3803ICT
Big Data Analysis

Lab 11 – Data Analytics in Cloud
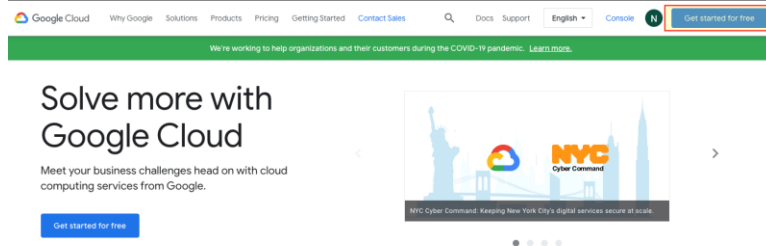
Table of Contents

ATTENTION—DO NOT FORGET TO STOP THE VM INSTANCE!!

I. Account Registration

Click Try Free in https://cloud.google.com/ after you logged in with your google account.

Provide your information as needed.

Input your payment information. You will have $300 for your trial. **Then you can CANCEL this after the course.**

Payment method ⓘ

\#

Card number

MM   YY   CVC
            /

Card number is required

Month   Year   CVC is
is   is   required
required required

Cardholder name

Cardholder name is required

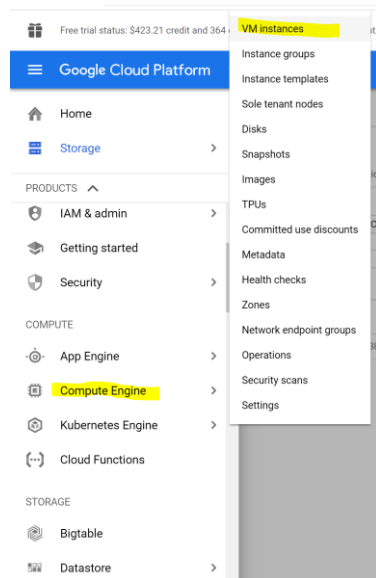☑ Credit or debit card address is same as above

II. Deployment

In this section, you need to deploy a website. The csv file contains emotion analysis of tweet posts over time.

A. Tutorial

1. Create a Compute Engine

2. Upload file to Google Storage
   + Click Cloud Storage in left panel.
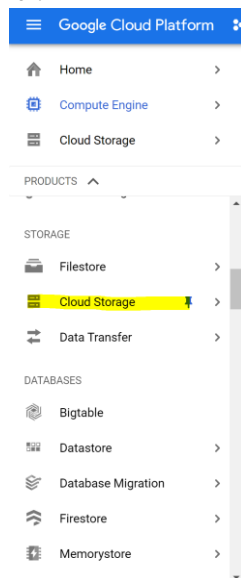
   + Create bucket.
           . Put name.
           . Choose 'regional' as storage class.
           . Select 'australia-southeast1' as location.
           . Default for other information.

+ Upload file emotion_tweets.csv to your new bucket.

3. Create Cloud Sql instance

+You can choose your own password, but you should remember it for later steps.
+ Write down your instance IP (ex: **34.87.227.43**) for later usages.

+ Create new database 'emotions'

4.   Setup permission

a. Allow sql service to access bucket to run import job later

Get Service account from Cloud Sql (ex: p626940988214-bfvjtb@gcp-sa-cloud-sql.iam.gserviceaccount.com)

Open Cloud Storage => Your bucket (ex: emotion_tweets) => Permissions

Add sql service account above (p626940988214-bfvjtb@gcp-sa-cloud-sql.iam.gserviceaccount.com) account with role Storage Object Viewer.

b. Allow VM instance to access sql cloud.

From left panel, click compute engine => copy your instance's external IP. (ex: **34.87.235.255**)

