Cricket World Cup T20 Tweet Analysis

In this project we collected twitter tweets on cricket world cup t20 and analyzed them using **Apache Spark** and the results of the query are **dynamically stored in MongoDB in JSON format.**

The **Data from MongoDB is retrieved using mongo api** and visualized using **Google Charts** and **High Charts.**

The Whole source code is **deployed and hosted in IBM Bluemix** so that we can access visualization globally.

Tools Technologies Used:

1. IntelliJ Idea
2. Web storm,
3. Spark 1.6.1
4. Python
5. Java REST API
6. Scala
7. Angular js for front end
8. Google Charts, High Charts for visualization
9. IBM Bluemix for hosting Application

IBM Bluemix Link:
http://bigdataspringdeploymentfinal.mybluemix.net/

Mlab data:

Query1:
https://api.mlab.com/api/1/databases/pb/collections/query1?apiKey=Q_u73BV4oOdMGpnu3WFGmJ8YH_lxHDHO

Query2:
https://api.mlab.com/api/1/databases/pb/collections/query2?apiKey=Q_u73BV4oOdMGpnu3WFGmJ8YH_lxHDHO

Query3:
https://api.mlab.com/api/1/databases/pb/collections/query3?apiKey=Q_u73BV4oOdMGpnu3WFGmJ8YH_lxHDHO

Query4:
https://api.mlab.com/api/1/databases/pb/collections/query4?apiKey=Q_u73BV4oOdMGpnu3WFGmJ8YH_lxHDHO

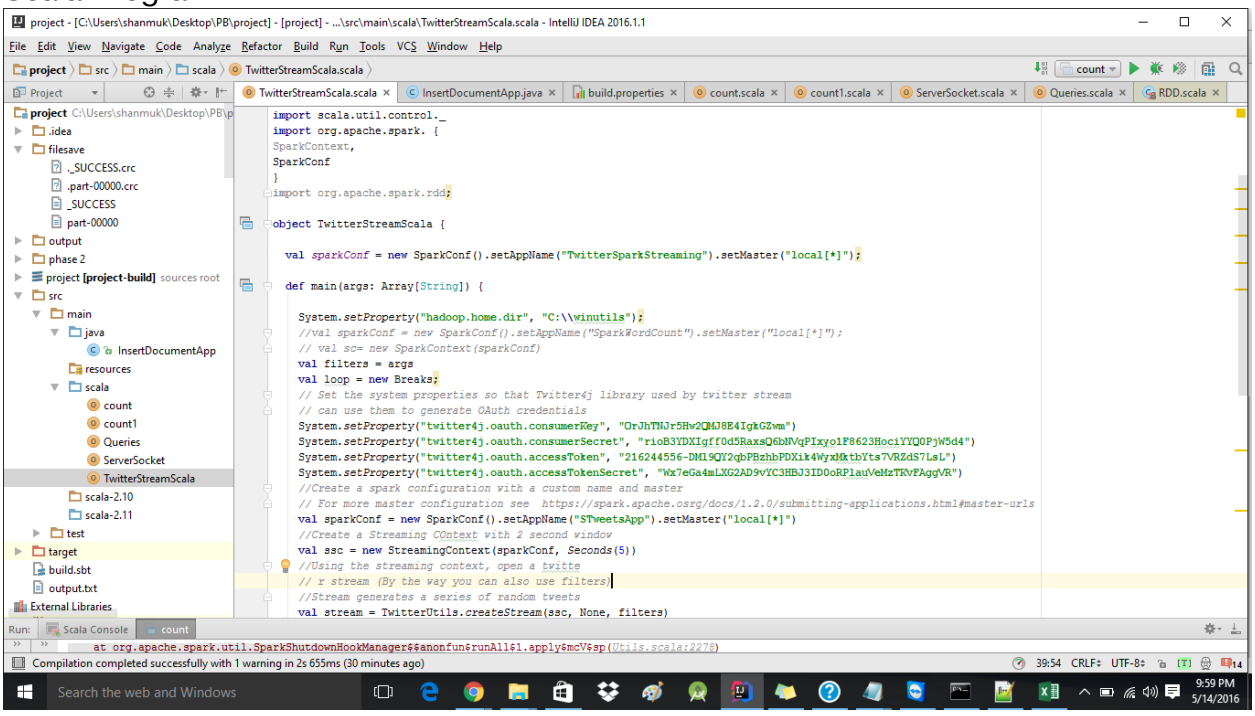
Query5:
https://api.mlab.com/api/1/databases/pb/collections/query5?apiKey=Q_u73BV4oOdMGpnu3WFGmJ8YH_lxHDHO

