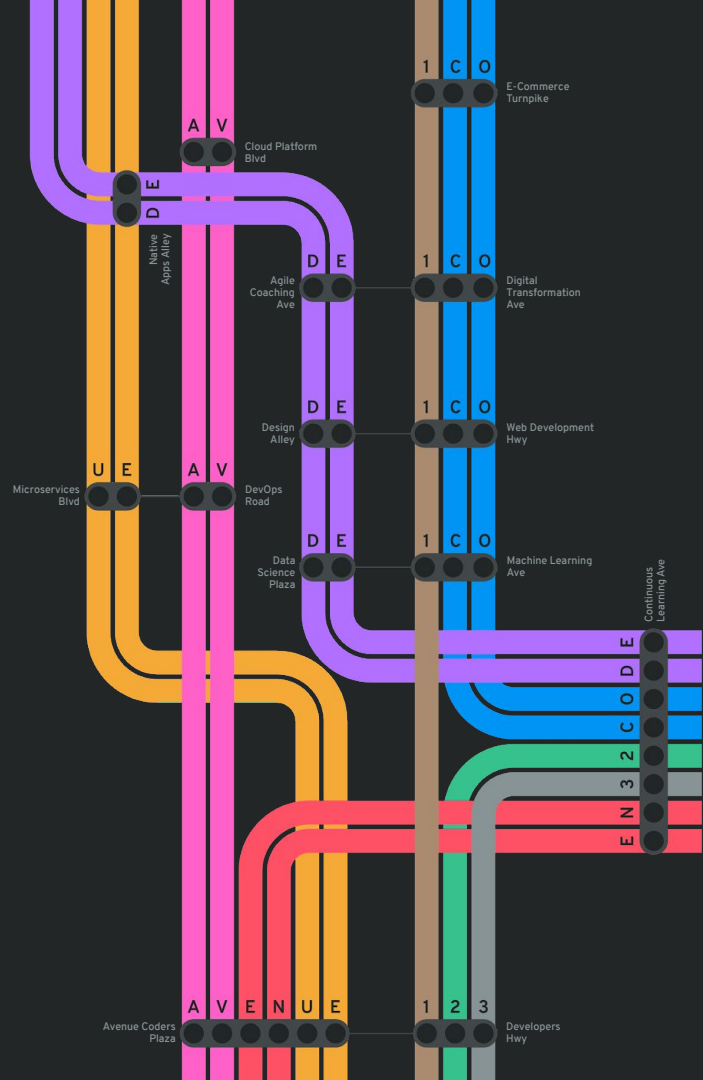
# AvenueCode


1 C O
E-Commerce Turnpike
A V
Cloud Platform Blvd
D E
Native Apps Alley
D E
Agile Coaching Ave
1 C O
Digital Transformation Ave
D E
Design Alley
1 C O
Web Development Hwy
U E
Microservices Blvd
A V
DevOps Road
D E
Data Science Plaza
1 C O
Machine Learning Ave
Continuous Learning Ave
E N 3 2 C O D E
A V E N U E
Avenue Coders Plaza
1 2 3
Developers Hwy

