Yog Chaudhary

11727095

ADTA 5240 Week 3'rd (harvesting, Storing, And Retrieving Data)

Professor: Dr. Zeynep Orhan **Sep 15, 2023**

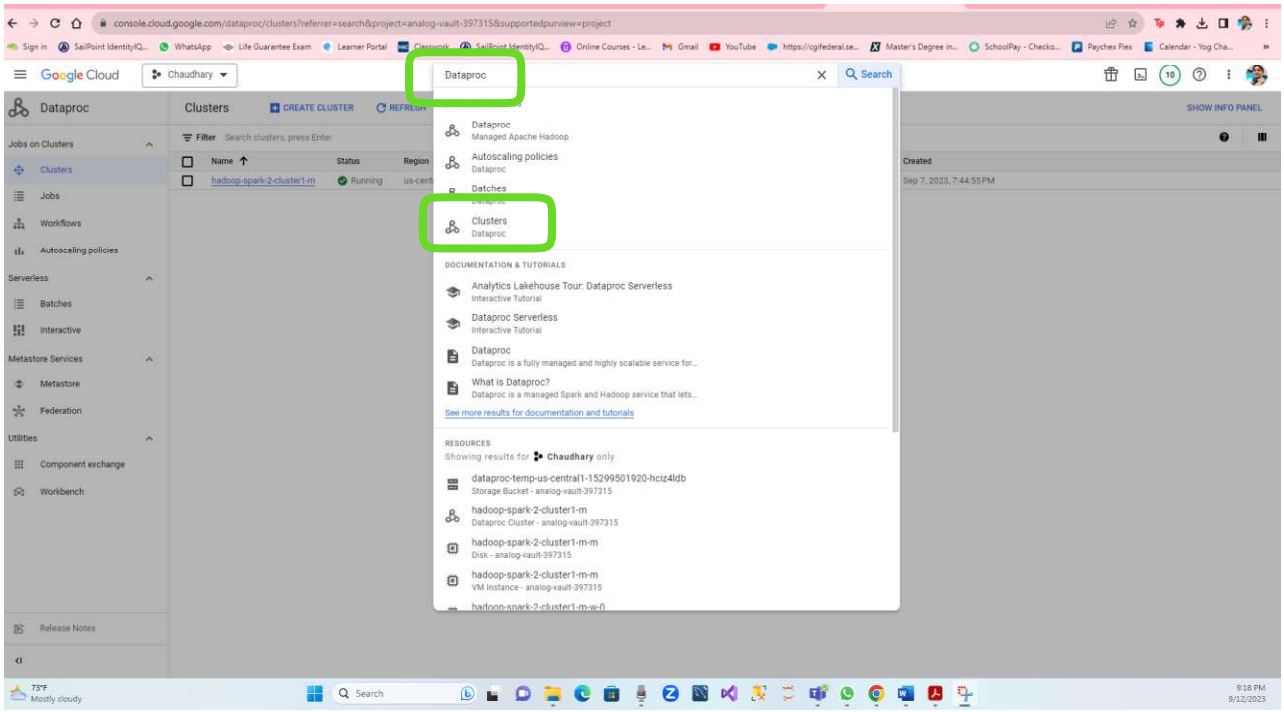
University Of North Texas

Exploring Remote Virtual Machine in the Cloud. Hadoop Ecosystem with Simple Linux Commands. Pdf

