**Part 3:**

**Approach:**

1. As a part of this question I have written a program to partition the tweets based on hashtags and check if the tweets consist of the keyword needed (Covid or Olympics). If the keyword matches the hashtag is selected and we count the number of such tweets for each keyword, hashtag pair. Finally, limit the tags to 10 and print out the hashtags for both the Covid and Olympics keywords.
2. The Mapper class takes the tweet data ‘Text’ as key and value as 1, loops through the text, and checks if the keywords are present. If the keywords are present then from the tokens we extract the hashtags and then concatenate the hashtag with the keyword to form a key and value pair.
3. The practitioner class then makes a partition for each keyword and separates the key values pairs sending them to their respective reducers.
4. The Reducer class counts the total number of tweets for a particular keyword, hashtag pair and prints out the final output with the top 10 hashtags, keyword and count values respectively.
5. Compiled the java program (Tharuni\_Samineni\_Program\_3.java).
6. Created a jar file.
7. Used a command to run the MapReducePartitioner program.
8. Once the program is compiled and run successfully, the final output of the program can be viewed by writing a command.

**Instructions to run the code:**

3,1) To compile the (Tharuni\_Samineni\_Program\_2.java)map reduce partition program to Partition the tweets based on the hashtag and count the

number of rows in each partiton based on the keywords ‘Covid’ and ‘Olympics’ use the following command:

- $HADOOP\_HOME/bin/hadoop com.sun.tools.javac.Main Tharuni\_Samineni\_Program\_3.java

Note: There is a note shown after compilation command as unchecked or unsafe operation, which can be ignored as they might be caused due to error checking.

