

CI/CD for Kubeflow Pipelines on Al Platform

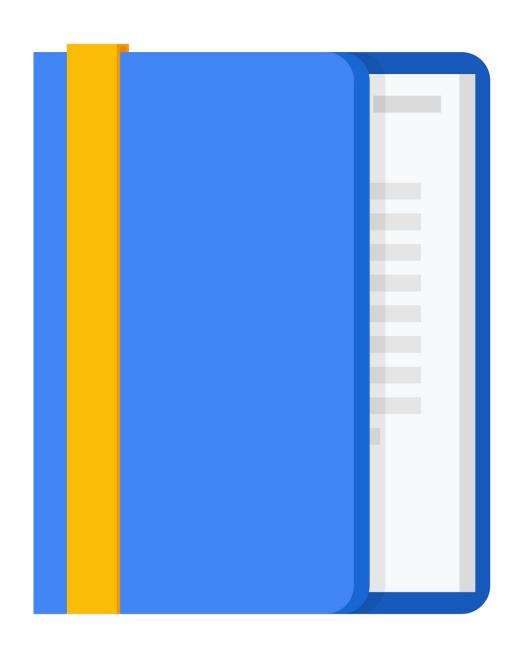
Agenda

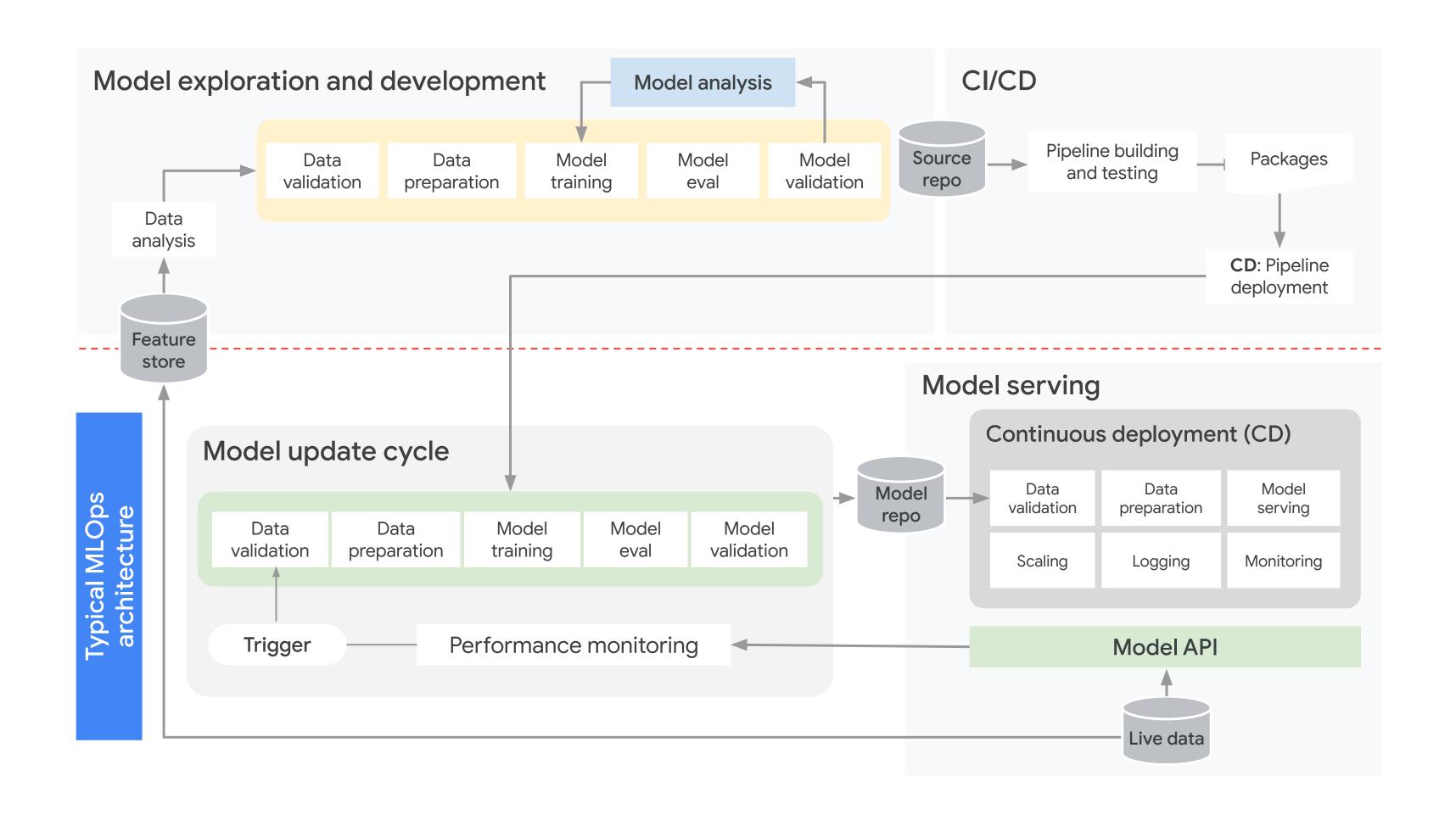
Concept Overview

Cloud Build Builders

Cloud Build Configuration

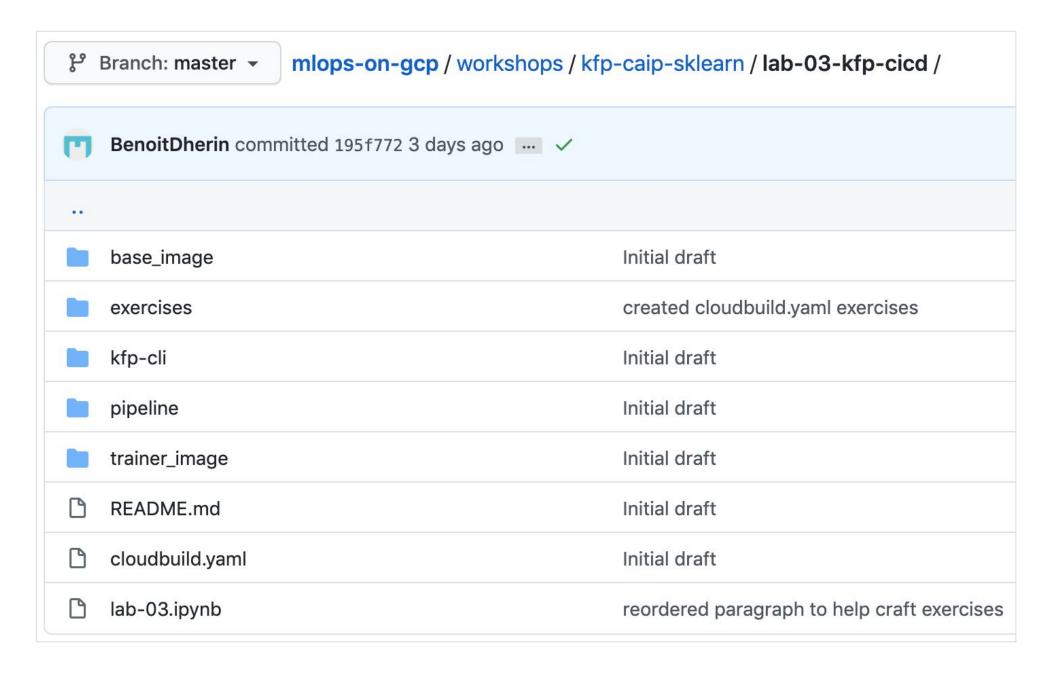
Cloud Build Triggers





Every container is a self-contained directory in repo

If any of these files is changed, you need to rebuild and push the Docker image.

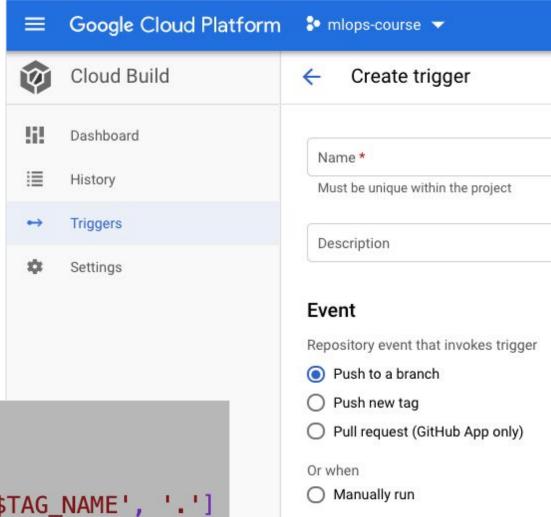


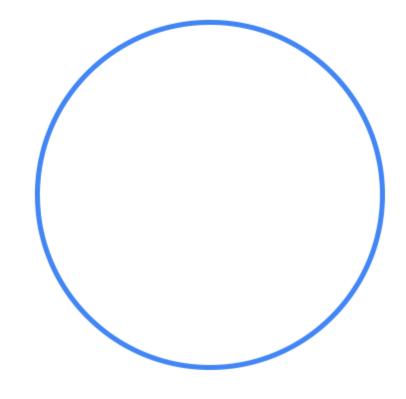
For CI/CD, use a GitHub trigger to rebuild ML artifacts

Connect the GitHub repository to your Google Cloud account, and then trigger a Cloud Build from a GitHub trigger.

