Yog Chaudhary

11727095

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

Professor: Dr. Zeynep Orhan   **Sep 15, 2023**
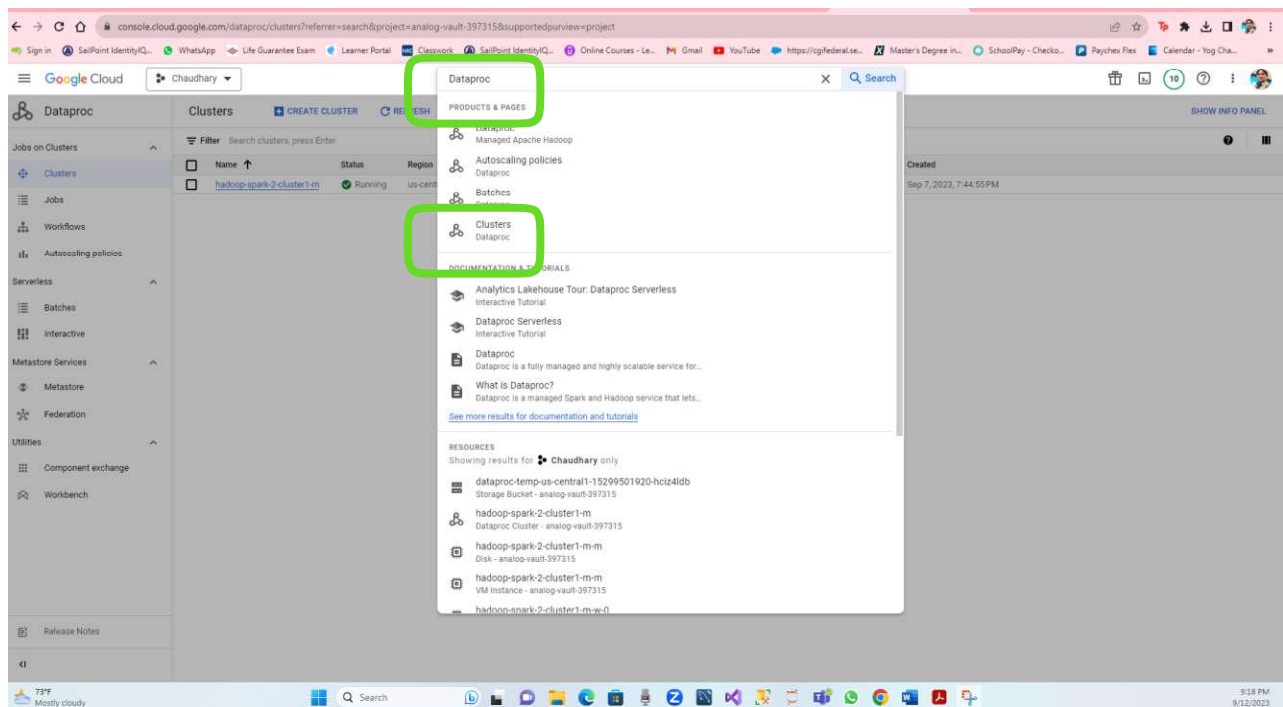
University Of North Texas

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

- ❖ **Monitoring the cluster in GCP:**
1. **For a Google 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.**
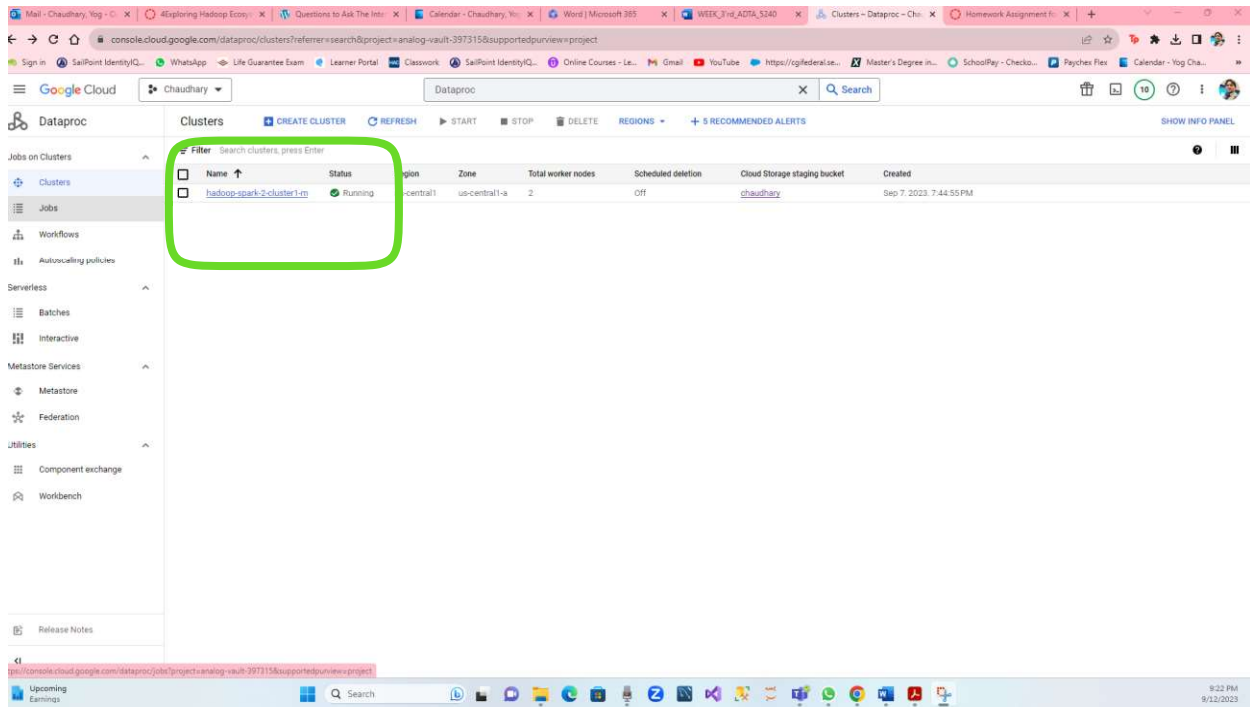
- **While doing this we need to start nodes by checking on the navigation panel.**
