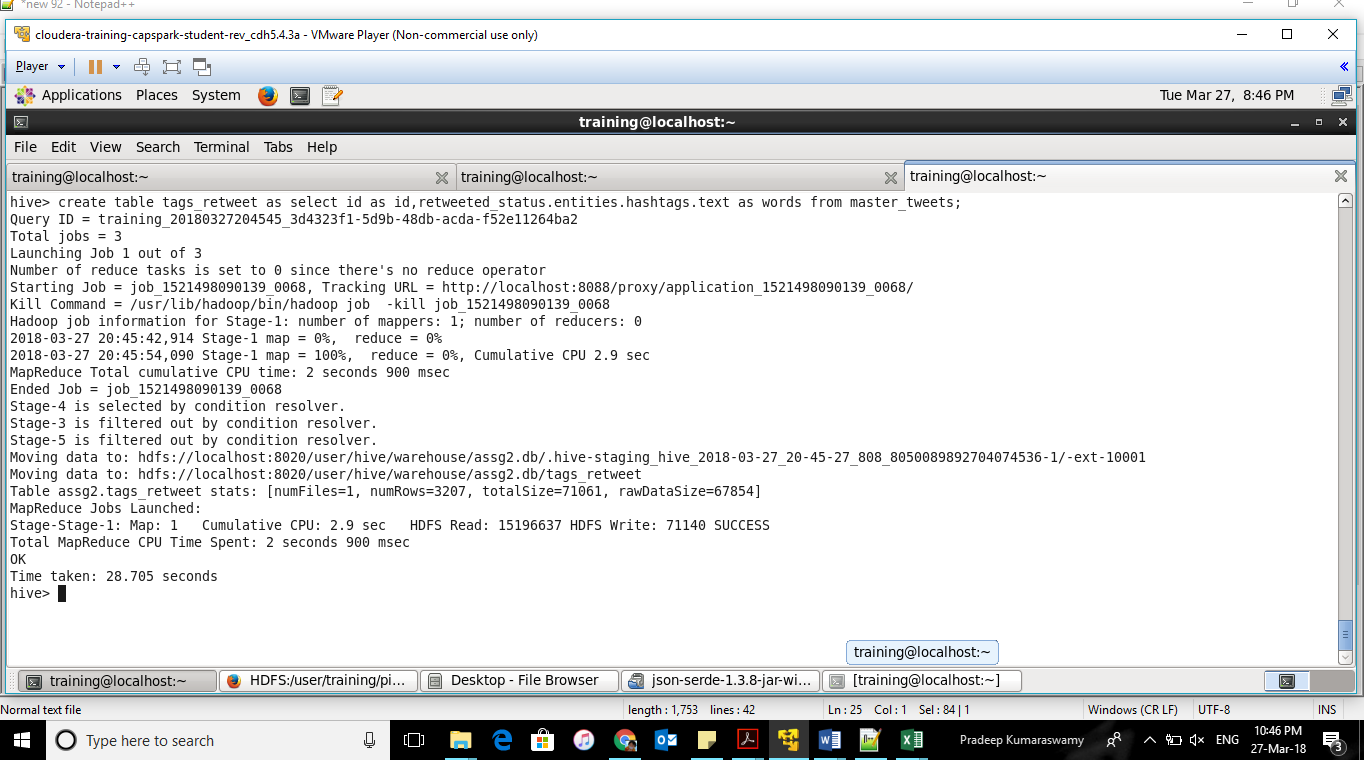
Pradeep Kumaraswamy

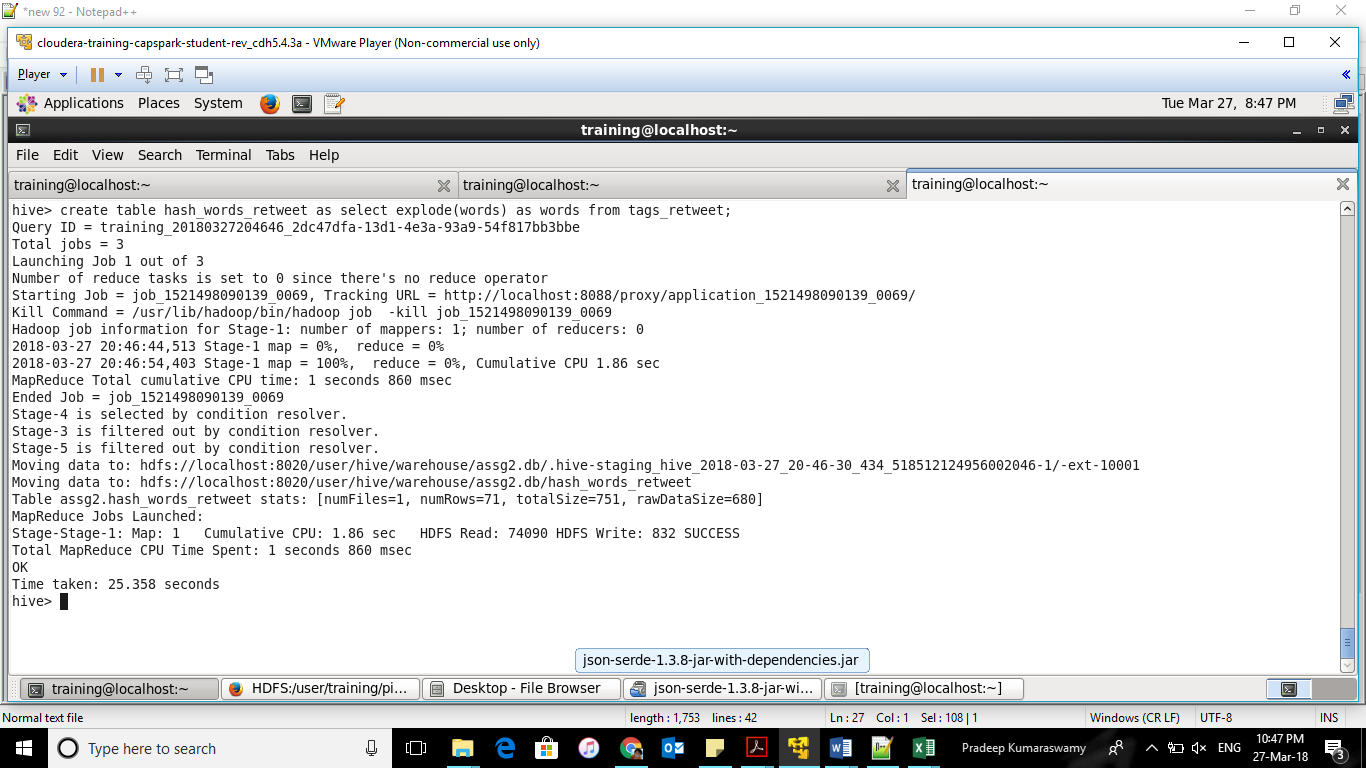
Assignment 2

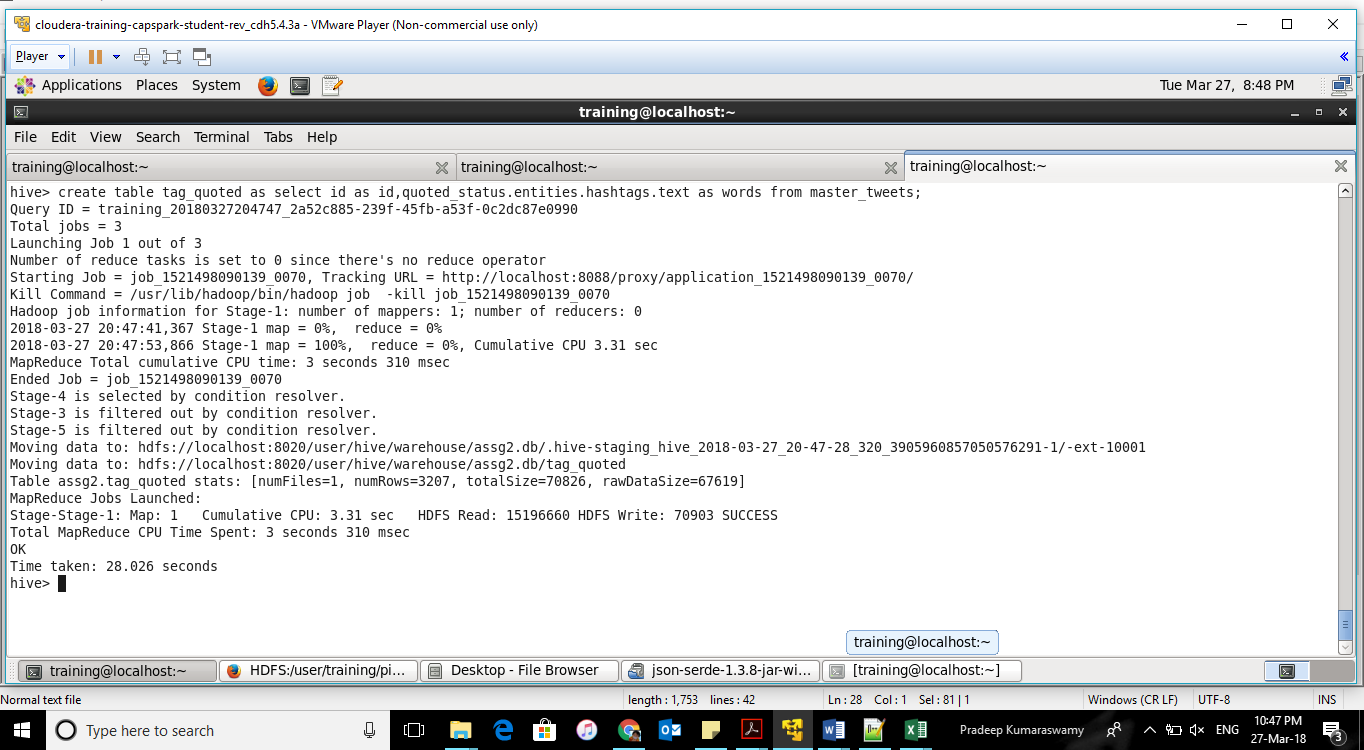
1 . a) 1. What are the hashtags used and how many times each are used?

