**ADTA-5240 - Assignment 2 – Sri Charan(11661666)**

**Task 1: Create Storage Bucket, 3 folders, and Load Data into the Bucket in GCP**

Task 1 has been divided into 3 sub tasks – Create storage bucket, creating 3 folders and loading data into bucket.

**Sub task 1: Create storage bucket:**

1. Navigate to the project on which we need to create Cloud Storage bucket ( I selected project: ADTA-5240-Sri-Charan) and select Cloud Storage

Graphical user interface, application

Description automatically generated

1. Now, click on create button on the top,

A screenshot of a computer

Description automatically generated

1. Provide bucket configurations:
   1. Give the name to the bucket you need to create. Name must be unique globally.
   2. choose where to store your bucket – selecting multi-region option gives higher availability and low latency of the bucket.
   3. Then, choose storage class – which determines type of bucket you are creating. Based on this, costs for storage and operations will be calculated. This can be selected as per your requirement. For now, I selected default class of standard type.
   4. Next, choose access control of the objects – which means whether objects stored in the bucket are accessible via public internet, here I selected uniform since I can set up permissions to all buckets created uniformly
   5. Now, we choose how to protect object data in the bucket. By default, GCP provides you protection against data loss however if one needs to have more secure way of storage they can store more advanced cost applicable security policies. I selected now which means I use default security mechanism provided by GCP
   6. Finally, click on CREATE button to create your bucket

Graphical user interface, text, application, email

Description automatically generated

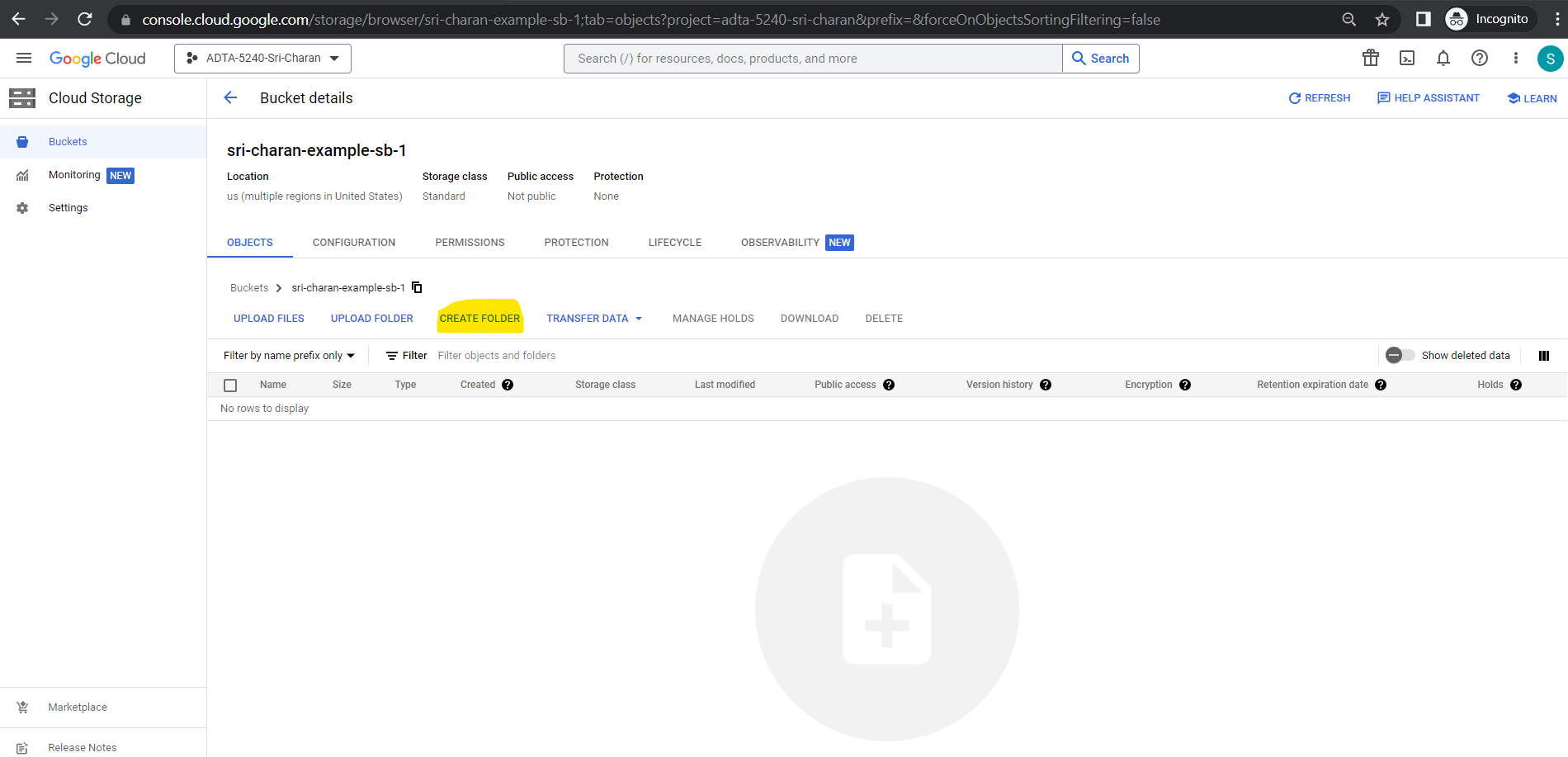
1. Once the bucket creates successfully, you can see the options for your new bucket as below

