)

Open-source
Version Control System
for Machine Learning Projects

[Download \(Windows\)](#) [Watch video How it works](#)

[We're on Github](#) ★ 2884

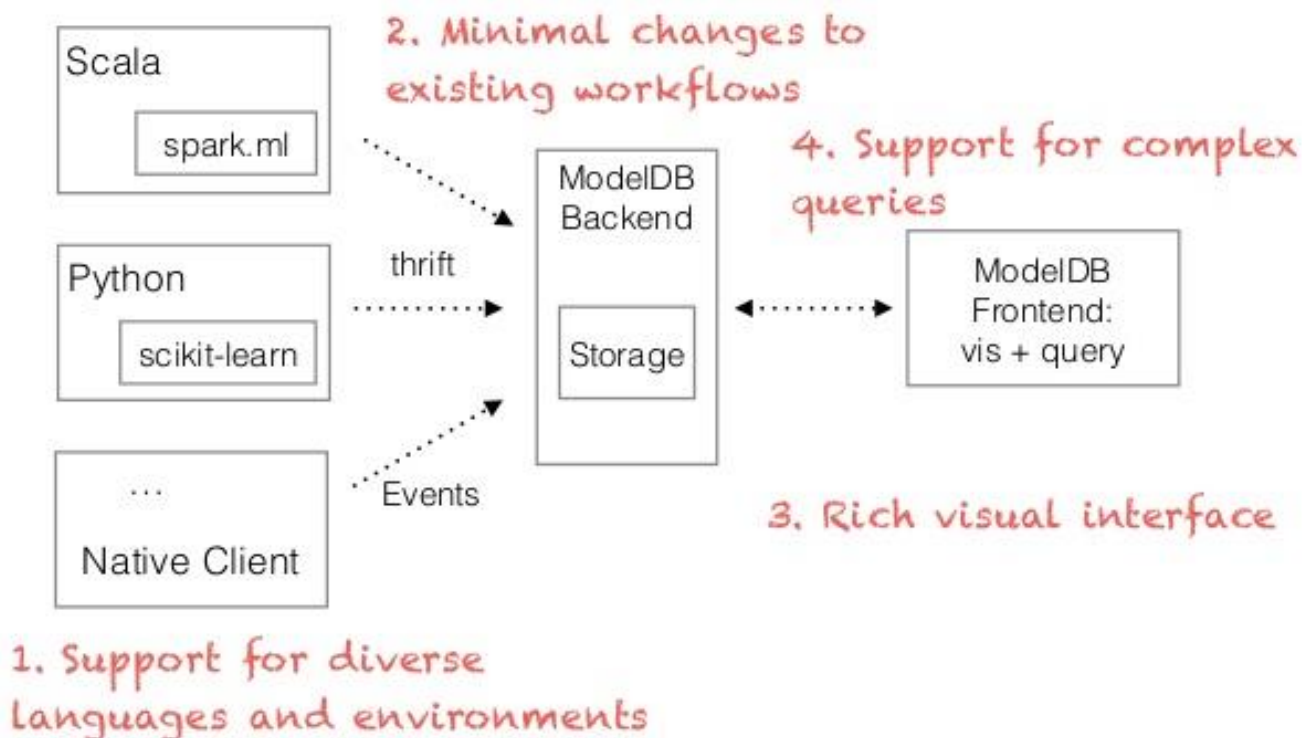
↓
[Learn more](#)

DVC tracks ML models and data sets

DVC is built to make ML models shareable and reproducible. It is designed to handle large files, data sets, machine learning models, and metrics as well as code.