Query6:
https://api.mlab.com/api/1/databases/pb/collections/query6?apiKey=Q_u73BV4oOdMGpnu3WFGmJ8YH_lxHDHO

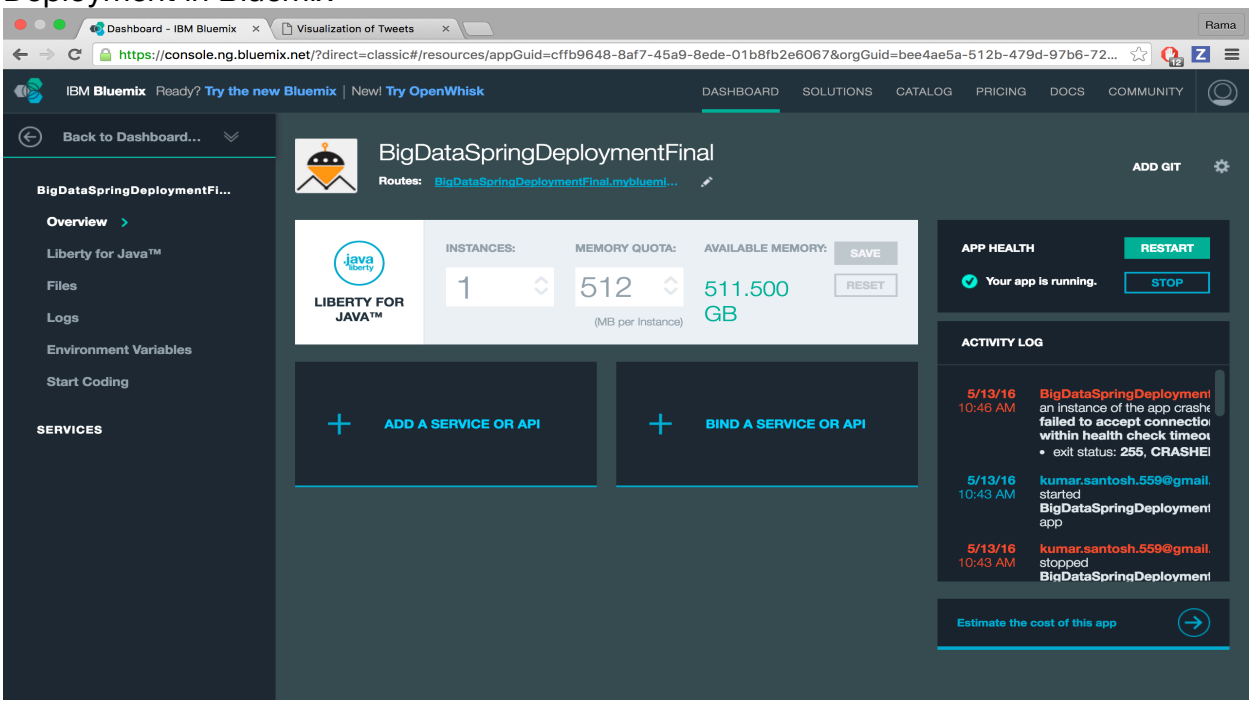
Query7:
https://api.mlab.com/api/1/databases/pb/collections/query7?apiKey=Q_u73BV4oOdMGpnu3WFGmJ8YH_lxHDHO

Query8:
https://api.mlab.com/api/1/databases/pb/collections/query8?apiKey=Q_u73BV4oOdMGpnu3WFGmJ8YH_lxHDHO