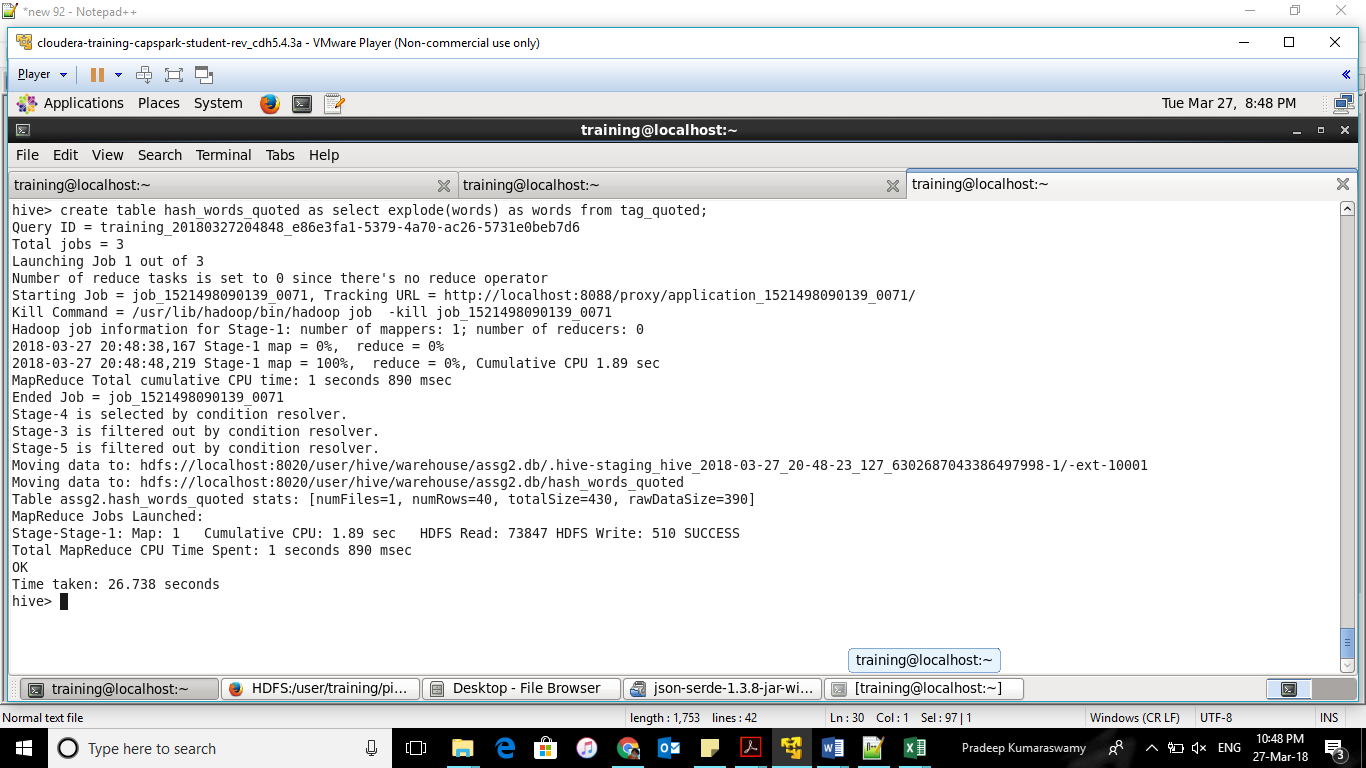


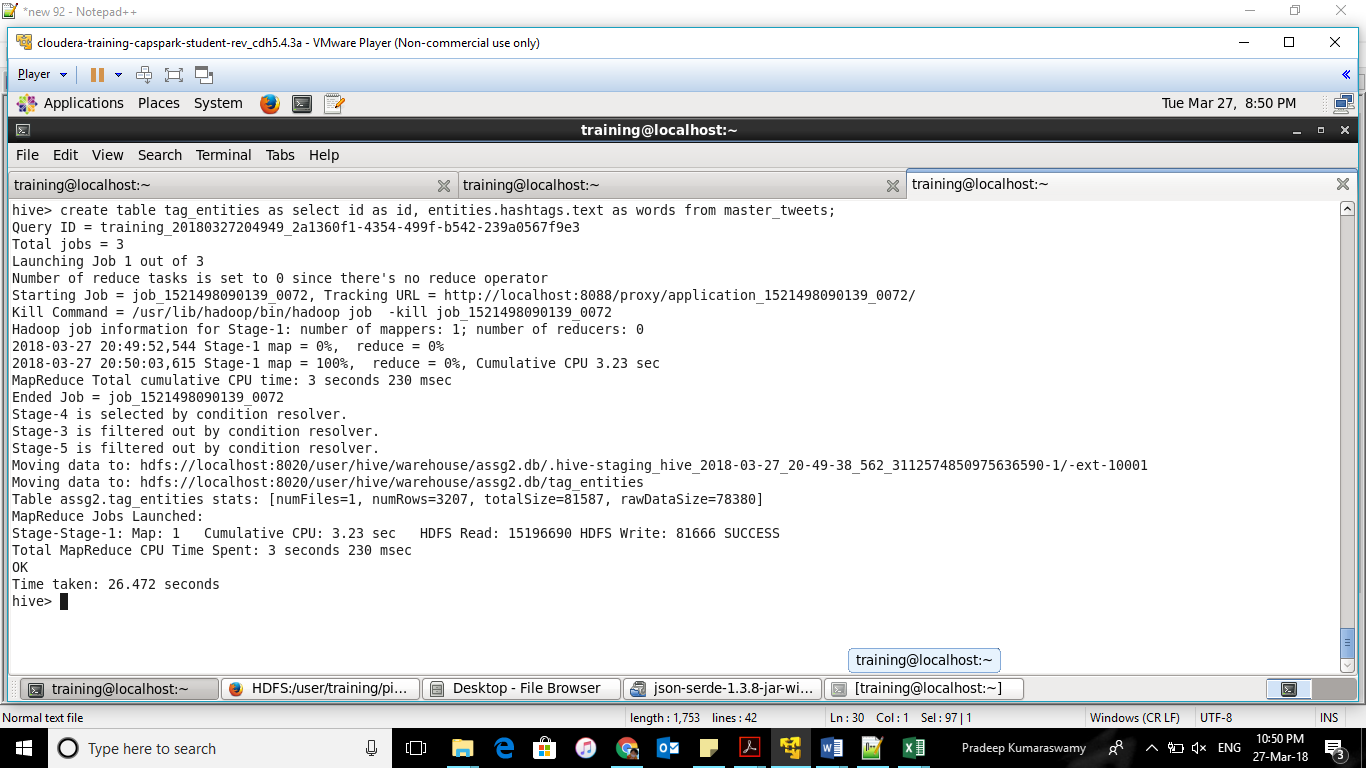
Table creation

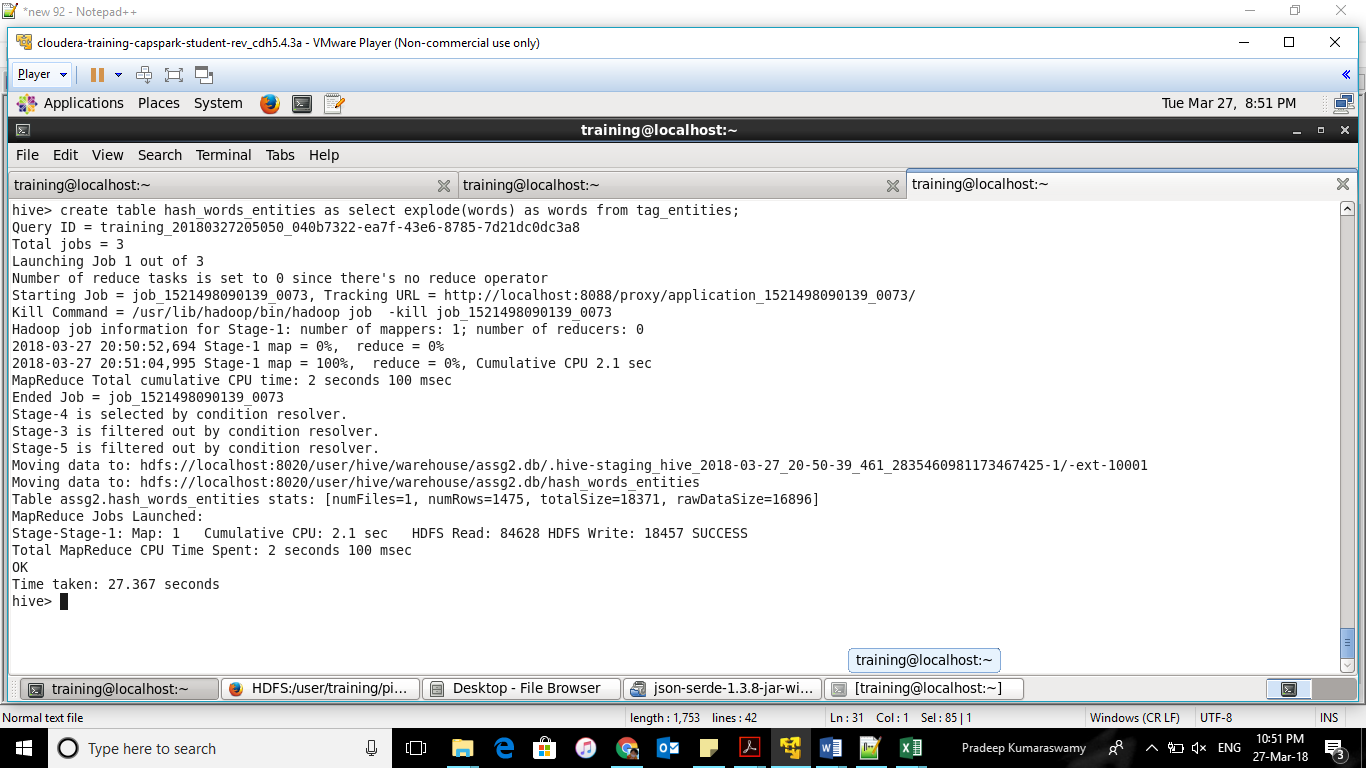


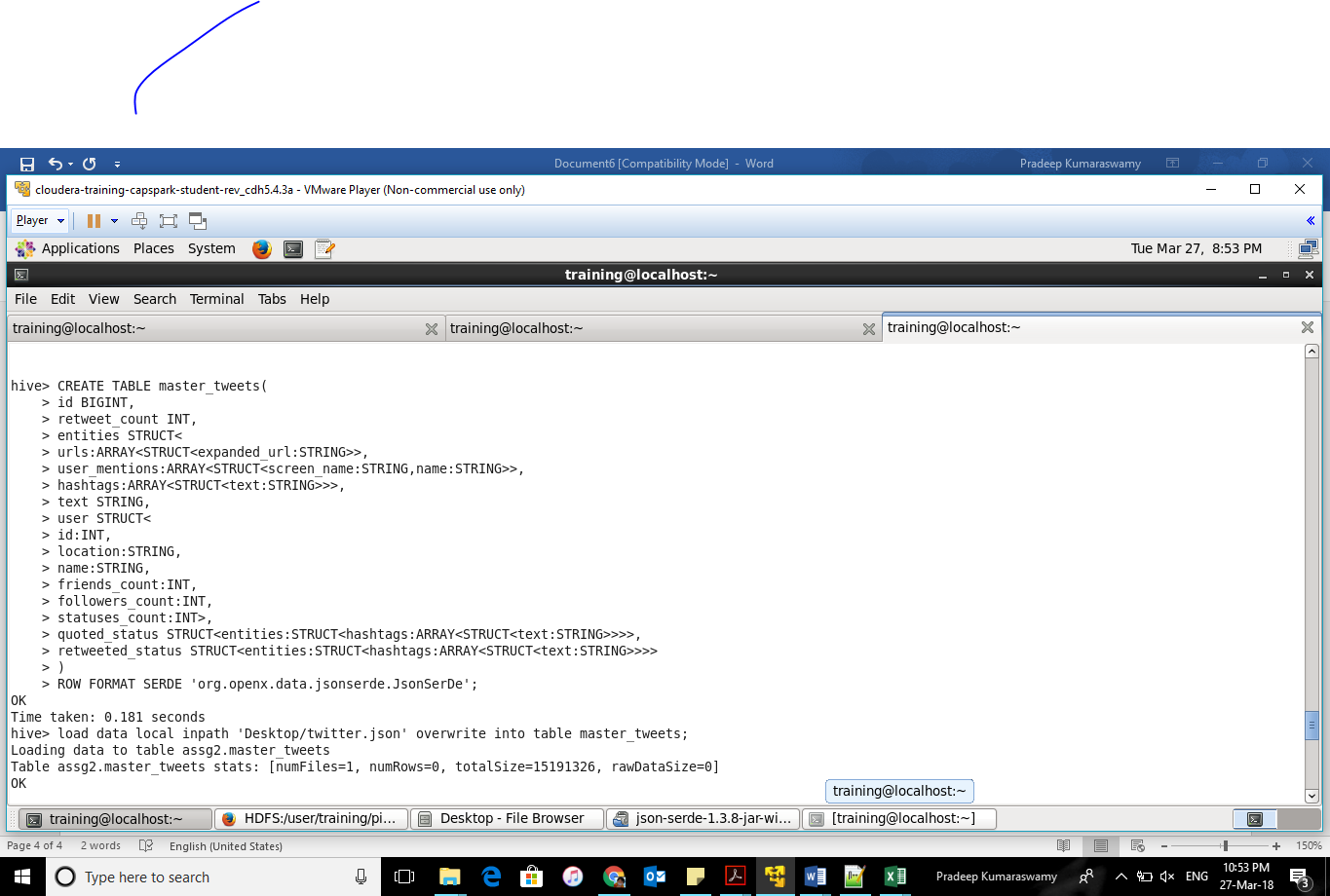


