AIOPS PROJECTS

Neha Purohit

AGENDA

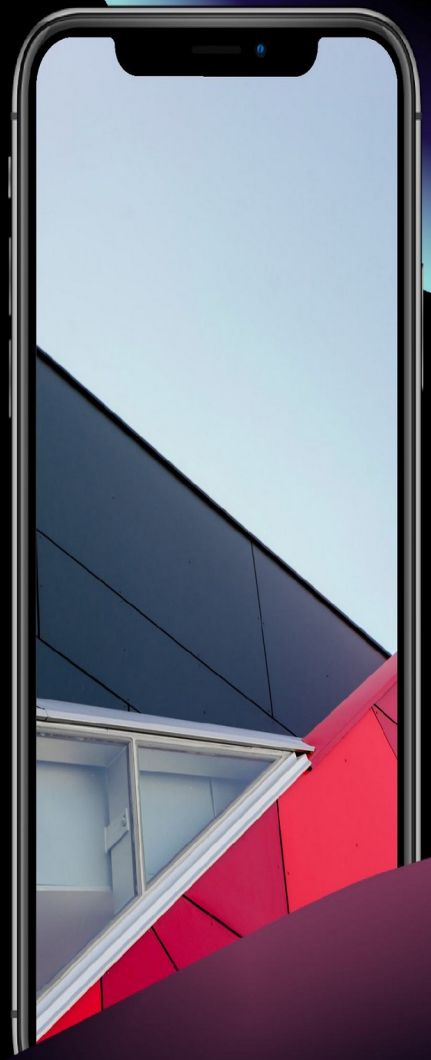
TOPICS COVERED

INTRODUCTION

- Machine Learning Pipeline
- Orchestration of pipeline
- Tensorflow extend

INTRODUCTION

Adoptaion of AIOPS to move machine learning model from an experiment to a robust production system.



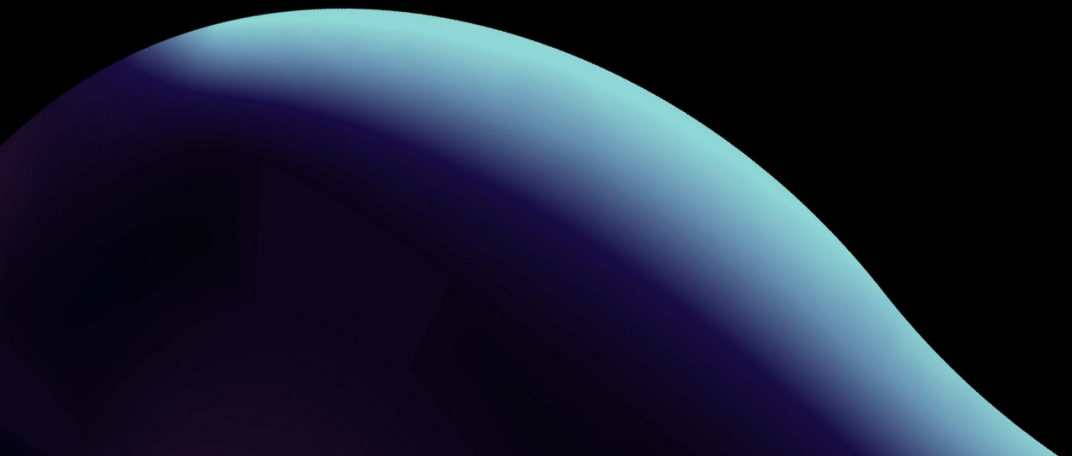
WHY AUTOMATED MACHINE LEARNING PIPELINE?

Maintenance of entire machine learning pipeline becomes easy.