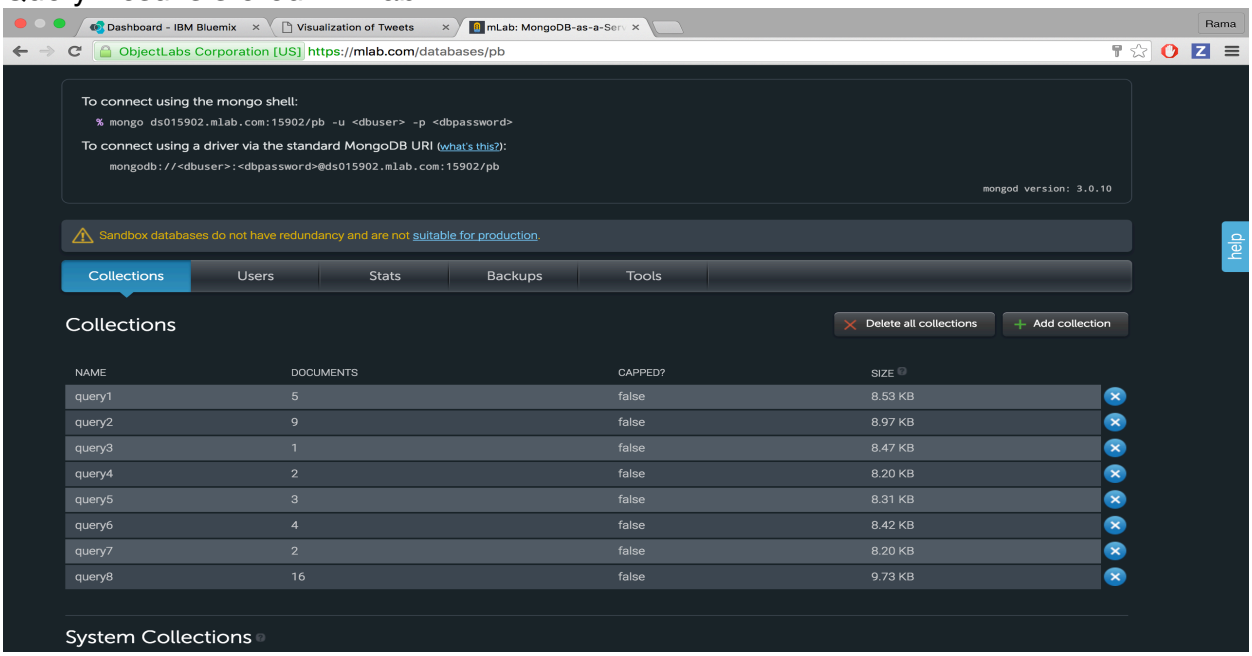
Screenshots of the Projects:
Scala Program