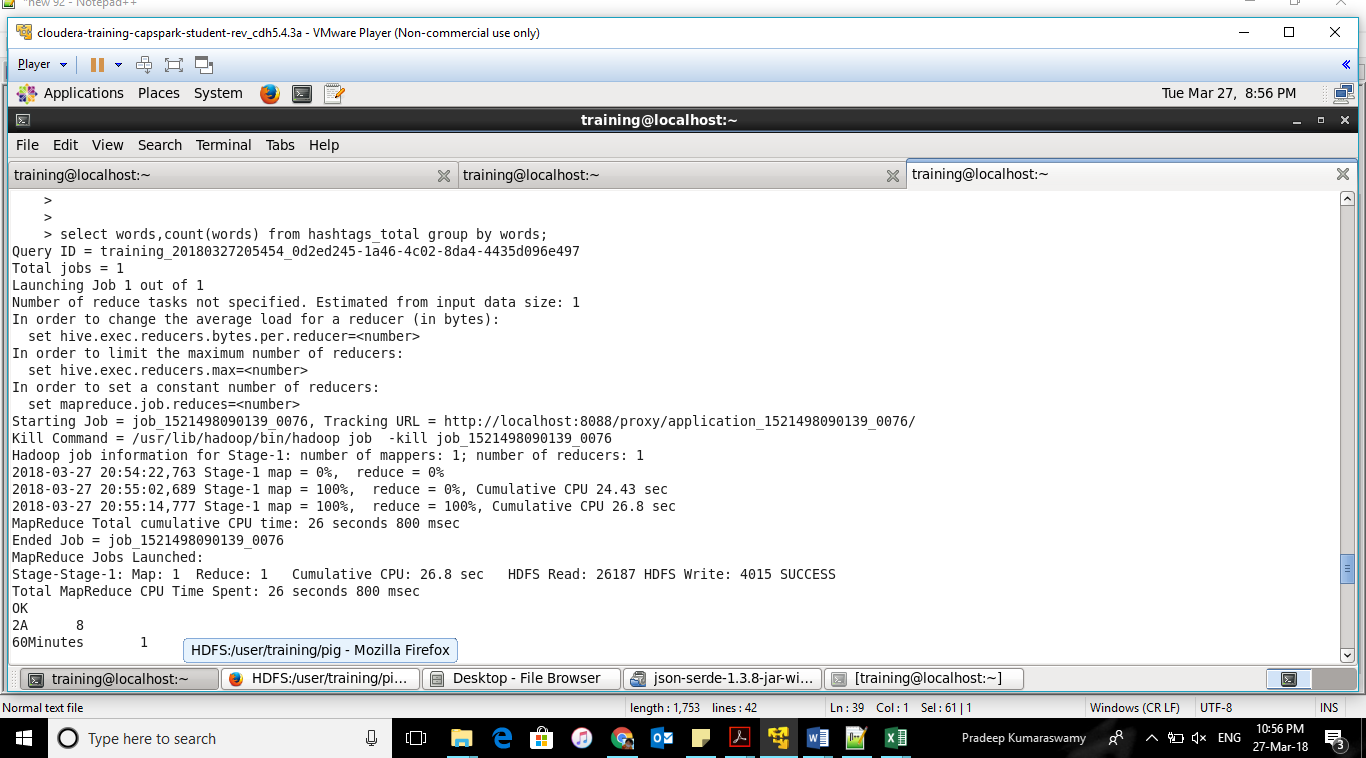




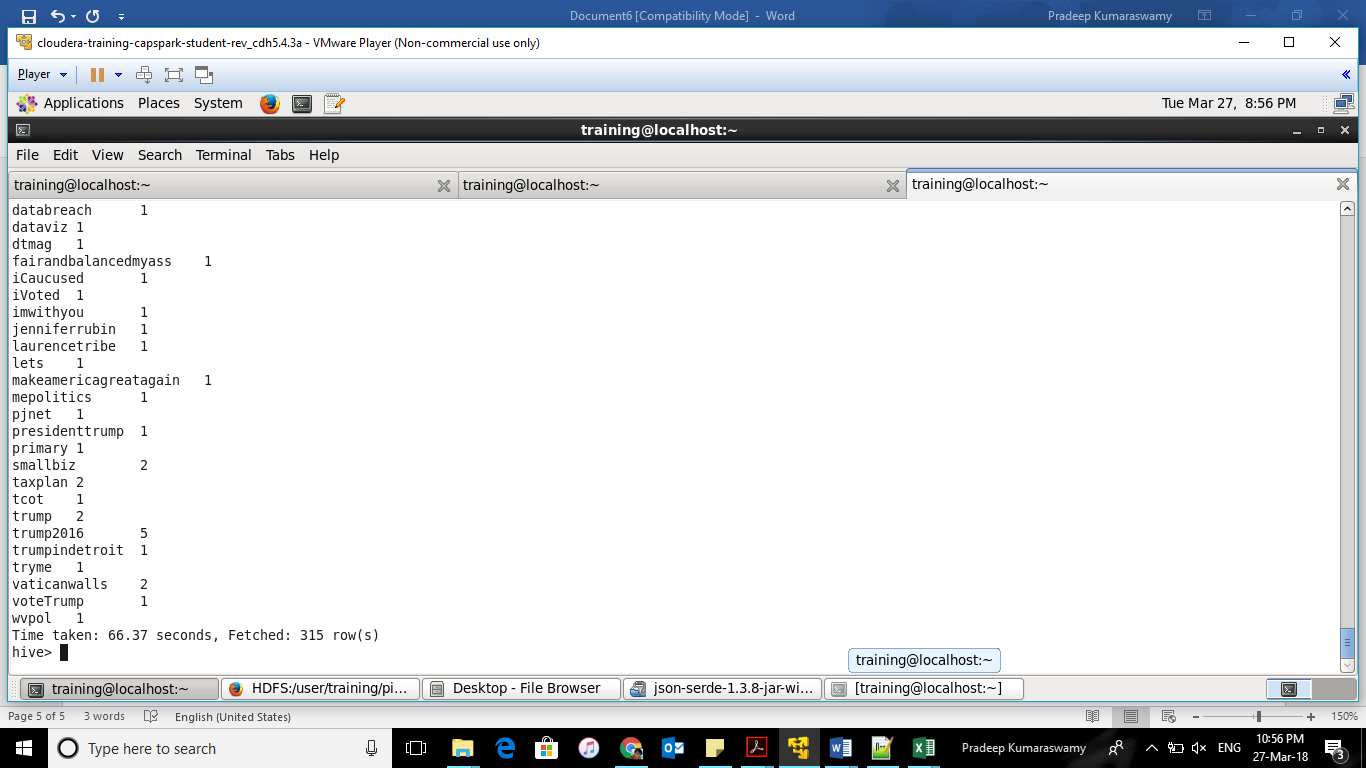








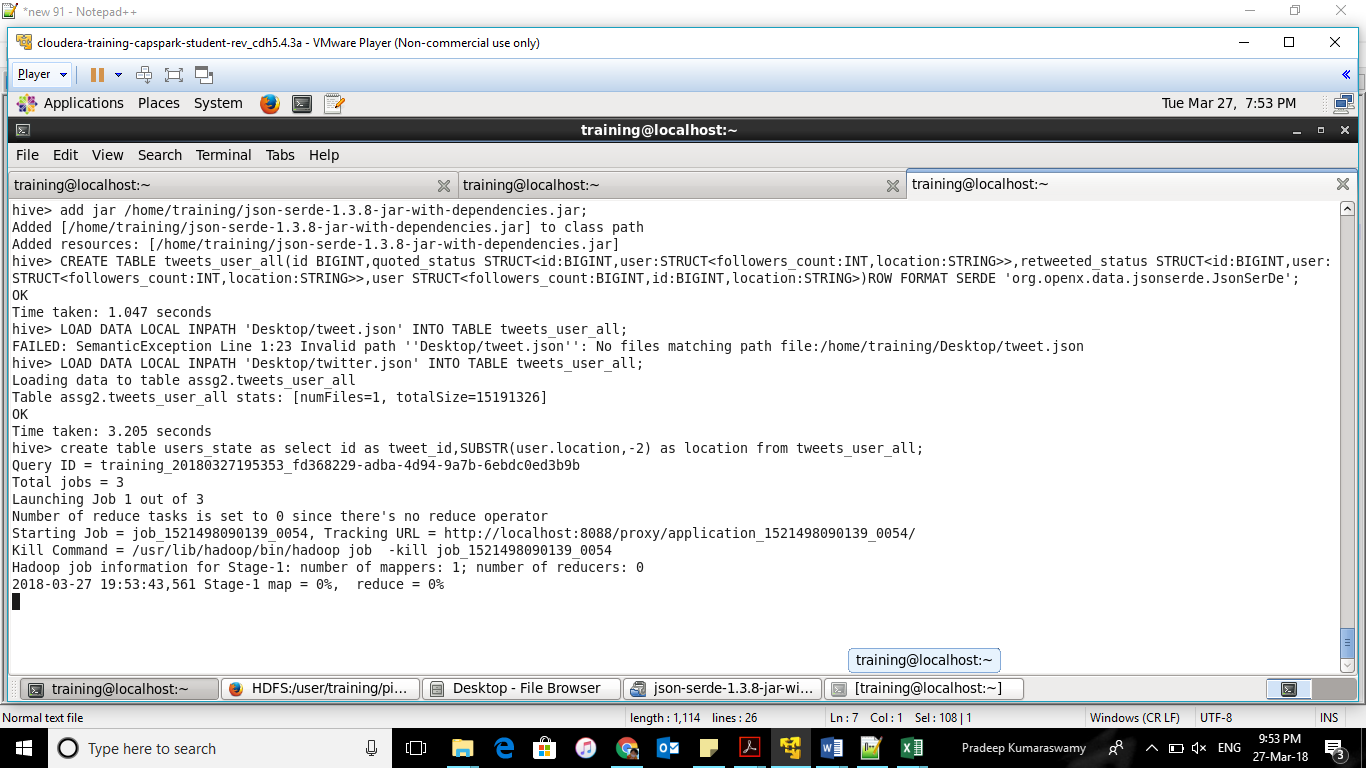
Totally 315 hashtags

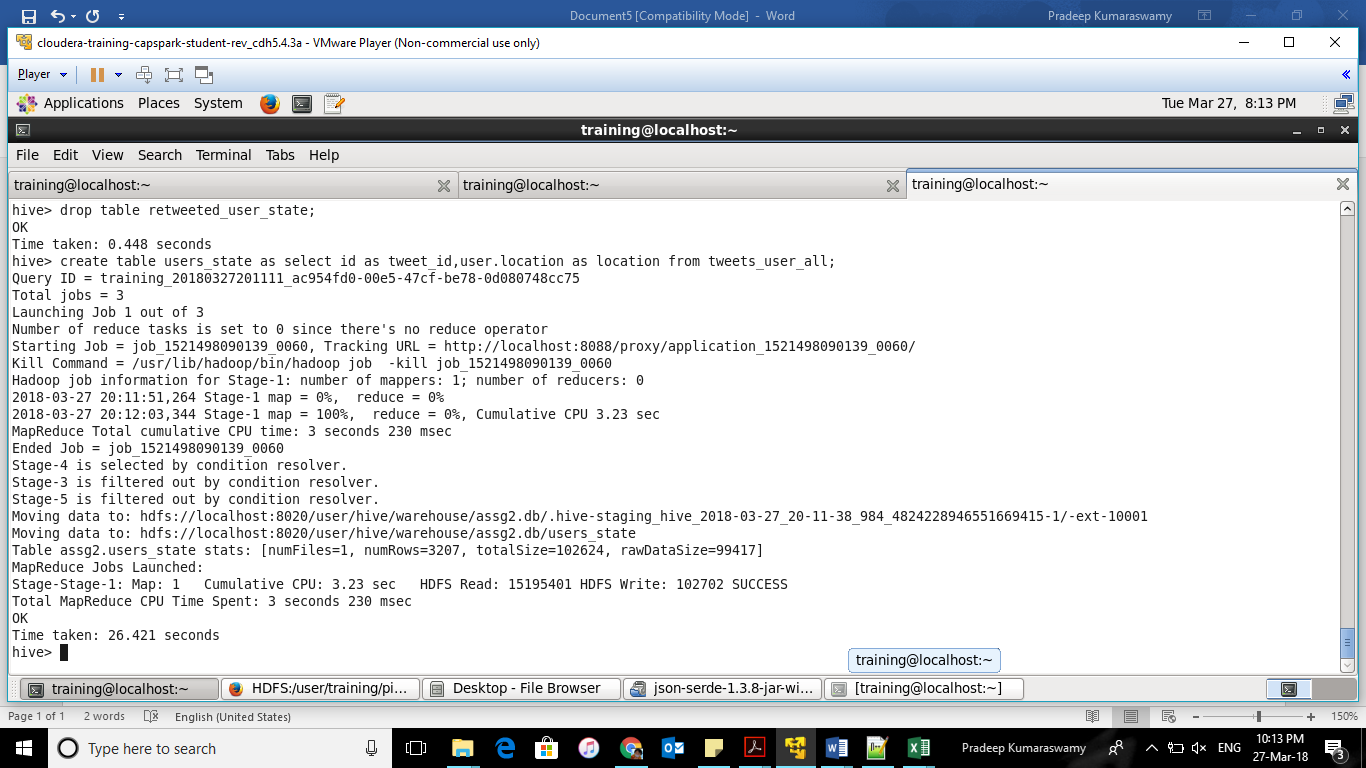


