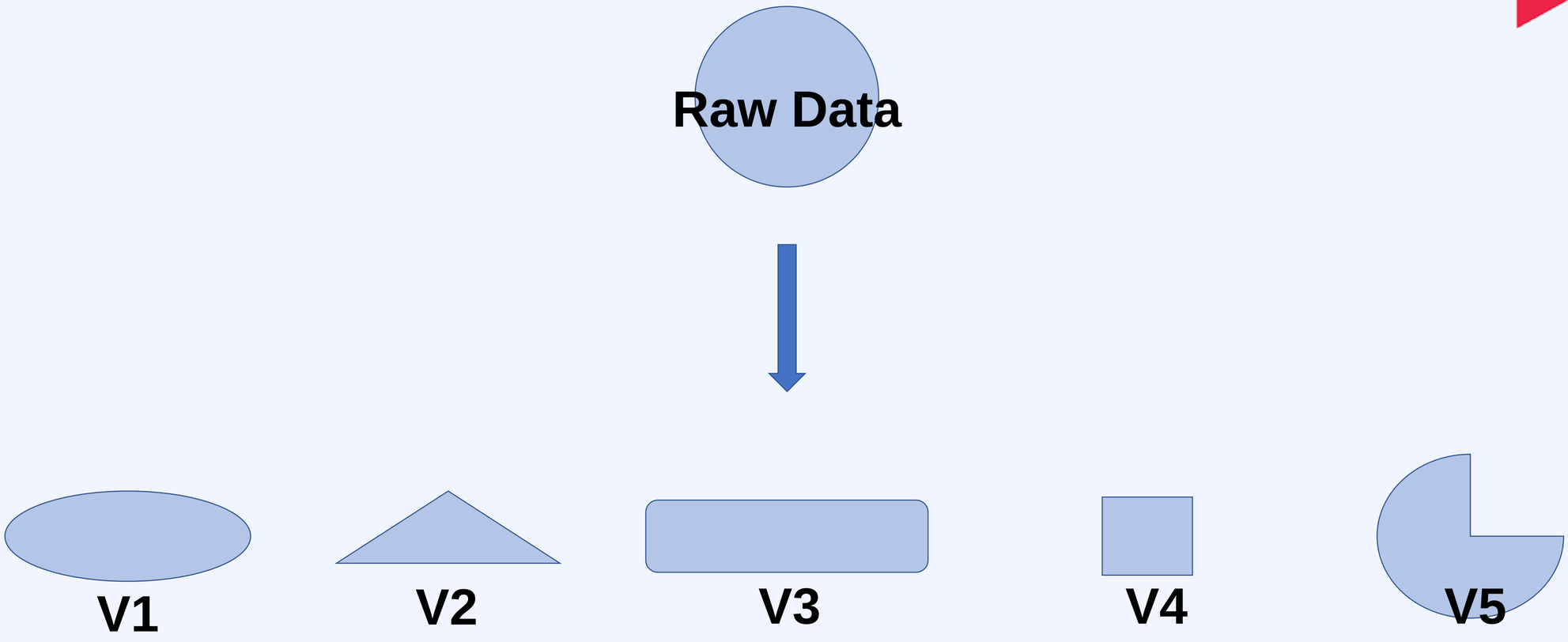
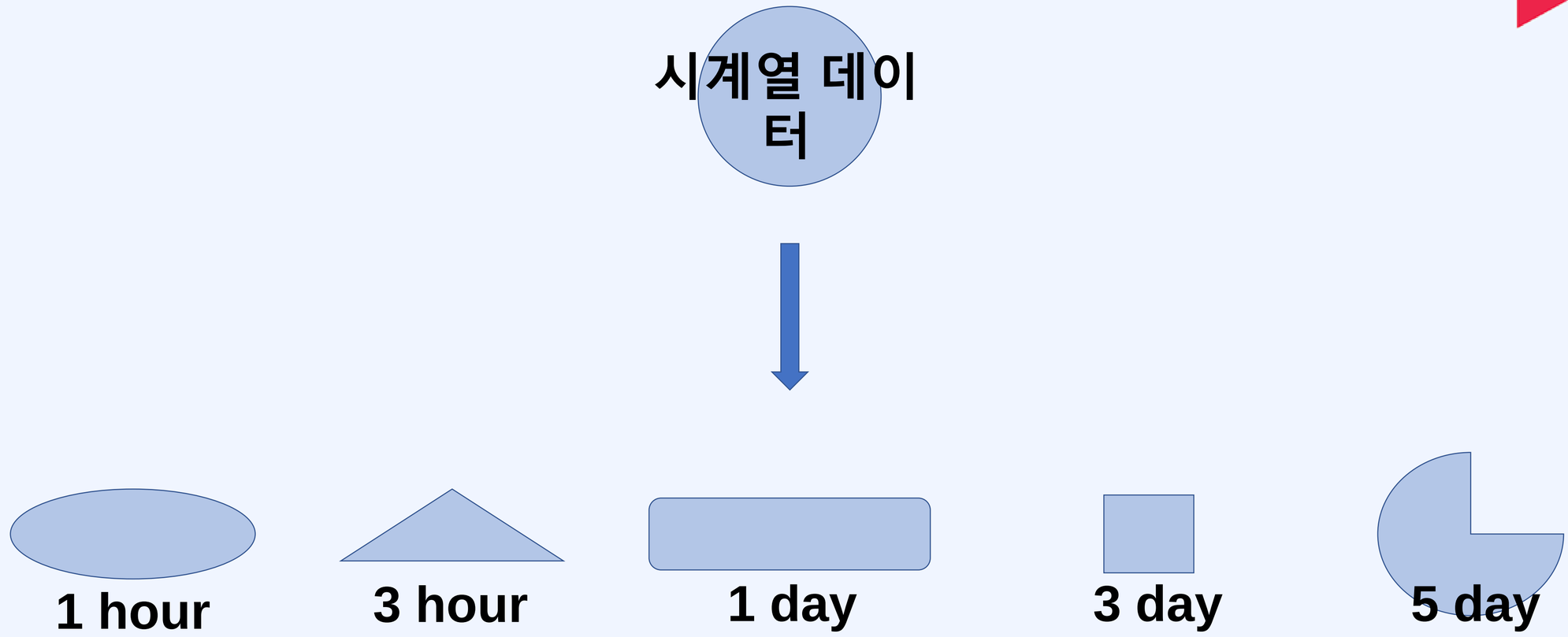
오픈소스를 통해 알아보는 MLOps 의 구성요소

