Yog Chaudhary

11727095

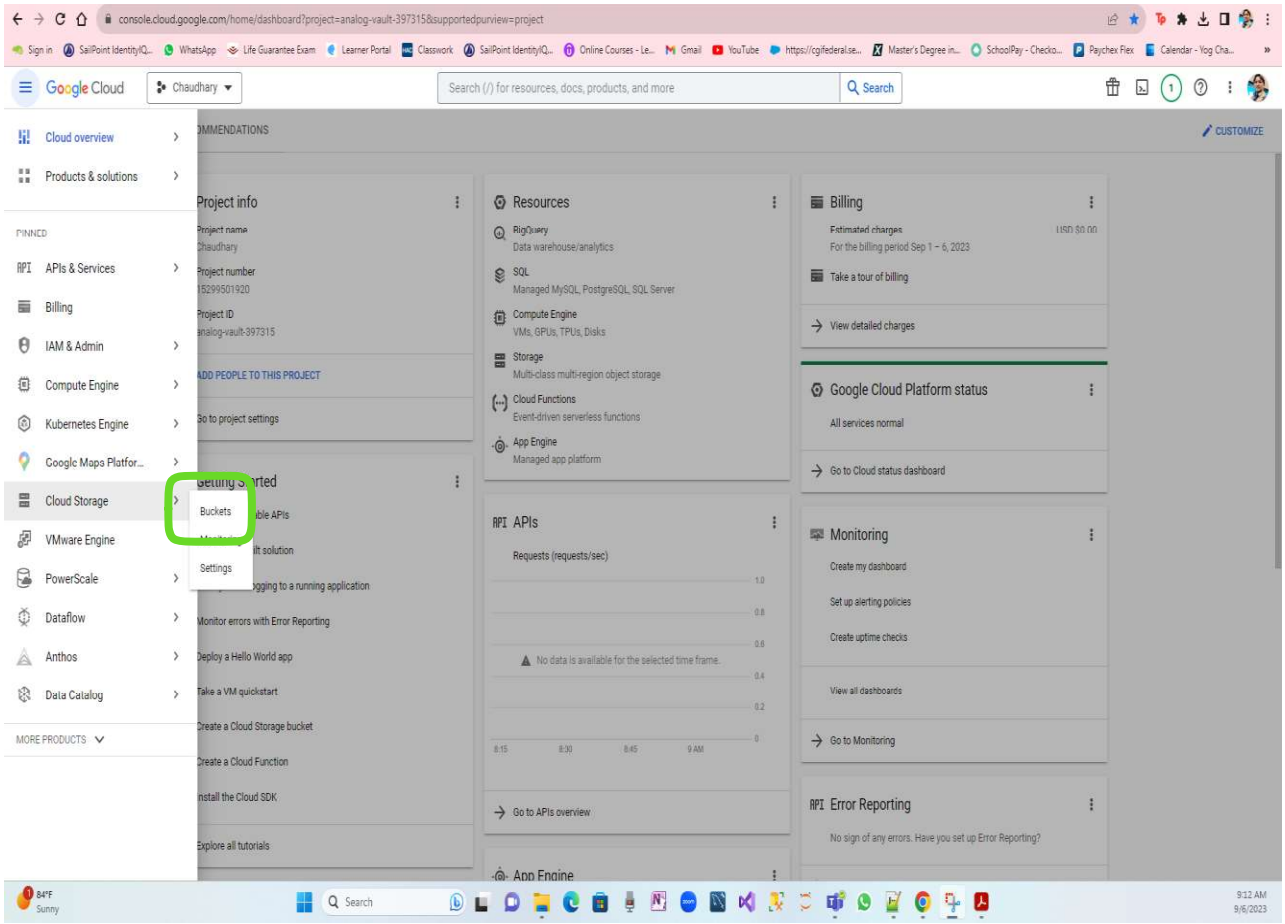
ADTA 5240 Week 2'nd (harvesting, Storing, And Retrieving Data)

Professor: Dr. Zeynep Orhan **Sep 09, 2023**

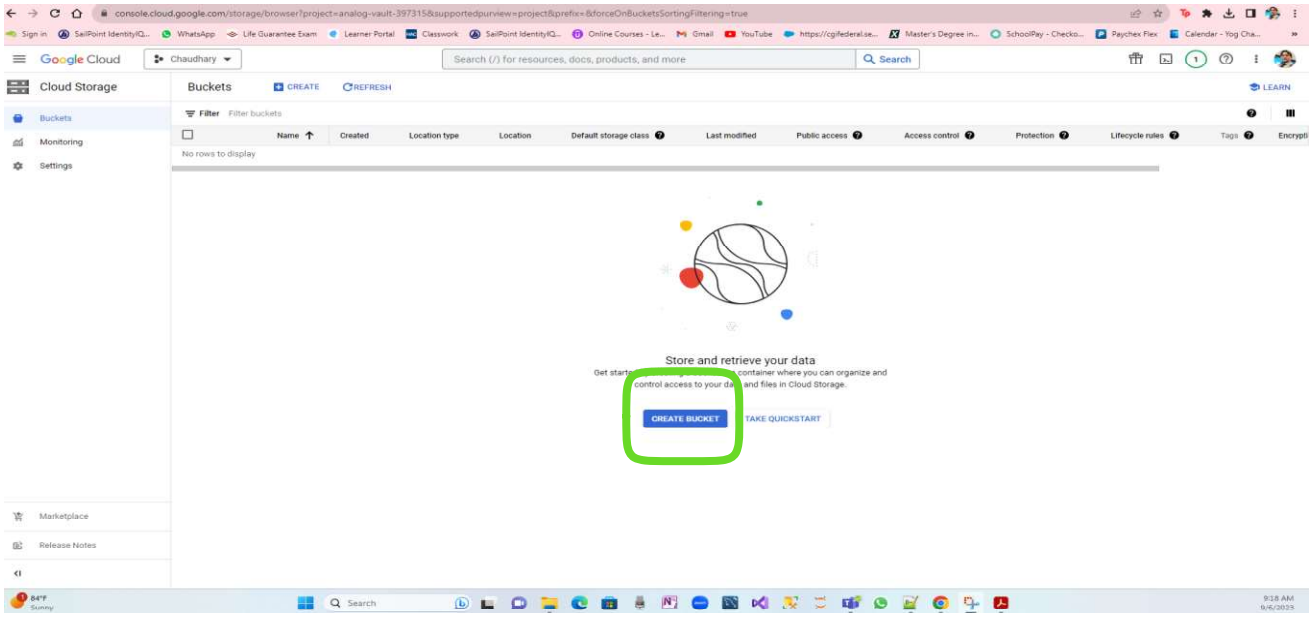
University Of North Texas

- **How to Create a New Storage Bucket, 3 Folder, and Load Data into a Folder in GCP.pdf**

1. In this, I clicked on the navigation pane after that, I scrolled down and clicked on cloud storage and buckets.
2. Click on create Bucket.

