Homework_1

Name: Rahul Purushottam Gaonkar(rpg283)

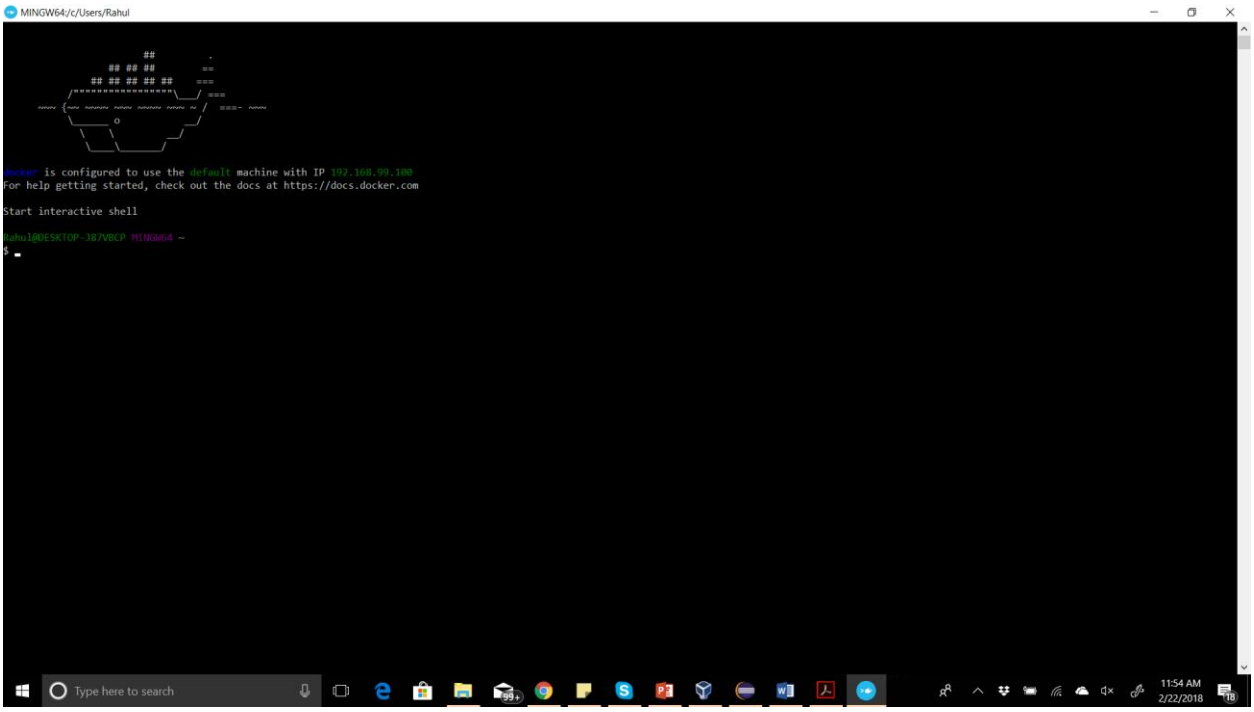

Figure1: Docker Running

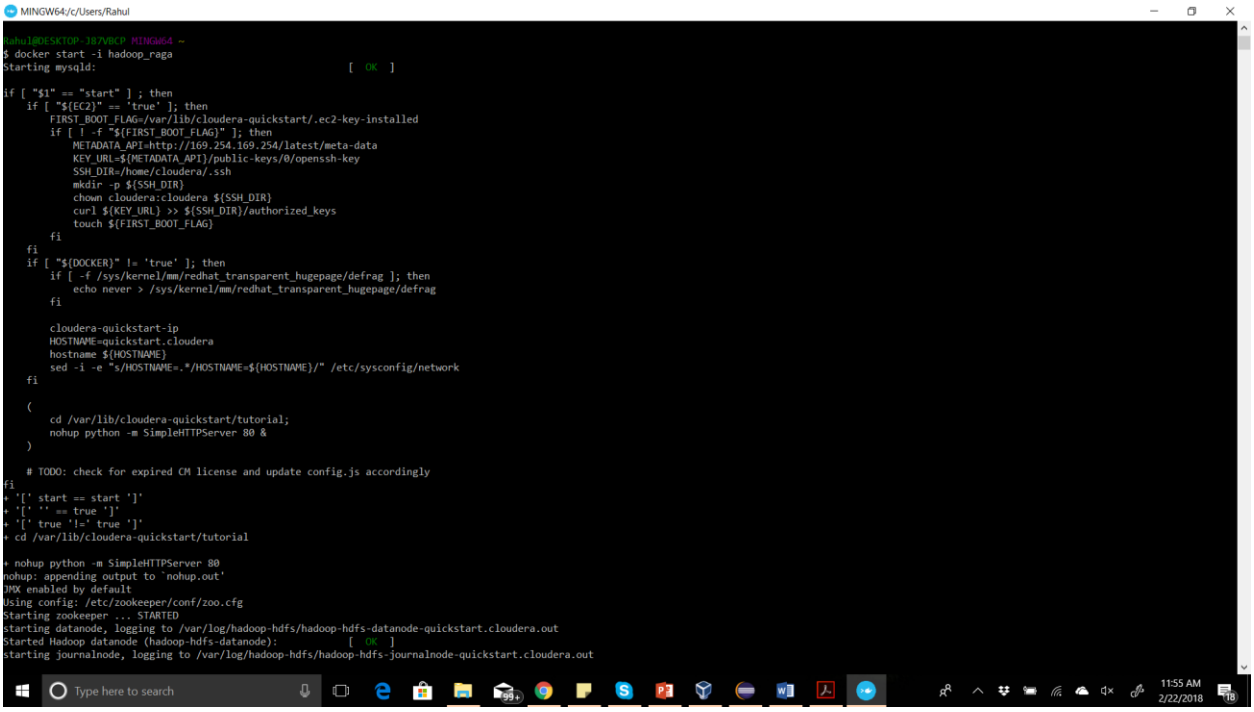

