Girish Rajani
A20503736
CSP-554 Big Data Technologies
Homework 2

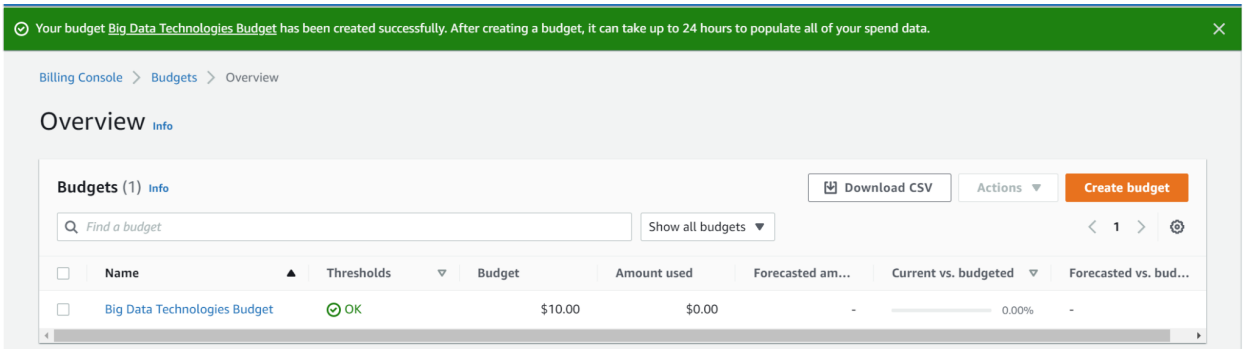

Figure 1 - Screenshot showing the creation of a usage budget with alerts

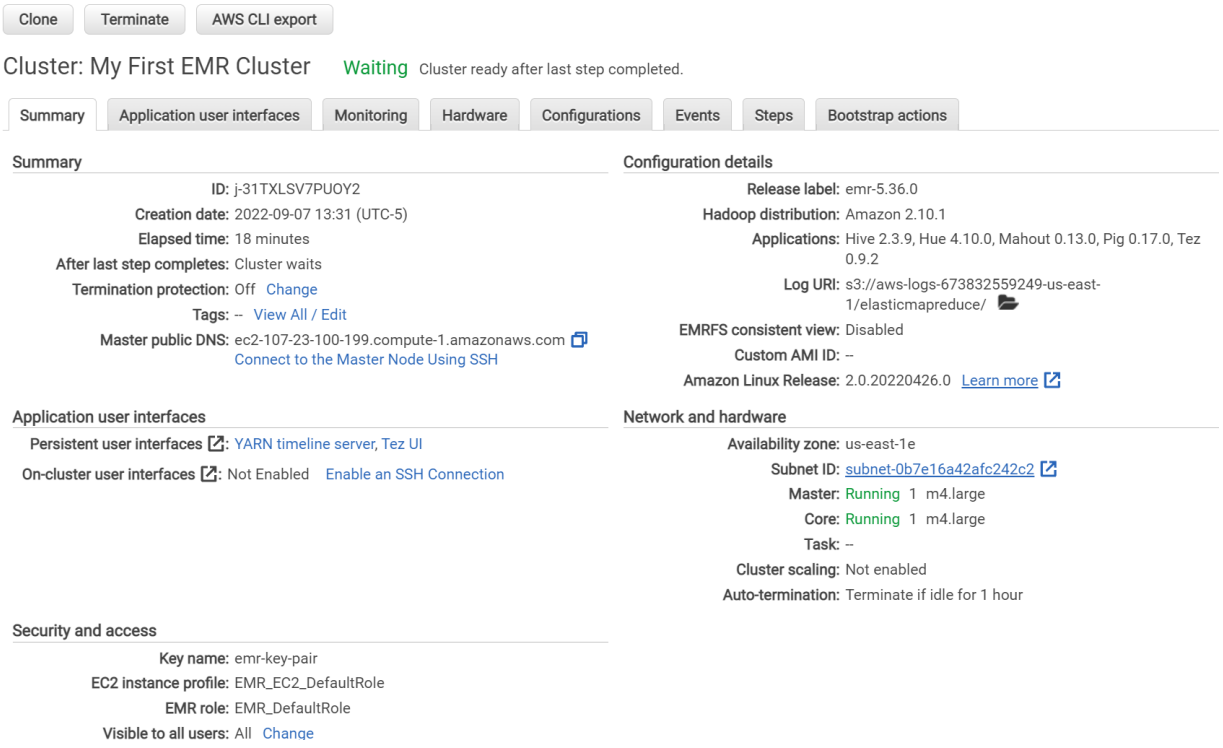

Figure 2 - Screenshot showing the successful creation of the Hadoop Cluster

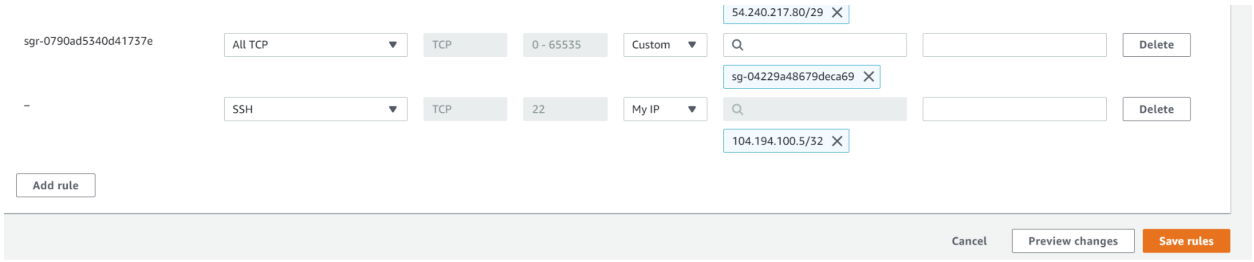