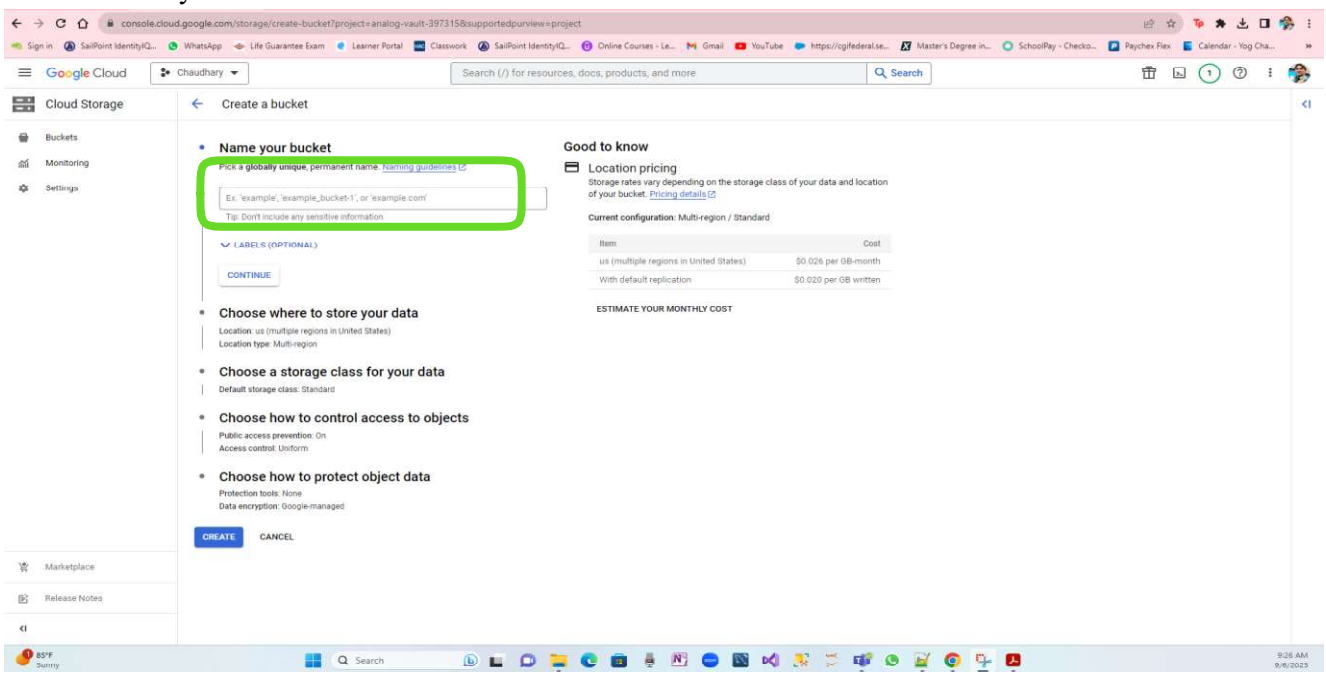


3. By clicking on the bucket I went to this page, and then I clicked on create a bucket.

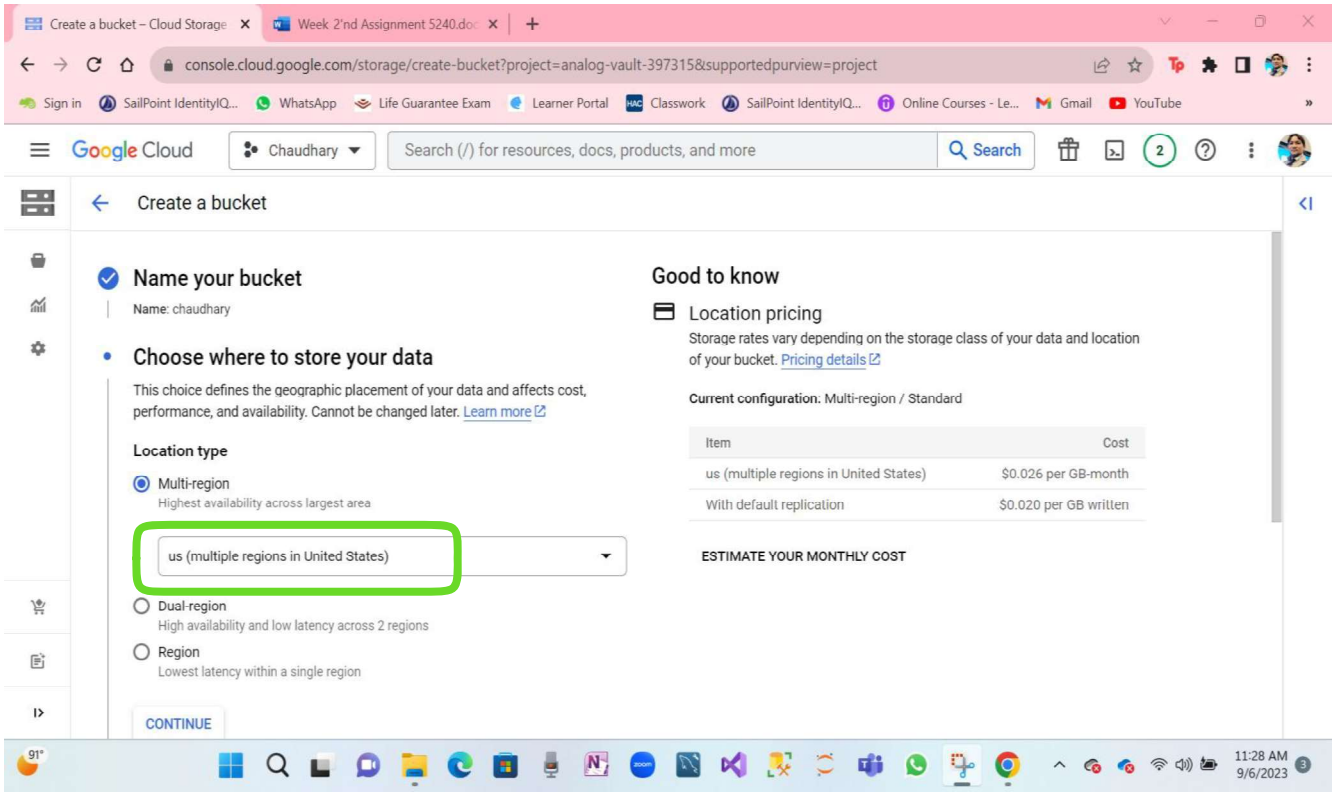


4. After I was redirected to this page.

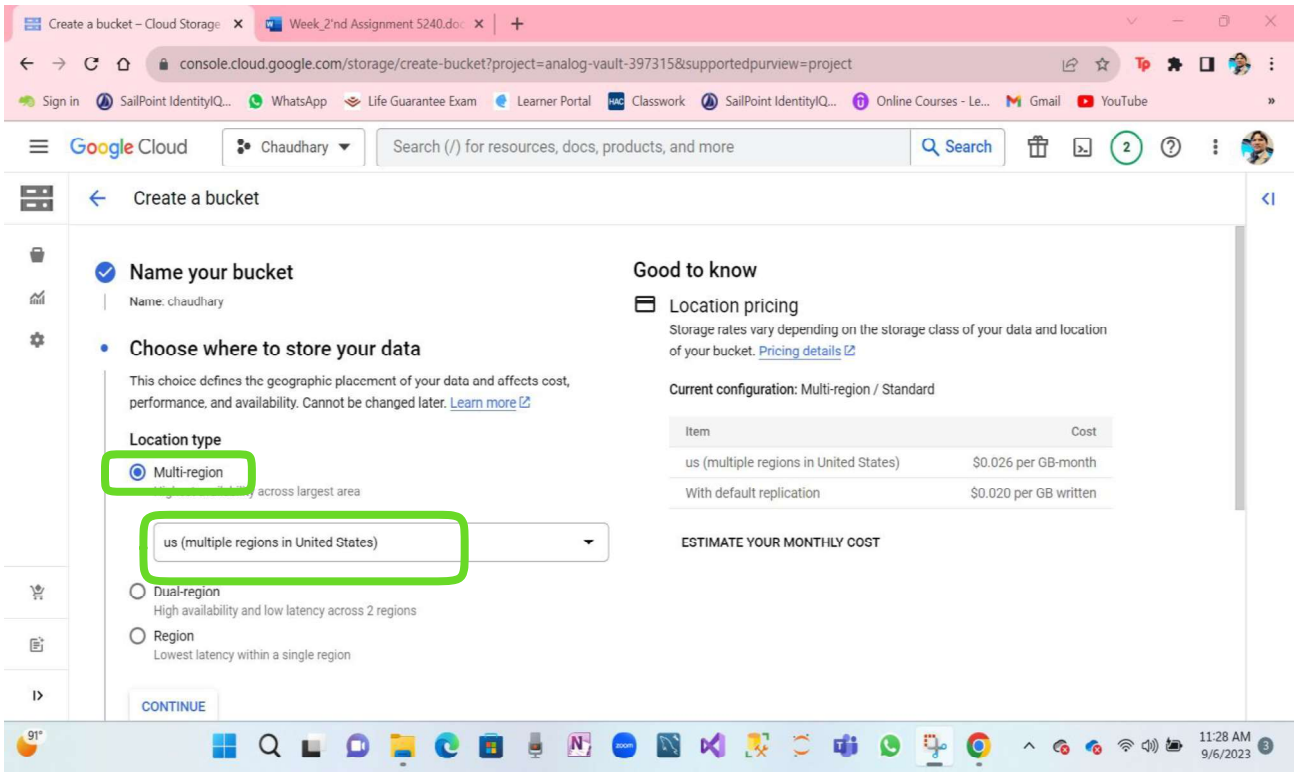
→ Name your bucket.



→ I named my bucket chaudhary without having any spaces/uppercase letters and clicked on continue.

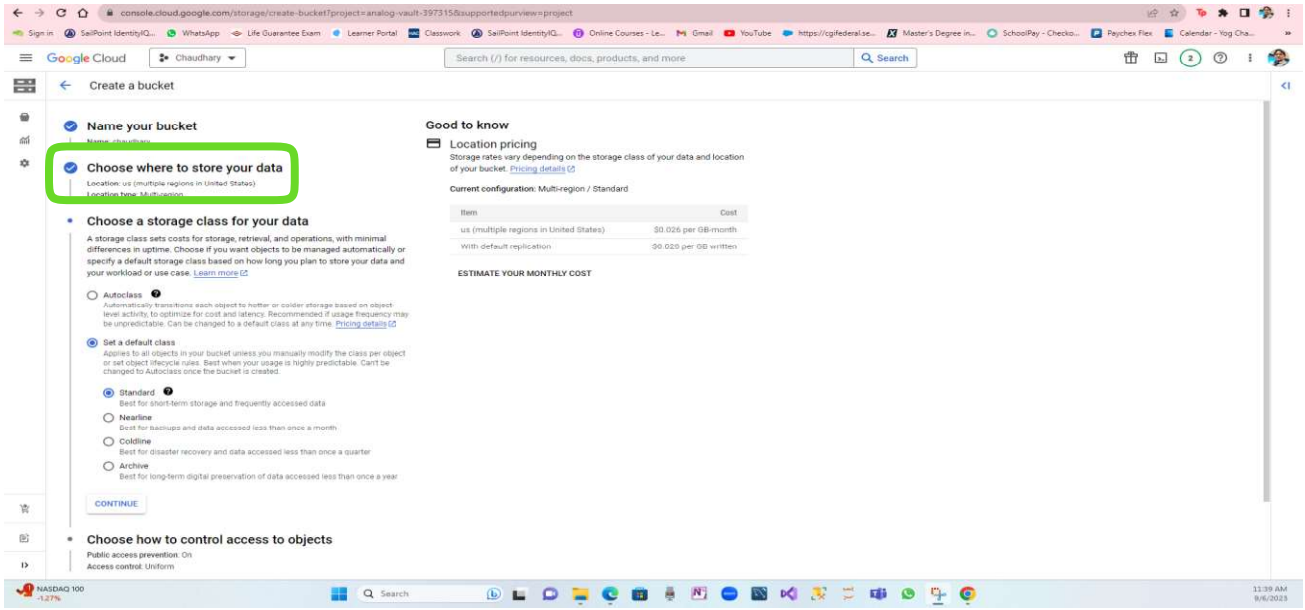


❖ **Choose where to store your data.**

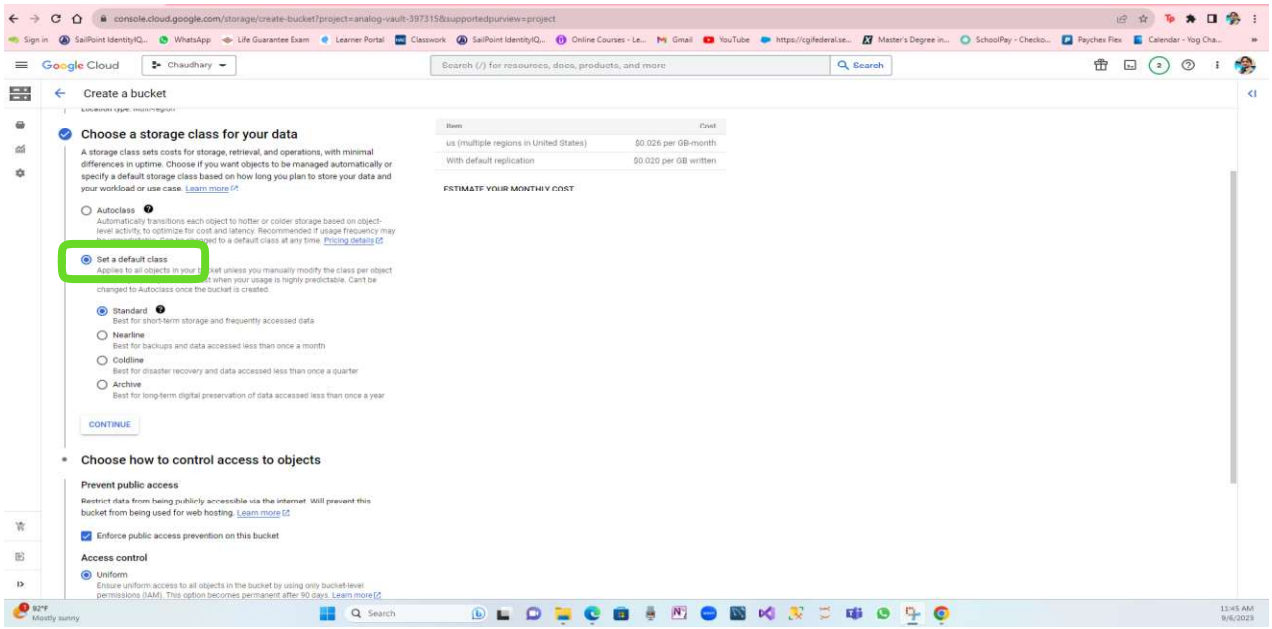


→ Keep “Choose where to store your data” as the

→ Click on Continue.

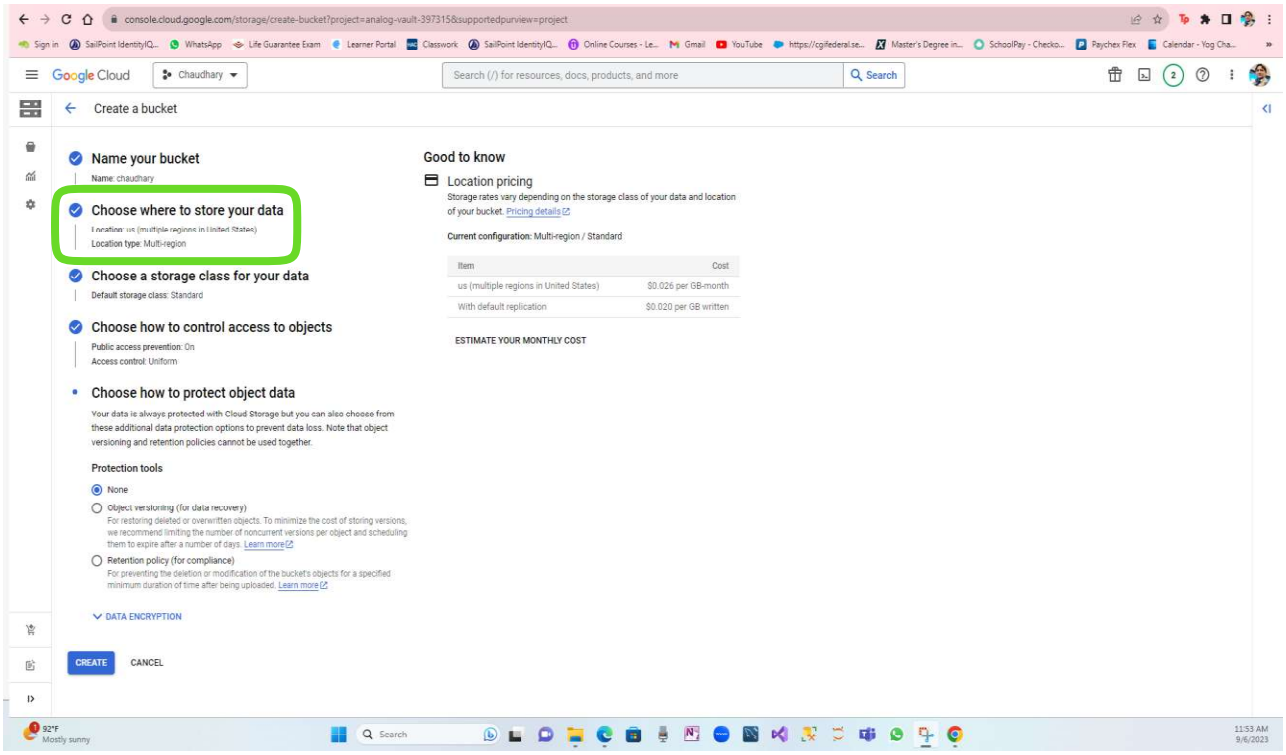


❖ **Keep "Choose a default storage class for your data" as the default "Standard."**

