

ML Pipelines on Google Cloud

Course ·

6% complete

[Professional Machine Learning Engineer Certification Learning Path](#) navigate_next [ML Pipelines on Google Cloud](#) navigate_next Introduction to TFX Pipelines

TFX Standard Components Walkthrough

2 hours Free

Overview

In this lab, you will work with the Covertypes Data Set and use TFX to analyze, understand, and pre-process the dataset and train, analyze, validate, and deploy a multi-class classification model to predict the type of forest cover from cartographic features.

Objectives

- Develop a high level understanding of TFX pipeline components.
- Learn how to use a TFX Interactive Context for prototype development of TFX pipelines.
- Work with the Tensorflow Data Validation (TFDV) library to check and analyze input data.
- Utilize the Tensorflow Transform (TFT) library for scalable data preprocessing and feature transformations.
- Employ the Tensorflow Model Analysis (TFMA) library for model evaluation.

Setup

For each lab, you get a new Google Cloud project and set of resources for a fixed time at no cost.


1. Sign in to Qwiklabs using an **incognito window**.
2. Note the lab's access time (for example, 1:15:00), and make sure you can finish within that time.
There is no pause feature. You can restart if needed, but you have to start at the beginning.
3. When ready, click **Start lab**.
4. Note your lab credentials (**Username** and **Password**). You will use them to sign in to the Google Cloud Console.
5. Click **Open Google Console**.
6. Click **Use another account** and copy/paste credentials for **this** lab into the prompts.
If you use other credentials, you'll receive errors or **incur charges**.

7. Accept the terms and skip the recovery resource page.

Note: Do not click **End Lab** unless you have finished the lab or want to restart it. This clears your work and removes the project.

Activate Cloud Shell

Cloud Shell is a virtual machine that contains development tools. It offers a persistent 5-GB home directory and runs on Google Cloud. Cloud Shell provides command-line access to your Google Cloud resources. `gcloud` is the command-line tool for Google Cloud. It comes pre-installed on Cloud Shell and supports tab completion.

1. Click the **Activate Cloud Shell** button () at the top right of the console.
2. Click **Continue**.

It takes a few moments to provision and connect to the environment. When you are connected, you are also authenticated, and the project is set to your *PROJECT_ID*.

Sample commands

- List the active account name:

```
gcloud auth list
```

(Output)

```
Credentialed accounts: - <myaccount>@<mydomain>.com (active)
```

(Example output)

```
Credentialed accounts: - google1623327_student@qwiklabs.net
```

- List the project ID:

```
gcloud config list project
```

(Output)

```
[core] project = <project_ID>
```

(Example output)

[core] project = qwiklabs-gcp-44776a13dea667a6 **Note:** Full documentation of **gcloud** is available in the [gcloud CLI overview guide](#).

Task 1. Enable Cloud Services

1. In Cloud Shell, run the command below to set the project id to your Google Cloud Project:

```
PROJECT_ID={{project_0.project_id|Project ID}} gcloud config set project $PROJECT_ID
```

2. Next, execute the following commands to enable the required cloud services:

```
gcloud services enable \ cloudbuild.googleapis.com \ container.googleapis.com \ cloudresourcemanager.googleapis.com \ iam.googleapis.com \  
containerregistry.googleapis.com \ containeranalysis.googleapis.com \ ml.googleapis.com \ dataflow.googleapis.com \ notebooks.googleapis.com
```

3. Next, add the Editor permission for your Cloud Build service account:

```
PROJECT_NUMBER=$(gcloud projects describe $PROJECT_ID --format="value(projectNumber)")  
CLOUD_BUILD_SERVICE_ACCOUNT="${PROJECT_NUMBER}@cloudbuild.gserviceaccount.com" gcloud projects add-iam-policy-binding  
$PROJECT_ID \ --member serviceAccount:$CLOUD_BUILD_SERVICE_ACCOUNT \ --role roles/editor
```

4. Now, create a custom service account to give CAIP training job access to AI Platform Vizier service for pipeline hyperparameter tuning:

```
SERVICE_ACCOUNT_ID=tfx-tuner-caip-service-account gcloud iam service-accounts create $SERVICE_ACCOUNT_ID \ --description="A  
custom service account for CAIP training job to access AI Platform Vizier service for pipeline hyperparameter tuning." \ --display-name="TFX  
Tuner CAIP Vizier"
```