# TensorFlow on Google Cloud

May / 2020

AvenueCode

AvenueCode

1. Context
2. Why TensorFlow on GCP?
3. Getting Started with Tensorflow
4. Tensorflow Extended
5. Hands On Qwiklabs

AvenueCode

# 1

# Context

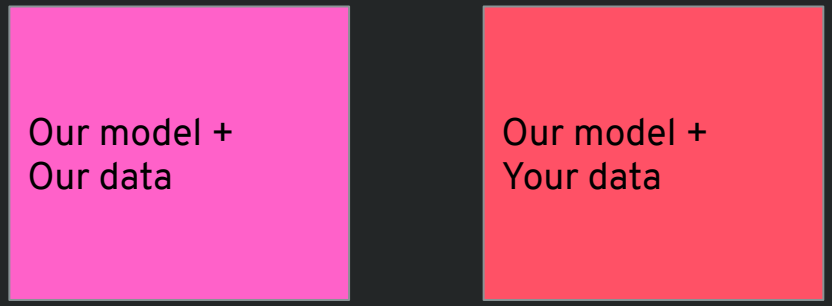
# ML on Google Cloud

AvenueCode


API's

# ML on Google Cloud

AvenueCode

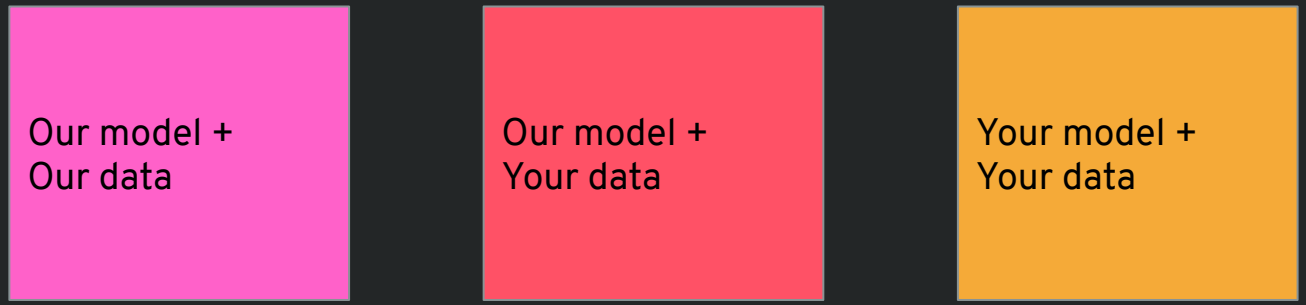



API's

AutoML

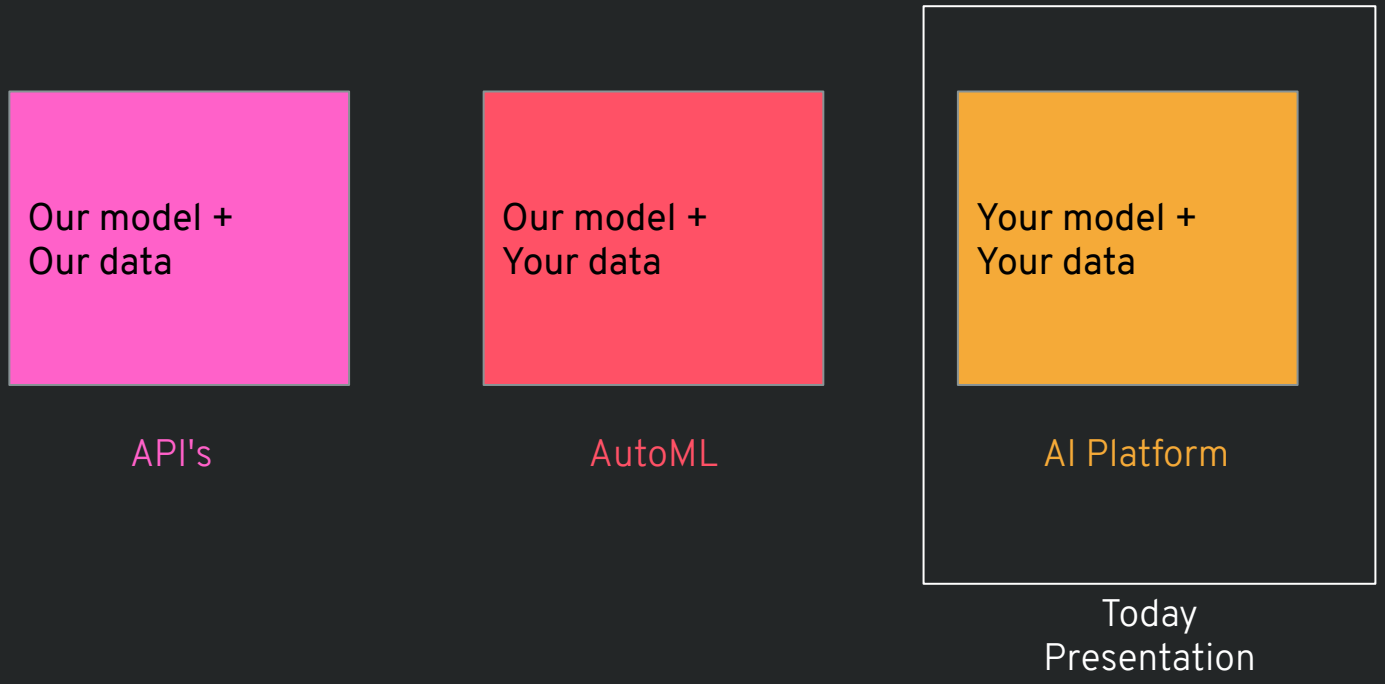
# ML on Google Cloud

AvenueCode



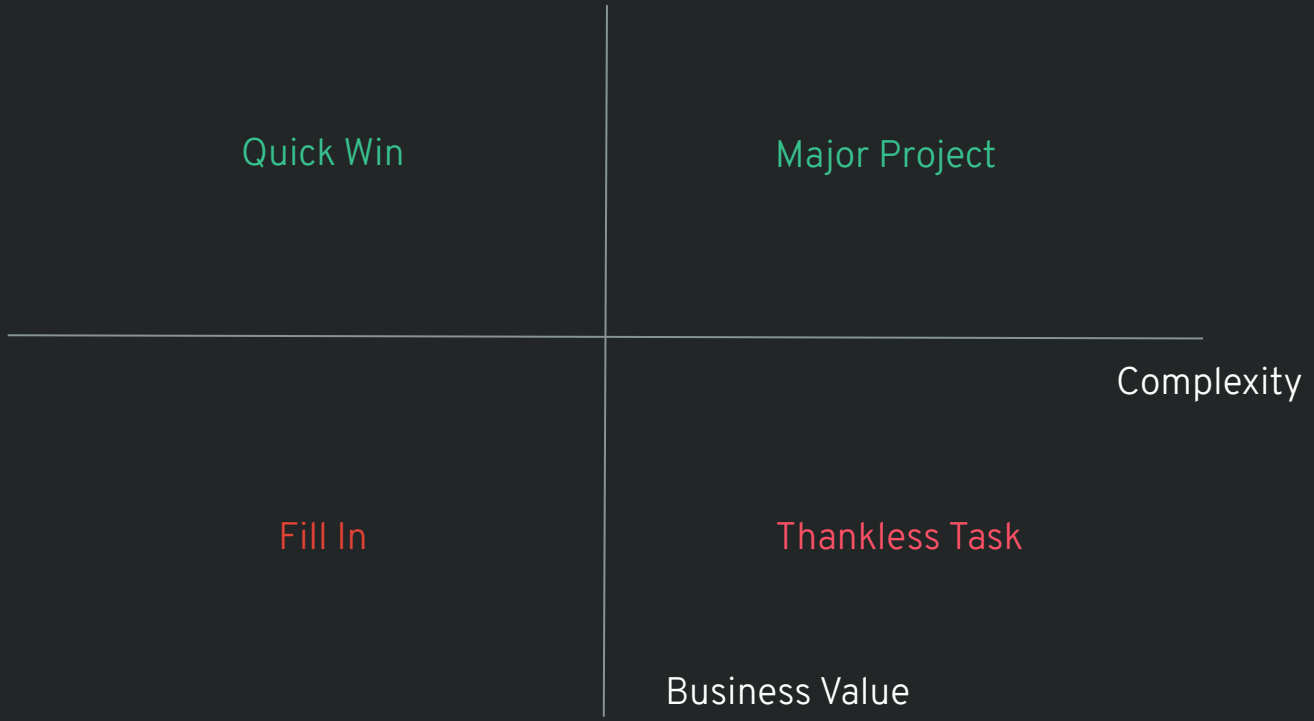
# ML on Google Cloud

AvenueCode