1 Data Management 란

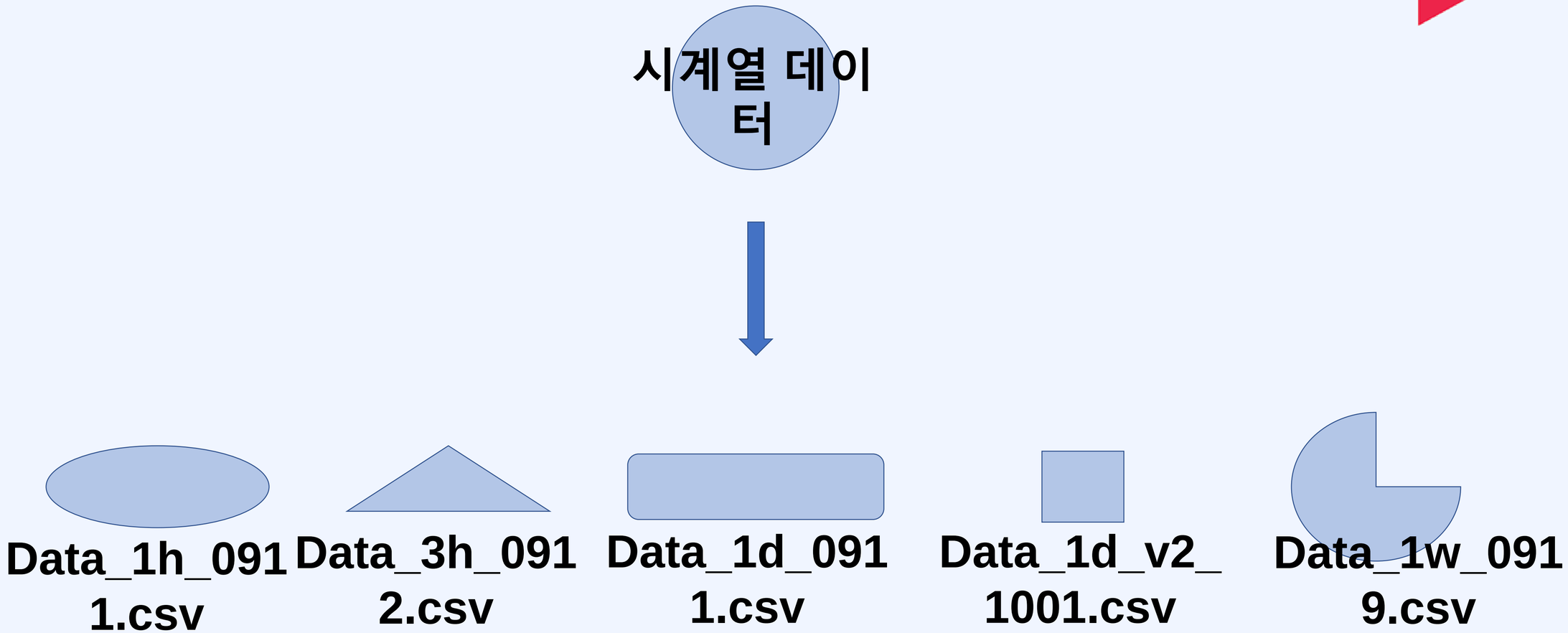
Data Management 란



For Example



파일 이름으로 구분하면 안 되나요?



