HOCHSCHULE DER MEDIEN

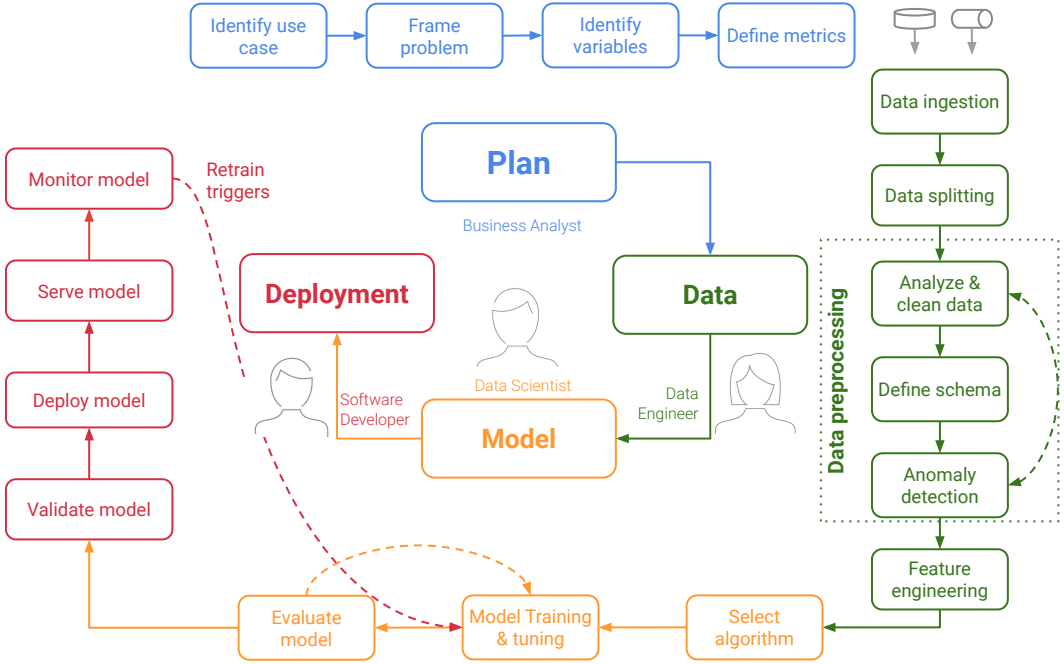
Data Science Lifecycle

& TensorFlow Extended

Prof. Dr. Jan Kirenz
HdM Stuttgart

Lifecycle
of an ML System

Plan | Data | Model | Deployment

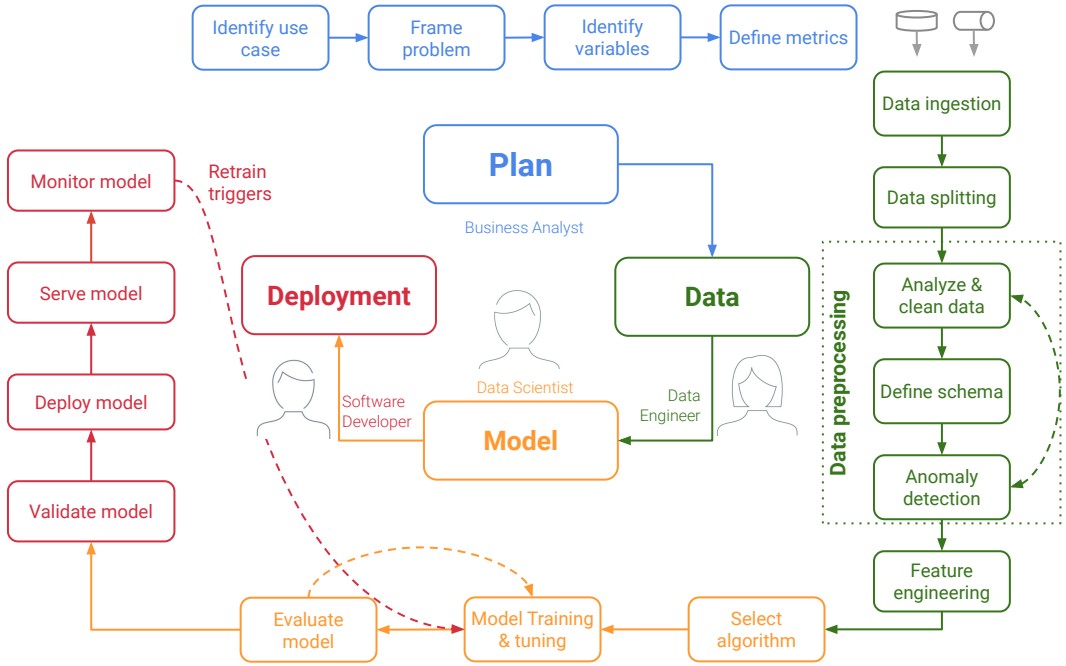


Common issues

- Lack of reuse and duplication
- Inconsistency (data, code, models)
- Manual and slow transition from PoC to production

Lifecycle
of an ML System

Plan | Data | Model | Deployment

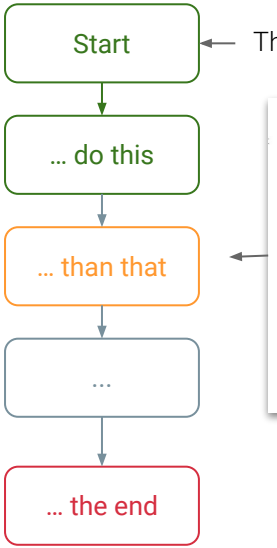


Components to solve these problems

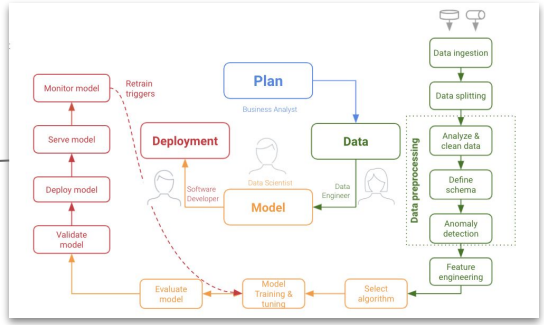
Data and feature management	Model management	Pipeline management	Metadata management
Feature store	Model registry	Pipeline orchestration	Metadata store

What is a pipeline?