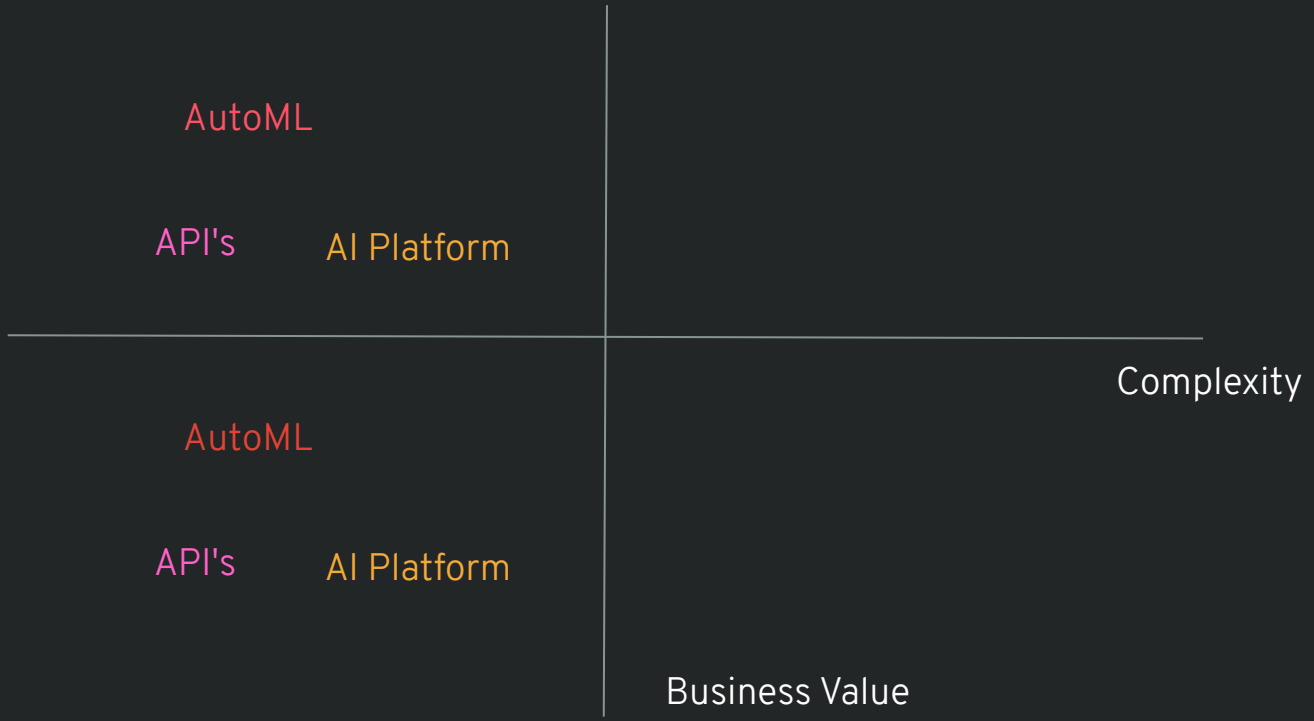
# AvenueCode Process

AvenueCode



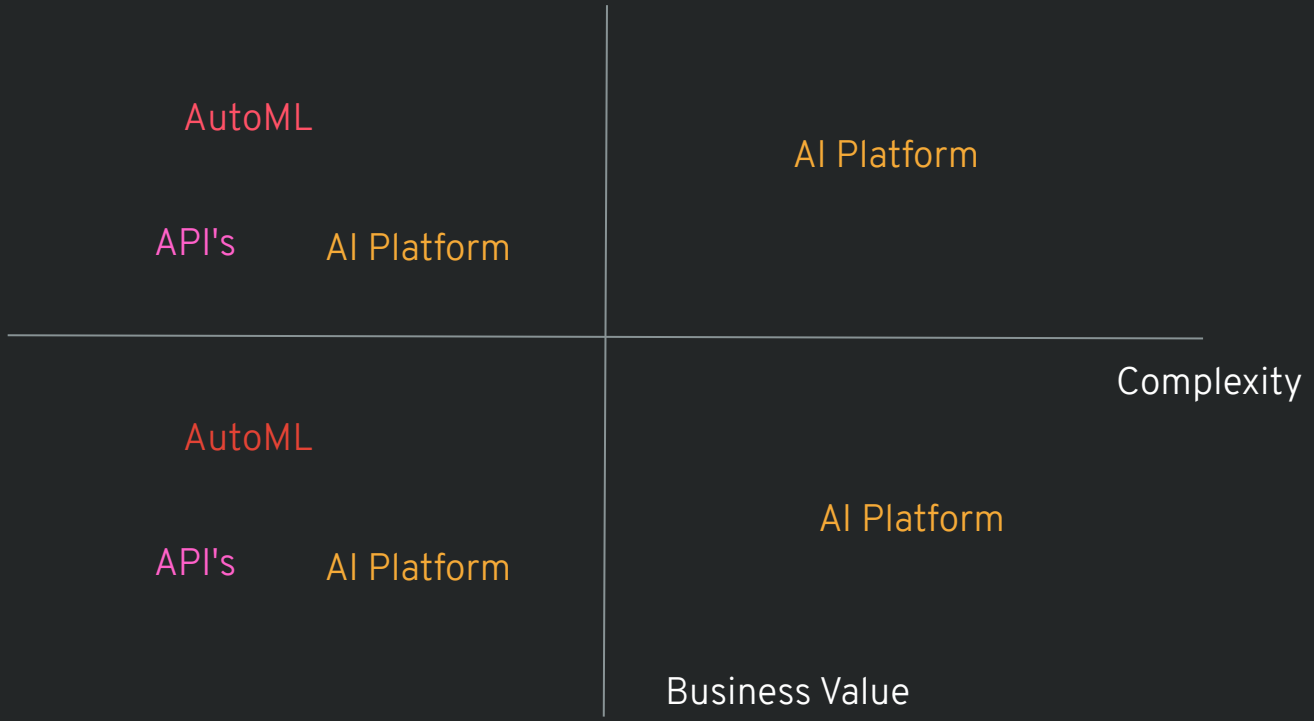
# AvenueCode Process

AvenueCode



# AvenueCode Process

AvenueCode



# Problem Example

AvenueCode

# Problem Example

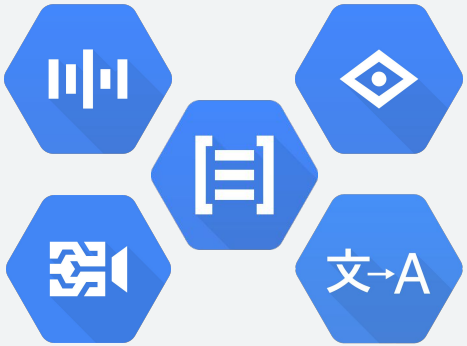
AvenueCode

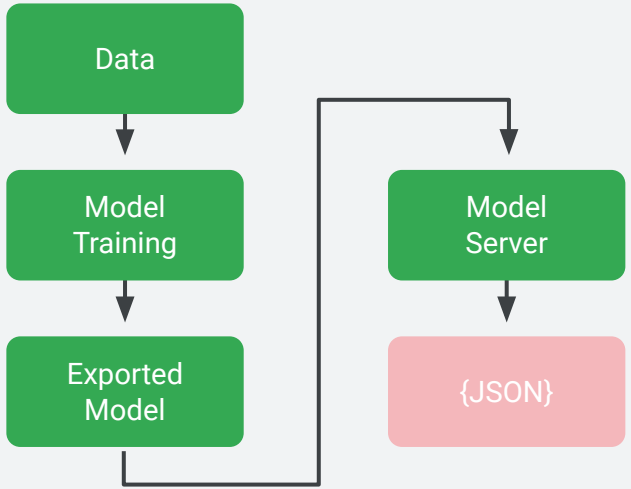
## Is this a plastic bag?

Our model +
Our data

API's

# Machine Learning APIs




# Problem Example

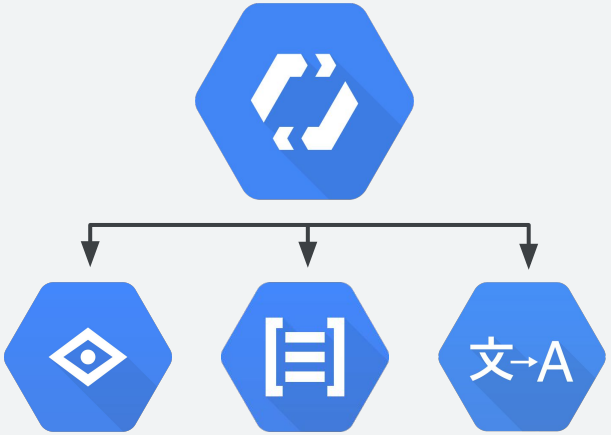
AvenueCode

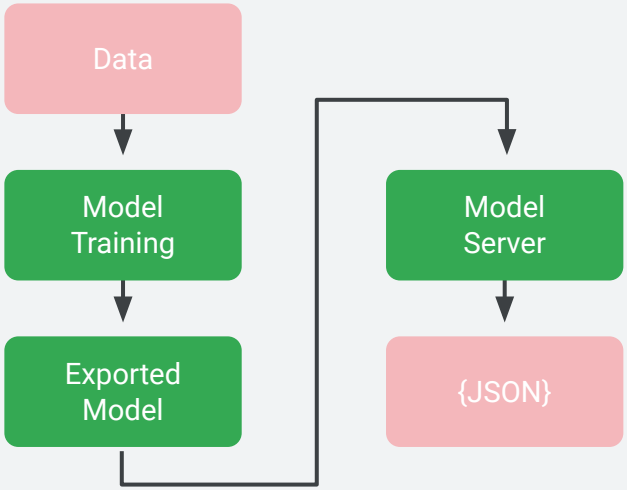
## What's the plastic type?