5. Grant your AI Platform service account additional access permissions to the AI Platform Vizier service for pipeline hyperparameter tuning:

```
PROJECT_NUMBER=$(gcloud projects describe $PROJECT_ID --format="value(projectNumber)") CAIP_SERVICE_ACCOUNT="service-  
${PROJECT_NUMBER}@cloud-ml.google.com.iam.gserviceaccount.com" gcloud projects add-iam-policy-binding $PROJECT_ID \ --member  
serviceAccount:$CAIP_SERVICE_ACCOUNT \ --role=roles/storage.objectAdmin gcloud projects add-iam-policy-binding $PROJECT_ID \ --  
member serviceAccount:$CAIP_SERVICE_ACCOUNT \ --role=roles/ml.admin
```

6. Grant service account access to Storage admin role:

```
SERVICE_ACCOUNT_ID=tfx-tuner-caip-service-account gcloud projects add-iam-policy-binding $PROJECT_ID \ --  
member=serviceAccount:${SERVICE_ACCOUNT_ID}@${PROJECT_ID}.iam.gserviceaccount.com \ --role=roles/storage.objectAdmin
```

7. Grant service account access to AI Platform Vizier role:

```
gcloud projects add-iam-policy-binding $PROJECT_ID \ --  
member=serviceAccount:${SERVICE_ACCOUNT_ID}@${PROJECT_ID}.iam.gserviceaccount.com \ --role=roles/ml.admin
```

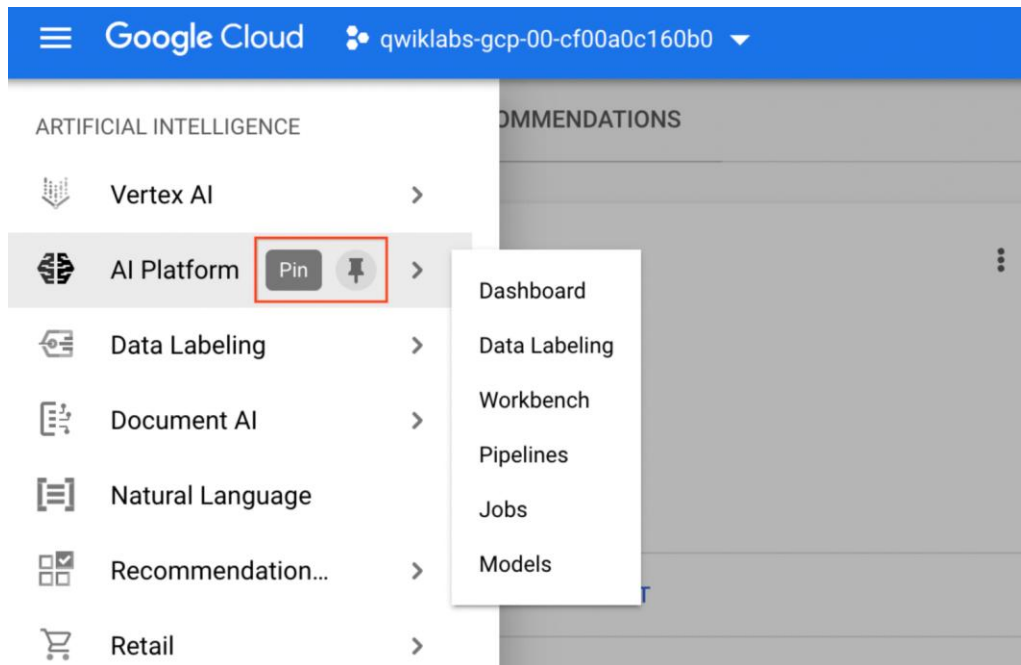
8. Grant your project's AI Platform Google-managed service account the Service Account Admin role for your AI Platform service account:

```
gcloud iam service-accounts add-iam-policy-binding \ --role=roles/iam.serviceAccountAdmin \ --member=serviceAccount:service-  
${PROJECT_NUMBER}@cloud-ml.google.com.iam.gserviceaccount.com \  
${SERVICE_ACCOUNT_ID}@${PROJECT_ID}.iam.gserviceaccount.com
```

