GPU INSTANCES CAN BE PROVISIONED EITHER BY USING A CLOUD MANAGED SERVICE OR COULD
EVEN BE PROVISIONED AS A SINGLE VM WITH ATTACHED GPU

OFFLOADING TRAINING TO GPU

Name
GPU

BASH
THE BOURNE-AGAIN SHELL

fluentd

KUBEFLOW PIPELINE TRAINING

Cloud
Storage

python

TensorBoard

IMAGE TRANSFORMER

SSD/OBJECT DETECTION MODEL

OCR MODEL

INTEGRATION TESTING ON SAMPLE
IMAGES, WHILE FETCHING THE
GOOD THRESHOLD FROM MLFLOW

Kubeflow

TRIGGERING A RETRAINING JOB BASED ON THE RETRAINING TRIGGER (FETCHING
NEW IMAGES FOR TRAINING)

RETRAINING Trigger
Cloud Functions

mlflow

Cloud
Storage

ARTIFACT STORE

SLZ (Safe landing zone)

IMAGE FEEDBACK

FOR EVERY IMAGE THAT IS IDENTIFIED AS A DRIFT THAT EVENT IS RECORDED

Pub/Sub

Trigger
Cloud Functions

SHADOW DEPLOYMENT OF MODELS BEFORE REPLACING

SHADOW DEPLOYMENT (ALL MODELS)

KUBEFLOW PIPELINE INFERENCE

ACTUAL INFERENCE PIPELINE CONTAINING CURRENT SERVING MODELS

DRIFT DETECTOR PILLAR

TFDV

AutoEncoder model

Structural Similarity Index

ScANN/Faiss

OCR model with positional output
(x1,y1,x2,y2)

Decision API
Cloud Endpoints

DECISION STEP

IMAGE TRANSFORMER

SSD/OBJECT DETECTION MODEL

OCR MODEL

PREDICTION RESULTS

SQL

Cloud
Storage

The Decision Step is a simply script that fetches distribution results for all the
"GOOD CANDIDATE" images from MLFlow artifacts and then compares it with the
current image