Our model +
Your data

AutoML

# AutoML


文→A


Data
Model Training
Exported Model
Model Server
{JSON}


API

# Problem Example

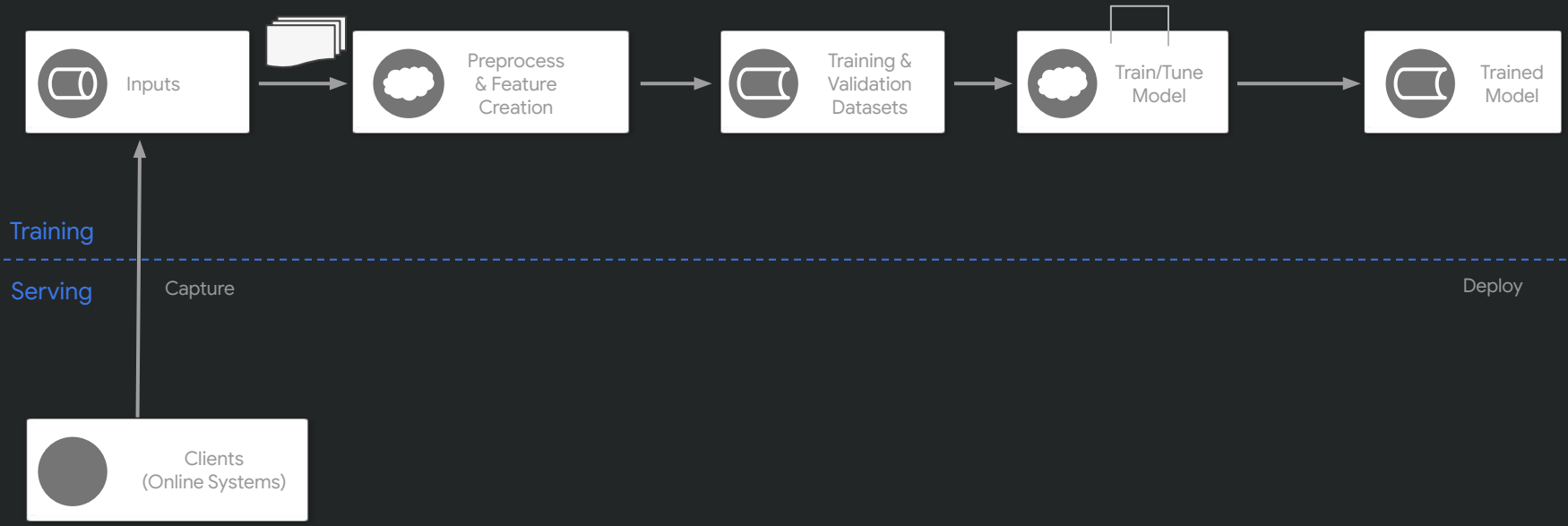
AvenueCode

**Will this plastic bag be recycled? From today how much time will take ?**

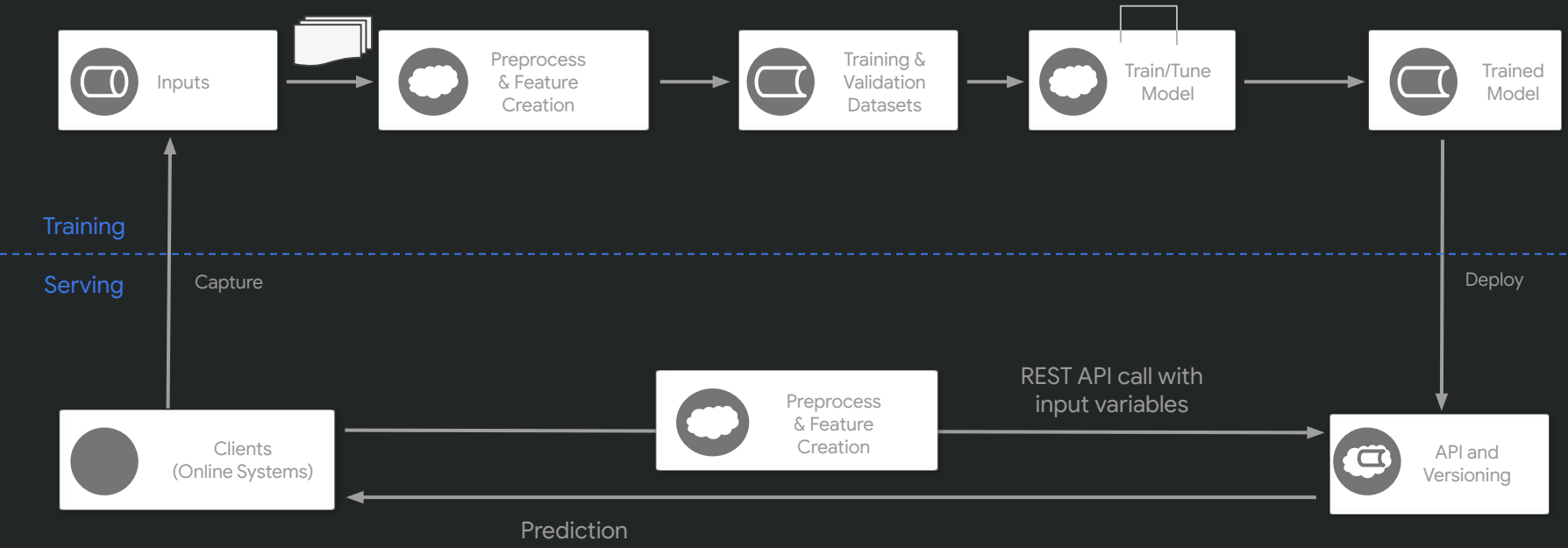
Your model +
Your data

AI Platform

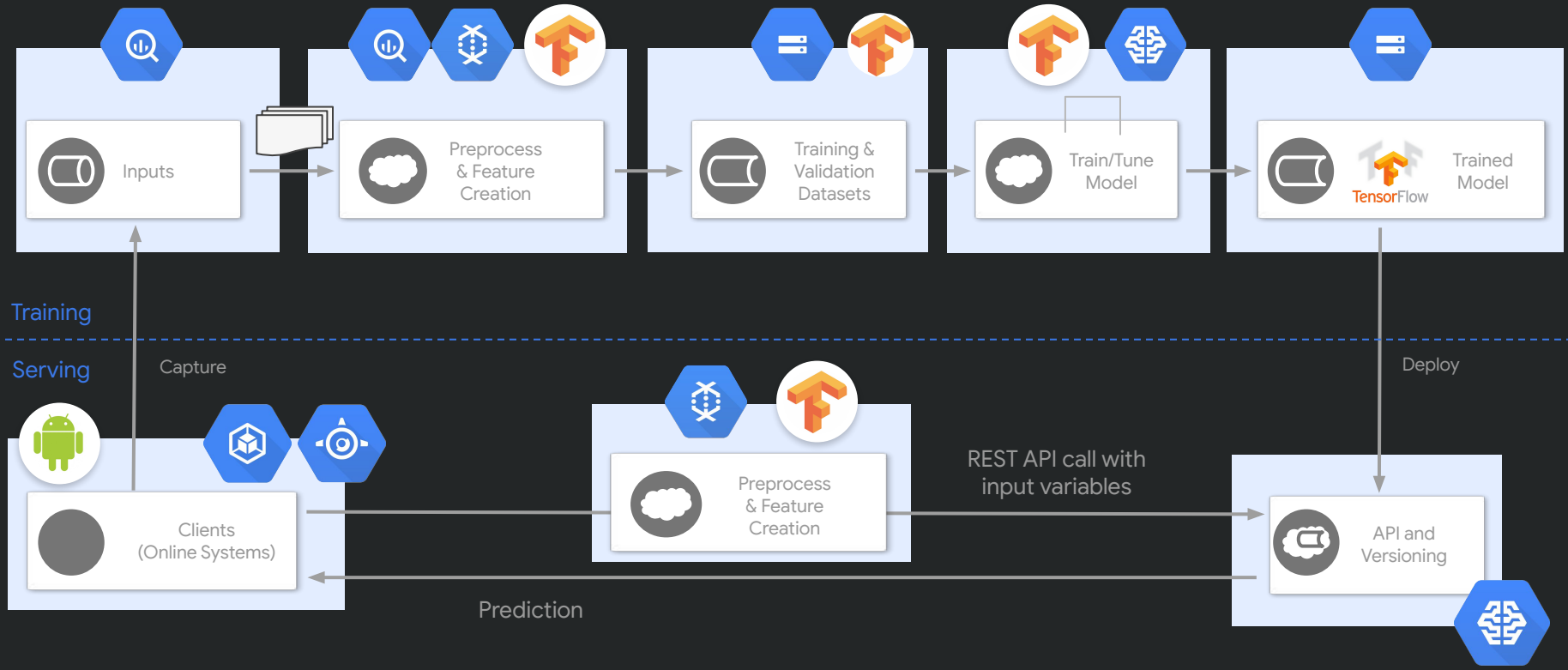
# Machine learning pipeline @ GCP



# Machine learning pipeline @ GCP



# Machine learning pipeline @ GCP



AvenueCode

2

# Why TensorFlow on GCP?

# “More than 87% of data science projects never make it into production”

- Multiple studies and surveys -

Source:
https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/

# Hidden Technical Debt in Machine Learning Systems

**D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips**
{dsculley,gholt,dgg,edavydov,toddphillips}@google.com
Google, Inc.

**Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, Dan Dennison**
{ebner,vchaudhary,mwyoung,jfcrespo,dennison}@google.com
Google, Inc.