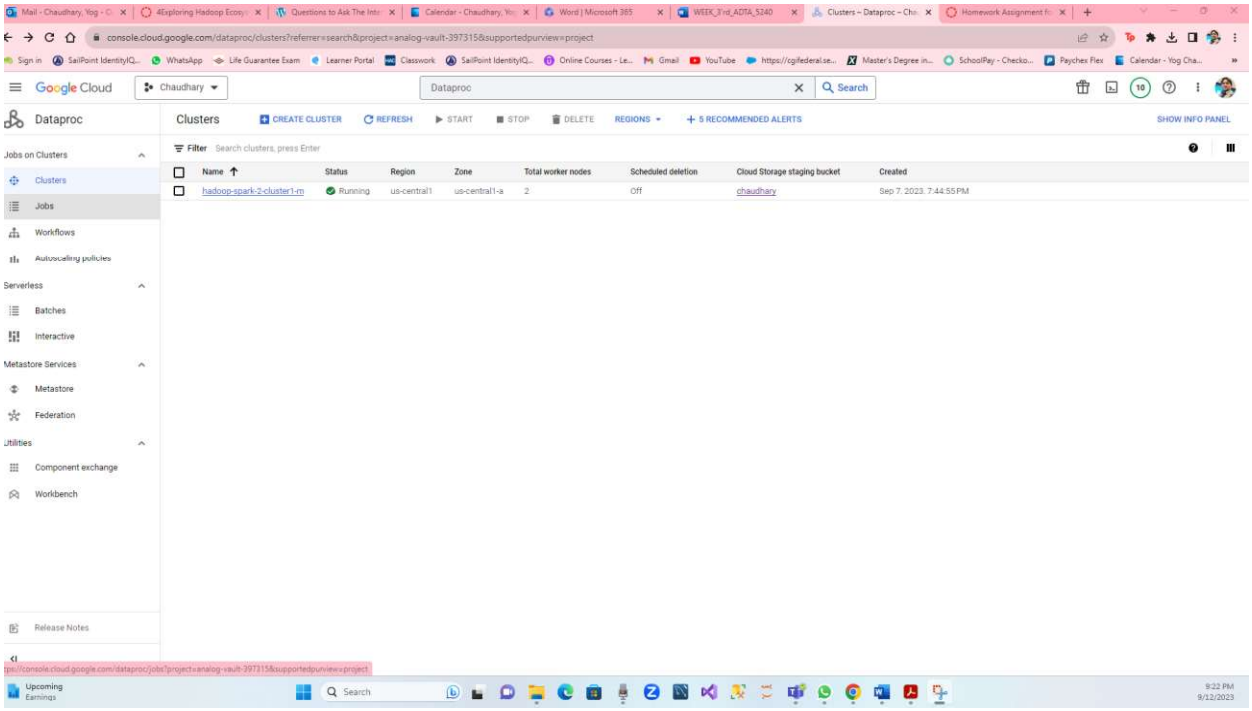
- ❖ **Monitoring the cluster in GCP:**

1. **For setting up a virtual machine will Linux as the operating system, I have signed into a Google GCP console.**

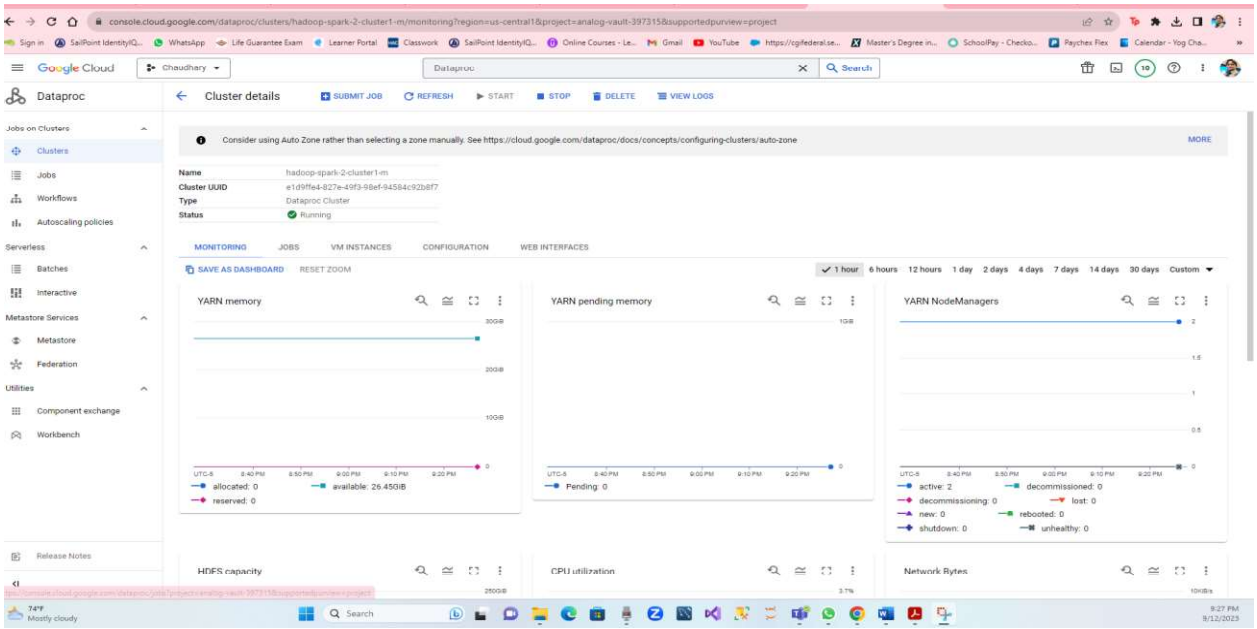
- **Then I clicked on three horizontal lines.**
- **In the search bar I typed "Cluster" and clicked on "Clusters Dataproc."**
- **Hare is screenshot**



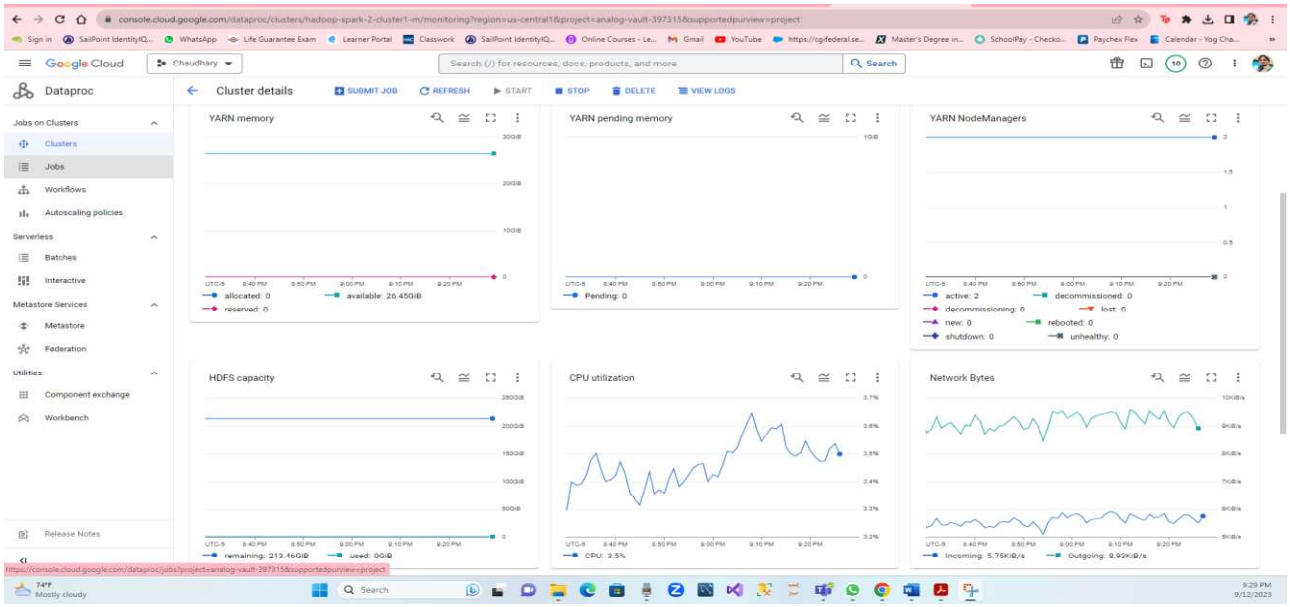
- **After clicking on clusters data proc it will mention the cluster that I have previously created, and it will show that it was running (with a green check mark).**
- **Below shows a screenshot of that.**