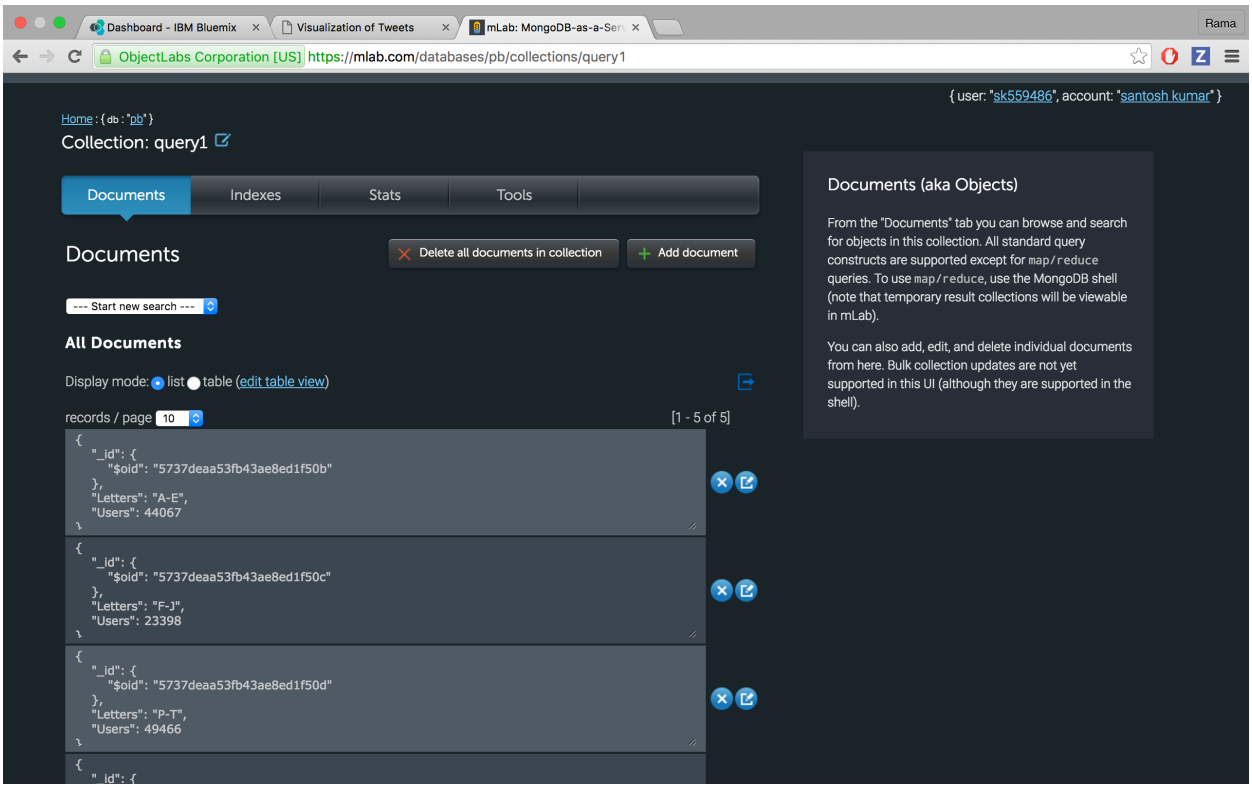


Deployment in Bluemix

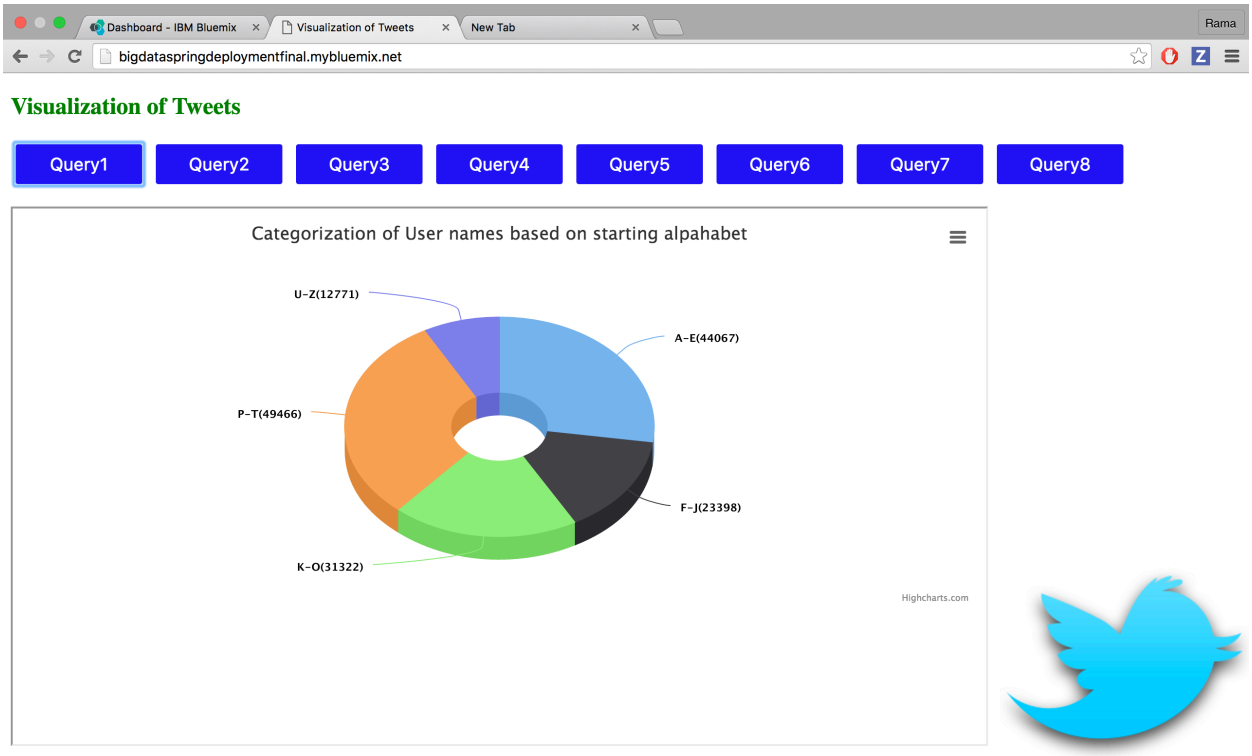


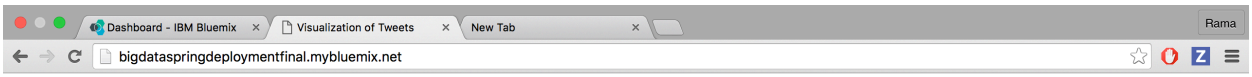
Query Results stored in Mlab





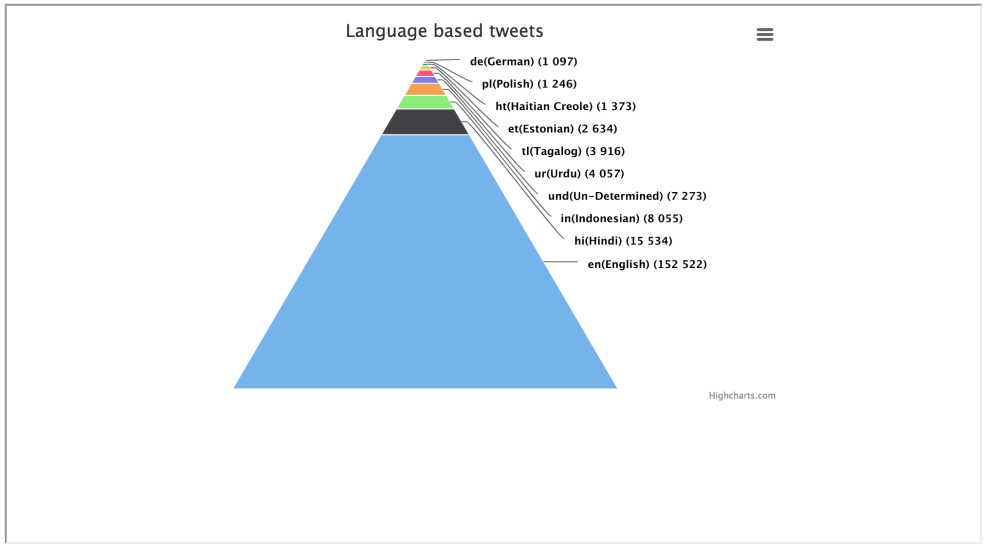