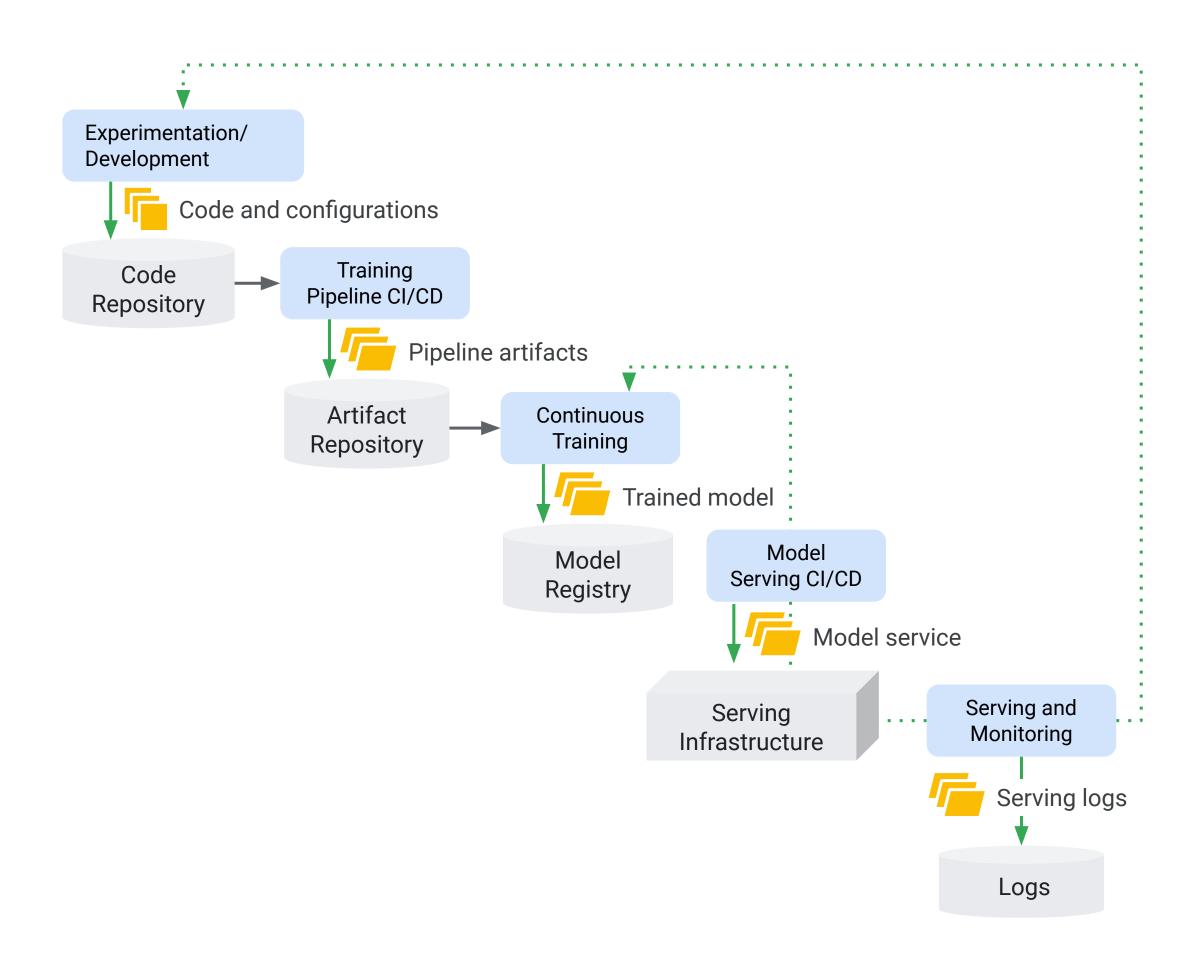
Do this for every container in your ML pipeline.

```
steps:
   # Build the trainer image
   - name: 'gcr.io/cloud-builders/docker'
     args: ['build', '-t', 'gcr.io/$PROJECT_ID/$_TRAINER_IMAGE_NAME:$TAG_NAME', '.']
     dir: $_PIPELINE_FOLDER/trainer_image
24
25
   # Build the base image for lightweight components
   - name: 'gcr.io/cloud-builders/docker'
     args: ['build', '-t', 'gcr.io/$PROJECT_ID/$_BASE_IMAGE_NAME:$TAG_NAME', '.']
     dir: $_PIPELINE_FOLDER/base_image
30
   # Compile the pipeline
   - name: 'gcr.io/$PROJECT_ID/kfp-cli'
33
     args:
     - '-c'
34
```





Where we are headed next



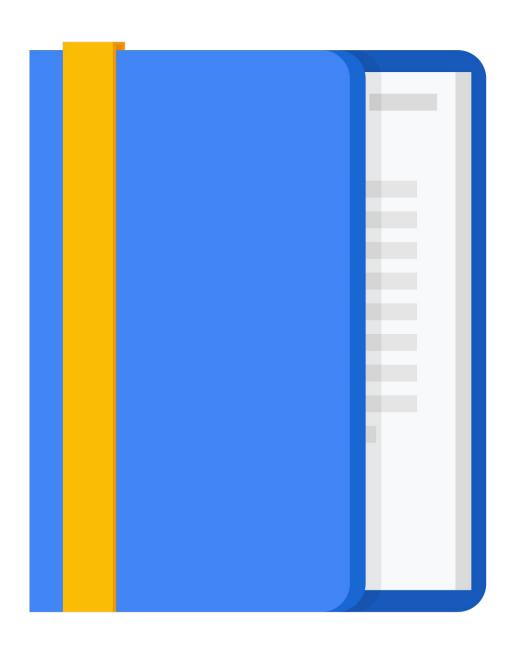
Agenda

Concept Overview

Cloud Build Builders

Cloud Build Configuration

Cloud Build Triggers



What are cloud builders?

Cloud configuration/provisioning actions that are packaged as Docker containers

Typical cloud builder actions:

- Building a Docker image from a Dockerfile
- Pushing a Docker image into a Google Cloud project registry
- Deploying a VM instance on Compute Engine
- Uploading a Kubeflow pipeline on CAIP Pipelines

What are cloud builders?

Standard

Already packaged config actions

Docker Registry:

gcr.io/cloud-builders/

Container Code:

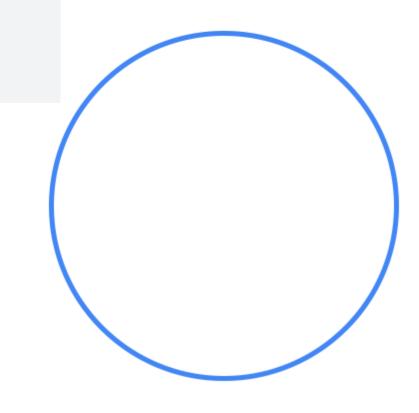
GoogleCloudPlatform/cloud-builders

Custom

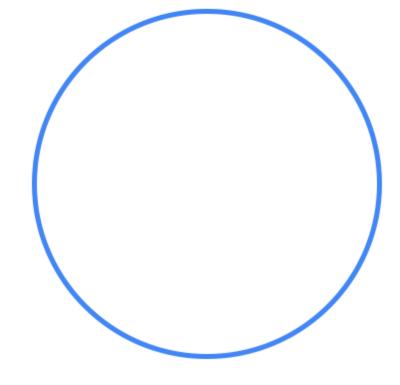
config actions packaged by you

Docker Registry:

gcr.io/<YOUR_PROJECT>/



Standard Cloud Builders



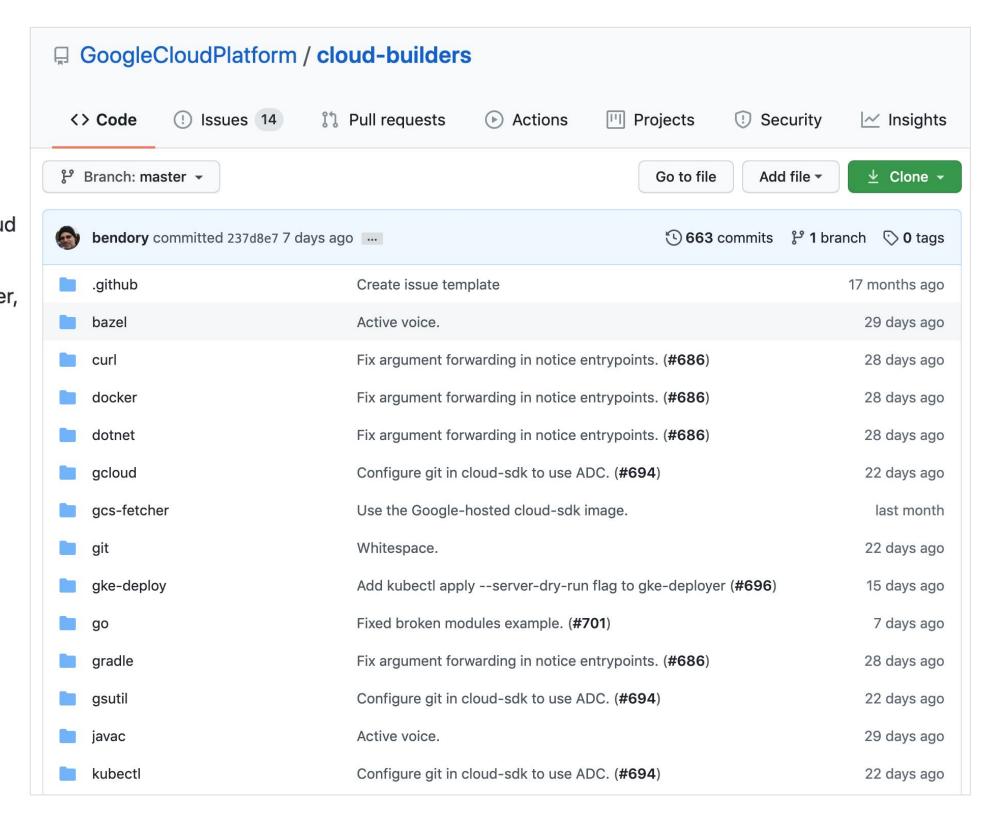
Standard cloud builders

Builder	Name	Example
bazel	gcr.io/cloud-builders/bazel	bazel example
docker	gcr.io/cloud-builders/docker	docker example
git	gcr.io/cloud-builders/git	git example
gcloud	gcr.io/cloud-builders/gcloud	gcloud example
gke-deploy	gcr.io/cloud-builders/gke-deploy	gke-deploy example

...wrap standard config tools as Docker containers

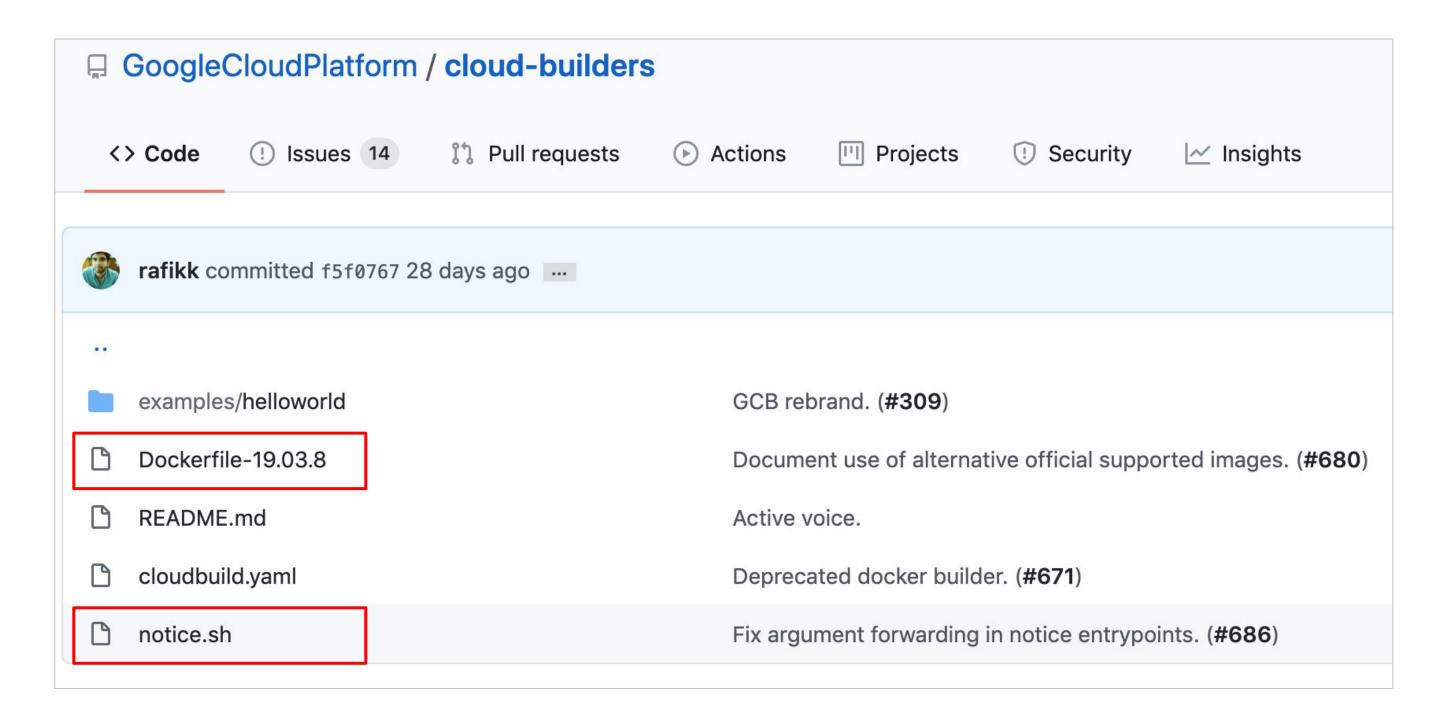
bazel: runs the bazel tool curl: runs the curl tool docker: runs the docker tool dotnet: run the dotnet tool gcloud: runs the gcloud tool gcs-fetcher: efficiently fetches objects from Google Cloud git: runs the git tool gke-deploy: deploys an application to a Kubernetes cluster, go: runs the go tool gradle: runs the gradle tool gsutil: runs the gsutil tool javac: runs the javac tool kubectl: runs the kubectl tool mvn: runs the mayen tool npm: runs the npm tool wget: runs the wget tool

yarn: runs the yarn tool



https://github.com/GoogleCloudPlatform/cloud-builders/tree/master/docker

Example: Docker cloud builder



```
FROM launcher.gcr.io/google/ubuntu16_04
ARG DOCKER_VERSION=5:19.03.8~3-0~ubuntu-xenial
RUN apt-get -y update && \
    apt-get -y install \
        apt-transport-https \
        ca-certificates \
        curl \
        make \
        software-properties-common && \
    curl -fsSL https://download.docker.com/linux/ubuntu/gpg | apt-key add - && \
    apt-key fingerprint 0EBFCD88 && \
    add-apt-repository \
       "deb [arch=amd64] https://download.docker.com/linux/ubuntu \
       xenial \
       edge" && \
    apt-get -y update && \
    apt-get -y install docker-ce=${DOCKER_VERSION} docker-ce-cli=${DOCKER_VERSION}
COPY notice.sh /usr/bin
```

ENTRYPOINT ["/usr/bin/notice.sh"]

When ran, /usr/bin/notice.sh is executed

Example: Docker cloud builder

notice.sh

```
#!/bin/sh
echo '
```

***** NOTICE ****

Alternative official `docker` images, including multiple versions across multiple platforms, are maintained by the Docker Team. For details, please visit https://hub.docker.com/_/docker.

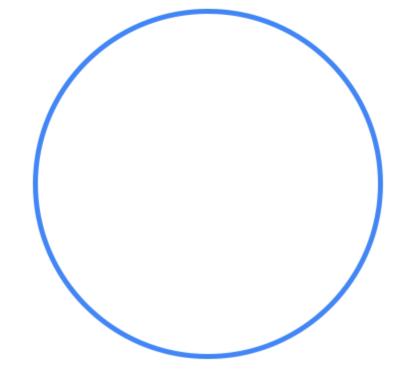
**** END OF NOTICE ****

1

/usr/bin/docker "\$@"

When run, the container passes its args to the docker command.

Custom Cloud Builders



Custom Cloud Build builder

Your custom Docker file

Dockerfile

FROM gcr.io/deeplearning-platform-release/base-cpu

RUN pip install kfp==0.2.5

ENTRYPOINT ["/bin/bash"]

IMAGE_URI='gcr.io/PROJECT_ID/IMAGE_NAME:TAG'

Your project registry

gcloud builds submit --timeout 15m --tag IMAGE_URI PATH ←

Path to the directory where Docker package code is located

Agenda

Concept Overview

Cloud Build Builders

Cloud Build Configuration

Cloud Build Triggers



Cloud Build configuration file

cloudbuild.yaml

Describe the cloud builders to be run step by step

```
steps:
- name: 'gcr.io/cloud-builders/docker' 
args: ['build', '-t', 'gcr.io/$PROJECT_ID/$_TRAINER_IMAGE_NAME:$TAG_NAME', '.']
dir: $_PIPELINE_FOLDER/trainer_image
- name: 'gcr.io/cloud-builders/docker' 
args: ['build', '-t', 'gcr.io/$PROJECT_ID/$_BASE_IMAGE_NAME:$TAG_NAME', '.']
dir: $_PIPELINE_FOLDER/base_image
```

Running Cloud Build

Path to the build code directory

Cloud Build config file describing the cloud builders to be run

Cloud Build config file describing the cloud builders to be run

Cloud Build config file describing the cloud config file

A simple Cloud Build step

steps:

The cloud builder container URI to be run

```
- name: 'gcr.io/cloud-builders/docker' ←
args: ['build', '-t', 'gcr.io/$PROJECT_ID/$_TRAINER_IMAGE_NAME:$TAG_NAME', '.']
dir: $_PIPELINE_FOLDER/trainer_image
```

Substitution

The CWD in the Docker container from which the entrypoint is executed

The arguments to be passed to the container entrypoint

Persistence dir:

```
steps:
- name:
   args:
   dir: <rel path>
```

A cloud builder container has its current working directory:

- Set by default to /workspace in the container.
- Shared between steps.