Git

Source Code Version Control Tool

Github is not for Large Data

Git

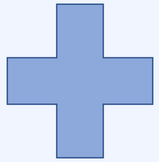
GitHub

GitLab

Bitbucket

For Large Data : Git + a

Git



- **DVC**
- **Pachyderm**
- **Delta Lake**
- **Dolt**

DVC

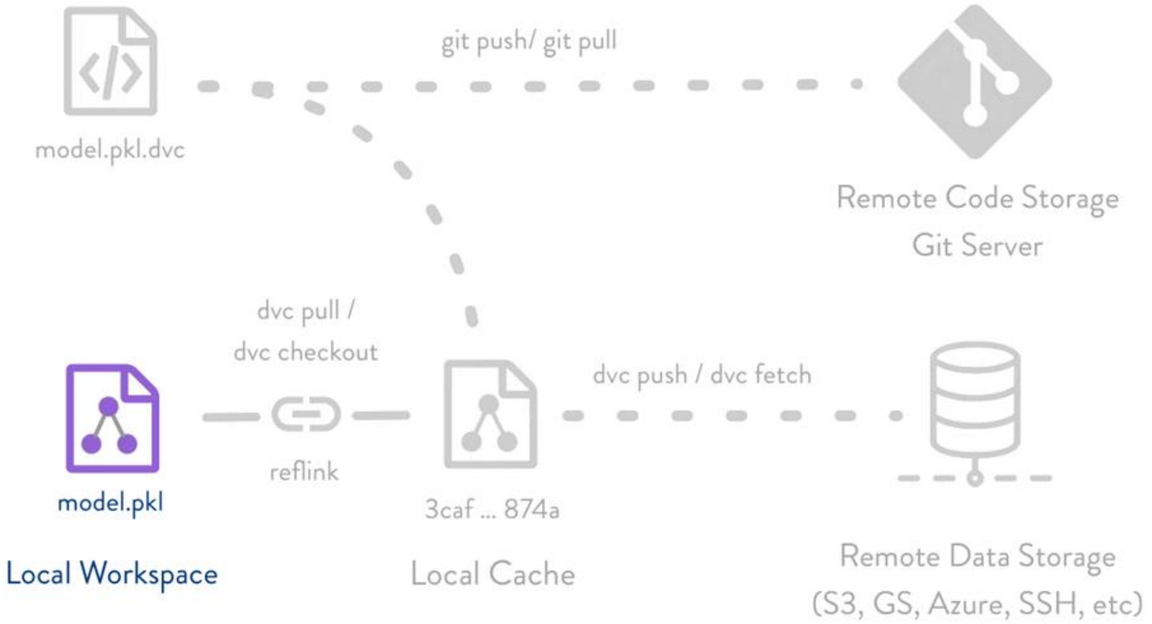
Data Version Control

https://dvc.org

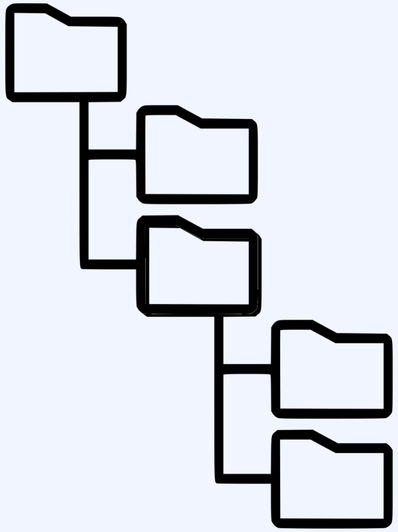
DVC 특징

- 대부분의 스토리지와 호환 (amazon s3, google drive, …)
- GitHub 외의 GitLab, Bitbucket 등의 대부분의 git 호스팅 서버와 연동
- Data Pipeline 을 DAG 로 관리
- Git 과 유사한 인터페이스