- **Clicked on computing engine.**
- **Then clicked on three vertical dots and clicked on start/resume.**
- **Now click.**
- **Below screenshot of that.**
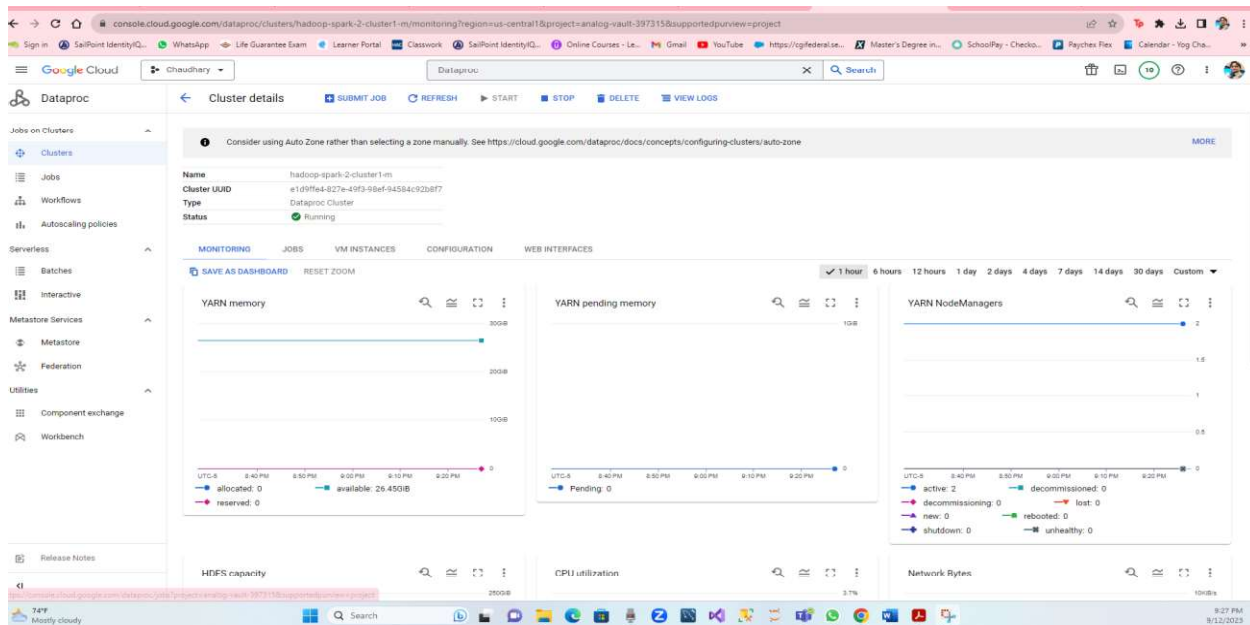
- **When I scroll down it will show how the monitoring is changing according to the usage.**
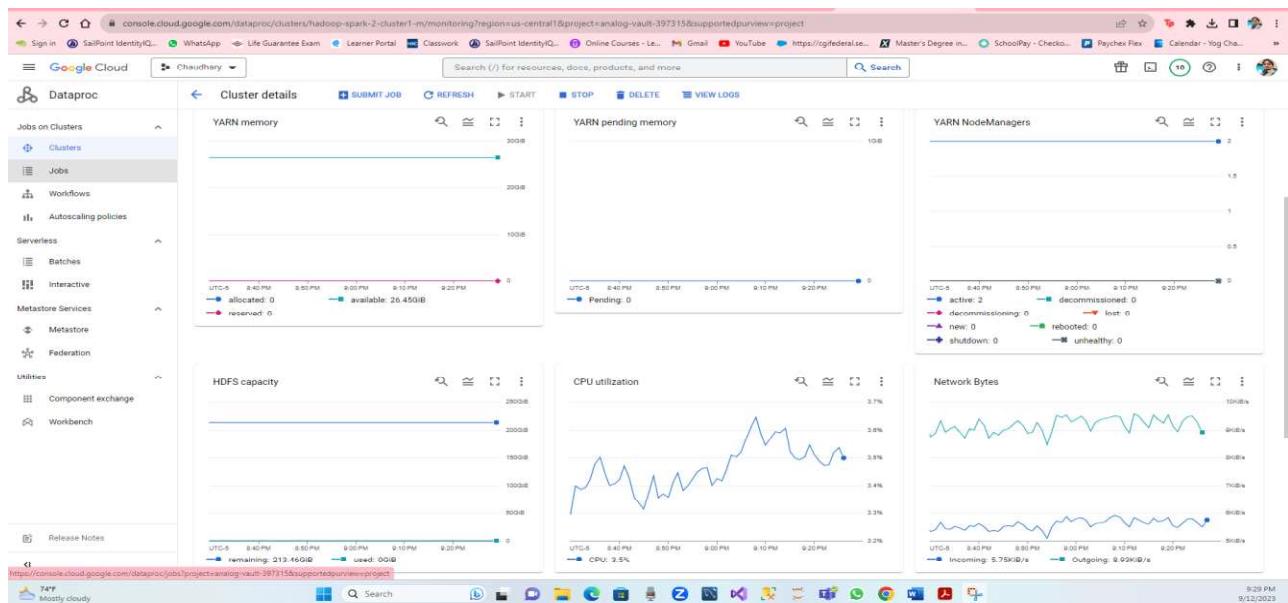
- By scrolling up and clicking on the virtual machines the screen will show the cluster details.