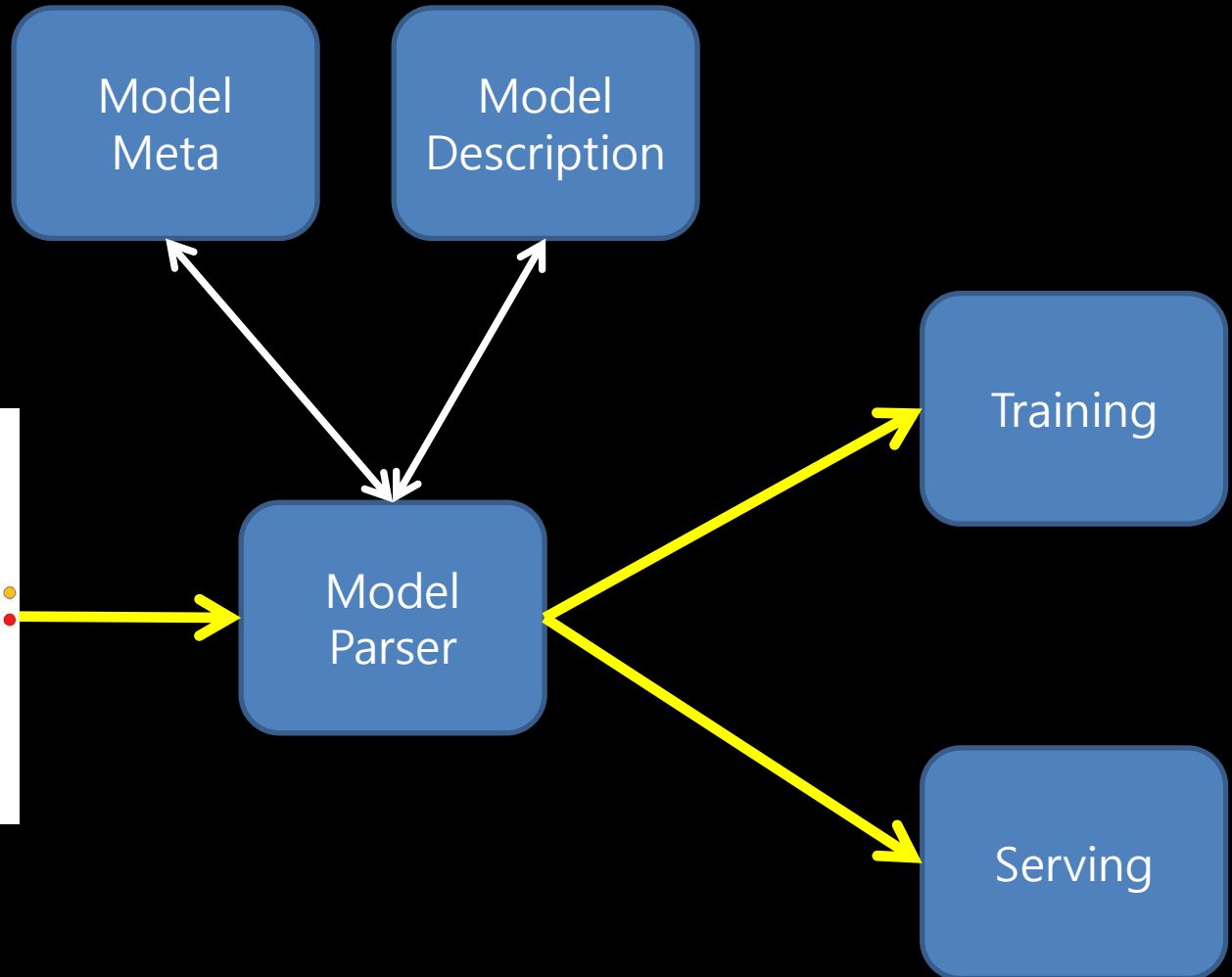
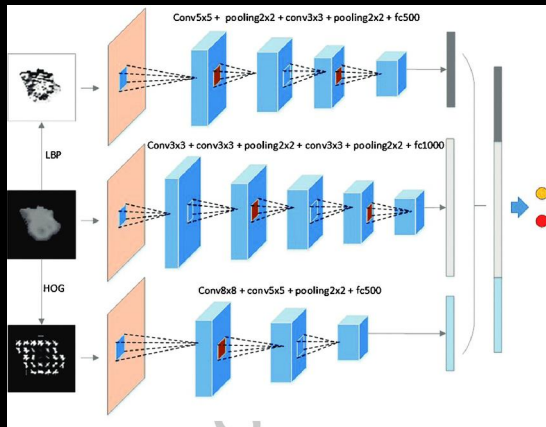
어떻게 하면 이 문제를 해결할까?

