DSC 650

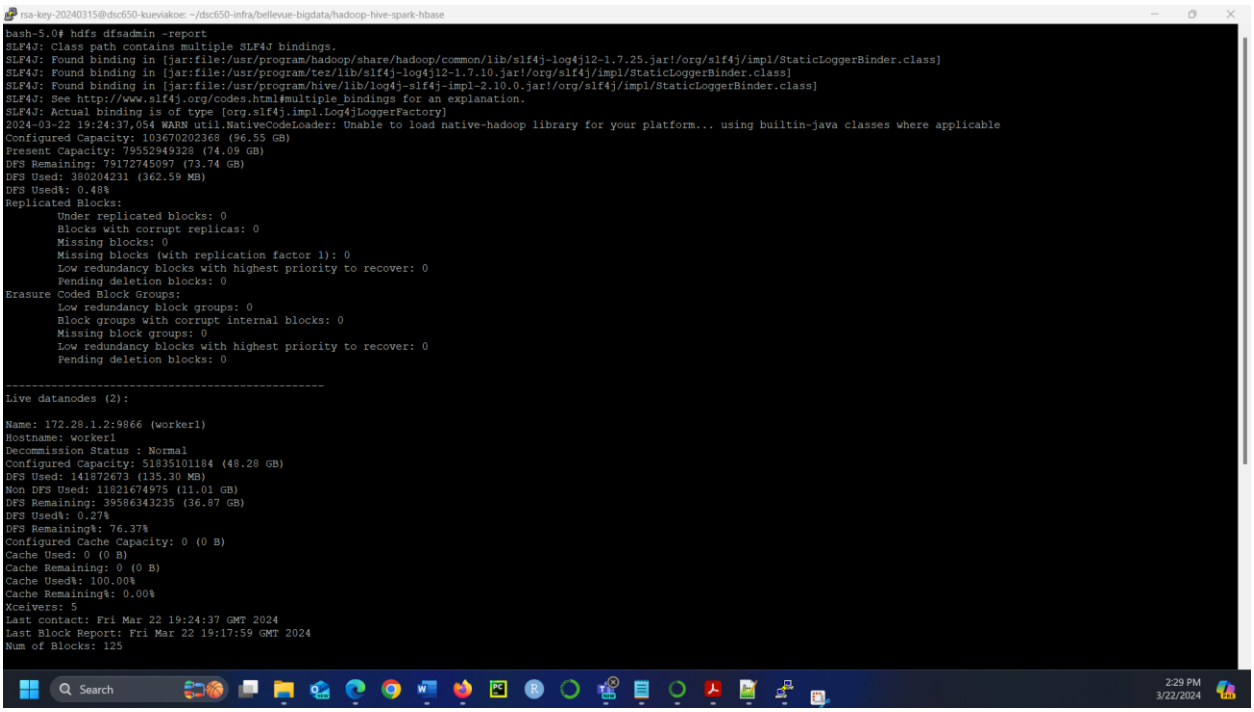
Week 2 Assignment

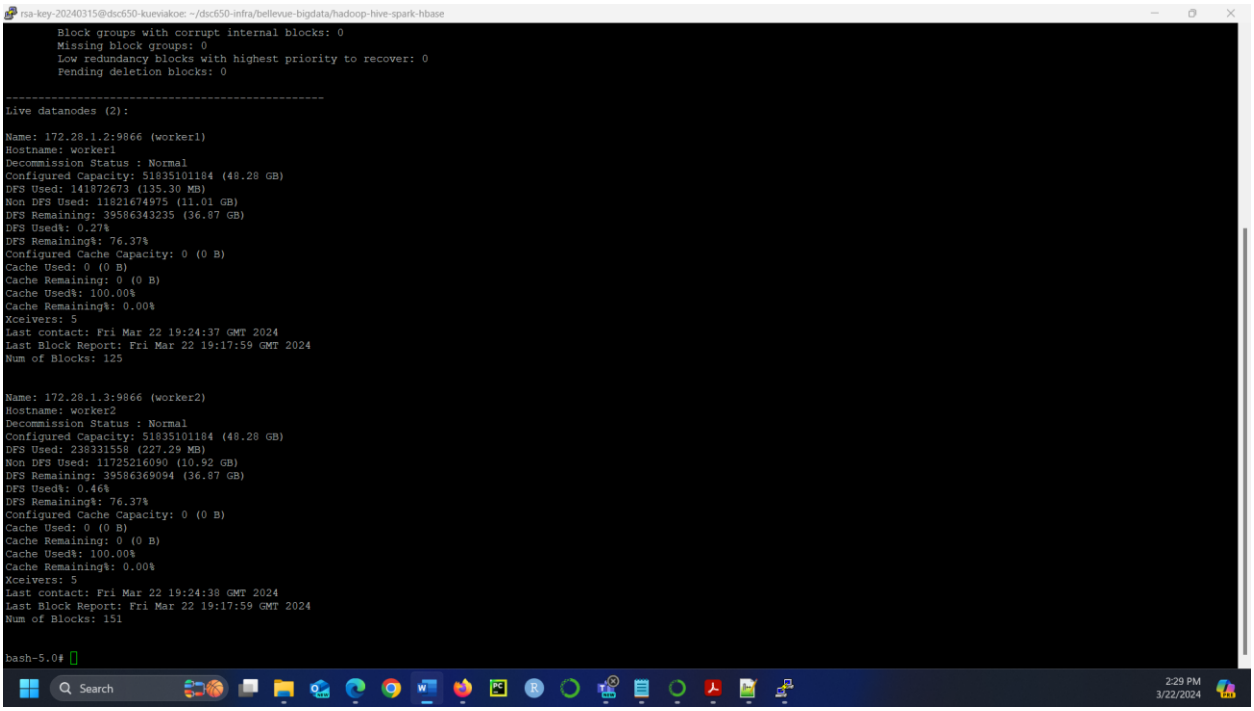
Eyram Kueviakoe

March 22, 2024

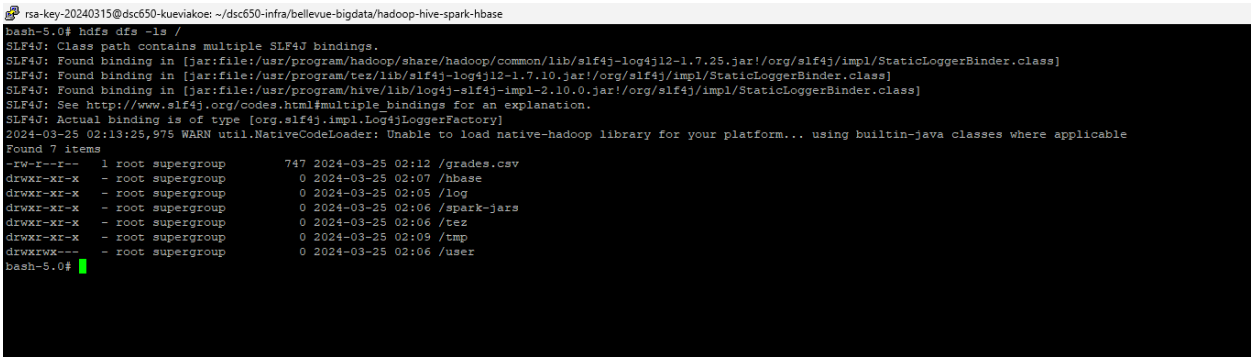