- We will see one master node and one worker node.
- After that I accessed the master node through "SSH."
- Click on "SSH."
- Click on "Open in browser."
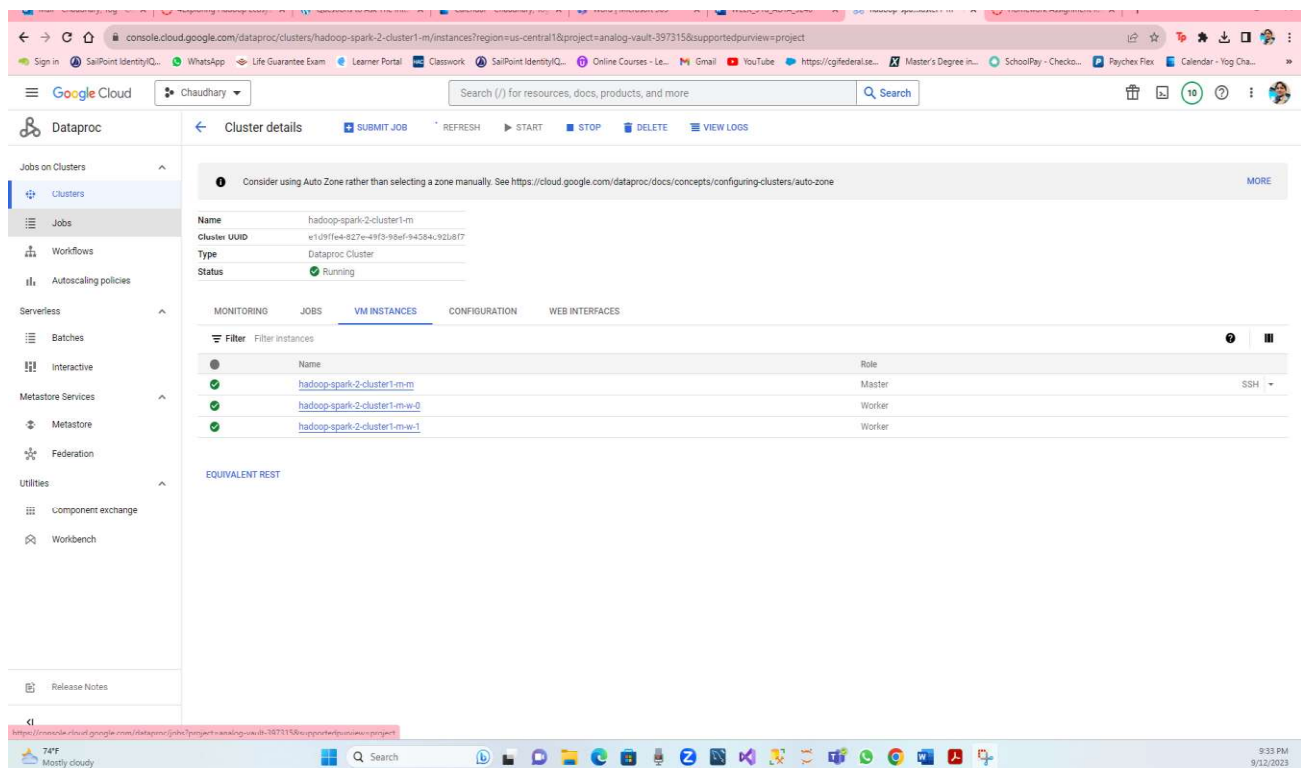
- Click on the drop-down button next to "SSH".
- Click on open in a new browser window.
