GA-1

#### **Introduction**

This report documents the completion of the Week 2 assignment for counting the number of lines in a file stored in Google Cloud Storage (GCS). The task involves writing a Python script, running it on a virtual machine (VM), and capturing the output. The report includes the steps for setting up the VM, running the script, and the results obtained.

### **Steps to Set Up the VM and Necessary Environment**

Detail the steps taken to set up the VM and install the necessary tools.

#### **Steps to Set Up the VM and Necessary Environment**

1. **Choose a Cloud Service Provider**
   * For this assignment, Google Cloud Platform (GCP) was selected.
2. **Create a VM Instance**
   * Logged into the GCP console.
   * Navigated to the "Compute Engine" section.
   * Created a new VM instance with the following specifications:
     + Machine type: e2-medium (2 vCPUs, 4 GB memory)
     + Operating system: Ubuntu 20.04 LTS
3. **Install Necessary Tools**
   * After the VM was up and running, connected to it via SSH.

Updated the package list and installed Python:  
  
sudo apt update

sudo apt install python3 python3-pip

Installed Google Cloud SDK:  
  
sudo apt install apt-transport-https ca-certificates gnupg

echo "deb [signed-by=/usr/share/keyrings/cloud.google.gpg] http://packages.cloud.google.com/apt cloud-sdk main" | sudo tee -a /etc/apt/sources.list.d/google-cloud-sdk.list

sudo apt update && sudo apt install google-cloud-sdk

gcloud init

Configured the gcloud CLI for the project:  
  
gcloud auth login

gcloud config set project <your\_project\_id>

### **Python Code Explanation**

Provide the Python code and explain each part.

#### **Python Code Explanation**

The following Python script counts the number of lines in a file stored in GCS.

from google.cloud import storage

def count\_lines\_in\_gcs\_file(bucket\_name, file\_name):

# Initialize a Google Cloud Storage client

client = storage.Client()

# Get the bucket containing the file

bucket = client.get\_bucket(bucket\_name)

# Get the blob (file object) from the bucket

blob = bucket.blob(file\_name)

# Download the file's content as a string

file\_content = blob.download\_as\_text()

# Count the number of lines in the file

line\_count = len(file\_content.splitlines())

return line\_count

# Example usage

if \_\_name\_\_ == "\_\_main\_\_":

bucket\_name = '21f3002806'

file\_name = 'GA-1/GA-1-IBD.txt'

line\_count = count\_lines\_in\_gcs\_file(bucket\_name, file\_name)

print(f"Number of lines in {file\_name}: {line\_count}")

# Save the result to a text file

with open('output.txt', 'w') as f:

f.write(f"Number of lines in {file\_name}: {line\_count}")

**Explanation:**

* **Imports**: Imports the Google Cloud Storage library.
* **count\_lines\_in\_gcs\_file function**:
  + Initializes a Google Cloud Storage client.
  + Retrieves the bucket and the blob (file) from GCS.
  + Downloads the file content as a string.
  + Splits the content by lines and counts the number of lines.
* **Main block**:
  + Defines the bucket name and file name.
  + Calls the function and prints the number of lines.
  + Writes the result to an output text file.

### **Output of the Script**

Run the script and capture the output.

#### **Output of the Script**

After setting up the VM and running the script, the following output was obtained:

sh

Copy code

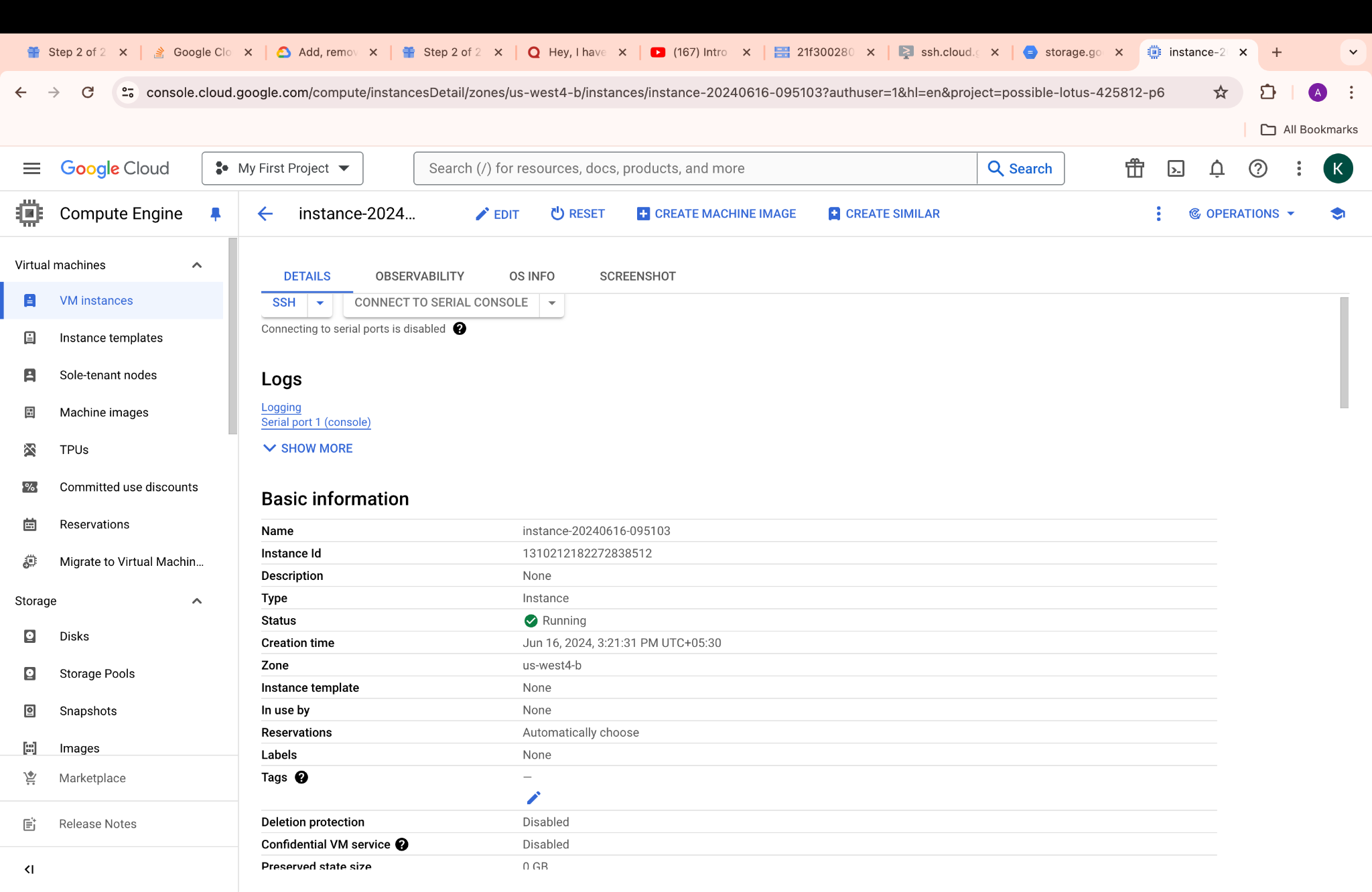
Number of lines in GA-1/GA-1-IBD.txt: 279