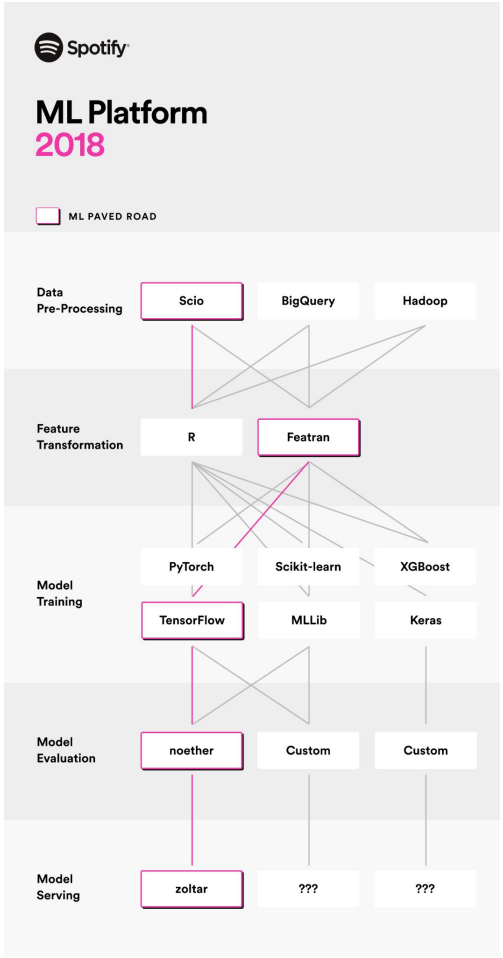
- Description of an ML **workflow**
- A pipeline **component** is a self-contained set of user code that performs one step in the pipeline
- Includes the definition of the **configuration** and **inputs** required to run the pipeline (e.g. model hyperparameters)

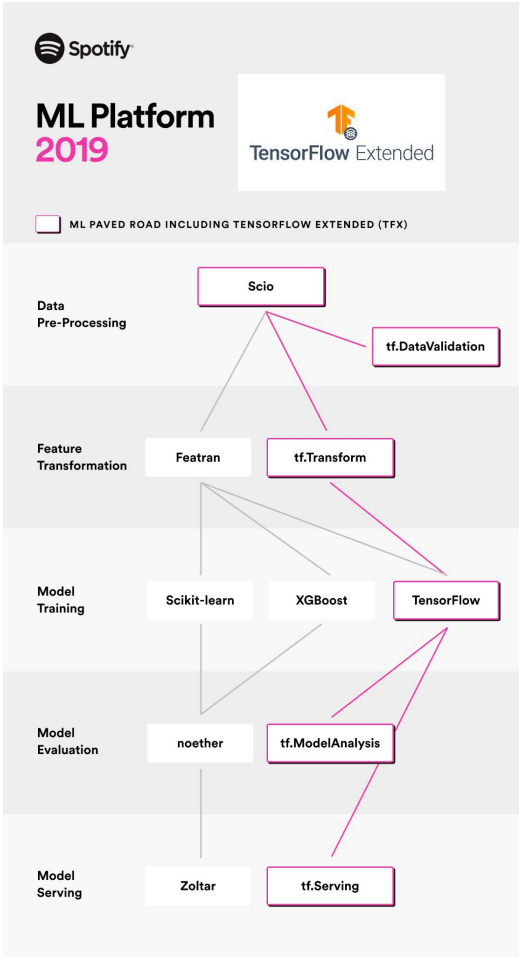


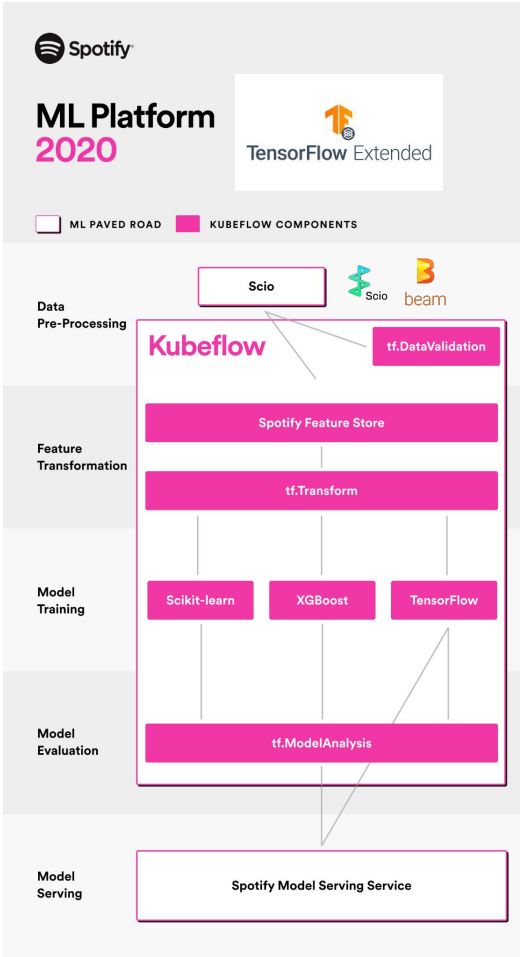
The workflow is also called directed acyclic graph (**DAG**)

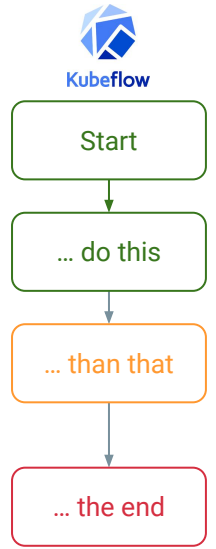


Complete workflow of the ML system lifecycle









Source: Baer & Ngahane (2019)

TensorFlow Extended (TFX)