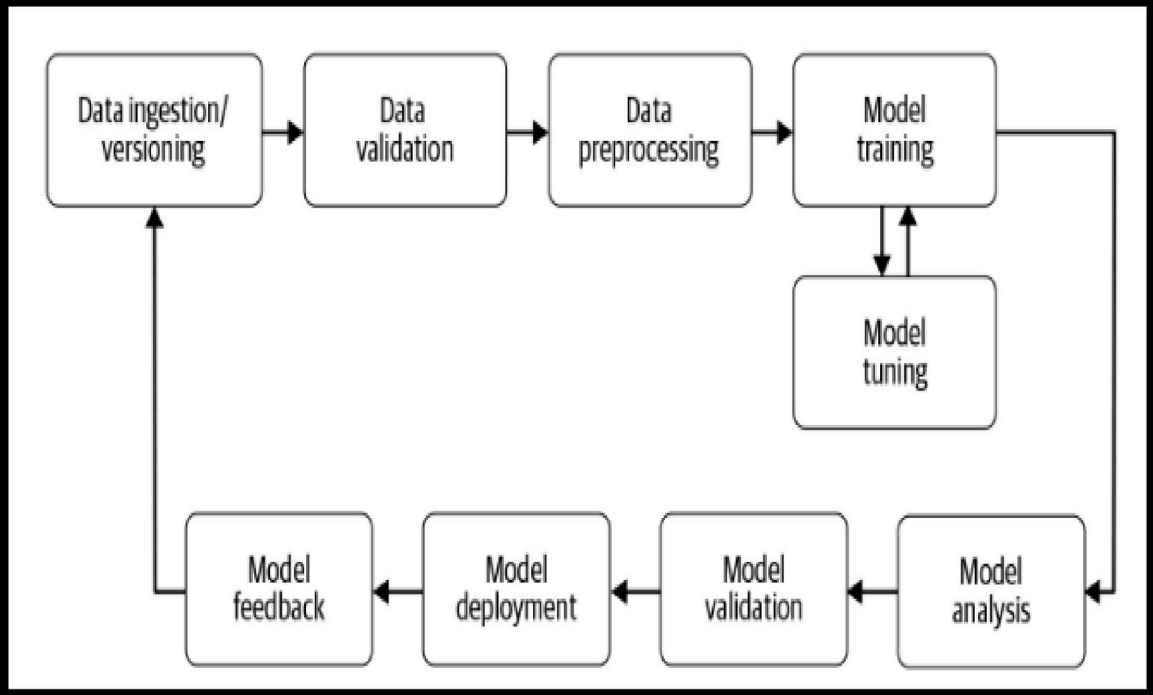
Retraining is simple due to automated pipeline.

Experiment tracking and model release management

Easy to update existing models



Machine Learning Pipeline



Data Ingestion

Defining the way to extract data from various sources

Perform data wrangling to transform data into a format that the following components of machine learning pipeline can digest.

Integrate data version control.

Data Validation

Check statistic of datasets such as range, number of categories, and distribution of categories.

Detecting anomalies in dataset such difference between training and validation set.

Stopping the machine learning pipeline if validation highlights anything out of ordinary

Data Preprocessing

Data can not be used directly to train machine learning model. Preprocessing is must on dataset so that machine learning model can be trained

Model Training and Tuning

Model training becomes difficult to manage on large training set. The efficient distribution of model training is crucial.

Model tuning help us to find out optimal model hyper parameters for our model.

Model Evaluation/Analysis

We can use metric to evaluate model performance

Observing model prediction on different groups of dataset.

This workflow may require manual review by data scientist to evaluate reliability of model.

Model Versioning

Model versioning allow us to keep track of which model, set of hyperparamteres and datasets have been selected to produce model.

Model Deployment

Once model is trained, tuned and evaluated, we deploy model as an API.

Feedback loop

We can then measure the effectiveness and performance of the newly deployed model. During this step, we can capture valuable information about the performance of the model.