DVC 저장 방식



Git 특징



보고서_final

보고서_final_final

보고서_final_진짜_final

보고서_final_진짜_진짜_final

Git 과 GitHub

Git

GitHub

Git 과 GitHub

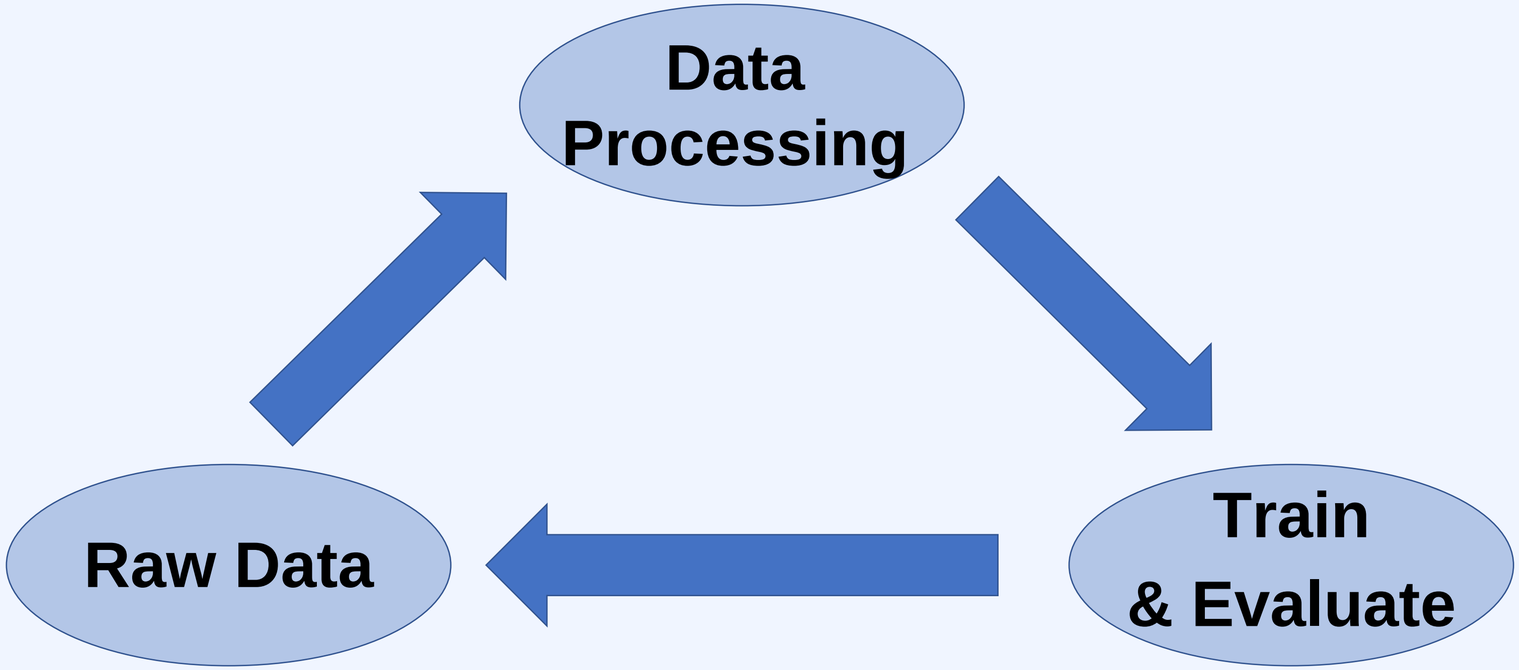
Git

- **git pull**
- **git push**
- **git add**
- **git commit**

오픈소스를 통해 알아보는 MLOps 의 구성요소

2 Model Management 란

ML Model 의 Life Cycle



어떤 정보를 기록해두어야 할까요?

Model 소스코드

Evaluation Metric 결과

사용한 parameters

Model.pkl 파일

학습에 사용한 data

데이터 전처리용 코드

전처리된 data

……..

모델 관련 정보의 저장 방식

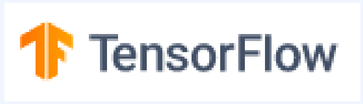
저장 방식이 제각각 다르다면
통합하고 비교하기 쉬울까요?

ML Model Life Cycle 관리의 어려움

- 비슷한 작업이 반복적으로 일어난다
- Dependency 패키지들이 많으며, 버전 관리가 어렵다
- 사람 Dependency 가 생긴다
- 테스트하기 어렵다
- Reproduce 되지 않는 경우가 많다
- Model 학습용 코드를 구현하는 사람과 Serving 용 코드를 구현하는 사람이 분리되어 있다.
- .....

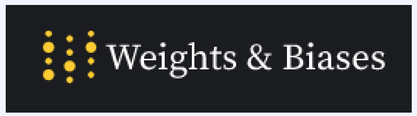
다양한 Model Management Tools

MLflow



Tensorboard

Neptune



Weights & Biases

Comet.ml

MLflow 의 구성 요소

- mlflow tracking
- mlflow projects
- mlflow models
- mlflow model registry