Deep Dive into HDFS

Screenshot 1: *hdfs dfsadmin -report*





Screenshot 2: Screenshot proving the data has been loaded
hdfs dfs -ls /

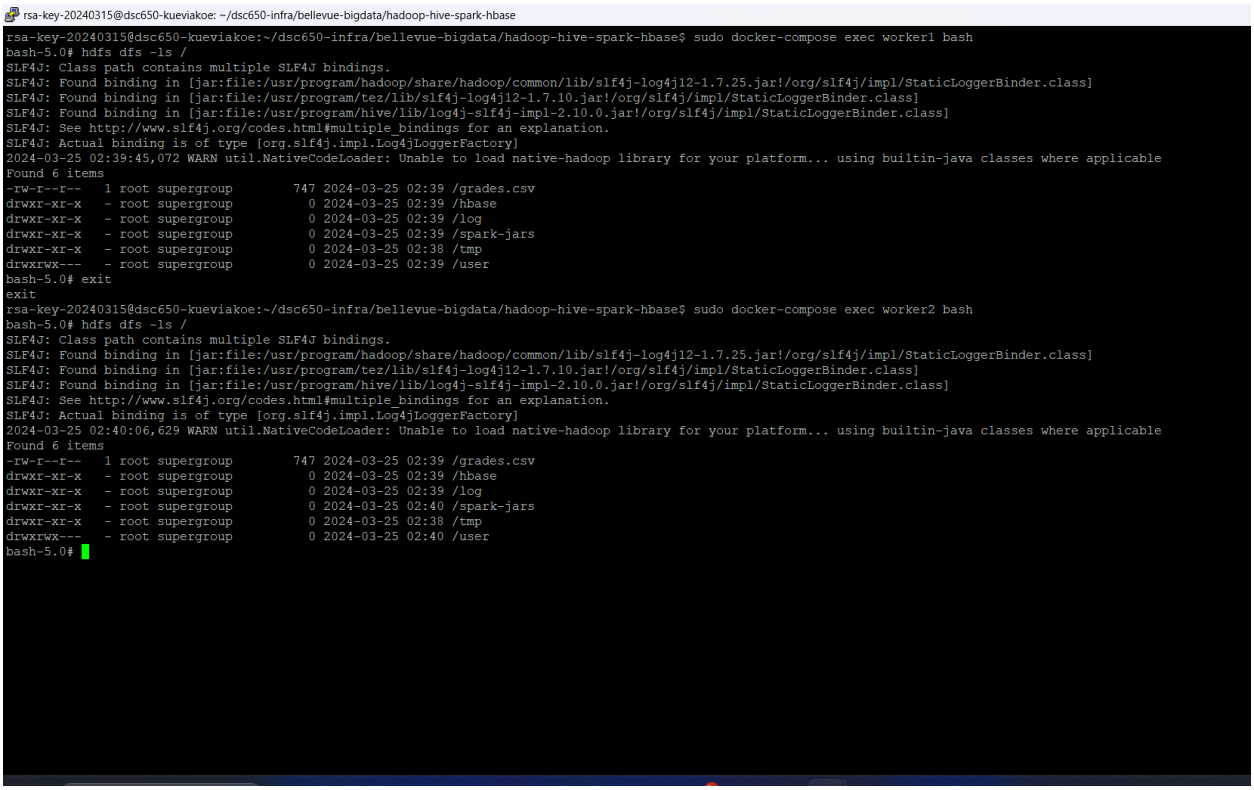


Screenshot:

docker-compose exec worker1 bash
hdfs dfs -ls /

and

docker-compose exec worker2 bash
hdfs dfs -ls /



Screenshots of the three chosen HDFS command outputs.