### **Screenshots Showing the VM Setup, Running of the Script, and the Result**

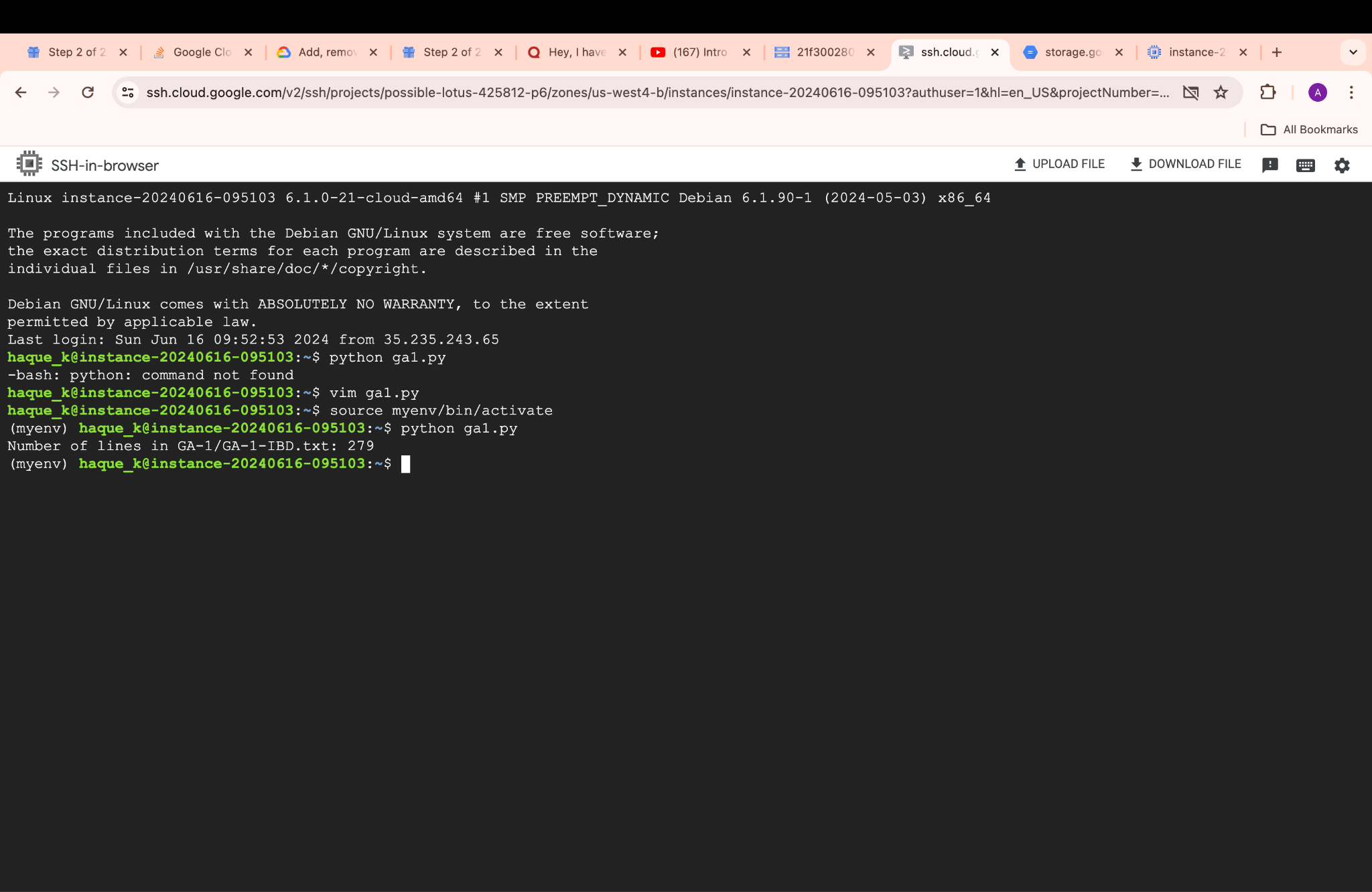
Include relevant screenshots.

#### **Screenshots**

1. **VM :**

****

1. **Running the Script**:

****