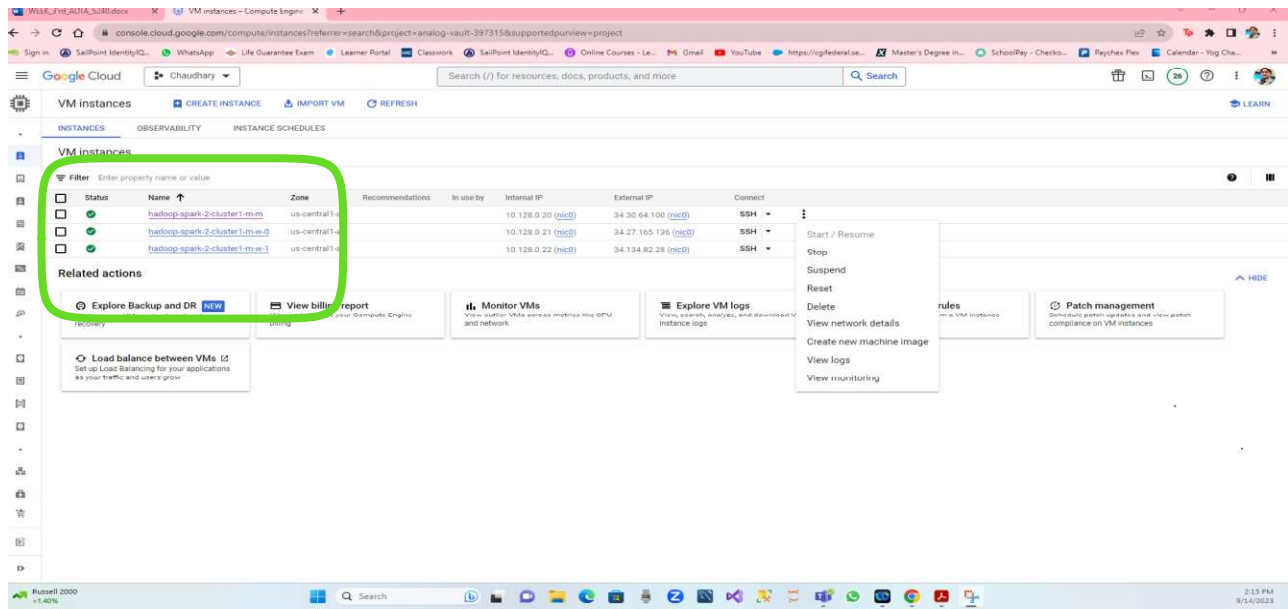
- Click authorized.
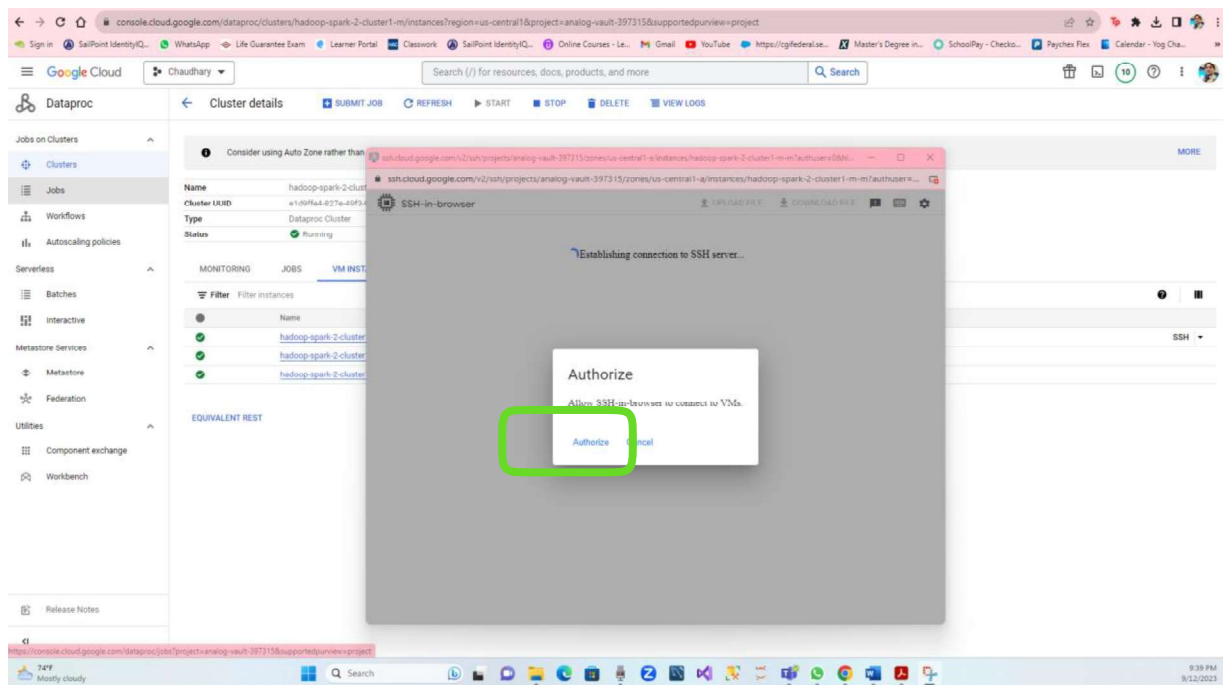
- ❖ See all the services of Hadoop in our cluster.
- ❖ Use the command.