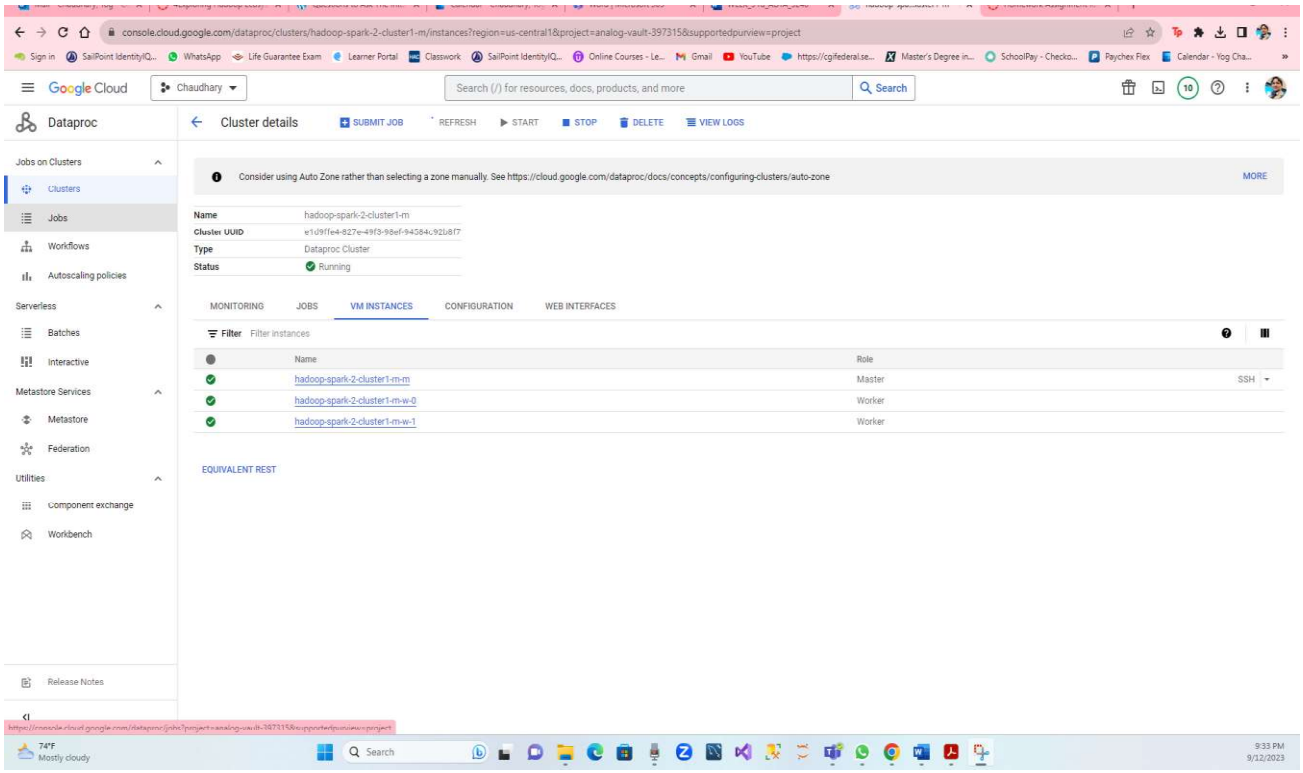
- **While doing this we need to start nodes by checking on the navigation panel.**
- **Clicked on computing engine.**
- **Then clicked on three vertical dots and clicked on start/resume.**
- **Now click on the cluster, and we can see a dashboard that monitors the cluster.**
- **Below shows the screenshot of that.**