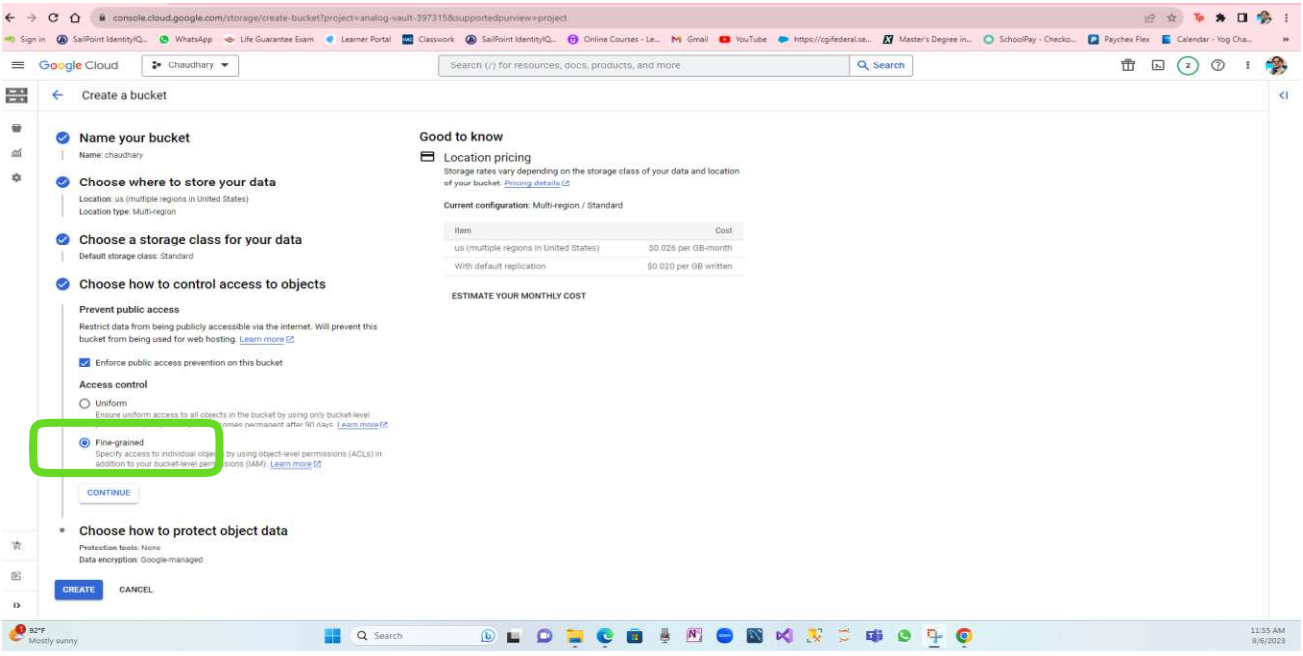


❖ **I have chosen my data as**

❖ **I am continued.**



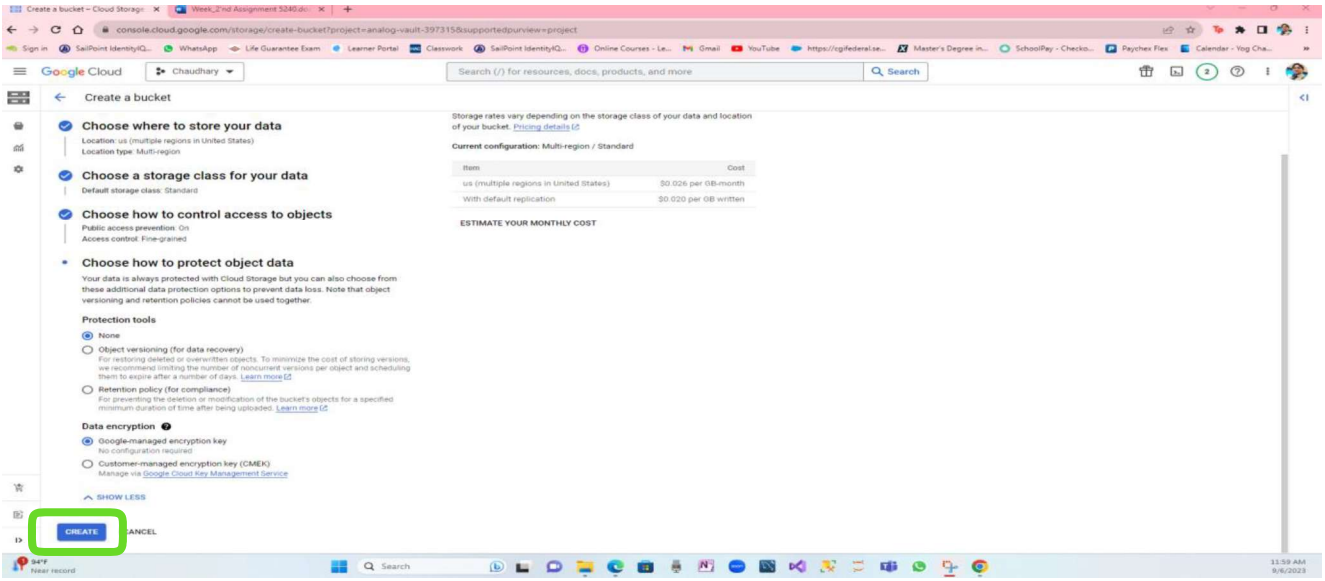
❖ **Choose how to control access to objects.**



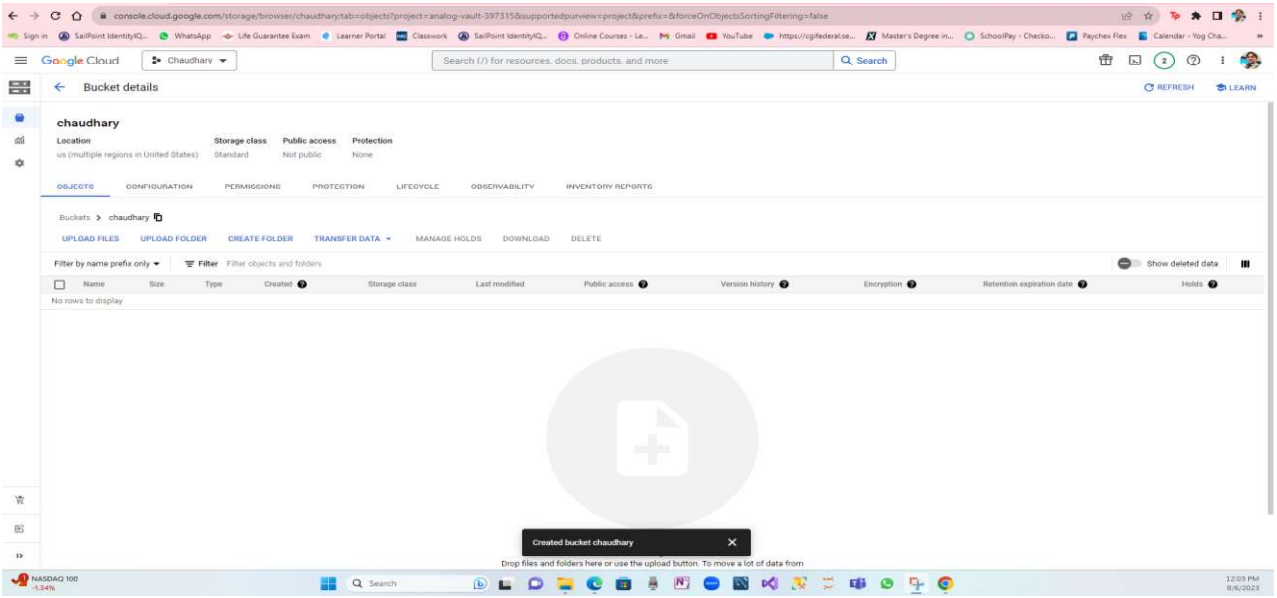
→ Then I chose fine-grained and clicked on continue.

❖ Then for advanced setting(option), I kept the default as a Google-managed key.

→ Then I clicked on create.



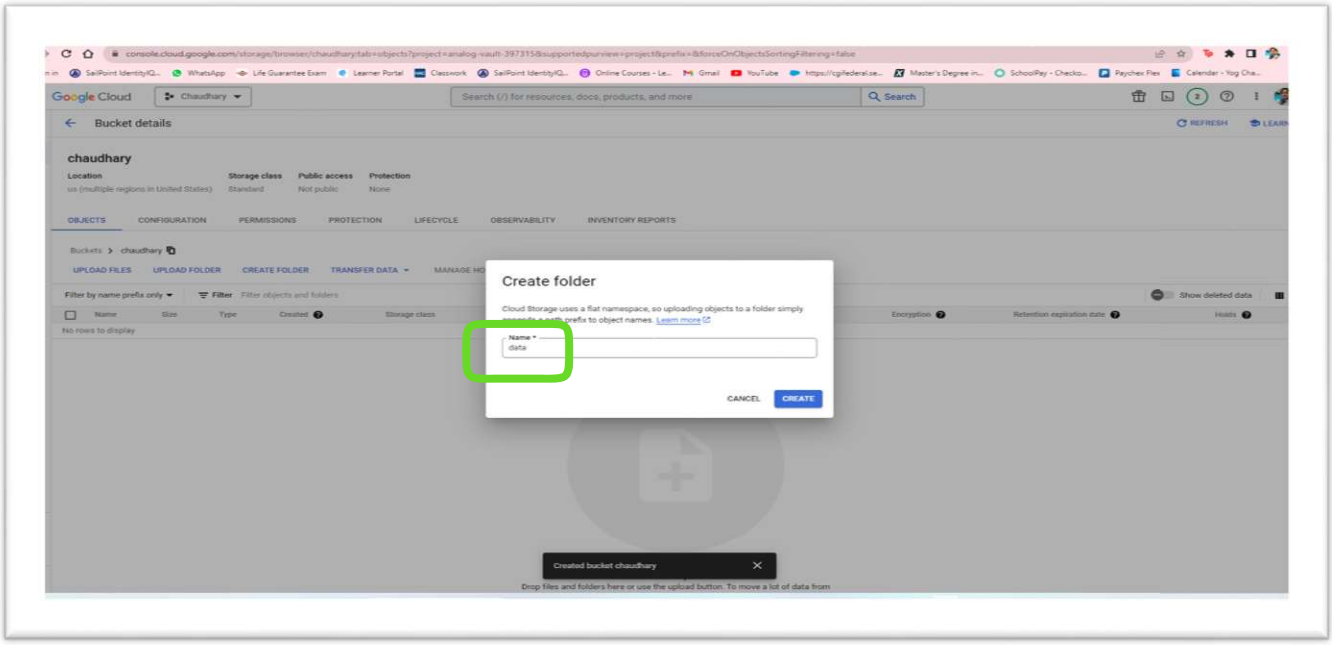
5. After clicking create, I was redirected to this page as shown below.



→ Now I must create 3 folders in this bucket such as **"data", "logs" and "outputs".**

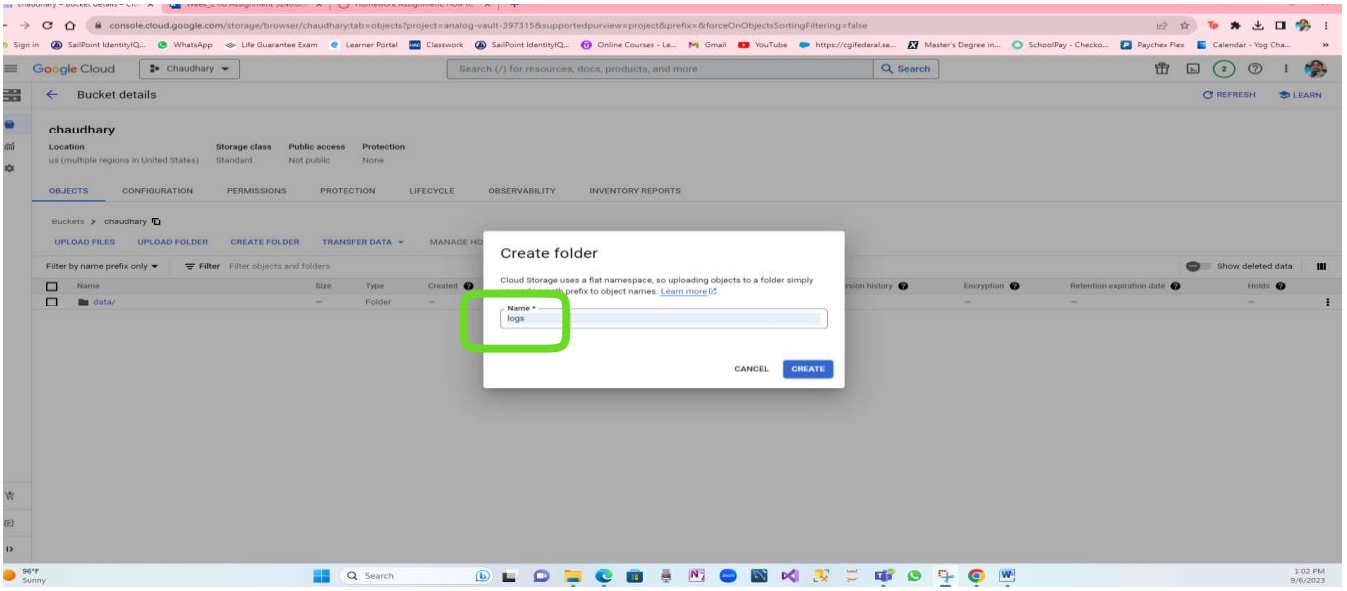
→ For this I have clicked on Created a folder.

→ After Entering the folder name as **"data,"** I clicked on create.



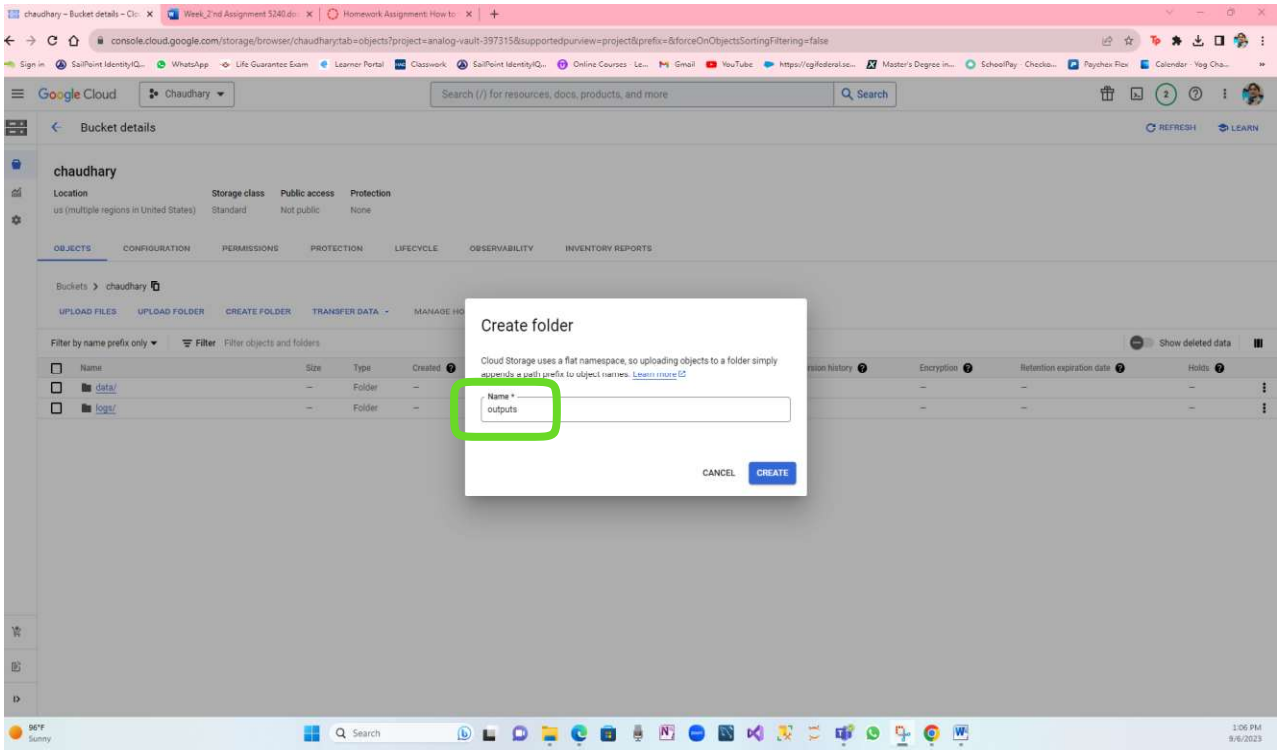
→ The same as followed for creating “logs”

→ Again, clicked on Create a folder, entered “logs” clicked.