- whoami
- pwd
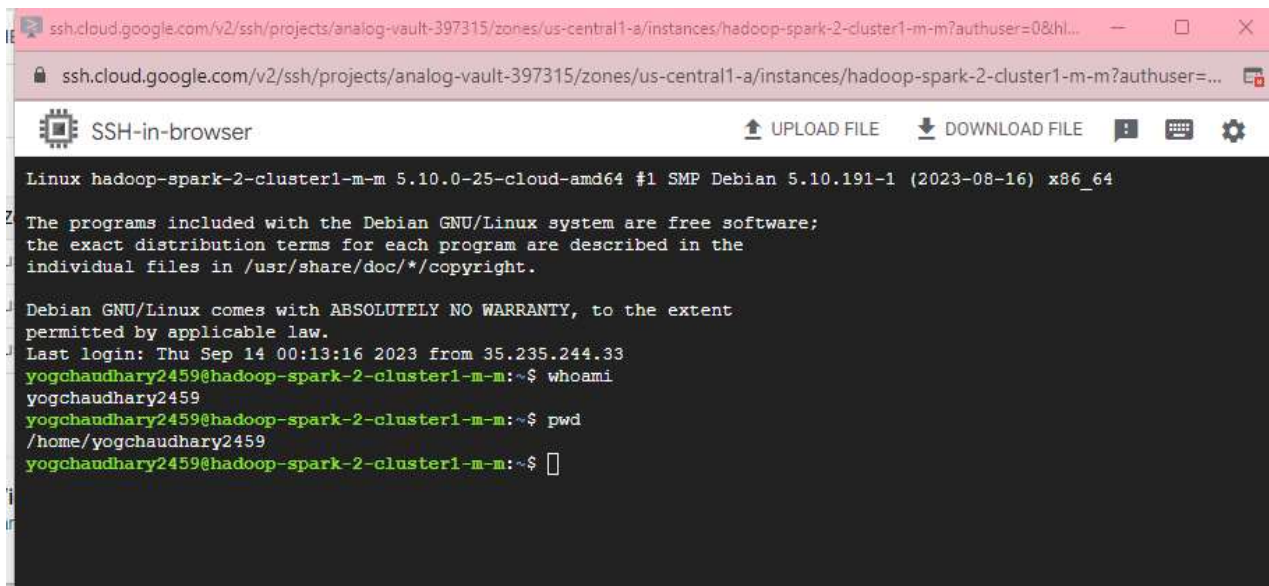
- ❖ use other commands.
- • **ps –ef | grep -I hadoop**
- • Hare screenshot

- • Click on "Show Scrollbar" to see the scrollbar.

Hadoop with Linux Commands is successfully screenshot

```
yogchaudhary2459@hadoop-spark-2-cluster1-m-m:~$ ps -ef | grep -i hadoop
hive        775     1  0 Sep13 ?        00:04:03 /usr/lib/jvm/temurin-11-jdk-amd64/bin/java -Dproc_jar
Dhive.log.file=hive-metastore.log -Dhive.log.threshold=INFO -Xmx4002m -Dproc_metastore -Dlog4j2.formatMsg
ationFile=hive-log4j2.properties -Djava.util.logging.config.file=/usr/lib/hive/conf/parquet-logging.prope
adoop/logs -Dyarn.log.file=hadoop.log -Dyarn.home.dir=/usr/lib/hadoop-yarn -Dyarn.root.logger=INFO,consol
hadoop/lib/native -Dhadoop.log.dir=/usr/lib/hadoop/logs -Dhadoop.log.file=hadoop.log -Dhadoop.home.dir=/us
ive -Dhadoop.root.logger=INFO,console -Dhadoop.policy.file=hadoop-policy.xml -Dhadoop.security.logger=INFO
p.util.RunJar /usr/lib/hive/lib/hive-metastore-3.1.3.jar org.apache.hadoop.hive.metastore.HiveMetaStore
hive        783     1  0 Sep13 ?        00:09:51 /usr/lib/jvm/temurin-11-jdk-amd64/bin/java -Dproc_jar
Dhive.log.file=hive-server2.log -Dhive.log.threshold=INFO -Xmx4002m -Dproc_hiveserver2 -Dlog4j2.formatMsg
:time,level,tags -Dlog4j.configurationFile=hive-log4j2.properties -Djava.util.logging.config.file=/usr/li
perties -Dyarn.log.dir=/usr/lib/hadoop/logs -Dyarn.log.file=hadoop.log -Dyarn.home.dir=/usr/lib/hadoop-ya
ole -Djava.library.path=/usr/lib/hadoop/lib/native -Dhadoop.log.dir=/usr/lib/hadoop/logs -Dhadoop.log.fil
/usr/lib/hadoop -Dhadoop.id.str=hive -Dhadoop.root.logger=INFO,console -Dhadoop.policy.file=hadoop-policy.