ModelDB Architecture & Design Decisions

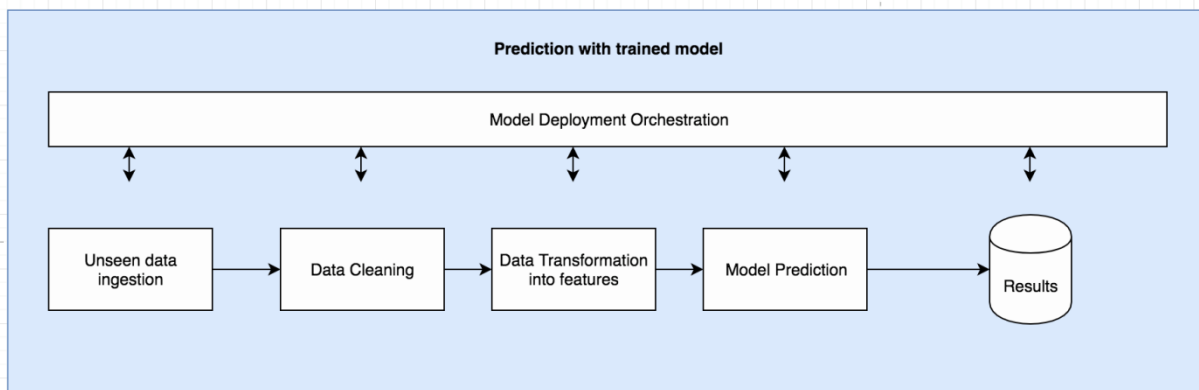


어떻게 하면 이 문제를 해결할까?

Model Description Script



Orchestration



일반 웹 CI/CD/모니터링
하고 비슷하다고 생각할 수 있는데

좀 다릅니다.

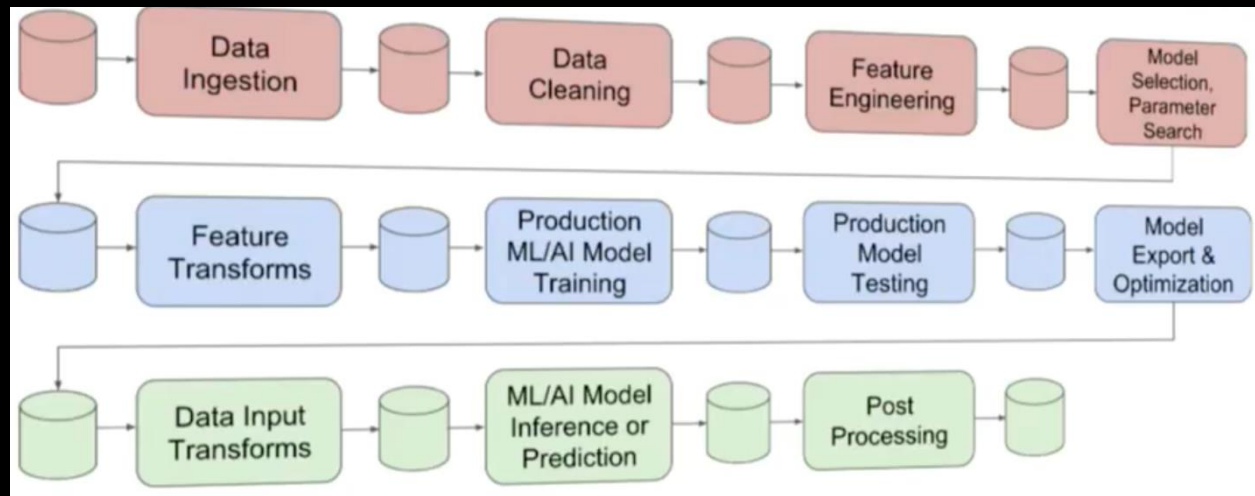
모니터링

소프트웨어 모니터링 외에

- 모델 퀄리티
- 데이터의 분포
- 데이터의 형태

WE CAN STOP PRAYING TO THE DEMO GODS

My program didn't work, and I was desperate. So...



