04

Production ML Pipelines

Production ML Pipelines

01 Ways to do ML on Google Cloud

02 Vertex AI Pipelines

03 AI Hub

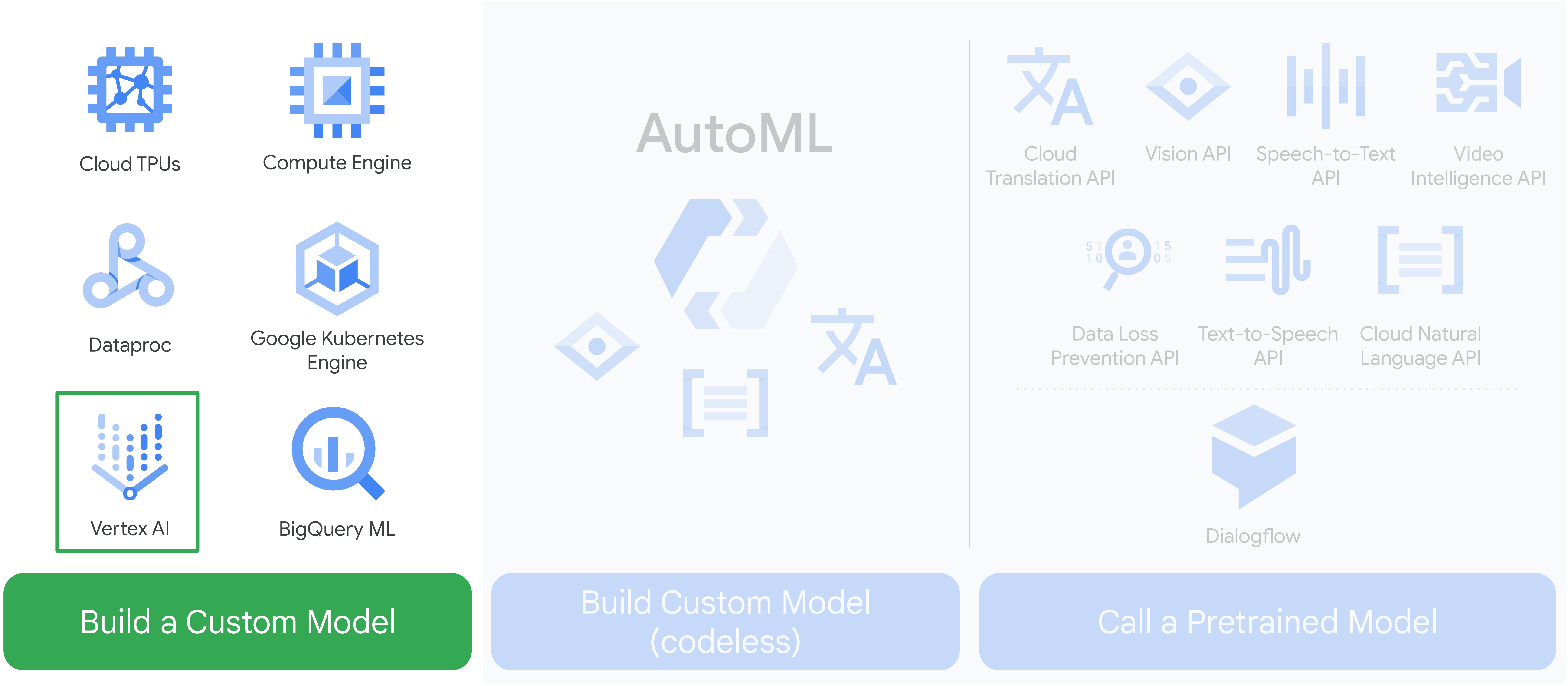
Production ML Pipelines

01 Ways to do ML on Google Cloud

02 Vertex AI Pipelines

03 AI Hub

Create and deploy custom models with Vertex AI



Cloud TPUs

Compute Engine

Dataproc

Google Kubernetes Engine

Vertex AI

BigQuery ML

Build a Custom Model

AutoML

Build Custom Model (codeless)

