Principles of BigData

Phase 2

**Project Team:**

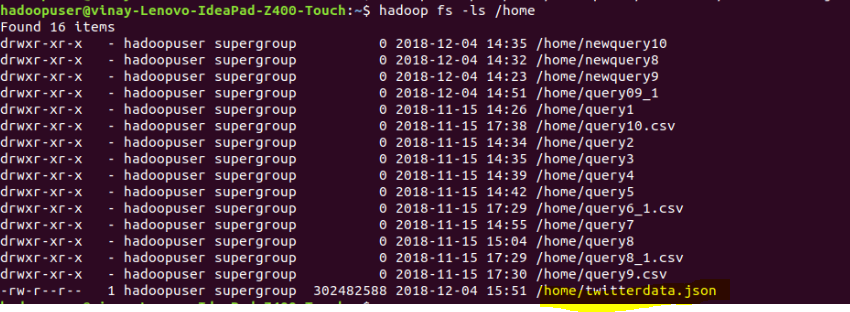
Sravya Para,

Hemasree Koganti,

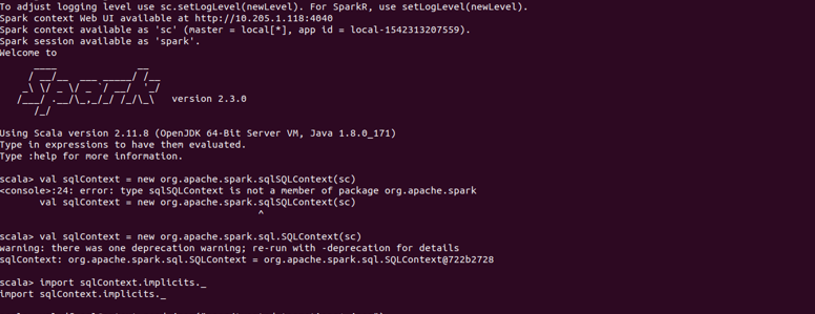
Sai Priyanka Gurajada.

# Setting up the spark:

Before executing the queries in spark we have uploaded the json file to HDFS directory:



We imported the required libraries ams started running our queries in spark.





In the below screenshot, the json file is converted into a table

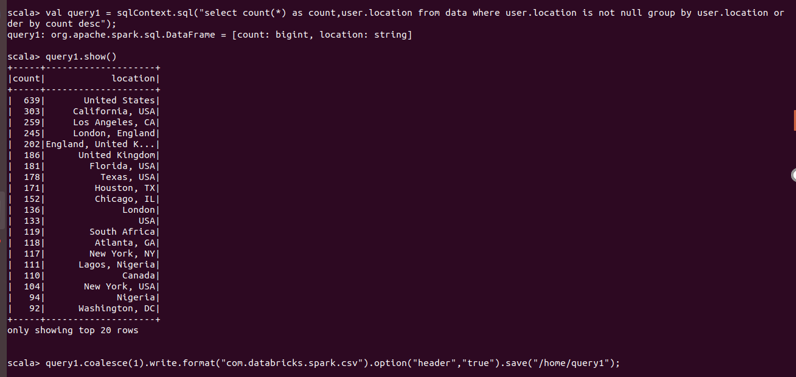


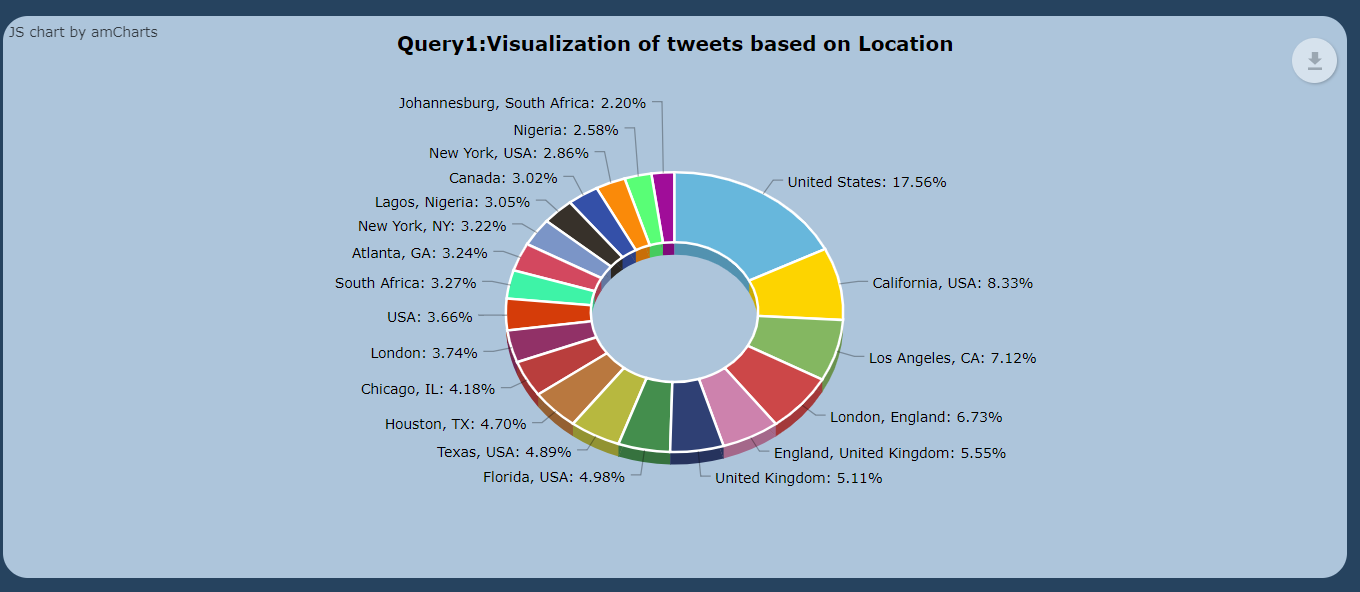
# Queries:

For visualization we used Am charts and ran the script in WebStorm using java script

## Query1:

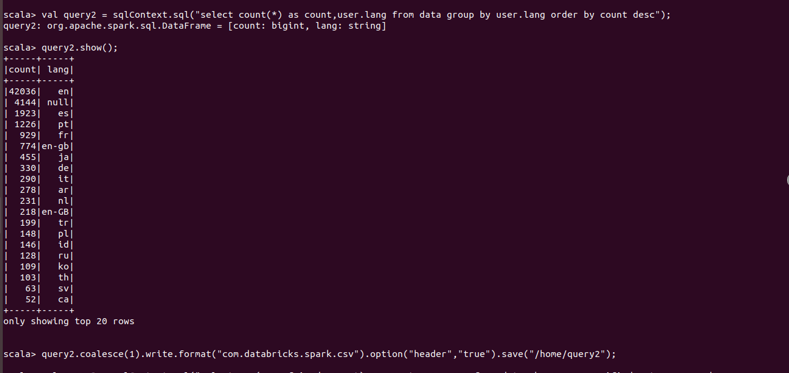
The below query gives the count of number of users in a location. The query displays the overall count of people from a location.

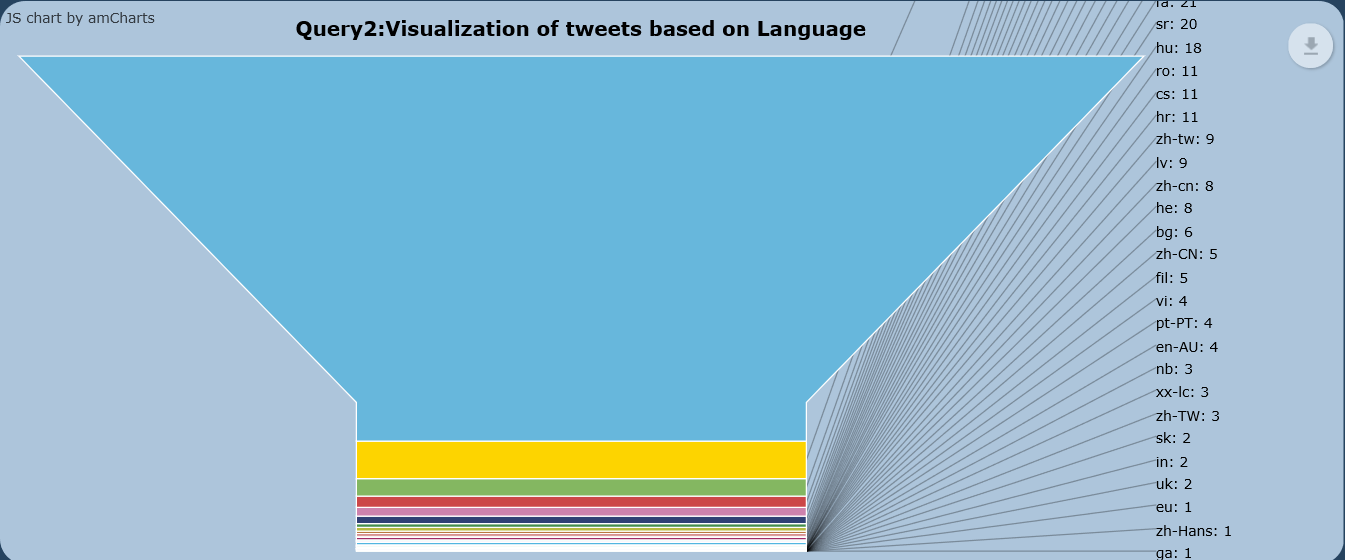




## Query 2:

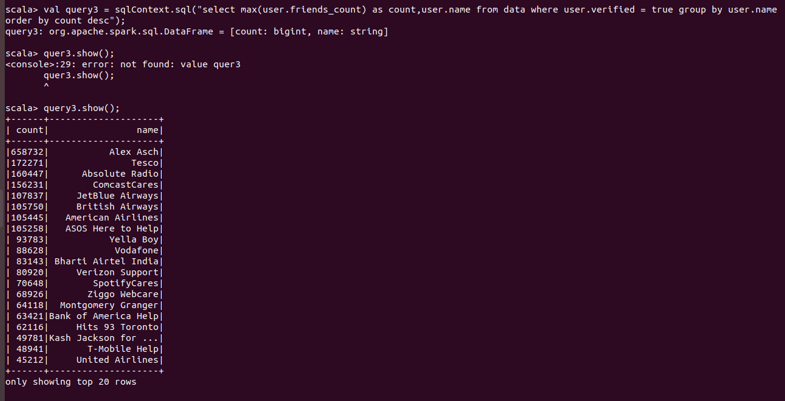
The below query shows the total count of users on a specific language .