결국 기도메타

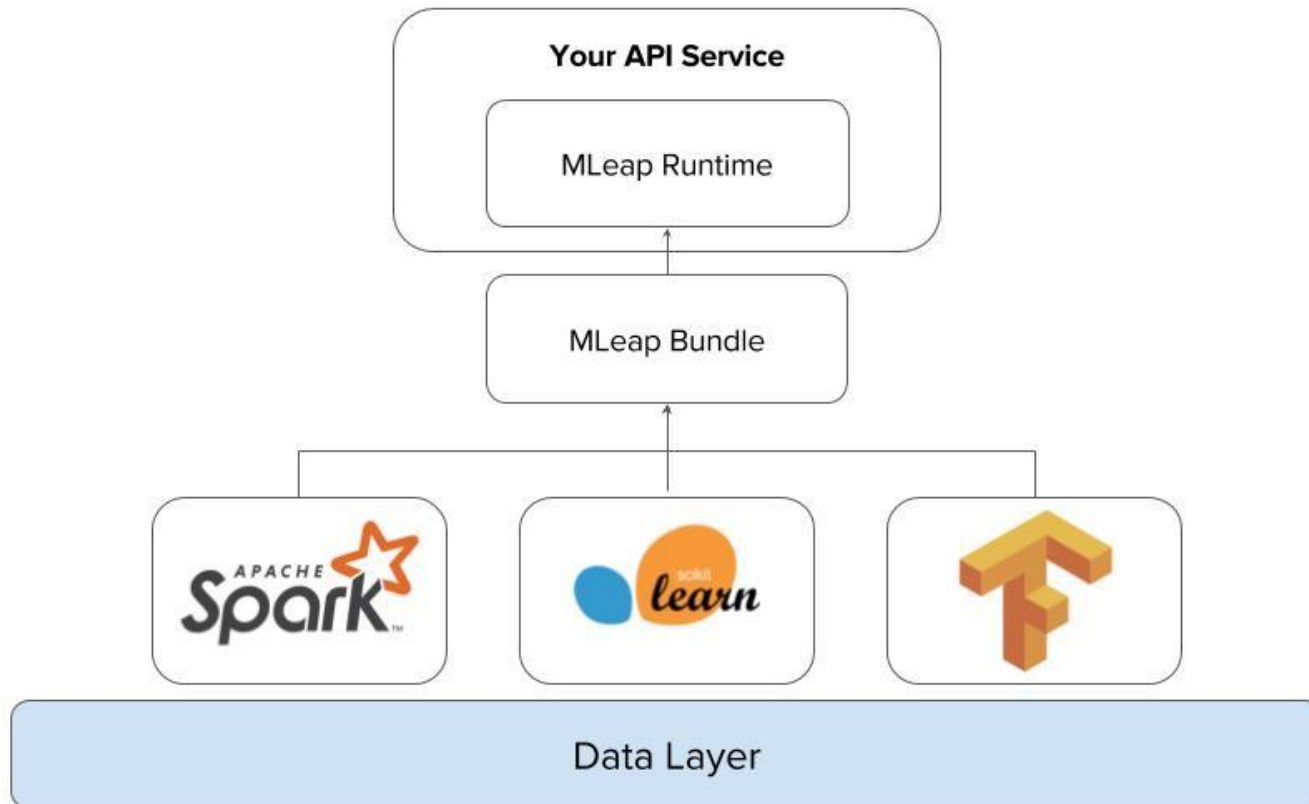
COMPUTATIONAL RESOURCE ALLOCATION

THIS IS A HARD PROBLEM

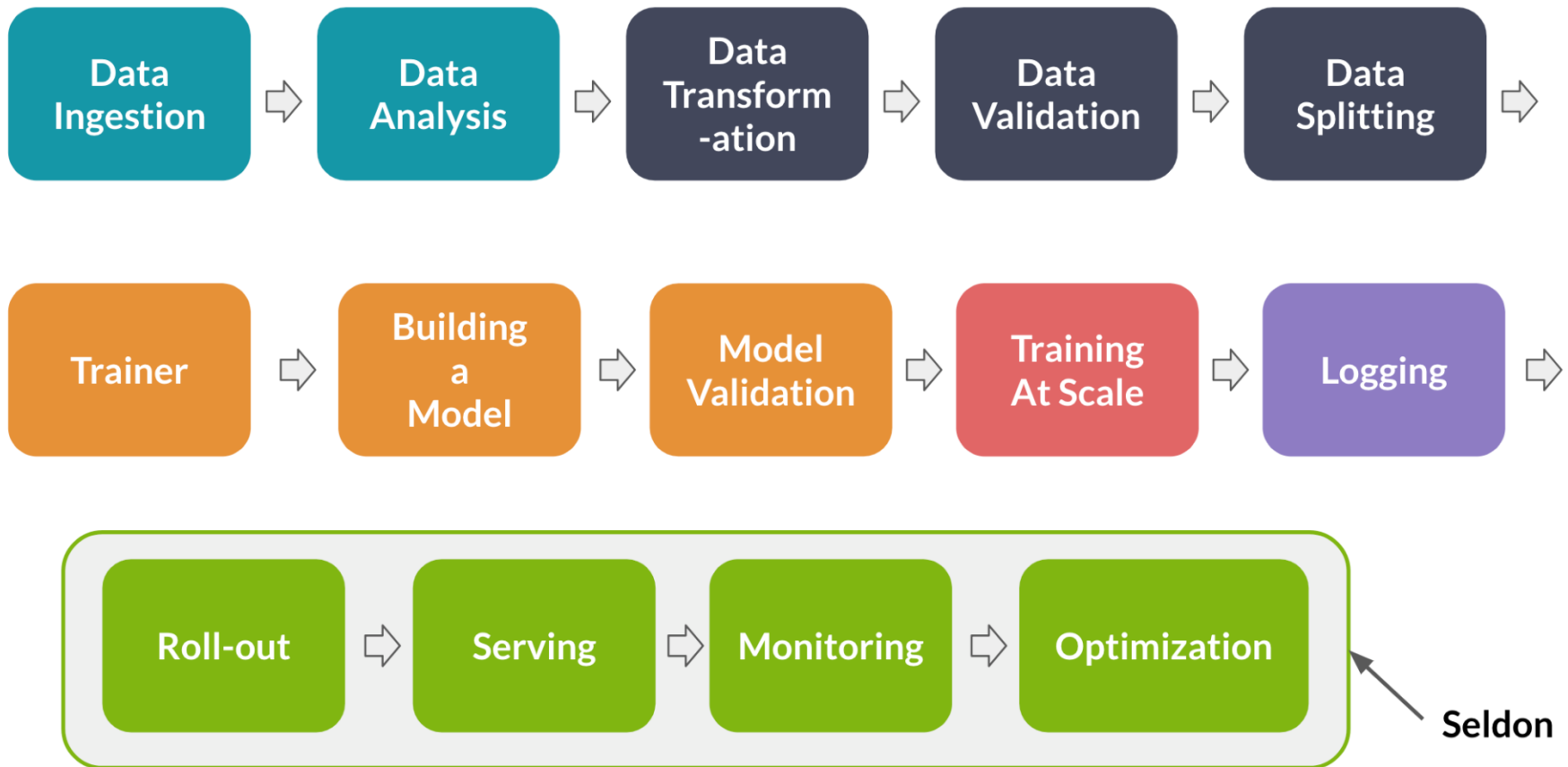
오케스트레이션 레벨에서 리소스 추상화

- An ETL framework
- A HDFS-based service
- A Kubernetes cluster
- Any distributed framework!

MLEAP



SELDON-CORE



SELDON-CORE

1. Package

Create REST or gRPC dockerized microservice .

2. Describe Deployment

Create/update Kubernetes resource manifest for deployment graph.

3. Deploy

Manage and analyze the performance of live deployments.