From left panel, click SQL => select your instance => Connections => Click "Add Network"
Copy above IP (ex: **34.87.235.255**) to Network.
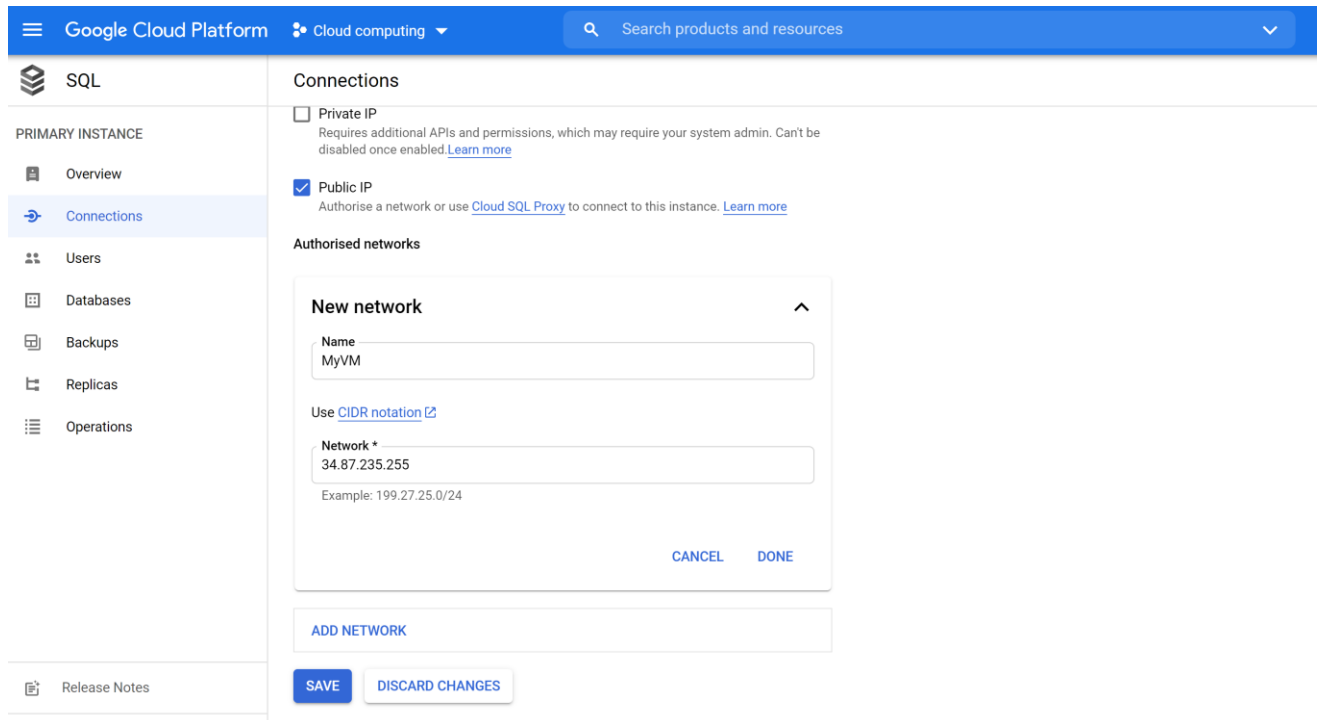
Enable Cloud SQL API: https://console.cloud.google.com/flows/enableapi?apiid=sqladmin. Select your project and then enable:
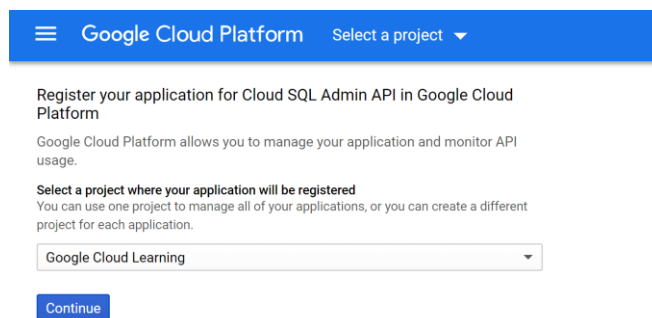
Follow the intruction to create the service account:
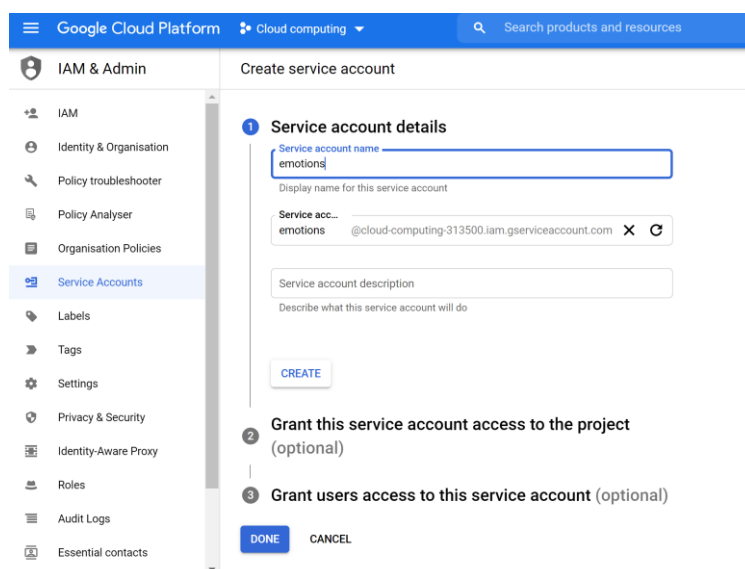
5. Connect My SQL in VM instance and import data
   From left panel => choose Compute Engine => connect SSH
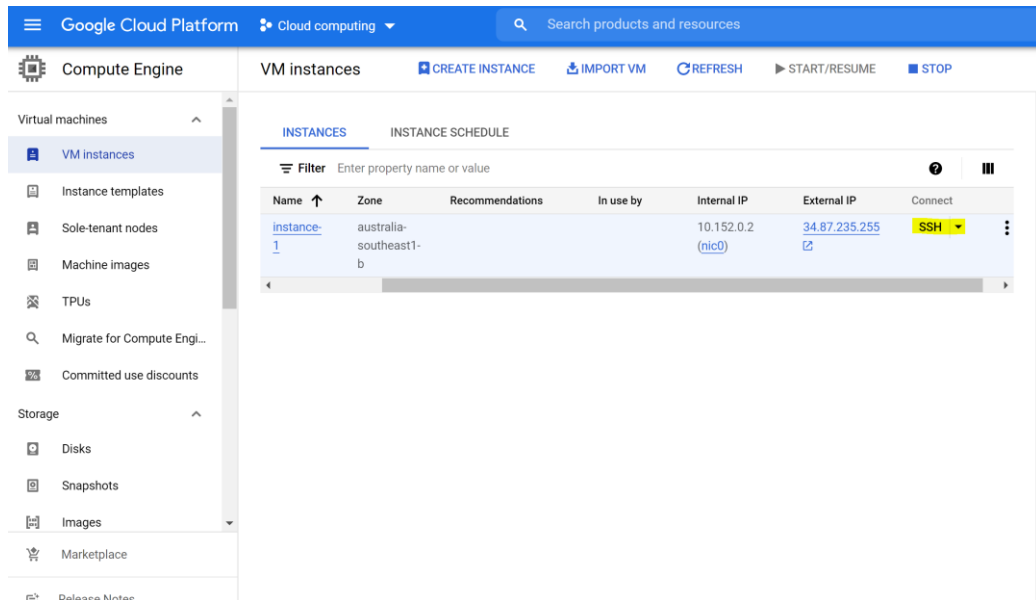
Install mysql client using command: sudo apt-get install mysql-client