Cloud Translation API

Vision API

Speech-to-Text API

Video Intelligence API

Data Loss Prevention API

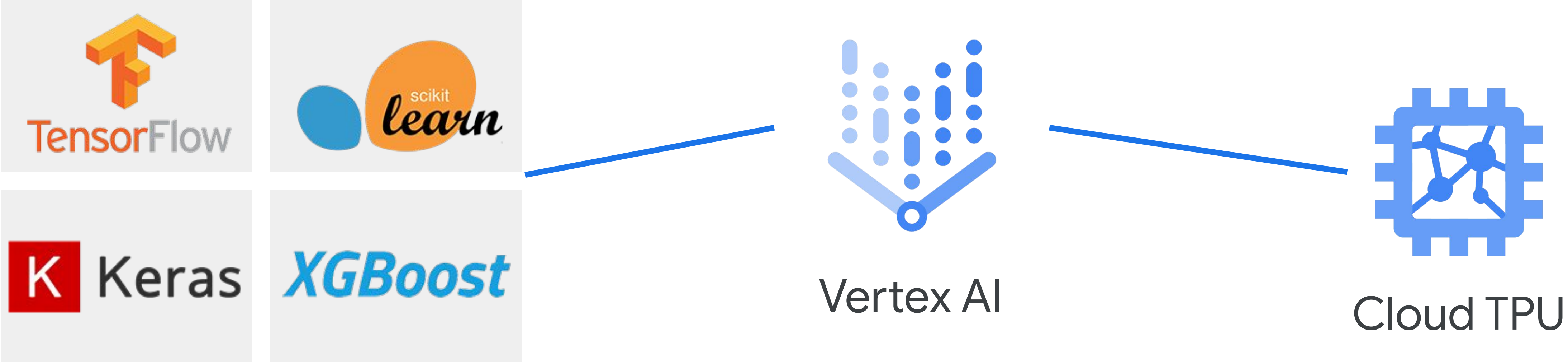
Text-to-Speech API

Cloud Natural Language API

Dialogflow

Call a Pretrained Model

Vertex AI is a fully managed service for custom machine learning models

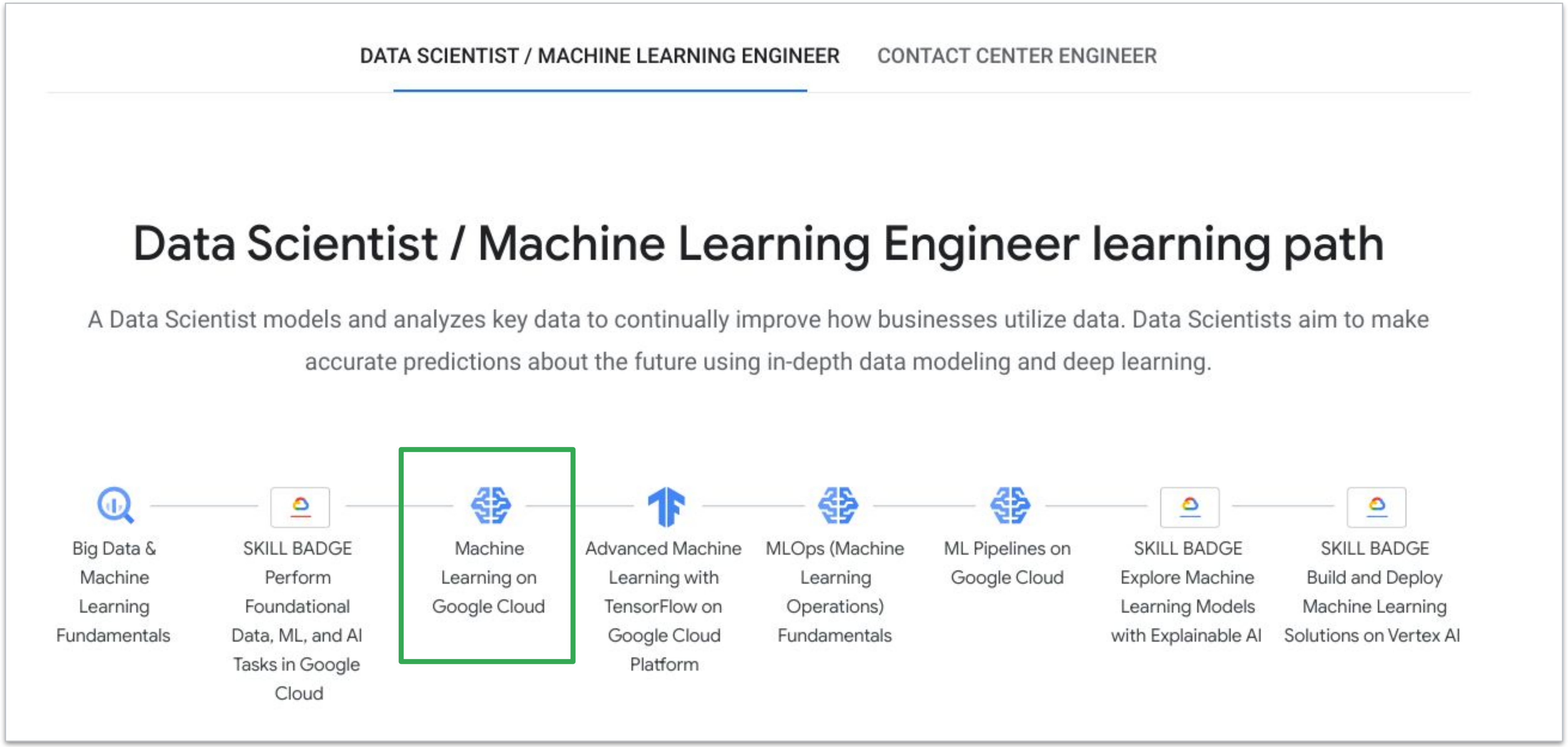


Vertex AI

Cloud TPU

- Scales to production
- Batching and distribution of model training
- Performs transformations on input data
- Hyper-parameter tuning
- Host and autoscale predictions
- Serverless - self-tuning - manages overhead

In this course, we don't cover writing TensorFlow models, only ways to operationalize them



DATA SCIENTIST / MACHINE LEARNING ENGINEER   CONTACT CENTER ENGINEER

Data Scientist / Machine Learning Engineer learning path

A Data Scientist models and analyzes key data to continually improve how businesses utilize data. Data Scientists aim to make accurate predictions about the future using in-depth data modeling and deep learning.

Big Data & Machine Learning Fundamentals
SKILL BADGE Perform Foundational Data, ML, and AI Tasks in Google Cloud
Machine Learning on Google Cloud
Advanced Machine Learning with TensorFlow on Google Cloud Platform
MLOps (Machine Learning Operations) Fundamentals
ML Pipelines on Google Cloud
SKILL BADGE Explore Machine Learning Models with Explainable AI
SKILL BADGE Build and Deploy Machine Learning Solutions on Vertex AI