This can be modified by specifying a dir path that:

- Will resolve to /workspace/<path>.
- Will still be shared between steps if the path is relative.

Caution: If the path is absolute, the CWD will not persist between steps.

Substitutions: \$_VARIABLE_NAME

```
steps:
- name: 'gcr.io/cloud-builders/docker'
   args: ['build', '-t', 'gcr.io/$PROJECT_ID/$_TRAINER_IMAGE_NAME:$TAG_NAME', '.']
   dir: $_PIPELINE_FOLDER/$_IMAGE_NAME

gcloud builds submit . --config cloudbuild.yaml \
   --substitutions '_PIPELINE_FOLDER=.,_IMAGE_NAME=trainer_base'
```

Custom cloud builder

```
# Upload the pipeline
- name: 'gcr.io/$PROJECT_ID/kfp-cli'
  args:
    - '-c'
        kfp --endpoint $_ENDPOINT pipeline upload \
            -p ${_PIPELINE_NAME}_$TAG_NAME $_PIPELINE_PACKAGE
  dir: $_PIPELINE_FOLDER/pipeline
```

No different from the standard cloud builders (but the registry is your own project registry)

Passing environment variables

```
- name: 'gcr.io/$PROJECT_ID/kfp-cli'
                                                    This script takes its input
 args:
                                                    from env variables
    - '-c'
    - 'dsl-compile --py $_PIPELINE_DSL --output $_PIPELINE_PACKAGE'
 env:
    - 'BASE_IMAGE=gcr.io/$PROJECT_ID/$_BASE_IMAGE_NAME:$TAG_NAME'
    - 'TRAINER_IMAGE=gcr.io/$PROJECT_ID/$_TRAINER_IMAGE_NAME:$TAG_NAME'
    - 'RUNTIME_VERSION=$_RUNTIME_VERSION'
    - 'PYTHON_VERSION=$_PYTHON_VERSION'
    - 'COMPONENT_URL_SEARCH_PREFIX=$_COMPONENT_URL_SEARCH_PREFIX'
    - 'USE_KFP_SA=$_USE_KFP_SA'
                                                     That's how the env variables
 dir: $_PIPELINE_FOLDER/pipeline
                                                     are passed to the script
```

Pushing images to Container Registry

Build the image locally on the build node steps: - name: 'gcr.io/cloud-builders/docker' args: ['build', '-t', 'gcr.io/\$PROJECT_ID/\$_TRAINER_IMAGE_NAME:\$TAG_NAME', '.'] dir: \$_PIPELINE_FOLDER/trainer_image Images: 'gcr.io/\$PROJECT_ID/\$_TRAINER_IMAGE_NAME:\$TAG_NAME' 'gcr.io/\$PROJECT_ID/\$_BASE_IMAGE_NAME:\$TAG_NAME' Push the image to the registry

Agenda

Concept Overview

Cloud Build Builders

Cloud Build Configuration

Cloud Build Triggers



Manually executing a Cloud Build

- 1. Clone the Github repo with your cloudbuild.yaml and ML code on a build node.
 - The build node can be any node with the gcloud sdk properly authenticated:
 - JupyterLab VM
 - Cloud Shell (if the containers to build are small)
 - Your laptop
 - A dedicated build VM on Cloud Compute
- 2. Run gcloud builds submit on the cloudbuild.yaml with the proper substitutions.

Manually executing a Cloud Build

```
SUBSTITUTIONS="""
In [ ]:
         2 ENDPOINT={},\
         3 TRAINER IMAGE NAME=trainer image,\
           BASE IMAGE NAME=base image, \
         5 TAG NAME=test,\
         6 PIPELINE FOLDER=.,\
         7 PIPELINE_DSL=covertype_training_pipeline.py,\
         8 PIPELINE PACKAGE=covertype training pipeline.yaml, \
         9 PIPELINE NAME=covertype continuous training,\
        10 RUNTIME VERSION=1.15,\
        11 PYTHON VERSION=3.7,\
        12 USE KFP SA=True,\
        13 COMPONENT URL SEARCH PREFIX=https://raw.githubusercontent.com/kubeflow/pipelines/0.2.5/components/gcp/
        14 """.format(ENDPOINT).strip()
         1 !gcloud builds submit . --config cloudbuild.yaml --substitutions {SUBSTITUTIONS}
In [ ]:
```

CI/CD: Automated Cloud Build triggers

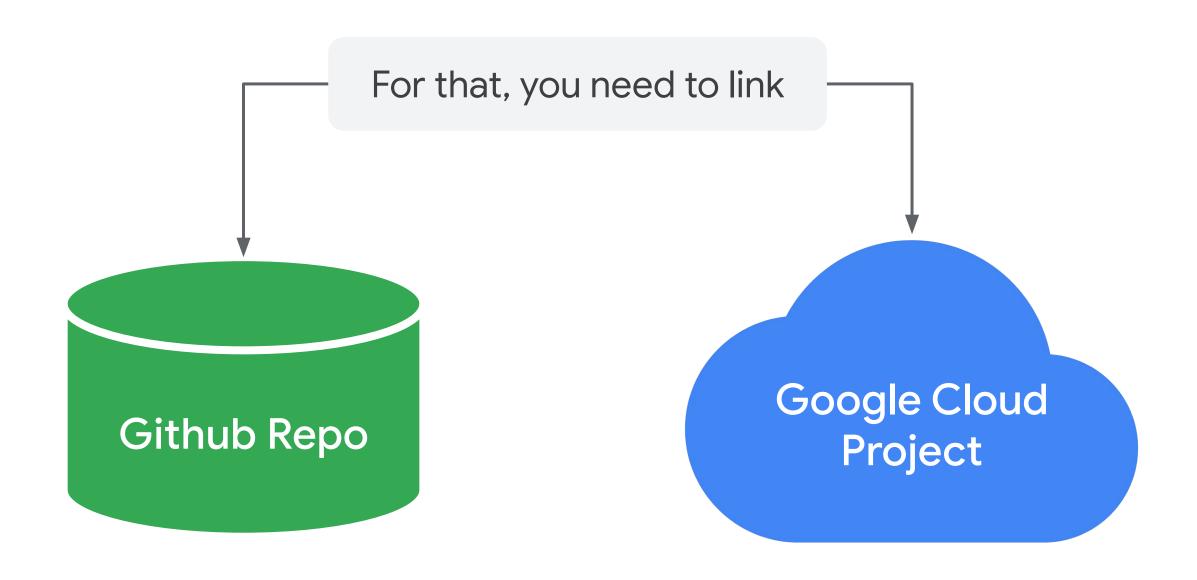
When the ML code changes in Github via:

- Push on a branch
- New tag
- Pull request
- etc.

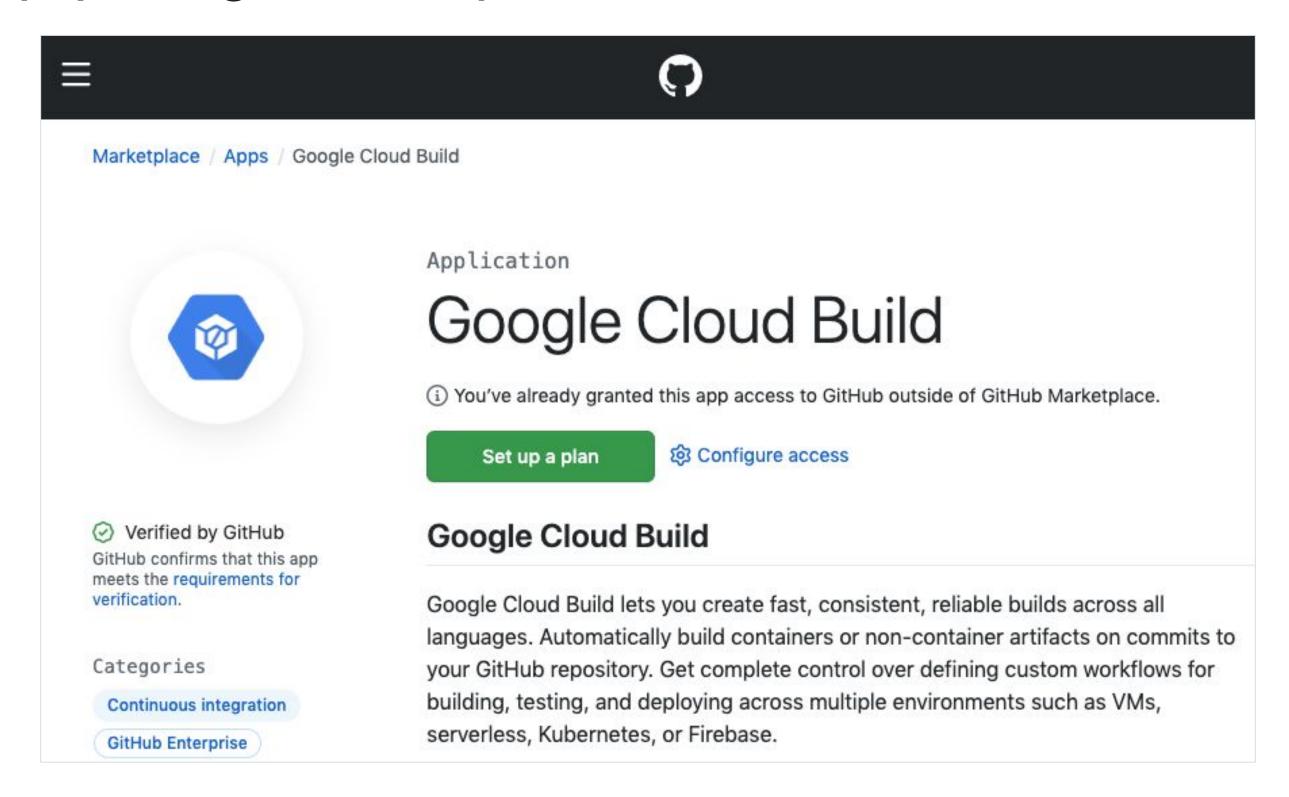
...you want a Cloud Build to be triggered automatically, so that the training artifacts are updated from the new code:

- Training containers
- Kubeflow pipelines
- etc.