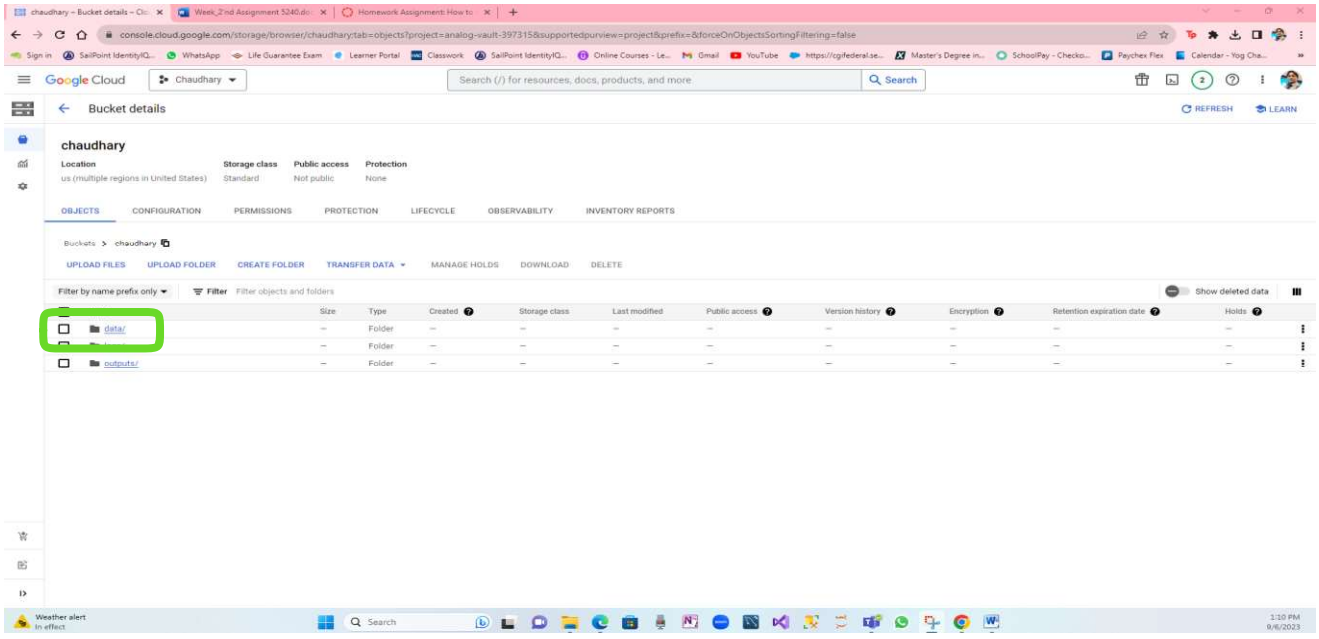
→ As following for creating “outputs.”

→ Clicked Entered “outputs.”



→ Hance, folders were created.

→ Below shows the screenshot that the three folders were created.



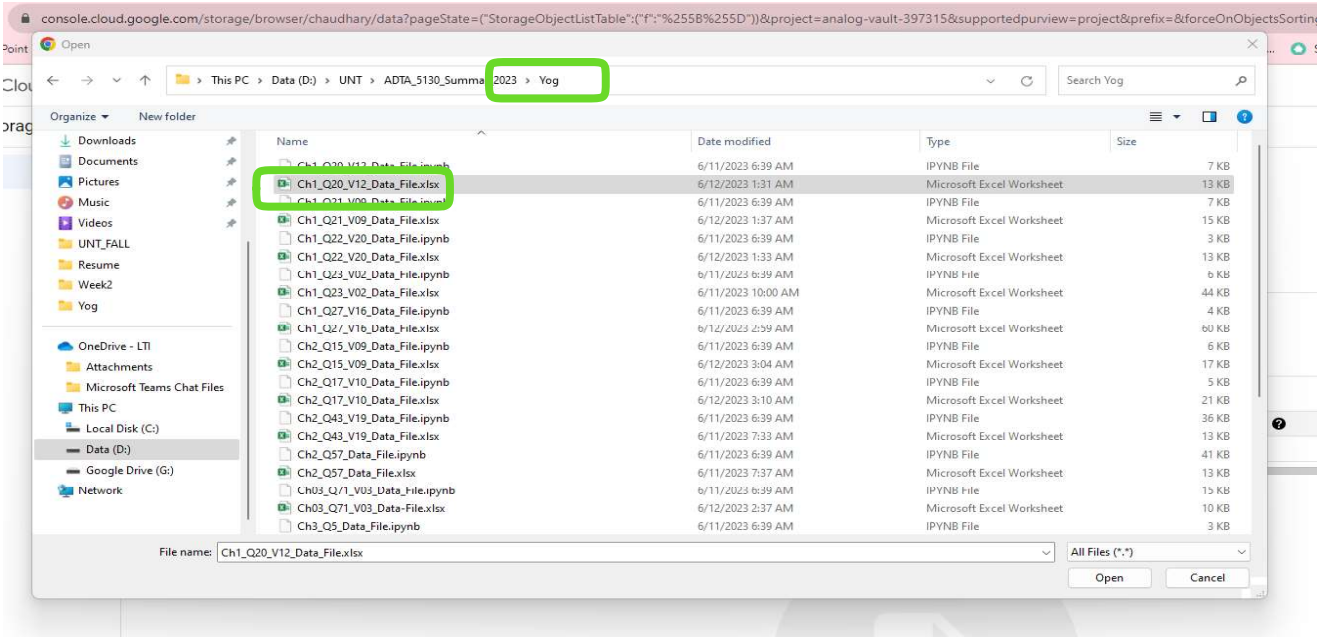
6. After this I must upload data to GCP storage buckets.

❖ **This is yog data. File.xlsx**

→ For this I have clicked on data.

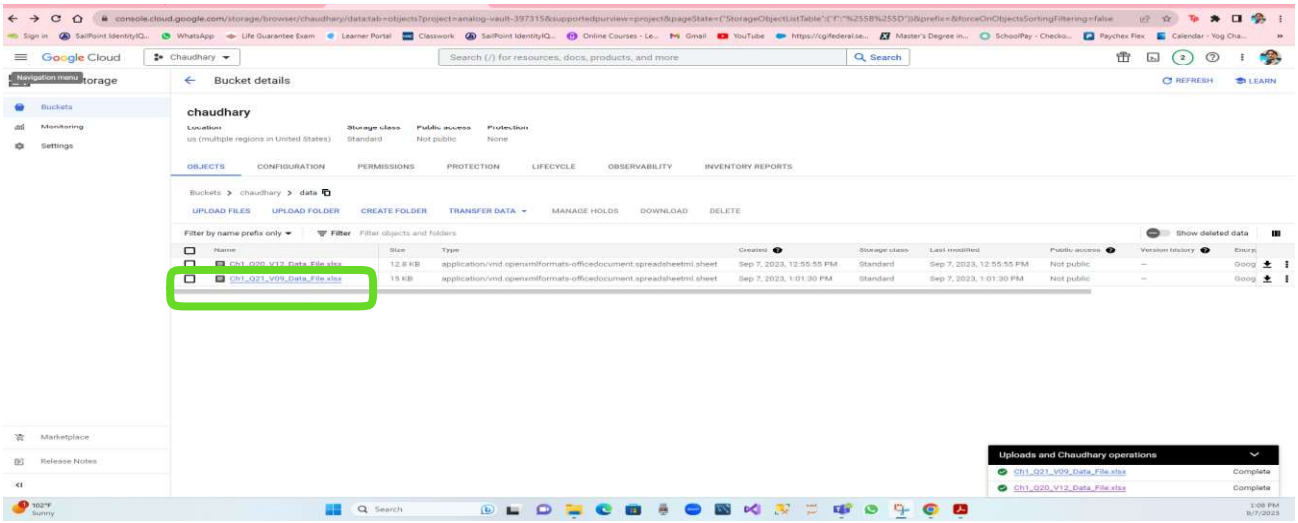
→ In this I clicked on update file.

→ After that, I browsed the file that must be uploaded which is user data File.xlsx



→ Again, I browed for another file that must be uploaded which is user Ch1_Q21_V09_Data_File

7. Hence, the two file are unloaded.



❖ **<u>How To Create a Hadoop and Cluster in GCP. Pdf.</u>**

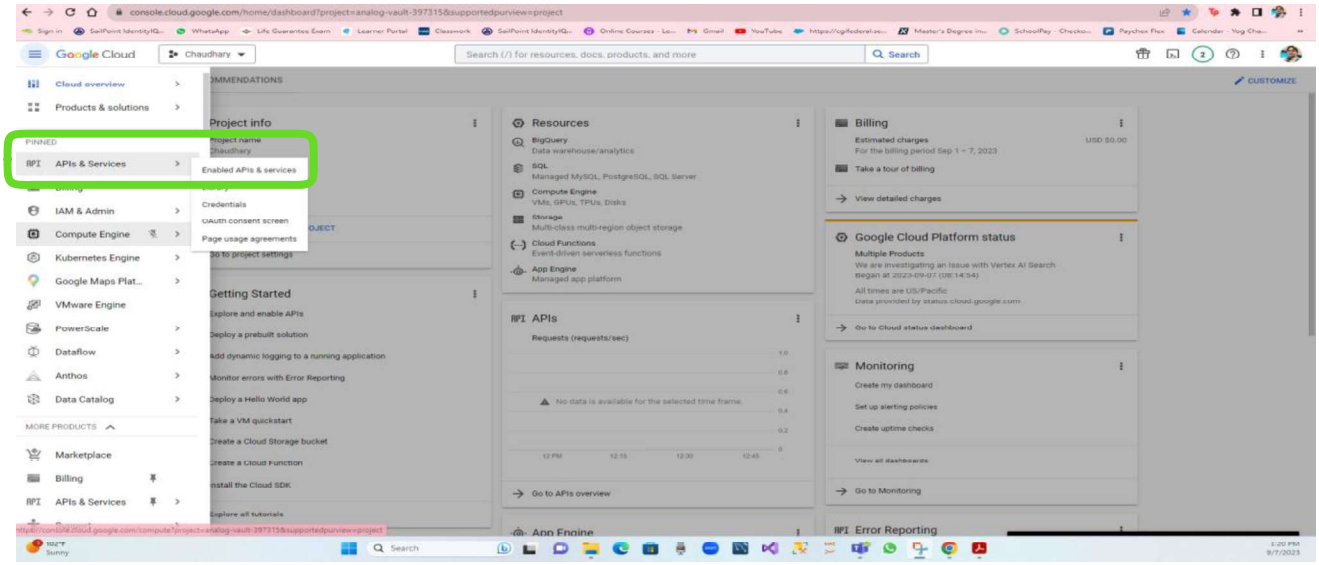
1. In this first I must enable GCP APIs (Compute Engine).

A) For this I must enable, and steps followed compute engine:

→ Clicked on the navigation menu

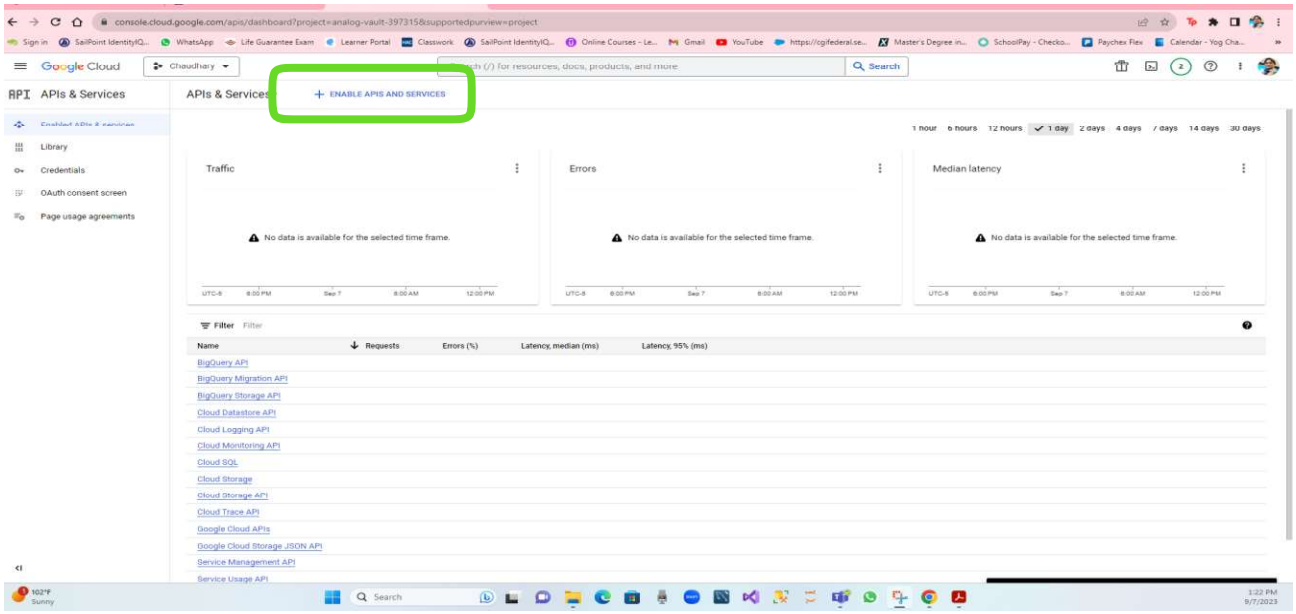
→ Scrolled down and clicked on API & services

→ Below shows the screenshot of the step I have followed.