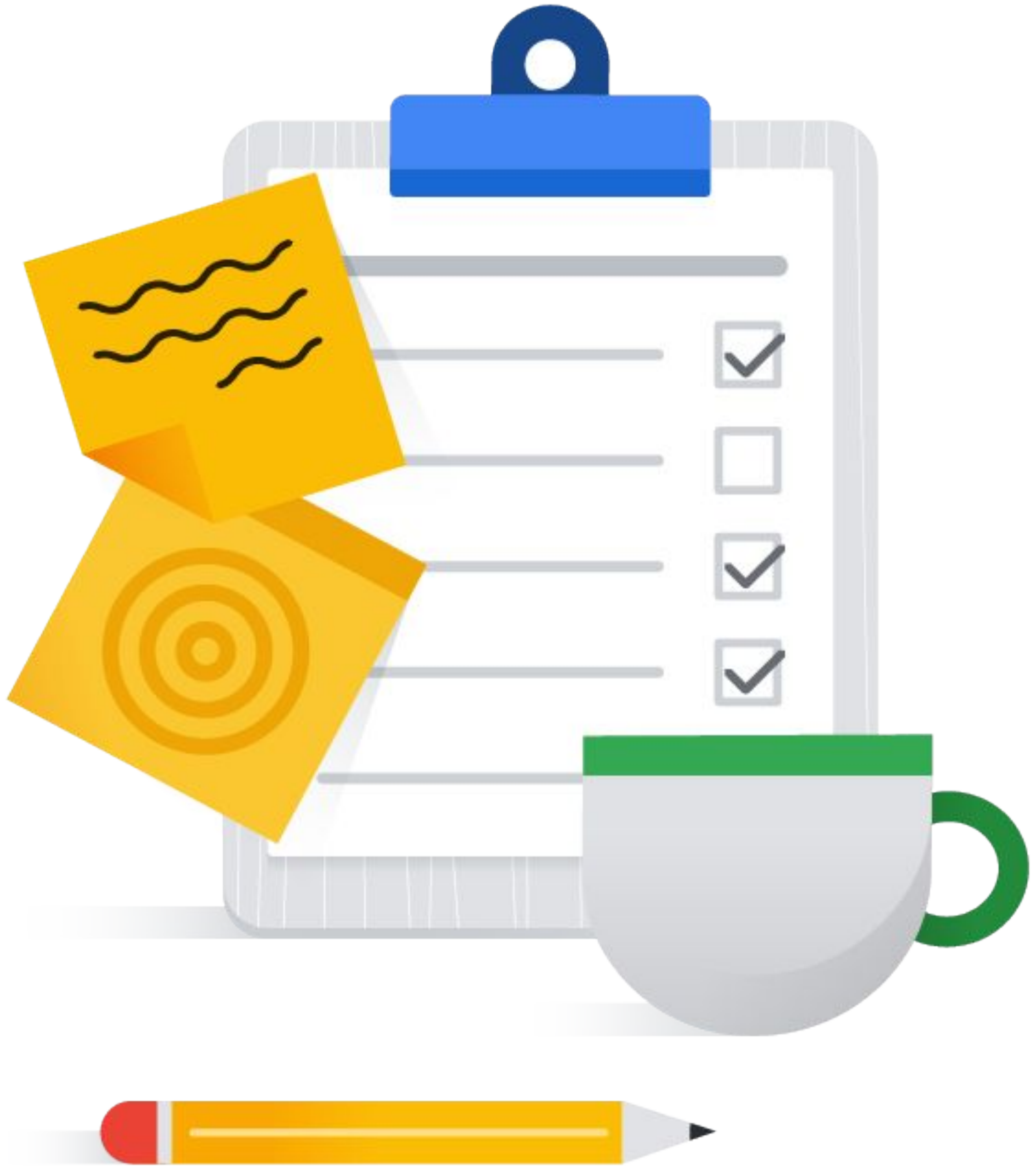
Google Cloud Training - Machine Learning and AI