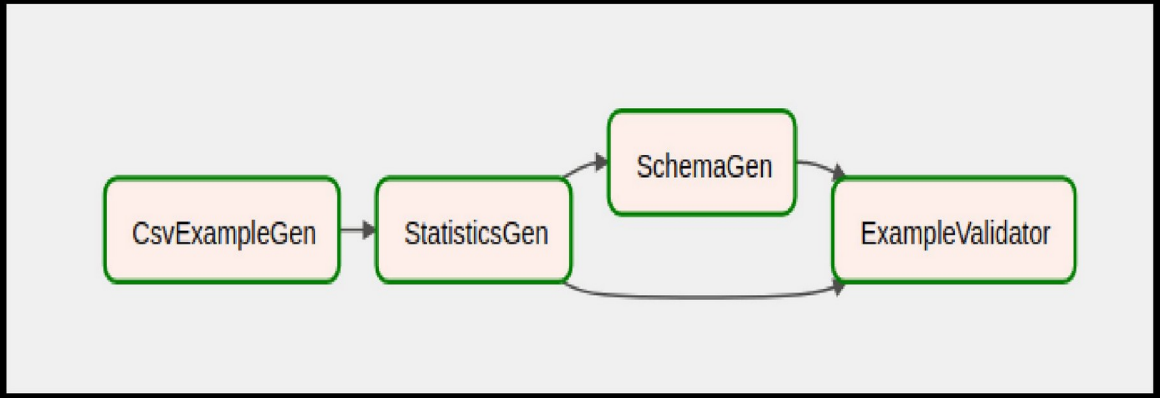
Pipeline Orchestration

All component of machine learning pipeline discussed in previous slides need to be executed or ochestrated so that execution take place in correct order.

We have many tool for orchestration:

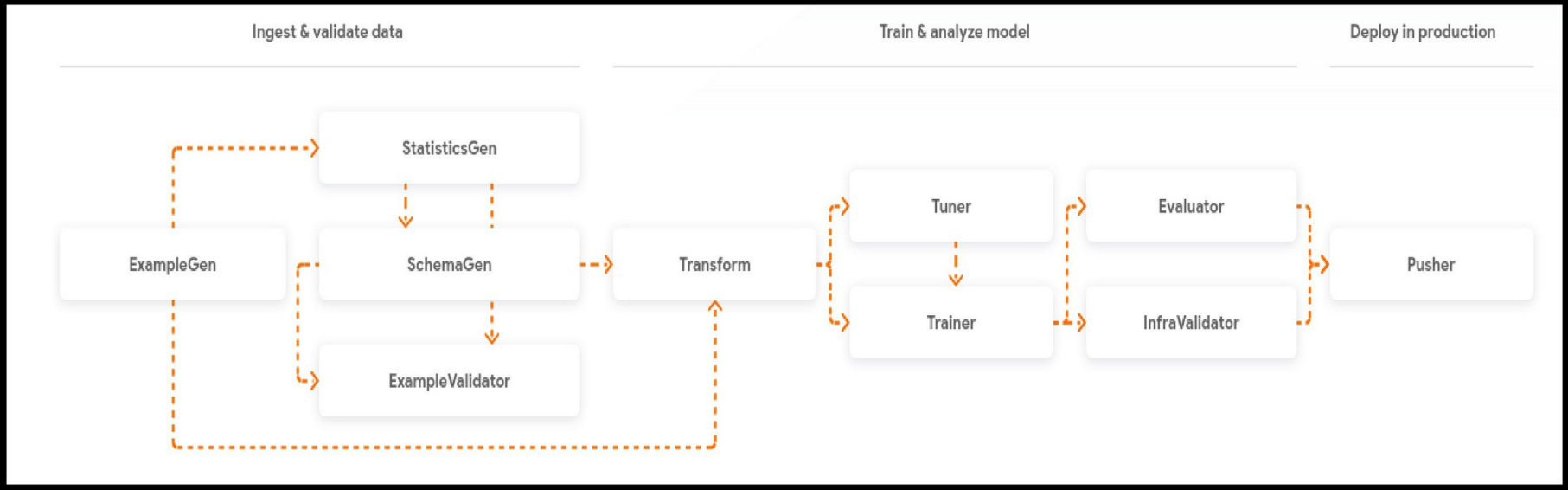
- Apache Airflow
- DVC
- Apache Beam

DAG (Directed Acyclic Graph)