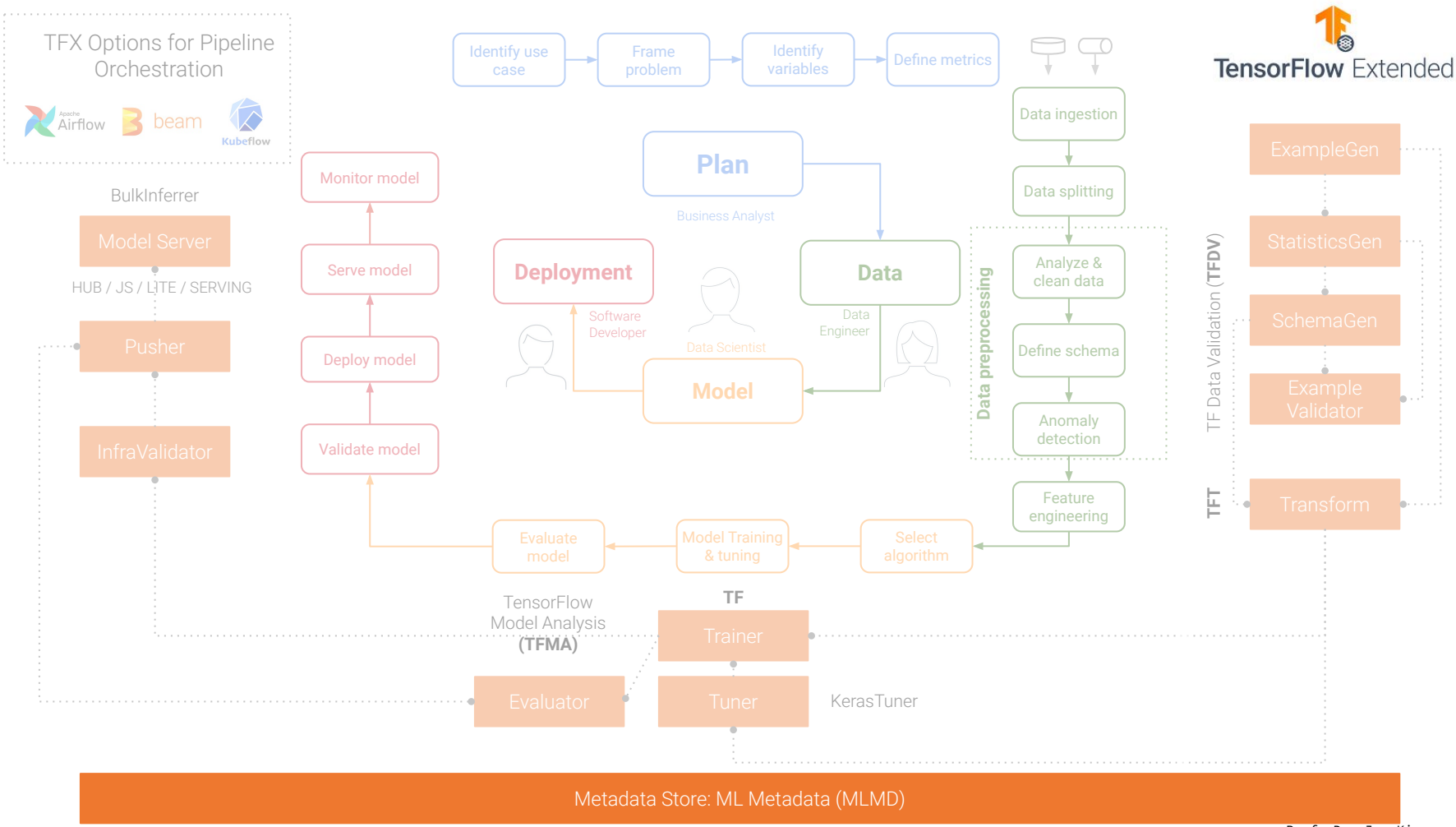
- Google-production-scale machine learning (ML) platform based on TensorFlow
- Portable to multiple environments (Azure, AWS, Google Cloud, IBM, ...)
- Python based toolkit; can be used with notebooks
- Helps you orchestrate your ML process: Apache Airflow, Apache Beam or Kubeflow pipelines

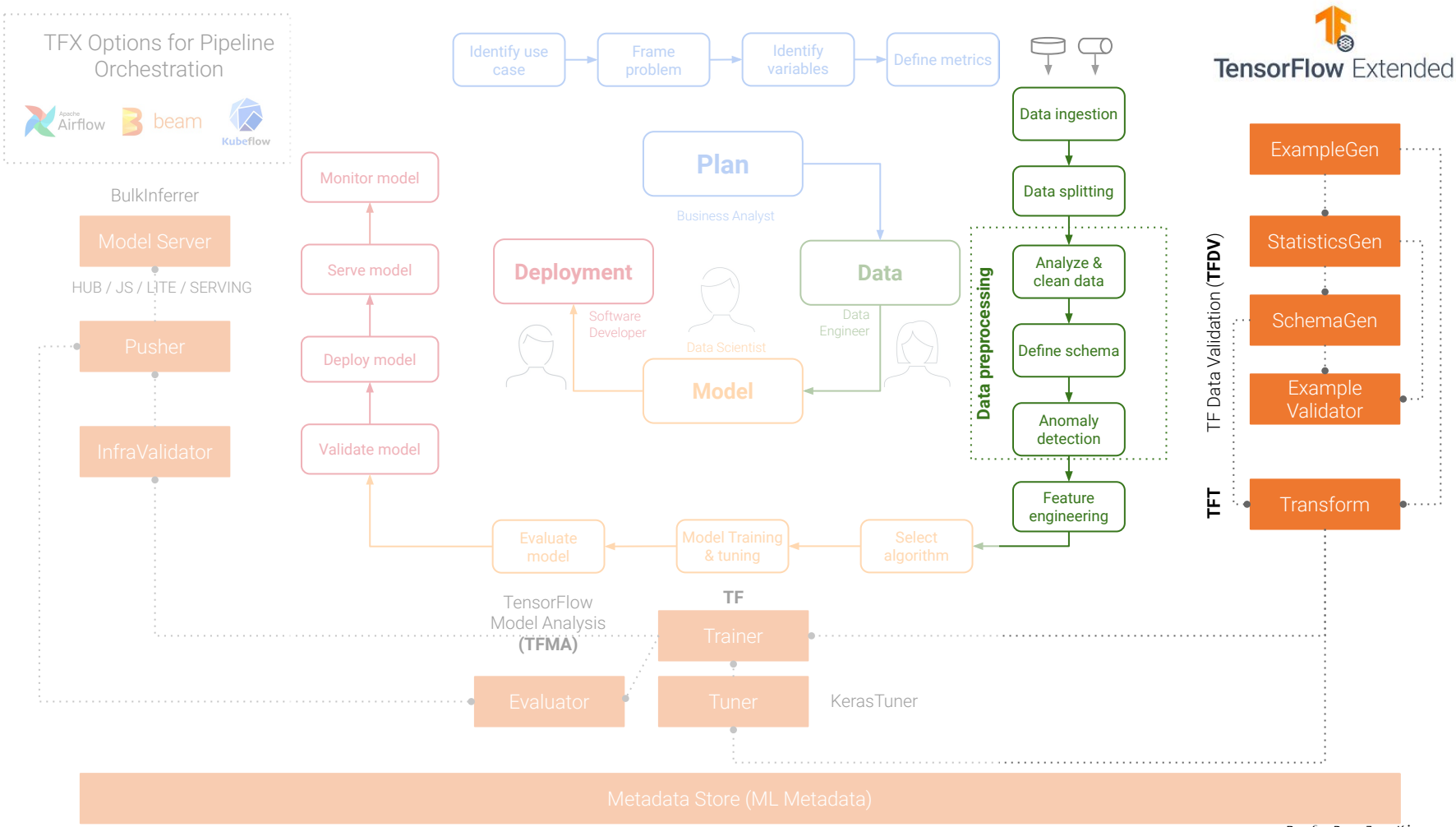
Source: TensorFlow (2021)