Figure 2: Starting Hadoop Container

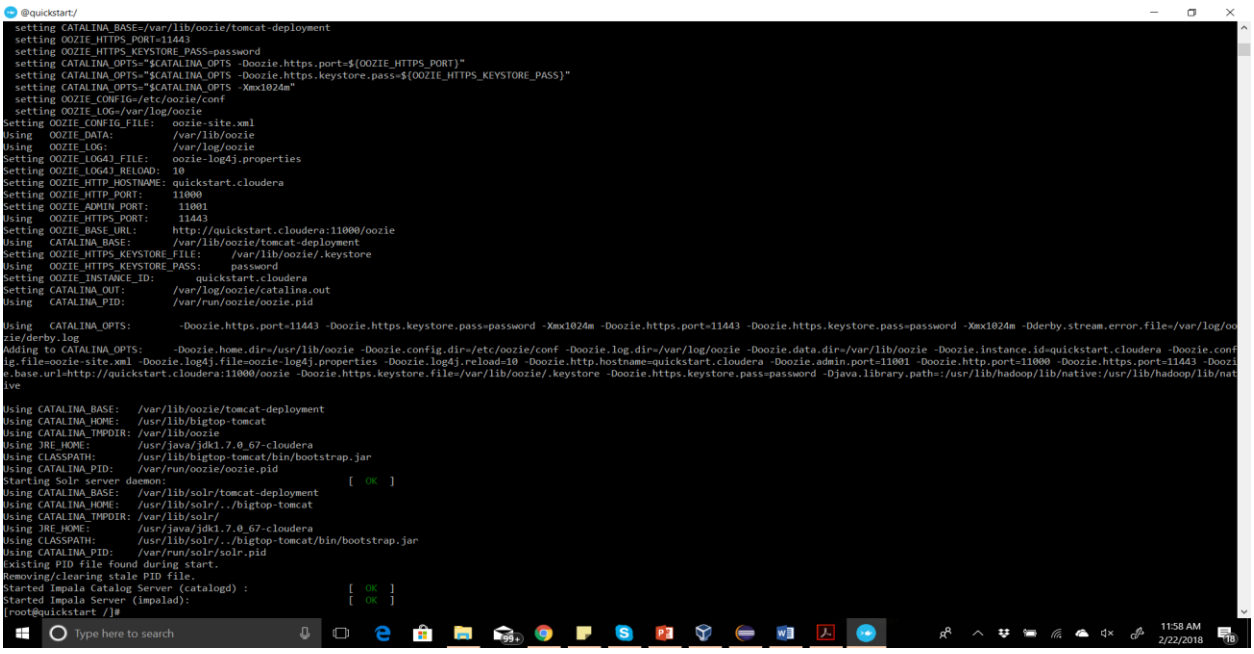

Figure 3: Hadoop Container Started

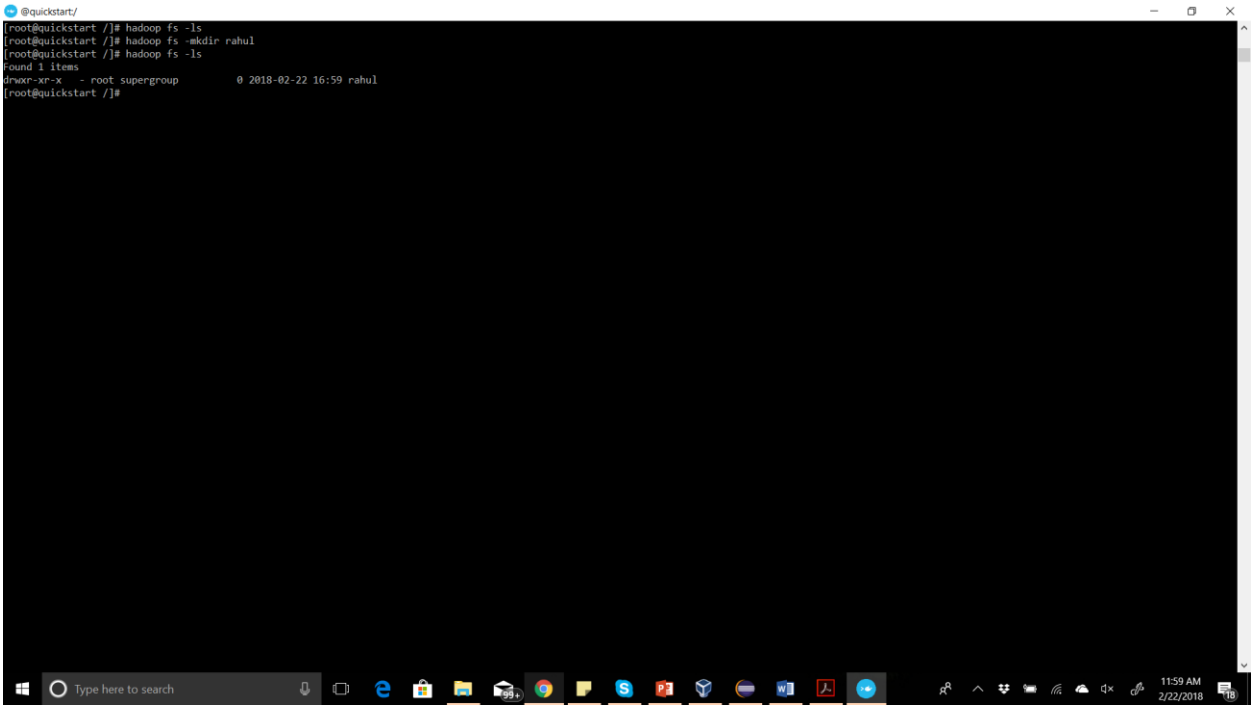

