CSP 554 Big Data Technologies

Assignment – #2(Modules 02a & 02b)

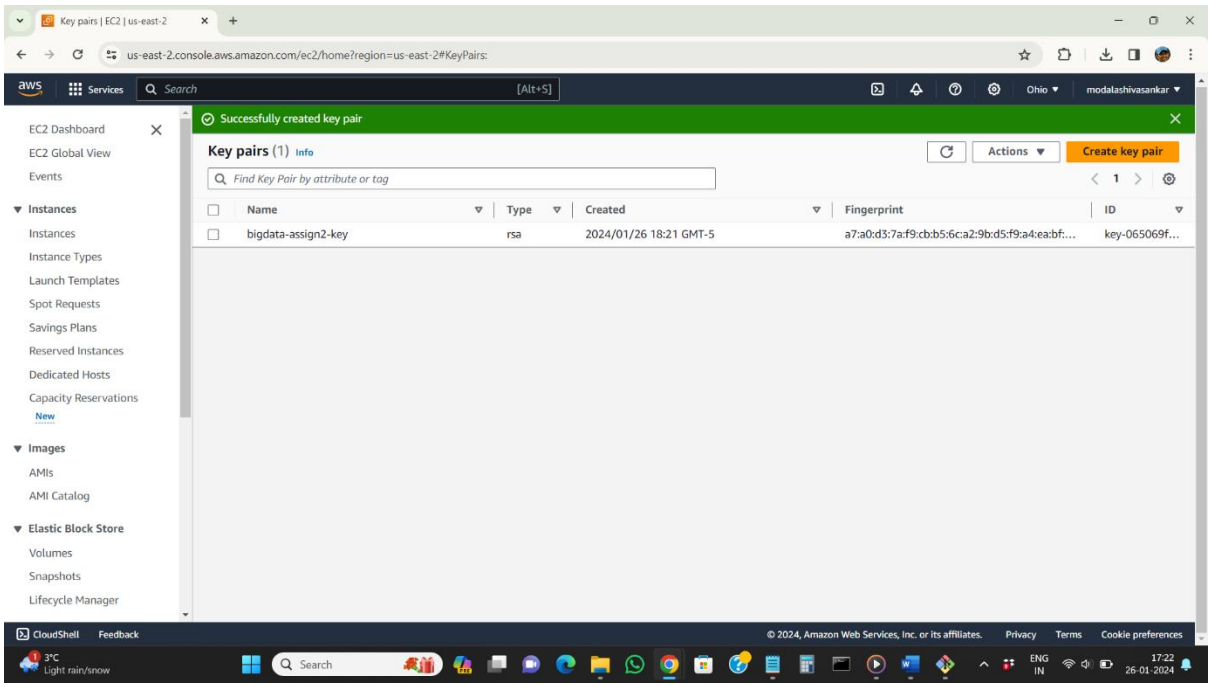
Shiva Sankar Modala(A20517528)

In this assignment, we need to create an EMR cluster to manipulate hadoop file system