# Why ML deployment is hard?

Deploying ML models at scale is very challenging

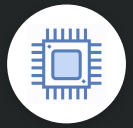
**In IT System**

Behavior of system is defined by Code

**In AI/ML System**

Behavior of system is defined by Data

**Organization** - Managing different ecosystems like programming languages

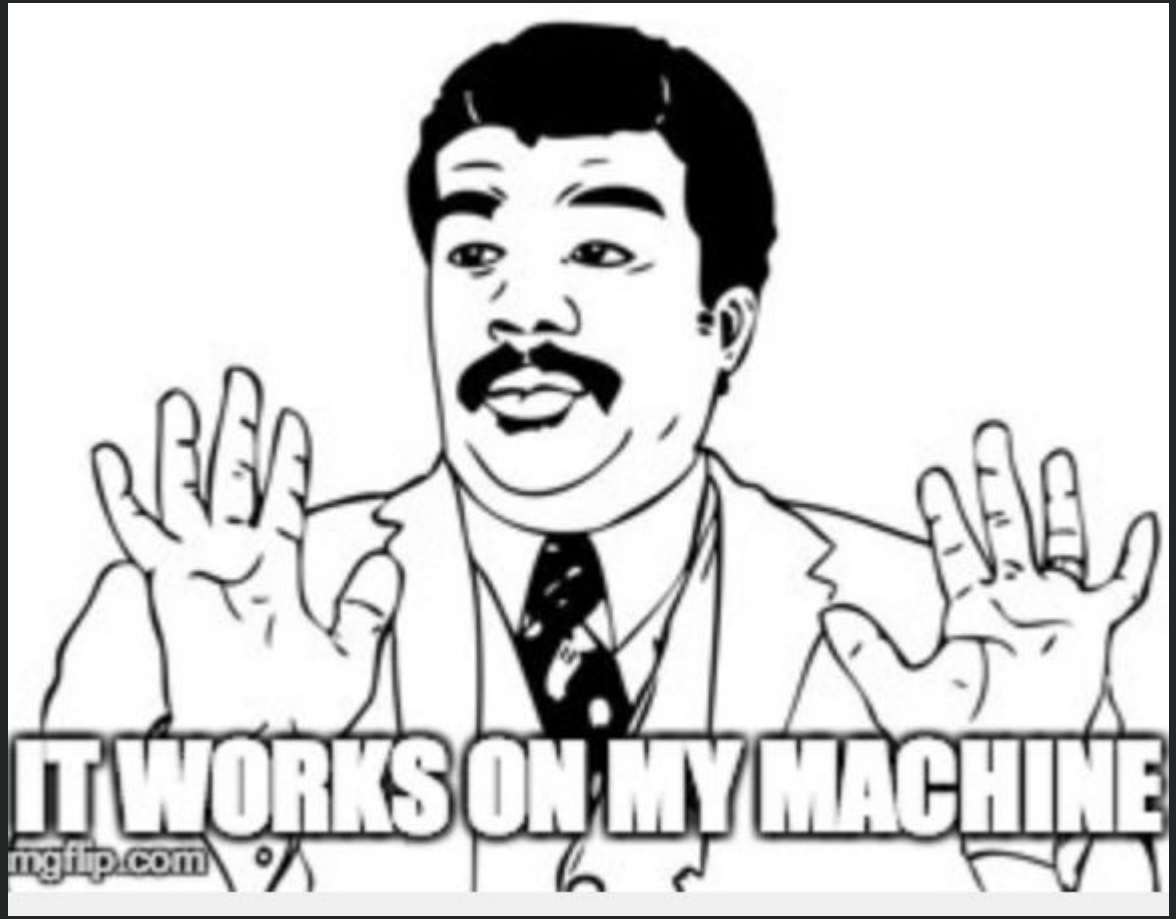


**Compute** - Continuous and reliable compute resources (dedicated servers or cloud only option)

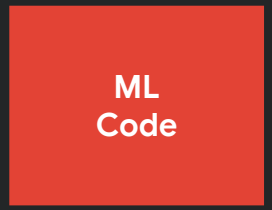
**Portability** - Huge dependencies on legacy systems