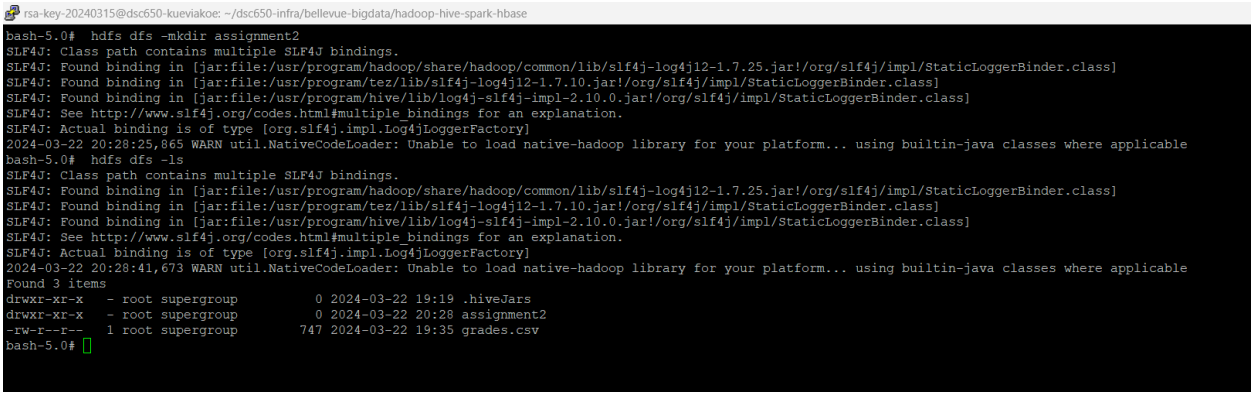
Command 1

hdfs dfs -mkdir assignment2

and

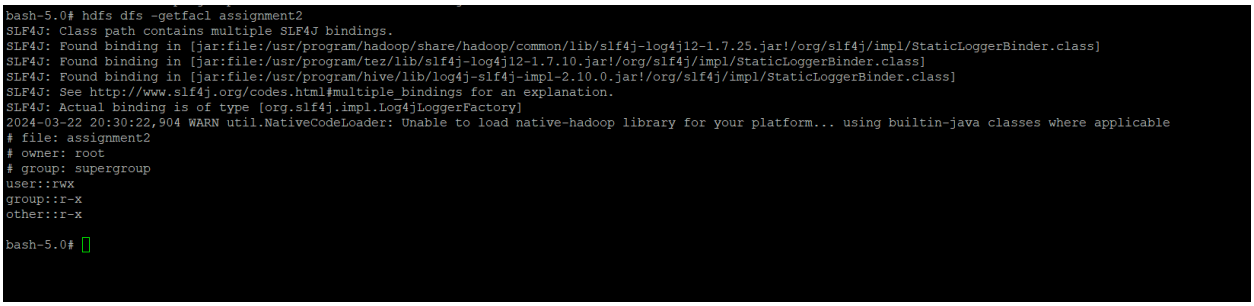
hdfs dfs -ls

to show the folder was created



Command 2

hdfs dfs -getfacl assignment2



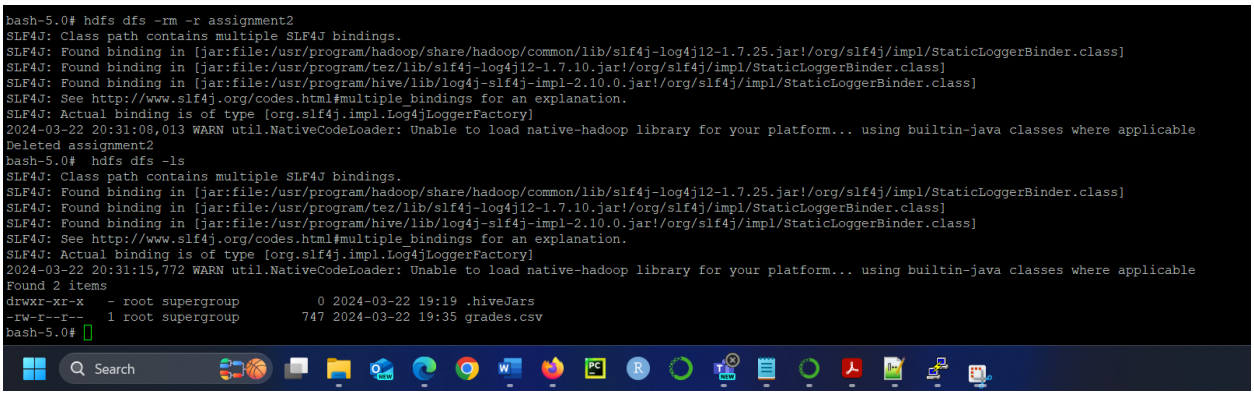
Command 3

hdfs dfs -rm -r assignment2

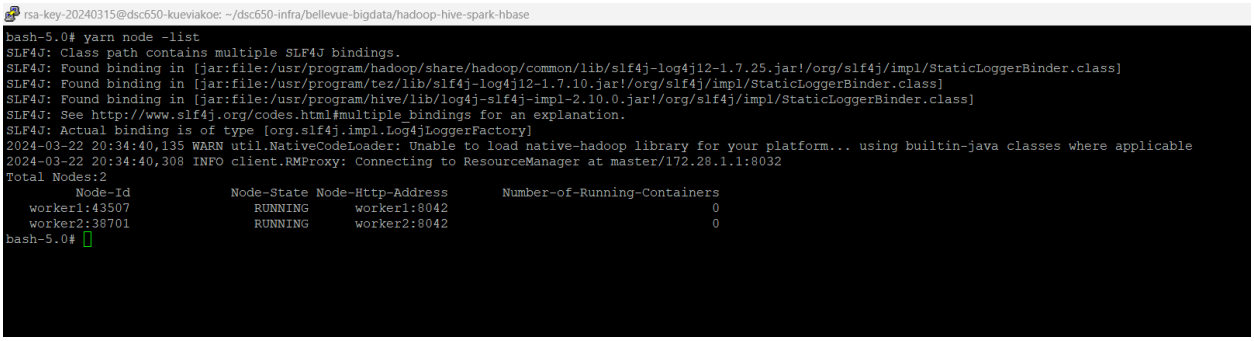
and

hdfs dfs -ls

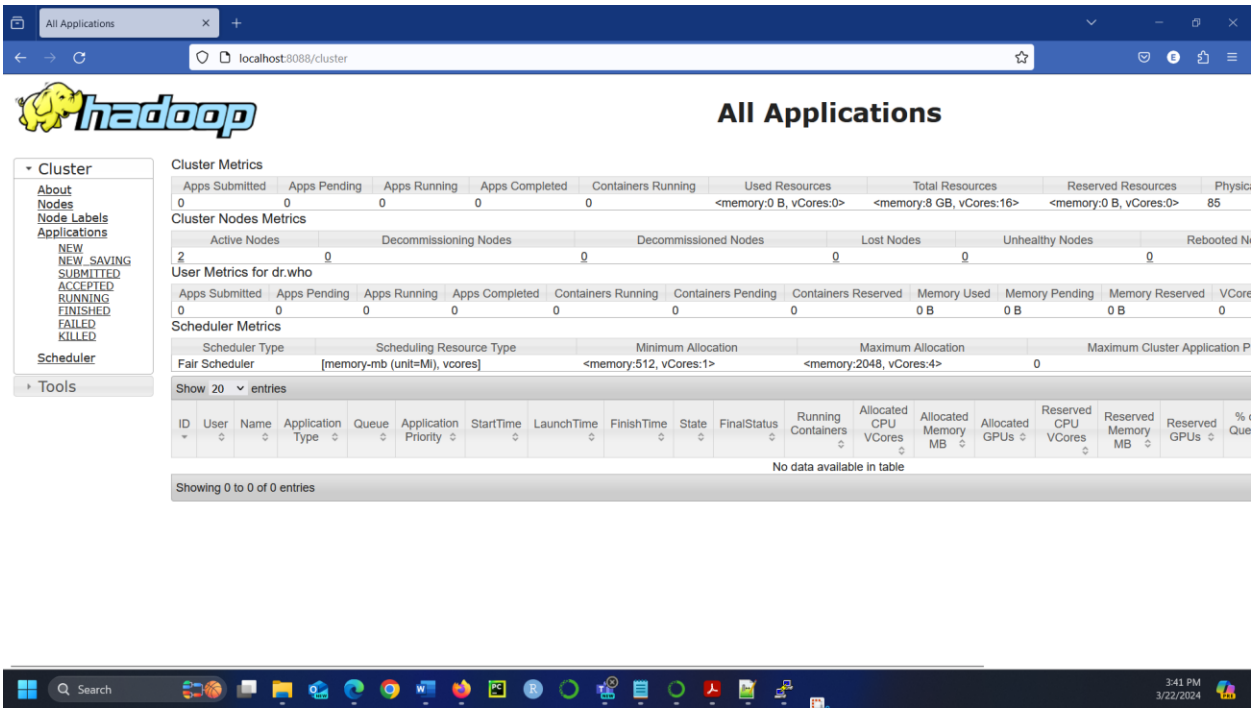