NFO,NullAppender org.apache.hadoop.util.RunJar /usr/lib/hive/lib/hive-service-3.1.3.jar org.apache.hive.se
mapred     1072     1  1 Sep13 ?        00:38:43 /usr/lib/jvm/temurin-11-jdk-amd64/bin/java -Dproc_hist
logger=INFO,RFA -XX:+UseConcMarkSweepGC -Xlog:gc*:stdout:time,level,tags -Dyarn.log.dir=/var/log/hadoop-ma
-mapred-historyserver-hadoop-spark-2-cluster1-m-m.log -Dyarn.home.dir=/usr/lib/hadoop-yarn -Dyarn.root.log
y.path=/usr/lib/hadoop/lib/native -Xmx4000m -Dhadoop.log.dir=/var/log/hadoop-mapreduce -Dhadoop.log.file=h
oop-spark-2-cluster1-m-m.log -Dhadoop.home.dir=/usr/lib/hadoop -Dhadoop.id.str=mapred -Dhadoop.root.logger
=hadoop-policy.xml -Dhadoop.security.logger=INFO,NullAppender org.apache.hadoop.mapreduce.v2.hs.JobHistory
yarn       1074     1  1 Sep13 ?        00:14:44 /usr/lib/jvm/temurin-11-jdk-amd64/bin/java -Dproc_reso
usr/lib/hadoop-yarn/./,/usr/lib/hadoop-yarn/lib,/usr/lib/hadoop-hdfs/./,/usr/lib/hadoop-hdfs/lib,/usr/lib/hadoop/./,/usr/lib/hadoop/lib
-Xmx6403m -Dyarn.log.dir=/var/log/hadoop-yarn -Dyarn.log.file=hadoop-yarn-resourcemanager-hadoop-spark-2-cluster1-m-m.log -Dyarn.home.di
r=/usr/lib/hadoop-yarn -Dyarn.root.logger=INFO,console -Djava.library.path=/usr/lib/hadoop/lib/native -Dhadoop.log.dir=/var/log/hadoop-y
arn -Dhadoop.log.file=hadoop-yarn-resourcemanager-hadoop-spark-2-cluster1-m-m.log -Dhadoop.home.dir=/usr/lib/hadoop -Dhadoop.id.str=yarn
-Dhadoop.root.logger=INFO,RFA -Dhadoop.policy.file=hadoop-policy.xml -Dhadoop.security.logger=INFO,NullAppender org.apache.hadoop.yarn.
server.resourcemanager.ResourceManager
hdfs       1077     1  1 Sep13 ?        00:38:36 /usr/lib/jvm/temurin-11-jdk-amd64/bin/java -Dproc_namenode -Dhdfs.audit.logger=INFO,
NullAppender -Xmx3201m -XX:+UseConcMarkSweepGC -Xlog:gc*:stdout:time,level,tags -Dyarn.log.dir=/var/log/hadoop-hdfs -Dyarn.log.file=hado
op-hdfs-namenode-hadoop-spark-2-cluster1-m-m.log -Dyarn.home.dir=/usr/lib/hadoop-yarn -Dyarn.root.logger=INFO,console -Djava.library.pat
h=/usr/lib/hadoop/lib/native -Dhadoop.log.dir=/var/log/hadoop-hdfs -Dhadoop.log.file=hadoop-hdfs-namenode-hadoop-spark-2-cluster1-m-m.lo
g -Dhadoop.home.dir=/usr/lib/hadoop -Dhadoop.id.str=hdfs -Dhadoop.root.logger=INFO,RFA -Dhadoop.policy.file=hadoop-policy.xml -Dhadoop.s
ecurity.logger=INFO,NullAppender org.apache.hadoop.hdfs.server.namenode.NameNode
yarn       1080     1  0 Sep13 ?        00:02:48 /usr/lib/jvm/temurin-11-jdk-amd64/bin/java -Dproc_timelineserver -XX:+UseConcMarkSwe
epGC -Xlog:gc*:stdout:time,level,tags -Djava.util.logging.config.file=/etc/hadoop/conf/yarn-timelineserver.logging.properties -Dyarn.log
.dir=/var/log/hadoop-yarn -Dyarn.log.file=hadoop-yarn-timelineserver-hadoop-spark-2-cluster1-m-m.log -Dyarn.home.dir=/usr/lib/hadoop-yar
n -Dyarn.root.logger=INFO,console -Djava.library.path=/usr/lib/hadoop/lib/native -Xmx4000m -Dhadoop.log.dir=/var/log/hadoop-yarn -Dhadoo
p.log.file=hadoop-yarn-timelineserver-hadoop-spark-2-cluster1-m-m.log -Dhadoop.home.dir=/usr/lib/hadoop -Dhadoop.id.str=yarn -Dhadoop.ro
ot.logger=INFO,RFA -Dhadoop.policy.file=hadoop-policy.xml -Dhadoop.security.logger=INFO,NullAppender org.apache.hadoop.yarn.server.appli
cationhistoryservice.ApplicationHistoryServer
hdfs       1082     1  1 Sep13 ?        00:33:51 /usr/lib/jvm/temurin-11-jdk-amd64/bin/java -Dproc_secondarynamenode -Dhdfs.audit.log
ger=INFO,NullAppender -Xmx3201m -XX:+UseConcMarkSweepGC -Xlog:gc*:stdout:time,level,tags -Dyarn.log.dir=/var/log/hadoop-hdfs -Dyarn.log.