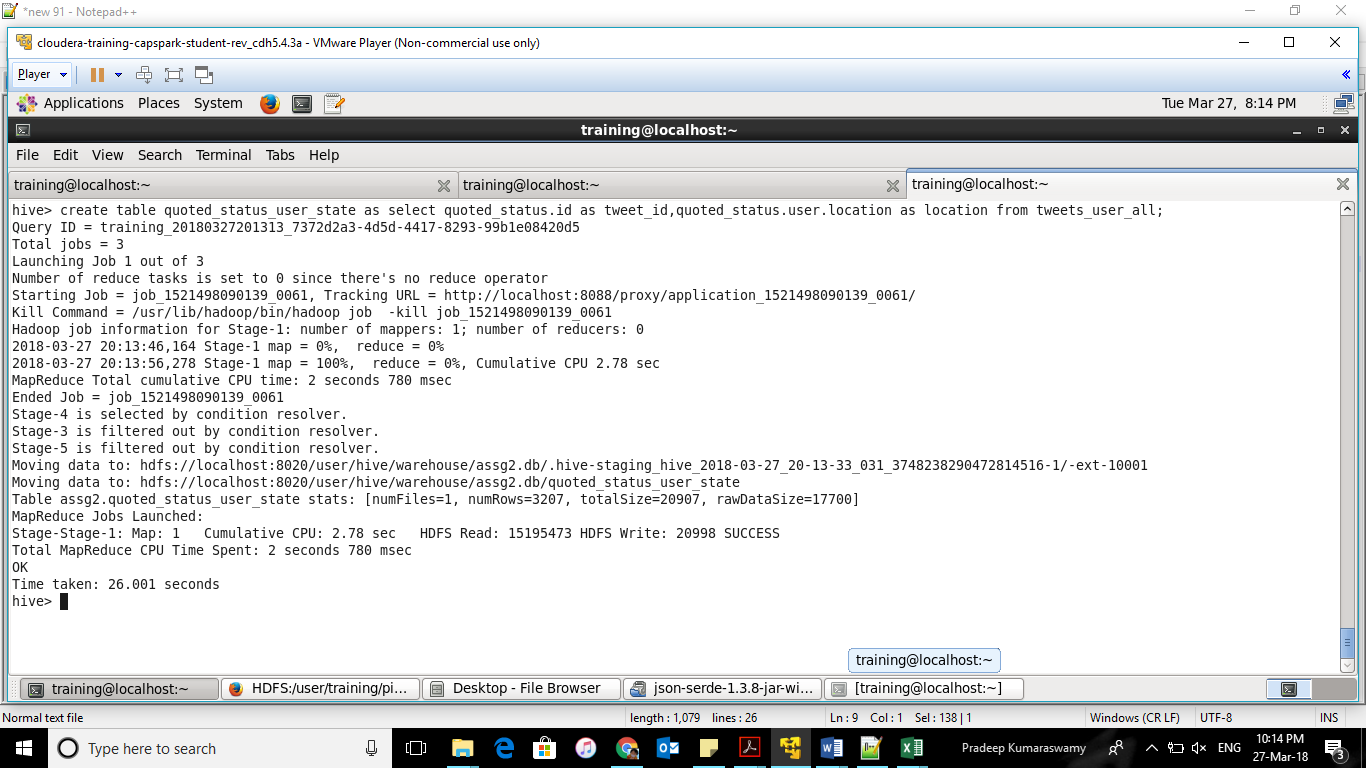
continued

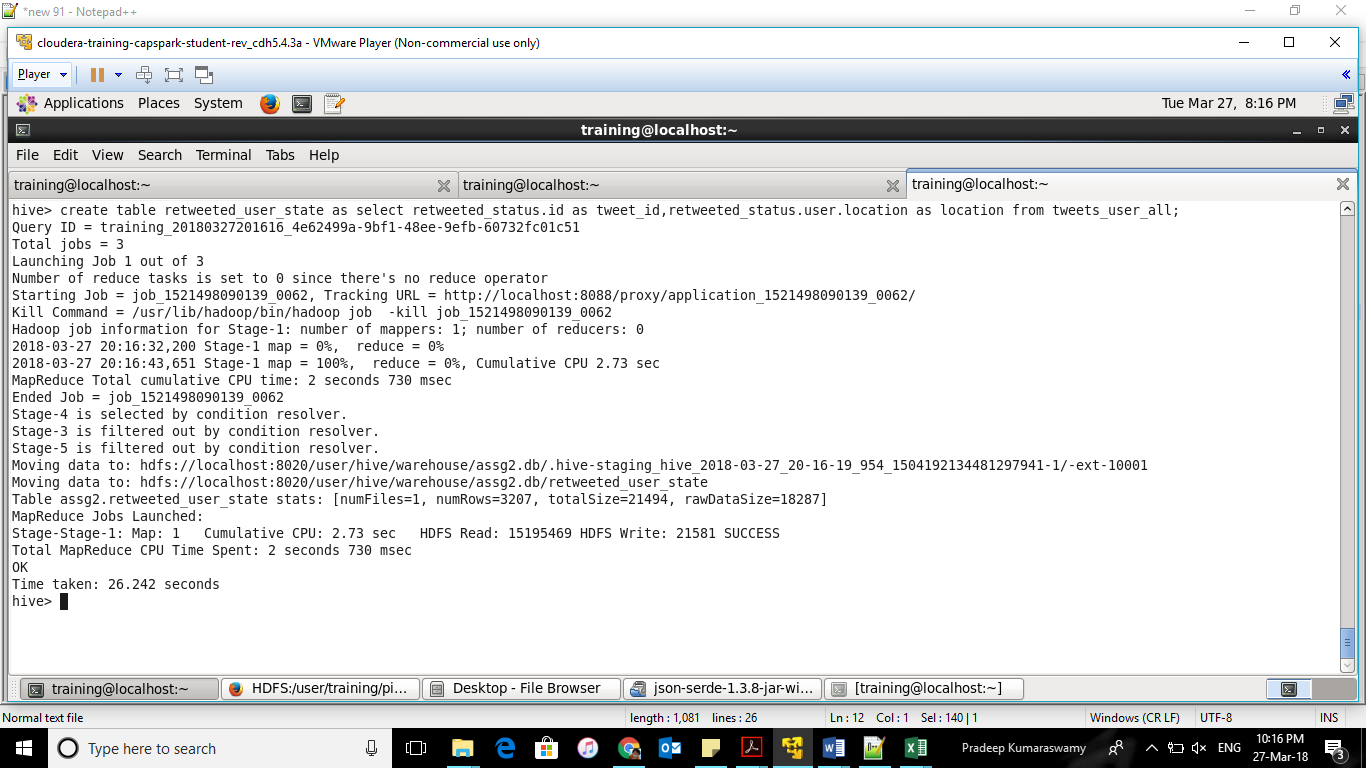
b) Which State have the most active users and how many tweets are posted by State?