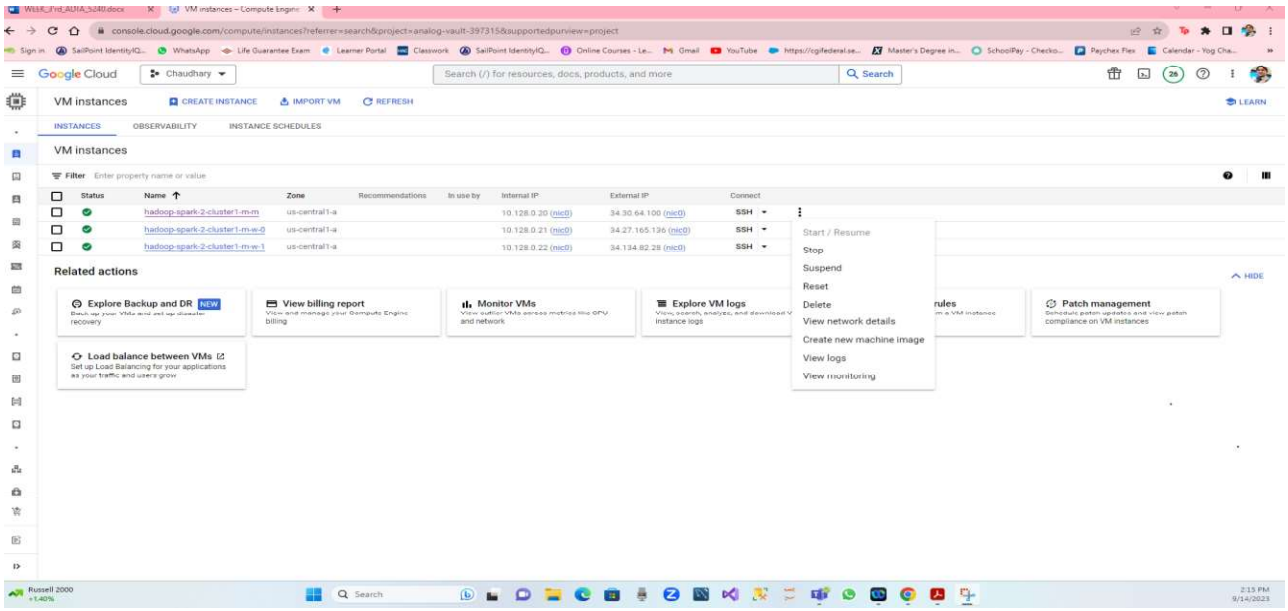
- **When I scroll down it will show how the monitoring is changing according to the usage.**



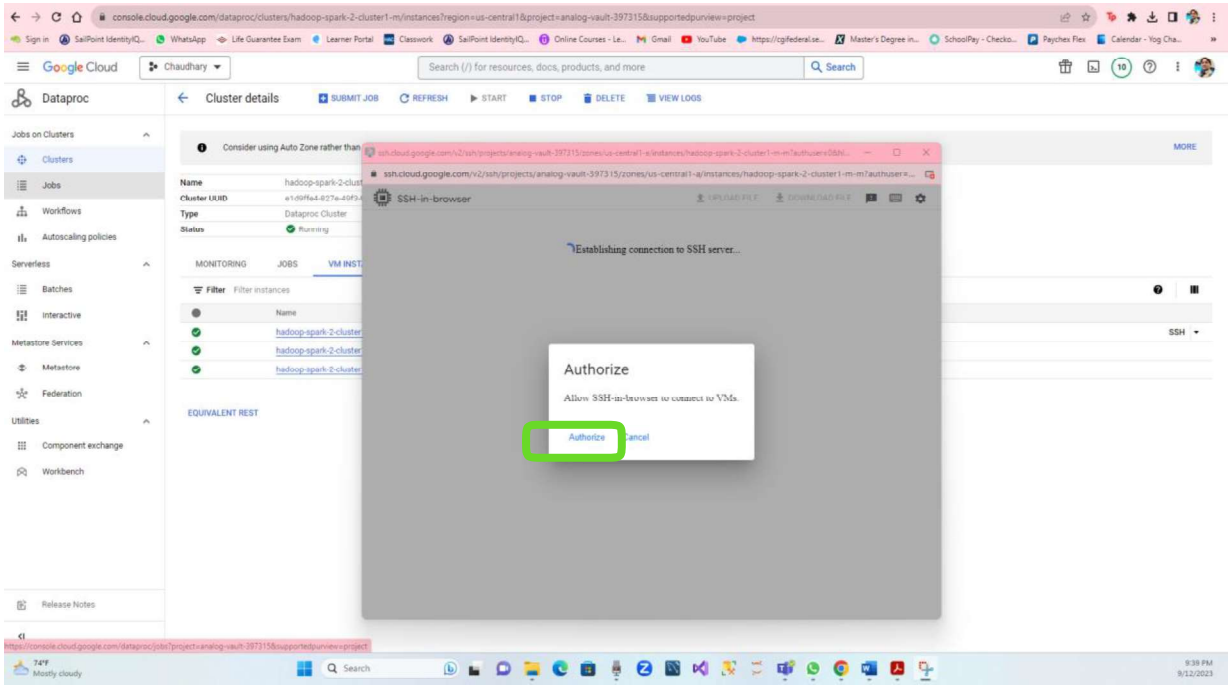
- **By scrolling up and clicking on the virtual machines the screen will show the cluster details.**
- **We will see one master node and one worker node.**
- **After that I accessed the master node through "SSH."**
- **Click on "SSH."**
- **Click on "Open in browser."**



- **Then I clicked on the drop-down button next to SSH, then clicked on open in new**
- browser window.