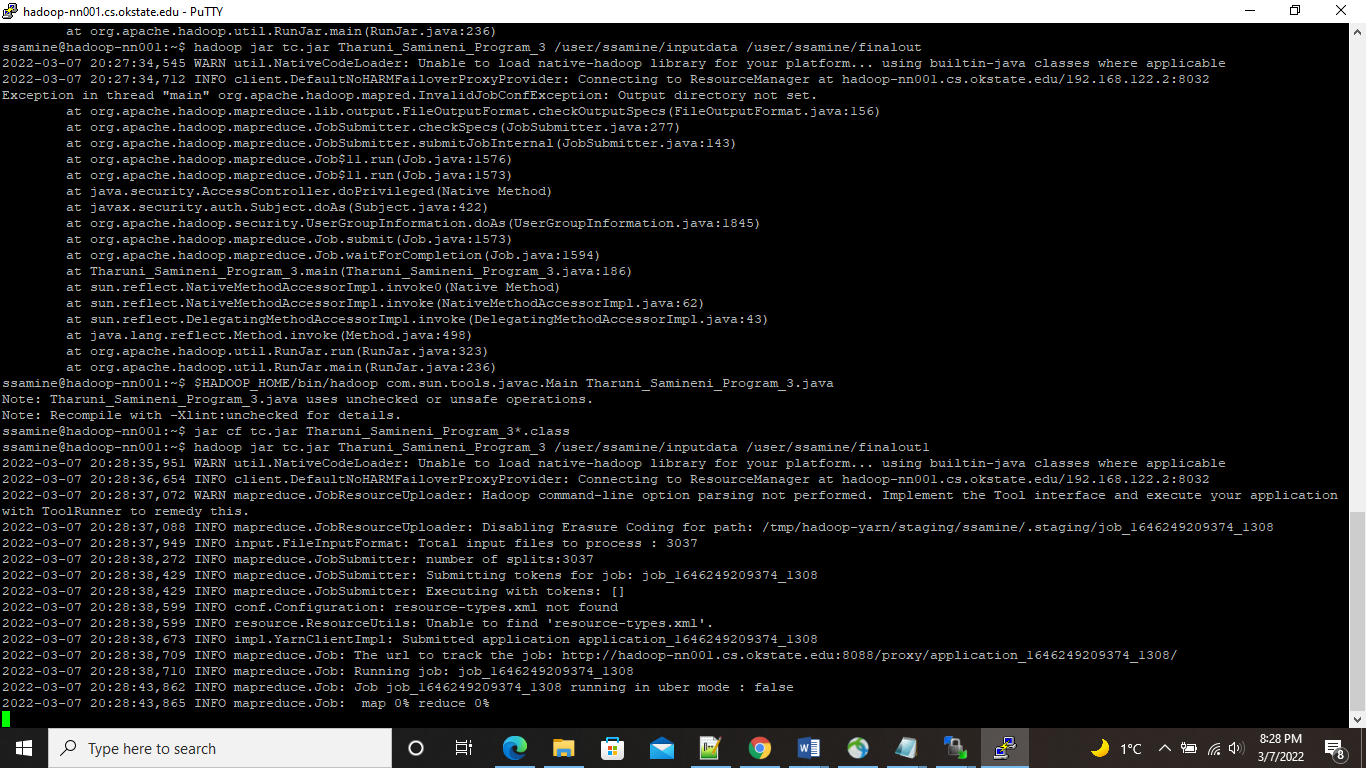


Figure 1: Shows the command the compile the program

3,2) To create a jar file use the following command:

- jar cf tc.jar Tharuni\_Samineni\_Program\_3\*.class

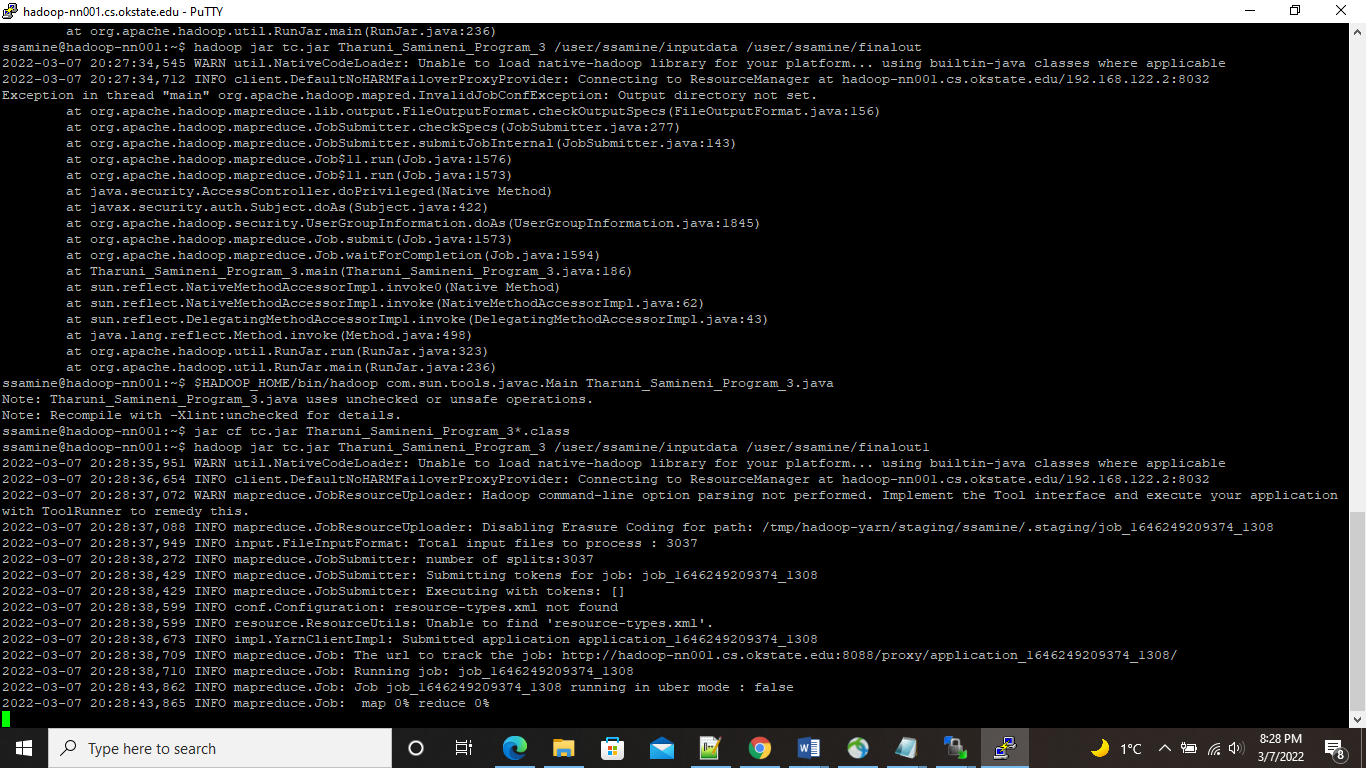


Figure 2: Shows the command to create a jar file

3,3) To run the map reduce partition program use the following command:

