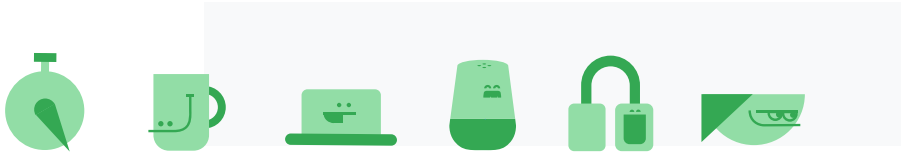
Study

Cloud Study Jams

Baseline: Data, ML, AI

Agenda:

6:00 pm	Registration / Food / Network
6:30 pm	Welcome
6:35 pm	Get free access to Qwiklabs
6:50 pm	Lab 1 - Cloud ML Engine Qwik Start
7:50 pm	End of Lab 1 - Lessons Learnt
8:30 pm	Event ends - THANK YOU!



Study

Upcoming GDG Events

WTM 2019 - April 13th
- Tickets over @GDGDublin meetup

Call for speakers IO Extended 2019
- Info @GDGDublin

Other events

CoderDojo's Coolest Projects International is looking for

Volunteers

When : May 5th (Bank Holiday Weekend)
Where : RDS

Time: 08.30 am to 16.00 pm

Social Media

@gdgclouddublin
#cloudstudyjams #gdgclouddublin

meetup.com/GDG-Cloud-Dublin

gdgclouddublin.slack.com

Together, as a team

You are encourage to participate in these channels and finish the labs at home.

slack

#ml-studyjams

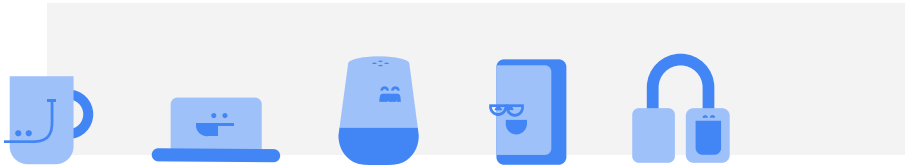
@gdgclouddublin

#CloudStudyJam
#GDGCloudDublin

Study

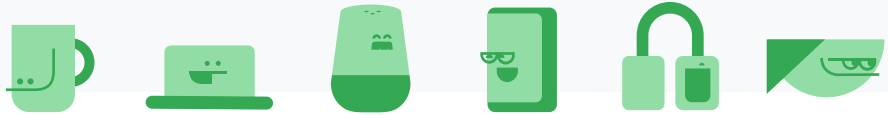
What is a Study Jam?

- Study Jams are community-run study groups for developers.
- Great opportunity for us to to learn something together
- Trainings for personal development or career advancement.



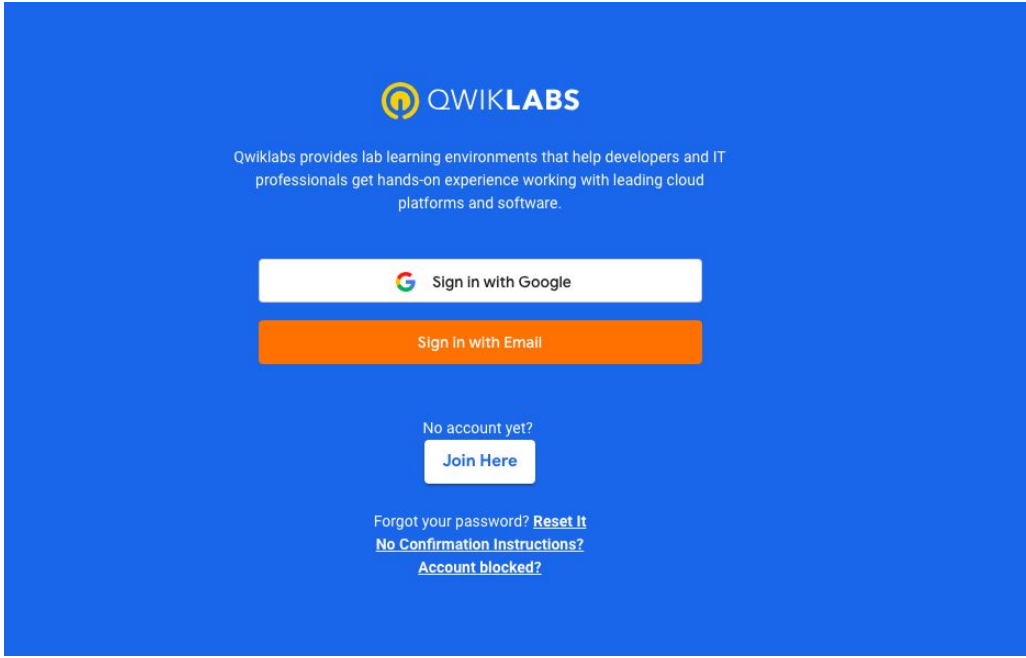
Study

Setup Qwiklabs



Study

Step 1: Make sure to sign out of your Qwiklabs account



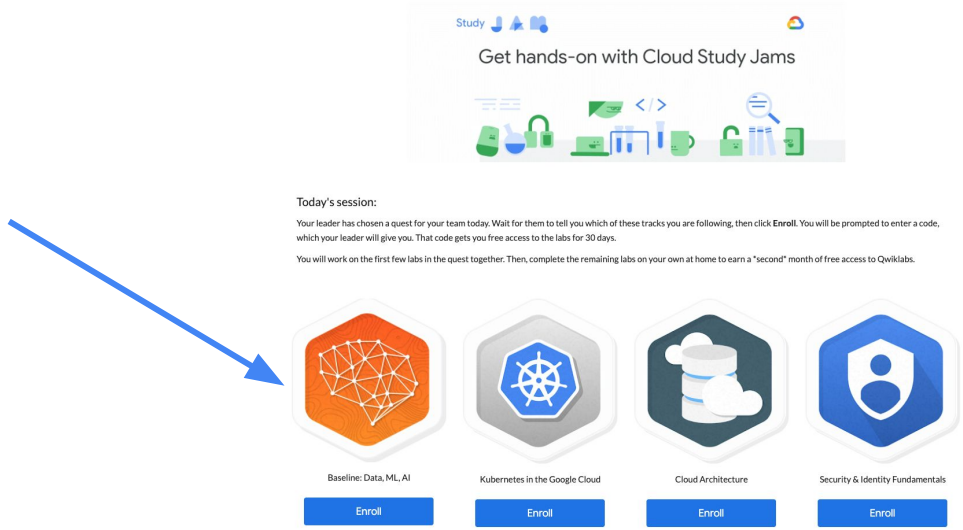
Study

Step 2:

Go to
http://bit.ly/studyjam2019

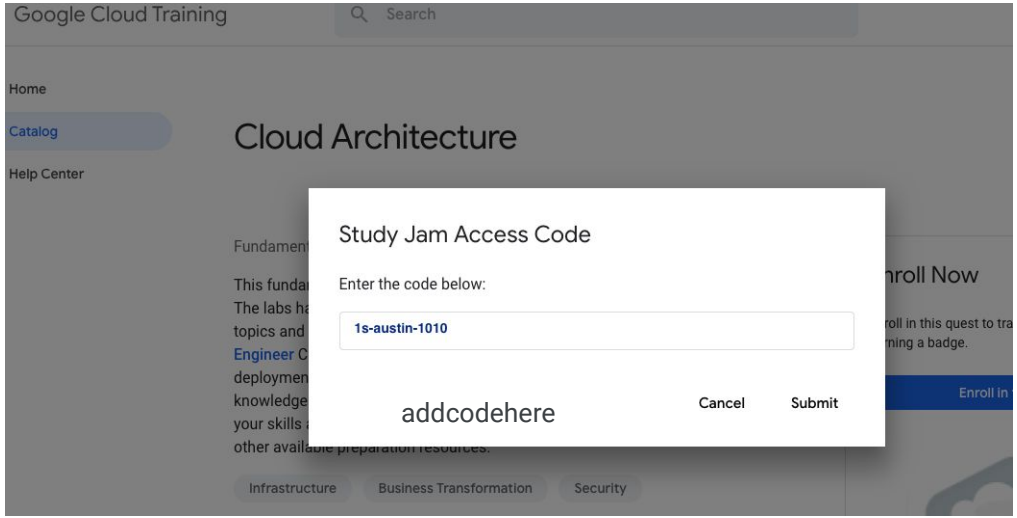
and click “Enroll” on:

Baseline: Data, ML, AI



Study

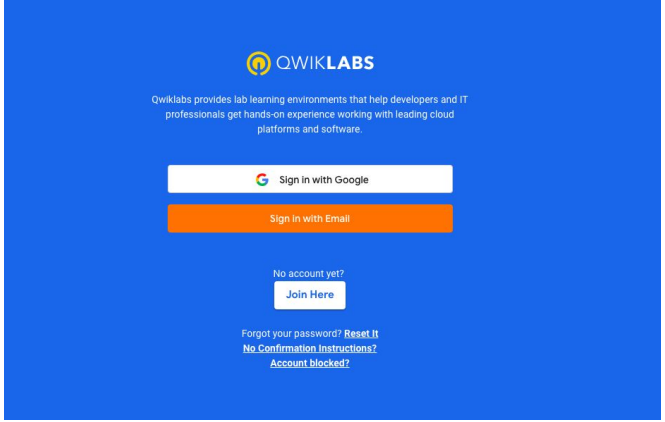
Step 3:
Enter the Study Jam Access Code

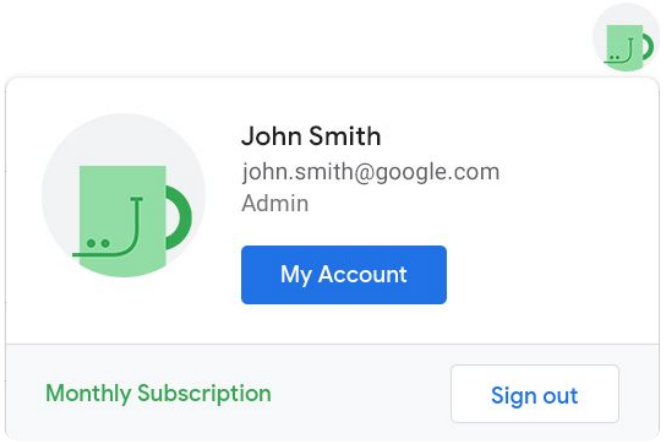


Study

Step 4: Sign into your account

1. Sign in with your Google account or email
2. Hover over avatar on top right and make sure it says **“monthly subscription”.**





Study

Cloud ML Engine Qwik start

Study

Overview

- Hands-on practice with TensorFlow

- What you will learn?
  - Learn how to deploy a model on Cloud ML Engine
  - Train your model to predict income category of a person

Study 

Steps

- Create a TensorFlow training application and validate it locally.
- Run your training job on a single worker instance in the cloud.
- Run your training job as a distributed training job in the cloud.
- Optimize your hyperparameters by using hyperparameter tuning.
- Deploy a model to support prediction.
- Request an online prediction and see the response.
- Request a batch predictions