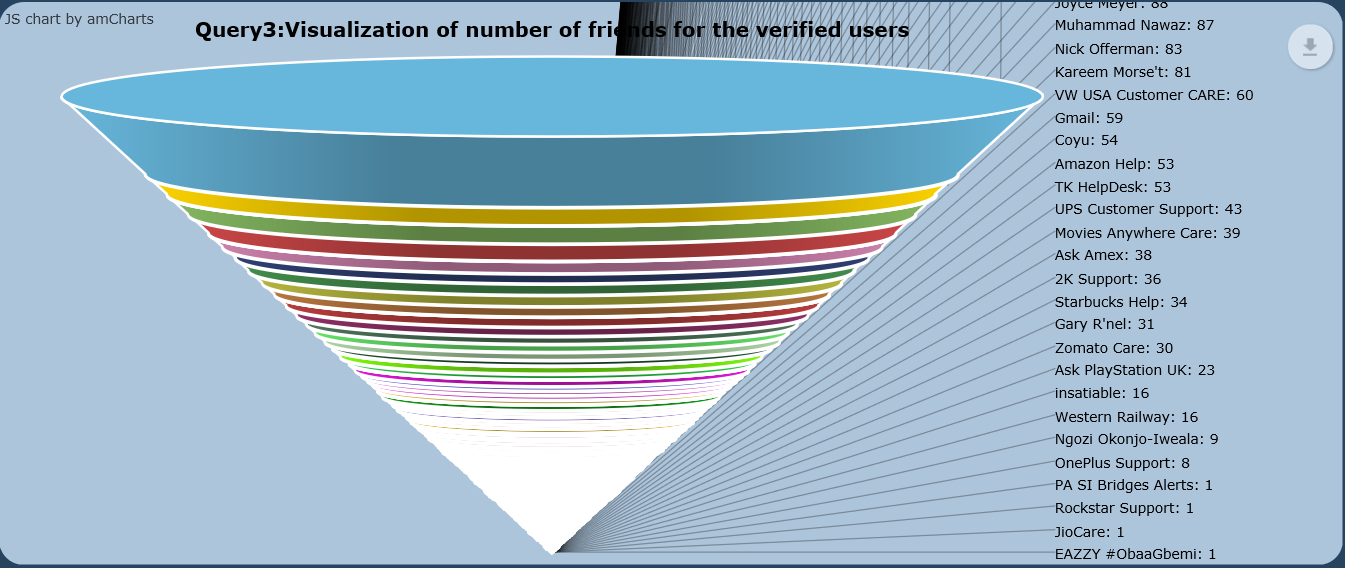




## Query3:

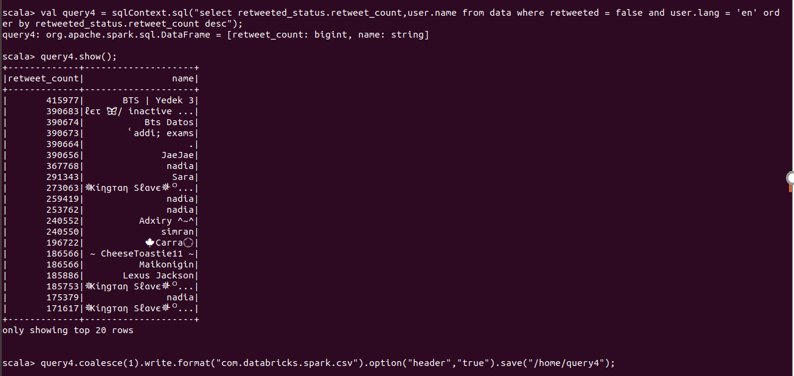
The below query shows verified users with number of friends count for verified users. The output is the user name and their friends count

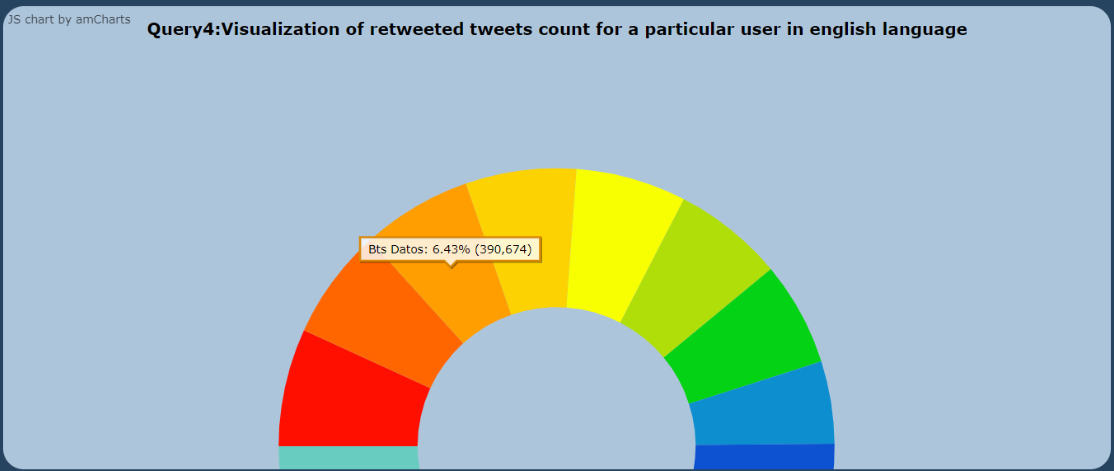




## Query 4 :

The below query displays the most retweeted tweet in English language and the count of number of times it has been retweeted.