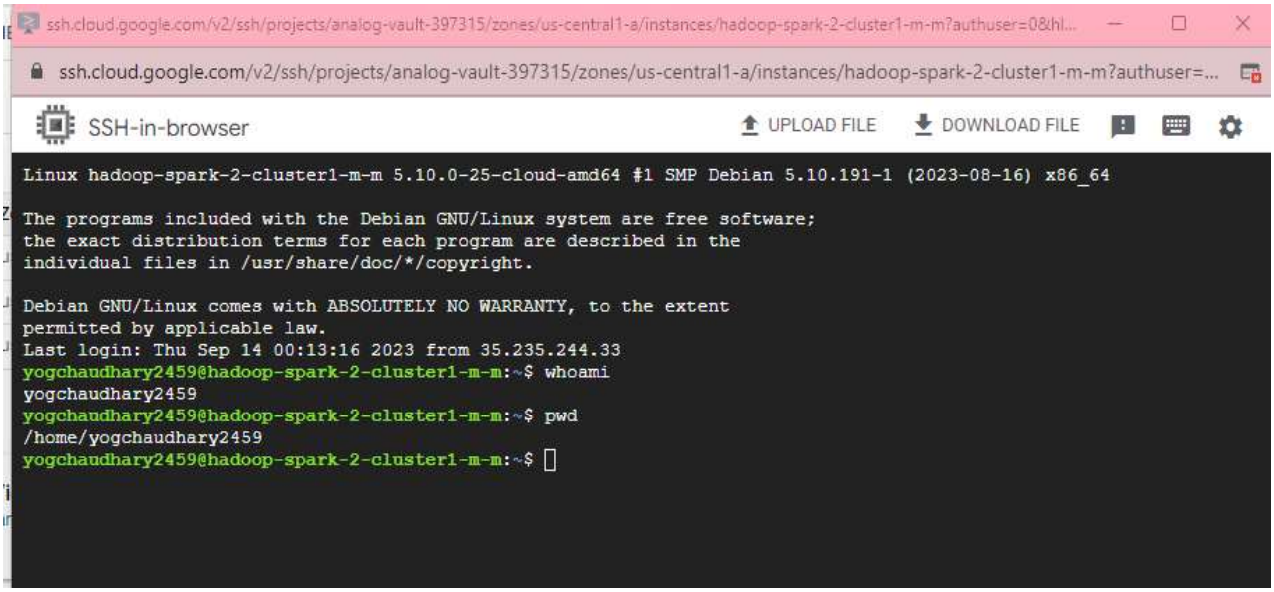
- Click authorized.