to show the folder is deleted



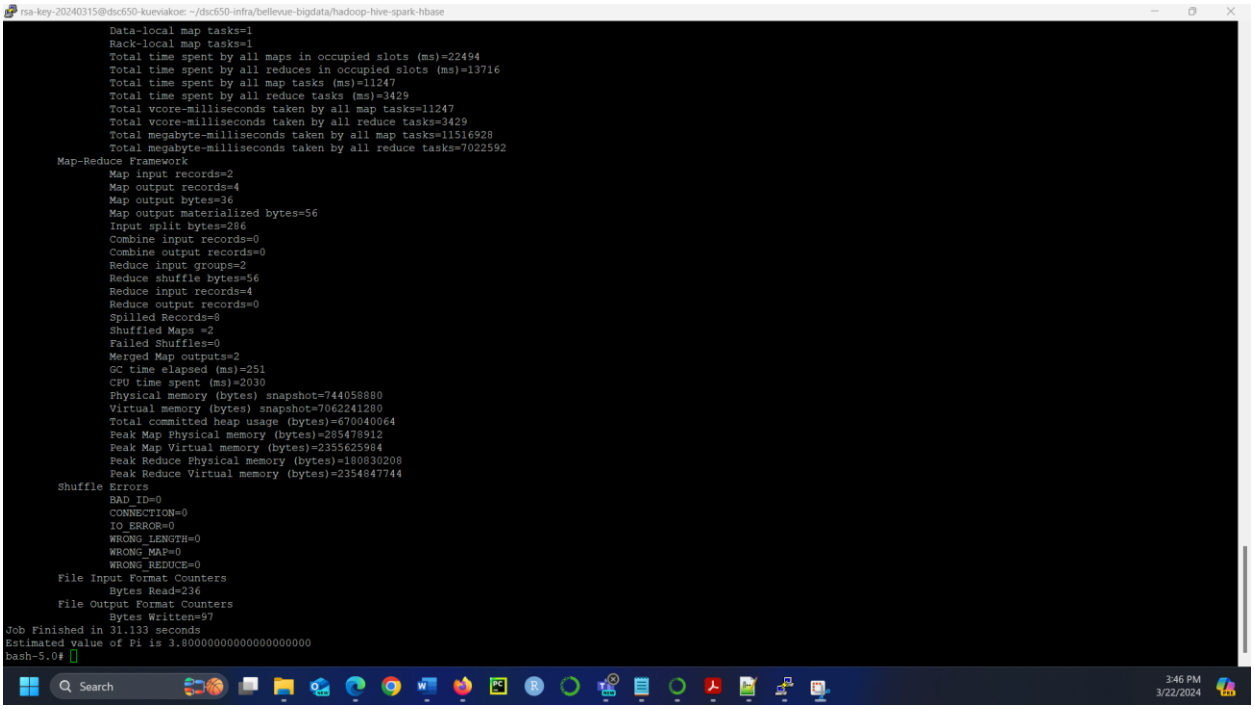
Screenshot
yarn node -list

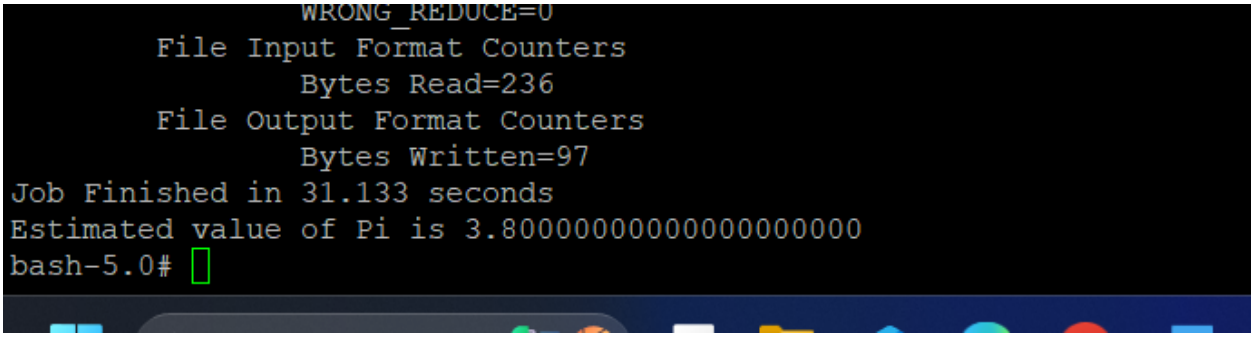


Screenshot from the YARN UI showing the updated maximum memory



Experimenting with MapReduce





Summary and significance of the result:

The estimated Pi value of 3.8 from the MapReduce job is higher than the actual Pi value. This means that the calculation wasn't very accurate.

There are many reasons that can explain the inaccuracy:

- We are not using enough samples or data points
- The random numbers are not properly generated.

To get an accurate result, we will need a larger sample. This will reduce the estimation error. Also the random number generation should be tuned to generate points that are evenly distributed.