What you will build

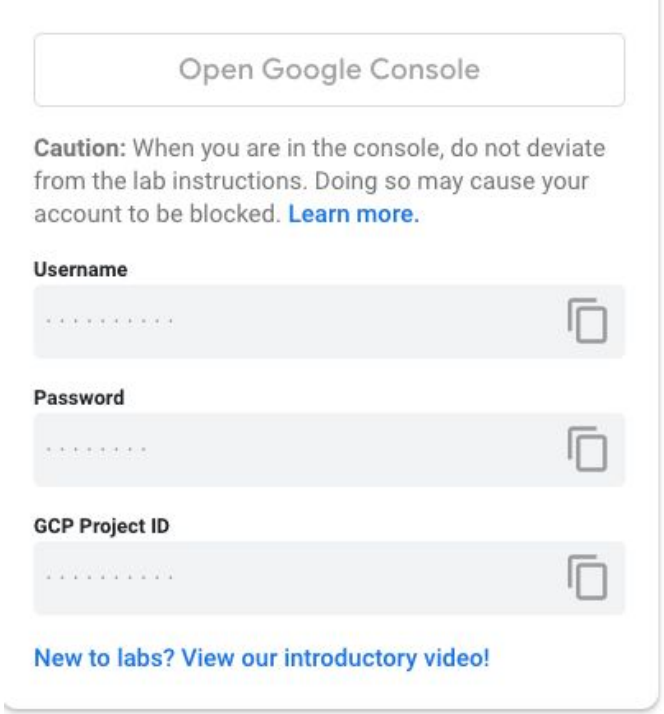
- A deep model for predicting income category
  - >50K — Greater than 50,000 dollars
  - <=50K — Less than or equal to 50,000 dollars

Using :

- Deep neural network
- Linear regression
- **DNNCombinedLinearClassifier** class from Tensorflow

How to use the labs?

1. Labs are timed, you cannot paused them
2. The timer is also the amount of time your cloud resources will be running on
3. Every time you hit start, you will get new credentials in the top left corner



Study

Concepts

- **gsutil** is a Python application that lets you access Cloud Storage from the command line.
- **Google Cloud Storage Buckets** are the basic containers that hold your data. Everything that you store in Cloud Storage must be contained in a bucket.
- **gcloud** command-line interface is a tool that provides the primary CLI to Google Cloud Platform
- **Tensorboard** set of visualization tools to make it easier to understand, debug, and optimize TensorFlow programs

Hints / Snippets

```
1
2   # Use gsutil help <command or topic> for detailed help.
3   gsutil help <command>
5   # gcloud help and any search terms for help
6   gcloud help <search terms>
8   # ml-engine Manage Cloud ML Engine jobs and models.
9   # train run a Cloud ML Engine training job locally
10  gcloud ml-engine local train [options]
```


Hints / Snippets

```
12    # Creating buckets with mb ( Make buckets )
13    gsutil mb gs://<bucket name>

15    # Submit a Cloud Machine Learning training job
16    gcloud ml-engine jobs submit training

18    # Create a new Cloud ML Engine model
19    gcloud ml-engine models create <model name>
```


Dataprep Qwik start

Study

Overview

- Hands-On practice with Google Cloud Dataprep

- What you will learn?
  - Manipulate a dataset
  - Correct, Transform and Join Data