- I used the commands

→ **Whoami**

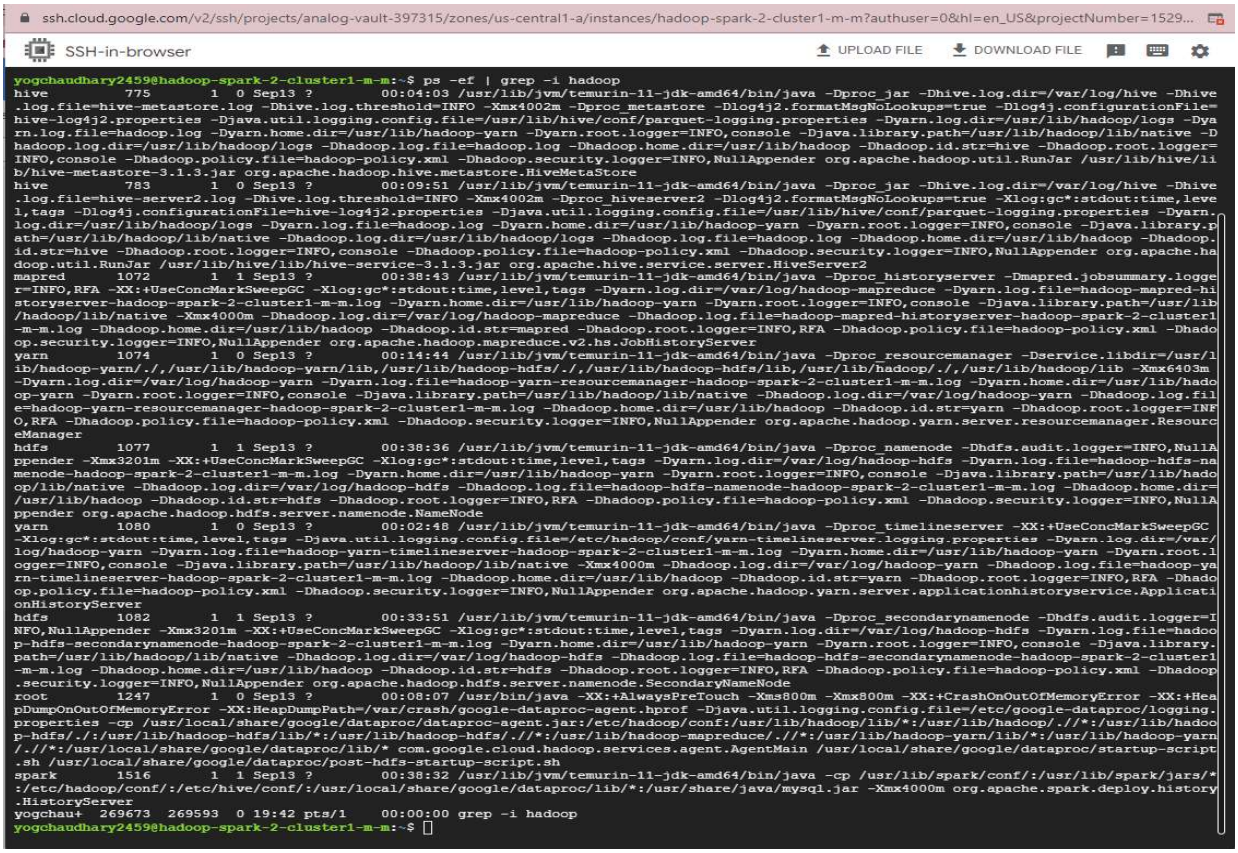
→ **Pwd**



- Try to another commands

→ **Ps –ef | grep -I hadoop**

- This will list all the processing currently running.
- Hare screenshot

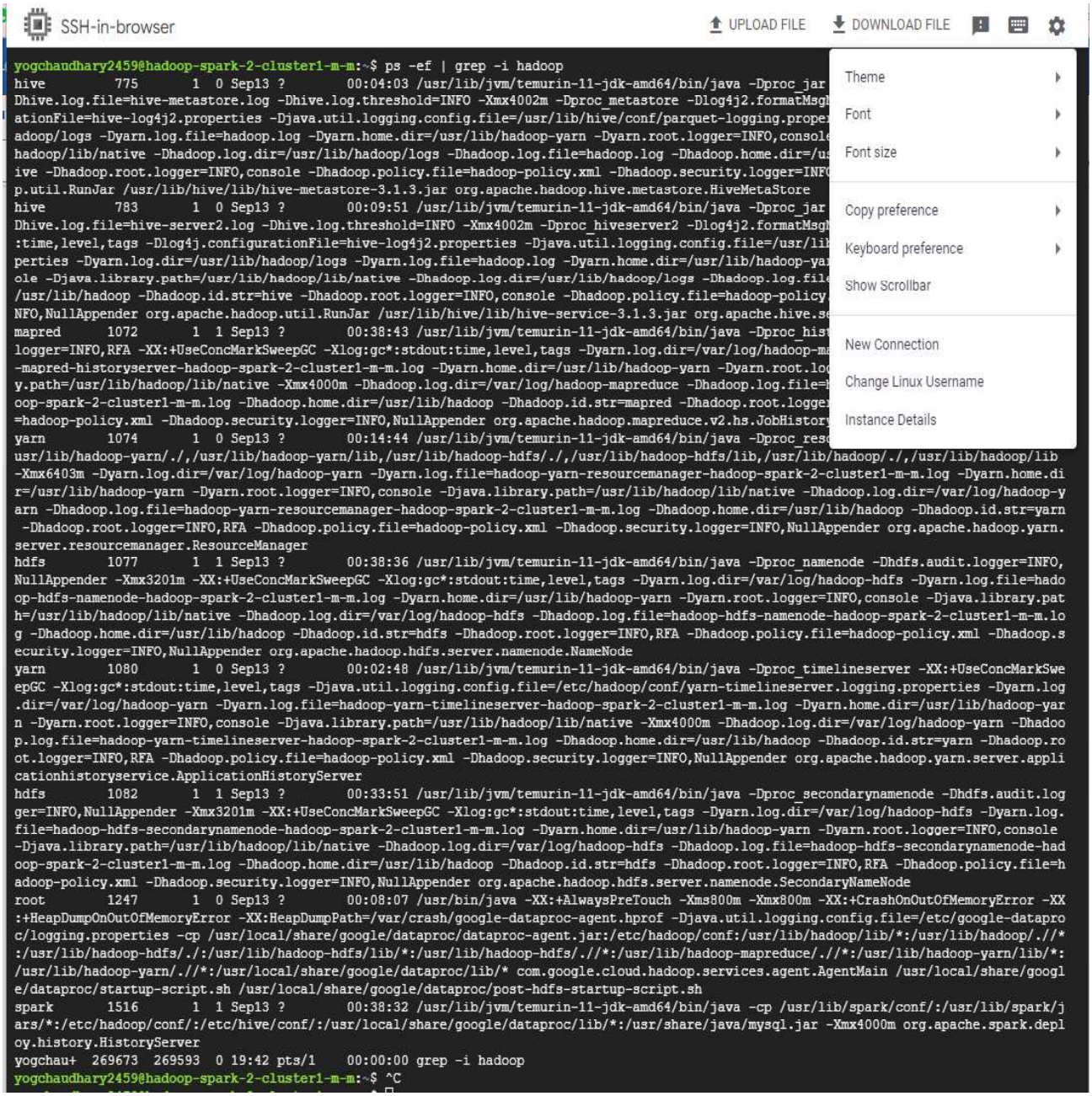


- Then, I move up and down on the terminal window.

→ Click on the setting icon in the upper right-hand side of the terminal

→ Click on “Show Scrollbar” to see the scrollbar.

Include screenshots that record the steps, and the Hadoop Ecosystem with Linux Commands is successfully screenshot