Figure 4: Created a directory of name Rahul in Hadoop filesystem

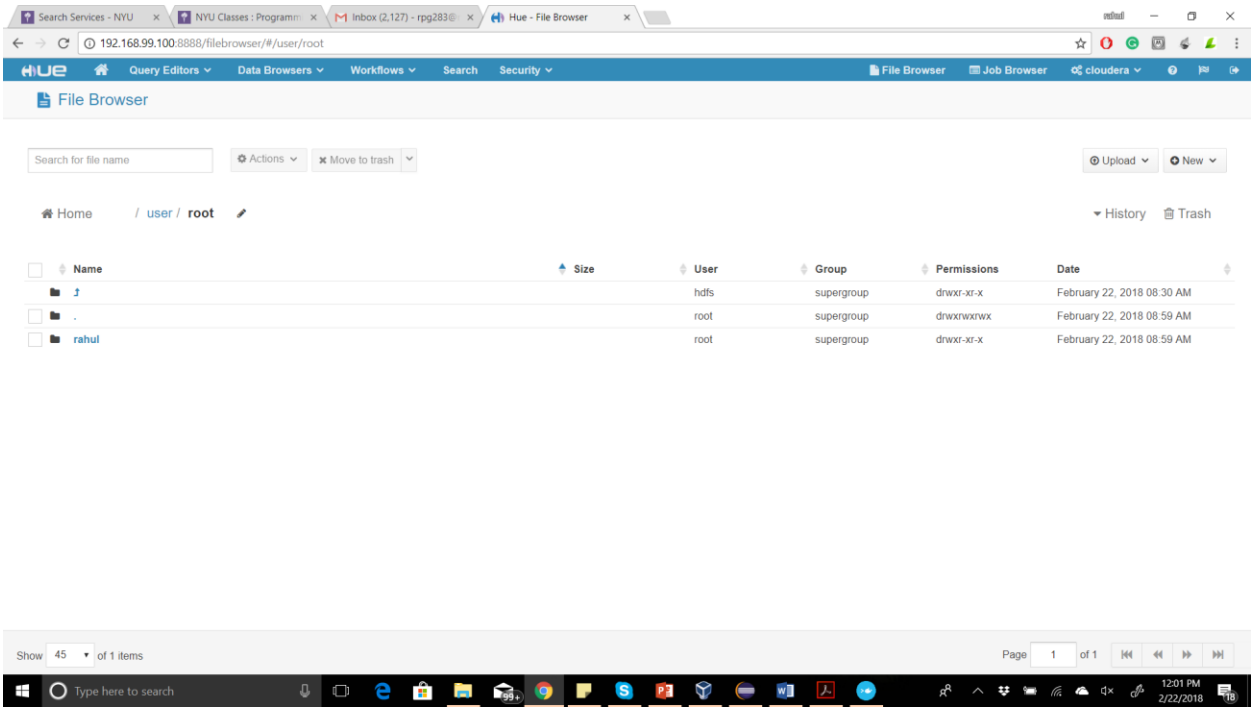

Figure 5: Directory name Rahul reflected in Hue