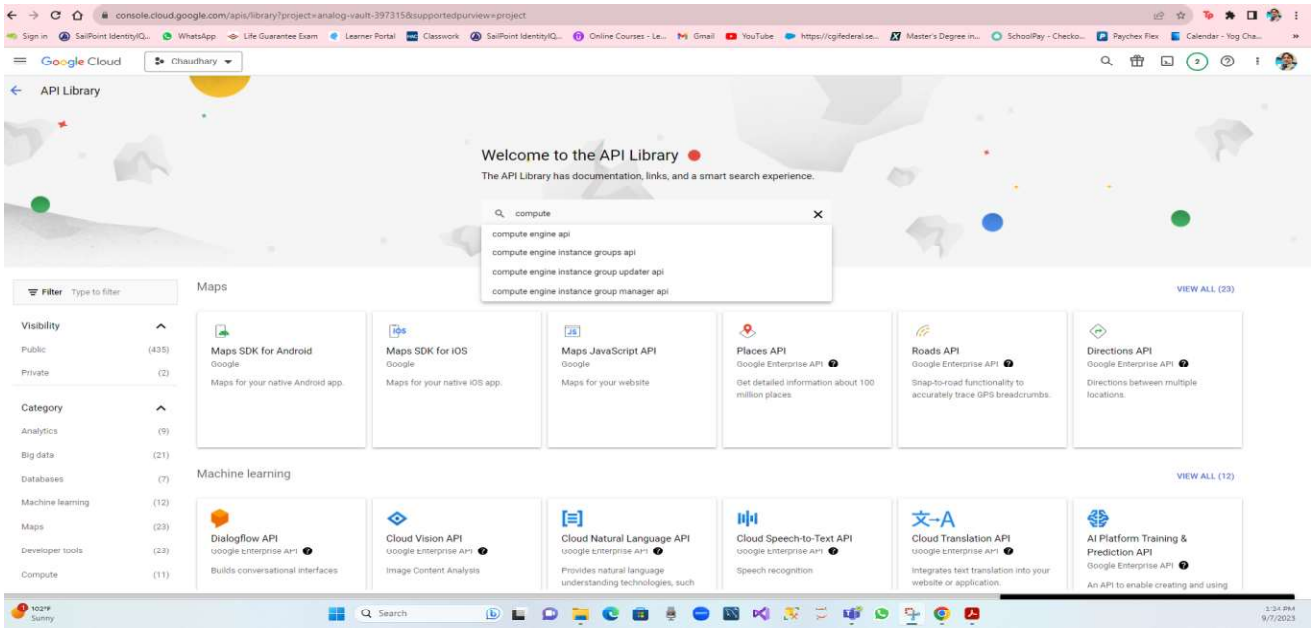


→ By clicking on APIs and services I was redirections to this page shows below.

→ Then I clicked on Enable APIs and Services.

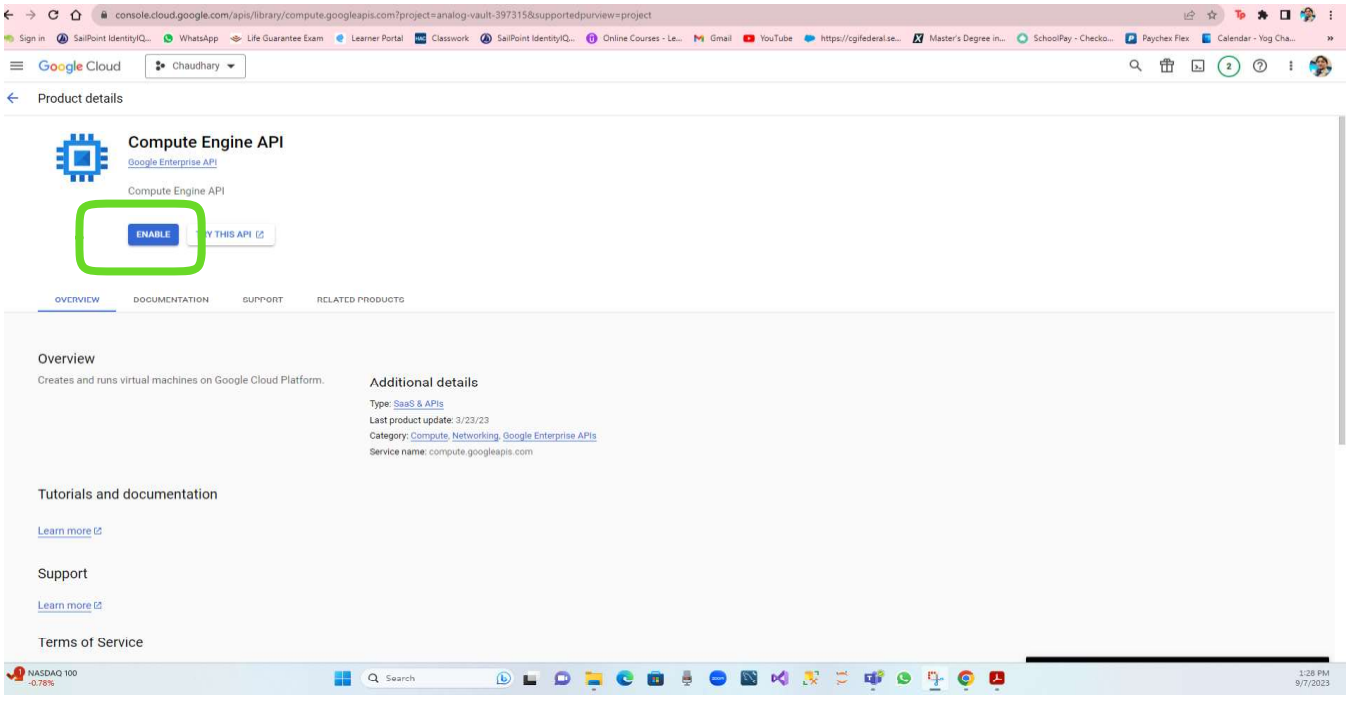


→ Gone to this page.



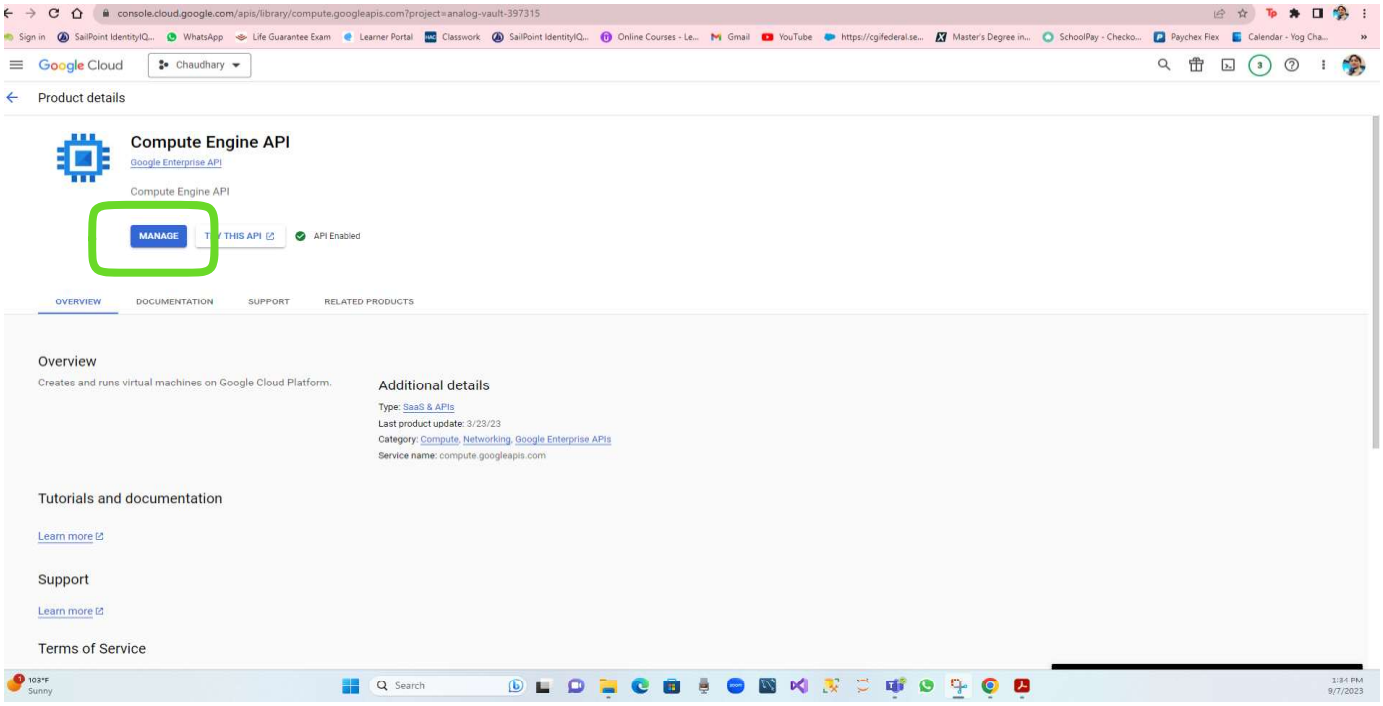
→ Then I entered.

→ By clicking, I was directly taken to this page, that is shown below

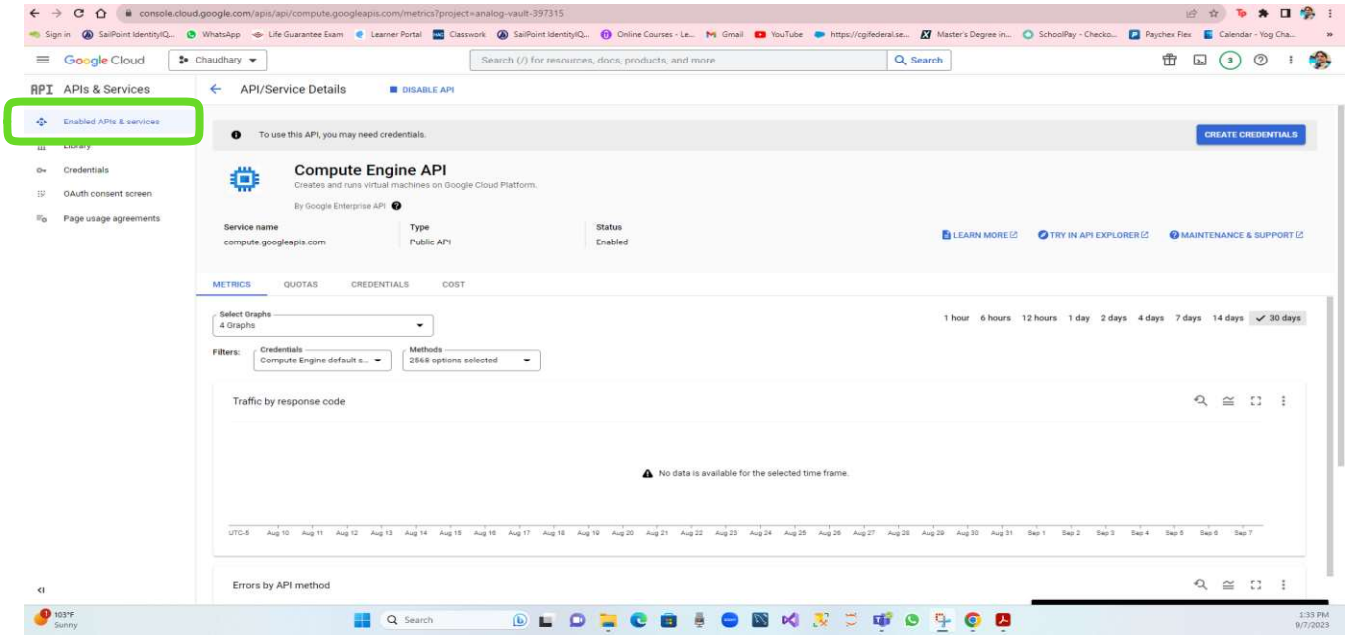


→ Then I clicked on enable.

→ Hare is the screenshot that shows my computer engine API has been successfully enabled.



→ Therefore, the screenshot has been successfully Compute Engine API and Services.