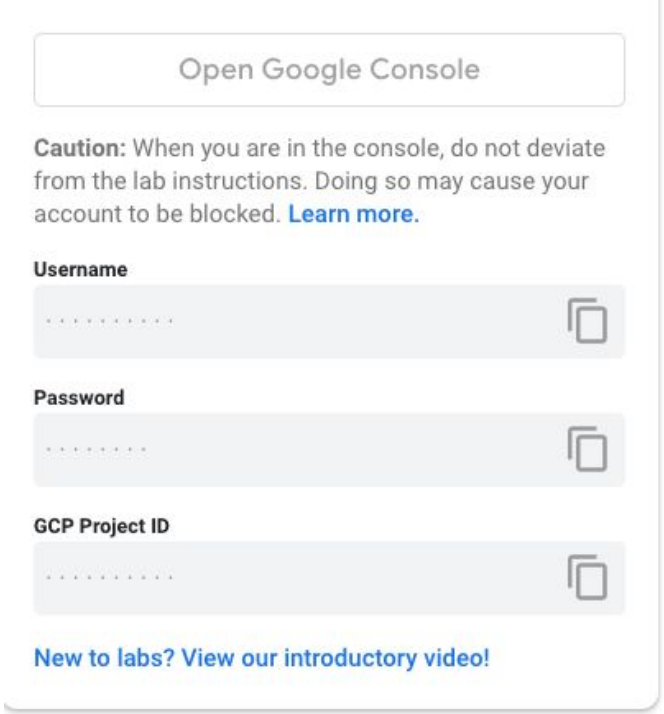
Study

Steps

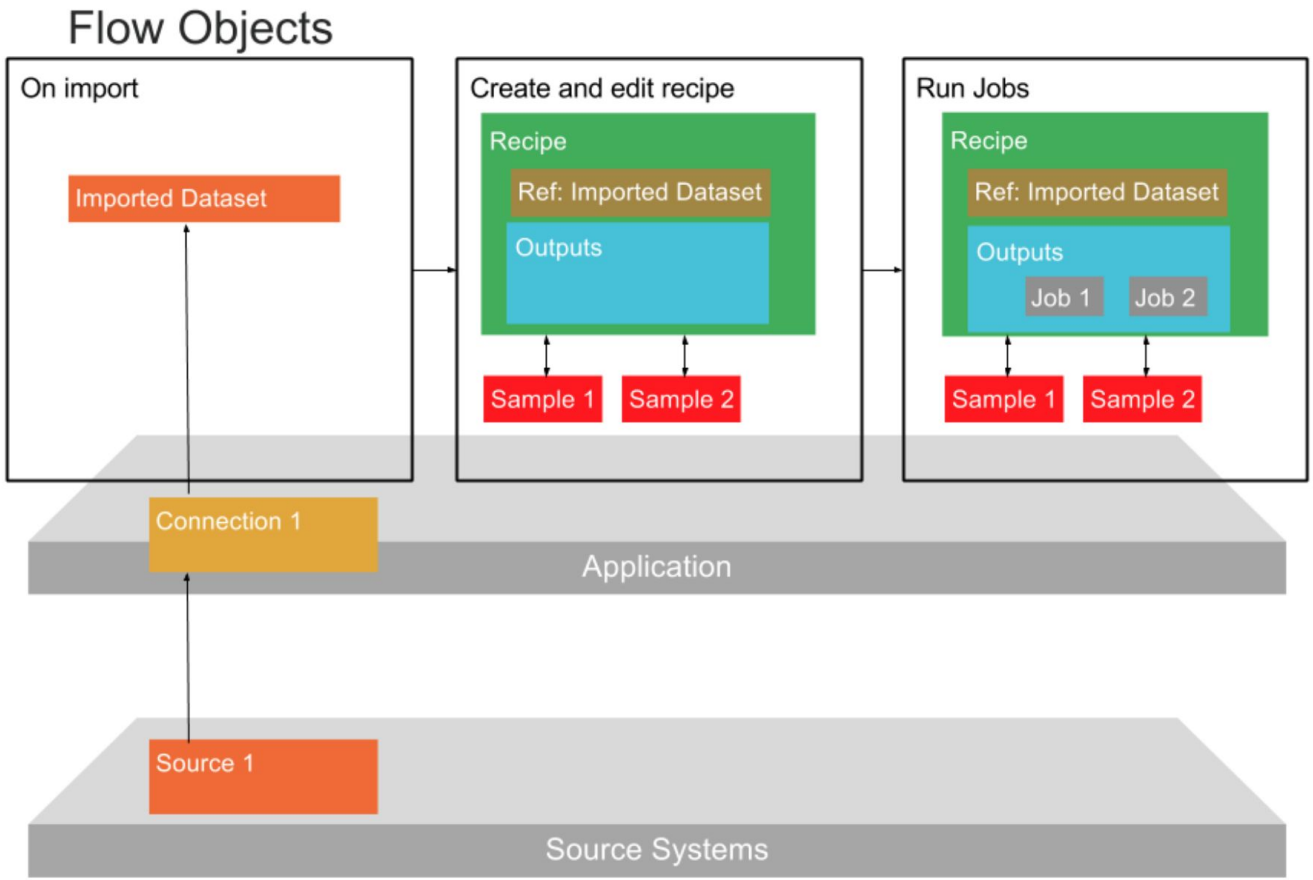
- Create a Google Cloud storage Bucket
- Initialize Cloud Dataprep
- Create a flow
- Import Dataset
- Prep the files
- Summary of Data
- Rename Columns

How to use the labs?

1. Labs are timed, you cannot paused them
2. The timer is also the amount of time your cloud resources will be running on
3. Every time you hit start, you will get new credentials in the top left corner



Study


Flow Objects
On import
Imported Dataset
Create and edit recipe
Recipe
Ref: Imported Dataset
Outputs
Sample 1
Sample 2
Run Jobs
Recipe
Ref: Imported Dataset
Outputs
Job 1
Job 2
Sample 1
Sample 2
Connection 1
Application
Source 1
Source Systems

Concepts

- A **flow** is a container for holding one or more datasets, associated recipes and other objects.
- An **imported dataset** is simply a reference to the original data; the data does not exist within the platform
- A **recipe** is a user-defined sequence of steps that can be applied to transform a dataset.
- **Outputs** contain one or more publishing destinations, which define the output format, location.

Dataflow: Qwik Start - Templates

Study

Overview

- Hands-On practice with Google Cloud Dataflow templates

- What you will learn?
  - Use Cloud Pub/Sub
  - Read messages from a Topic
  - Push messages to BigQuery Table

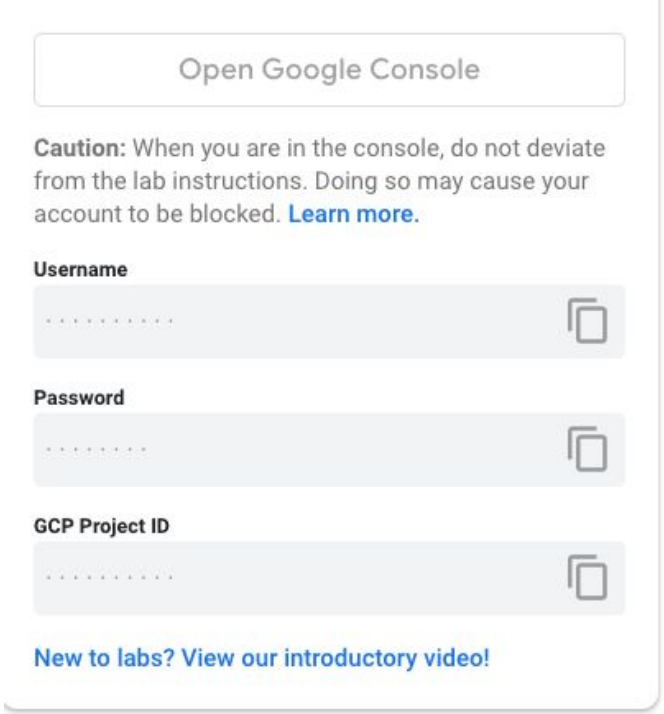
Study

Steps

- Create a BigQuery Dataset and Table using **CloudShell**
- Create a BigQuery Dataset and Table using **GCP Console**
- Run the pipeline
- Submit a query