We require EC2 key pair, S3 Object, EMR Cluster

First, I created a key pair from EC2

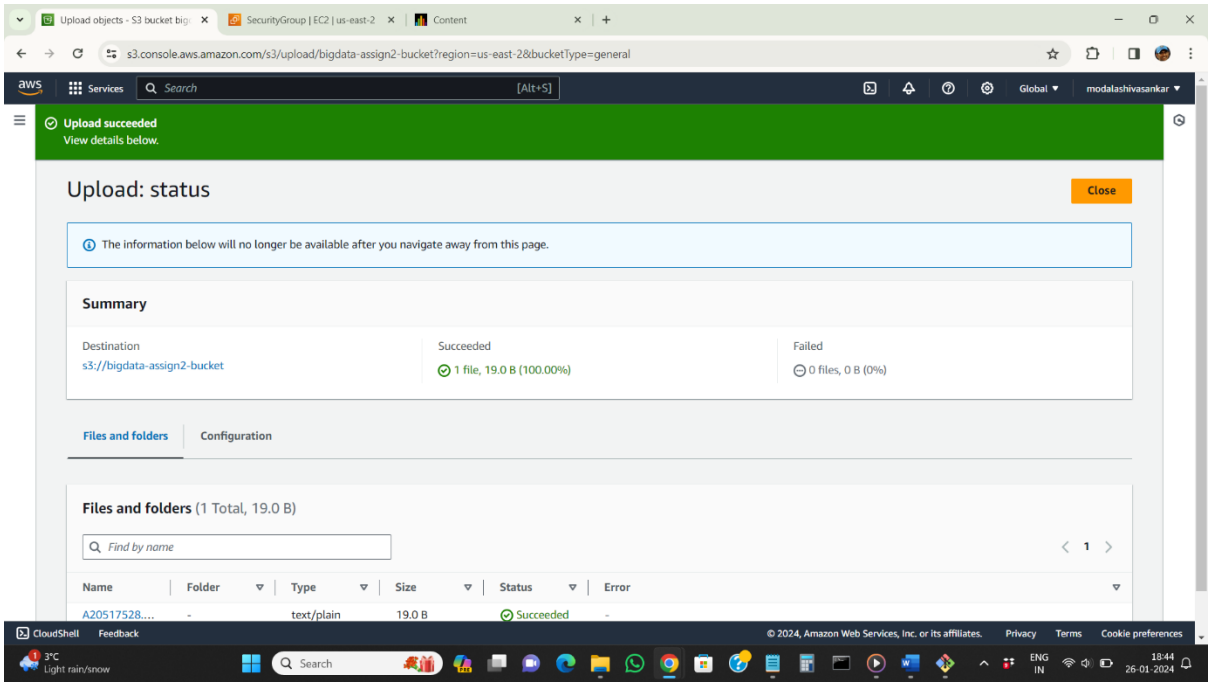
Key pair: bigdata-assign2-key



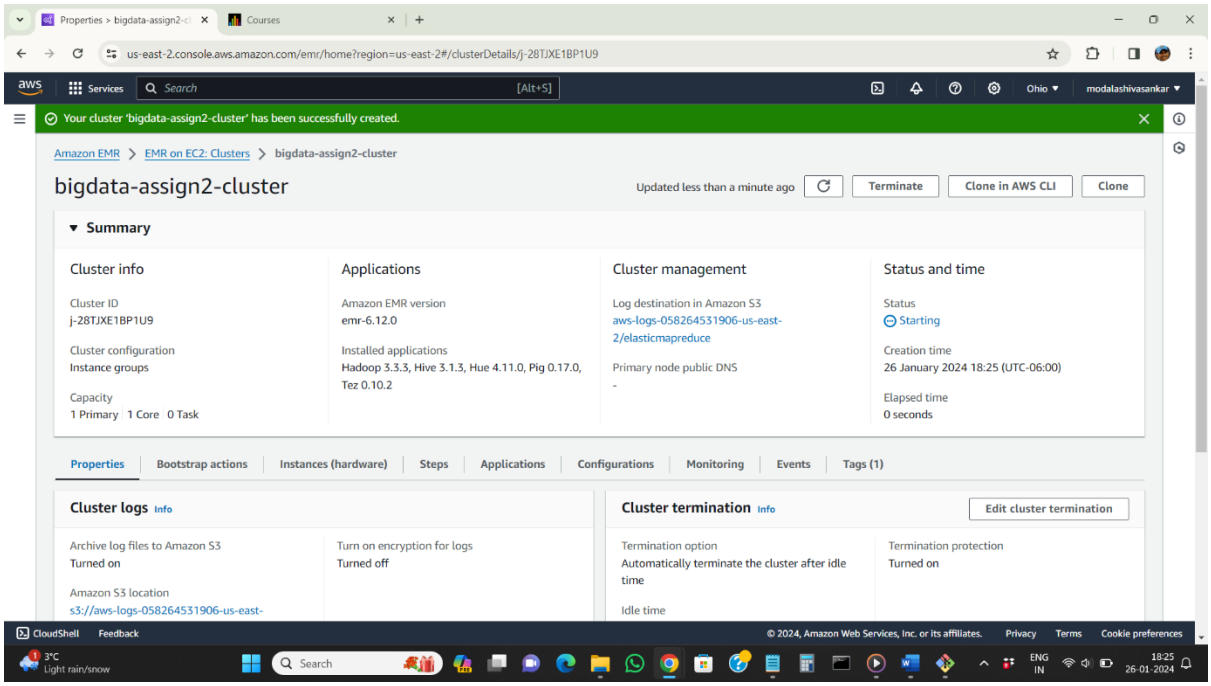
Next, I uploaded .txt file in the S3 bucket