Graphical user interface, text, application

Description automatically generated

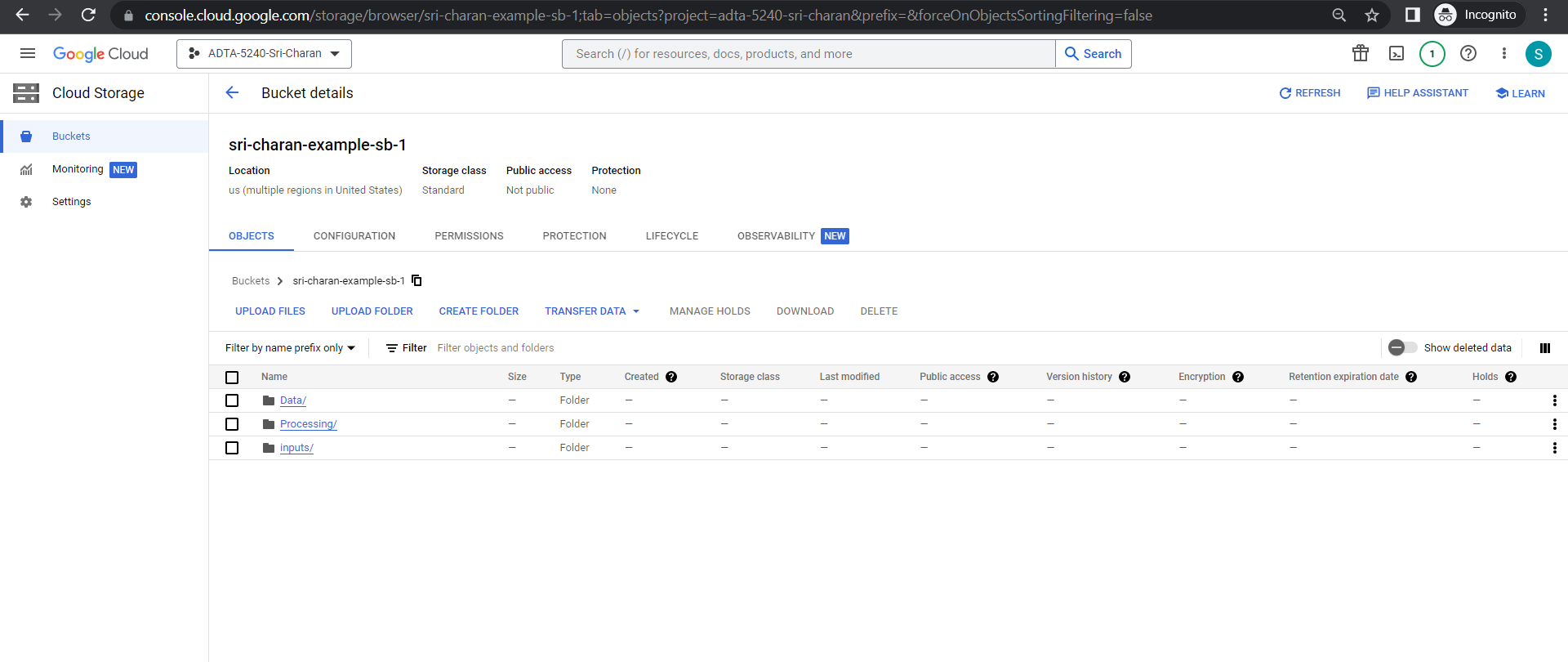
**Sub task 2: Creating 3 Folders in the bucket:**