2. Now I must enable GCP APIs (DataProc).

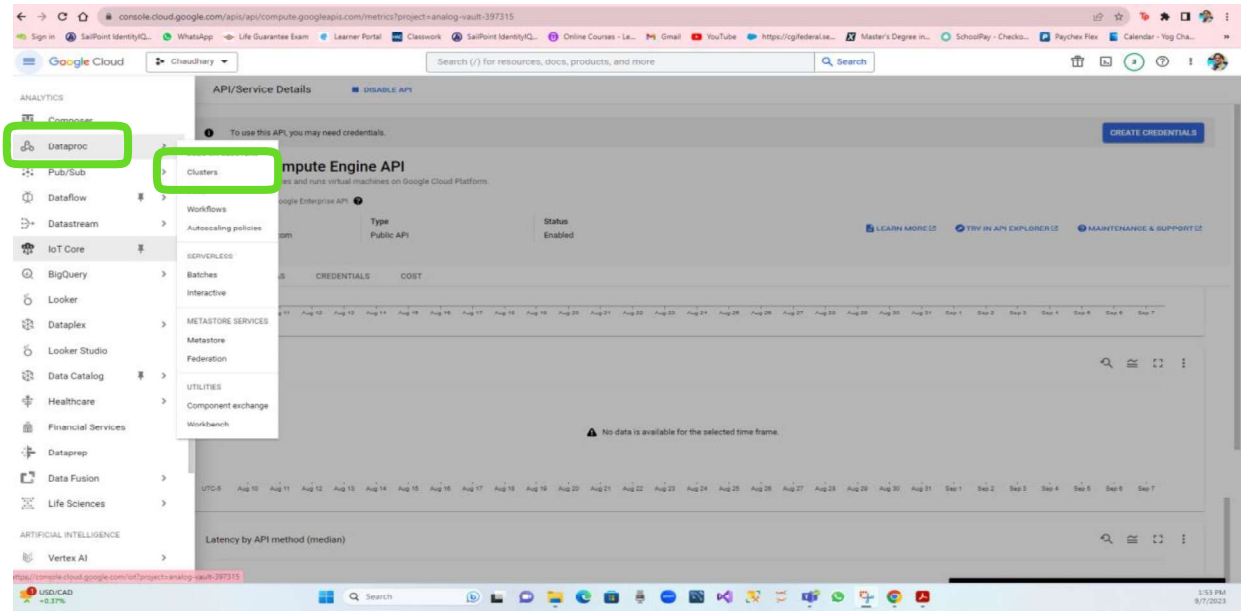
A) For this I must enable DataProc API.

B) Steps followed for creating DataProc API:

→ Scrolled down clicked on DataProc.

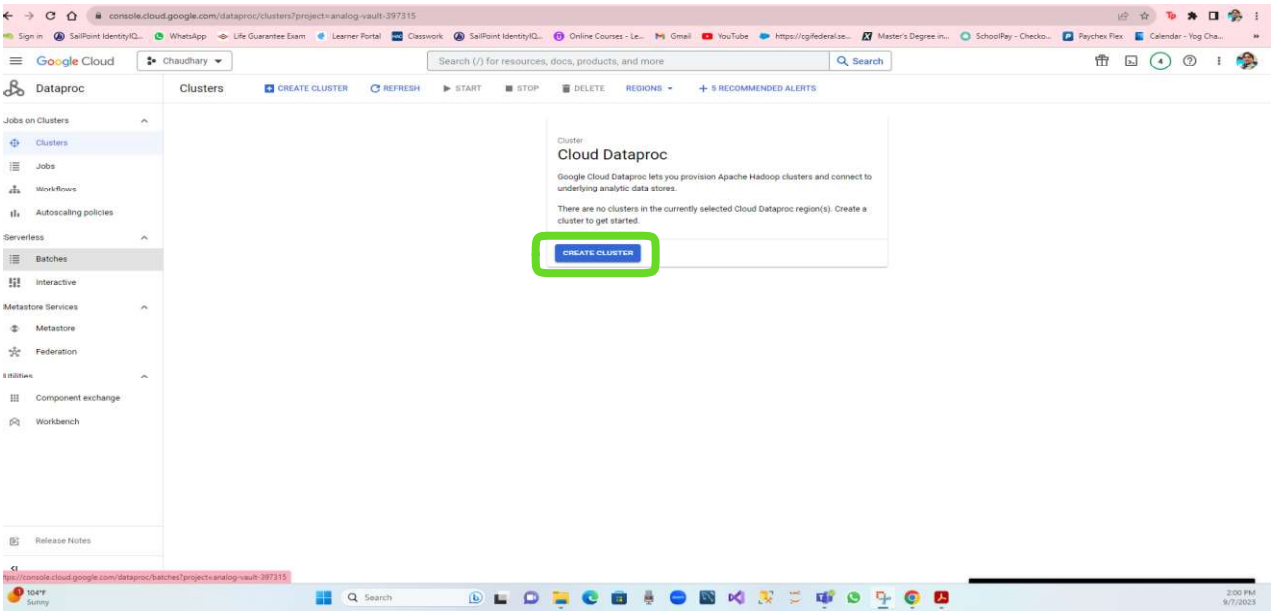
→ Click on clusters.

→ Below shows screenshot of the steps, I have followed.



→ After that I have taken to this page, then I clicked on create a cluster.

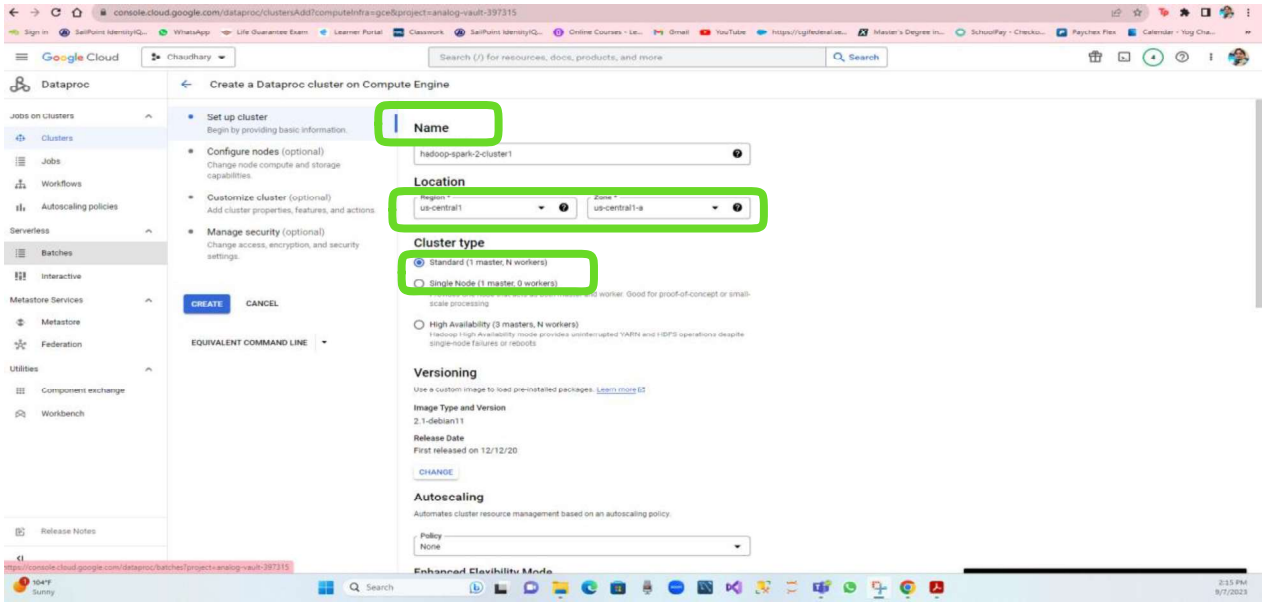
→ Below shows a screenshot



A) Now I must set up Cluster Cloud DataProc.

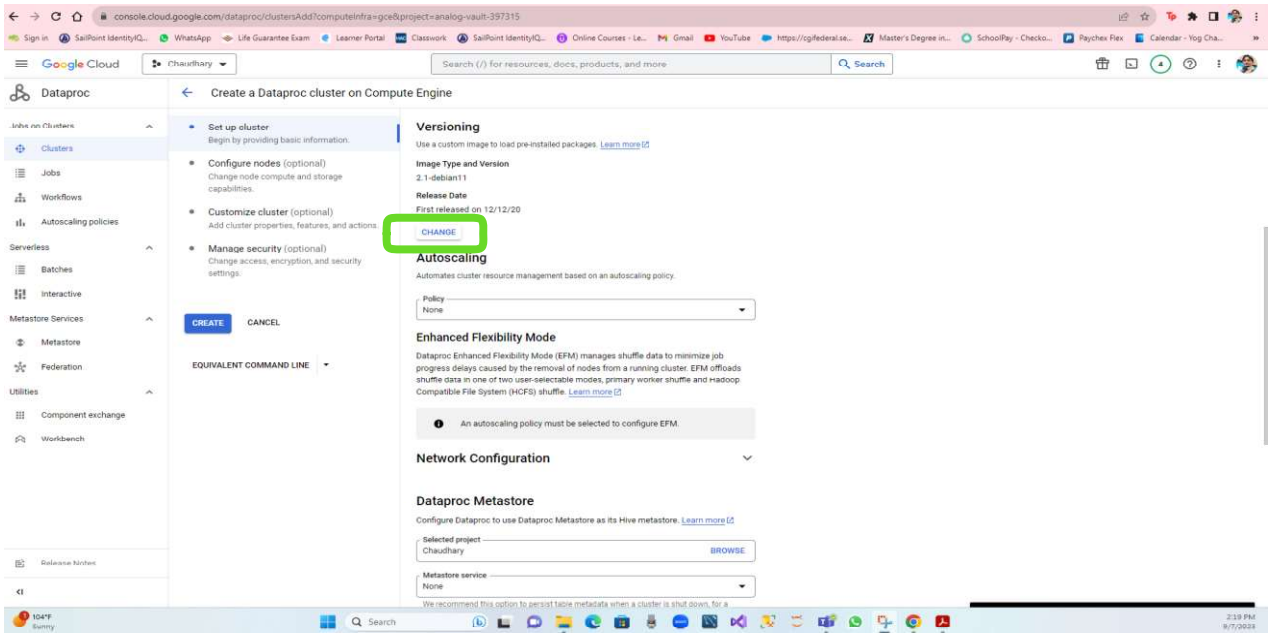
After clicking on create cluster, I have taken to this page, where I have entered.

→ Here is screenshot of this.



→ Now I should change the system version, for this, I must scroll down to “versioning” and click on change.

→ Hare is this screenshot.

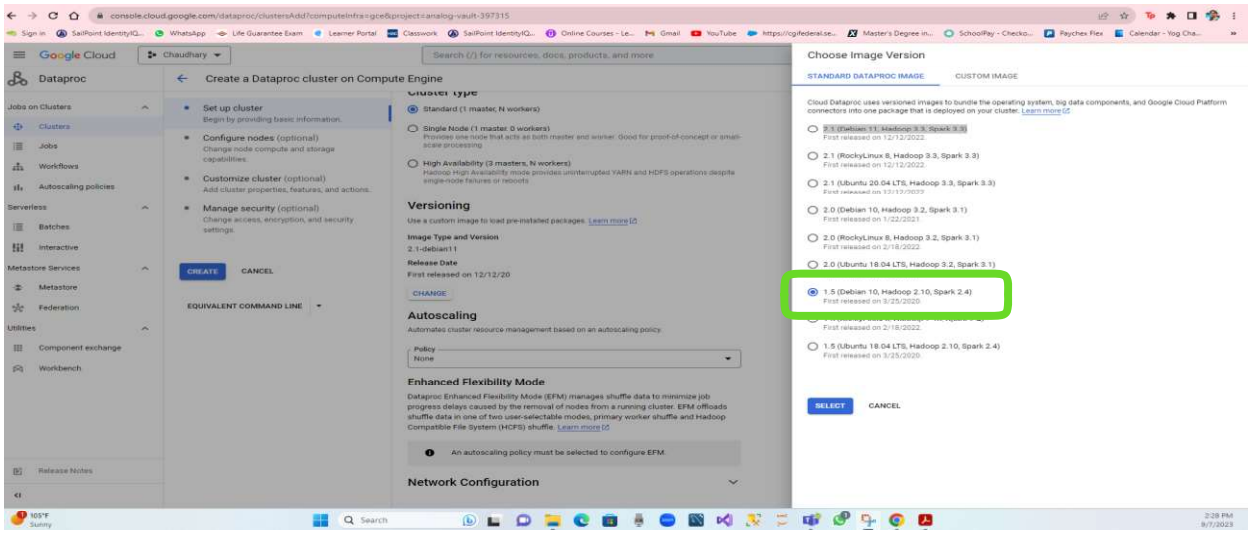


→ By clicking on the change, I was taken to this page.