Production ML Pipelines

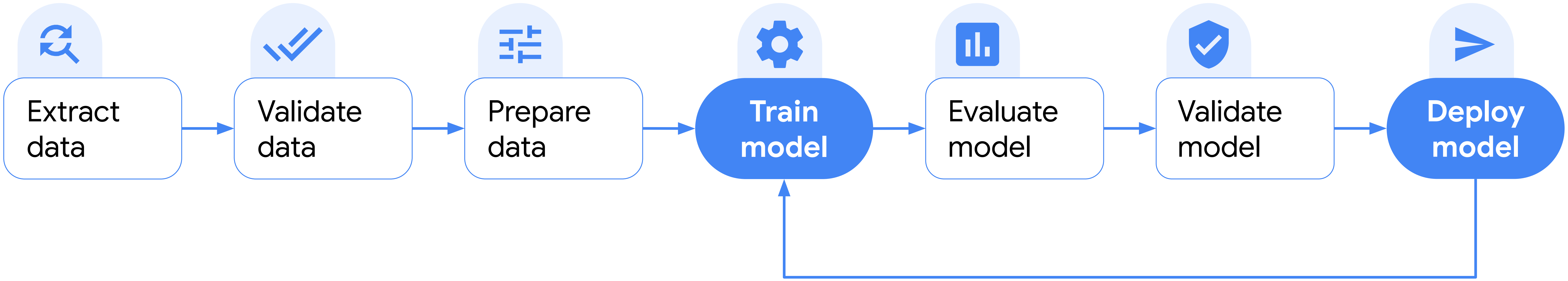
01 Ways to do ML on Google Cloud

02 Vertex AI Pipelines

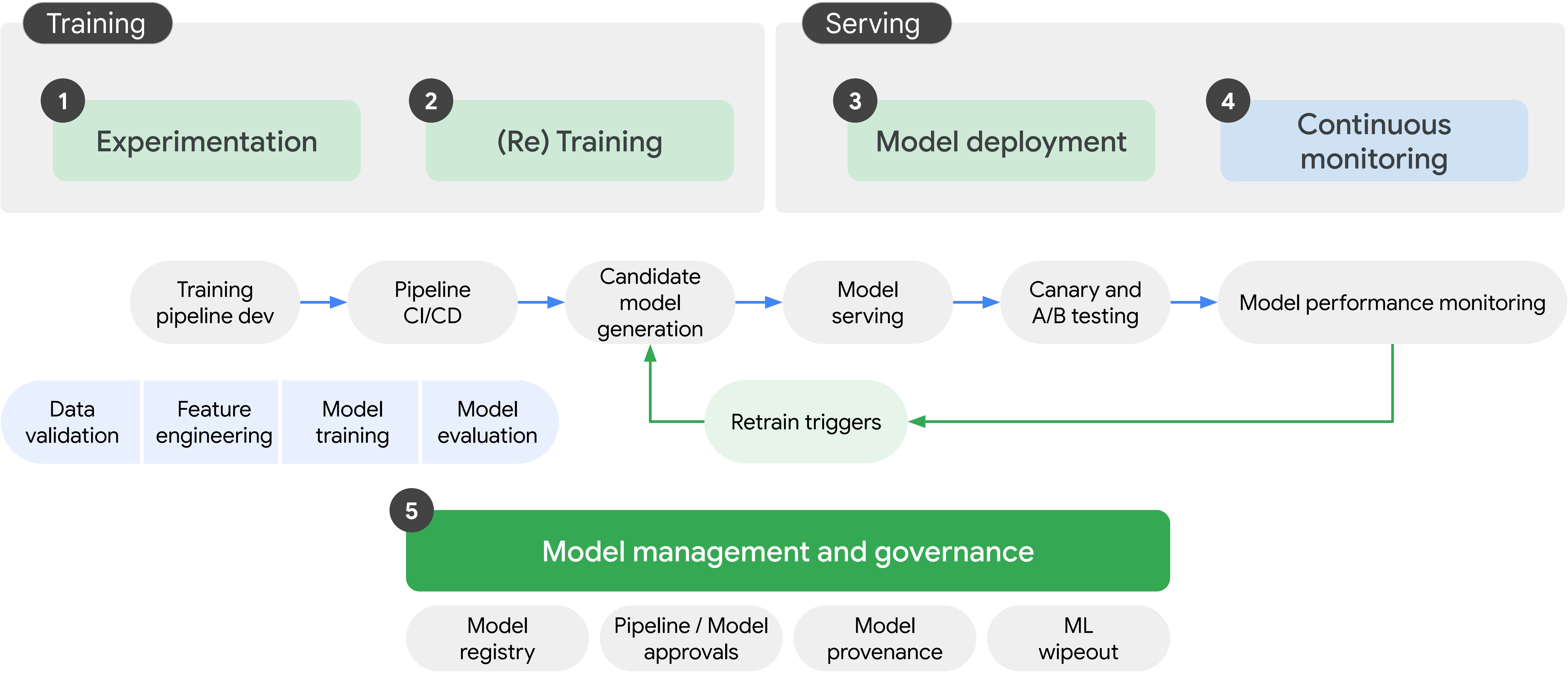
03 AI Hub



Pipelines automate the training and deployment of models



Pipelines are the backbone of production ML systems



Pipelines product portfolio

Kubeflow

Kubeflow Pipelines

- Kubernetes-native.
- Open source.
- The industry standard for running ML Pipelines.

Google Cloud

AI Platform Pipelines - Hosted Beta

- Kubeflow pipelines running on Google Cloud.
- Optimized for GKE.
- Integrated with Google Cloud services.

Vertex Pipelines - Managed PREVIEW

- Fully managed and serverless.
- Allows users to focus on building their pipelines, scale easily, and pay only for the resources they use.