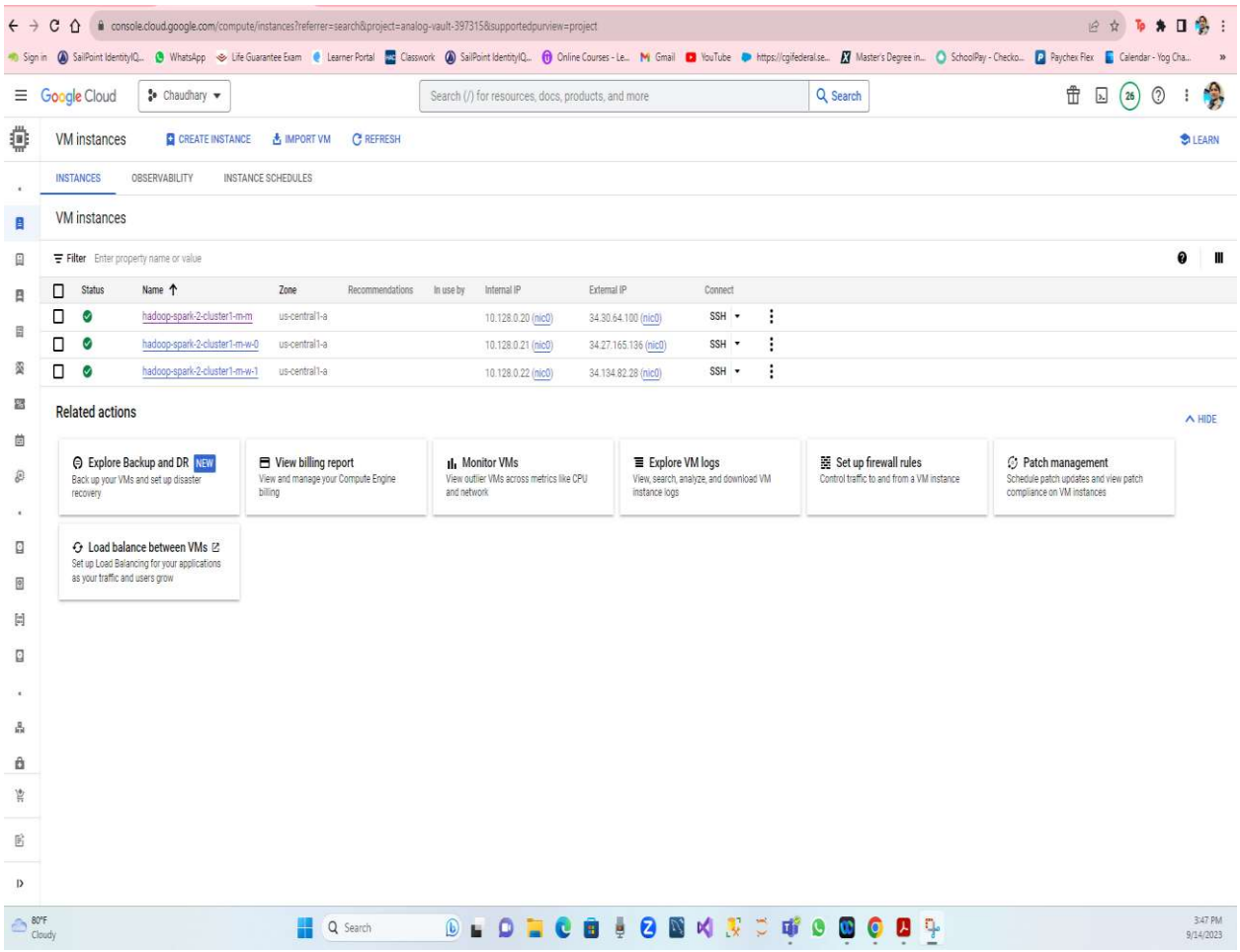
- They are all components of the Hadoop Ecosystem (ex. hive). Include the process ID number (ex. 744) and the subsystem currently running in the component (ex. Job History).

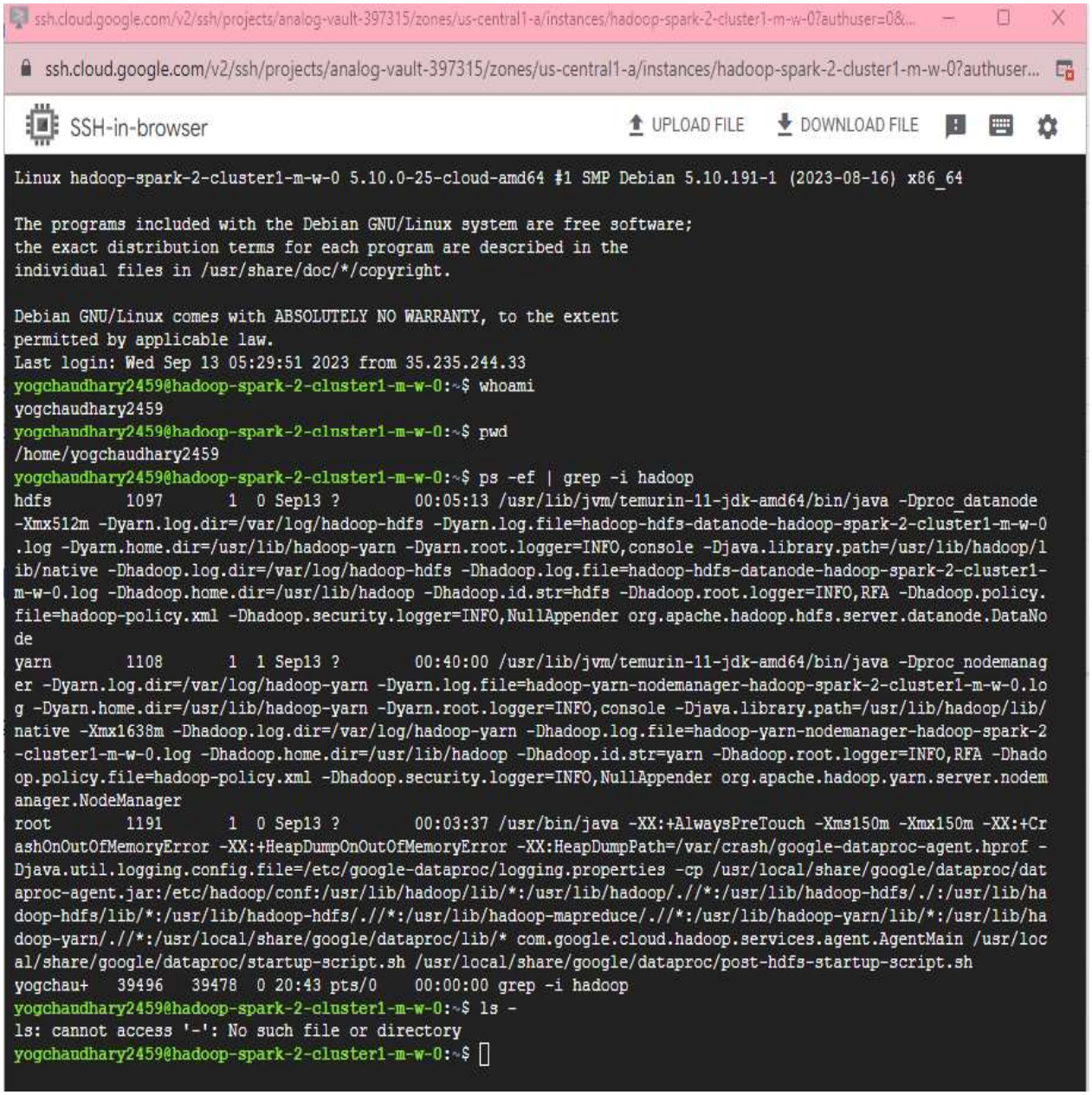
- hive. Process number 775 running Hive MetaStor
- hive. Process number 783 running HiveServer2
- Mapred. Process number 1072 running JobHistoryServer
- Yarn process number 1074 running ResourceManager
- Hdfs process number 1077 running NameNode.