1. Select the option CREATE FOLDER and give the name to the folder,



Graphical user interface, application

Description automatically generated



**Sub task 3: Load data into the bucket**

1. Select the folder to which dataset need to be uploaded. I have selected “Data” folder

A screenshot of a computer

Description automatically generated

1. Click on “UPLOAD FILES” and select file you want to upload

A screenshot of a computer

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. You can see the file uploaded,

Graphical user interface, application

Description automatically generatedI have attached below screenshots to depict files have been uploaded into other 2 folders in the storage bucket

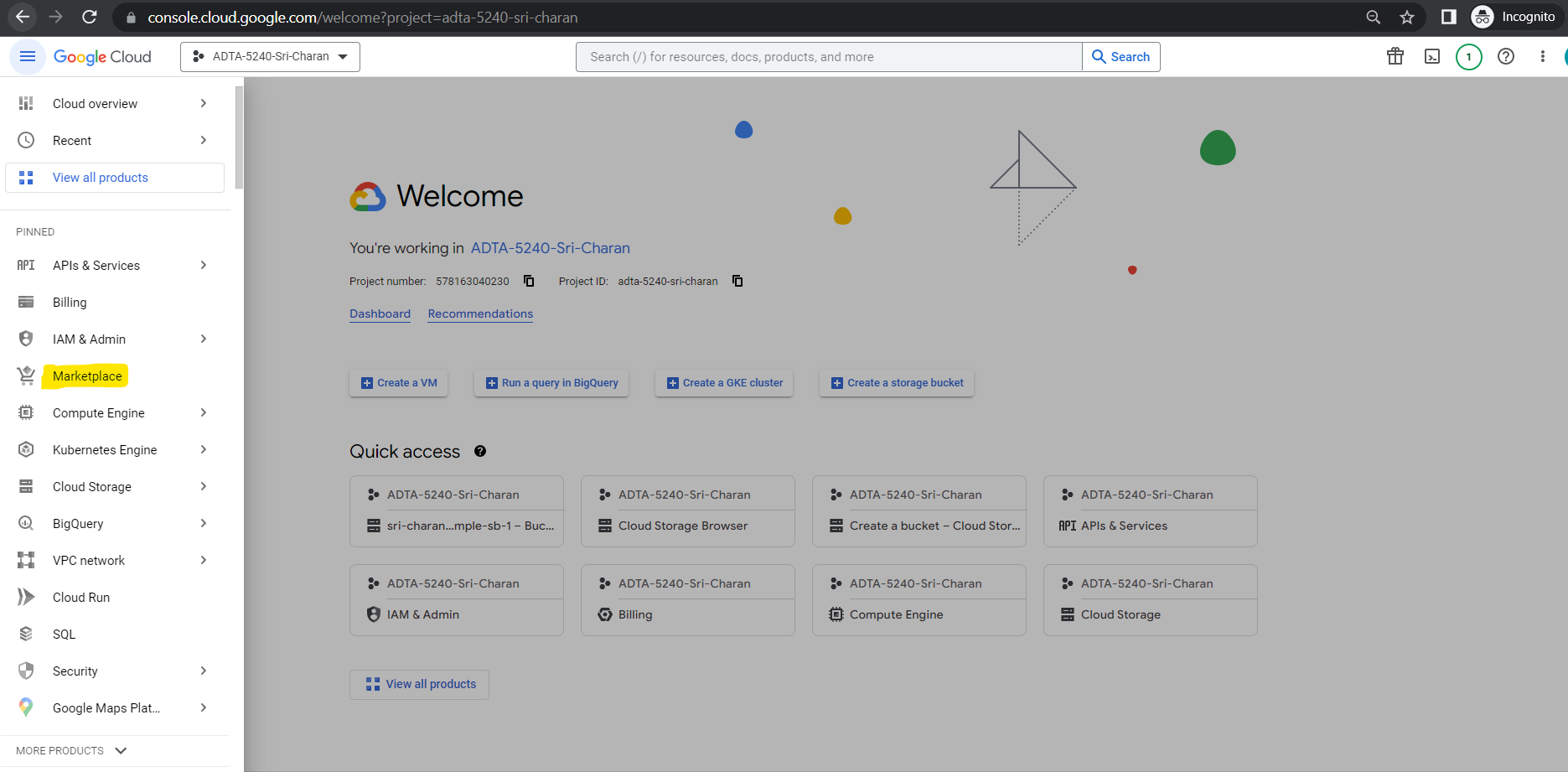
Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