TFX 1.0 (19.05.21)

- Enterprise-grade support
- Security patches and select bug fixes for up to three years
- Guaranteed API & Artifact backward compatibility

Source: Google (2021)


TFX Options for Pipeline Orchestration
Apache Airflow
beam
Kubeflow
TensorFlow Extended
Identify use case
Frame problem
Identify variables
Define metrics
Plan
Business Analyst
Data ingestion
Data splitting
Analyze & clean data
Define schema
Anomaly detection
Feature engineering
Data preprocessing
Data
Data Engineer
Deployment
Software Developer
Data Scientist
Model
Monitor model
Serve model
Deploy model
Validate model
BulkInferrer
Model Server
HUB / JS / LITE / SERVING
Pusher
InfraValidator
Evaluate model
Model Training & tuning
Select algorithm
TensorFlow Model Analysis (TFMA)
TF
Trainer
Evaluator
Tuner
KerasTuner
ExampleGen
StatisticsGen
SchemaGen
Example Validator
TF Data Validation (TFDV)
TFT
Transform
Metadata Store: ML Metadata (MLMD)


TFX Options for Pipeline Orchestration
Apache Airflow
beam
Kubeflow
Identify use case
Frame problem
Identify variables
Define metrics
TensorFlow Extended
Data ingestion
Data splitting
Analyze & clean data
Define schema
Anomaly detection
Feature engineering
Data preprocessing
ExampleGen
StatisticsGen
SchemaGen
Example Validator
Transform
TF Data Validation (TFDV)
TFT
Plan
Business Analyst
Data
Data Engineer
Deployment
Software Developer
Data Scientist
Model
Monitor model
Serve model
Deploy model
Validate model
BulkInferrer
Model Server
HUB / JS / LITE / SERVING
Pusher
InfraValidator
Evaluate model
Model Training & tuning
Select algorithm
TensorFlow Model Analysis (TFMA)
TF
Trainer
Tuner
KerasTuner
Evaluator
Metadata Store (ML Metadata)