CI/CD: Automated Cloud Build triggers



Set up your github repo to work with Cloud Build



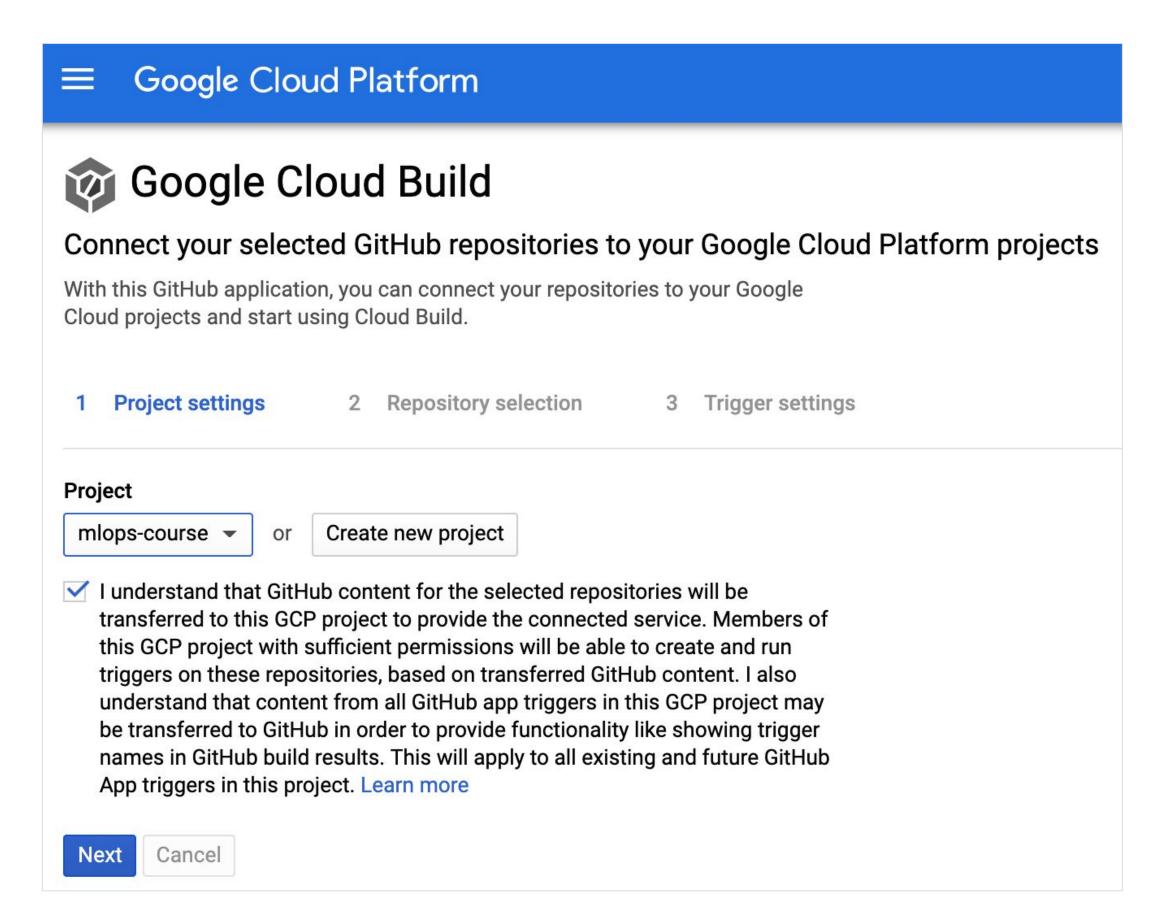
Allow your repo to be accessed by Google Cloud Build