**Seasonality** - ML workloads works in patches; this need auto scaling capabilities

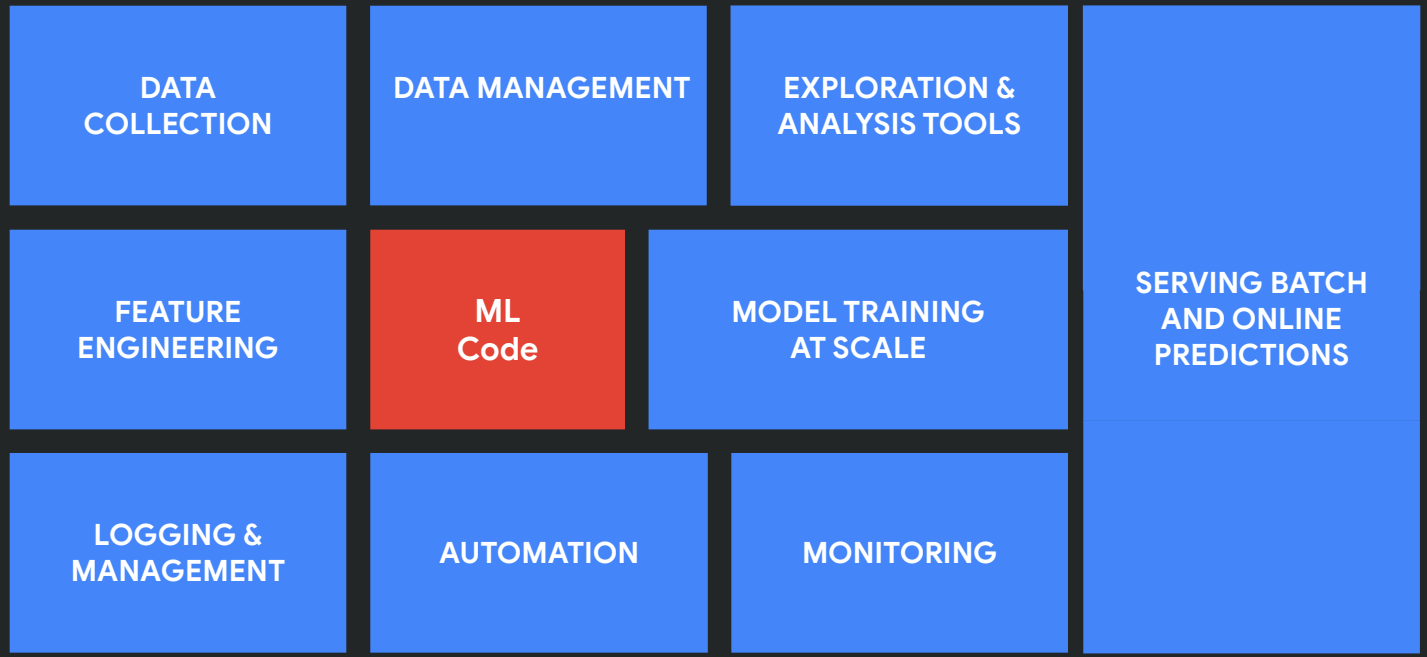
AvenueCode



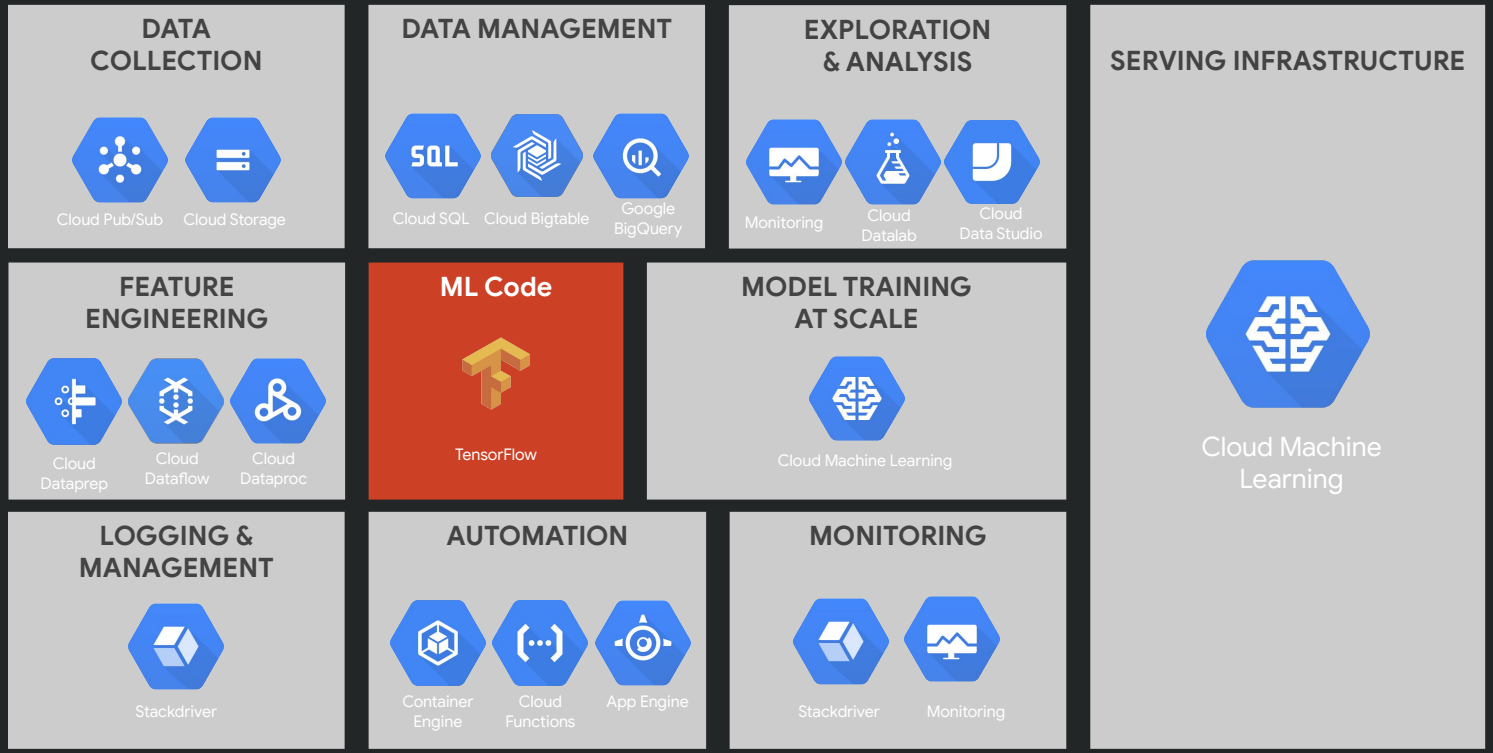
# Operational ML : What you first think



# Operational ML : What really is



# Operational ML - end-to-end ML solution on GCP



AvenueCode

# Excuse me sir, what about TensorFlow?

# TensorFlow deals with all kinds of data

Audio

Image

Text

Tabular

# TensorFlow deals with main machine learning tasks

Audio

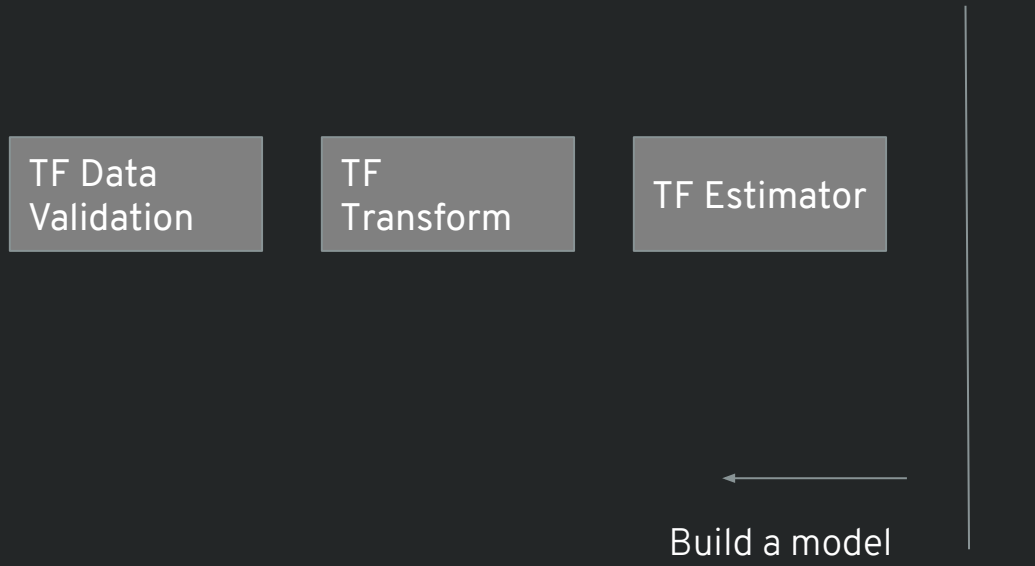
Image

Text

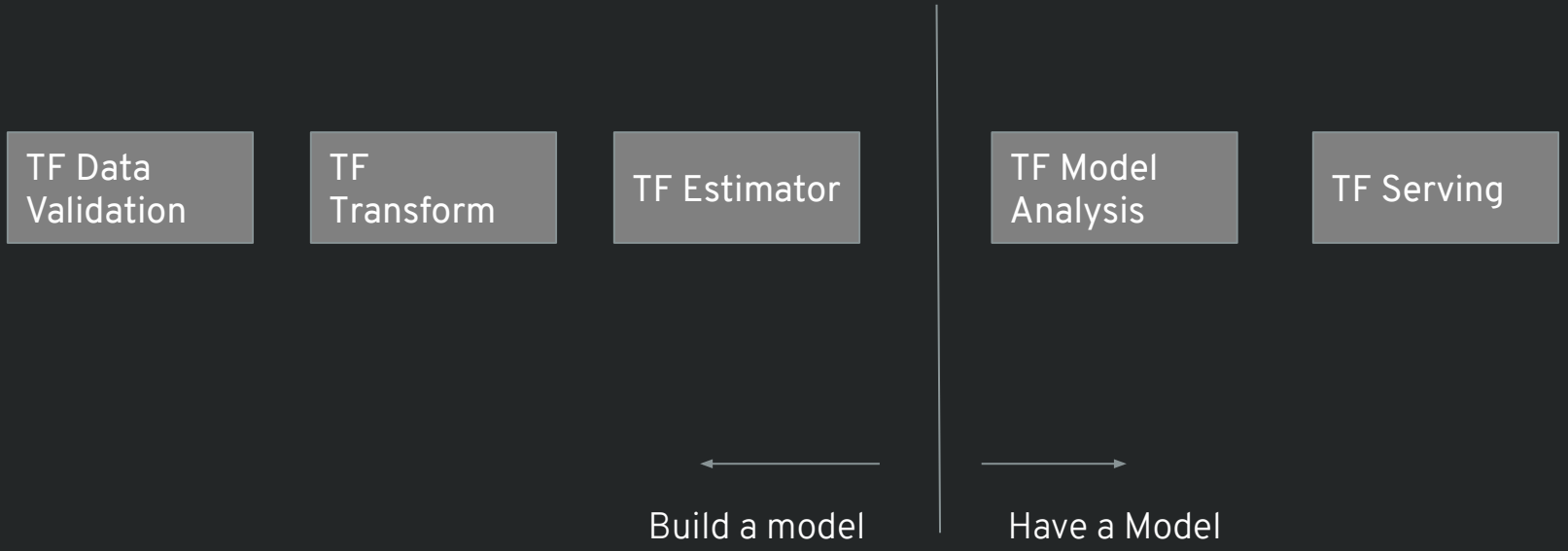
Tabular

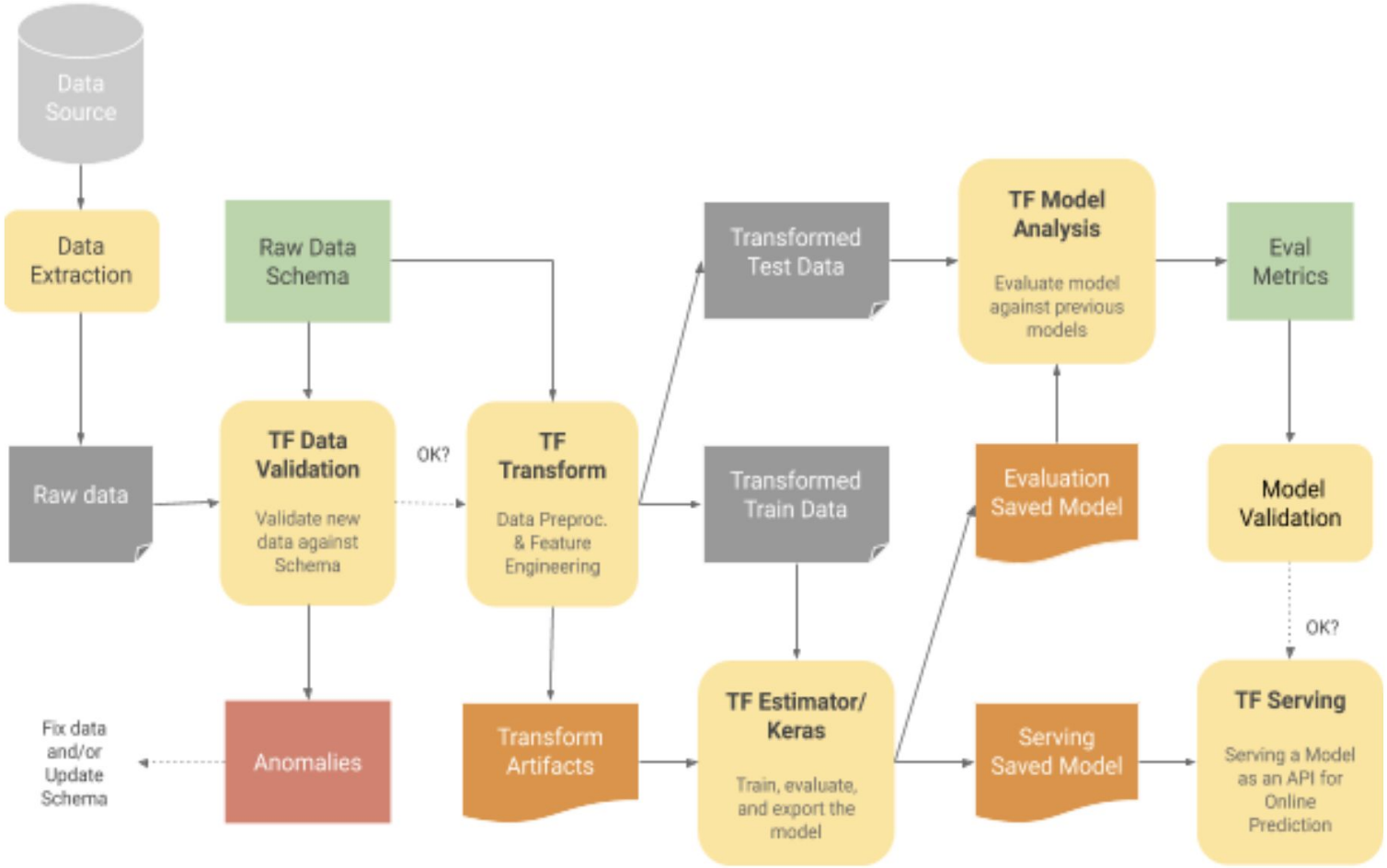
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Transfer Learning

