

# Lab - 3 Task 1

## Objectives:

Spin Up a VM and write a python program to count lines of a file placed in GCS.

## Instructions:

1. On Google Cloud Console, go to Compute Engine->VM instances.
2. Click on the Create Instance button.
3. Select allow API access to all the Google Cloud APIs in the window that appears.
4. Then click on Create.
5. Click on the created VM instance under the VM instances tab.
6. In the remote access, select the option SSH (open in the browser window).

The instructions from here on have to be executed on the VM window that opens:

7. In the window that pops up, download the file using `wget https://storage.googleapis.com/bdl2022/task1.txt`.
8. Install google-cloud-sdk using the instructions on the link <https://cloud.google.com/sdk/docs/quickstart-linux>
9. Using `gsutil cp`, copy the downloaded file to your bucket.
10. Install pip3 - `sudo apt-get install python3-pip`
11. Install google-cloud-storage using `pip3 install google-cloud-storage` ( These are python APIs for interacting with GCS )
12. Run the following code:

```
from google.cloud import storage

client = storage.Client()

bucket = client.get_bucket('word_count_gcs')

blob = bucket.get_blob('addresses.csv')

x = blob.download_as_string()
x = x.decode('utf-8')

# count the lines in x
```