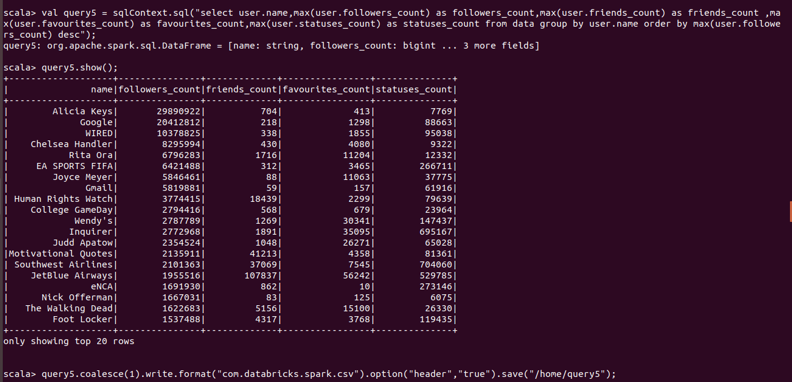


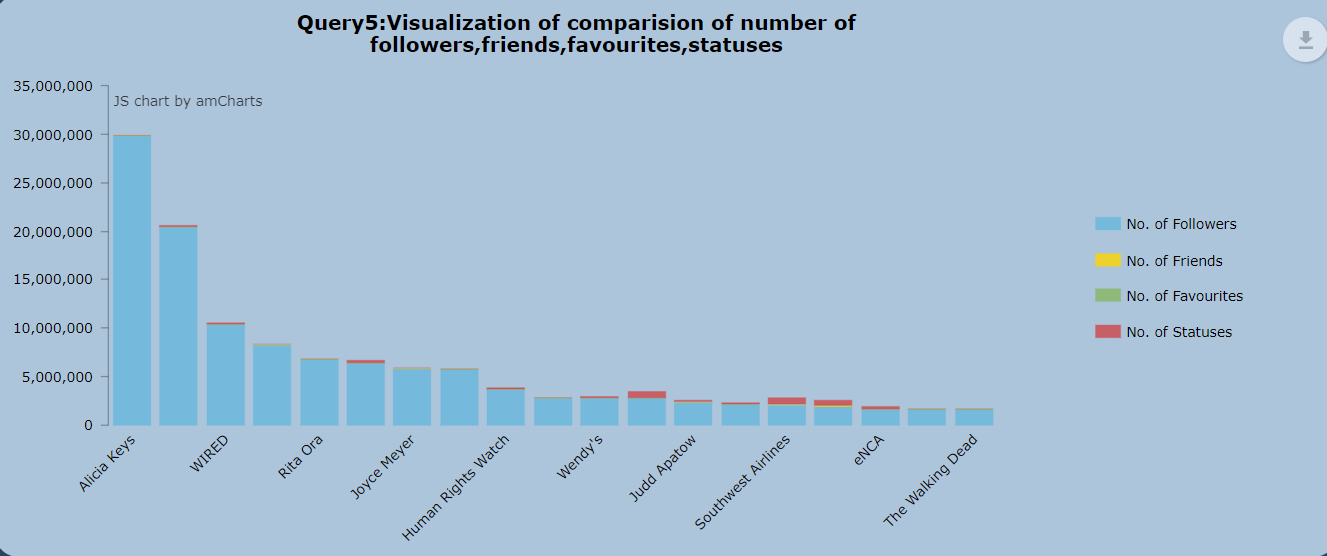


## Query 5:

The below query displays user name with maximum followers count, friends count, favourites count,

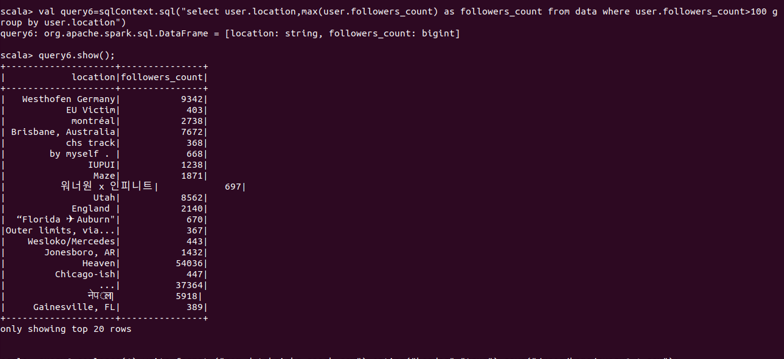
And the number of status they have posted.