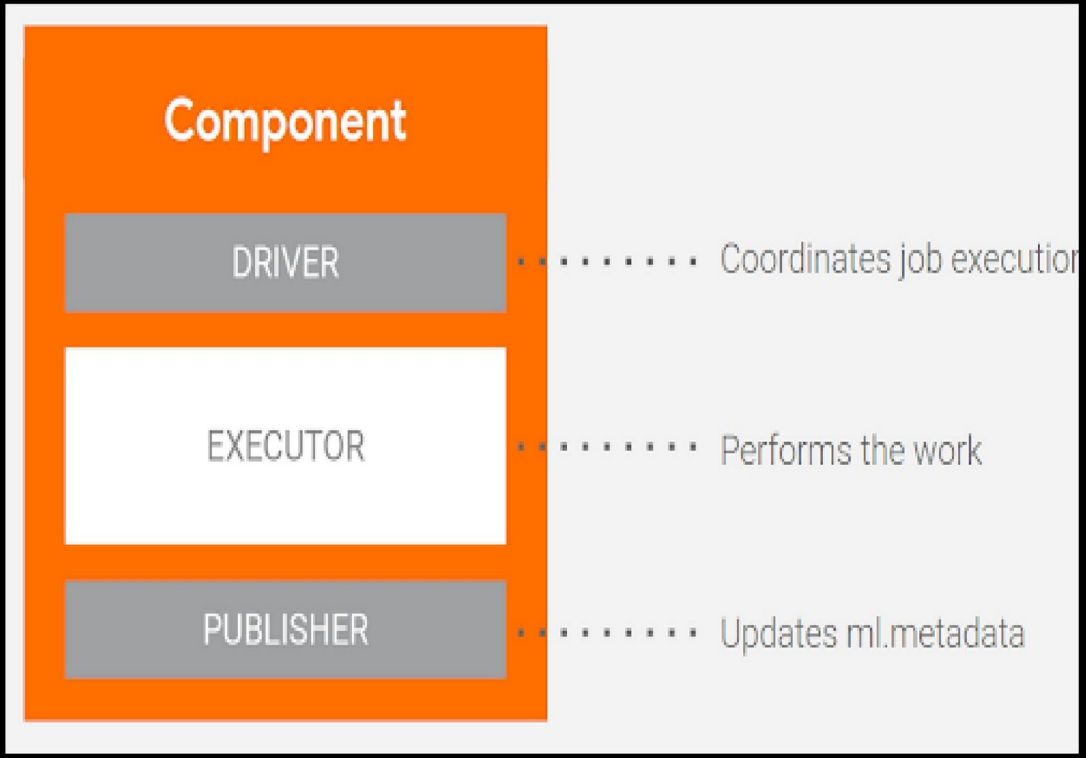
TensorFlow Extended (TFX) is an end-to-end platform for deploying production ML pipelines

A TFX pipeline is a sequence of components that implement an ML pipeline which is specifically designed for scalable, high-performance machine learning tasks. Components are built using TFX libraries which can also be used individually.



Installation of tfx

Overview of TFX Components



ML Metadata

The components of TFX "communicate" through metadata.

Meta data store information about references to artifact.

Audit trail of machine learning learning pipeline.

MLMD supports three types of backends

- In-memory SQLite database
- SQLite
- MySQL

Interactive Pipelines

We will implement a machine learning pipeline step by step and demonstrate its implementations through an interactive pipeline. The pipeline runs in a Jupyter Notebook, and the components' artifacts can be immediately reviewed.