

# Twitter Analysis on Music

Shuchita Khandelwal, Anuja Sawant, Tarun Chandrashekar, Shubhabrata Mukherjee  
University of Missouri Kansas City

## INTRODUCTION

Python was used to obtain tweets from Twitter API. Analysis of tweets help us in understanding various constraints of user responses. The hashtag used was '#Music'. Illustrations have been shown below to characterize the different patterns and inferences can be made to obtain statistical data.

## ARCHITECTURE

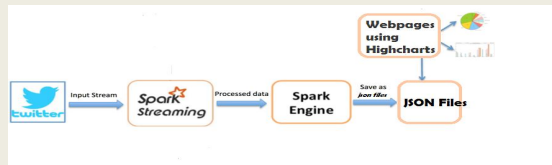
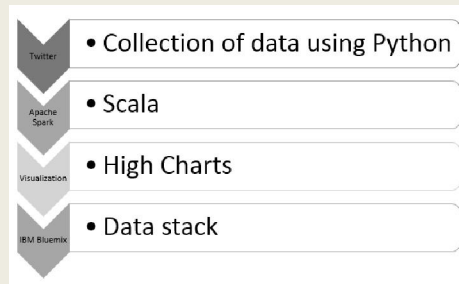


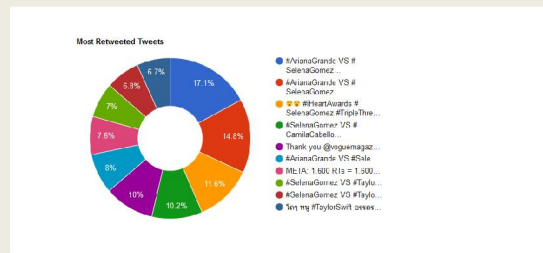
Figure 1.a. Architecture



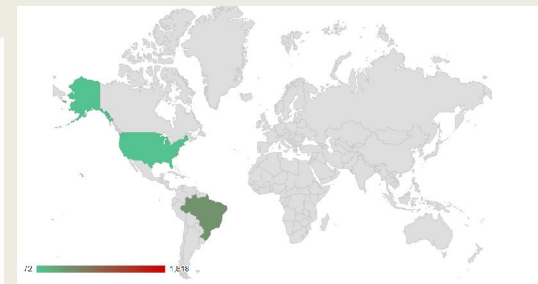
## ANALYSIS

Twitter served as the framework from which we were able to collect data in the form of tweets. The input stream was collected in the form of tweets using Spark Streaming. We use the Spark engine to analyze the tweets that are taken from the streaming and generate a schema for the tweets collected. Tweets are analyzed using the sql queries run in the spark engine. We have taken the results of the queries and written those in json files which were later processed to represent diagrammatically. The basic interpretation of this project was to understand how big data was obtained and how analysis can be done by grouping them. Interpretation of data can be implemented using illustrations such as charts and graphs that add significance to the respective characteristic of data.

## TWITTER ANALYSIS OUTPUTS



```
val r7 = sqlContext.sql("SELECT MAX(retweeted_status.retweet_count) as maxretweetcnt , retweeted_status.text FROM querytable GROUP BY retweeted_status.text ORDER BY maxretweetcnt DESC LIMIT 10")<br>r7.coalesce(1).write.json("srt")
```



```
sqlContext.sql("SELECT user.time_zone, SUBSTR(created_at, 0, 9), COUNT(*) AS total_count FROM querytable WHERE user.time_zone IS NOT NULL GROUP BY user.time_zone, SUBSTR(created_at, 0, 9) ORDER BY total_count DESC LIMIT 15")<br>r2.coalesce(1).write.json("time")
```

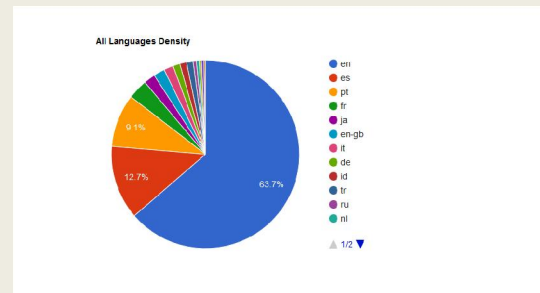
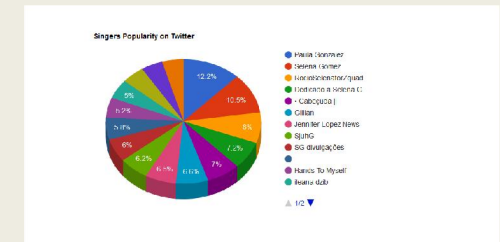


Figure 3: User Devices

```
sqlContext.sql("SELECT user.lang , COUNT(*) as cnt FROM querytable WHERE user.lang IS NOT NULL GROUP BY user.lang ORDER BY cnt DESC LIMIT 15")<br>r8.coalesce(1).write.json("lang")
```



```
val query = sqlContext.sql("select user.name, count(user.followers_count) as followersCount from querytable1 group by user.name order by followersCount desc limit 10")<br>r3.write.json("followers")
```

## Execution Time of queries

Queries	Execution time	Total tasks
1. Locations Query	26s	12
2. User Devices	130s	60
3. Languages	26s	12
4. Celebrities	25s	12
5. Hash Tags	16s	12
6. Followers	26s	12
7. Favorites	26s	12
8. Retweeted Queries	26s	12
9. Most tweeted users	26s	12

## FUTURE ASPECTS

The future aspects of this analysis is to integrate with the third party application to use this analysis and use this results to show more complicated analysis. Even write the more complex queries such as to find the sentimental analysis on ISIS attacks.