- hadoop jar tc.jar Tharuni\_Samineni\_Program\_2 /user/ssamine/inputdata /user/ssamine/finalout5

(Use different directories for output, otherwise an error will be shown that directory already exist)

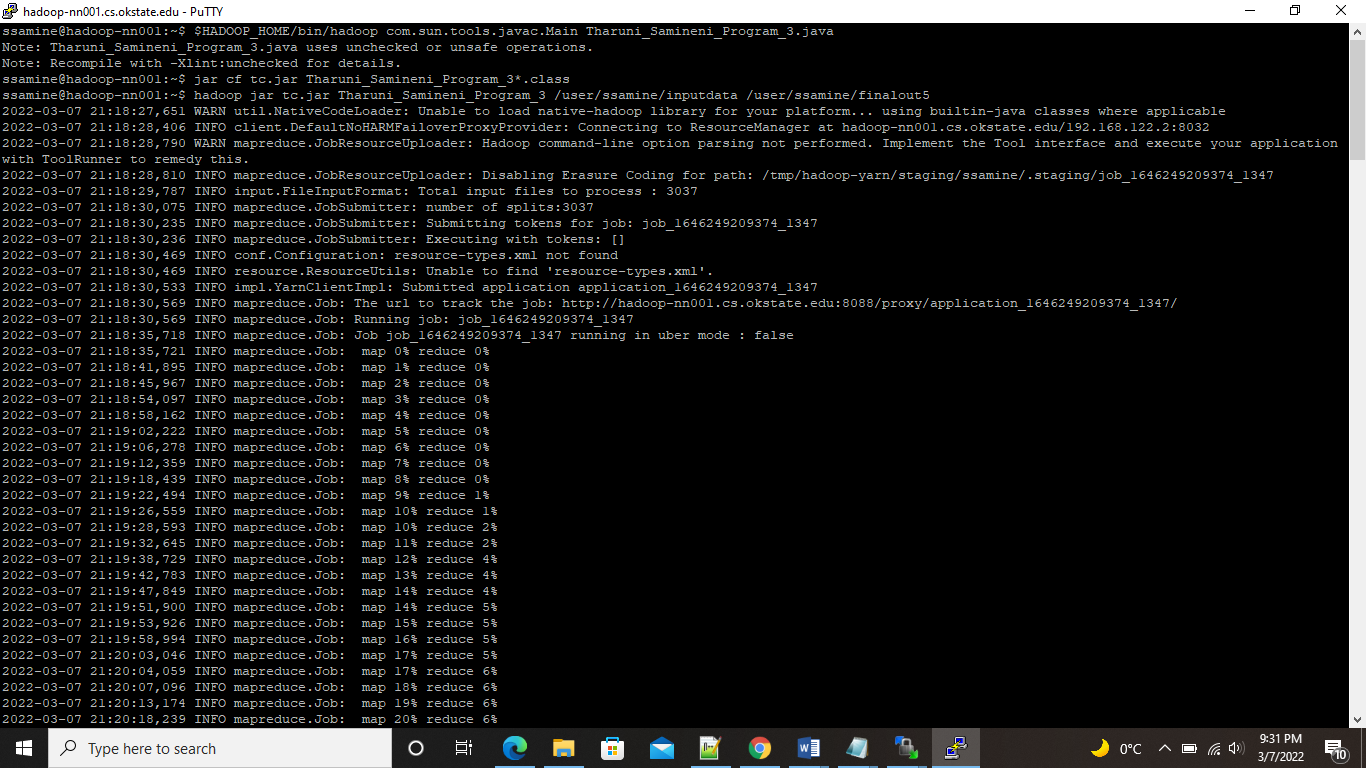


Figure 3: Shows the command to run the program after successful compilation

3,4) To check the output of the program use the command:

- hdfs dfs -cat /user/ssamine/test/part-r-00000

- hdfs dfs -cat /user/ssamine/test/part-r-00001

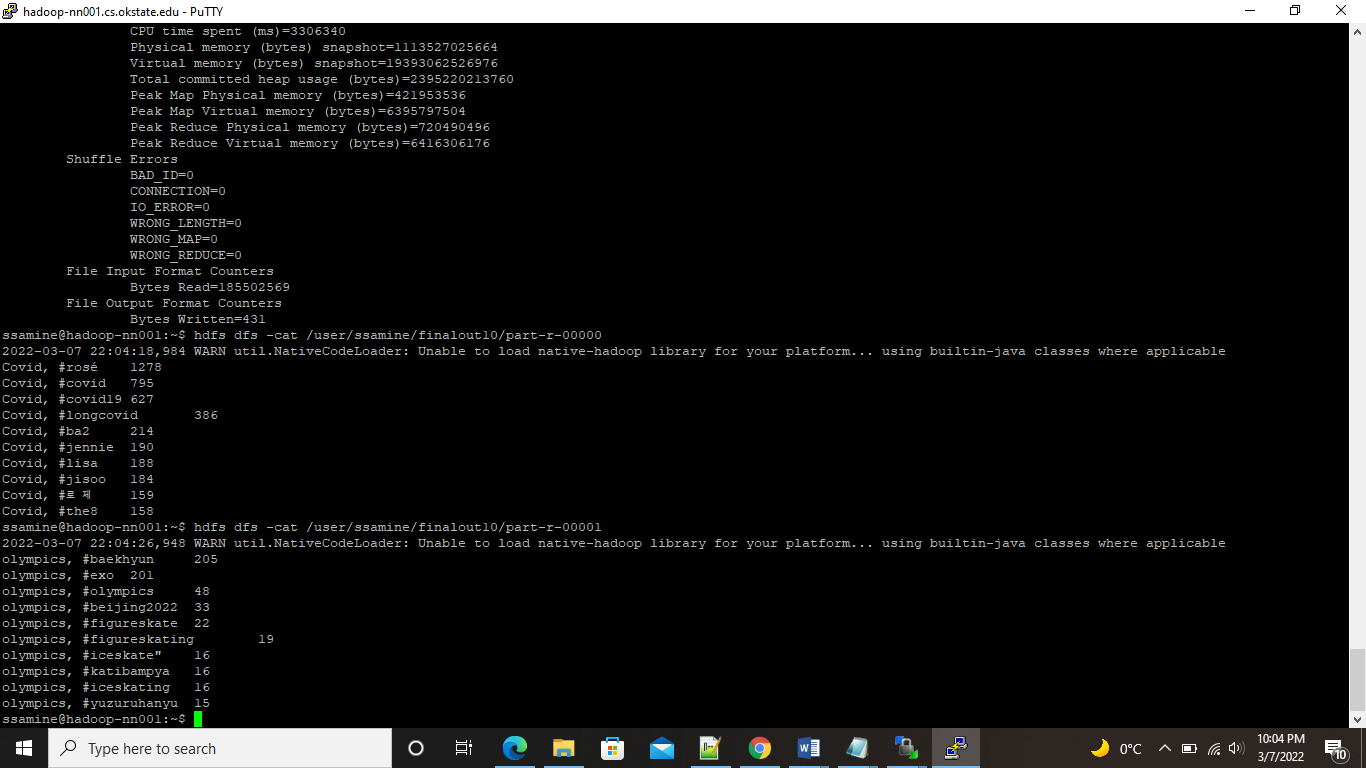


Figure 4: Shows the command to print the final output on the console for respective keyword

**Expected Final Output:**

**For Keyword ‘olympics’**

olympics, #baekhyun 205

olympics, #exo 201

olympics, #olympics 48

olympics, #beijing2022 33

olympics, #figureskate 22

olympics, #figureskating 19

olympics, #iceskate" 16

olympics, #katibampya 16

olympics, #iceskating 16

olympics, #yuzuruhanyu 15

**For Keyword ‘covid’**

Covid, #rosé 1278

Covid, #covid 795

Covid, #covid19 627

Covid, #longcovid 386

Covid, #ba2 214

Covid, #jennie 190

Covid, #lisa 188

Covid, #jisoo 184

Covid, #로제 159

Covid, #the8 158