Connect to your mysql using comman: mysql --host=[your sql IP] --user=root --password=[your root password]

mysql --host=34.87.227.43 --user=root --password=griffith_cloud_learning

Create table by running these commands in console.

+ use emotions;
+ CREATE TABLE IF NOT EXISTS emotions (

tweet_id bigint,
sentiment text,
author text,
content text
);

+ quit

Change gcloud authentication to run import task: gcloud auth login => then follow their instruction

```
@instance-1:~$ gcloud auth login

You are running on a Google Compute Engine virtual machine.
It is recommended that you use service accounts for authentication.

You can run:

  $ gcloud config set account `ACCOUNT`

to switch accounts if necessary.

Your credentials may be visible to others with access to this
virtual machine. Are you sure you want to authenticate with
your personal account?

Do you want to continue (Y/n)?  Y

Go to the following link in your browser:

    https://accounts.google.com/o/oauth2/auth?redirect_uri=urn%3Aietf%3Awg%3Aoauth%3A2.0%3Aoob&prompt=select_account&
response_type=code&client_id=32555940559.apps.googleusercontent.com&scope=https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fu
serinfo.email+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fcloud-platform+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fappe
ngine.admin+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fcompute+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Faccounts.reau
th&access_type=offline


Enter verification code: 4/8wCRu-cwLjjLfkD9eWbv-aINxEmkF-AiGjLxuaX9_TEn8_qL2yYMJLk
WARNING: `gcloud auth login` no longer writes application default credentials.
If you need to use ADC, see:
  gcloud auth application-default --help
```


Import data from emotion_tweets.csv file(in your Storeage bucket) by running this command:

gcloud sql import csv [INSTANCE_NAME] gs://[BUCKET_NAME]/[FILE_NAME] \

--database=[DATABASE_NAME] --table=[TABLE_NAME]