MLflow tracking

mlflow tracking

mlflow projects

mlflow models

mlflow model registry

MLflow tracking 의 구조

- https://medium.com/engineering-at-ooba/machine-learning-lifecycle-management-using-mlflow-64d3bd75b6bd

MLflow 장점

쉬운 설치

쉬운 Migration

대시보드 제공

다양한 Client API 제공

다양한 Backend Storage 연동 지원

다양한 Artifact Storage 연동 지원

……..

Let‘s use MLflow

- https://github.com/mlflow/mlflow

오픈소스를 통해 알아보는 MLOps 의 구성요소

3 Model Serving 이란

I have a model, now what ?

Serving 이란?

ML Model 을 서비스화하는 것

서비스 제공 방식

HTTP API Request

챗봇과의 대화

Netflix 영상 좋아요 버튼

Youtube 구독 버튼

네이버 길찾기 버튼

……..

서빙 단계에서 막히는 이유

모델 개발과 소프트웨어 개발의 방법의 괴리

모델 개발과정과 소프트웨어 개발 과정의 파편화

모델 평가 방식 및 모니터링 구축의 어려움

서빙의 간편화를 도와주는 도구

Seldon Core

TFServing

KFServing

Torch Serve

BentoML

Flask & Seldon Core

Flask

Seldon Core

오픈소스를 통해 알아보는 MLOps 의 구성요소

4 Model Monitoring

I deployed Model, Now What ?

Why Monitoring ?

Possible Issues

Inference Server 가 다운된다면 ?

거래소 서버가 터진다면 ?

갑자기 전쟁이 난다면 ?

