

1. Complete file *project3.py* by filling in the blanks (...) for each task with one or more lines of code.
 - Specify the Twitter account that you want to analyze in line 27 of *project3.py* (the default Twitter account is CNN).
2. Create a Twitter account and apply for a Twitter developer account.
3. Retrieve your Twitter credentials (key and access token). For detailed instructions, see <https://developer.twitter.com/en/docs/basics/authentication/guides/access-tokens>.
 - Complete file *twitter_auth.py* by filling in your Twitter credentials.
4. Open the Google Cloud Platform (GCP) console and log in to your account.
5. Create a bucket.
 - Go to *Storage > Browser > Create bucket*.
 - Choose a name for the bucket. This name must be globally unique.
 - Leave default values for all other fields and create the bucket.
6. Upload file *gcprun.sh* to the bucket.
7. Create a cluster.
 - Go to *Dataproc > Clusters > Create a cluster*.
 - Choose a name for the cluster.
 - Go to *Advanced Options*.

- In *Cloud Storage staging bucket*, choose the bucket created in step 5.
 - In *Image*, select 1.4 Ubuntu 18.04.
 - In *Initialization actions*, click on *Add initialization action*. Then, browse the bucket created in step 5 and select file *gcprun.sh*.
 - Leave default values for all other fields and create the cluster.
8. Select the cluster created in step 7. Then, go to the *VM Instances* tab and SSH into the master node of the cluster.
 9. Create new file *twitter_auth.py* in the master node and copy the contents of *twitter_auth.py*.
 10. Create new file *project3.py* in the master node and copy the contents of *project3.py*.
 11. To retrieve and analyze the Tweets, run the command:

```
sudo -s spark-submit project3.py
```

Note: This may take a few minutes to run. Also, the number of Tweets retrieved may change in every run. If not enough Tweets are retrieved to create the word clouds, you should run this command again.

12. To save the results to the bucket, run the command:

```
sudo gsutil -m cp -r Good.png Bad.png Neutral.png /var/log/dataproc-  
initialization-script-0.log gs://#BUCKETNAME#
```

(Replace #BUCKETNAME# with the name of the bucket).

Note: This project will be difficult to test locally, so you will be spending most of the time working in GCP. Do not forget to delete the cluster when you are not using it. Otherwise, you will be charged and you may use all you

- Python file *project3.py* with your completed code for all tasks.
- Log file of the execution of the cluster (saved to the bucket).
- Report (1-2 pages, saved as a PDF file) containing the word clouds for “good,” “neutral,” and “bad” tweets, your analysis of the word clouds, and a brief explanation of the reasons for using GCP vs. a local environment for this project.