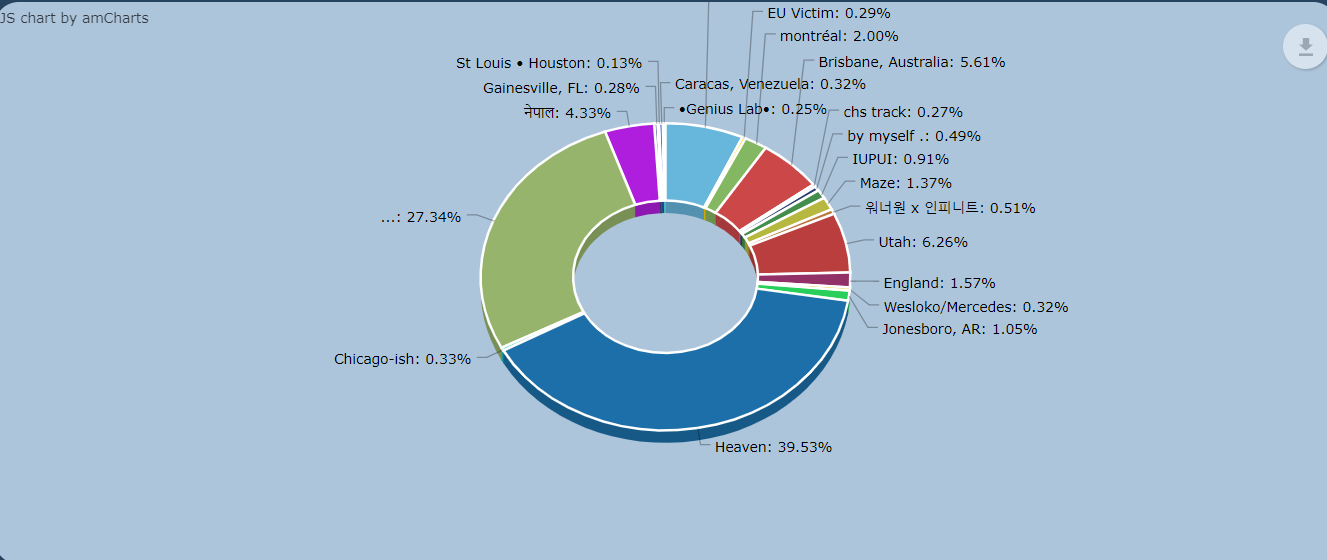




## Query 6:

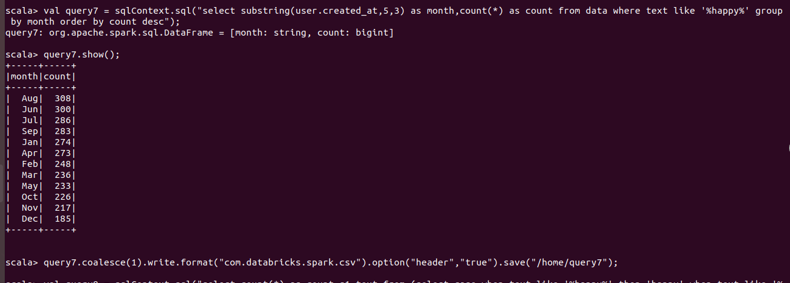
The query shows the the users locatons and followerscount in that location in which followers count is more than 100 grouped with location.

