**Create a Storage Bucket, three folders, and Load Data into the Bucket in GCP.**

* Steps I took to create a new bucket in GCP

1. In the Google Cloud console navigation menu, click **Buckets**.
2. On the **Buckets** page, click **Create**.
3. In the **Name your bucket** field, enter a globally unique name **bucket\_adta5240bin**.
4. Click **Continue**.
5. In the **Choose where to store your data** section, select **Region** as the location type, and then Select **us-central1**.
6. Click **Continue**.
7. In the **Choose a default storage class for your data** section, select **Standard** as the Class.
8. Click **Continue**.
9. In the **Choose how to control access to object** section, keep the Enforce public access prevention on this bucket checked, and from Access control, select **Uniform**.
10. Click **Continue**.
11. In the **Choose how to protect object data: No change**
12. Click **Create**.

* Screenshots documenting the steps I have taken.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a bucket

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* Successfully created a bucket in GCP.

A screenshot of a computer

Description automatically generated

* Steps I took to create three folders in GCP

1. Click **Create folder**.
   1. In the **Create folder** window, do the following:
   2. In the **Name** field, enter **data**.
   3. Click **Create**.
2. Click **Create folder**.
   1. In the **Create folder** window, do the following:
   2. In the **Name** field, enter **log**.
   3. Click **Create**.
3. Click **Create folder**.
   1. In the **Create folder** window, do the following:
   2. In the **Name** field, enter **outputs**.
   3. Click **Create**.

* Screenshots documenting the steps you have taken.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* successfully created three folders in GCP.

A screenshot of a computer

Description automatically generated

* Include steps I took to upload 2 data sets into the newly created "Data" folder
* include screenshots documenting the steps you have taken.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* Screenshot that shows I have successfully uploaded two datasets into the "data" folders in GCP.   
    
  A screenshot of a computer

  Description automatically generated

**Create a Hadoop Cluster**

* Steps I took to enable the Compute Engine API
  1. Went to the project's dashboard
  2. Click Enable and Manage APIs
  3. Click Enable API
  4. Click Compute Engine API
  5. Click Enable to enable the Compute Engine API
* Steps I took to enable the DataProc API
  1. Navigate to Menu > APIs & Services > Enable Apis and Services
  2. Search for DataProc API
  3. Click Cloud Dataproc API
  4. Click the Enable API button
  5. Wait for the API to enable
  6. Navigate to Menu > Product & Solutions
  7. Under Analytics, pin [Dataproc](https://console.cloud.google.com/dataproc?authuser=1&project=adta5240bin) to Manu
* Screenshots documenting the steps you have taken.
  + Enable the Compute Engine API

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* + Enable the DataProc API

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## Create a Hadoop and Spark Cluster in GCP Go to the Dataproc Clusters page.

1. Go to **Clusters**
2. Click Create **cluster**.
3. In the **Create** Dataproc cluster dialog, click Create in the Cluster on the Compute engine row.
4. **Set up cluster**
   1. In the **Cluster** **Name** field, enter **hadoop-spark-2-cluster**.
   2. In the **Location** **section**: In the Region, select **us-central1**, and in Zone lists, select **us-central-a**.
   3. In the **Cluster type field,** select **Standard (1 master, N workers)**
   4. In the **Versioning > Change > 1.5 (Ubuntu 18.04 LTS, Hadoop 2.10, Spark 2.4)**A screenshot of a computer

      Description automatically generated
5. **Configure node**
   1. In the **Manager node** field
      1. Series> E2
      2. Machine type > e2 standard-8
   2. CPU > Primary disc size : 128GB > Primary disc type > Standard persistent disk  
      A screenshot of a computer

      Description automatically generated
   3. In the **Worker Node filed**
      1. Series> E2
      2. Machine type > e2 standard-4
      3. CPU > number of nodes > 2 > Primary disc size : 128GB > Primary disc type > Standard persistent disk

A screenshot of a computer program

Description automatically generated

1. **Customize cluster**
   1. In the **Cloud Storage staging bucket** field, browse > select your bucket
2. For all the other options, use the default settings.
3. To create the cluster, click Create.

A screenshot of a computer

Description automatically generated

* all three nodes are successfully running.

A screenshot of a computer

Description automatically generated

* three nodes are turned off in GCP

A screenshot of a computer

Description automatically generated