Figure 3 - Showing the creation of a new inbound rule to allow SSH connection

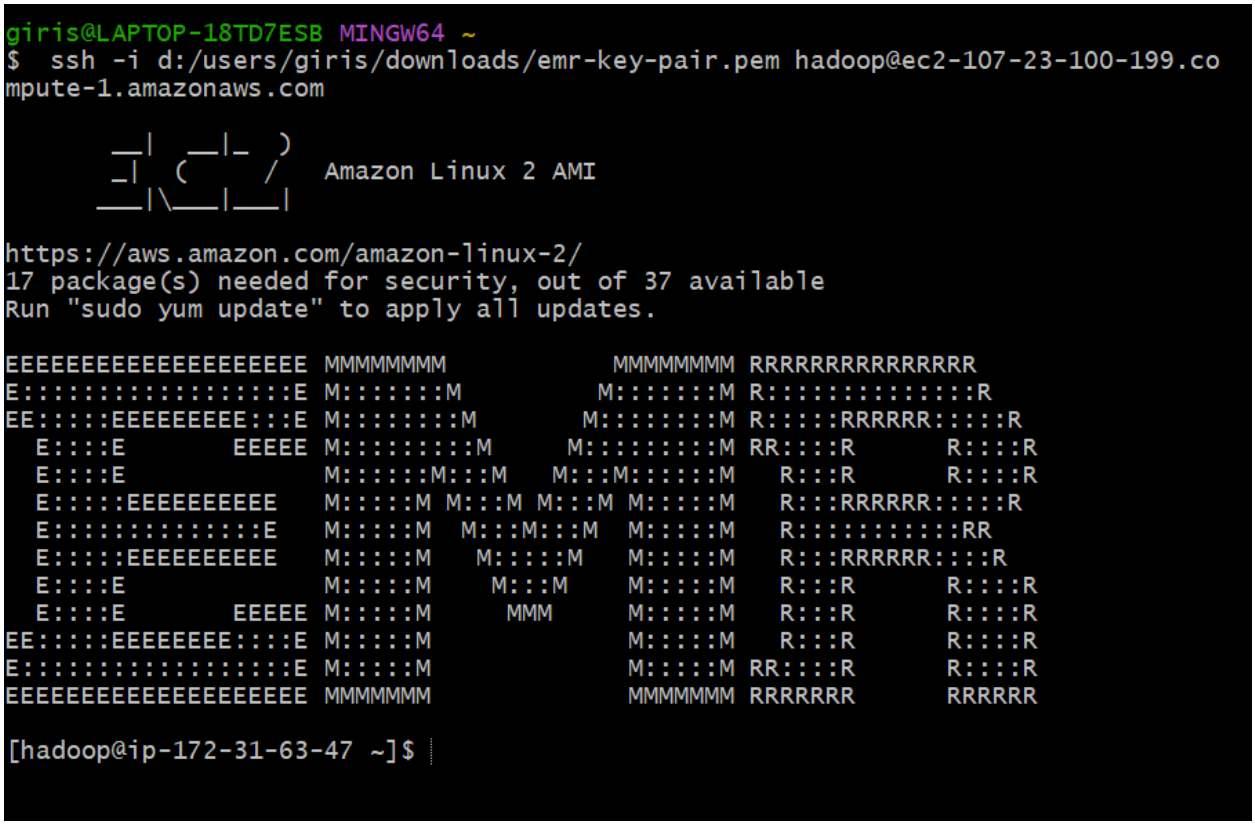

Figure 4 - Showing successful connection to Master Node using SSH

Figure 5 - Showing the successful upload of the myname.txt file from the local machine to the home directory of the Hadoop master node account

```
[hadoop@ip-172-31-63-47 ~]$ ls
girishrajani.txt
[hadoop@ip-172-31-63-47 ~]$
```