+ gcloud sql import csv googlecloudsql gs://emotion_tweets/emotion_tweets.csv \
 --database=emotions --table=emotions

```
-1:~$ gcloud sql import csv googlecloudsql gs://emotion_tweets/emotion_tweets.csv  --database=
emotions --table=emotions
```

6. Deploy sample app

+ Unzip app.zip file
+ Open main.py and change your mysql information.

```
### TO EDIT ###
sql_host = '35.197.176.134'
sql_connection_name='lexical-archery-231806:australia-southeast1:googlecloudsql'
sql_port = 3306
sql_database = 'currencies'
sql_user = 'root'
sql_password = 'griffith_cloud_learning'
### END TO EDIT ###
```

+ Open app.yaml and change to your mysql information.

```
#[START gae_flex_mysql_env]
env_variables:
    # Replace user, password, database, and instance connection name with the values obtained
    # when configuring your Cloud SQL instance.
    SQLALCHEMY_DATABASE_URI: >-
      mysql+pymysql://root:griffith_cloud_learning@/currencies?unix_socket=/cloudsql/lexical-archery-231806:australia-southeast1:googlecloudsql
#[END gae_flex_mysql_env]

#[START gae_flex_mysql_settings]
# Replace project and instance with the values obtained  when configuring your
# Cloud SQL instance.
beta_settings:
    cloud_sql_instances: lexical-archery-231806:australia-southeast1:googlecloudsql
#[END gae_flex_mysql_settings]
```

+ Push it to your github.
+ Install git in your cloud.
       .sudo apt-get update.
       .sudo apt-get install git.
+ In VM console, clone your repository to your local directory (ex: Cloud/Demo).
+ cd to your local directory
+ Install pip3
       . sudo apt-get update
       . sudo apt-get -y install python3-pip
+ Run "pip3 install -t lib -r requirements.txt" to install necessary libraries.
+ Upgrade google-cloud-sdk.

```
# Create environment variable for correct distribution
export CLOUD_SDK_REPO="cloud-sdk-$(lsb_release -c -s)"

# Add the Cloud SDK distribution URI as a package source
echo "deb http://packages.cloud.google.com/apt $CLOUD_SDK_REPO main" | sudo tee -a
/etc/apt/sources.list.d/google-cloud-sdk.list

# Import the Google Cloud Platform public key
curl https://packages.cloud.google.com/apt/doc/apt-key.gpg | sudo apt-key add -

# Update the package list and install the Cloud SDK
sudo apt-get update && sudo apt-get install google-cloud-sdk
```
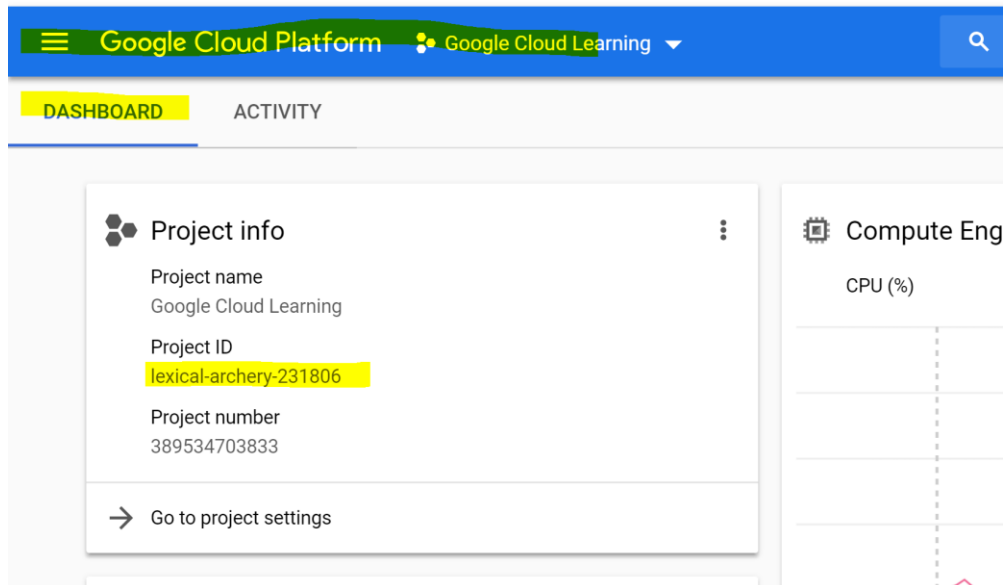
+ Following these steps to activate your server.
       . gcloud auth application-default login

. gcloud config set project [YOUR PROJECT_ID-]

. gcloud init => choose 1. Then follow the instruction (you need to select 38-australia-southeast1-b as your region).
. gcloud beta app deploy (this will take time)

+ Browse your web in browser.
+ If you have any errors, go to Error Reporting in left panel of cloud.