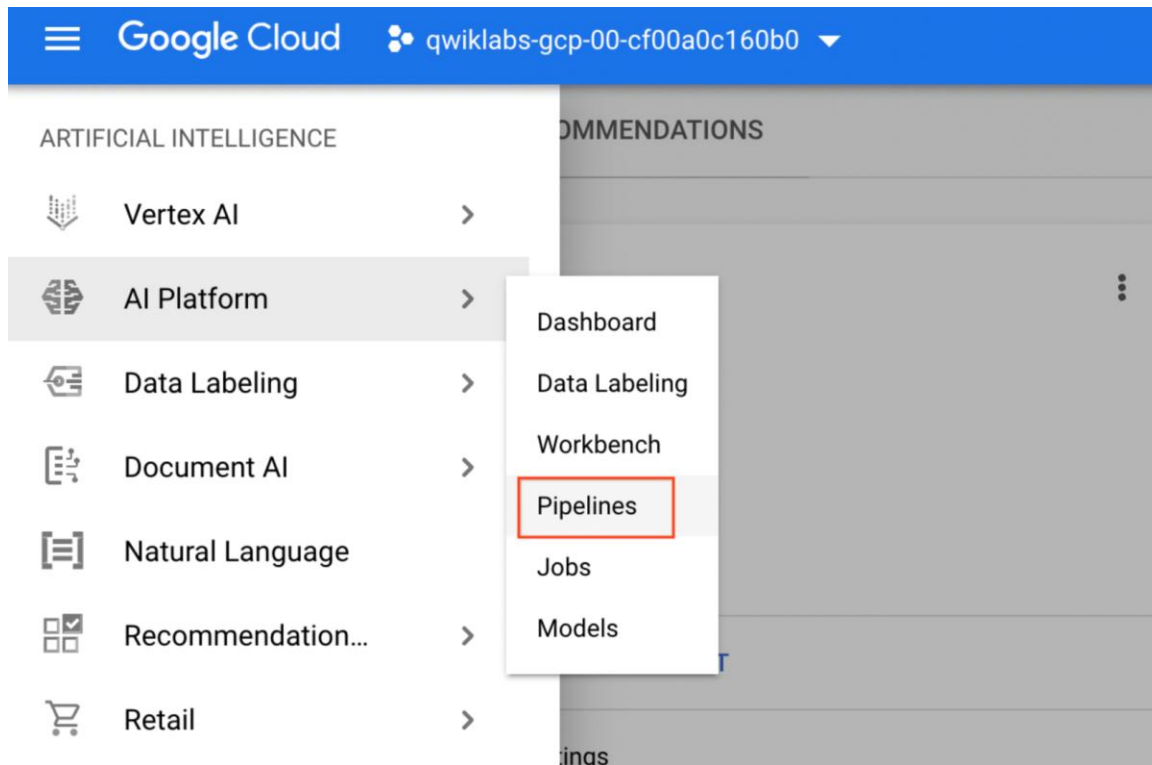
Click **Check my progress** to verify the objective. Enable cloud services

Task 2. Create an instance of AI Platform Pipelines

1. From the **Navigation menu**, scroll down to **AI Platform** and pin the section for easier access later in the lab.



2. Click **Pipelines**.



3. Then click **New Instance**.

Google Cloud qwiklabs-gcp-00-cf00a0c160b0

Search Products, resources, docs (/)

AI Platform

AI Platform Pipelines BETA

+ NEW INSTANCE REFRESH DELETE

Vertex AI is our next generation AI Platform, with many new features that are unavailable in the current platform. productionize models with MLOps. [Learn More](#)

MIGRATE TO VERTEX AI GO TO VERTEX AI

Filter Filter

Status	Name	Zone	Version	Cluster	Namespace
No rows to display					

4. Click **Configure**.



Kubeflow Pipelines

Version: Kubeflow Pipelines 1.7.1

[Google Cloud AI Platform](#)

Reusable end-to-end ML workflow platform



5. On the Kubeflow Pipelines page, check **Allow access to the following Cloud APIs**, leave the name as is, and then click **Create New Cluster**.

This should take 2-3 minutes to complete. Wait for the cluster to finish before proceeding to the next step. In the first tab opened, you can view the Cluster Creation taking place in the [GKE section of the Cloud Console](#), or see the individual VMs spinning up in the [GCE section of the Cloud Console](#).