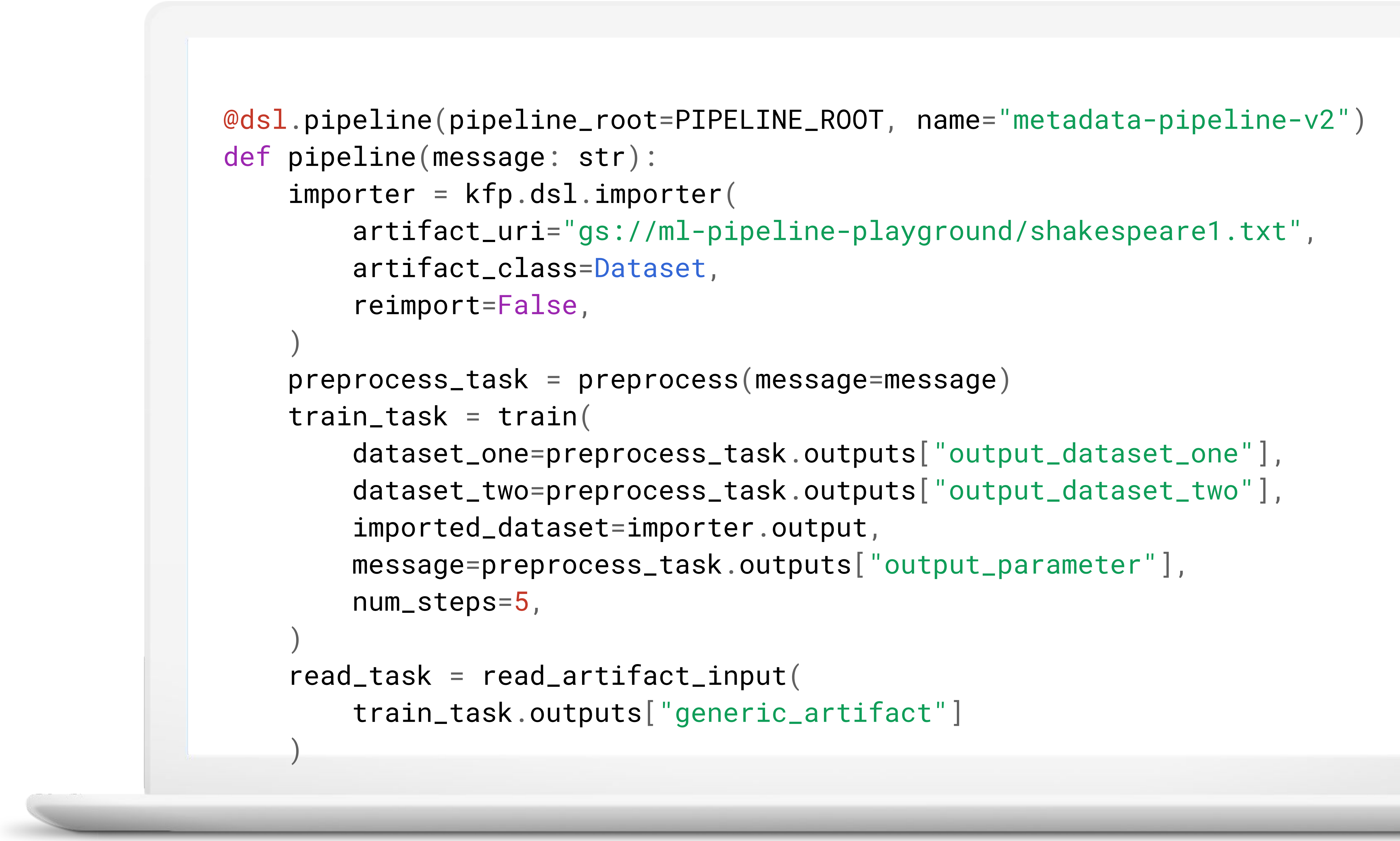
Write your pipeline

Easy to use Python SDKs

Build pipelines using Data Scientist friendly SDKs like TensorFlow Extended and Kubeflow Pipelines.

Rich, scalable pre-built components

We provide a rich set of pre-built components for common ML tasks, which leverage Google Cloud services.



```
@dsl.pipeline(pipeline_root=PIPELINE_ROOT, name="metadata-pipeline-v2")
def pipeline(message: str):
    importer = kfp.dsl.importer(
        artifact_uri="gs://ml-pipeline-playground/shakespeare1.txt",
        artifact_class=Dataset,
        reimport=False,
    )
    preprocess_task = preprocess(message=message)
    train_task = train(
        dataset_one=preprocess_task.outputs["output_dataset_one"],
        dataset_two=preprocess_task.outputs["output_dataset_two"],
        imported_dataset=importer.output,
        message=preprocess_task.outputs["output_parameter"],
        num_steps=5,
    )
    read_task = read_artifact_input(
        train_task.outputs["generic_artifact"]
    )
```

Key capabilities

Python SDKs

Data scientist friendly Python SDKs

1

Scalable

Run as many pipelines on as much data as you want.

2

Metadata and lineage

Store metadata for every artifact produced by the pipeline.

3

Monitoring UIs and APIs

Track and debug pipelines executions.

4

Security

Supports IAM, VPC-SC, and CMEK.

5

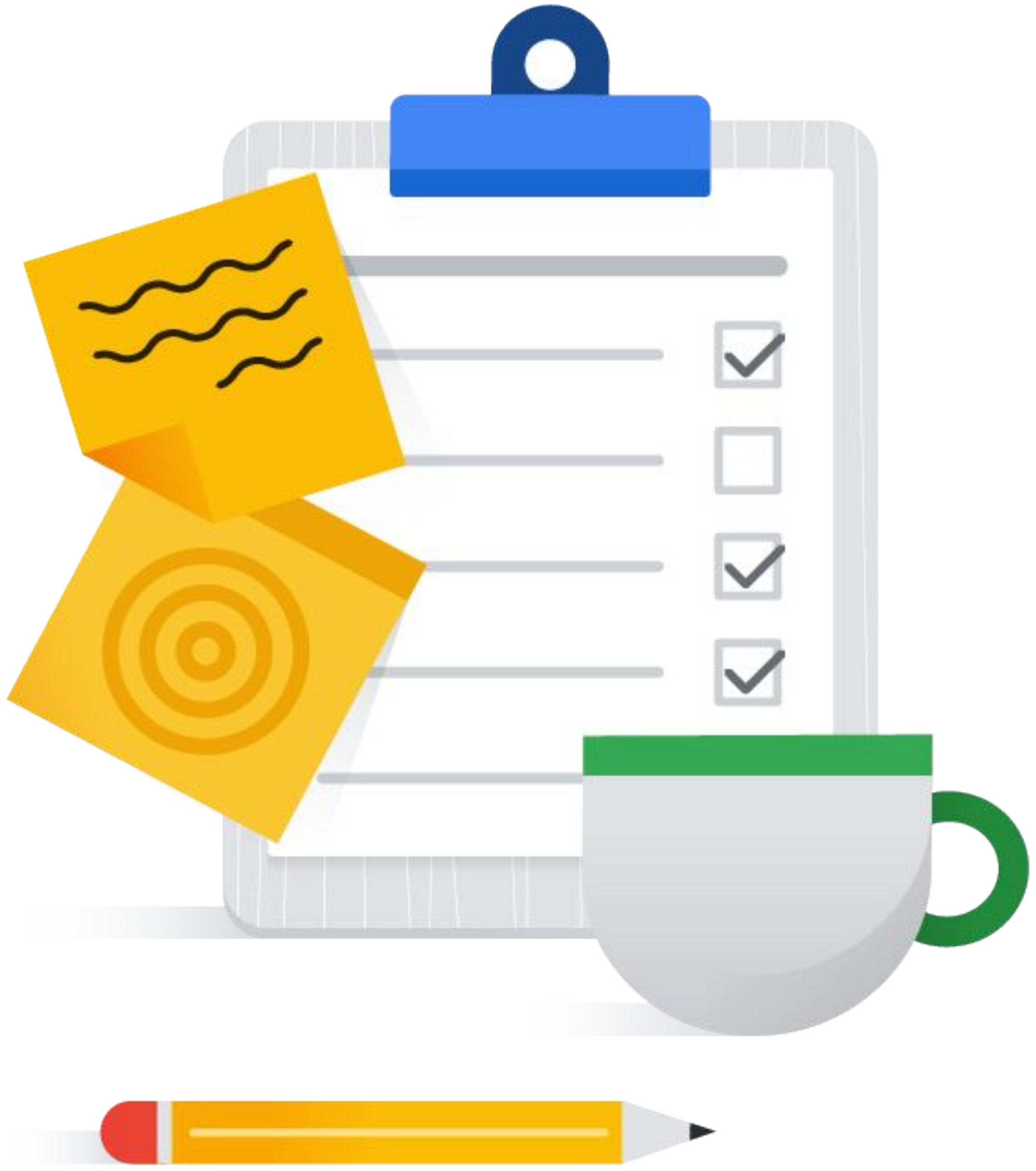
Cost-effective

Only pay for the pipelines you run and the resources they use.