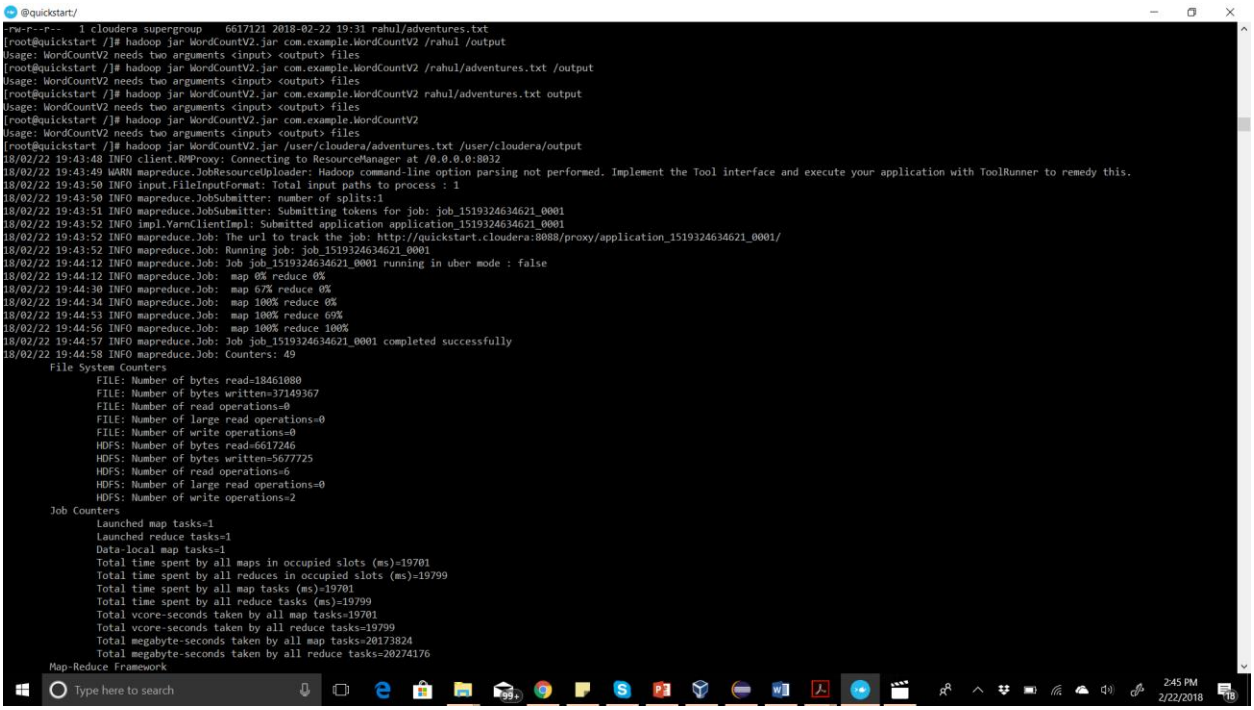

Figure 6: Running the MapReduce program(Job) of WordCountV2 to get the Phrase Count (i.e. Word Pair Count)

Command Used: hadoop jar WordCountV2.jar /user/cloudera/adventures.txt /user/cloudera/output

Placed the input file in user/cloudera and set the output path as /user/cloudera/output.