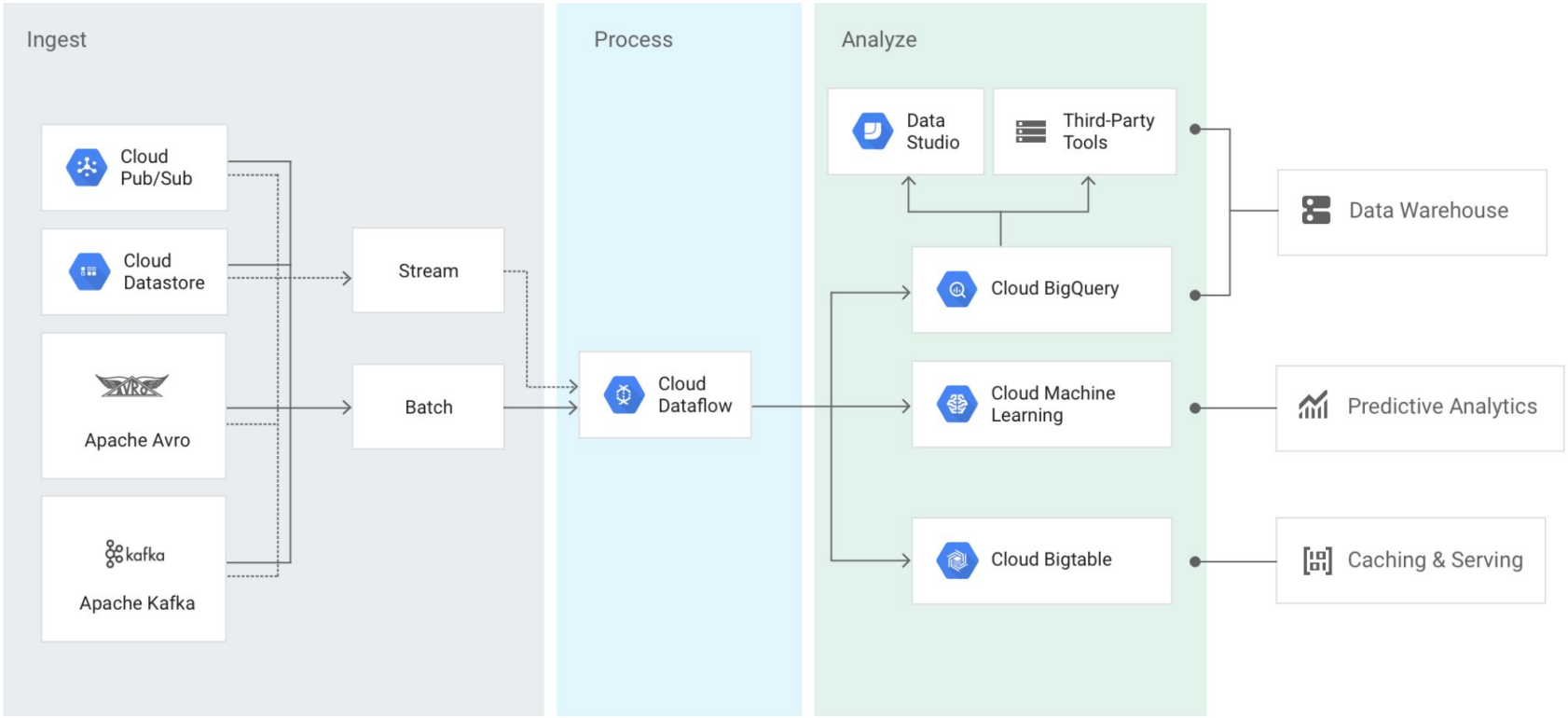
How to use the labs?

1. Labs are timed, you cannot paused them
2. The timer is also the amount of time your cloud resources will be running on
3. Every time you hit start, you will get new credentials in the top left corner

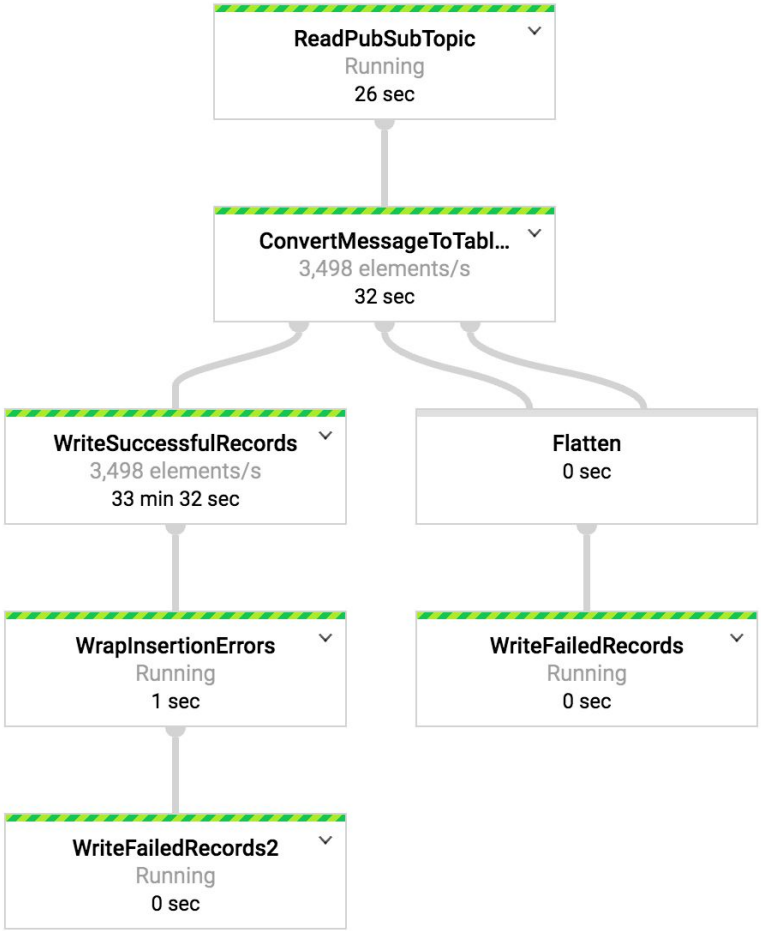


Study

Data Transformation with Cloud Dataflow



Study


ReadPubSubTopic
Running
26 sec
ConvertMessageToTabl...
3,498 elements/s
32 sec
WriteSuccessfulRecords
3,498 elements/s
33 min 32 sec
Flatten
0 sec
WrapInsertionErrors
Running
1 sec
WriteFailedRecords
Running
0 sec
WriteFailedRecords2
Running
0 sec