TFX Options for Pipeline Orchestration

Apache Airflow · beam · Kubeflow

BulkInferrer

Model Server

HUB / JS / LITE / SERVING

Pusher

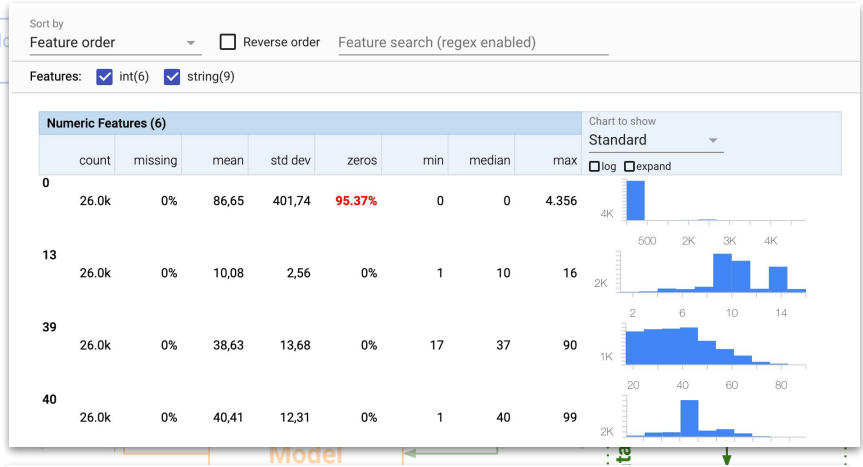
InfraValidator

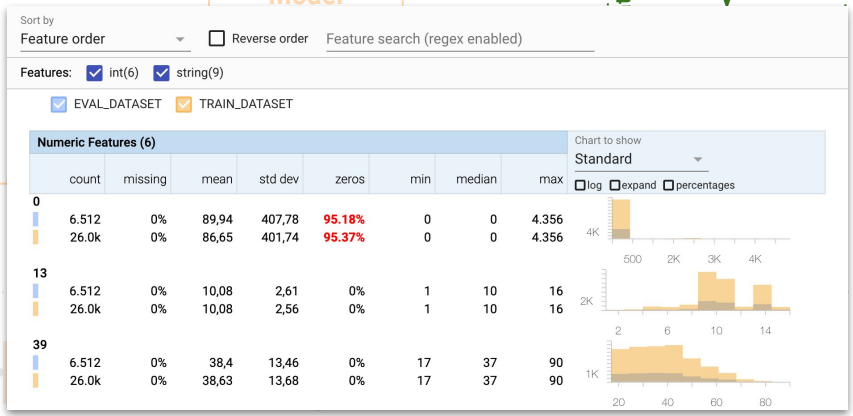
Monitor model

Serve model

Deploy model

Validate model





TensorFlow Extended

ExampleGen

StatisticsGen

SchemaGen

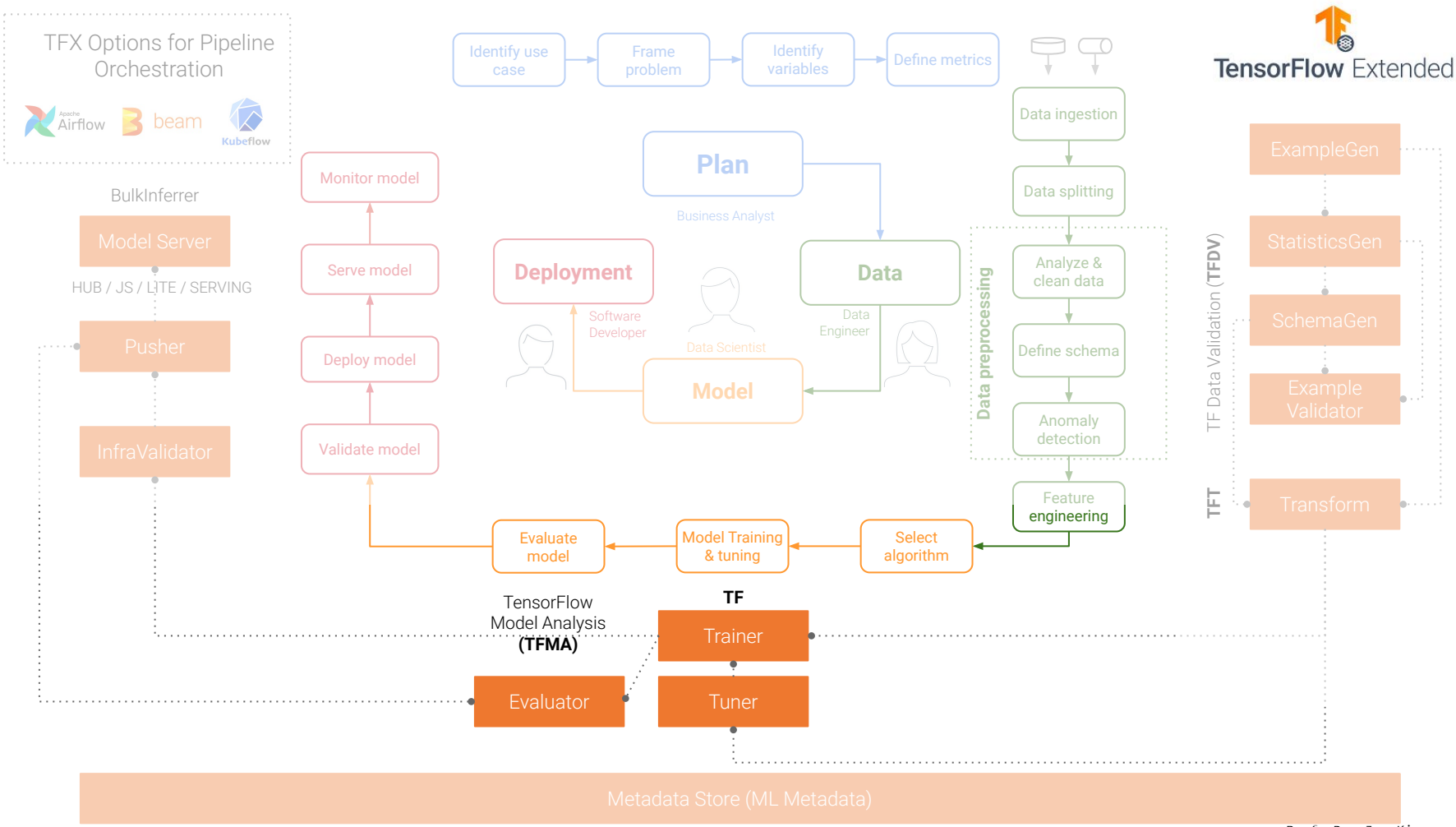
Example Validator

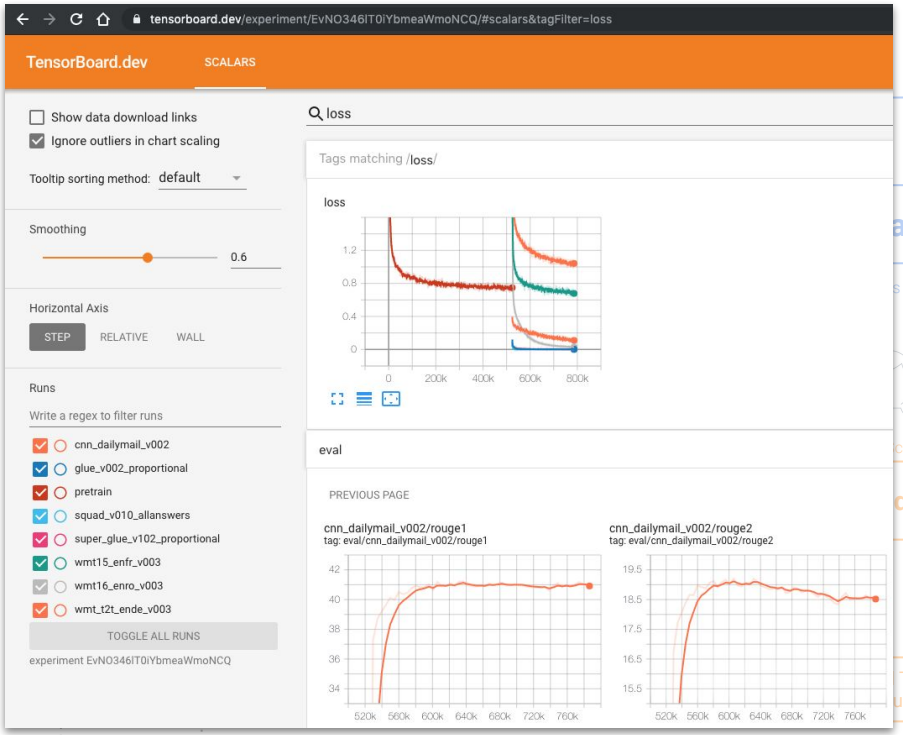
TF Data Validation (TFDV)