Repository access

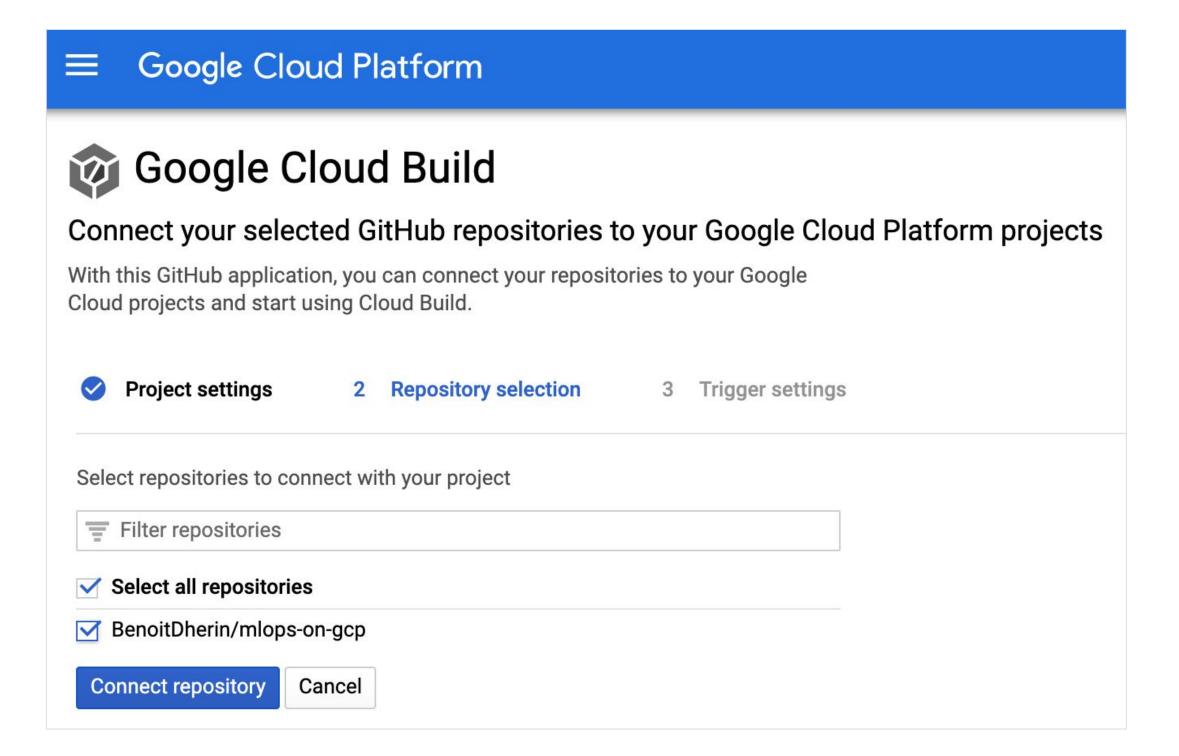
Save

Cancel

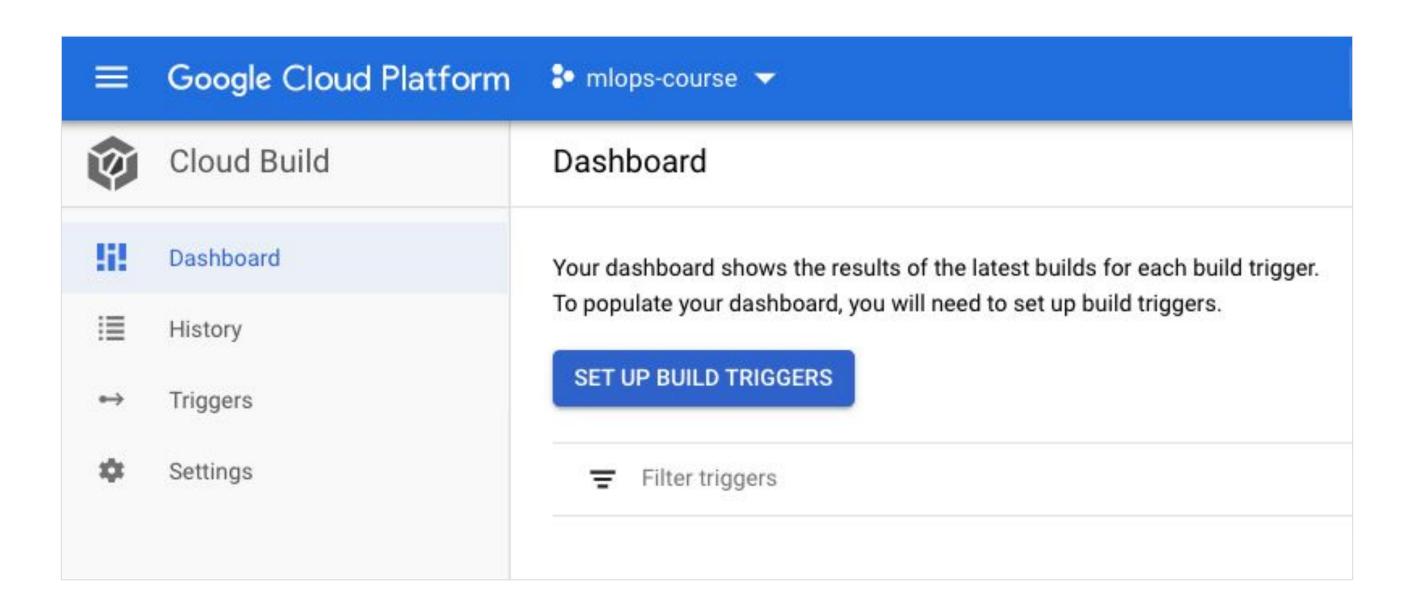
Add your repo to Cloud Build



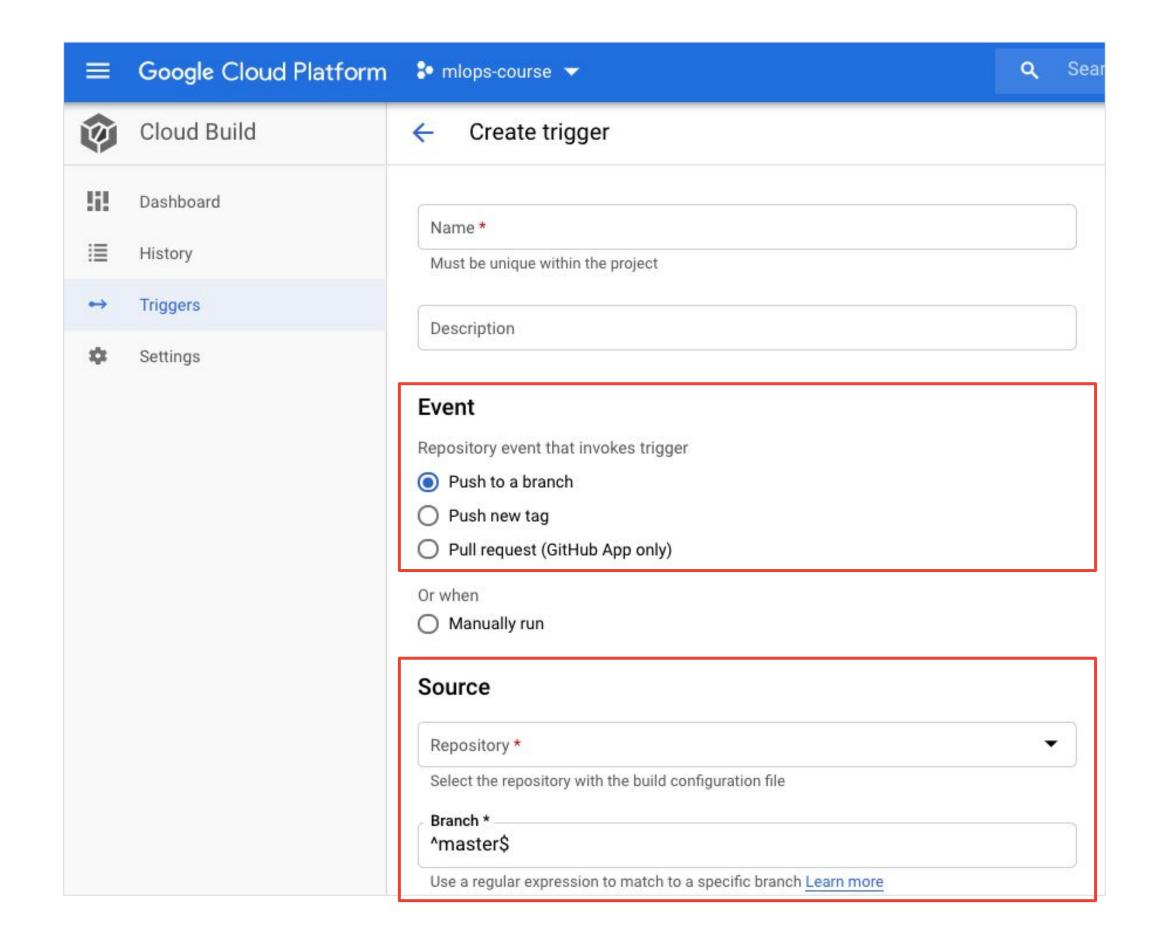
Add your repo to Cloud Build



Set up triggers



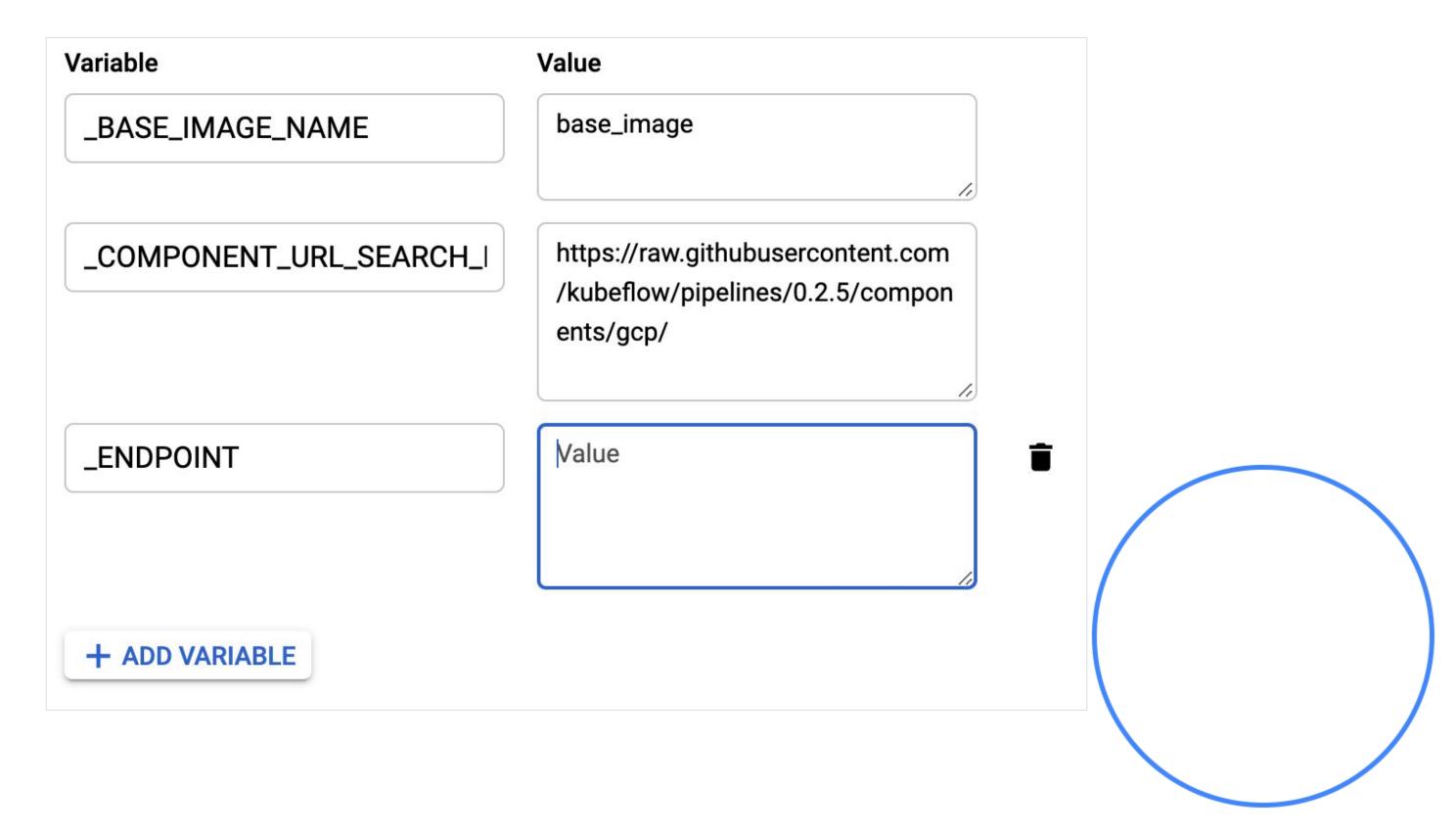
Cloud triggers



Specify the location of the cloudbuild.yaml file

Field	Value
Name	[YOUR TRIGGER NAME]
Description	[YOUR TRIGGER DESCRIPTION]
Event	Tag
Source	[YOUR FORK]
Tag (regex)	*
Build Configuration	Cloud Build configuration file (yaml or json)
Cloud Build configuration file location	/ workshops/kfp-caip-sklearn/lab-03-kfp-cicd/cloudbuild.yaml