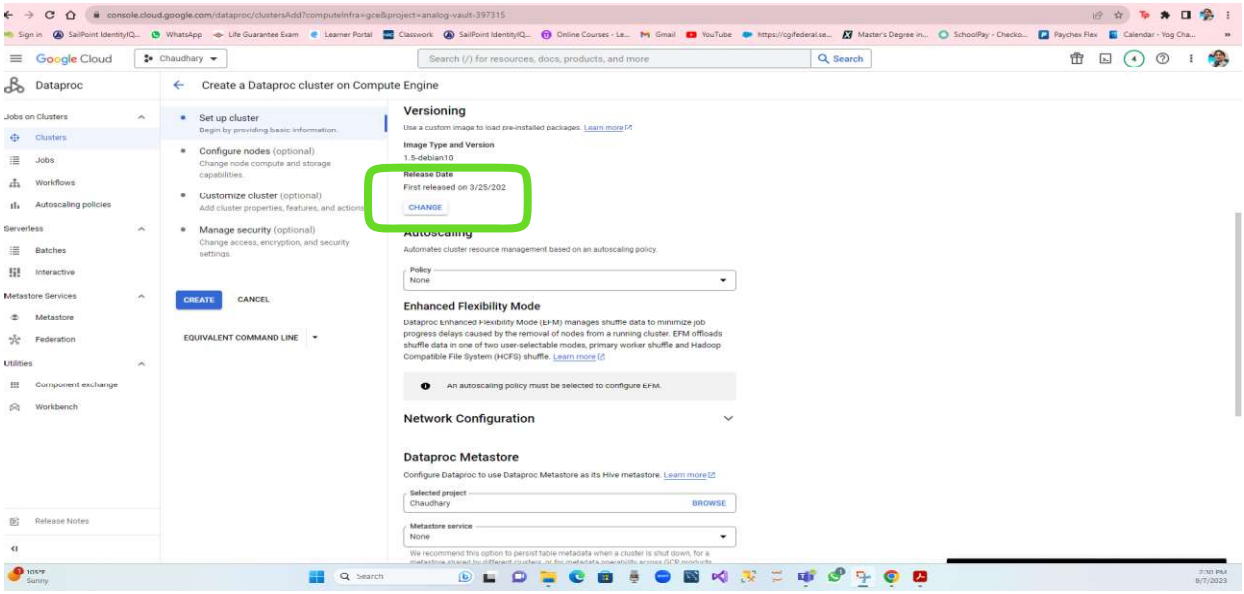
→ Leave “STANDARED DATAPROC IMAGE.”

→ I click on 2.1 (Debian 11, Hadoop 3.3, Spark 3.3) instead of

→ Below was screenshot.



→ Here the screenshot after the change.



A. Configure Master and Worker Nodes.

→ I clicked on configure nodes.

→ I selected “General Purpose” for the “machine family”

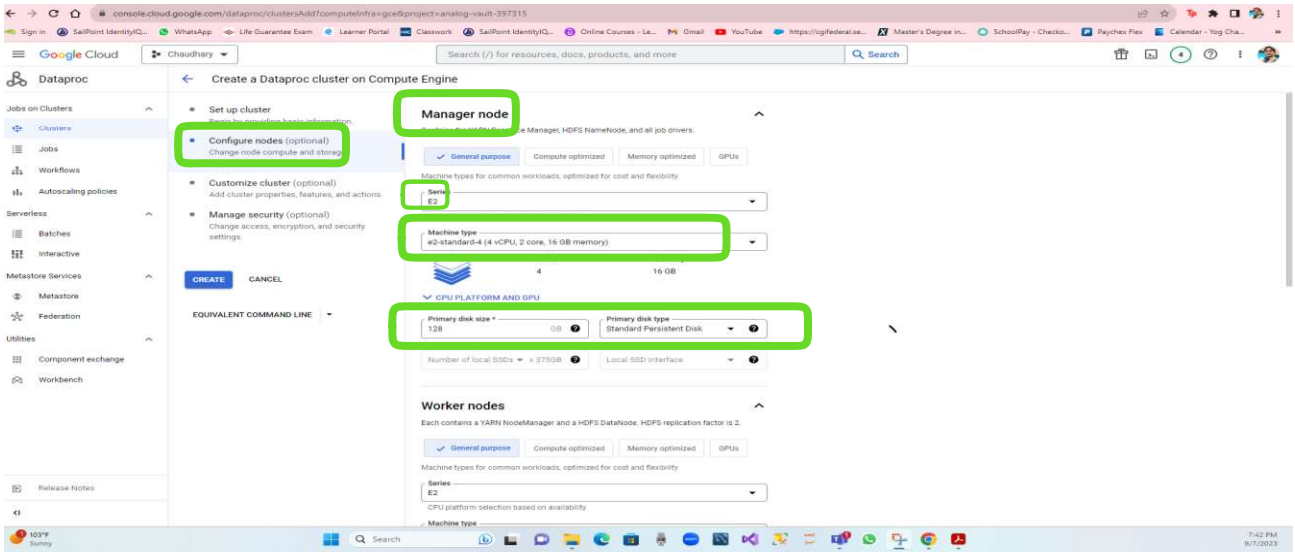
→ I selected “E2” for “series”.

→ I selected “e2-standard-8 (8 vCPUs, 16 GB memory)” for the machine type.

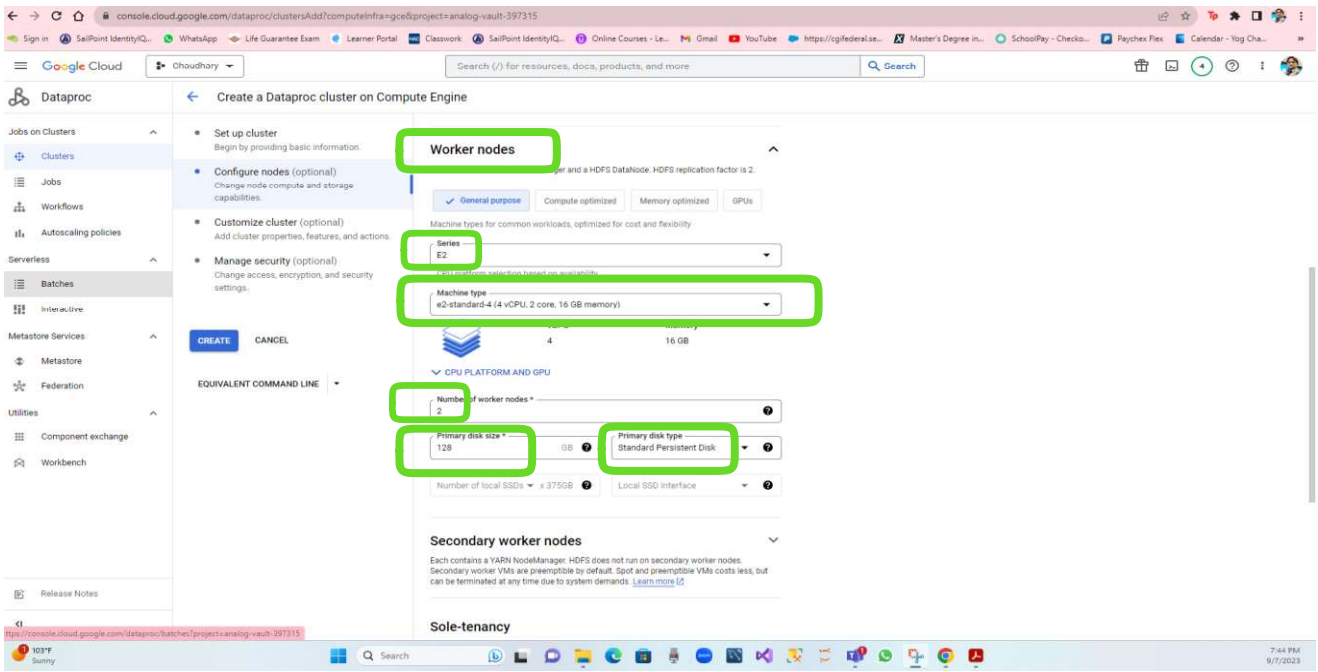
→ I selected “128 GB” for the “primary disk size (min 10GB)”.

→ I selected “standard persistent disk” for the “primary disk type”

→ Here is the screenshot.



→ After that, I scrolled down to the Workers nodes.
→ I selected “General Purpose” for the “machine family”
→ I selected “E2” for “series”.
→ I selected “e2-standard-4 (4 vCPUs, 16 GB memory)” for the machine type.
→ I kept “2” for the “number of worker nodes”.
→ I selected “128 GB” for the “primary disk size (min 10GB)”.
→ I selected “standard persistent disk” for the “primary disk type”
→ Here is screenshot.



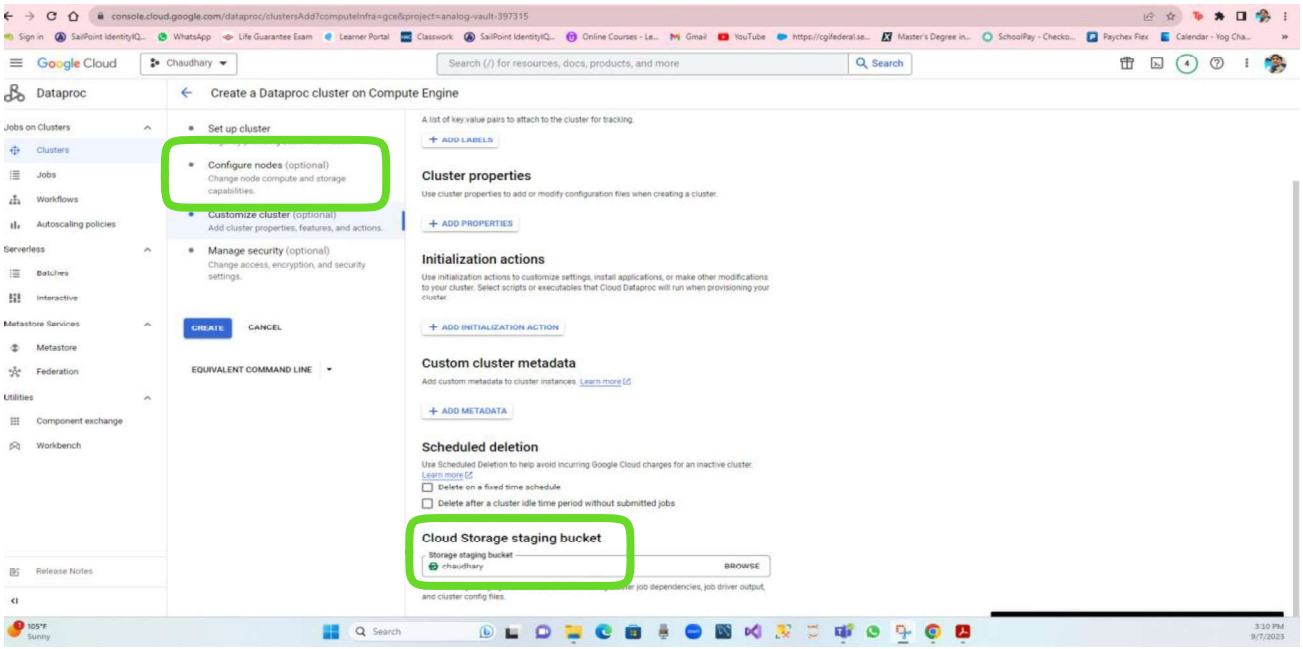
B. Customize cluster

→ Clicked on customize cluster.

→ Then which wan name chaudhary

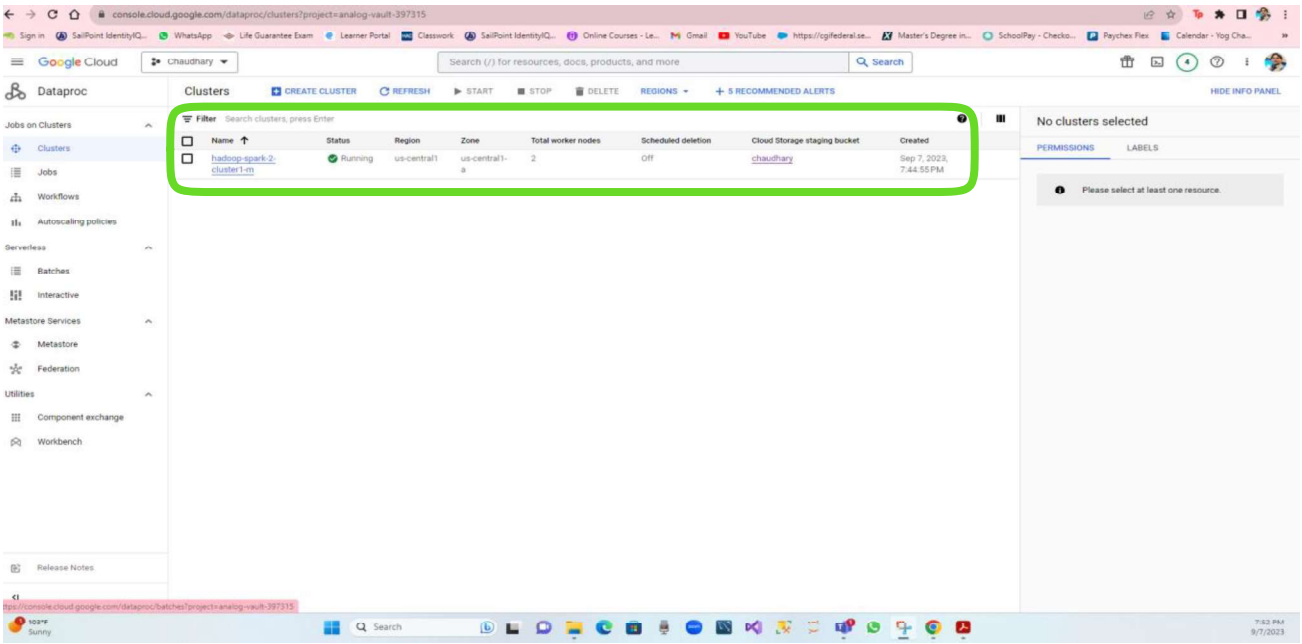
→ Then clicked on select.

→ It “chaudhary.”