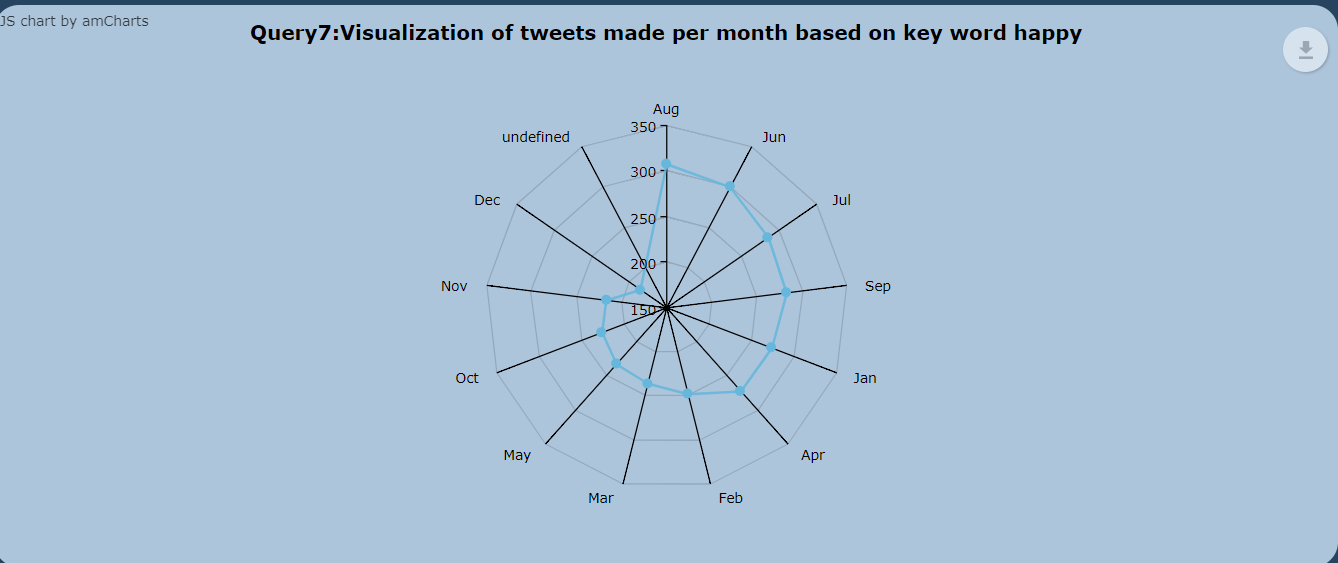




## Query 7:

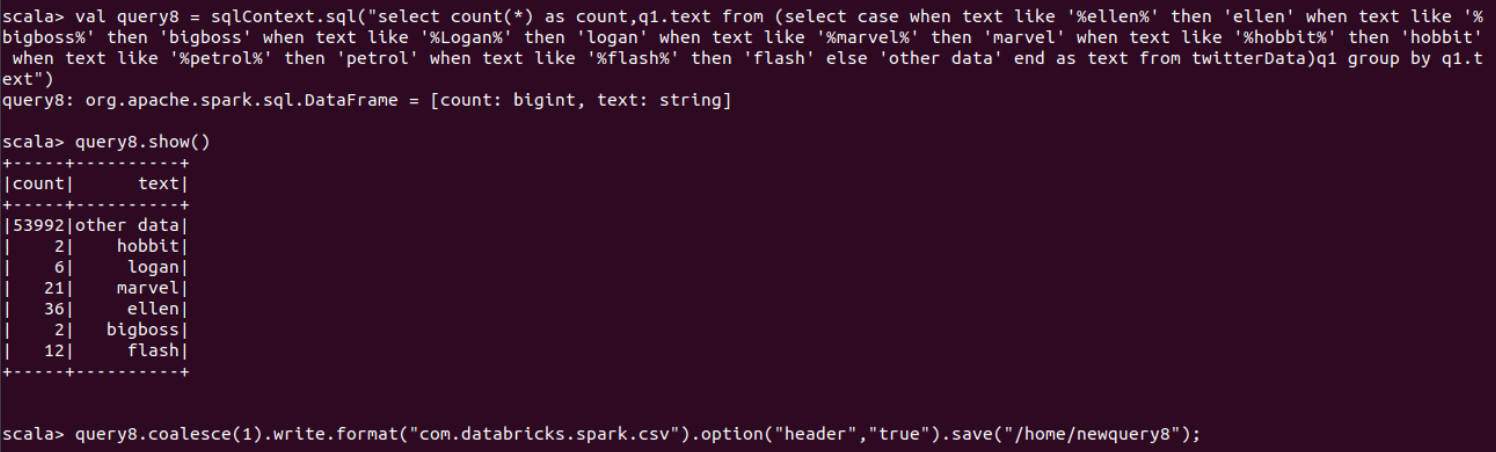
The below query displays number of users has tweeted ‘happy’ in the months

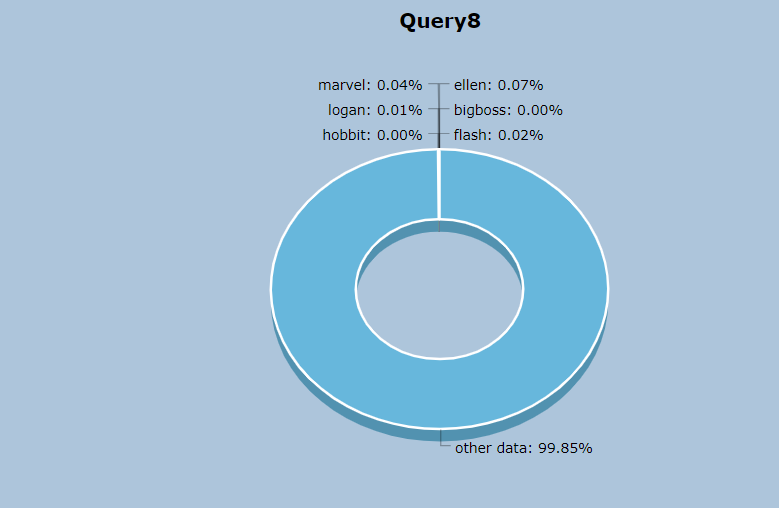




## Query 8:

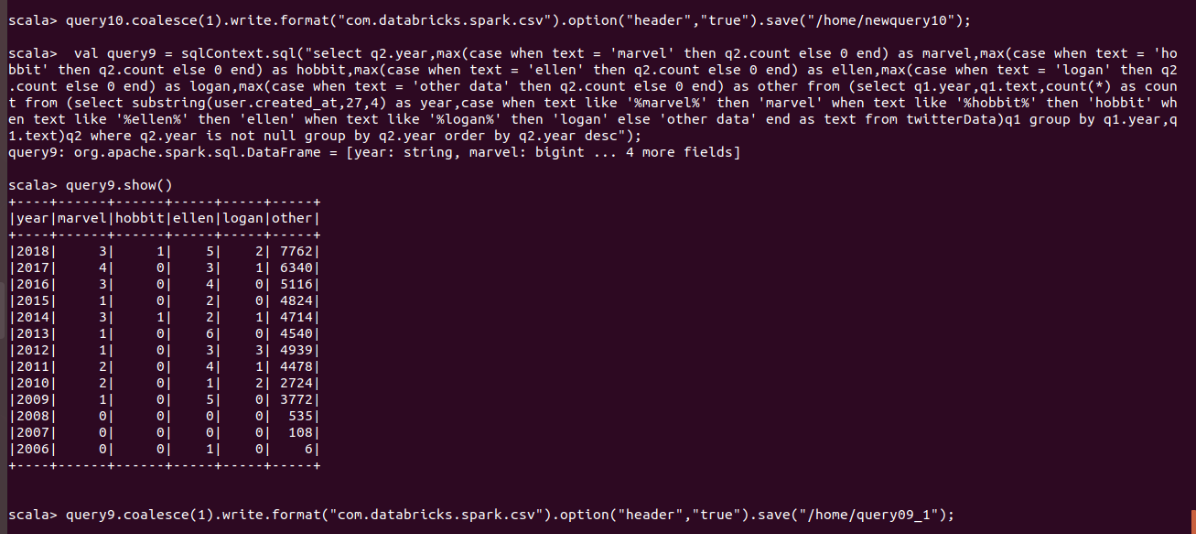
Count of the tweets that contained any of hobbit, logan, marvel, ellen, bigboss, flash and the rest any of the words