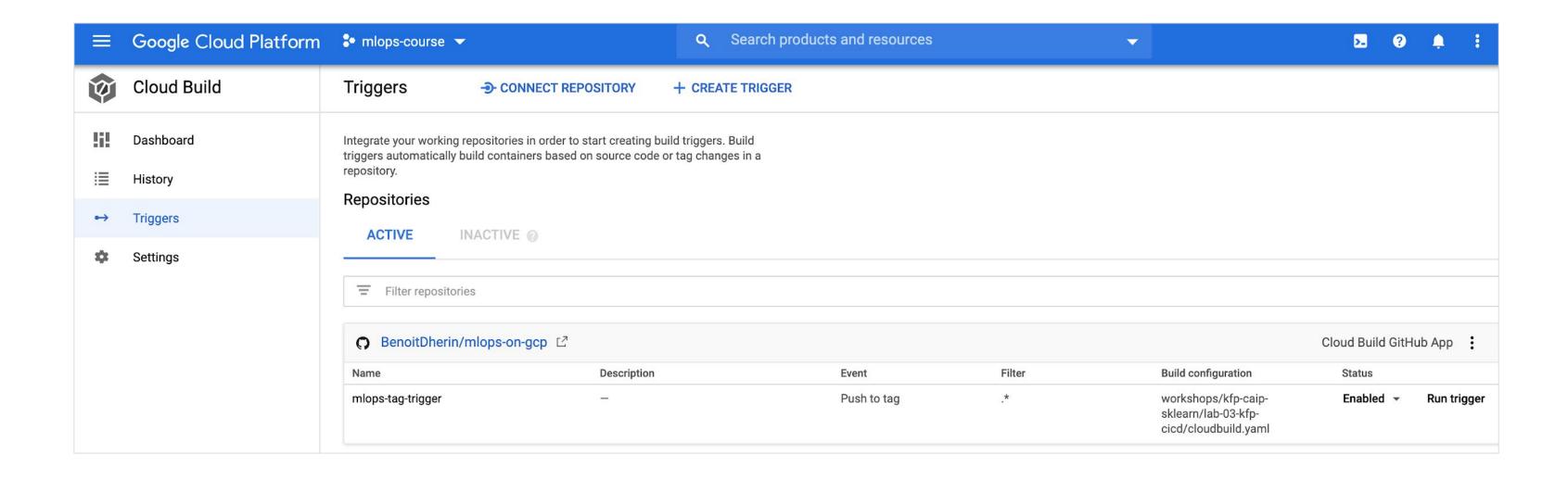
Set up the substitution variable values



Lab example

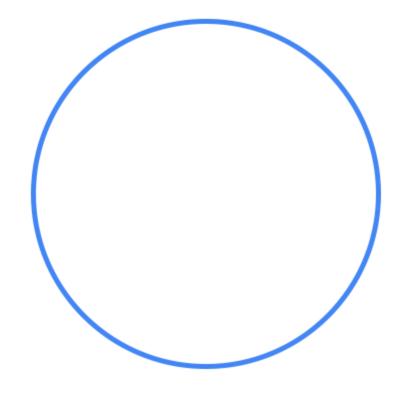
Variable	Value
_BASE_IMAGE_NAME	base_image
_COMPONENT_URL_SEARCH_PREFIX	https://raw.githubusercontent.com/kubeflow/pipelines/0.2.5/components/gcp/
_ENDPOINT	[Your inverting proxy host]
_PIPELINE_DSL	covertype_training_pipeline.py
_PIPELINE_FOLDER	workshops/kfp-caip-sklearn/lab-03-kfp-cicd
_PIPELINE_NAME	covertype_training_deployment
_PIPELINE_PACKAGE	covertype_training_pipeline.yaml
_PYTHON_VERSION	3.7
_RUNTIME_VERSION	1.15
_TRAINER_IMAGE_NAME	trainer_image
_USE_KFP_SA	False

A Cloud Build trigger is now listed

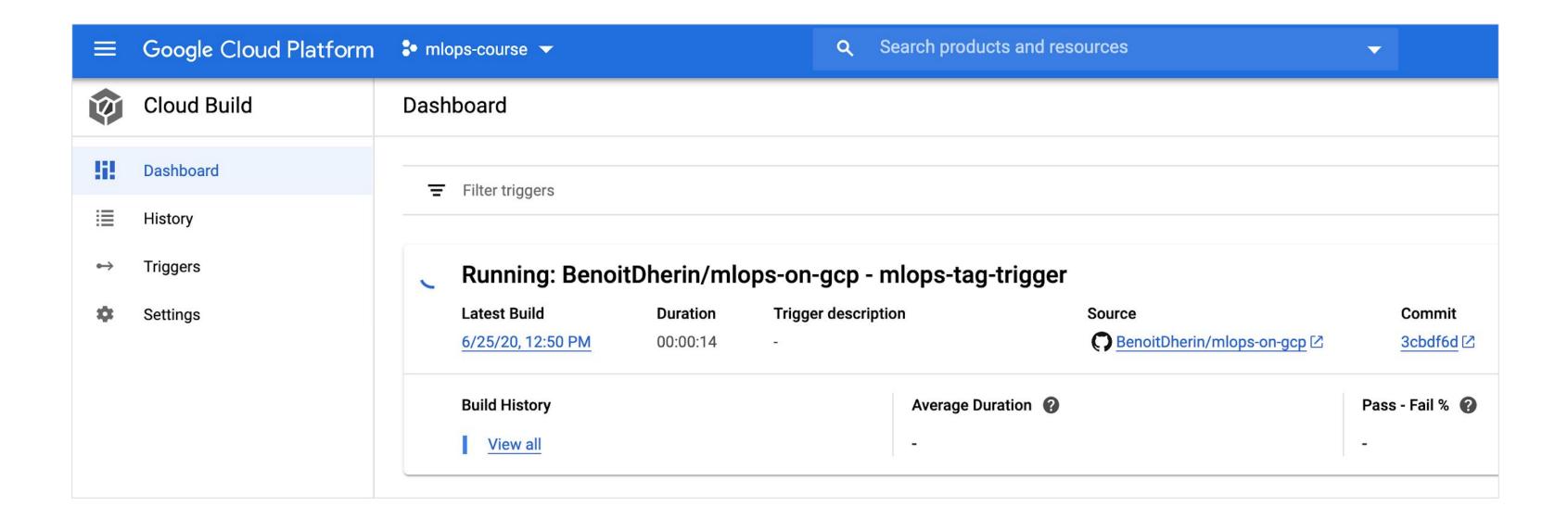


Pushing a new tag will now trigger a Cloud Build

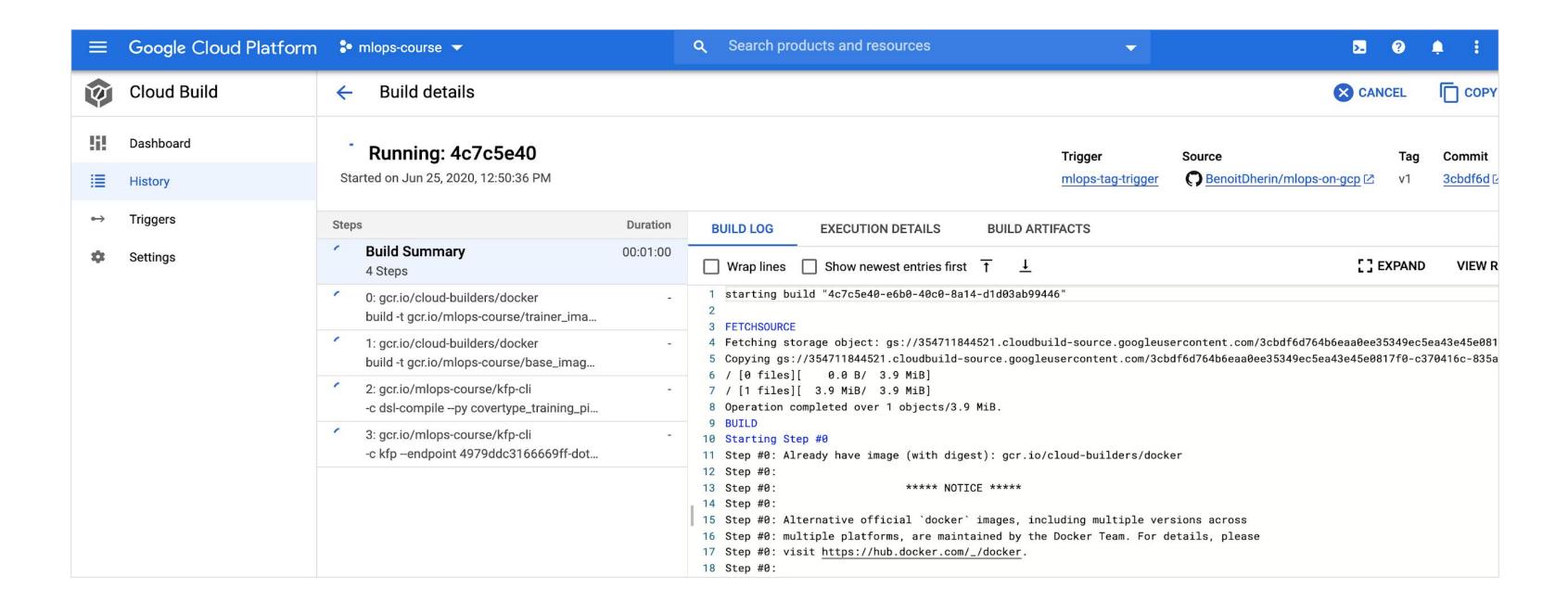
```
git tag [TAG NAME]
git push origin --tags
```