Figure 6 - using the ls command to confirm that the file is appearing in the Hadoop master node

9. (2 points) 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 fs –ls /

```
[hadoop@ip-172-31-63-47 ~]$ hadoop fs -ls /
Found 4 items
drwxr-xr-x   - hdfs hdfsadmingroup          0 2022-09-07 18:39 /apps
drwxrwxrwt   - hdfs hdfsadmingroup          0 2022-09-07 18:43 /tmp
drwxr-xr-x   - hdfs hdfsadmingroup          0 2022-09-07 18:39 /user
drwxr-xr-x   - hdfs hdfsadmingroup          0 2022-09-07 18:39 /var
[hadoop@ip-172-31-63-47 ~]$
```

Figure 7 - Using the hadoop fs -ls / command to view the files/directories

10. (2 points) Execute a command (you needed to figure out which one) to list the files and directories under the hdfs directory listed below: /user

Command used was hadoop fs -ls /user

```
[hadoop@ip-172-31-63-47 ~]$ hadoop fs -ls /user
Found 6 items
drwxrwxrwx   - hadoop hdfsadmingroup          0 2022-09-07 18:39 /user/hadoop
drwxr-xr-x   - mapred mapred                 0 2022-09-07 18:39 /user/history
drwxrwxrwx   - hdfs   hdfsadmingroup         0 2022-09-07 18:39 /user/hive
drwxrwxrwx   - hue    hue                    0 2022-09-07 18:39 /user/hue
drwxrwxrwx   - oozie  oozie                  0 2022-09-07 18:43 /user/oozie
drwxrwxrwx   - root   hdfsadmingroup         0 2022-09-07 18:39 /user/root
[hadoop@ip-172-31-63-47 ~]$
```

Figure 8 - Using the hadoop fs -ls /user command to view the files under the hdfs user directory

11. (2 points) Execute a command to create the following HDFS directory: /user/csp554

Command used was hadoop fs -mkdir /user/csp554

```
[hadoop@ip-172-31-63-47 ~]$ hadoop fs -mkdir /user/csp554
```

Figure 9 - Using the hadoop fs -mkdir /user/csp554 command to create the HDFS directory

12. (2 points) Execute a command to create the following HDFS directory: /user/csp554-2

```
[hadoop@ip-172-31-63-47 ~]$ hadoop fs -mkdir /user/csp554-2
```

Figure 10 - Using the hadoop fs -mkdir /user/csp554-2 command to create the HDFS directory

Command hadoop fs -ls /user was used to double check if the csp554 and csp554-2 directories were created

```
[hadoop@ip-172-31-63-47 ~]$ hadoop fs -ls /user
Found 8 items
drwxr-xr-x   - hadoop hdfsadmingroup          0 2022-09-07 19:00 /user/csp554
drwxr-xr-x   - hadoop hdfsadmingroup          0 2022-09-07 19:03 /user/csp554-2
drwxrwxrwx   - hadoop hdfsadmingroup          0 2022-09-07 18:39 /user/hadoop
drwxr-xr-x   - mapred mapred                  0 2022-09-07 18:39 /user/history
drwxrwxrwx   - hdfs   hdfsadmingroup          0 2022-09-07 18:39 /user/hive
drwxrwxrwx   - hue    hue                     0 2022-09-07 18:39 /user/hue
drwxrwxrwx   - oozie  oozie                   0 2022-09-07 18:43 /user/oozie
drwxrwxrwx   - root   hdfsadmingroup          0 2022-09-07 18:39 /user/root
[hadoop@ip-172-31-63-47 ~]$
```

Figure 11 - Using the hadoop fs -ls /user command to confirm that /user/csp554 and /user/csp554-2 directories were successfully created

13. (2 points) 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

```
[hadoop@ip-172-31-63-47 ~]$ hadoop fs -put /home/hadoop/girishrajani.txt /user/csp554
```

Figure 12 - Showing command executed to copy the girishrajani.txt file from the Hadoop home directory to the /user/csp554 hdfs directory

```
[hadoop@ip-172-31-63-47 ~]$ hadoop fs -ls /user/csp554
Found 1 items
-rw-r--r--   1 hadoop hdfsadmingroup         23 2022-09-07 19:46 /user/csp554/girishrajani.txt
[hadoop@ip-172-31-63-47 ~]$
```

Figure 13 - Using hadoop fs -ls /user/csp554 command to see if girishrajani.txt has been successfully copied to the /user/csp554 directory

Note: From this question onwards, my EMR cluster was terminated due to timeout so I had to clone the cluster which gave me a different Master public DNS and I had to repeat the above steps using this new connection, hence the reason for a different Hadoop IP.