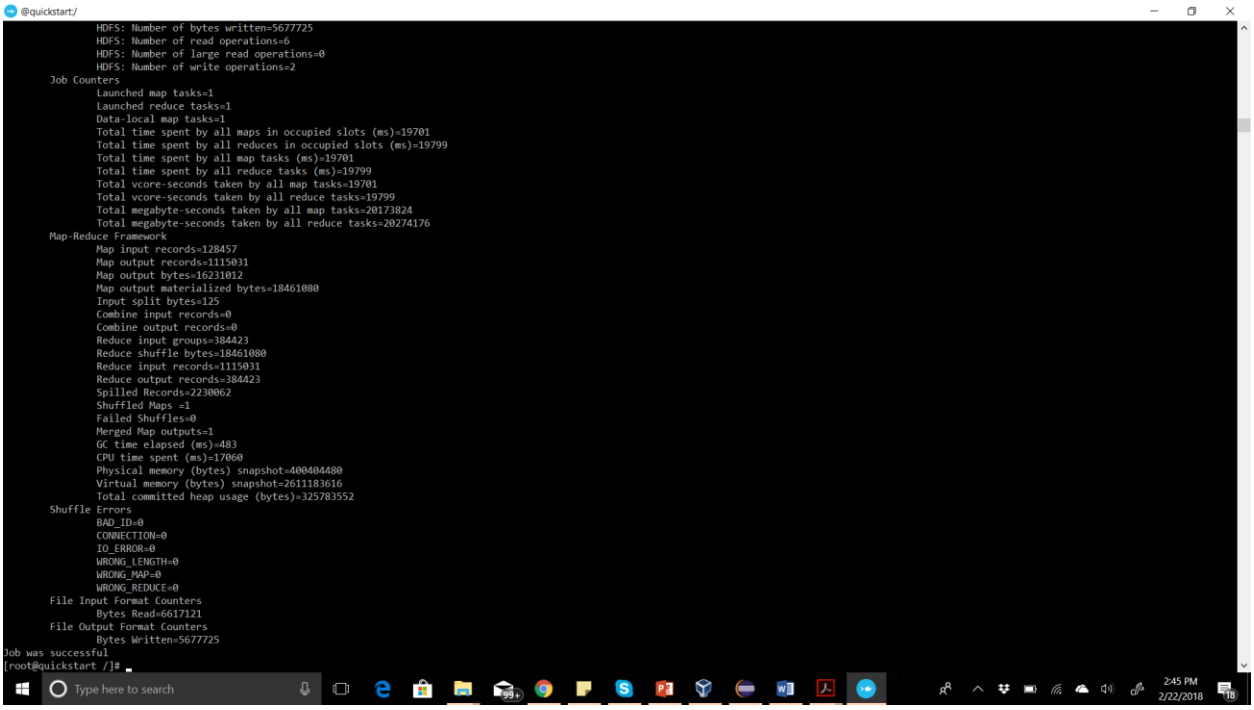

Figure 7: Job was Successful

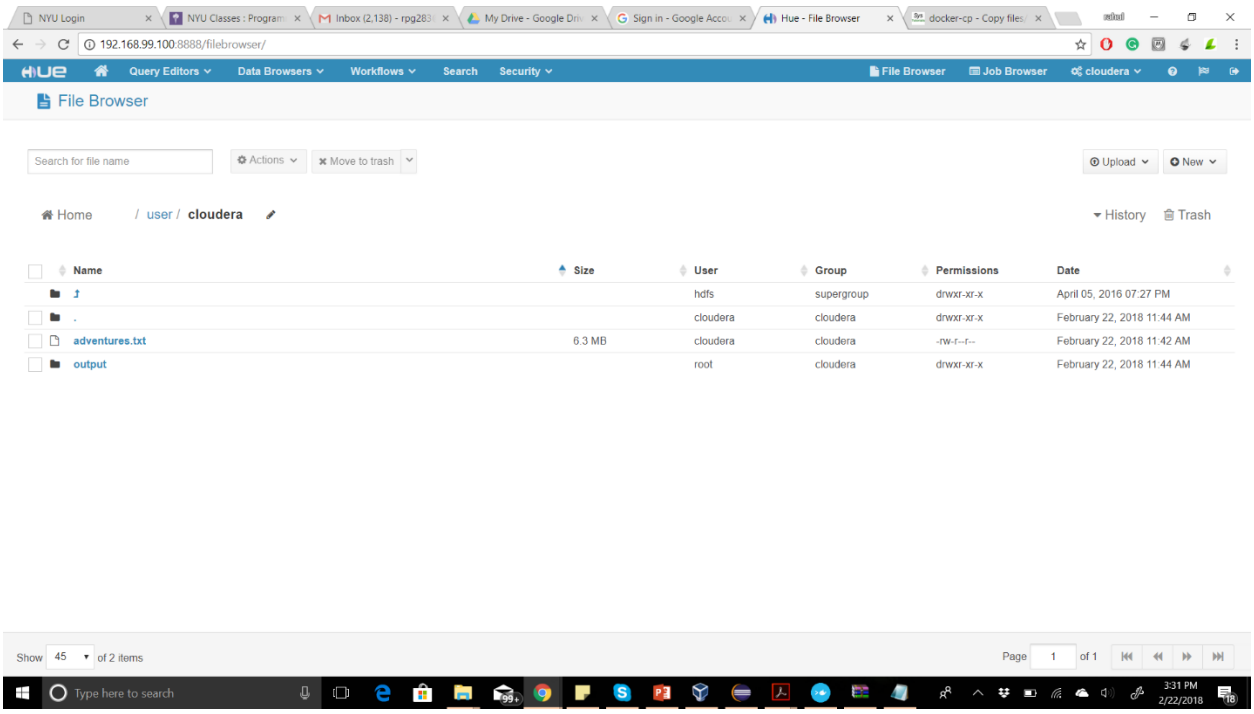

Figure 8: Output folder generated in file HDFS (shown in hue)