file=hadoop-hdfs-secondarynamenode-hadoop-spark-2-cluster1-m-m.log -Dyarn.home.dir=/usr/lib/hadoop-yarn -Dyarn.root.logger=INFO,console
-Djava.library.path=/usr/lib/hadoop/lib/native -Dhadoop.log.dir=/var/log/hadoop-hdfs -Dhadoop.log.file=hadoop-hdfs-secondarynamenode-had
oop-spark-2-cluster1-m-m.log -Dhadoop.home.dir=/usr/lib/hadoop -Dhadoop.id.str=hdfs -Dhadoop.root.logger=INFO,RFA -Dhadoop.policy.file=h
adoop-policy.xml -Dhadoop.security.logger=INFO,NullAppender org.apache.hadoop.hdfs.server.namenode.SecondaryNameNode
root       1247     1  0 Sep13 ?        00:08:07 /usr/bin/java -XX:+AlwaysPreTouch -Xms800m -Xmx800m -XX:+CrashOnOutOfMemoryError -XX
:+HeapDumpOnOutOfMemoryError -XX:HeapDumpPath=/var/crash/google-dataproc-agent.hprof -Djava.util.logging.config.file=/etc/google-datapro
c/logging.properties -cp /usr/local/share/google/dataproc/dataproc-agent.jar:/etc/hadoop/conf:/usr/lib/hadoop/lib/*:/usr/lib/hadoop/./*
:/usr/lib/hadoop-hdfs/./:/usr/lib/hadoop-hdfs/lib/*:/usr/lib/hadoop-hdfs/.//*:/usr/lib/hadoop-mapreduce/.//*:/usr/lib/hadoop-yarn/lib/*:
/usr/lib/hadoop-yarn/.//*:/usr/local/share/google/dataproc/lib/* com.google.cloud.hadoop.services.agent.AgentMain /usr/local/share/googl
e/dataproc/startup-script.sh /usr/local/share/google/dataproc/post-hdfs-startup-script.sh
spark      1516     1  1 Sep13 ?        00:38:32 /usr/lib/jvm/temurin-11-jdk-amd64/bin/java -cp /usr/lib/spark/conf/:/usr/lib/spark/j
ars/*:/etc/hadoop/conf/:/etc/hive/conf/:/usr/local/share/google/dataproc/lib/*:/usr/share/java/mysql.jar -Xmx4000m org.apache.spark.depl
oy.history.HistoryServer
yogchau+ 269673 269593  0 19:42 pts/1    00:00:00 grep -i hadoop
yogchaudhary2459@hadoop-spark-2-cluster1-m-m:~$ ^C
```
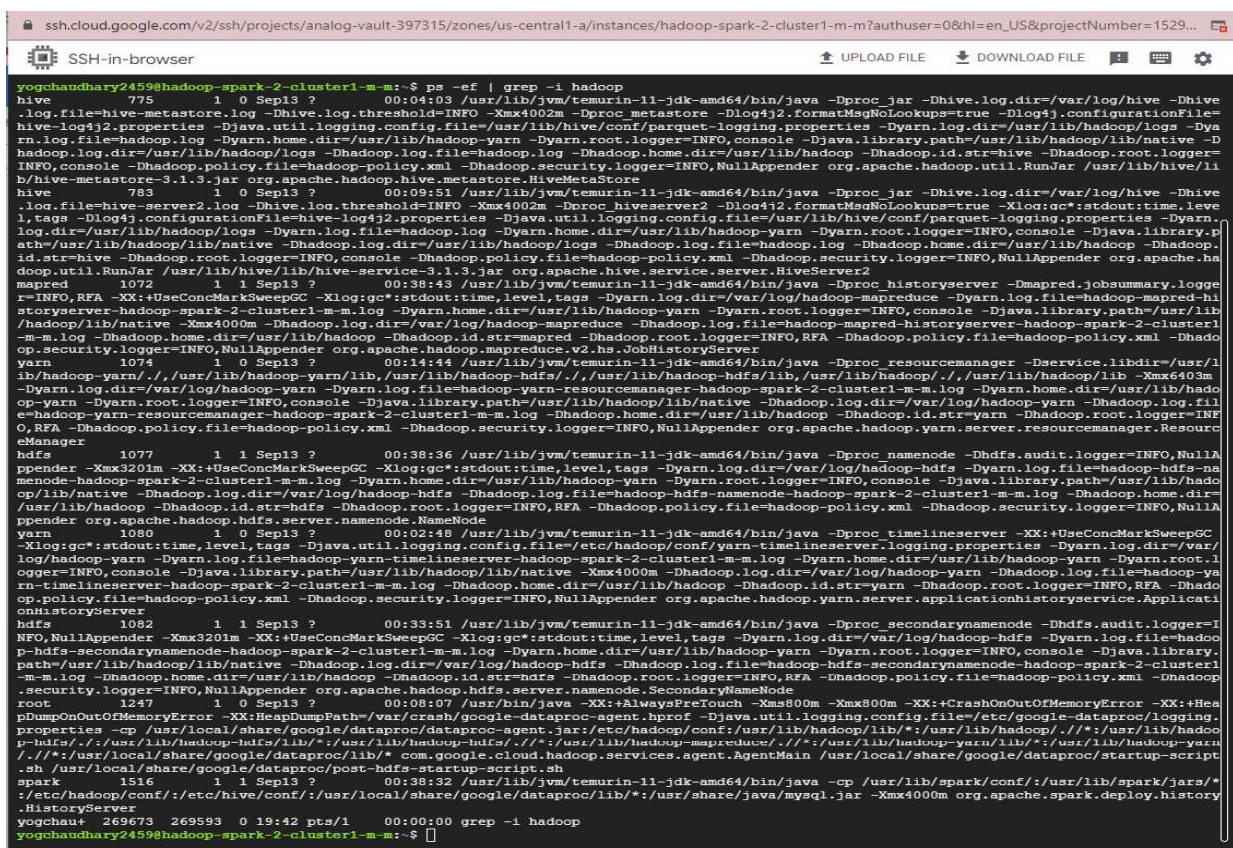
- ❖ They are all the components of Hadoop with Linux Commands.