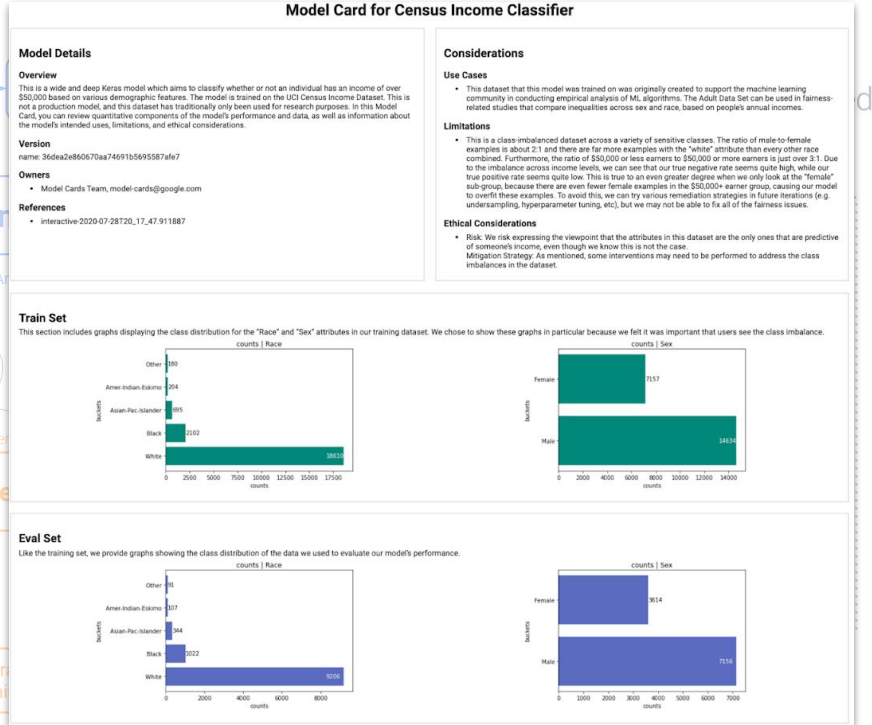
TFT

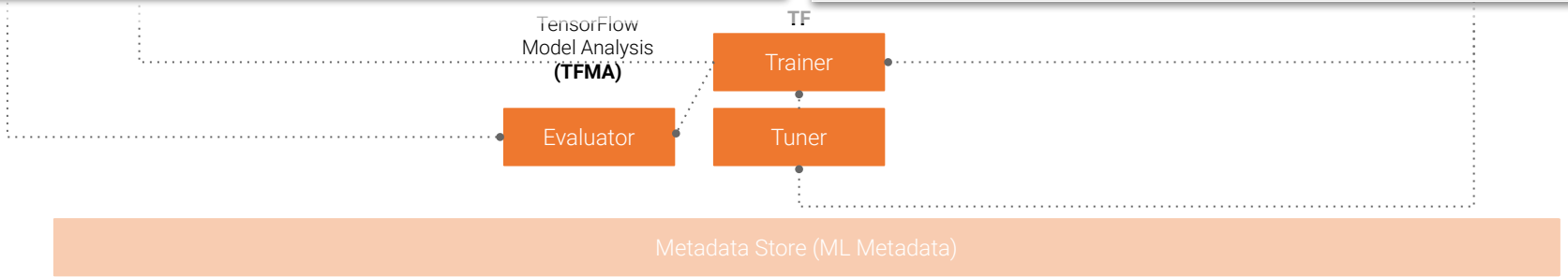
Transform

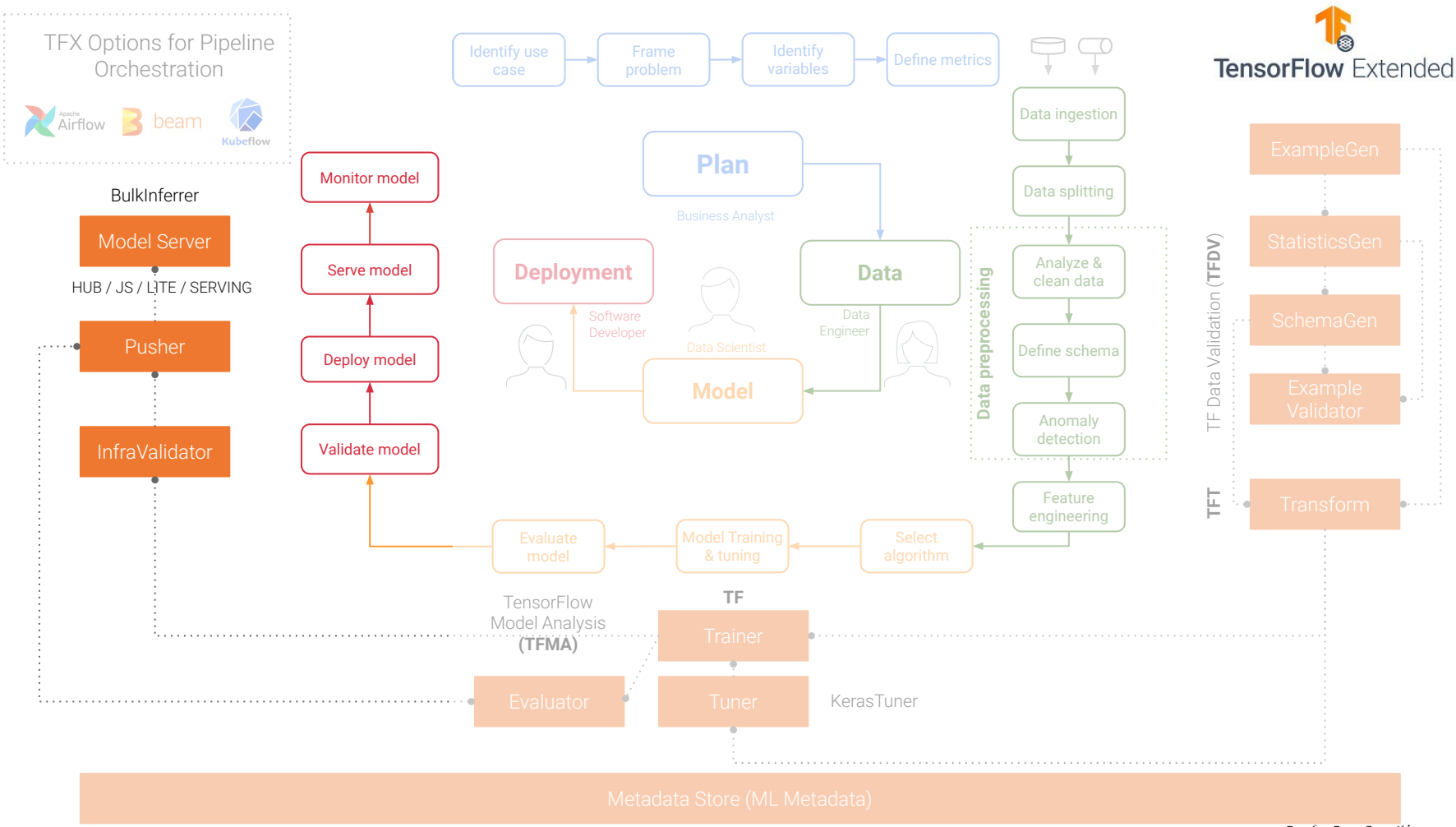
Metadata Store (ML Metadata)