6

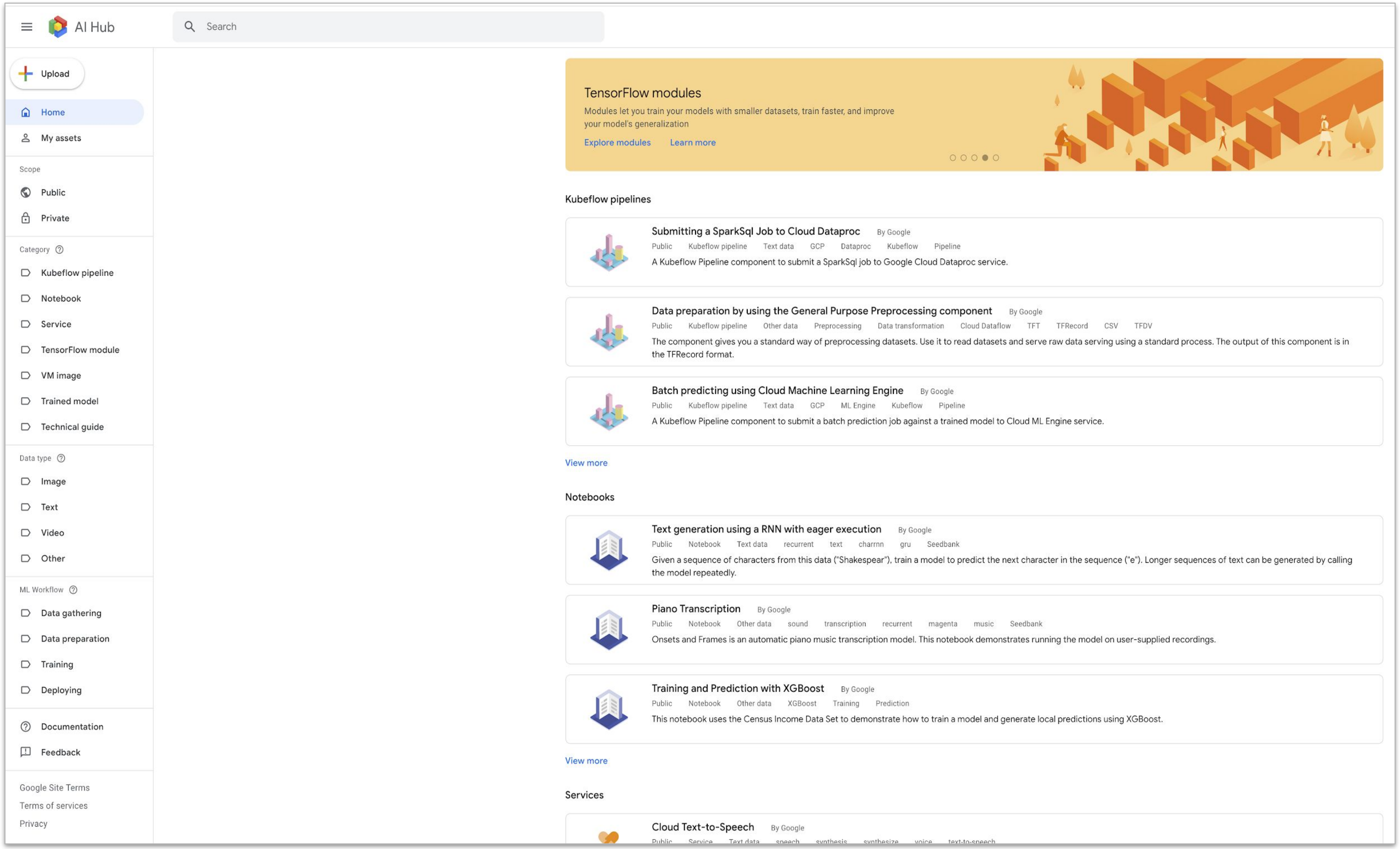
Production ML Pipelines with Kubeflow

01	Ways to do ML on Google Cloud
02	Vertex AI Pipelines
03	AI Hub



AI Hub is a repository for AI assets

Don’t reinvent the wheel! Find and deploy ML pipelines.



TensorFlow modules

Modules let you train your models with smaller datasets, train faster, and improve your model's generalization

Kubeflow pipelines

Submitting a SparkSql Job to Cloud Dataproc By Google
Public Kubeflow pipeline Text data GCP Dataproc Kubeflow Pipeline
A Kubeflow Pipeline component to submit a SparkSql job to Google Cloud Dataproc service.

Data preparation by using the General Purpose Preprocessing component By Google
Public Kubeflow pipeline Other data Preprocessing Data transformation Cloud Dataflow TFT TFRecord CSV TFDV
The component gives you a standard way of preprocessing datasets. Use it to read datasets and serve raw data serving using a standard process. The output of this component is in the TFRecord format.