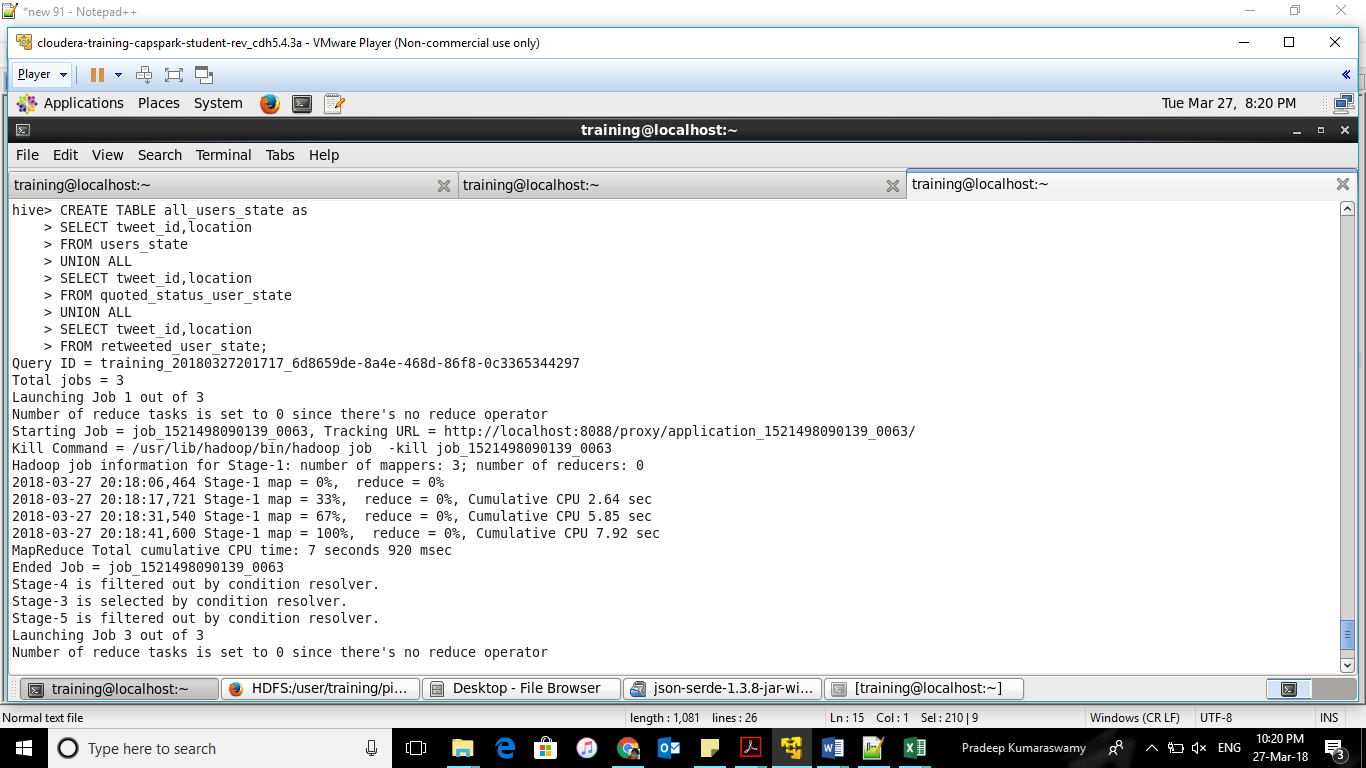
Newyork and 3221

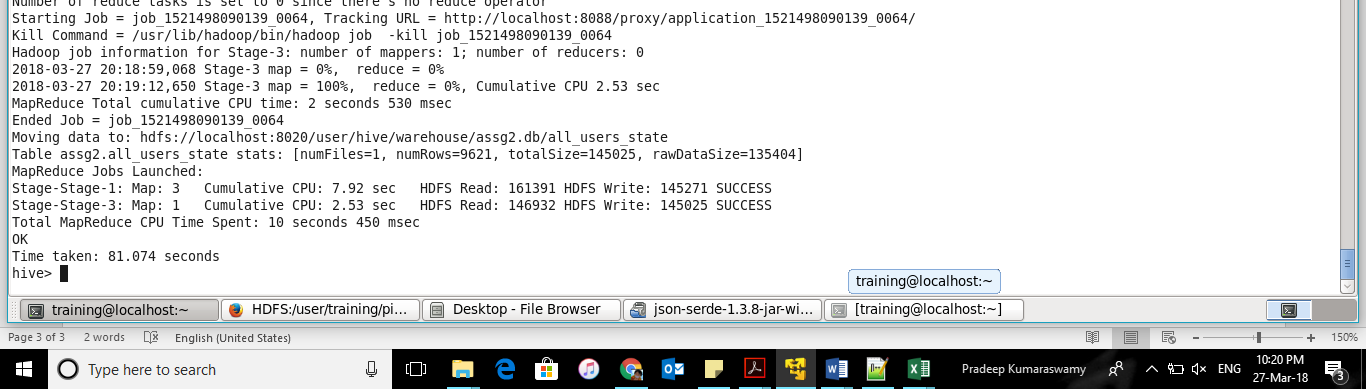


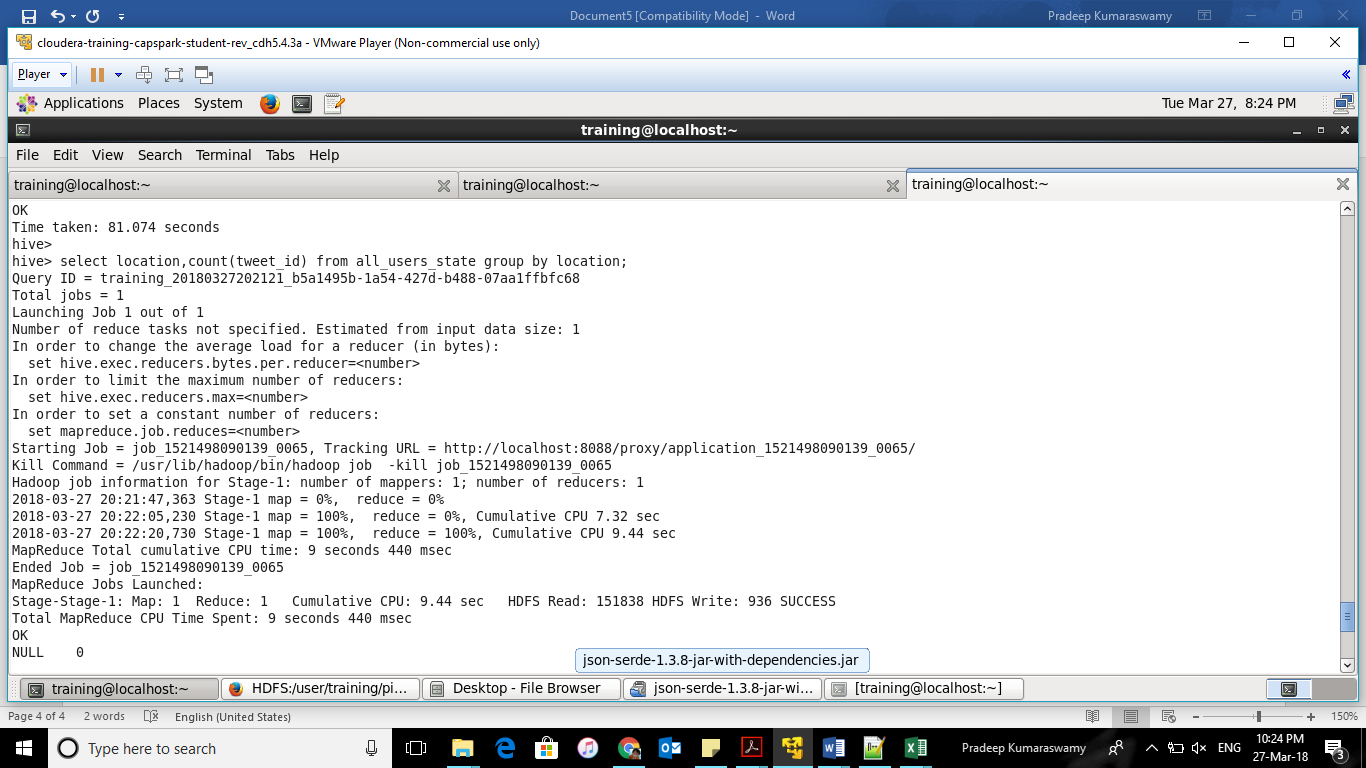


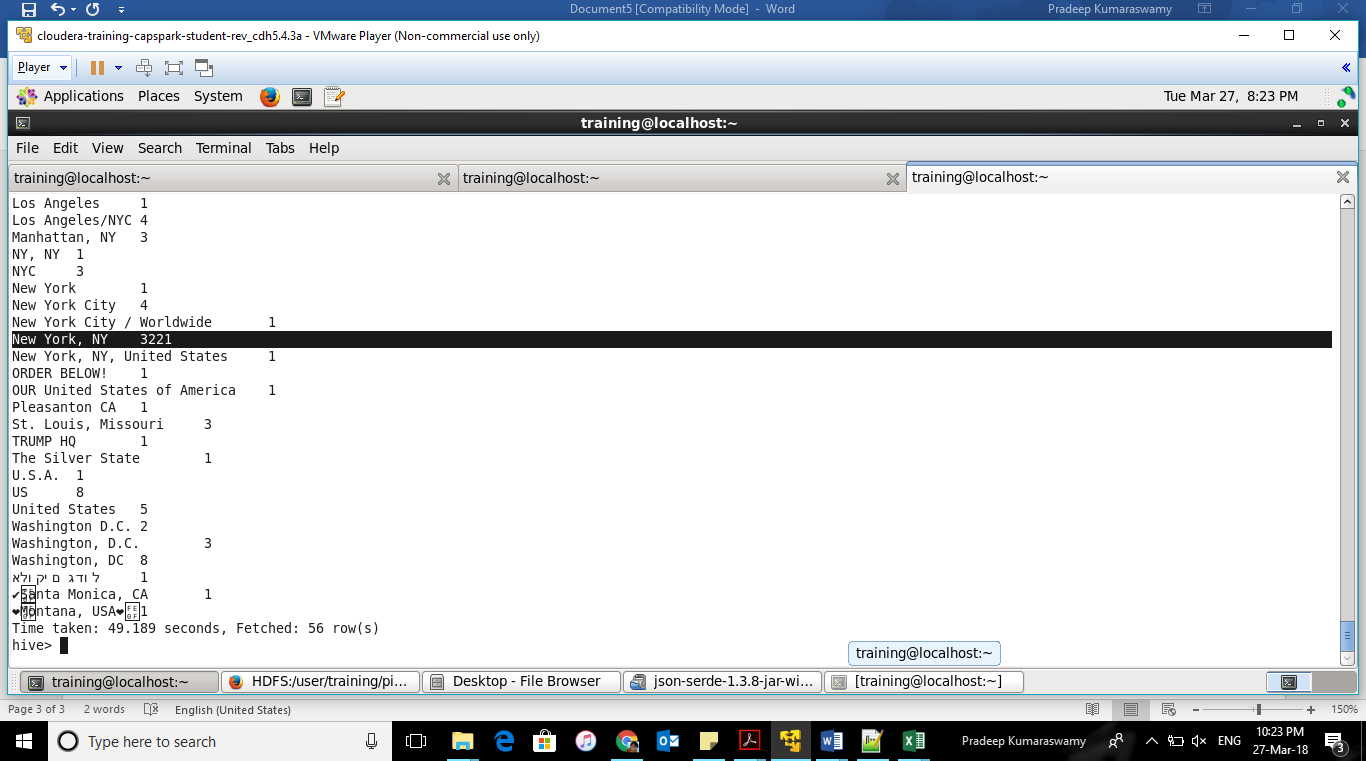