6. When the cluster creation is complete, check the **Terms of Service** box, leave other settings unchanged, and then click **Deploy**. You will see the individual services of KFP deployed to your GKE cluster.

Proceed to the next step while installation occurs.

Click **Check my progress** to verify the objective. Create an instance of AI Platform Pipelines

Task 3. Create an instance of Cloud AI Platform Notebooks

1. In Cloud Shell, run the following commands to create a user-managed notebook:

```
gcloud notebooks instances create tfx-on-googlecloud --vm-image-project=deeplearning-platform-release --vm-image-family=tf2-2-3-cpu --machine-type=e2-medium --location={{project_0.default_zone|zone_to_be_set}}
```

2. Click on the **Navigation Menu**. Navigate to **Vertex AI**, then to **Workbench**.
3. Click on **USER-MANAGED NOTEBOOKS**

The notebook creation will take 2-3 minutes to complete.

4. Click **Open JupyterLab** next to your notebook name. A JupyterLab window opens in a new tab.

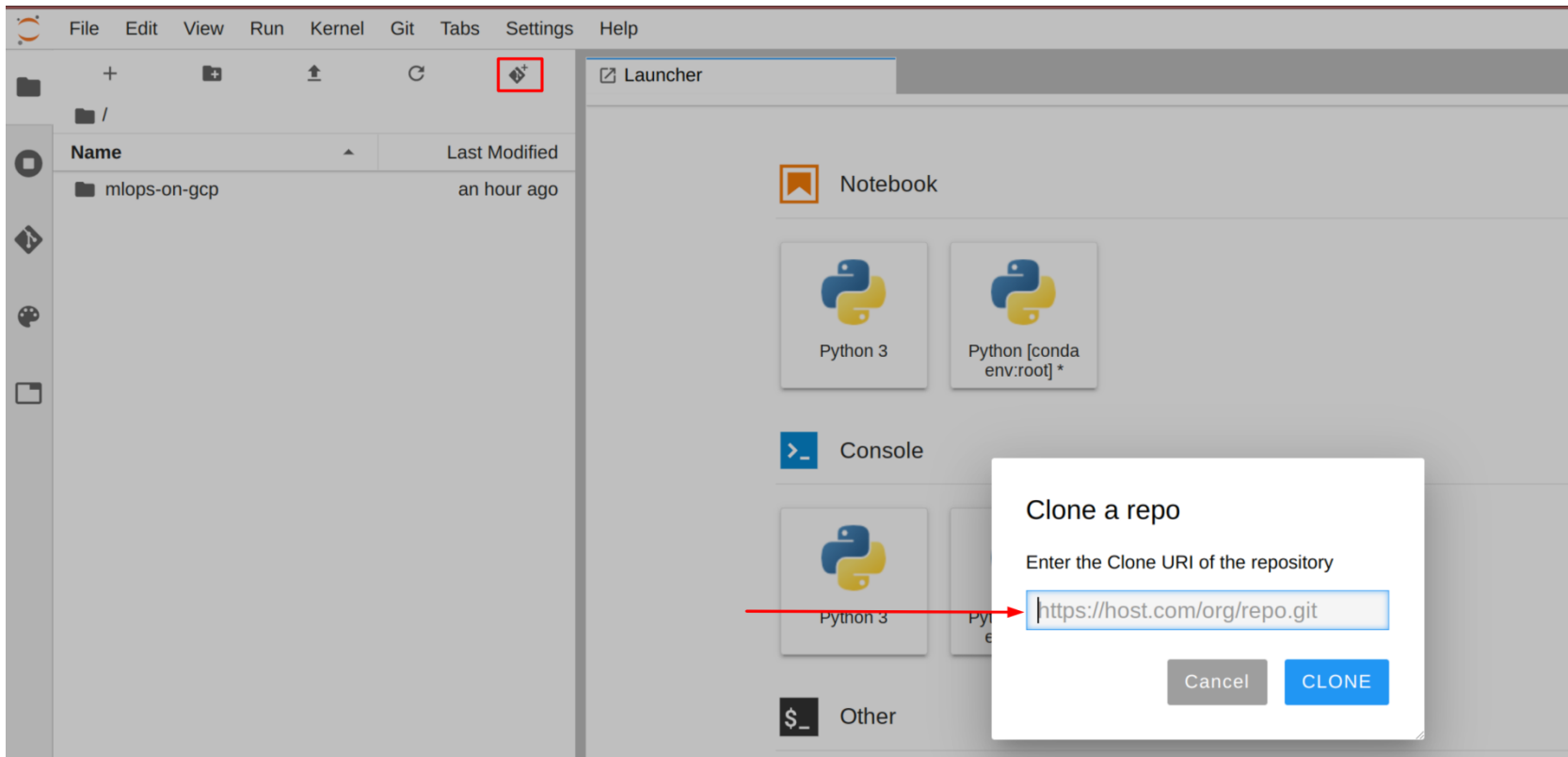
Click **Check my progress** to verify the objective. Create an instance of Cloud AI Platform Notebooks