Bucket name: bigdata-assign2-bucket

File uploaded: A20517528.txt



Now, I created an EMR cluster



Setting the permissions for the private key

shiva@LAPTOP-7EA2T3G6 MINGW64 ~

$ chmod 400 C:/Users/shiva/Downloads/bigdata-assign2-key.pem

I moved the .txt file from my local directory to the home directory of the Hadoop.

shiva@LAPTOP-7EA2T3G6 MINGW64 ~

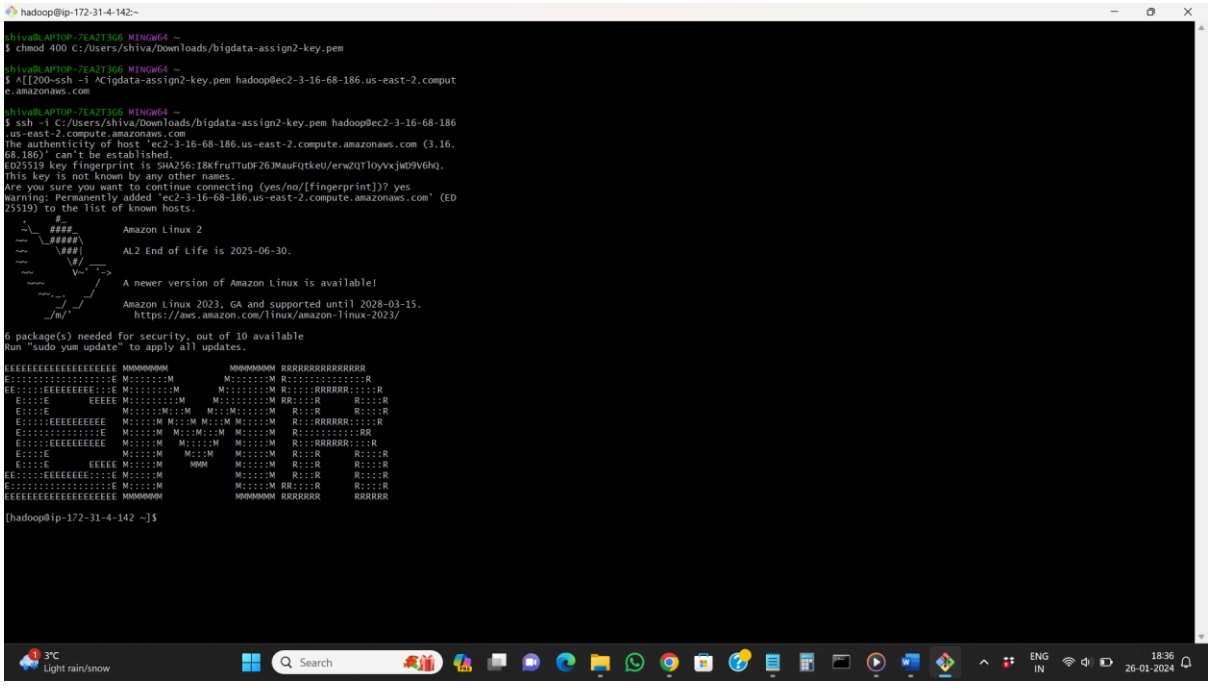
$ scp -i C:/Users/shiva/Downloads/bigdata-assign2-key.pem C:/Users/shiva/Downloads/Shiva_Sankar_Modala.txt hadoop@ec2-3-16-68-186.us-east-2.compute.amazonaws.com:/home/hadoop