14. (2 points) 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

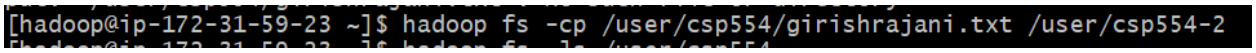

Figure 14 - Using the cp command to copy the girishrajani.txt file from the /user/csp554 directory to the /user/csp554-2 directory

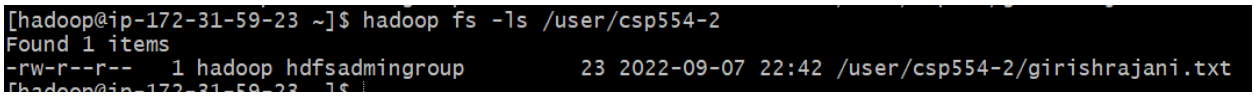

Figure 15 - Using hadoop fs -ls /user/csp554-2 to confirm that girishrajani.txt is in the /user/csp554-2 directory

15. (2 points) 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.

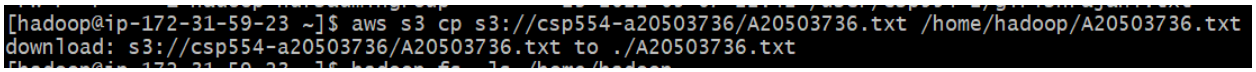

Figure 16 - Command used to copy the object A20503736.txt from my csp554-a20503736 S3 bucket into the Hadoop master node file system

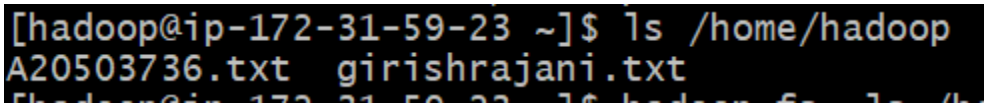

Figure 17 - ls /home/hadoop was used to see if A20503736.txt was successfully copied from the S3 bucket to the Hadoop master node file system

16. (2 points) 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:”

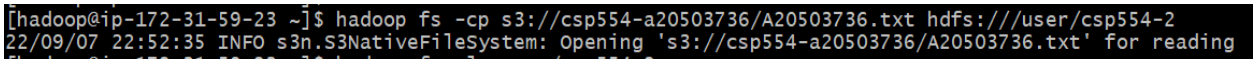

Figure 18 - Showing command used to copy the A20503736.txt from the S3 Bucket into the HDFS directory /user/csp554-2

To list the files and directories under the hdfs directory listed below: /user/csp554-2, the following code was used:
hadoop fs -ls /user/csp554-2

```
[hadoop@ip-172-31-59-23 ~]$ hadoop fs -ls /user/csp554-2
Found 2 items
-rw-r--r--   1 hadoop hdfsadmingroup         21 2022-09-07 22:52 /user/csp554-2/A20503736.txt
-rw-r--r--   1 hadoop hdfsadmingroup         23 2022-09-07 22:42 /user/csp554-2/girishrajani.txt
```

Figure 19 - Using hadoop fs -ls/user/csp554-2 to confirm that the A20503736.txt file is successfully appearing in the /user/csp554-2 directory

17. (2 points) 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

Code executed: hadoop fs -cat /user/csp554-2/A20503736.txt

```
[hadoop@ip-172-31-59-23 ~]$ hadoop fs -cat /user/csp554-2/A20503736.txt
this is the id file
```

Figure 20 - Showing the “cat” command used to display the contents of the A20503736.txt file

18. (2 points) 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.

Code used to remove the A20503736.txt file is:
hadoop fs -rm -r /user/csp554-2/A20503736.txt

```
[hadoop@ip-172-31-59-23 ~]$ hadoop fs -rm -r /user/csp554-2/A20503736.txt
Deleted /user/csp554-2/A20503736.txt
```

Figure 21 - Showing the “rm” command used to remove the A20503736.txt file

```
[hadoop@ip-172-31-59-23 ~]$ hadoop fs -ls /user/csp554-2
Found 1 items
-rw-r--r--   1 hadoop hdfsadmingroup         23 2022-09-07 22:42 /user/csp554-2/
girishrajani.txt
```

Figure 22 - Screenshot of the listed content of user/csp554-2 directory after the file was deleted