Task 4. Clone the example repo within your AI Platform Notebooks instance

To clone the `mlops-on-gcp` notebook in your JupyterLab instance:

1. In JupyterLab, click the **Terminal** icon to open a new terminal.
2. At the command-line prompt, type in the following command and press Enter:

`git clone https://github.com/GoogleCloudPlatform/mlops-on-gcp` **Note:** If the cloned repo does not appear in the JupyterLab UI, you can use the top line menu and under **Git > Clone a repository**, clone the repo (`https://github.com/GoogleCloudPlatform/mlops-on-gcp`) using the UI.



3. Confirm that you have cloned the repository by double clicking on the `mlops-on-gcp` directory and ensuring that you can see its contents. The files for all the Jupyter notebook-based labs throughout this course are available in this directory.

Navigate to the example notebook

1. In the JupyterLab, open a terminal and execute the following commands:

`cd mlops-on-gcp/workshops/tfx-caip-tf23 ./install.sh` **Note:** Please ignore pip incompatibility warnings and 404 Client Errors.

2. Now, in AI Platform Notebooks, navigate to `mlops-on-gcp/workshops/tfx-caip-tf23/lab-01-tfx-walkthrough/labs` and open `lab-01.ipynb`.
3. Clear all the cells in the notebook (Edit > Clear All Outputs) and then Run the cells one by one.
4. When prompted, come back to these instructions to check your progress.

If you need more help, you may take a look at the complete solution by navigating to `mlops-on-gcp/workshops/tfx-caip-tf23/lab-01-tfx-walkthrough/solutions`, open `lab-01.ipynb`.

Click **Check my progress** to verify the objective. Clone the example repo within AI Platform Notebooks instance

Task 5. Run your training job in the Cloud

Test completed tasks - Configure and run CsvExampleGen

Click **Check my progress** to verify the objective. Configure and run CsvExampleGen

Test completed task - Create and run training component

Click **Check my progress** to verify the objective. Create and run training component

Congratulations!

This concludes your introductory walkthrough through TFX pipeline components. In the lab, you used TFX to analyze, understand, and pre-process the dataset and train, analyze, validate, and deploy a multi-class classification model to predict the type of forest cover from cartographic features

End your lab

When you have completed your lab, click **End Lab**. Qwiklabs removes the resources you've used and cleans the account for you.

You will be given an opportunity to rate the lab experience. Select the applicable number of stars, type a comment, and then click **Submit**.

The number of stars indicates the following:

- 1 star = Very dissatisfied
- 2 stars = Dissatisfied
- 3 stars = Neutral
- 4 stars = Satisfied
- 5 stars = Very satisfied

You can close the dialog box if you don't want to provide feedback.

For feedback, suggestions, or corrections, please use the **Support** tab.

Copyright 2022 Google LLC All rights reserved. Google and the Google logo are trademarks of Google LLC. All other company and product names may be trademarks of the respective companies with which they are associated.

- [Overview](#)
- [Setup](#)
- [Task 1. Enable Cloud Services](#)
- [Task 2. Create an instance of AI Platform Pipelines](#)
- [Task 3. Create an instance of Cloud AI Platform Notebooks](#)
- [Task 4. Clone the example repo within your AI Platform Notebooks instance](#)
- [Task 5. Run your training job in the Cloud](#)
- [Congratulations!](#)
- [End your lab](#)