Study

Concepts

- **Cloud Pub/Sub** is a scalable, durable event ingestion and delivery system that serves as a foundation for modern stream analytics pipelines
- **The Cloud Pub/Sub to BigQuery template** is a streaming pipeline
  a. Reads JSON-formatted messages from a Cloud Pub/Sub **topic**
  b. Converts & write messages to **BigQuery Table**
- **Cloud Dataflow** is a fully-managed service for transforming and enriching data in stream (real time) and batch (historical) modes
- **Pipeline** encapsulates the entire series of computations involved in reading input data, transforming that data, and writing output data

Hints / Snippets

```
21    # Create a dataset, table, view, or transfer
22    # configuration with this name.
23    bq mk <dataset name>
```

```
25    # create a table with schema
26    bq mk --schema <[FIELD]:[DATA_TYPE],[FIELD]:[DATA_TYPE]> -t <dataset name>.<table name>
```


Dataflow: Qwik Start - Python

Study

Overview

- Hands-On practice with Cloud Dataflow Python SDK

- What you will learn?
  - Use Dataflow SDK for Python
  - Run example pipeline using GCP console

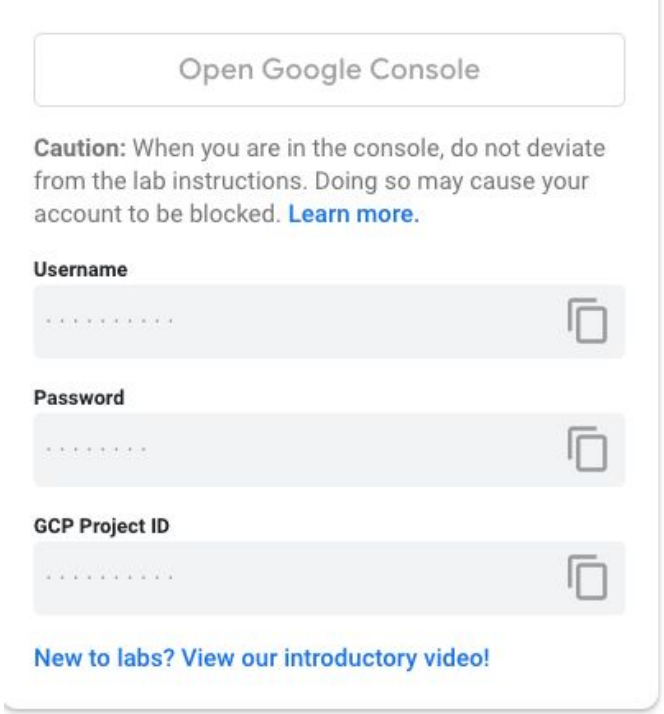
Study

Steps

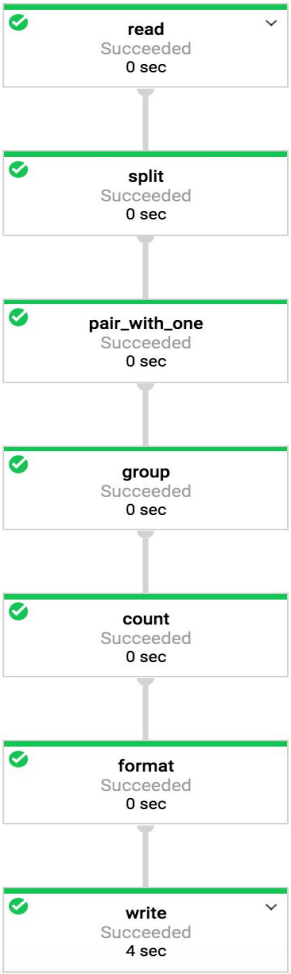
- Create a Cloud Storage Bucket
- Install Cloud Dataflow SDK for Python
- Run a pipeline with Direct Runner
- Run a pipeline Remotely
- Check the job succeeded

How to use the labs?

1. Labs are timed, you cannot paused them
2. The timer is also the amount of time your cloud resources will be running on
3. Every time you hit start, you will get new credentials in the top left corner



Study



Custom counters

Filter by counter name, value or step

Counter name	Value	Step
empty_lines	1,663	split
word_len_dist		split
word_len_dist[COUNT]	28,001	split
word_len_dist[MAX]	15	split
word_len_dist[MEAN]	4	split
word_len_dist[MIN]	1	split
word_lengths	117,723	split
words	28,001	split

Study

Dataproc: Qwik Start - Console

Study

Overview

- Hands-On practice with DataProc Console

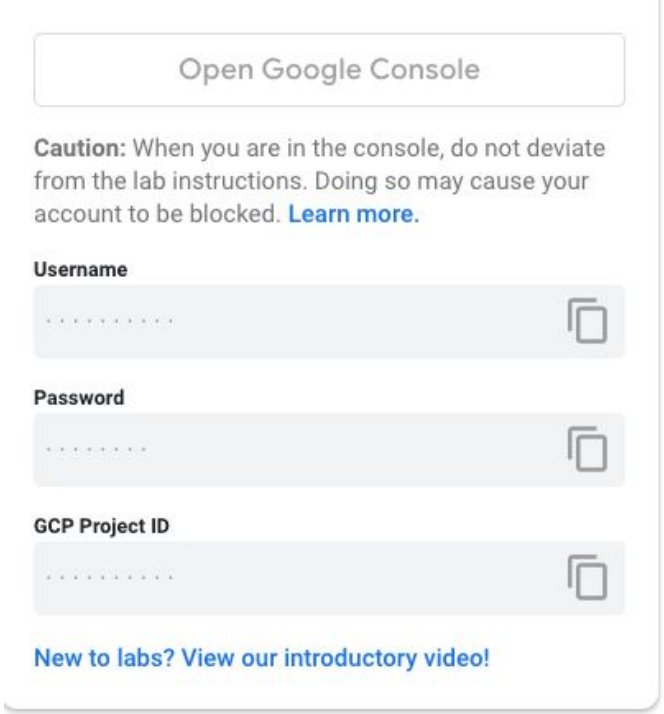
- What you will learn?
  - Create a dataproc cluster
  - Submit a Job to the Cluster
  - View the Job Output

Steps

- Enable dataproc API
- Create a Cluster
- Submit a Job
- Check the output
- Scale up your cluster

How to use the labs?