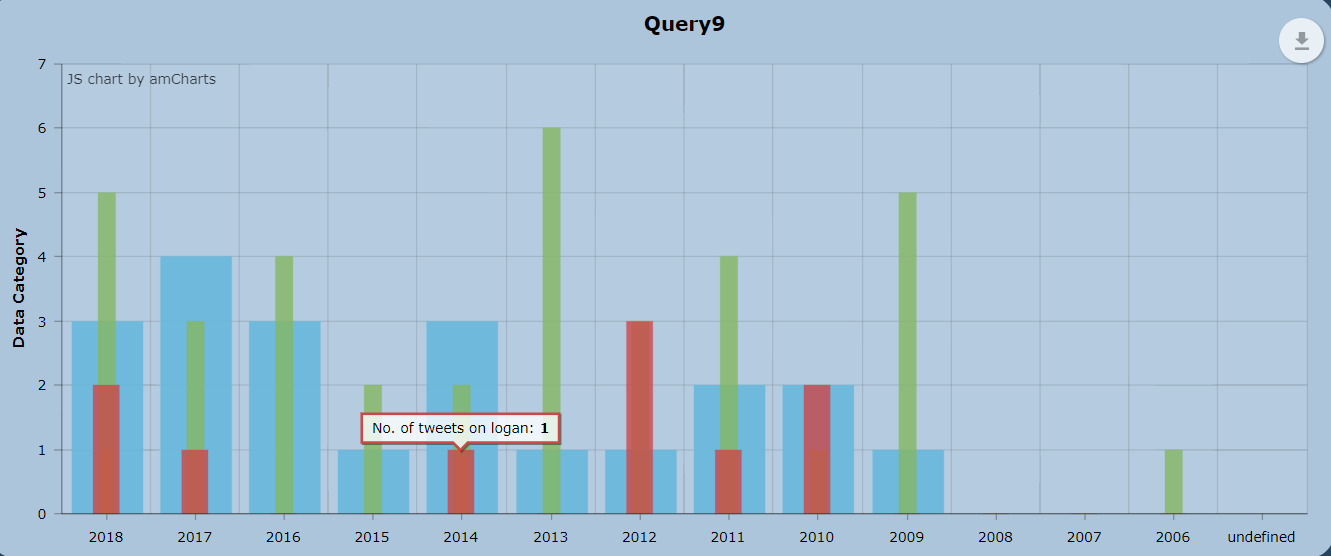




## Query9:

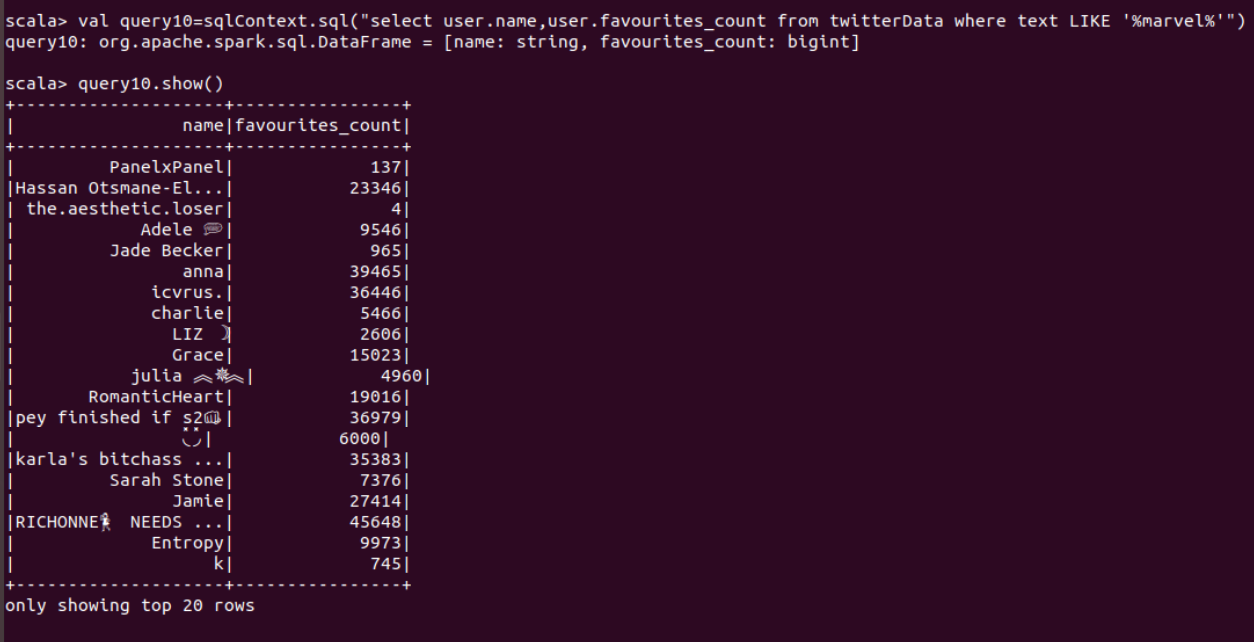
Query shows the number of times a word is tweeted in particular years.