Visualization screenshots

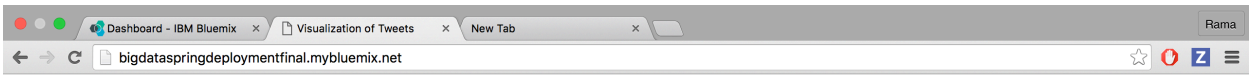




Visualization of Tweets

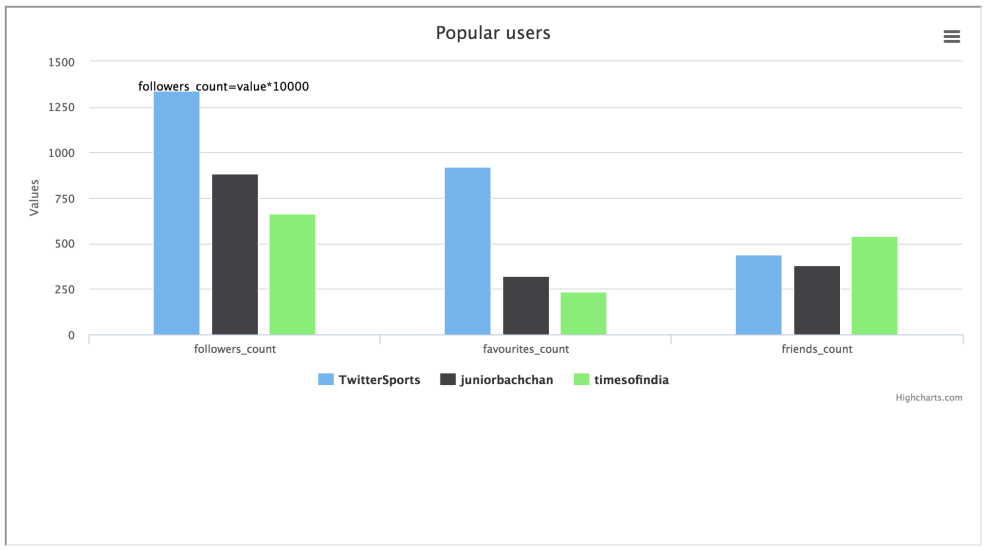
Query1 | Query2 | Query3 | Query4 | Query5 | Query6 | Query7 | Query8