1. Labs are timed, you cannot paused them
2. The timer is also the amount of time your cloud resources will be running on
3. Every time you hit start, you will get new credentials in the top left corner



Study

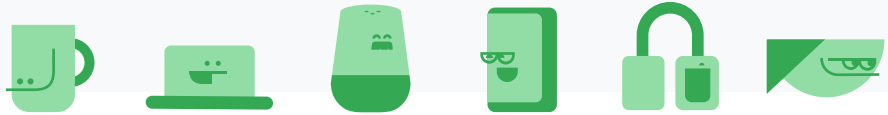
Concepts

- **Cloud DataProc** is a managed Spark and Hadoop service that lets you take advantage of open source data tools for batch processing, querying, streaming, and machine learning
- **Cloud dataproc cluster components:**
  a. Anaconda
  b. Jupyter
  c. Kerberos
  d. Zeppelin
  e. Zookeeper

Hints / Snippets

```
35  # Create a dataproc cluster from CLI with Jupyter enabled
36  gcloud dataproc clusters create cluster-name \
37    --optional-components=JUPYTER \
38    --image-version=1.3
```

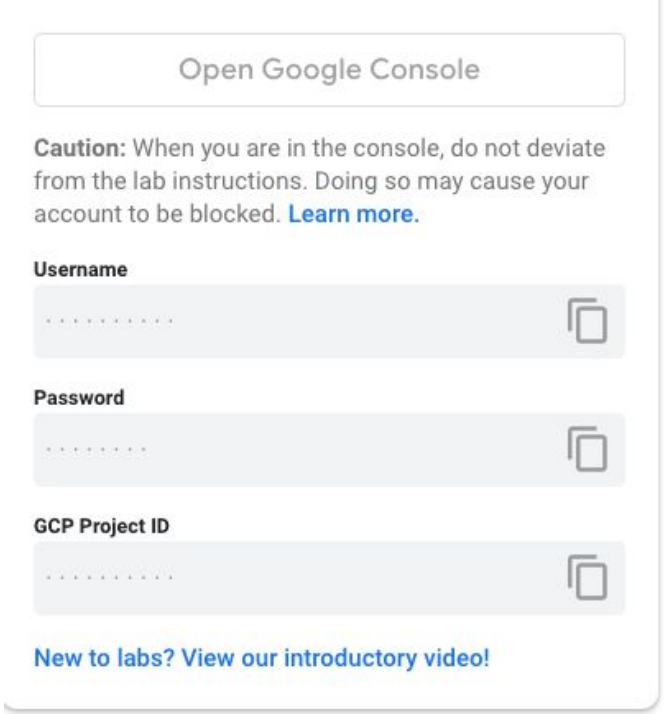

Dataproc: Qwik Start Command Line



Study

How to use the labs?

1. Labs are timed, you cannot paused them
2. The timer is also the amount of time your cloud resources will be running on
3. Every time you hit start, you will get new credentials in the top left corner



Study

Hints / Snippets

```
# Create a cluster for dataproc
gcloud dataproc clusters create [CLUSTER NAME]

# Create a new Job
gcloud dataproc jobs submit spark --cluster example-cluster \
  --class org.apache.spark.examples.SparkPi \
  --jars file:///usr/lib/spark/examples/jars/spark-examples.jar -- 1000

# scale up the cluster by adding more workers
gcloud dataproc clusters update example-cluster --num-workers 4

# scale down once you dont need it anymore
gcloud dataproc clusters update example-cluster --num-workers 2
```


Cloud Natural Language API: Qwik Start

Study

Concepts

- **Syntax Analysis:** Extract tokens and sentences, identify parts of speech (PoS) and create dependency parse trees for each sentence.
- **Entity Recognition:** Identify entities and label by types such as person, organization, location, events, products and media.
- **Sentiment Analysis:** Understand the overall sentiment expressed in a block of text.
- **Content Classification:** Classify documents in predefined 700+ categories.
- **Multi-Language:** Enables you to easily analyze text in multiple languages including English, Spanish, Japanese, Chinese (Simplified and Traditional), French, German, Italian, Korean and Portuguese.
- **Integrated REST API:** Access via REST API. Text can be uploaded in the request or integrated with Google Cloud Storage.

Overview

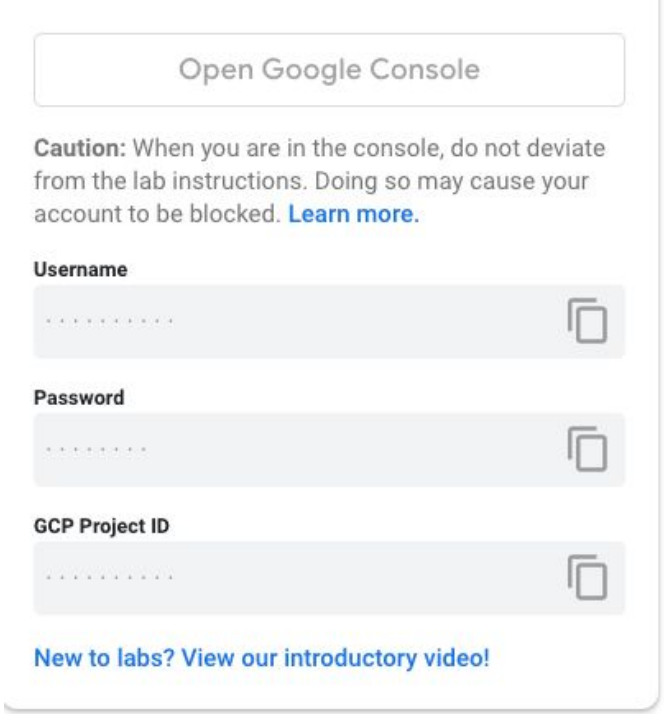
- You will use the analyze-entities method to ask the Cloud Natural Language API to extract "entities" (e.g. people, places, and events) from a snippet of text.