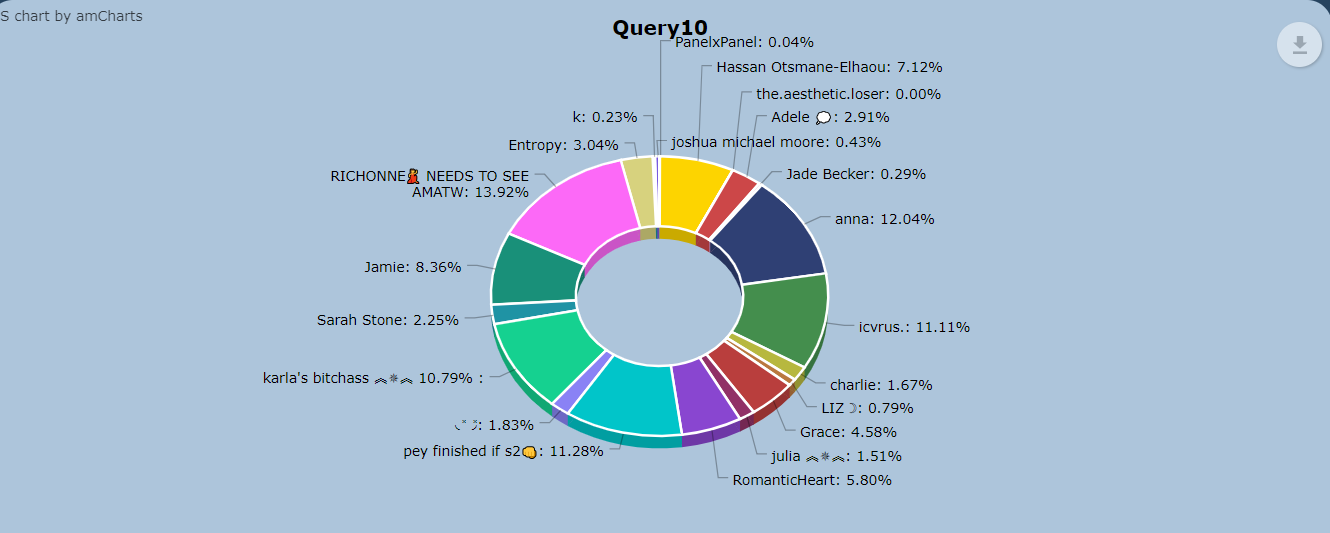




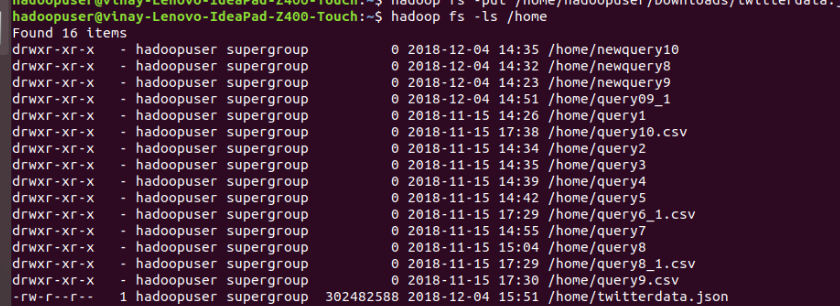
## Query 10:

Query shows the username , favourites count of a user whose data contains “marvel”

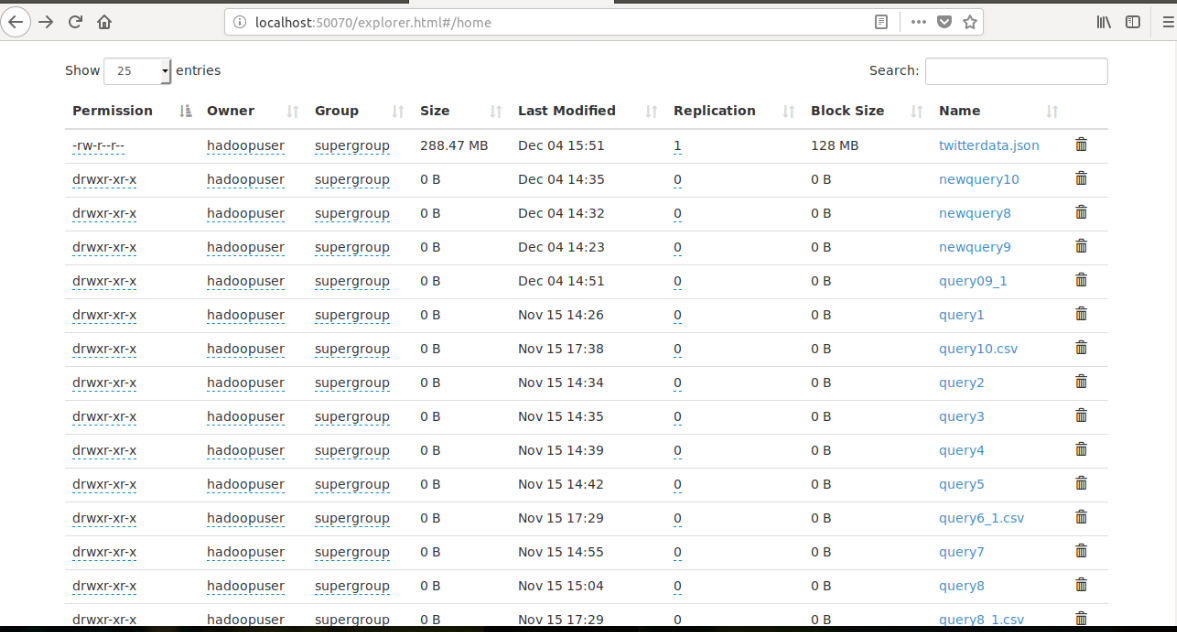




Location of the queries saved in HDFS



Browsing through HDFS :



The front end code and query scripts can be found in below the GitHub link: <https://github.com/sravyapara/bigdata_project>