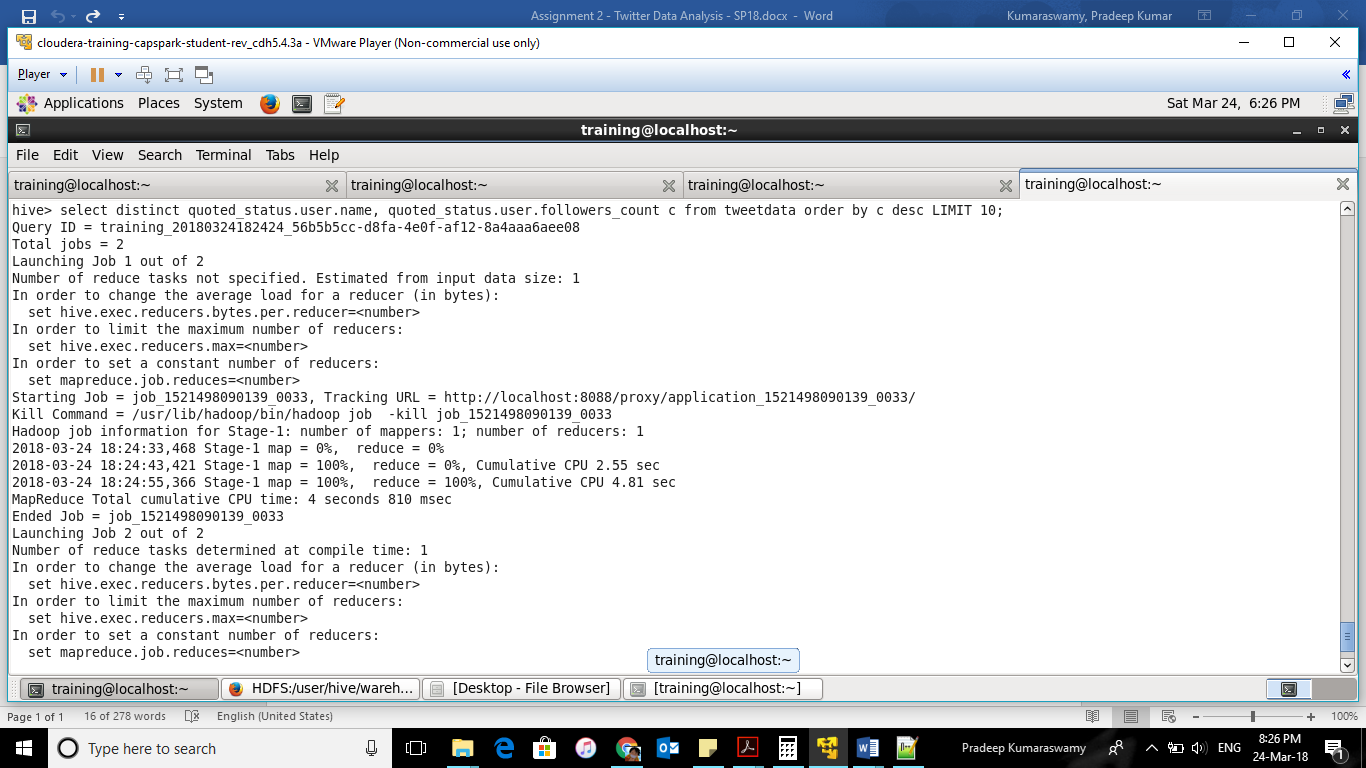




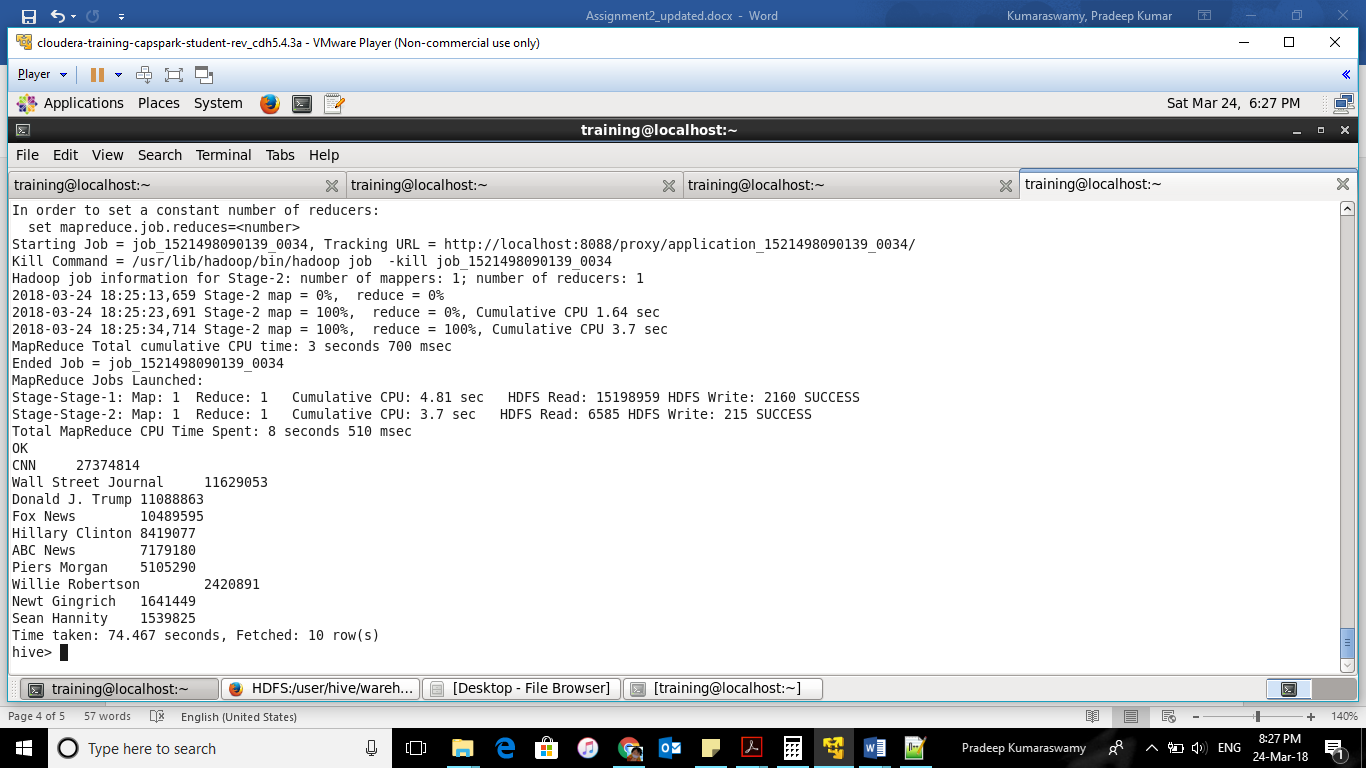




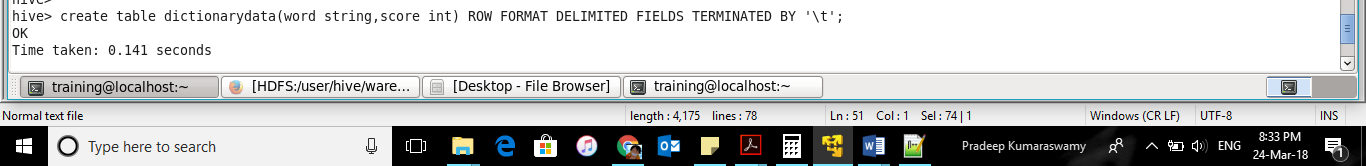
c) Based on the user’s followers count, who are the top ten users who have tweeted?