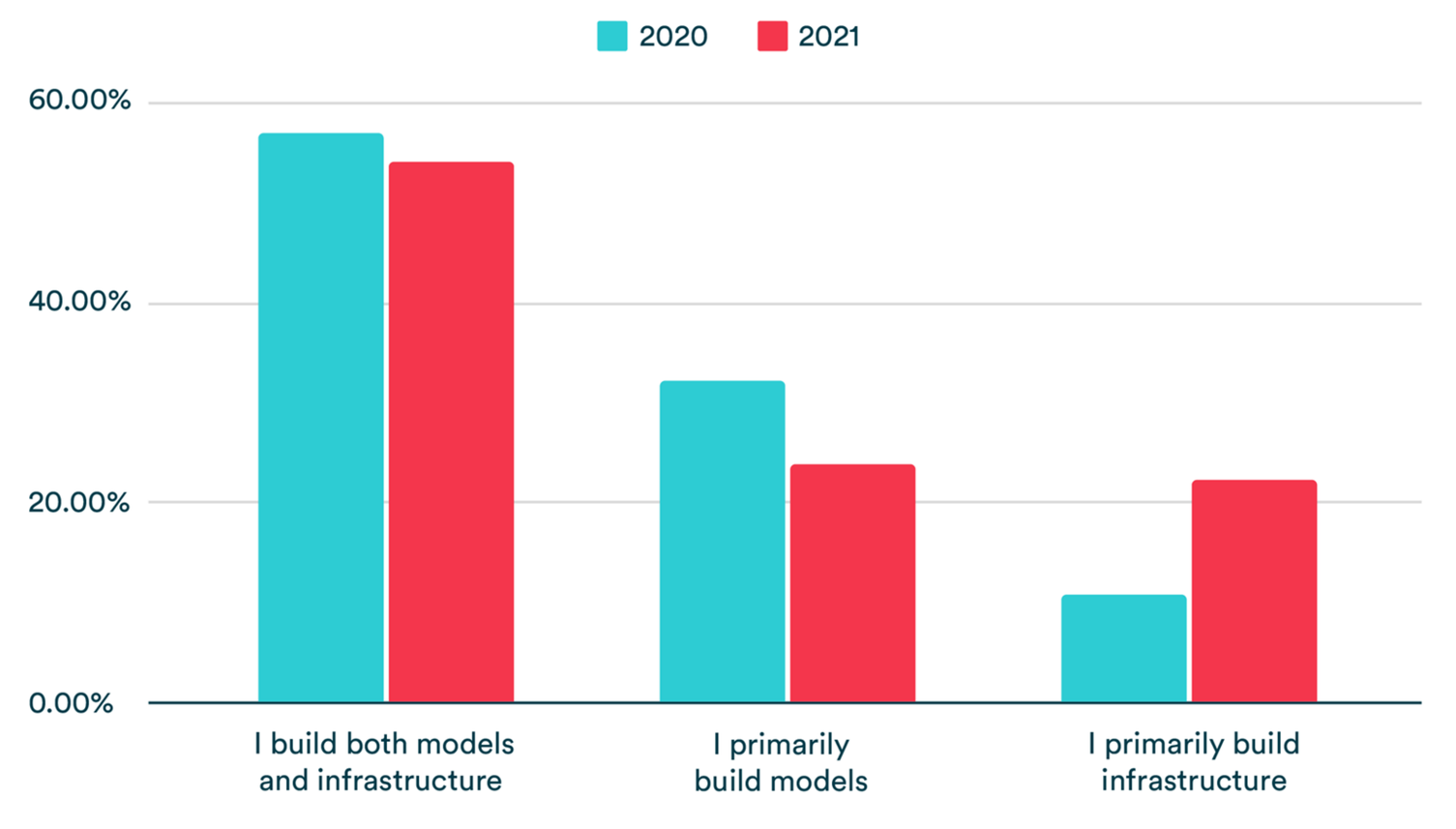
갑자기 미 연준이 테이퍼링 발표를 한다면 ?

갑자기 일론 머스크가 트위터를 한다면 ?