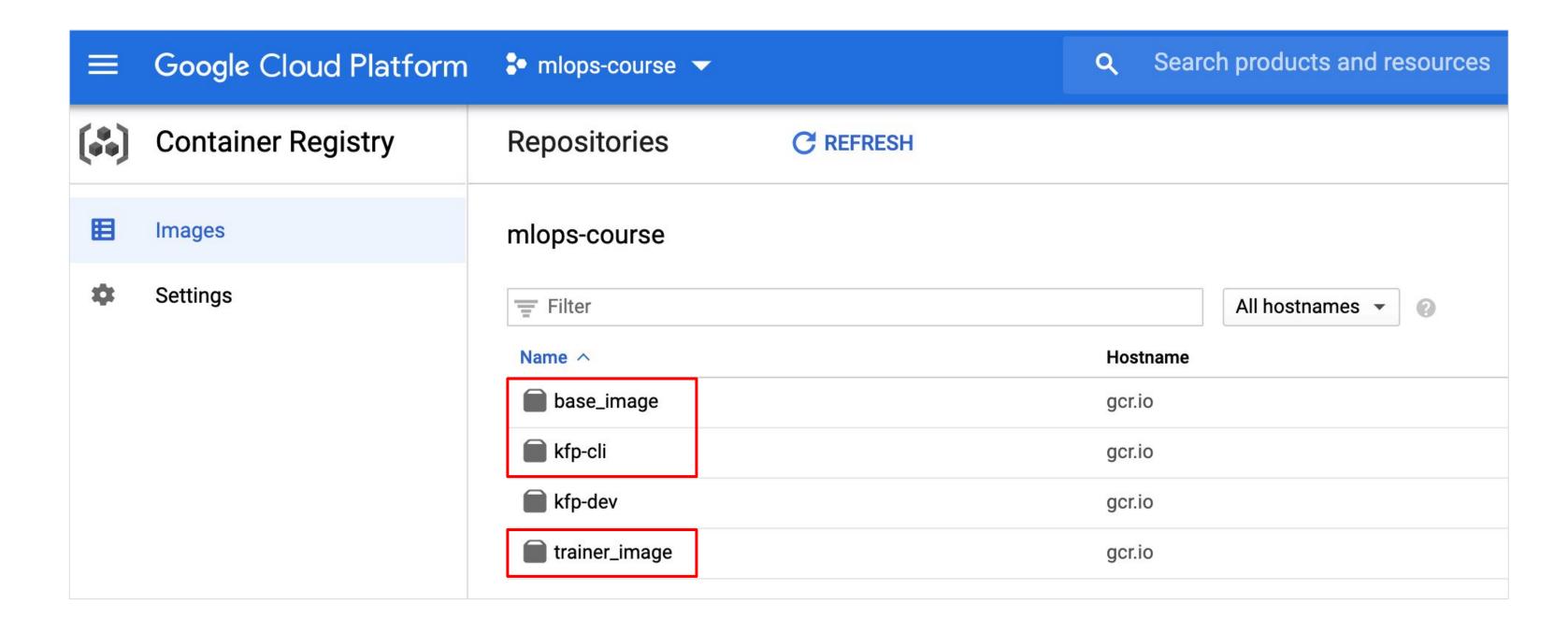
The build can be monitored



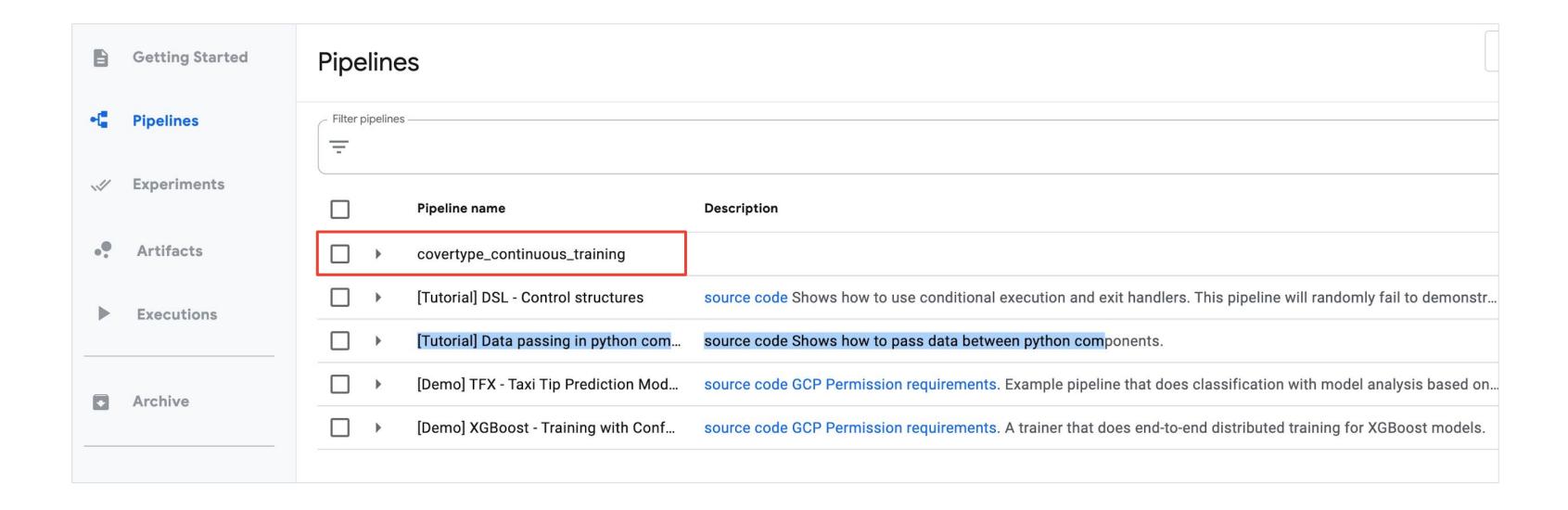
The build can be monitored



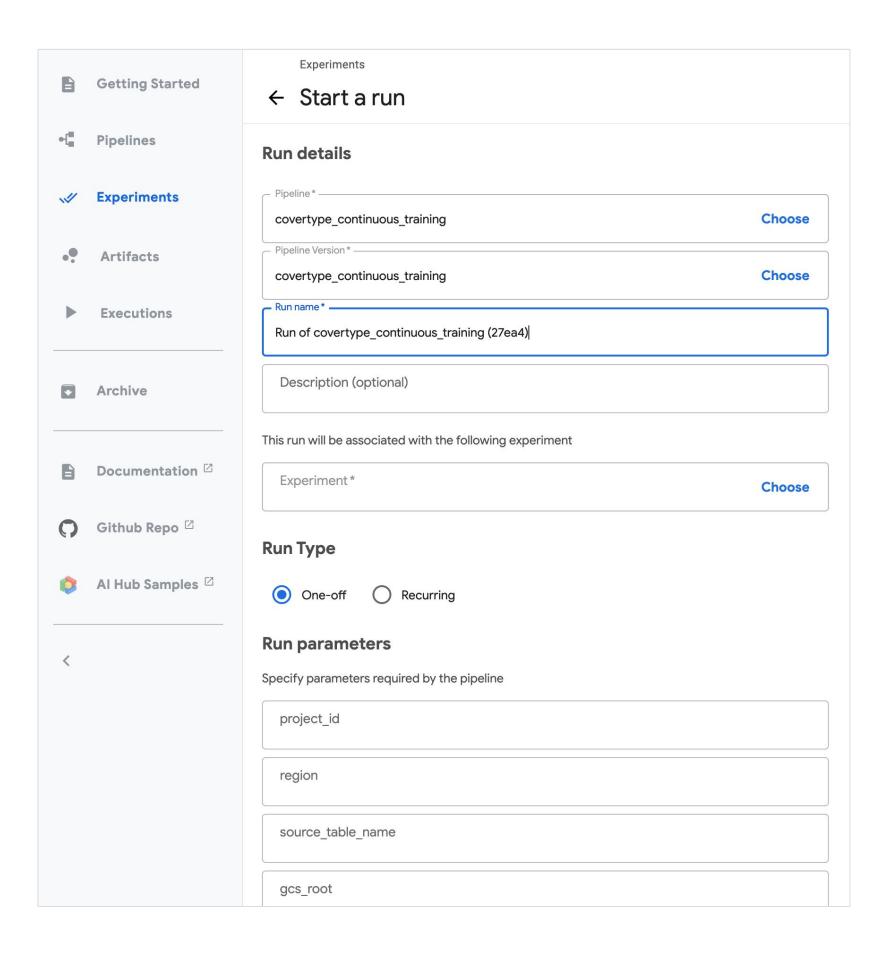
After the build, new artifacts are available



After the build, new artifacts are available



Now the new pipeline is ready to be run



Lab

CI/CD for a KFP Pipeline

In this lab you will walk through authoring a Cloud Build CI/CD workflow that automatically builds and deploys a KFP pipeline.

https://github.com/GoogleCloudPlatform/mlops-on-gcp/blob/master/workshops/kfp-caip-sklearn/lab-03-kfp-cicd/exercises/lab-03.ipynb

cloud.google.com