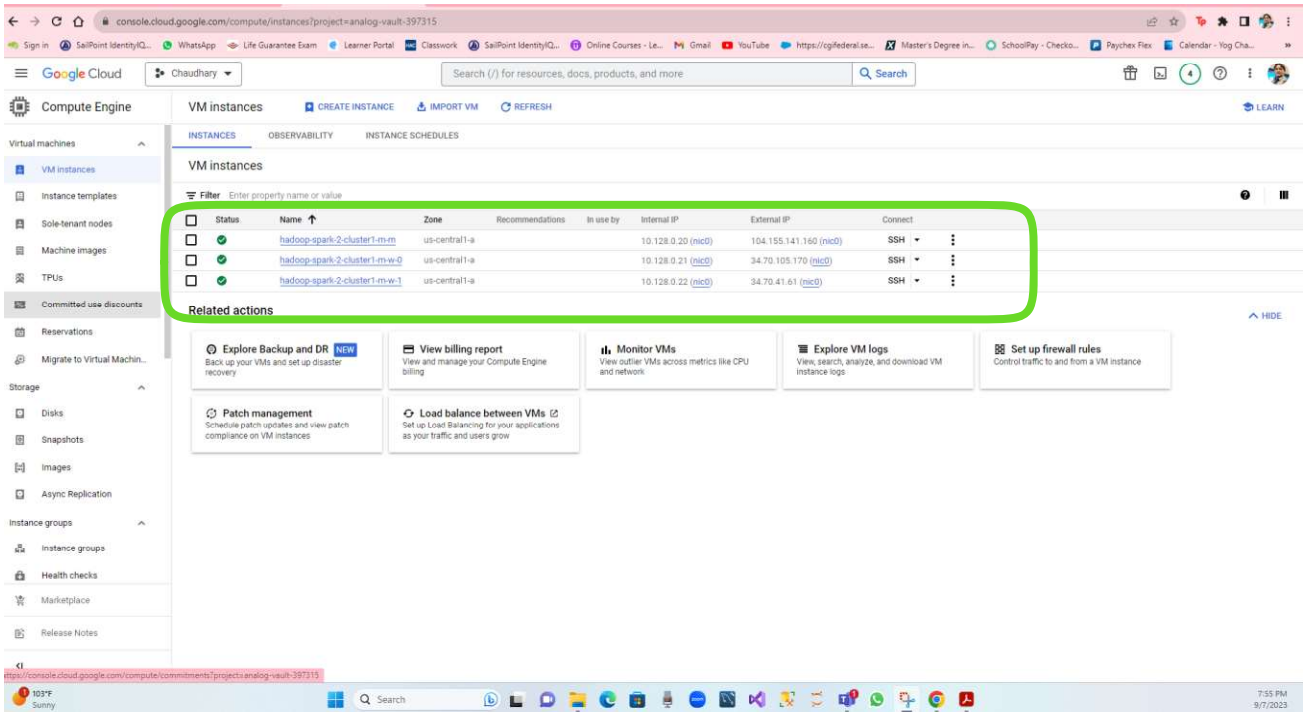
→ Then clicked on create which has taken me to the next page.

→ Below shows the screenshot of Hadoop-spark-2cluster1-m, I have created successfully. Where can be seen as new cluster was running.



→ Hence the new cluster: Haddop-spark-2-cluster, I have been successfully created. As we can see where the new cluster was running

→ Below is the screenshot that all three nodes are run.



A) Starting cluster nodes in GCP.

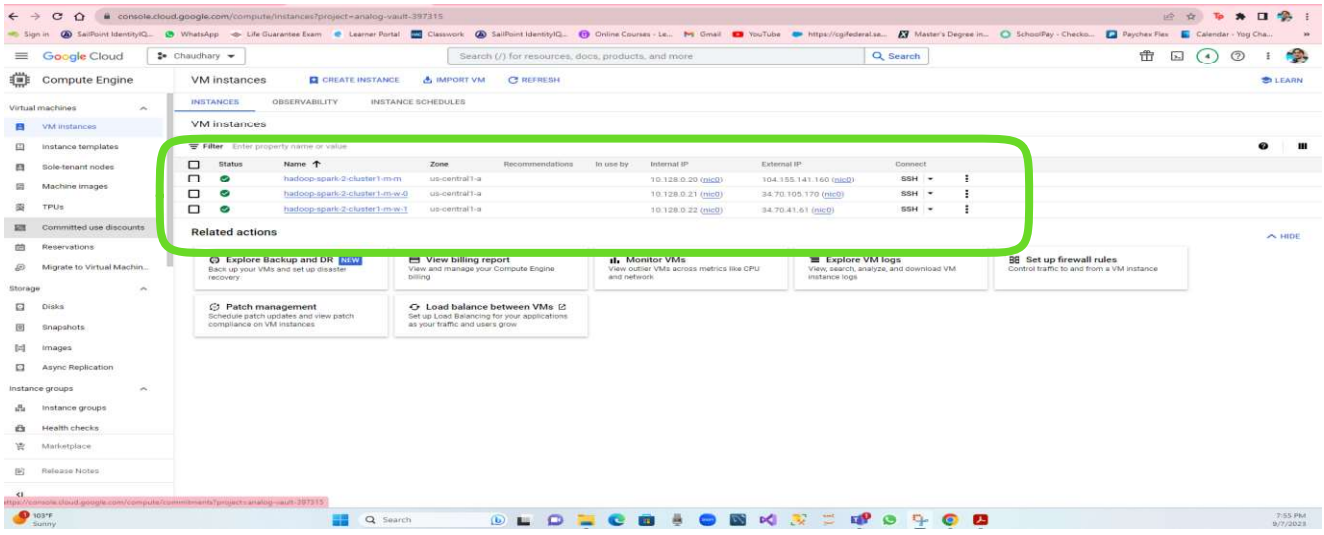
→ **T**o use the cluster nodes, we must start all the nodes of the cluster.

→ For this to start the master node we must click on the 3 vertical dots.

→ By clicking on that it will show the cluster start/resume tab.

→ By clicking on start/resume it will show start, click on that, we will see nodes was running.

→ Repeat the same thing for other two nodes.



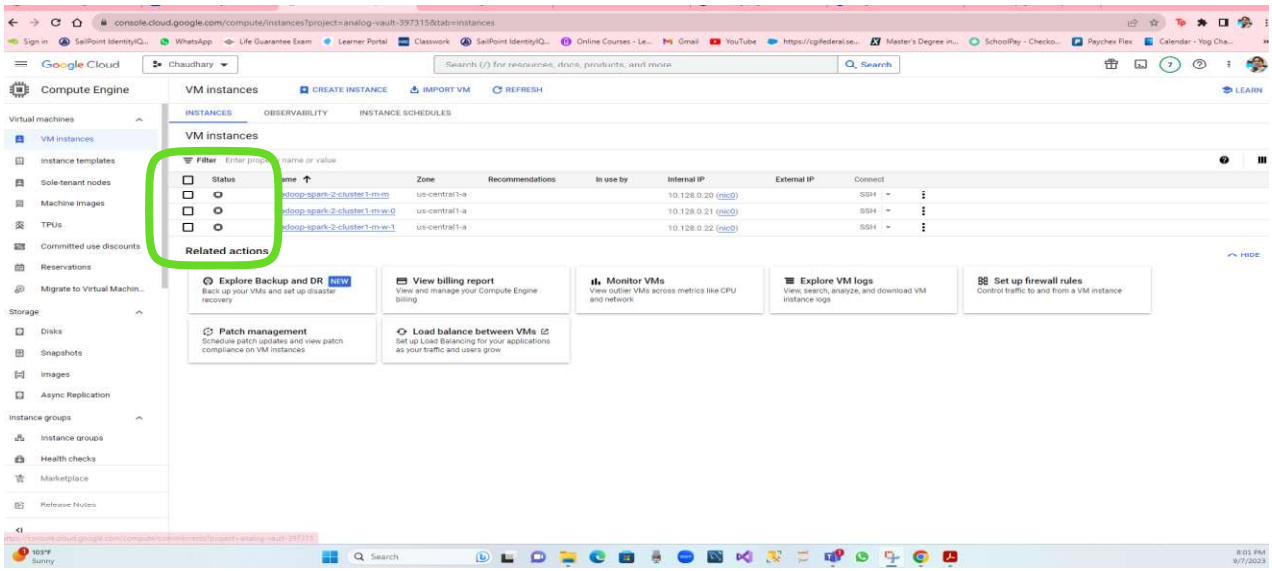
→ Here is the screenshot that shows all the three nodes are running.

B) Stopping cluster nodes in GCP.

→ For this to stop the master node we must click on the 3 vertical dots.

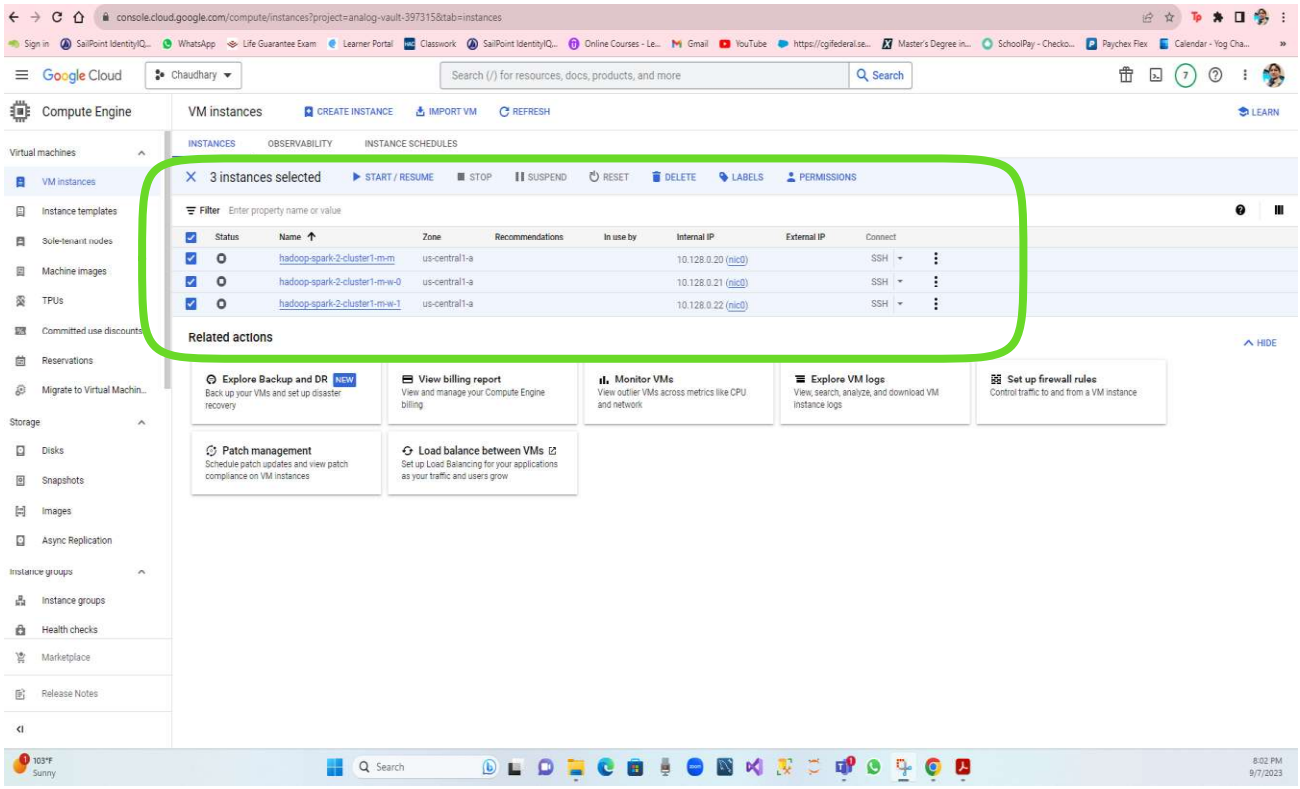
→ By clicking on that it will show the cluster stop tab.

→ By clicking on stop it will show stop, click on that, we will see node stopped running. Repeat the same thing for other two nodes.



→ Below is the screenshot that all three nodes are stopped.

❖ Finally, successfully VM instances is that 3 nodes are Start/Resume, Stop, Suspended, reset, Delete and Permission.


console.cloud.google.com/compute/instances?project=analog-vault-397315&tab=instances
Google Cloud
Chaudhary
Search (/) for resources, docs, products, and more
Search
Compute Engine
VM instances
CREATE INSTANCE
IMPORT VM
REFRESH
LEARN
INSTANCES
OBSERVABILITY
INSTANCE SCHEDULES
3 instances selected
START / RESUME
STOP
SUSPEND
RESET
DELETE
LABELS
PERMISSIONS
Filter
Enter property name or value
Status
Name
Zone
Recommendations
In use by
Internal IP
External IP
Connect
hadoop-spark-2-cluster1-m-m
us-central1-a
10.128.0.20 (nic0)
SSH
hadoop-spark-2-cluster1-m-w-0
us-central1-a
10.128.0.21 (nic0)
SSH
hadoop-spark-2-cluster1-m-w-1
us-central1-a
10.128.0.22 (nic0)
SSH
Related actions
HIDE
Explore Backup and DR
NEW
Back up your VMs and set up disaster recovery
View billing report
View and manage your Compute Engine billing
Monitor VMs
View outlier VMs across metrics like CPU and network
Explore VM logs
View, search, analyze, and download VM instance logs
Set up firewall rules
Control traffic to and from a VM instance
Patch management
Schedule patch updates and view patch compliance on VM instances
Load balance between VMs
Set up Load Balancing for your applications as your traffic and users grow
Virtual machines
VM instances
Instance templates
Sole-tenant nodes
Machine images
TPUs
Committed use discounts
Reservations
Migrate to Virtual Machin...
Storage
Disks
Snapshots
Images
Async Replication
Instance groups
Instance groups
Health checks
Marketplace
Release Notes
103°F
Sunny
Search
8:02 PM
9/7/2023