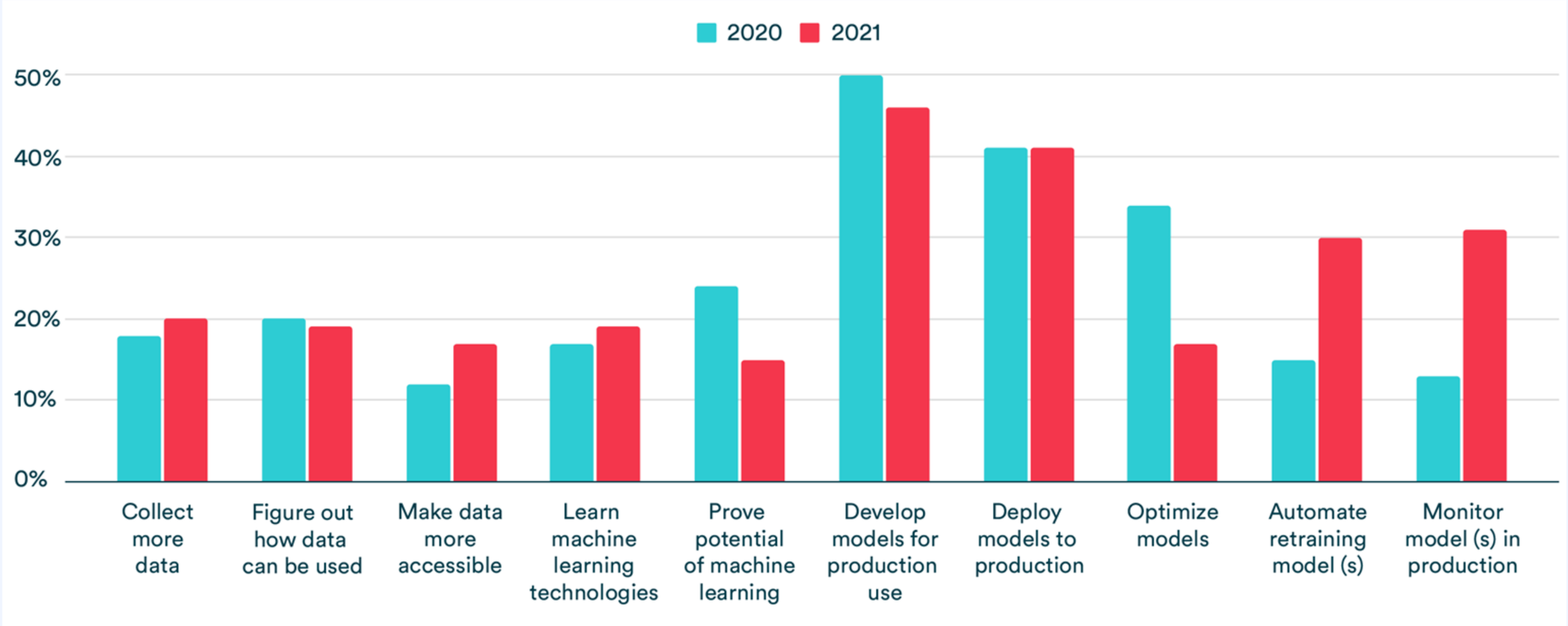
……..

최근에 사용한 툴은 어떤 분야인가요?

당신의 업무를 가장 잘 설명하는 문장은 무엇인가요 ?



다음 세 달 동안 어떤 분야에 집중할 계획인가요?



The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction, IEEE, Google

***Monitor 1*: Dependency changes result in notification**

***Monitor 2*: Data invariants hold in training and serving inputs**

***Monitor 3*: Training and serving features compute the same values**

***Monitor 4*: Models are not too stale**

***Monitor 5*: The model is numerically stable**

***Monitor 6*: The model has not experienced dramatic or slow-leak regressions in training speed, serving latency, throughput, or RAM usage**

***Monitor 7*: The model has not experienced a regression in prediction quality on served data**

서비스 제공 방식

ML 관련

- **Input Data Distribution**
- **Feature Distribution**
- **Output Data Distribution**
- **Performance (Evaluation)**
- **Model Stability**
- **……..**

Ops 관련

- **Request Latency**
- **Request Error Rate**
- **CPU, Memory Utilization**
- **Disk I/O**
- **Network Traffic**
- **……..**

I monitor Model, Now What ?

ML 서비스 모니터링이 어려운 이유

쇼핑몰 페이지 접속이
원활하지 않은 상황

- **Request Latency, Throughput**
- **Request Error Rate**
- **CPU, Memory Utilization**
- **Disk I/O**
- **Network Traffic**

쇼핑몰 고객 맞춤 추천 제품의
판매량 하락

?

모니터링을 위한 오픈소스