- Yarn. Process number 1080 running Application HistoryServer
- Hdfs. Process number 1082 running SecondaryNameNode
- Root. Process number 12347ResourceManager
- Spark. Process number 1516 running HistoryServer

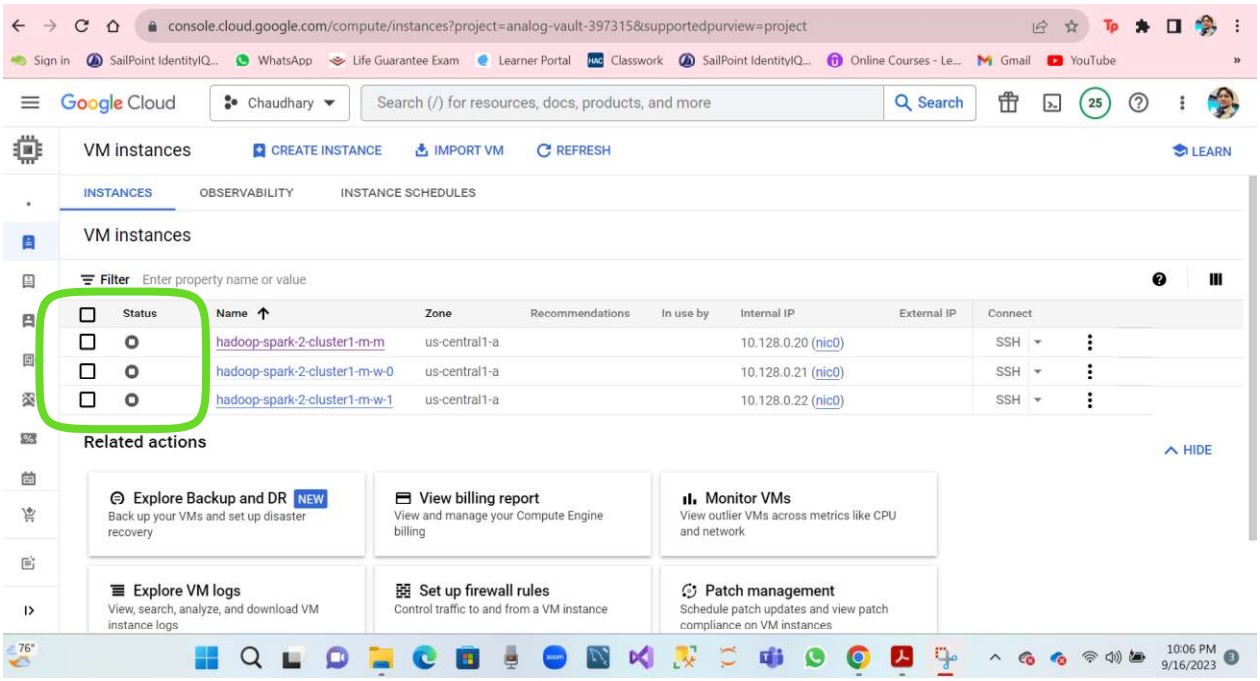
2. **Now we must open another new SSH terminal. For this, we need to go back to the master node to access another SSH terminal.**

- Going back to GCP and typing “computer engine.”
- New clicked on “SSH” then I clicked on “open in browser.”





3. Now finally all 3 virtual machine instances were stopped as shown below in GCP by selecting stop from the three dots present at the top right to the **SSH** each node.



4. The components of the Ecosystem for record and steps, included screenshot

→ Hare is the screenshot