TFX Options for Pipeline Orchestration
Apache Airflow
beam
Kubeflow
TensorFlow Extended
Identify use case
Frame problem
Identify variables
Define metrics
Plan
Business Analyst
Data
Data Engineer
Deployment
Software Developer
Data Scientist
Model
Data ingestion
Data splitting
Analyze & clean data
Define schema
Anomaly detection
Feature engineering
Data preprocessing
ExampleGen
StatisticsGen
SchemaGen
Example Validator
Transform
TF Data Validation (TFDV)
TFT
Select algorithm
Model Training & tuning
Evaluate model
Monitor model
Serve model
Deploy model
Validate model
BulkInferrer
Model Server
HUB / JS / LITE / SERVING
Pusher
InfraValidator
TensorFlow Model Analysis (TFMA)
TF
Trainer
Tuner
Evaluator
Metadata Store (ML Metadata)


tensorboard.dev/experiment/EvNO346IT0iYbmeaWmoNCQ/#scalars&tagFilter=loss
TensorBoard.dev
SCALARS
Show data download links
Ignore outliers in chart scaling
Tooltip sorting method: default
Smoothing
0.6
Horizontal Axis
STEP
RELATIVE
WALL
Runs
Write a regex to filter runs
cnn_dailymail_v002
glue_v002_proportional
pretrain
squad_v010_allanswers
super_glue_v102_proportional
wmt15_enfr_v003
wmt16_enro_v003
wmt_t2t_ende_v003
TOGGLE ALL RUNS
experiment EvNO346IT0iYbmeaWmoNCQ
loss
Tags matching /loss/
loss
eval
PREVIOUS PAGE
cnn_dailymail_v002/rouge1
tag: eval/cnn_dailymail_v002/rouge1
cnn_dailymail_v002/rouge2
tag: eval/cnn_dailymail_v002/rouge2


Model Card for Census Income Classifier
Model Details
Overview
Version
Owners
References
Considerations
Use Cases
Limitations
Ethical Considerations
Train Set
This section includes graphs displaying the class distribution for the "Race" and "Sex" attributes in our training dataset. We chose to show these graphs in particular because we felt it was important that users see the class imbalance.
counts | Race
counts | Sex
Eval Set
Like the training set, we provide graphs showing the class distribution of the data we used to evaluate our model's performance.
counts | Race
counts | Sex


TensorFlow
Model Analysis
(TFMA)
TF
Trainer
Evaluator
Tuner
Metadata Store (ML Metadata)


TFX Options for Pipeline Orchestration
Apache Airflow
beam
Kubeflow
TensorFlow Extended
Identify use case
Frame problem
Identify variables
Define metrics
Data ingestion
Data splitting
Analyze & clean data
Define schema
Anomaly detection
Feature engineering
Data preprocessing
ExampleGen
StatisticsGen
SchemaGen
Example Validator
Transform
TF Data Validation (TFDV)
TFT
Plan
Business Analyst
Data
Data Engineer
Deployment
Software Developer
Data Scientist
Model
Select algorithm
Model Training & tuning
Evaluate model
Monitor model
Serve model
Deploy model
Validate model
BulkInferrer
Model Server
HUB / JS / LITE / SERVING
Pusher
InfraValidator
TensorFlow Model Analysis (TFMA)
TF
Trainer
Tuner
KerasTuner
Evaluator
Metadata Store (ML Metadata)

TFX Options for Pipeline Orchestration

Apache Airflow beam Kubeflow

TensorFlow Extended

TensorFlow Lite is a set of tools that enables on-device machine learning by helping developers run their models on mobile, embedded, and IoT devices.

BulkInferrer

Model Server

HUB / JS / **LITE** / SERVING

Pusher

InfraValidator

ExampleGen

StatisticsGen

SchemaGen

Example Validator

TF Data Validation (**TFDV**)