Shiva_Sankar_Modala.txt 100% 21 0.9KB/s 00:00

Now, I connected my private key to the Hadoop EC2 instance

shiva@LAPTOP-7EA2T3G6 MINGW64 ~

$ ssh -i C:/Users/shiva/Downloads/bigdata-assign2-key.pem hadoop@ec2-3-16-68-186.us-east-2.compute.amazonaws.com



[hadoop@ip-172-31-4-142 ~]$ pwd

/home/hadoop

[hadoop@ip-172-31-4-142 ~]$ ls

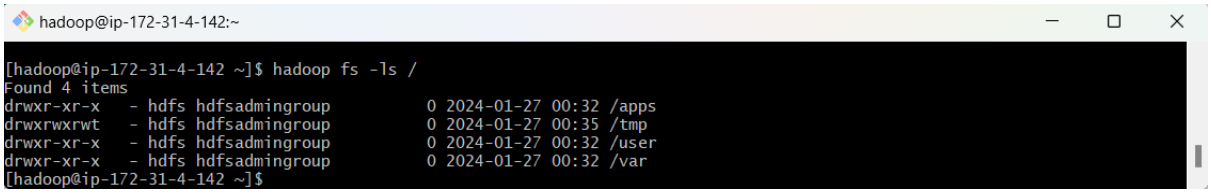
Shiva_Sankar_Modala.txt

pwd – Present Working Directory - It shows us which directory we currently using.

ls – gives the list of files in the present working directory

9) Execute the following hdfs command to list the files or directories that are listed (also indicating which is a file and which a directory)

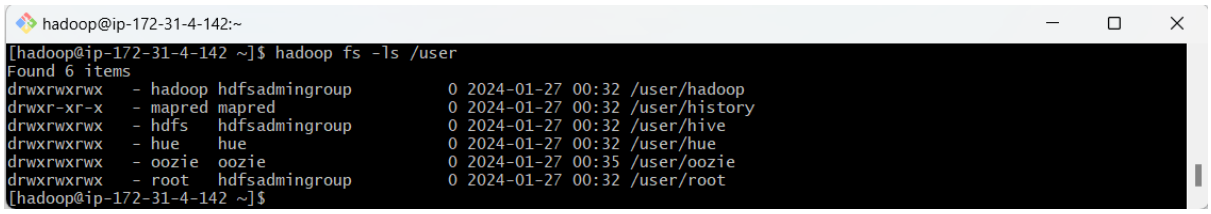
[hadoop@ip-172-31-4-142 ~]$ hadoop fs -ls /



10) Execute a command (you needed to figure out which one) to list the files and directories under the hdfs directory listed below: /user

Write down the command you executed and also take a screen snapshot of names of the files or directories that are listed and include it in your assignment submission.

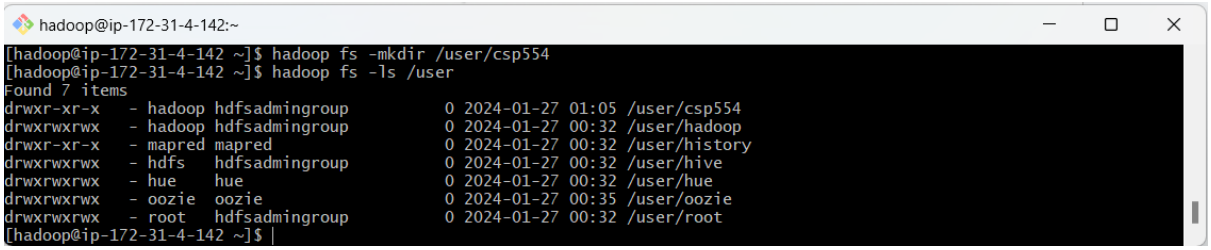
[hadoop@ip-172-31-4-142 ~]$ hadoop fs -ls /user



11) Execute a command to create the following HDFS directory: /user/csp554 Record the command you executed and include it in your assignment submission.