**Task 2: Create a Hadoop cluster**

1. Under the project home page, click on hamburger icon button towards left side of the screen and navigate to marketplace 
2. Now, search for Compute Engine API in the Marketplace search bar and select Compute Engine API

Graphical user interface, text, application

Description automatically generated

1. Now, click on enable button for the Compute Engine API. Which enables Compute Engine API for the project you have selected.

Graphical user interface, text, application, email

Description automatically generated

1. After few seconds, it gets enabled

Graphical user interface, text, application, Teams

Description automatically generated

1. Now, repeat the same steps for DataProc API

Graphical user interface, text, application, email

Description automatically generated

Sub task 2: Creating Dataproc Cluster

1. Click on hamburger icon on left, scroll to Dataproc under more apps and select Clusters,

A screenshot of a computer

Description automatically generated

1. Click on CREATE CLUSTER

Graphical user interface, text, application, email

Description automatically generated

1. Click Create on Cluster on Compute Engine

Graphical user interface, text, application

Description automatically generated

1. Under Setup cluster section, give the name to the cluster: hadoop-spark-2-cluster-11661666. Select region as us-central1 and zone as us-central1-a and cluster type as Standard

Graphical user interface, text, application

Description automatically generated

1. Under Versioning section, click change and select the Image type and version as Ubuntu 1.5 (Ubuntu 18.04 LTS, Hadoop 2.10, Spark 2.4)

Graphical user interface, text, application, email

Description automatically generated

1. Then select Configure nodes section, and provide manage node configurations as Machine Family GENERAL PURPOSE, Series E2, machine type e2-standard-8, etc. (refer below screenshot)

A screenshot of a computer

Description automatically generated

1. For worker node, select family GENERAL PURPOSE, series E2 and machine type e2-standard-4, etc.

Graphical user interface, text, application

Description automatically generated

1. Under Customize cluster, leave the network configuration as is and scroll down to the ‘Cloud Storage staging bucket’ and click on browse. Then select your bucket created earlier and click “SELECT”.

Graphical user interface, text, application

Description automatically generated

1. Finally, you can click on CREATE. Hang on it takes a moment to configureGraphical user interface, text, application, email

   Description automatically generated
2. After few minutes, you can see Hadoop cluster starts running,

A screenshot of a computer

Description automatically generated

1. Now, navigate to Compute Engine Service you can see 3 nodes are running(1 master, 2 worker nodes)

Graphical user interface, text, email

Description automatically generated

Note: Once activity is done, always remember to stop cluster so that you are not charged.

Graphical user interface, text, application

Description automatically generated