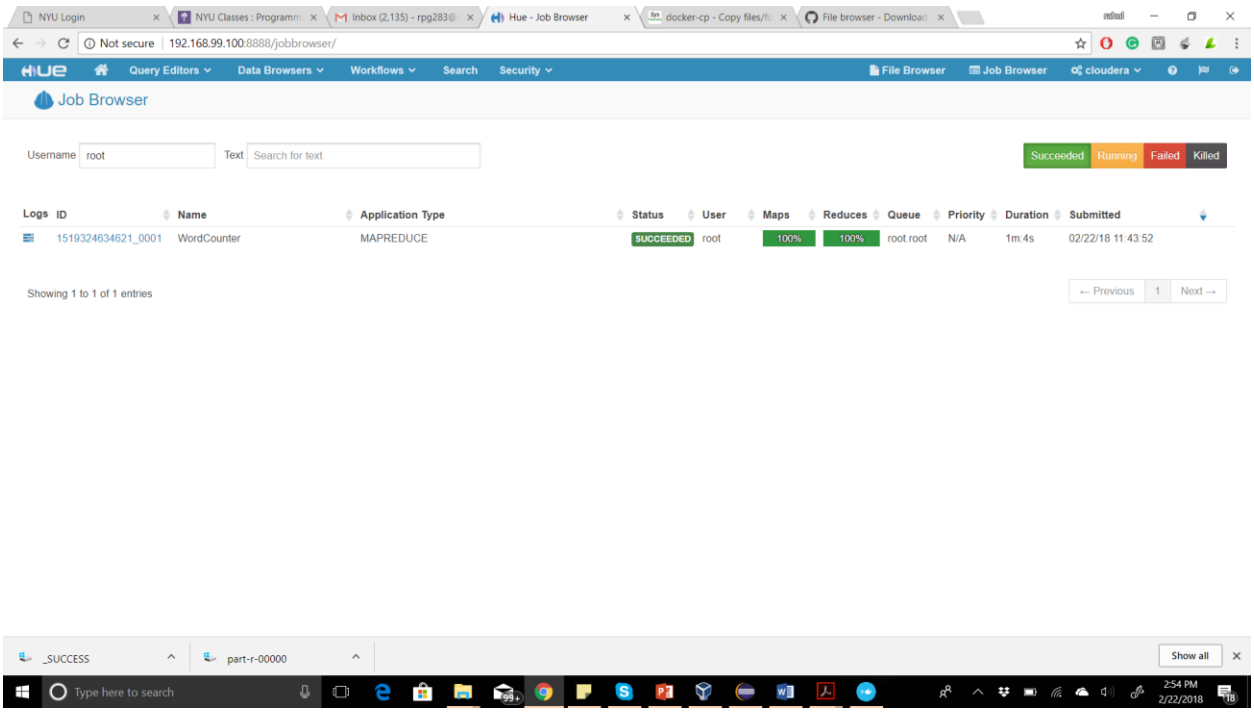

Figure 9: Job Successful Status in hue

Steps to run the Job:

1. Place the WordCountV2.jar file in the hadoop container.
2. Place the adventures.txt file in the HDFS. I had placed it in user/cloudera/adventures.txt
3. Then run the job by executing the command **hadoop jar WordCountV2.jar /user/cloudera/adventures.txt /user/cloudera/output**
4. The output path argument should be a unique path, or it will throw an exception. I have mentioned **/user/cloudera/output** where I will get the output file generated.