[hadoop@ip-172-31-4-142 ~]$ hadoop fs -mkdir /user/csp554

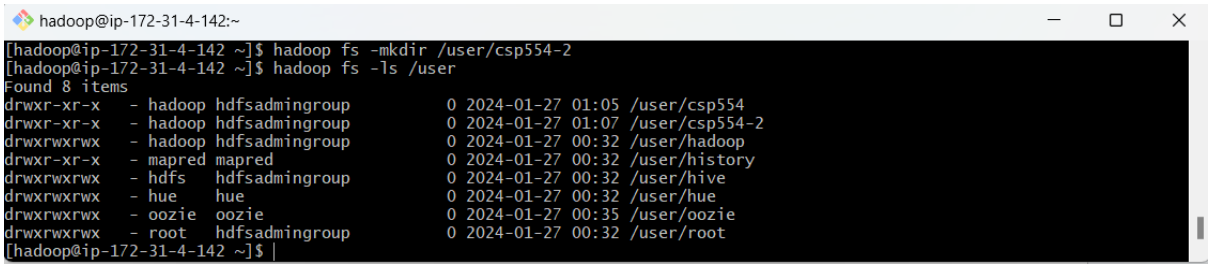
[hadoop@ip-172-31-4-142 ~]$ hadoop fs -ls /user



12) Execute a command to create the following HDFS directory: /user/csp554-2 Record the command you executed and include it in your assignment submission.

[hadoop@ip-172-31-4-142 ~]$ hadoop fs -mkdir /user/csp554-2

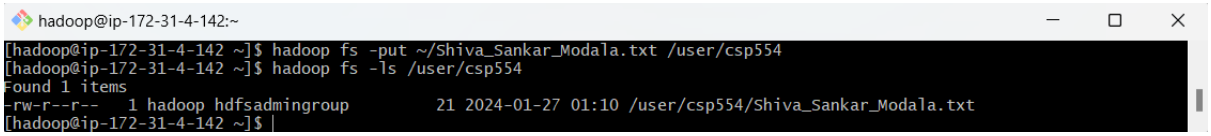
[hadoop@ip-172-31-4-142 ~]$ hadoop fs -ls /user



13) Execute a command that copies a given local file to the given hdfs directory: Source local file: /home/hadoop/myname.txt (where the actual name is your name as described above) Destination HDFS directory: /user/csp554 Record the command you executed and include it in your assignment submission.

[hadoop@ip-172-31-4-142 ~]$ hadoop fs -put ~/Shiva_Sankar_Modala.txt /user/csp554

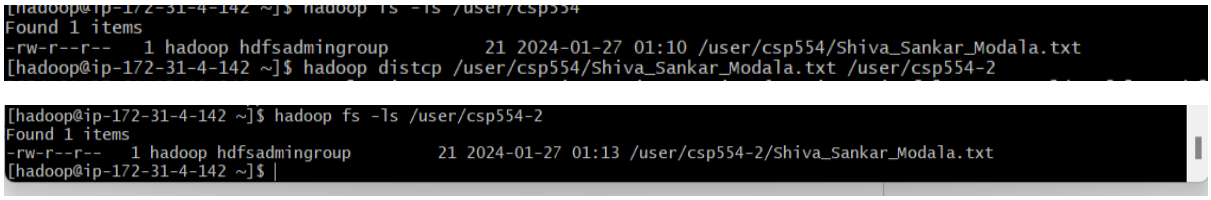
[hadoop@ip-172-31-4-142 ~]$ hadoop fs -ls /user/csp554



14) Copy a file from one hdfs directory to another hdfs directory and write down the command Source hdfs file: /user/csp554/myname.txt (where the actual name is your name as described above) Destination HDFS directory: /user/csp554-2

Record the command you executed and include it in your assignment submission.