Steps

- Create an API key
- Make an Entity Analysis request

How to use the labs?

1. Labs are timed, you cannot paused them
2. The timer is also the amount of time your cloud resources will be running on
3. Every time you hit start, you will get new credentials in the top left corner



Study

Hints / Snippets

```
# Get project ID
export GOOGLE_CLOUD_PROJECT=$(gcloud config get-value core/project)

# check the value of the project
echo $GOOGLE_CLOUD_PROJECT

# Create a service account
gcloud iam service-accounts create [SERVICE NAME] --display-name "Any display string"

# Create a private key for a service account
gcloud iam service-accounts keys create OUTPUT-FILE
    --iam-account=IAM_ACCOUNT
    [--key-file-type=KEY_FILE_TYPE; default="json"] [GCLOUD_WIDE_FLAG ...]

# Set the path to your credentials
# USER is your user find it by doing
echo $(whoami)
export GOOGLE_APPLICATION_CREDENTIALS="/home/USER/key.json"

# OR use
export GOOGLE_APPLICATION_CREDENTIALS="/home/$(whoami)/key.json"
```

Study

Resources

https://cloud.google.com/natural-language/docs/sentiment-tutorial

Google Cloud Speech API: Qwik Start

Study

Overview

The Google Cloud Speech API:

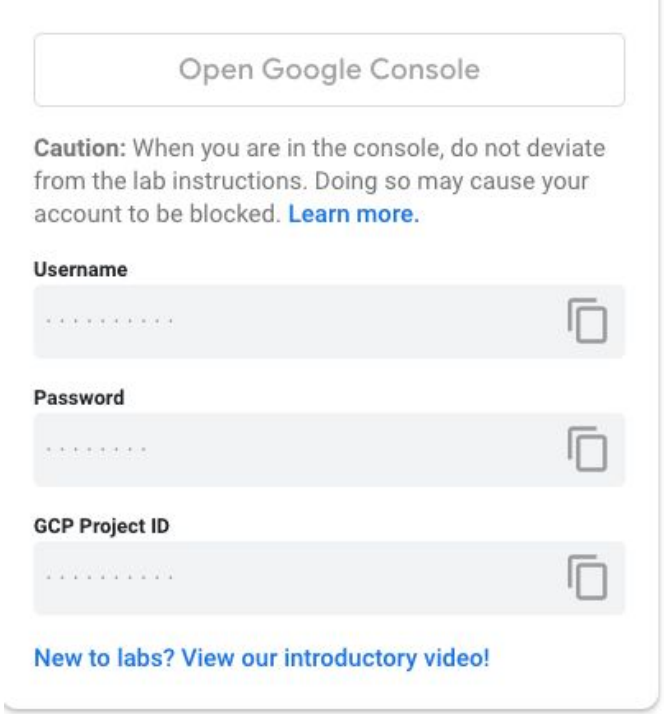
- Enable easy integration of Google speech recognition technologies into developer applications.
- Allows you to send audio and receive a text transcription from the service

Steps

- Create an API key
- Create a Speech API request
- Call the Speech API request

How to use the labs?

1. Labs are timed, you cannot paused them
2. The timer is also the amount of time your cloud resources will be running on
3. Every time you hit start, you will get new credentials in the top left corner



Study

Resources

This tutorial shows how to transcribe the audio recorded from a phone using Cloud Speech-to-Text.

https://cloud.google.com/speech-to-text/docs/phone-model

Study

Video Intelligence: Qwik Start

Study

Overview

Video Intelligence API:

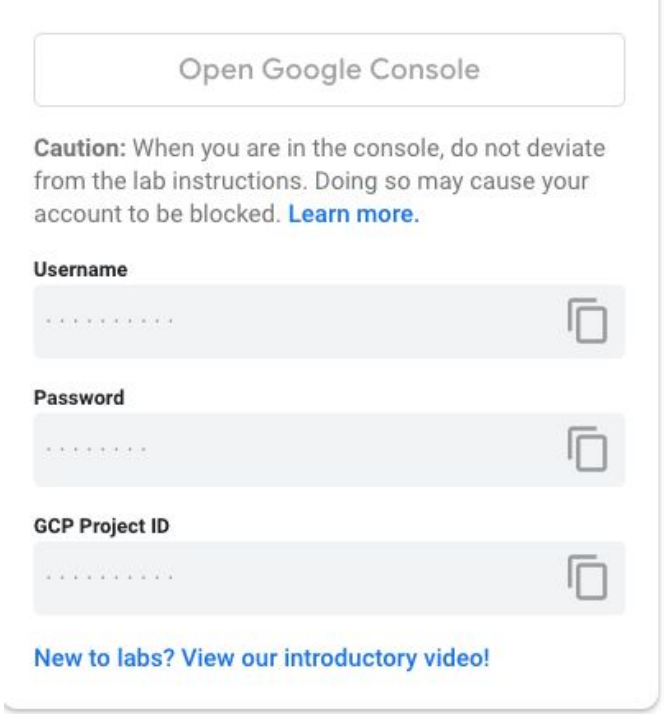
- Has pre-trained machine learning models
- Recognize a vast number of objects, places, and actions in stored and streaming video.
- It's highly efficient for common use cases and improves over time as new concepts are introduced.

Steps

- Enable the Video Intelligence API
- Set up authorization
- Make an annotate video request

How to use the labs?

1. Labs are timed, you cannot paused them
2. The timer is also the amount of time your cloud resources will be running on
3. Every time you hit start, you will get new credentials in the top left corner



Study

Resources

This tutorial walks you through a basic Video API application, using a `SHOT_CHANGE_DETECTION` request. A `SHOT_CHANGE_DETECTION` request provides the annotation results:

- List of all shots that occur within the video
- For each shot, provide the start and end time of the shot

https://cloud.google.com/video-intelligence/docs/shot_detection

Study

Study

That’s a wrap.