STRONG FOCUS ON MODEL ORCHESTRATION

2. Seldon Deploy

(UI, Collaboration, Control, Audit)

MAB
(Multi-Arm Bandits)

Outlier Detection

Explanation

Bias Detection

1. Seldon Core

(runtime ML graph engine)

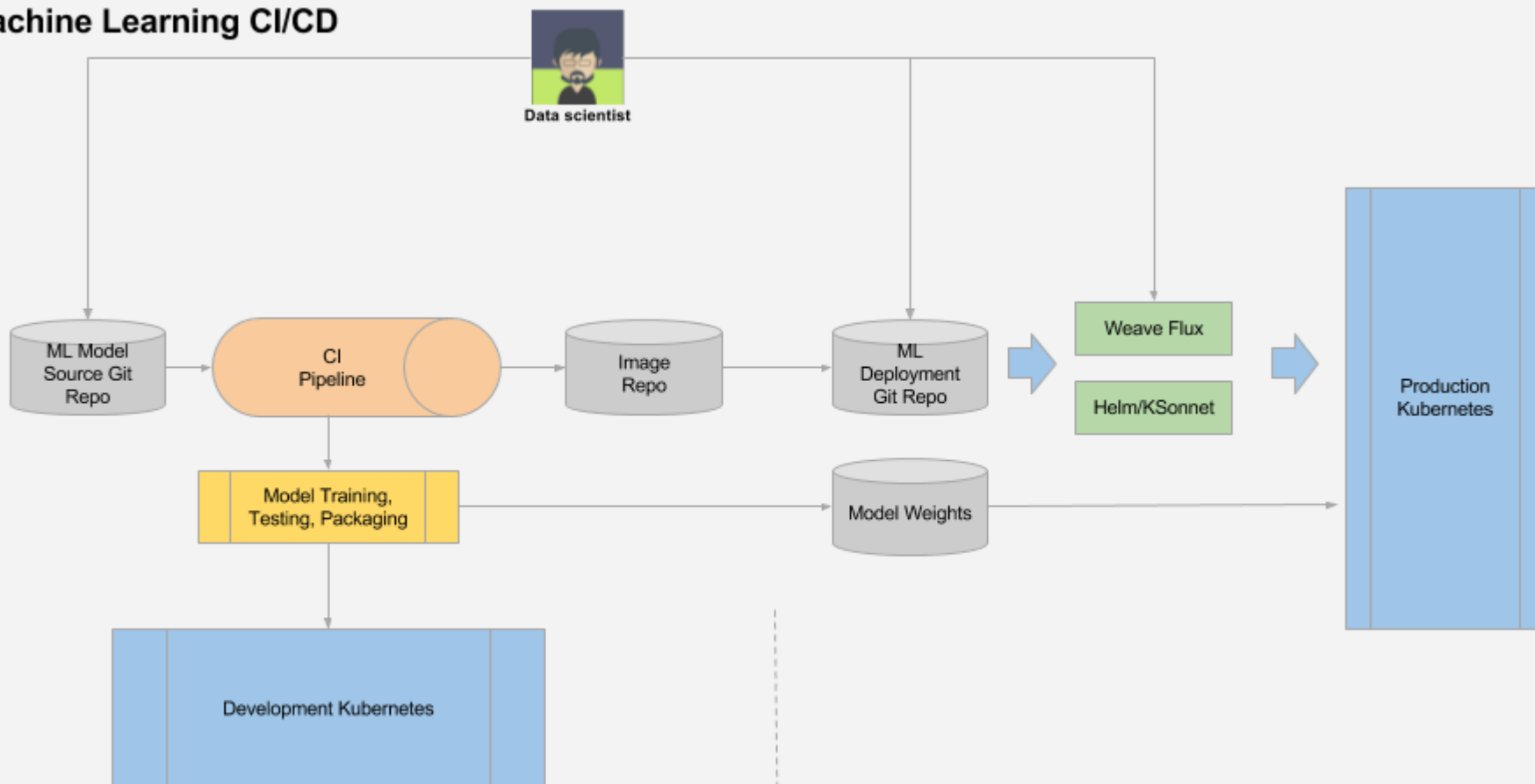
Microservices - Istio service mesh (optional)



kubernetes

TACKLING THE CI/CD CHALLENGE

Machine Learning CI/CD



이게 다 귀찮으시면

MLflow

kubeflow

Backend AI

아빠!
어디가?
추천은 얼음캠핑
저녁 재료 장보기

MBC

화색
아 그냥 이게
끝이예요?

네....