Batch predicting using Cloud Machine Learning Engine By Google
Public Kubeflow pipeline Text data GCP ML Engine Kubeflow Pipeline
A Kubeflow Pipeline component to submit a batch prediction job against a trained model to Cloud ML Engine service.

Notebooks

Text generation using a RNN with eager execution By Google
Public Notebook Text data recurrent text charrnn gru Seedbank
Given a sequence of characters from this data ("Shakespear"), train a model to predict the next character in the sequence ("e"). Longer sequences of text can be generated by calling the model repeatedly.

Piano Transcription By Google
Public Notebook Other data sound transcription recurrent magenta music Seedbank
Onsets and Frames is an automatic piano music transcription model. This notebook demonstrates running the model on user-supplied recordings.

Training and Prediction with XGBoost By Google
Public Notebook Other data XGBoost Training Prediction
This notebook uses the Census Income Data Set to demonstrate how to train a model and generate local predictions using XGBoost.

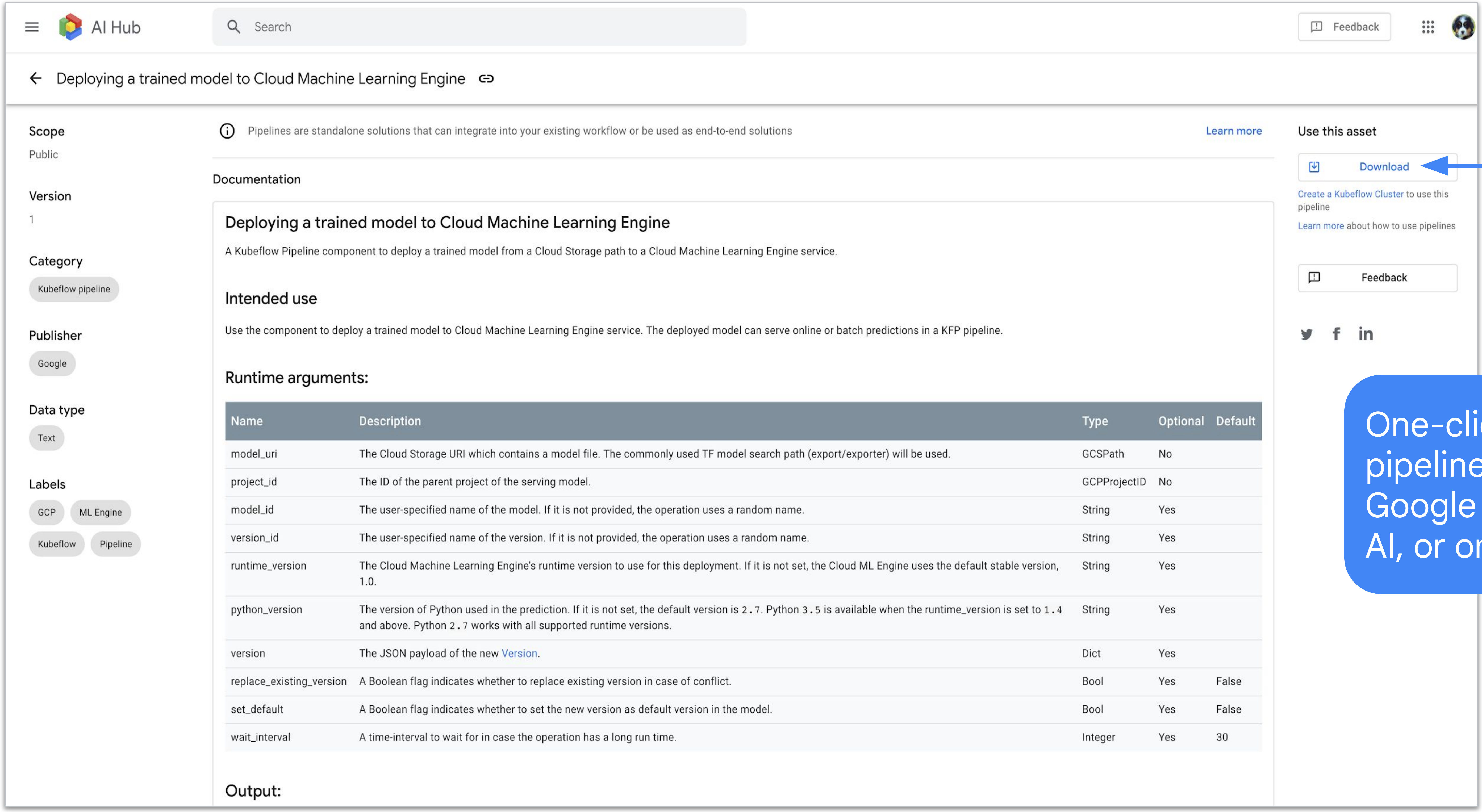
Services

Cloud Text-to-Speech By Google
Public Service Text data speech synthesis synthesize voice text-to-speech

AI Hub stores various asset types

- Kubeflow pipelines and components
- Jupyter notebooks
- TensorFlow modules
- Trained models
- Services
- VM images

This is what a typical asset looks like



Deploying a trained model to Cloud Machine Learning Engine

Scope

Public

Version

1

Category

Kubeflow pipeline

Publisher

Google

Data type

Text

Labels

GCP ML Engine Kubeflow Pipeline

Pipelines are standalone solutions that can integrate into your existing workflow or be used as end-to-end solutions Learn more

Documentation

Deploying a trained model to Cloud Machine Learning Engine

A Kubeflow Pipeline component to deploy a trained model from a Cloud Storage path to a Cloud Machine Learning Engine service.

Intended use

Use the component to deploy a trained model to Cloud Machine Learning Engine service. The deployed model can serve online or batch predictions in a KFP pipeline.