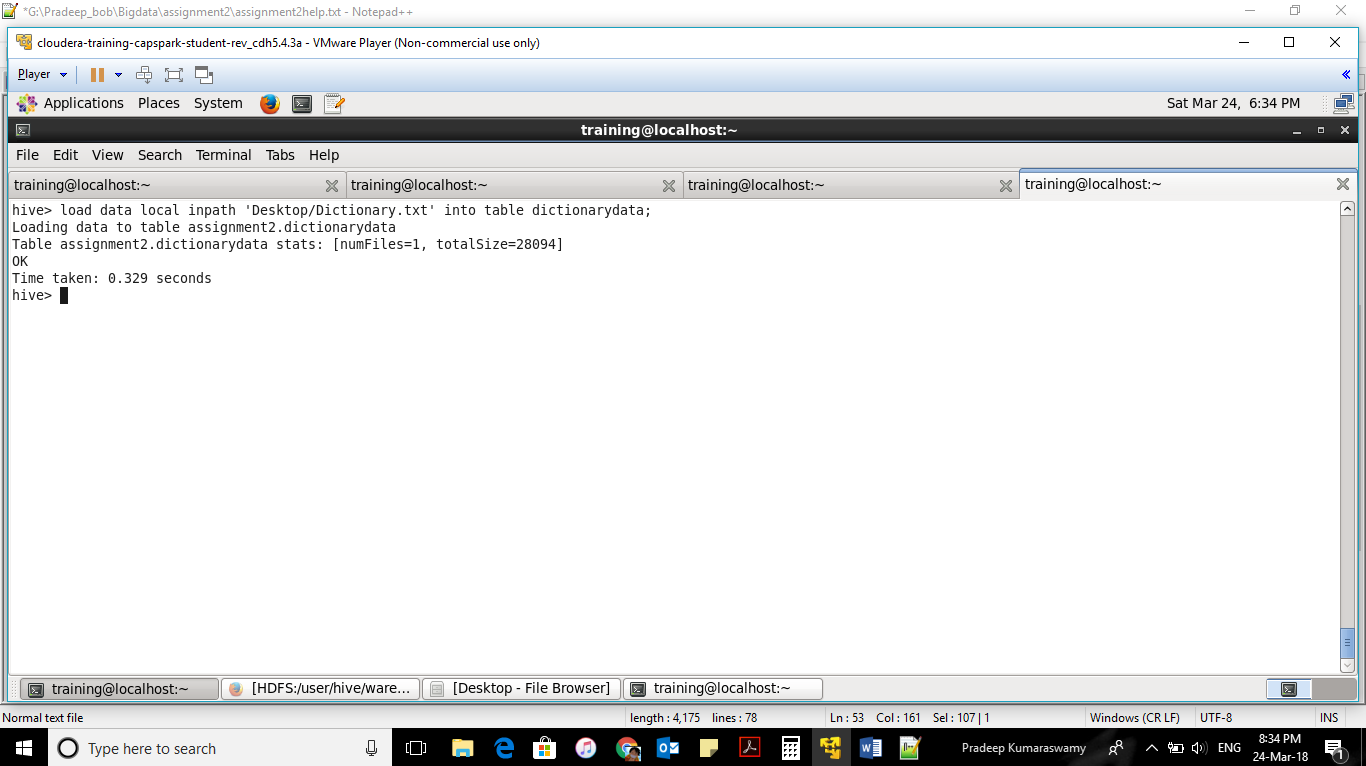


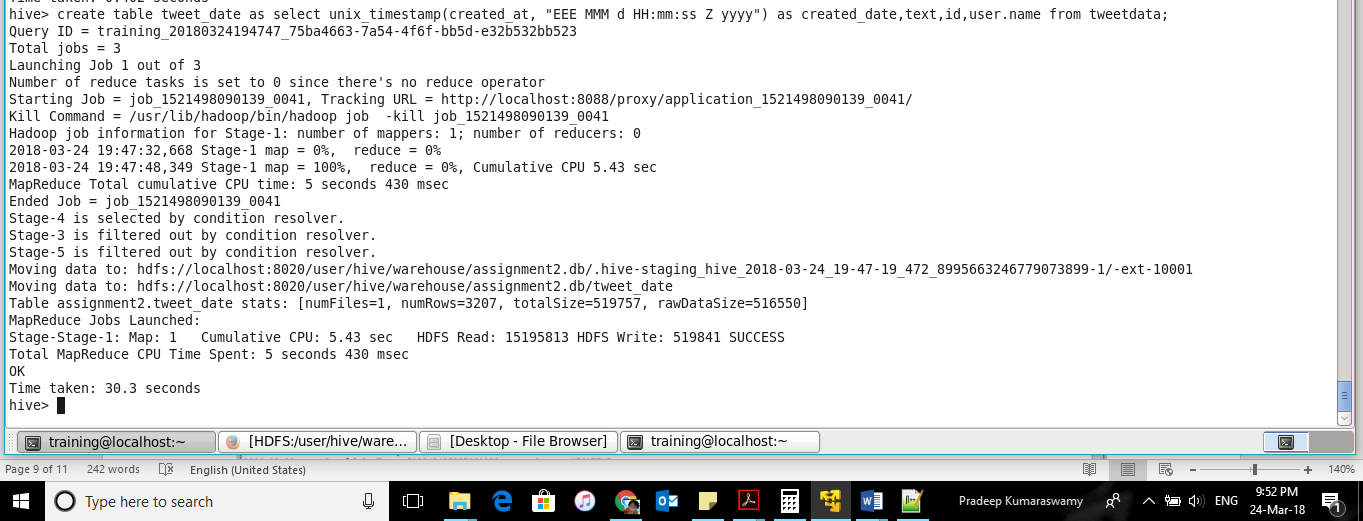
Top ten users

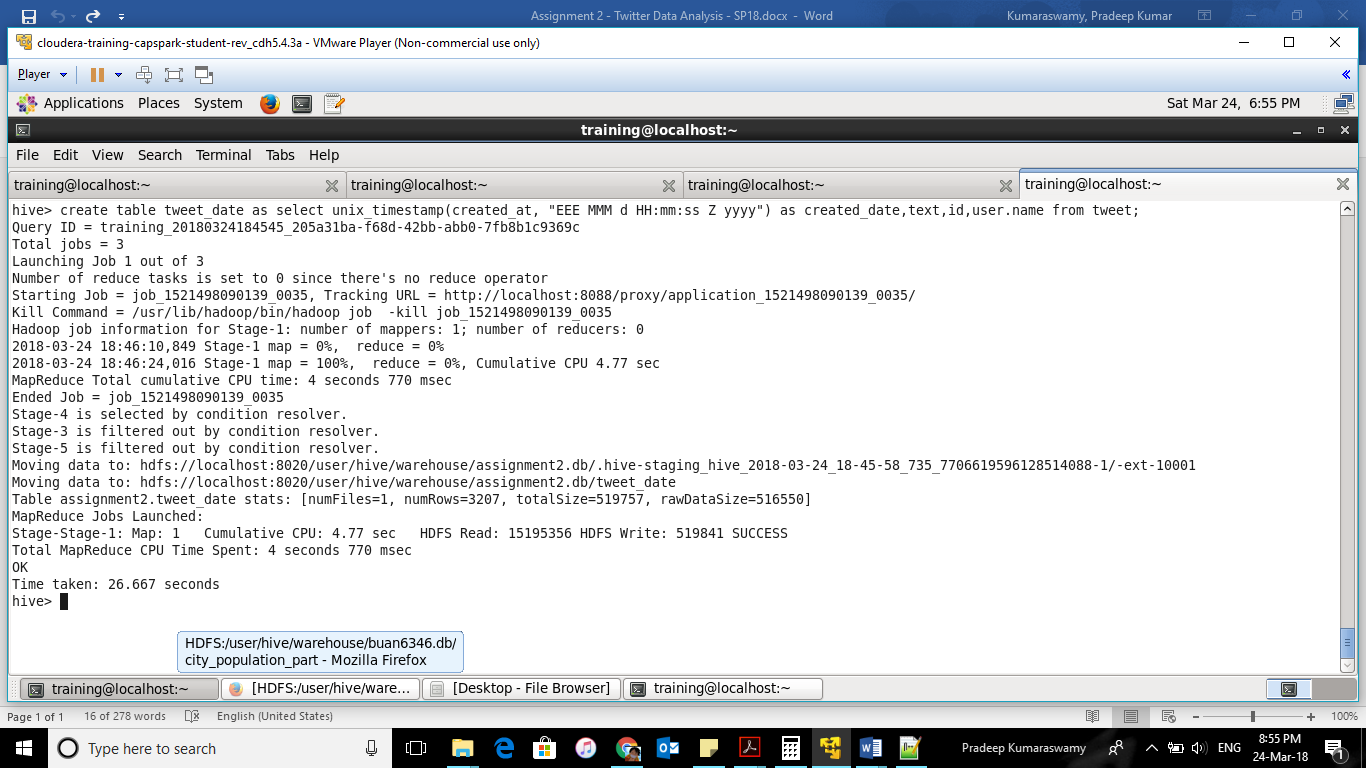


d) What is the polarity score for each tweet that was posted? Does the tweet have a positive or negative sentiment?

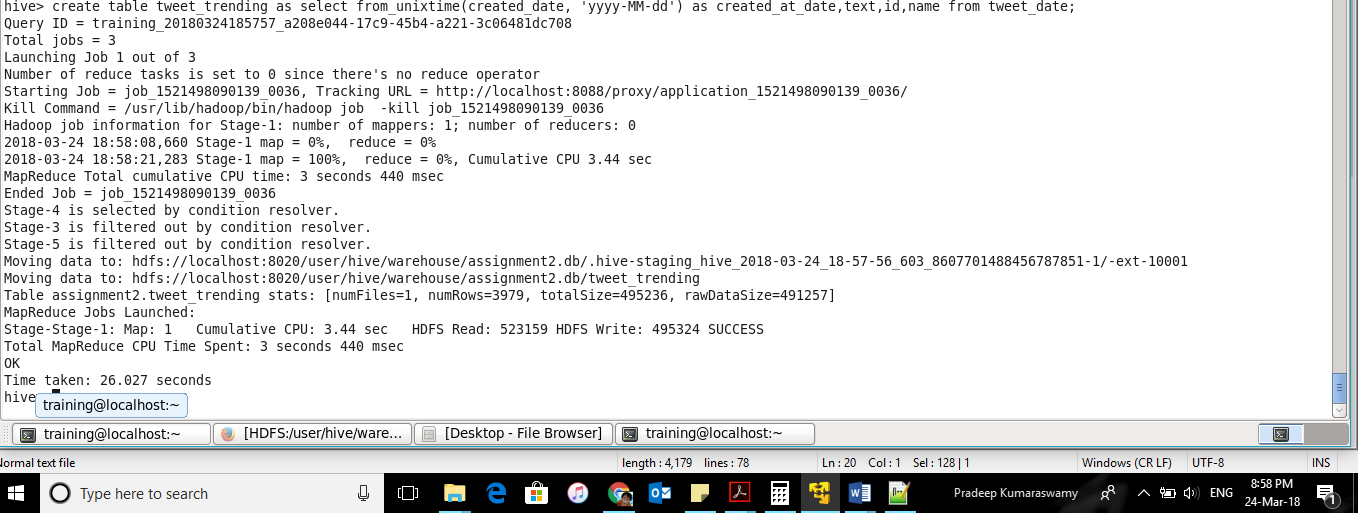


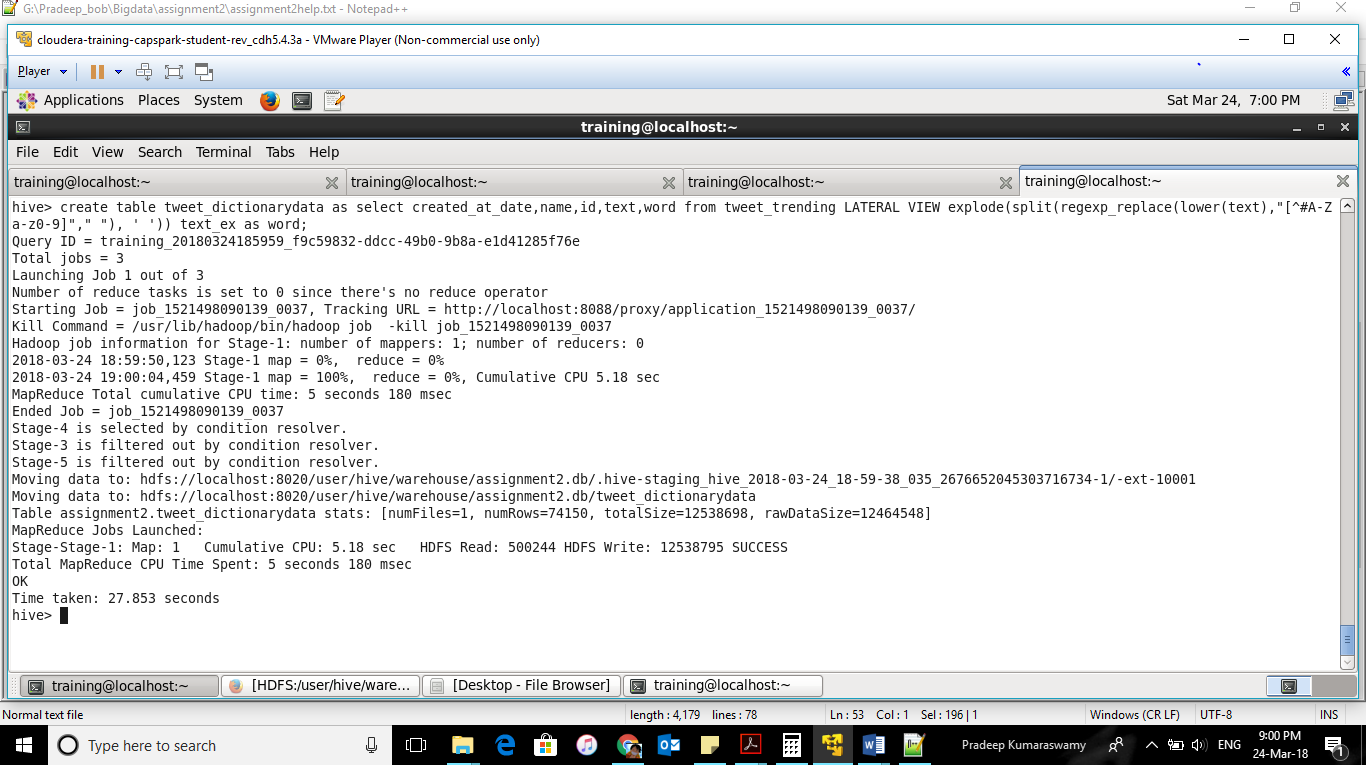


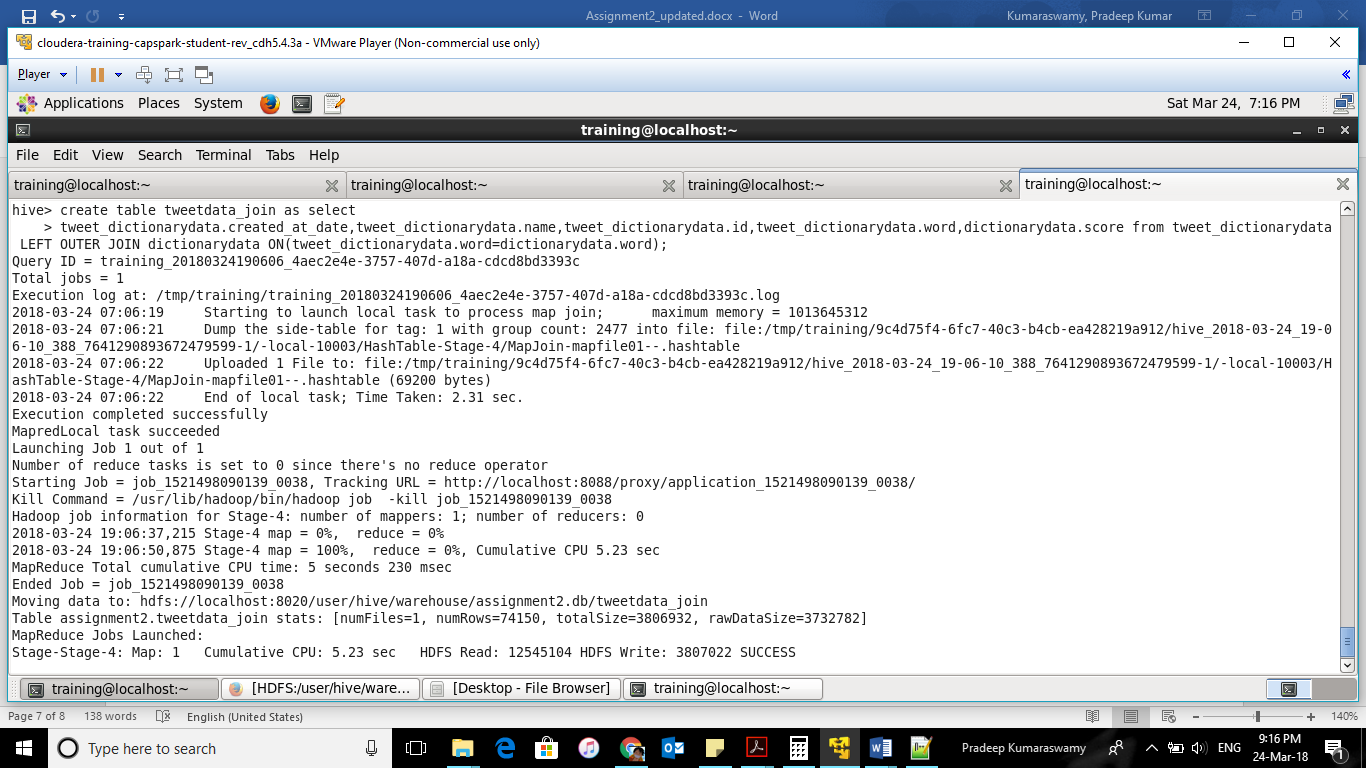
Converting Tweet time stamp to unix format as shown below