B. (Optional) Spark – Hello world

You will learn how to submit a job in Google Cloud and execute it.
1. Upload file to bucket
   + Create a bucket in your Google Storage like previous question (ex: **sparklearning**).
   + Upload file hello_world.py to your new bucket
2. Create a new Dataproc Cluster in Google Cloud
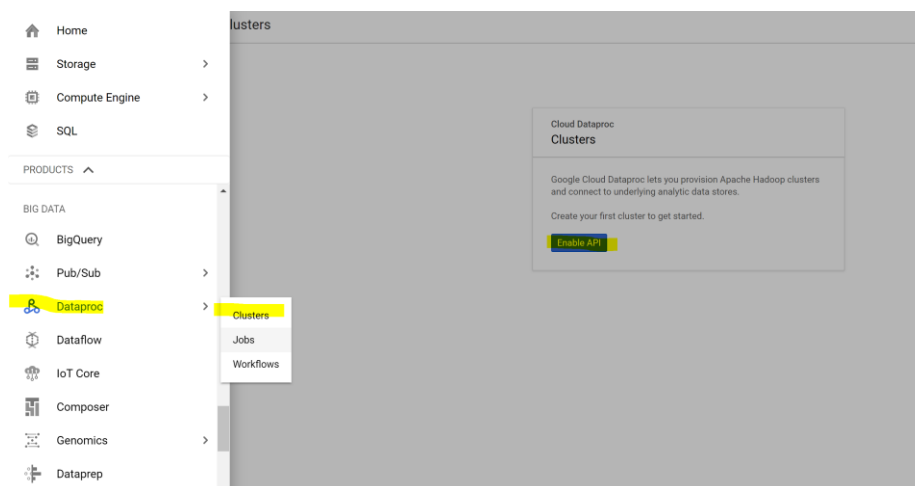
+ Then click create cluster.
   . Name: cluster-spark
   . Region: australia-southeast1, zone: australia-southeast1-a.
   . Choose 1vCPU for both master and worker nodes
   . Others as default

3. Submit a job via Cloud Shell.
   + Click icon on the top right of screen to activate Cloud Shell.
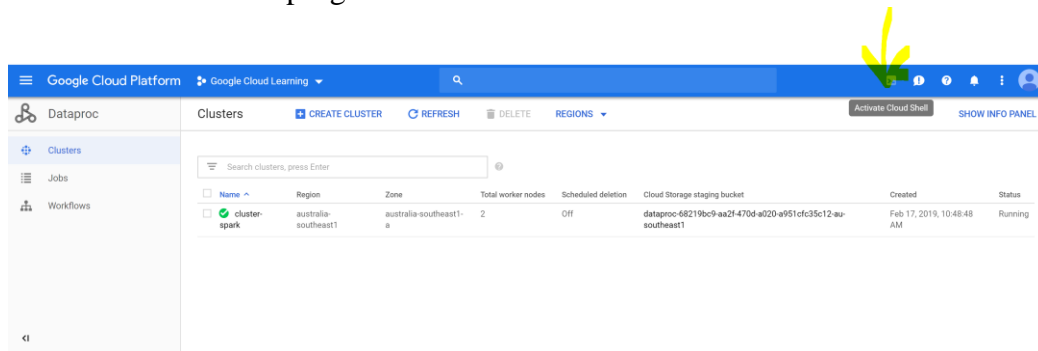
+ Submit a job via Cloud Shell with this syntax.

```
gcloud dataproc jobs submit job-command \

 --cluster cluster-name --region region \

job-specific flags and args
```

. gcloud dataproc jobs submit pyspark --cluster cluster-spark --region australia-southeast1 gs://sparklearning/hello_world.py

+ Wait and observe the result.

C. (Optional) Practices

- Upload pagerank.py and web-Standford.txt to bucket
- Run PySpark for them

gcloud dataproc jobs submit pyspark --cluster cluster-spark --region australia-southeast1 gs://sparklearning/pagerank.py -- gs://sparklearning/web-Stanford.txt 20