TFT

Transform

TensorFlow Model Analysis **(TFMA)**

TF

Trainer

Evaluator

Tuner

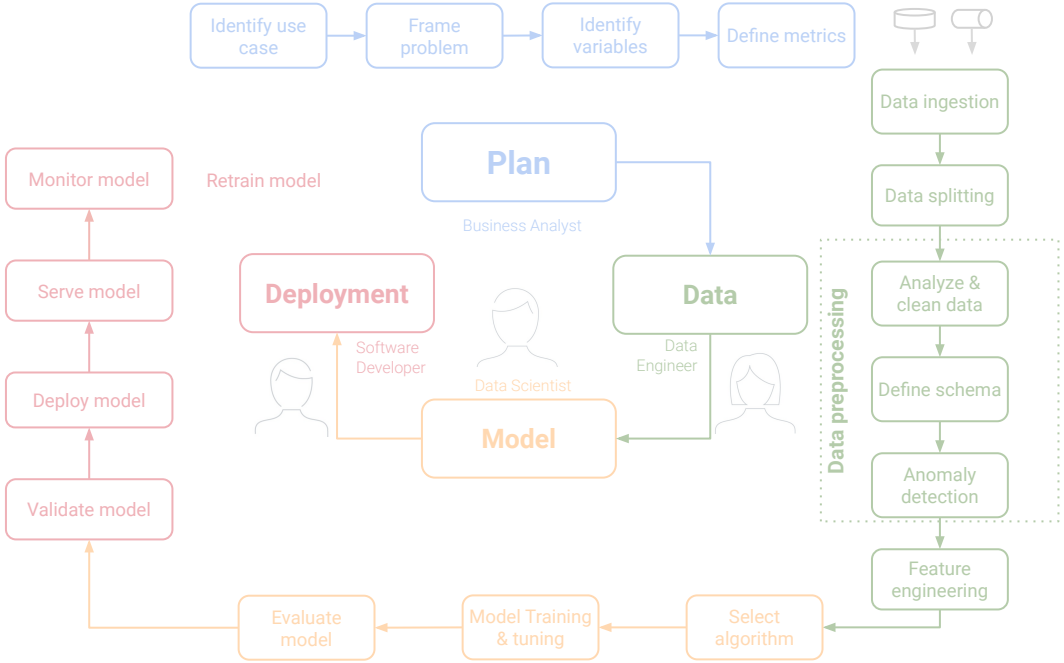
KerasTuner

Metadata Store (ML Metadata)

TFX Options for Pipeline Orchestration

Apache Airflow beam Kubeflow

TensorFlow Extended

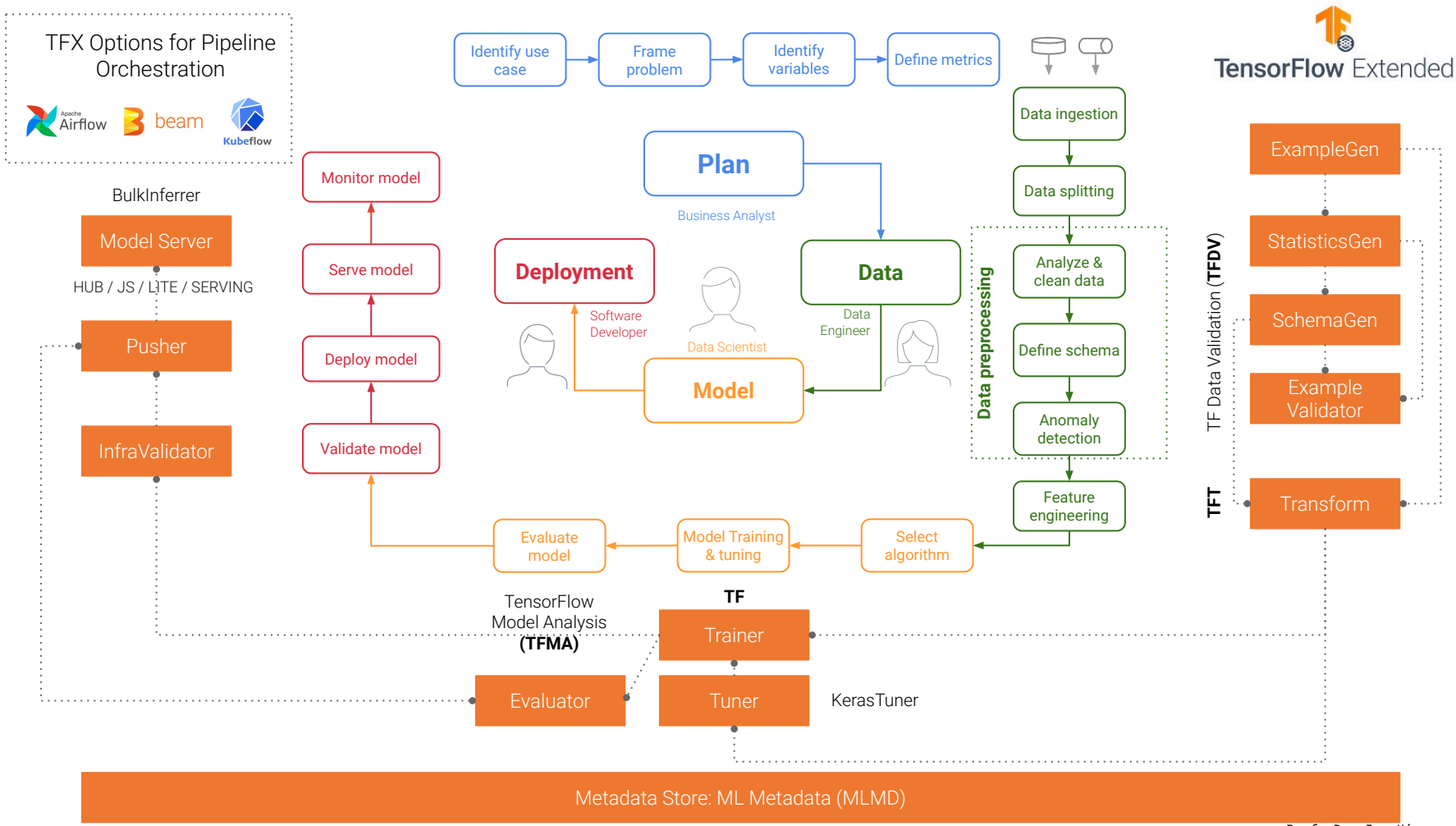


Production phase: automate the execution of the ML pipeline based on a schedule or certain triggering conditions.

Data preparation phase: automatically ingest, validate and transform data and provide features to models

Development phase: run the ML experiment, instead of manually executing each step.

Metadata Store (ML Metadata)


TFX Options for Pipeline Orchestration
Apache Airflow
beam
Kubeflow
Identify use case
Frame problem
Identify variables
Define metrics
TensorFlow Extended
Data ingestion
Data splitting
Analyze & clean data
Define schema
Anomaly detection
Feature engineering
Data preprocessing
ExampleGen
StatisticsGen
SchemaGen
Example Validator
Transform
TF Data Validation (TFDV)
TFT
Plan
Business Analyst
Data
Data Engineer
Deployment
Software Developer
Data Scientist
Model
Select algorithm
Model Training & tuning
Evaluate model
Monitor model
Serve model
Deploy model
Validate model
BulkInferrer
Model Server
HUB / JS / LITE / SERVING
Pusher
InfraValidator
TensorFlow Model Analysis (TFMA)
TF
Trainer
Tuner
KerasTuner
Evaluator
Metadata Store: ML Metadata (MLMD)