- hive. Process number 775 running Hive MetaStor.
- hive. Process number 783 running HiveServer2.
- Mapred. Process number 1072 running JobHistoryServer.
- Yarn process number 1074 running ResourceManager.
- Hdfs process number 1077 running NameNode.
- Yarn. Process number 1080 running Application HistoryServer.
- Hdfs. Process number 1082 running SecondaryNameNode.
- Root. Process number 12347 ResourceManager.
- Spark. Process number 1516 running HistoryServer.

2. Now we must open another new SSH terminal.
- Going back to GCP and typing "computer engine."
- New clicked on "SSH" then I clicked on "open in browser."

SSH-in-browser ⬆ UPLOAD FILE ⬇ DOWNLOAD FILE

```
Linux hadoop-spark-2-cluster1-m-w-0 5.10.0-25-cloud-amd64 #1 SMP Debian 5.10.191-1 (2023-08-16) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Wed Sep 13 05:29:51 2023 from 35.235.244.33
yogchaudhary2459@hadoop-spark-2-cluster1-m-w-0:~$ whoami
yogchaudhary2459
yogchaudhary2459@hadoop-spark-2-cluster1-m-w-0:~$ pwd
/home/yogchaudhary2459
yogchaudhary2459@hadoop-spark-2-cluster1-m-w-0:~$ ps -ef | grep -i hadoop
hdfs        1097       1  0 Sep13 ?        00:05:13 /usr/lib/jvm/temurin-11-jdk-amd64/bin/java -Dproc_datanode
-Xmx512m -Dyarn.log.dir=/var/log/hadoop-hdfs -Dyarn.log.file=hadoop-hdfs-datanode-hadoop-spark-2-cluster1-m-w-0
.log -Dyarn.home.dir=/usr/lib/hadoop-yarn -Dyarn.root.logger=INFO,console -Djava.library.path=/usr/lib/hadoop/l
ib/native -Dhadoop.log.dir=/var/log/hadoop-hdfs -Dhadoop.log.file=hadoop-hdfs-datanode-hadoop-spark-2-cluster1-
m-w-0.log -Dhadoop.home.dir=/usr/lib/hadoop -Dhadoop.id.str=hdfs -Dhadoop.root.logger=INFO,RFA -Dhadoop.policy.
file=hadoop-policy.xml -Dhadoop.security.logger=INFO,NullAppender org.apache.hadoop.hdfs.server.datanode.DataNo
de
yarn        1108       1  1 Sep13 ?        00:40:00 /usr/lib/jvm/temurin-11-jdk-amd64/bin/java -Dproc_nodemanag
er -Dyarn.log.dir=/var/log/hadoop-yarn -Dyarn.log.file=hadoop-yarn-nodemanager-hadoop-spark-2-cluster1-m-w-0.lo
g -Dyarn.home.dir=/usr/lib/hadoop-yarn -Dyarn.root.logger=INFO,console -Djava.library.path=/usr/lib/hadoop/lib/
native -Xmx1638m -Dhadoop.log.dir=/var/log/hadoop-yarn -Dhadoop.log.file=hadoop-yarn-nodemanager-hadoop-spark-2
-cluster1-m-w-0.log -Dhadoop.home.dir=/usr/lib/hadoop -Dhadoop.id.str=yarn -Dhadoop.root.logger=INFO,RFA -Dhado
op.policy.file=hadoop-policy.xml -Dhadoop.security.logger=INFO,NullAppender org.apache.hadoop.yarn.server.nodem
anager.NodeManager
root        1191       1  0 Sep13 ?        00:03:37 /usr/bin/java -XX:+AlwaysPreTouch -Xms150m -Xmx150m -XX:+Cr
ashOnOutOfMemoryError -XX:+HeapDumpOnOutOfMemoryError -XX:HeapDumpPath=/var/crash/google-dataproc-agent.hprof -
Djava.util.logging.config.file=/etc/google-dataproc/logging.properties -cp /usr/local/share/google/dataproc/dat
aproc-agent.jar:/etc/hadoop/conf:/usr/lib/hadoop/lib/*:/usr/lib/hadoop/.//*:/usr/lib/hadoop-hdfs/.//:/usr/lib/ha
doop-hdfs/lib/*:/usr/lib/hadoop-hdfs/.//*:/usr/lib/hadoop-mapreduce/.//*:/usr/lib/hadoop-yarn/lib/*:/usr/lib/ha
doop-yarn/.//*:/usr/local/share/google/dataproc/lib/* com.google.cloud.hadoop.services.agent.AgentMain /usr/loc
al/share/google/dataproc/startup-script.sh /usr/local/share/google/dataproc/post-hdfs-startup-script.sh
yogchau+   39496   39478  0 20:43 pts/0    00:00:00 grep -i hadoop
yogchaudhary2459@hadoop-spark-2-cluster1-m-w-0:~$ ls -
ls: cannot access '-': No such file or directory
yogchaudhary2459@hadoop-spark-2-cluster1-m-w-0:~$ □
```

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