Runtime arguments:

Name	Description	Type	Optional	Default
model_uri	The Cloud Storage URI which contains a model file. The commonly used TF model search path (export/exporter) will be used.	GCSPath	No	
project_id	The ID of the parent project of the serving model.	GCPProjectID	No	
model_id	The user-specified name of the model. If it is not provided, the operation uses a random name.	String	Yes	
version_id	The user-specified name of the version. If it is not provided, the operation uses a random name.	String	Yes	
runtime_version	The Cloud Machine Learning Engine's runtime version to use for this deployment. If it is not set, the Cloud ML Engine uses the default stable version, 1.0.	String	Yes	
python_version	The version of Python used in the prediction. If it is not set, the default version is `2.7`. Python `3.5` is available when the runtime_version is set to `1.4` and above. Python `2.7` works with all supported runtime versions.	String	Yes	
version	The JSON payload of the new Version.	Dict	Yes	
replace_existing_version	A Boolean flag indicates whether to replace existing version in case of conflict.	Bool	Yes	False
set_default	A Boolean flag indicates whether to set the new version as default version in the model.	Bool	Yes	False
wait_interval	A time-interval to wait for in case the operation has a long run time.	Integer	Yes	30

Output:

Use this asset

Download

Create a Kubeflow Cluster to use this pipeline

Learn more about how to use pipelines

Feedback

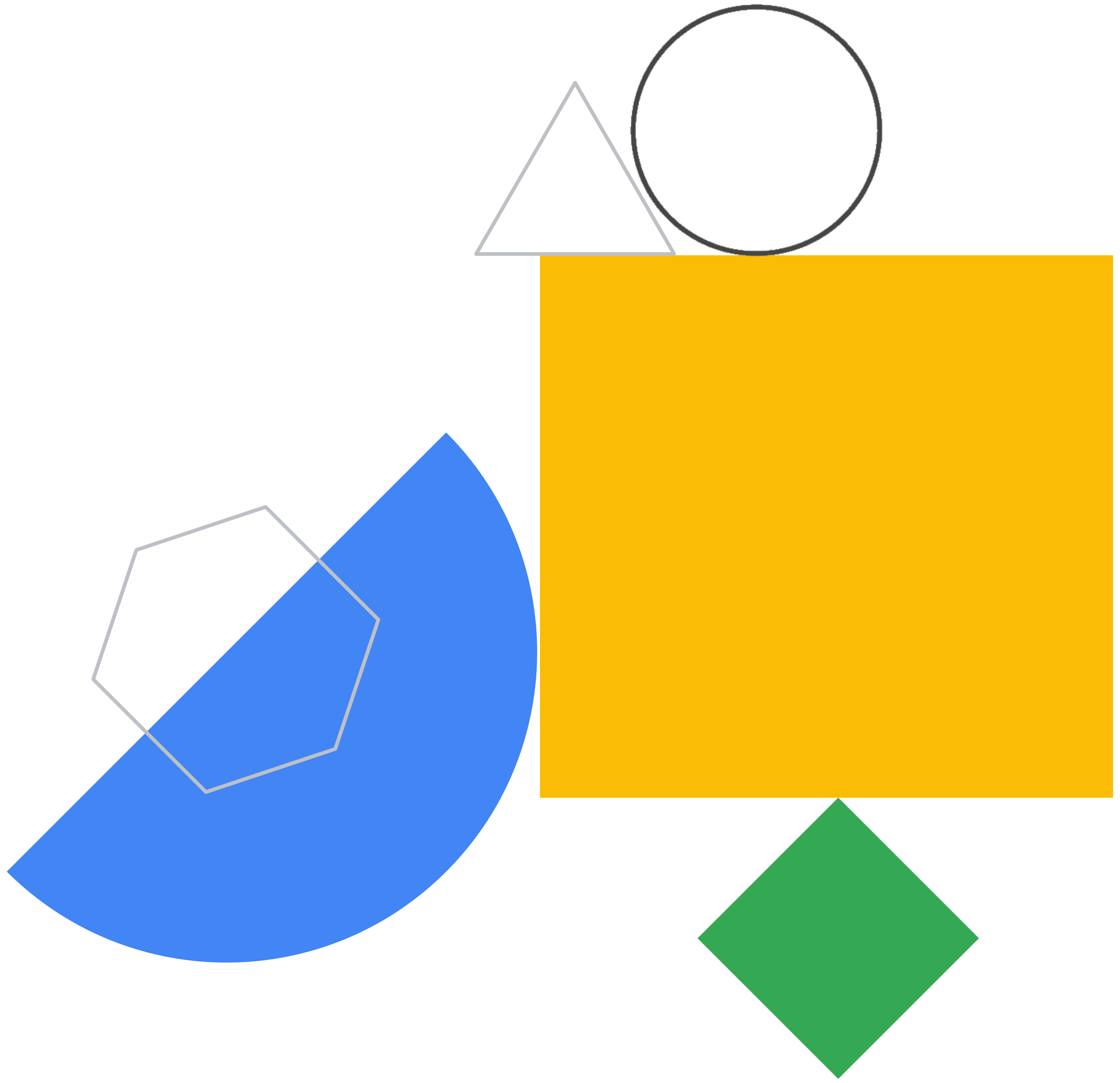
One-click deployment of ML pipelines via Kubeflow on Google Cloud as platform for AI, or on premise.

Assets on AI Hub are collected in two scopes: public assets and restricted assets

- Public scope are available to all AI Hub users.
- Restricted scope contains AI components that you have uploaded and assets that have been shared with you.

Lab Intro

Running Pipelines on Vertex AI



Lab objectives

01 Set up the project environment

02 Inspect and configure pipeline code

03 Execute the AI pipeline

Summary

- Use ML on Google Cloud using either:
  - Vertex AI (your model, your data)
  - AutoML (our models, your data)
- Use Vertex AI Pipelines to deploy end-to-end ML pipelines.
- Don’t reinvent the wheel for your ML pipeline! Leverage pipelines on AI Hub.