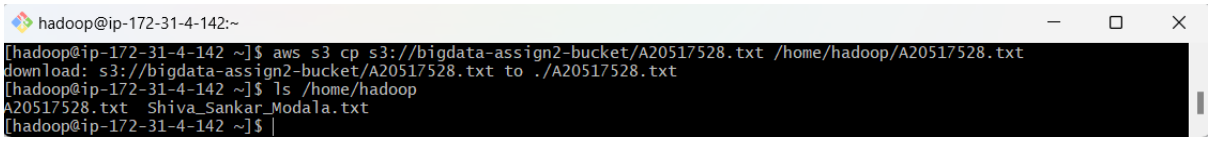
[hadoop@ip-172-31-4-142 ~]$ hadoop distcp /user/csp554/Shiva_Sankar_Modala.txt /user/csp554-2



15) Copy the object myid.txt you uploaded to an S3 bucket into the Hadoop master node Linux file system. The actual object includes your student id as above. Note, Amazon EMR and Hadoop provide a variety of file systems that you can use with EMR. You specify which file system to use with a file system prefix.

[hadoop@ip-172-31-4-142 ~]$ aws s3 cp s3://bigdata-assign2-bucket/A20517528.txt /home/hadoop/A20517528.txt

[hadoop@ip-172-31-4-142 ~]$ ls /home/Hadoop

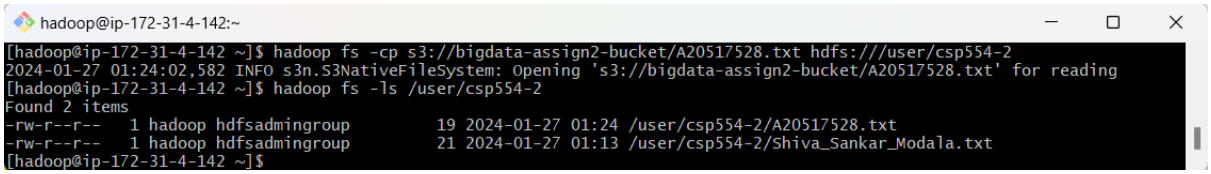


16) Copy the same object myid.txt you created in an S3 bucket into HDFS into the directory /users/csp554 hadoop fs -cp s3://mybucket/myid.txt hdfs:///user/csp554-2 Note, the three slashes after the “hdfs:” After you executed the above command, execute another command (you needed to figure out which one) to list the files and directories under the hdfs directory listed below: /user/csp554-2

Write down the command you executed and also take a screen snapshot of names of the files or directories that are listed and include it in your assignment submission.

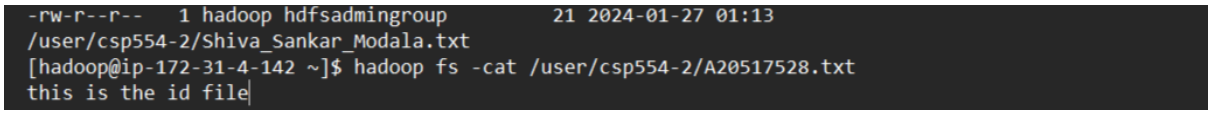
[hadoop@ip-172-31-4-142 ~]$ hadoop fs -cp s3://bigdata-assign2-bucket/A20517528.txt hdfs:///user/csp554-2

[hadoop@ip-172-31-4-142 ~]$ hadoop fs -ls /user/csp554-2



17) Execute a command to show the contents of the myid.txt file in the hdfs directory /user/csp554-2 Clue: look up about how to use the “cat” command in the file system shell document. Write down the command you executed and also take a screen snapshot of the listed content of the file and include it in your assignment submission.

[hadoop@ip-172-31-4-142 ~]$ hadoop fs -cat /user/csp554-2/A20517528.txt



18) Execute a command to remove the myid.txt file in the hdfs directory /user/csp554-2 Clue: look up about how to use the “rm” command in the file system shell document. Write down the command you executed, then list the content of the /user/csp554-2 HDFS directory and take a screen snapshot of the listed content of the directory and include it in your assignment submission.

[hadoop@ip-172-31-4-142 ~]$ hadoop fs -rm /user/csp554-2//A20517528.txt