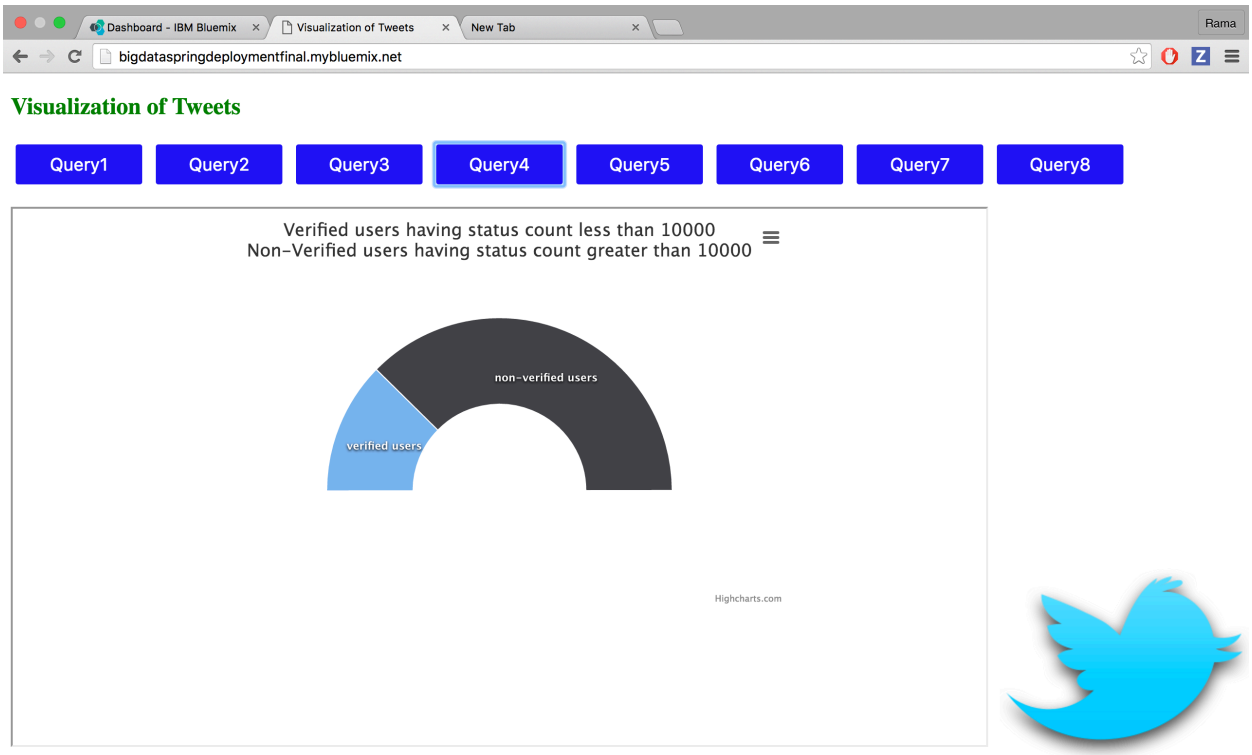


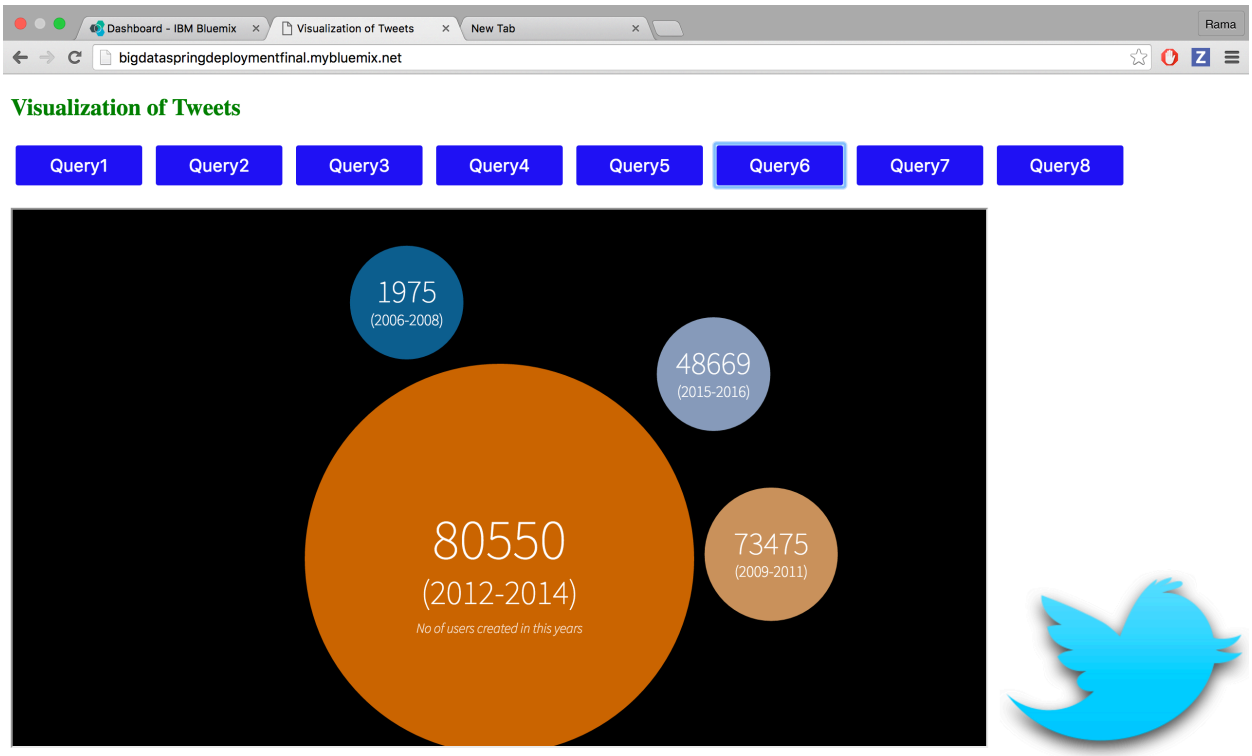


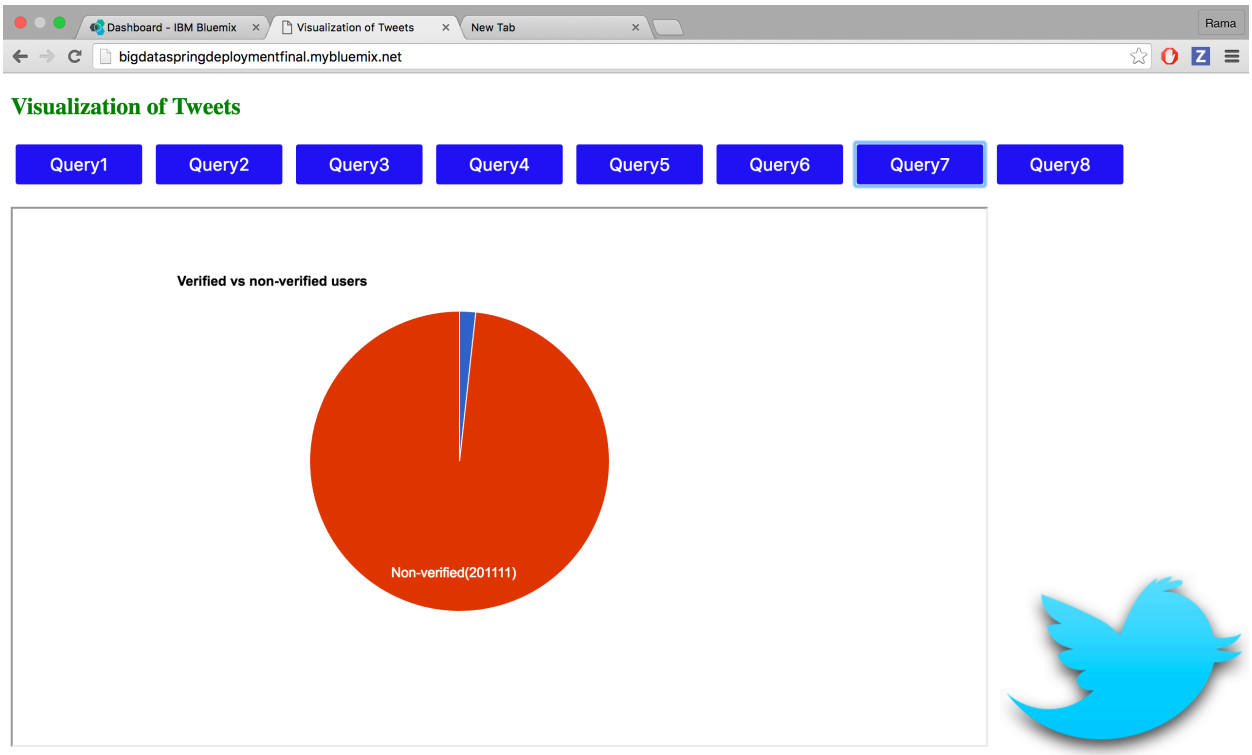
Visualization of Tweets

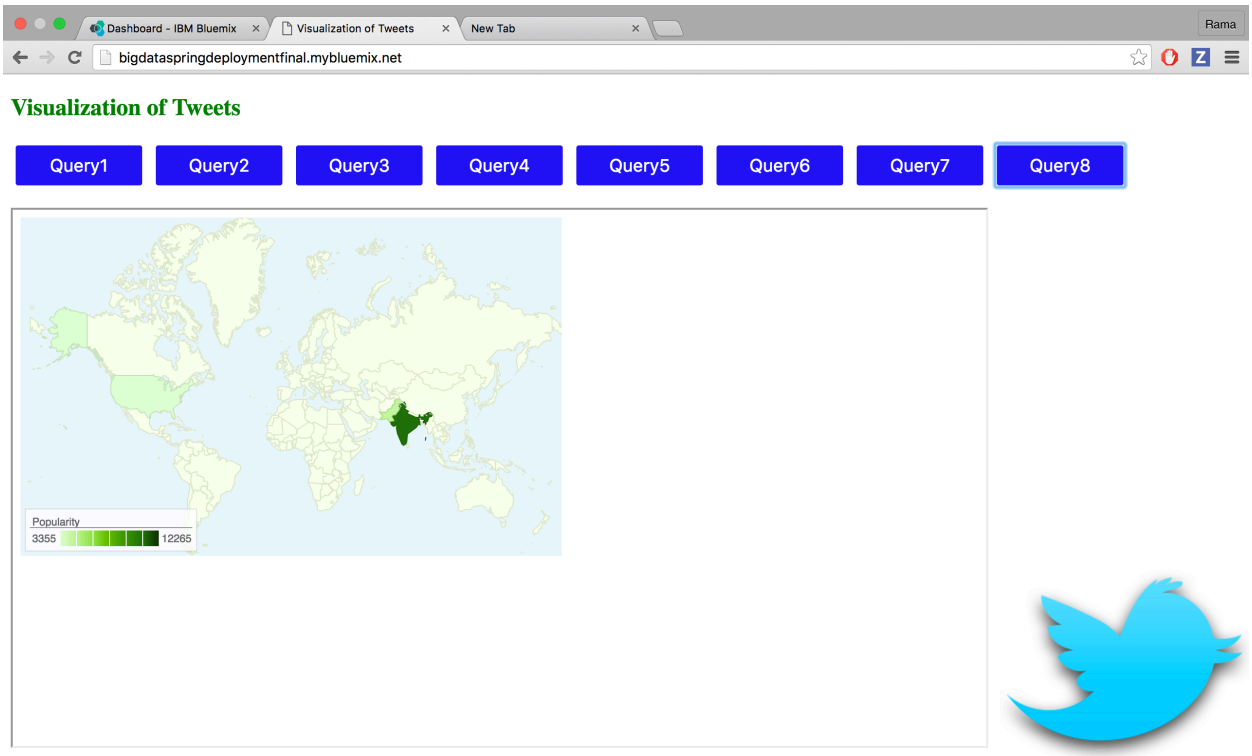
Query1 | Query2 | Query3 | Query4 | Query5 | Query6 | Query7 | Query8




Dashboard - IBM Bluemix
Visualization of Tweets
New Tab
Rama
bigdataspringdeploymentfinal.mybluemix.net
Visualization of Tweets
Query1
Query2
Query3
Query4
Query5
Query6
Query7
Query8
Verified users having status count less than 10000
Non-Verified users having status count greater than 10000
non-verified users
verified users
Highcharts.com


Dashboard - IBM Bluemix
Visualization of Tweets
New Tab
Rama
bigdataspringdeploymentfinal.mybluemix.net
Visualization of Tweets
Query1
Query2
Query3
Query4
Query5
Query6
Query7
Query8
1975
(2006-2008)
48669
(2015-2016)
80550
(2012-2014)
No of users created in this years
73475
(2009-2011)


Dashboard - IBM Bluemix
Visualization of Tweets
New Tab
Rama
bigdataspringdeploymentfinal.mybluemix.net
Visualization of Tweets
Query1
Query2
Query3
Query4
Query5
Query6
Query7
Query8
Verified vs non-verified users
Non-verified(201111)


Dashboard - IBM Bluemix
Visualization of Tweets
New Tab
Rama
bigdataspringdeploymentfinal.mybluemix.net
Visualization of Tweets
Query1
Query2
Query3
Query4
Query5
Query6
Query7
Query8
Popularity
3355
12265