# TensorFlow provides a end-to-end solution



# TensorFlow provides a end-to-end solution






Typical TFX-based ML system

AvenueCode

3

# Getting Started with TensorFlow

AvenueCode

# What are tensors?

- Pattern of data in TensorFlow
- N-dimensional array.
- Beyond the numeric representation, tensors carry informations as data types and array's shape.

# A tensor is an N-dimensional array of data

Rank 0 Tensor
scalar

Rank 1 Tensor
vector

Rank 2 Tensor
matrix

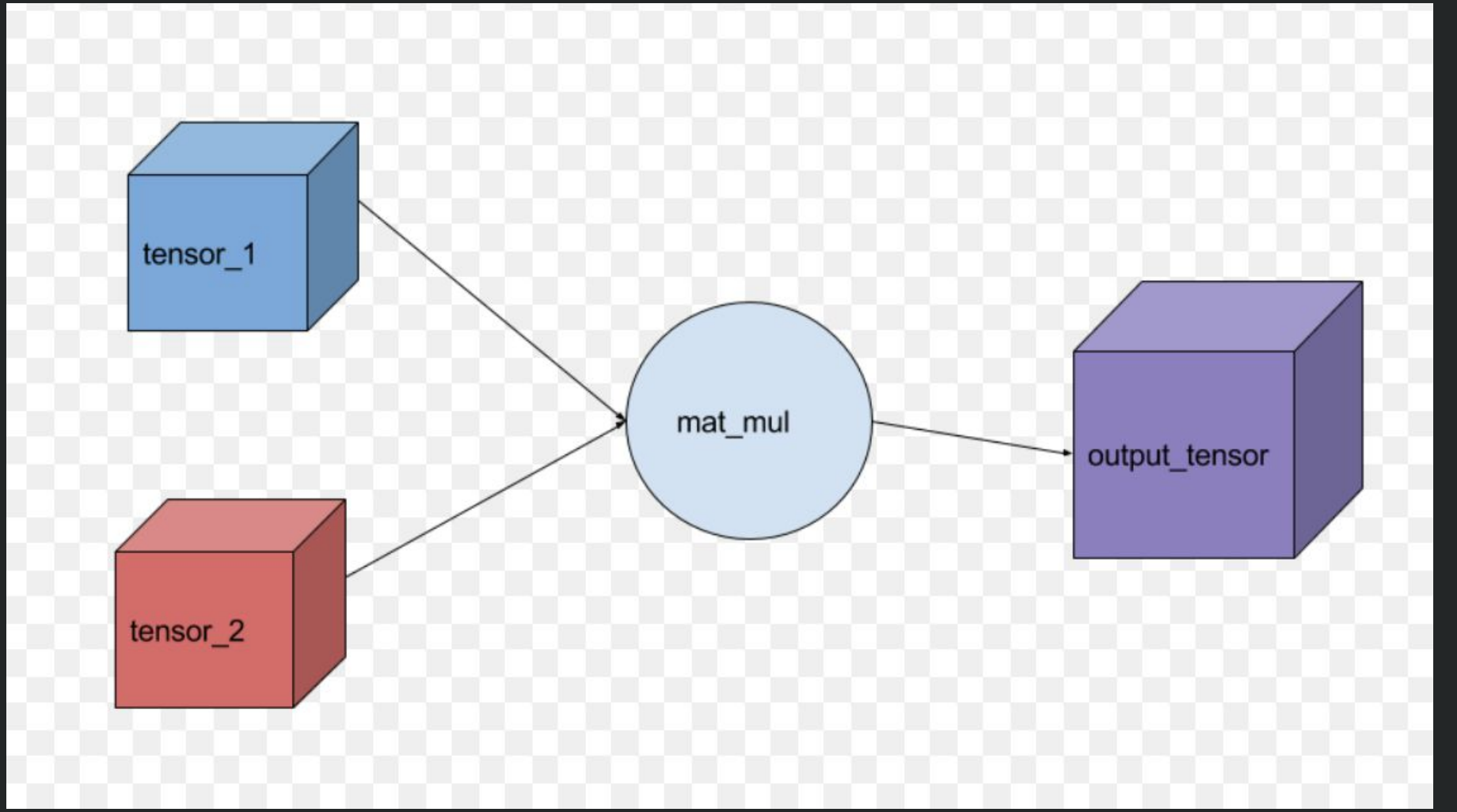
Rank 3 Tensor

Rank 4 Tensor

AvenueCode

# Graphs : Brief Introduction

- Math structure to compute relationship between objects. .
- **Tensorflow generate  graphs every time a operation is computed.**


tensor_1
tensor_2
mat_mul
output_tensor

AvenueCode

# Code

(Link)