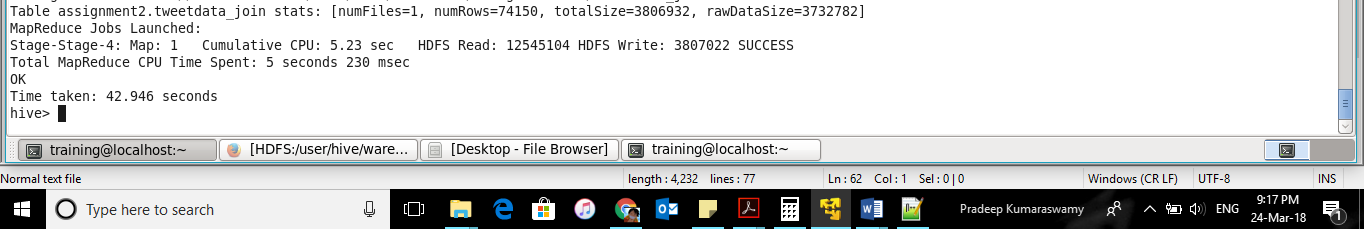


Extracting the Unix time stamp as desired in the query

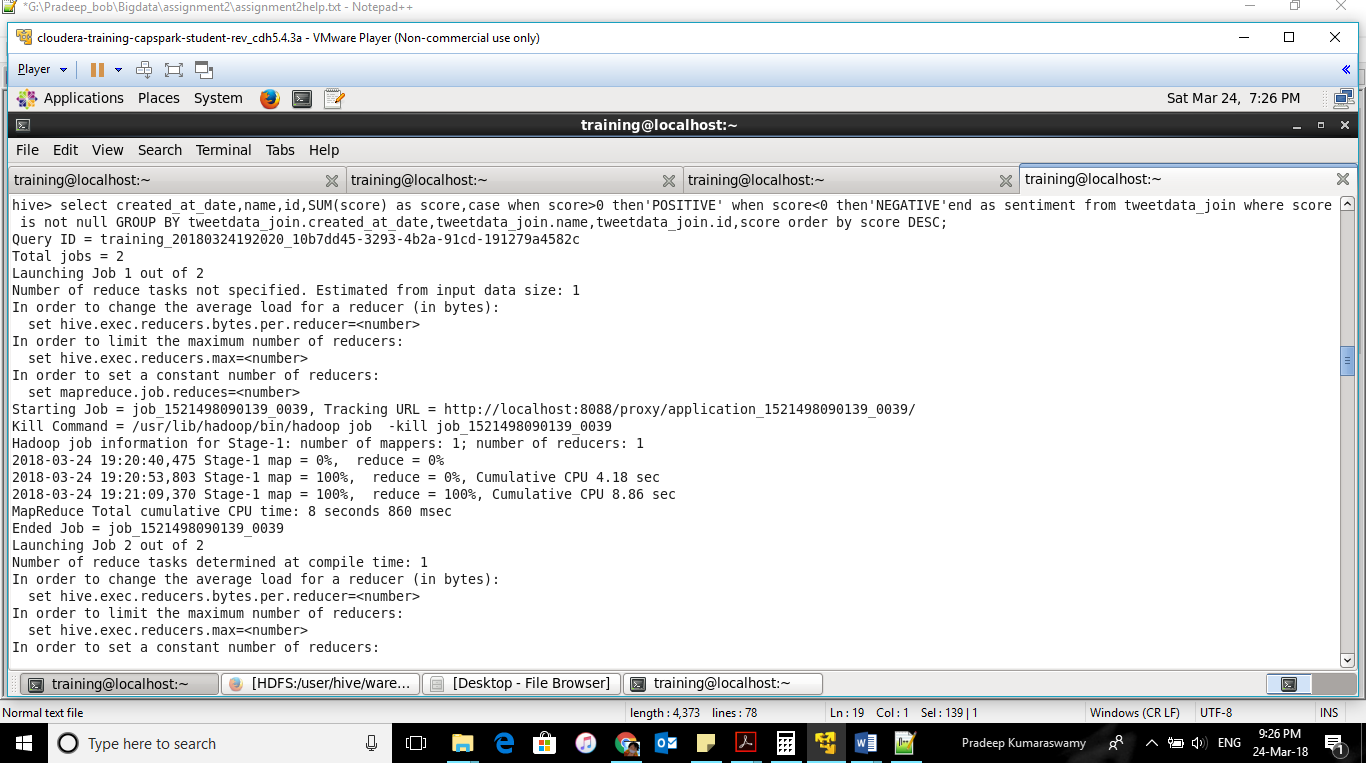


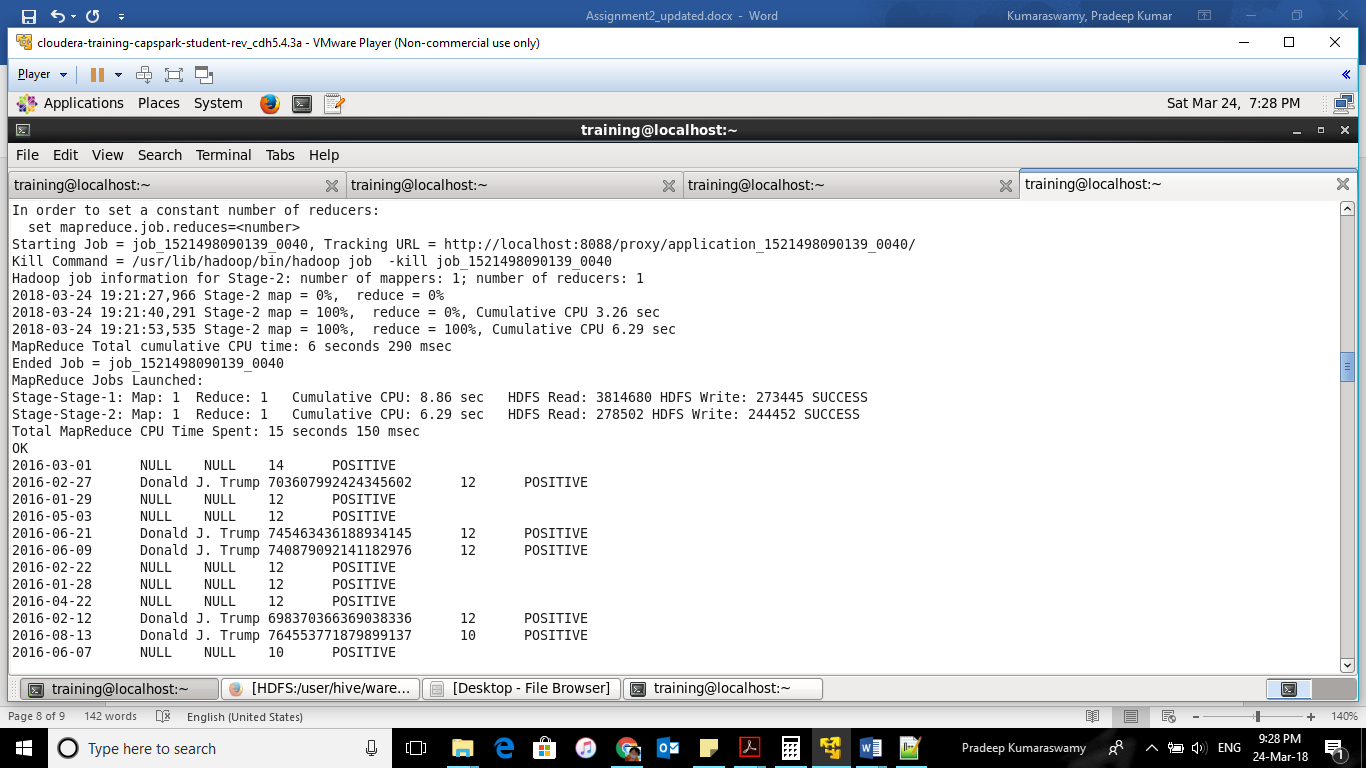
Creating the dictionary data table for comparison 

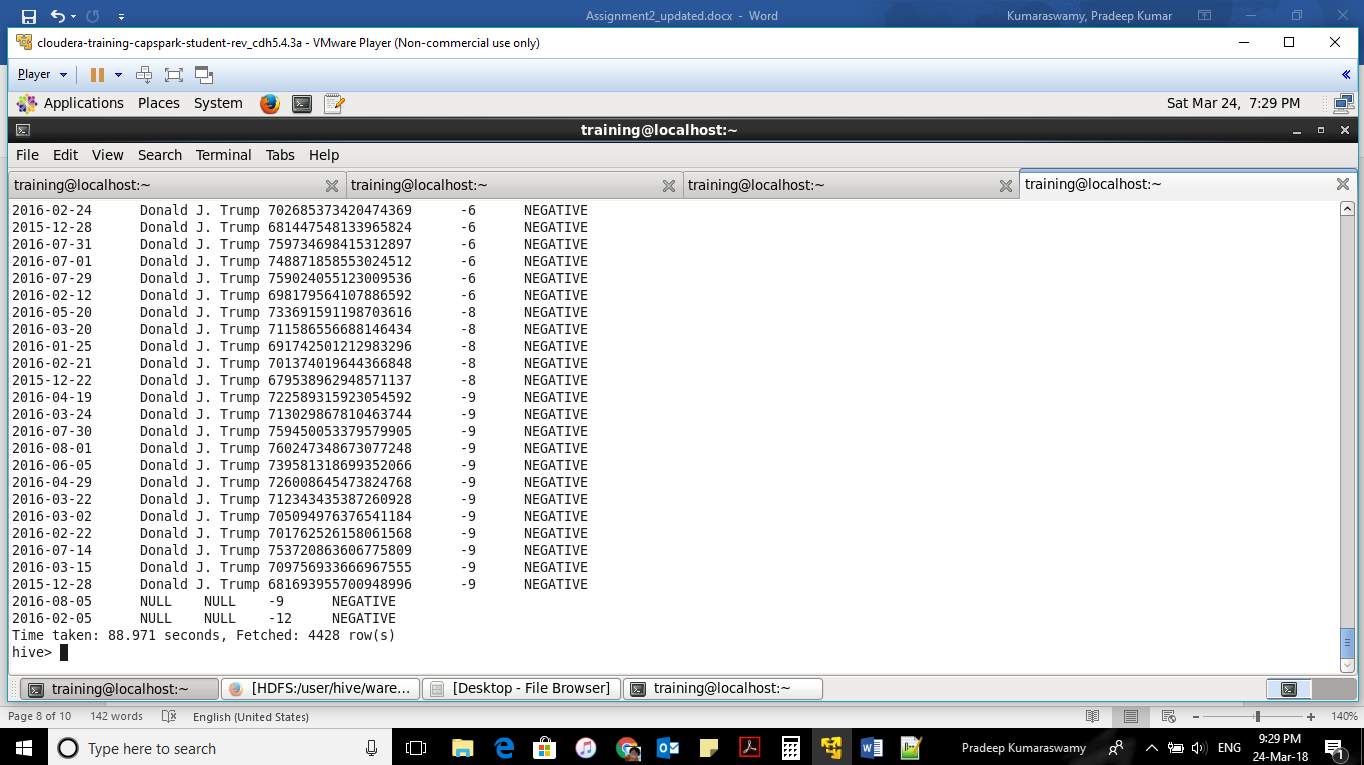
Creating join table to compare the tweet data with dictionary table created above



Sentiment scoring output query







Question 2)

Do you find any problem in the way sentiment analysis was performed in the previous question? If so, how will you improve it?

Problems observed:

* The scoring is limited to the content existing in Dictionary file. For example, tweets which does not have an entry in dictionary will be wrongly interpreted or taken as NULL value.
* Sarcasm tweets will be wrongly interpreted as positive sentiments in many cases.
* Misinterpretation of positive sentiments as Negative because of redundancy in negation words.

for example : ‘the exam was **not bad**’ will be interpreted as Negative here

* Misinterpretation of emotional tweets or ironic tweets.

Possible solutions

Dynamic update of dictionary file to improve sentiment analysis for unknown tweets

Ex: algorithm to add new words in the dictionary using synonyms and antonyms words.

Knowledge of person’s type of comments/behavioral comments over a period of time will help us understand the type of his/her tweets. It will help us in determining the sarcasm or ironic type of comments and his score his sentiment accordingly. This data can be used in the system to predict the sentiment.