AvenueCode

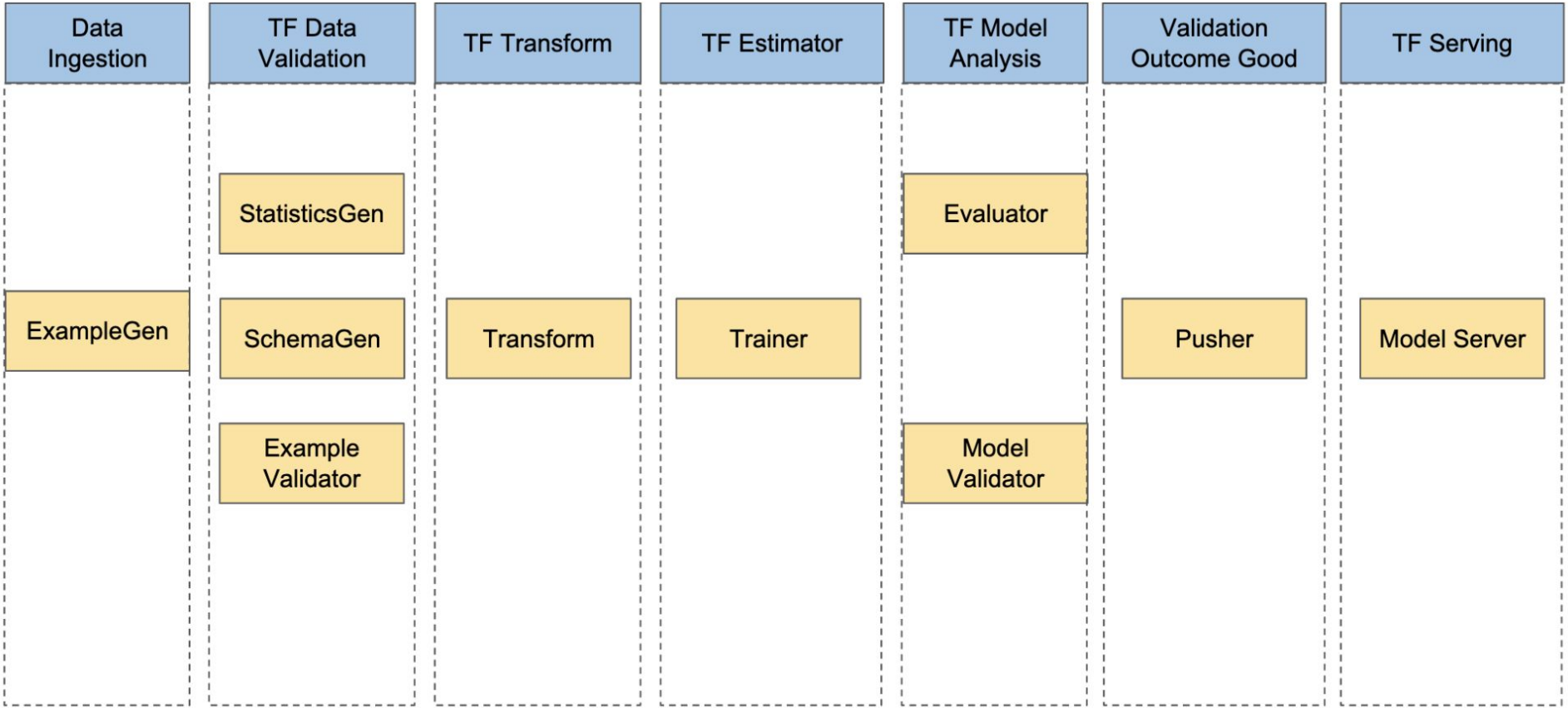
# 4

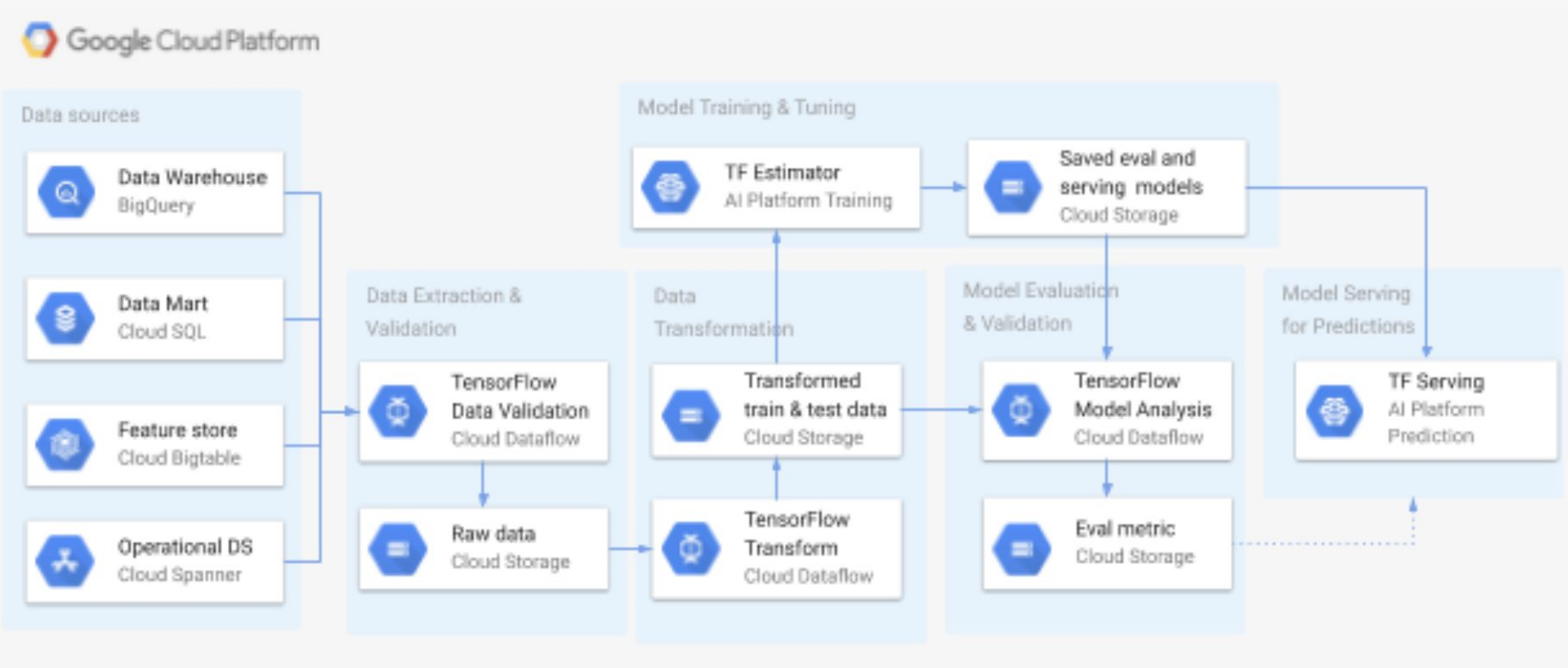
# TensorFlow Extended

AvenueCode

# What is TensorFlow Extended?

- General purpose open source  machine learning platform implemented by google.
- A set of gluable machine learning components in one platform simplifying machine learning deployment.
- Aims to reduce time to put machine learning in production and cover technical debts in those practices. .


Data Ingestion
TF Data Validation
TF Transform
TF Estimator
TF Model Analysis
Validation Outcome Good
TF Serving
ExampleGen
StatisticsGen
SchemaGen
Example Validator
Transform
Trainer
Evaluator
Model Validator
Pusher
Model Server


Google Cloud Platform
Data sources
Data Warehouse
BigQuery
Data Mart
Cloud SQL
Feature store
Cloud Bigtable
Operational DS
Cloud Spanner
Data Extraction & Validation
TensorFlow Data Validation
Cloud Dataflow
Raw data
Cloud Storage
Data Transformation
Transformed train & test data
Cloud Storage
TensorFlow Transform
Cloud Dataflow
Model Training & Tuning
TF Estimator
AI Platform Training
Saved eval and serving models
Cloud Storage
Model Evaluation & Validation
TensorFlow Model Analysis
Cloud Dataflow
Eval metric
Cloud Storage
Model Serving for Predictions
TF Serving
AI Platform Prediction

AvenueCode

# tf.data_validation

AvenueCode

# Real World Challenges

## Challenges

- Schema Validation
- Data contains anomalies
- Label invalid data
- Change feature skew

## How TFDV can help?

- tfdv.schema_validation()
- tfdv.display_anomalies()
- tfdv.display_anomalies()
- tfdv.visualize_statistics()

| Better Data, Better Models | Automated schema generation | Integration with Facets (live demo) | Anomaly detection |
|---|---|---|---|
| Compute summary statistics for train/test data | Input feature value ranges | Identify train/test/validation set skew | Detect missing values |
| Drift detection | Type imputation | Unexpected feature values | Out of range values |
| Training-Serving skew detection | Environment specific schema | Feature by feature analysis | Wrong feature types |
| | Schema validation | Compare statistics across two or more data sets | Correct non-conforming data |
| | | Supports visualization of large datasets responsively | |

AvenueCode

# Code

(Link)

AvenueCode

# tf.transform

AvenueCode

# About Transform

- Framework for running batch and streaming data processing jobs.
- Uses Apache Beam.
- Easily integrated with Kubeflow. (example)

| Feature Engineering @ Scale | Transformations |
| --- | --- |
| Transform data before it goes into a model. Output is exported as a TensorFlow graph, used for both training & serving. | tft.scale_by_min_max(), tft.scale_to_0_1(), tft.scale_to_z_score() |
| Create feature embeddings & enriching text features | tft.tfidf(), tft.ngrams(), tft.hash_strings() |
| Builds transformations into TF graph for your model, so same transformations are applied for train & serving. | Convert strings to integers by generating a vocabulary over all input values |
| Vocabulary generation | tft.compute_and_apply_vocabulary(), tft.string_to_int() |
| Normalize values & Bucketization | tft.bucketize() |

AvenueCode

AvenueCode

# Code

(Link)

AvenueCode

# tf.estimators

AvenueCode

# About Estimators

- Easily create Estimators using Keras API.
- Sklearn compatible syntax.
- Train models using CPU/ GPU / TPU.
- Perform distributed training.
- Save Summaries on TensorBoard.
- Model checkpoint and failure recover.

AvenueCode

# Code

(Link)

AvenueCode

# tf.model_analysis

AvenueCode

# About Model Analysis

- Evaluate model performance on large datasets.
- Check performance choosing differents metrics and slices of data.
- Track Performance over the time.
- Friendly Visualization tool.

AvenueCode

# tf.serving

AvenueCode

# About Serving

- Low Latency,
- Scalable horizontally.
- Reliable & Robust.
- Load/ host multiple versions of a model.
- Built in A/B testing.

AvenueCode

5

# Qwiklabs

link

# Thank you!
# Questions?

Tulio Vieira de Souza
tulio.souza@avenuecode.com
Data Science Engineer