

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_IxHDHO

Query2:

https://api.mlab.com/api/1/databases/pb/collections/query2?apiKey=Q_u73BV4oOdMGpnu3WFGmJ8YH_IxHDHO

Query3:

https://api.mlab.com/api/1/databases/pb/collections/query3?apiKey=Q_u73BV4oOdMGpnu3WFGmJ8YH_IxHDHO

Query4:

https://api.mlab.com/api/1/databases/pb/collections/query4?apiKey=Q_u73BV4oOdM Gpnu3WFGmJ8YH_IxHDHO

Query5:

https://api.mlab.com/api/1/databases/pb/collections/query5?apiKey=Q_u73BV4oOdM Gpnu3WFGmJ8YH_IxHDHO

Query6:

https://api.mlab.com/api/1/databases/pb/collections/query6?apiKey=Q_u73BV4oOdM Gpnu3WFGmJ8YH_IxHDHO

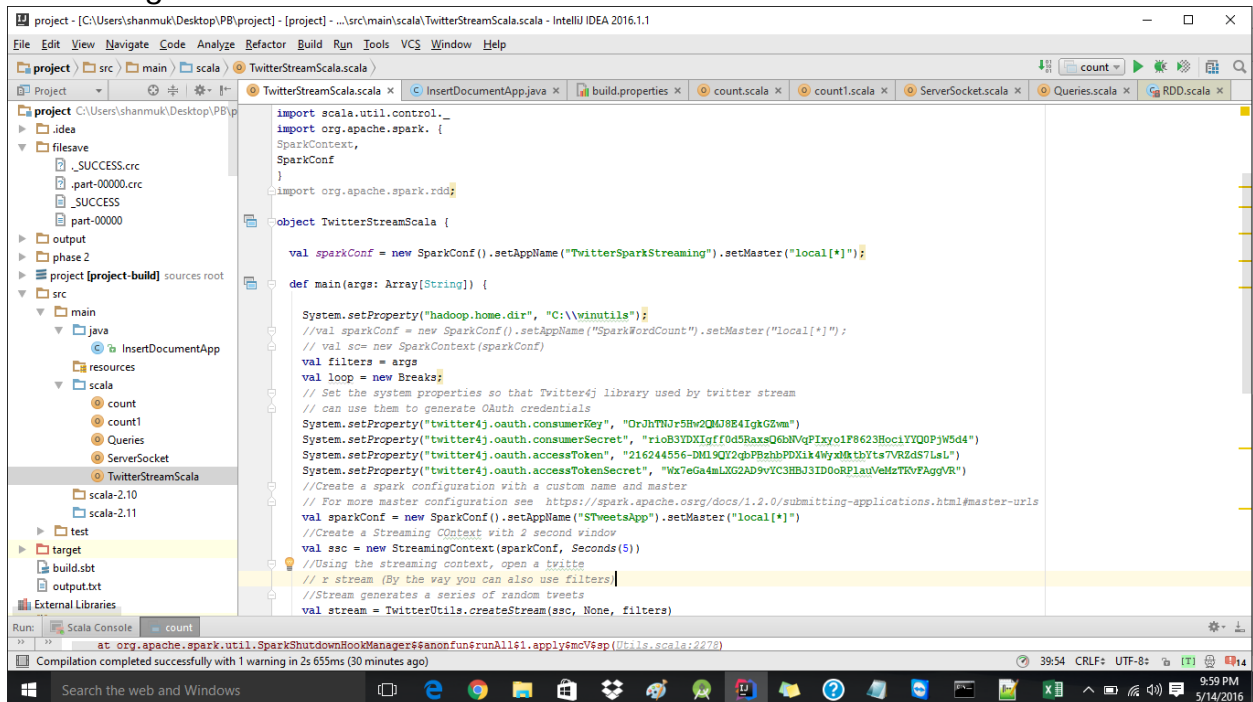
Query7:

https://api.mlab.com/api/1/databases/pb/collections/query7?apiKey=Q_u73BV4oOdM Gpnu3WFGmJ8YH_IxHDHO

Query8:

https://api.mlab.com/api/1/databases/pb/collections/query8?apiKey=Q_u73BV4oOdM Gpnu3WFGmJ8YH_IxHDHO

Screenshots of the Projects: Scala Program



Deployment in Bluemix

The screenshot shows the IBM Bluemix console interface. The top navigation bar includes links for Dashboard, Solutions, Catalog, Pricing, Docs, and Community. The main content area displays the deployment details for 'BigDataSpringDeploymentFinal'. The 'Overview' tab is selected, showing the app's status as 'Running'. The 'Instances' section shows 1 instance with a memory quota of 512 MB and available memory of 511.500 GB. The 'App Health' section shows a green checkmark indicating the app is running. The 'Activity Log' section shows a list of events, including a crash report for 'BigDataSpringDeploymentFinal' on 5/13/16 at 10:46 AM. The 'Services' section is visible on the left sidebar.

BigDataSpringDeploymentFinal

Routes: BigDataSpringDeploymentFinal.mybluemix.net

Overview

Liberty for Java™

Instances: 1

Memory Quota: 512 MB

Available Memory: 511.500 GB

APP HEALTH: Your app is running.

ACTIVITY LOG

- 5/13/16 10:46 AM: BigDataSpringDeploymentFinal: an instance of the app crashed failed to accept connection within health check timeout. exit status: 255, CRASHED
- 5/13/16 10:43 AM: kumar.santosh.559@gmail.com started BigDataSpringDeploymentFinal app
- 5/13/16 10:43 AM: kumar.santosh.559@gmail.com stopped BigDataSpringDeploymentFinal app

Estimate the cost of this app

Query Results stored in Mlab

The screenshot shows the Mlab MongoDB console interface. The top navigation bar includes links for Dashboard, Solutions, Catalog, Pricing, Docs, and Community. The main content area displays the 'Collections' tab for a database named 'pb'. The 'Collections' section shows a list of collections with columns for Name, Documents, Capped?, and Size. The 'query1' collection has 5 documents and a size of 8.53 KB. The 'query2' collection has 9 documents and a size of 8.97 KB. The 'query3' collection has 1 document and a size of 8.47 KB. The 'query4' collection has 2 documents and a size of 8.20 KB. The 'query5' collection has 3 documents and a size of 8.31 KB. The 'query6' collection has 4 documents and a size of 8.42 KB. The 'query7' collection has 2 documents and a size of 8.20 KB. The 'query8' collection has 16 documents and a size of 9.73 KB. The 'System Collections' section is visible at the bottom.

To connect using the mongo shell:

```
% mongo ds015902.mlab.com:15902/pb -u <dbuser> -p <dbpassword>
```

To connect using a driver via the standard MongoDB URI (what's this?):

```
mongodb://<dbuser>:<dbpassword>@ds015902.mlab.com:15902/pb
```

Sandbox databases do not have redundancy and are not suitable for production

Collections

NAME	DOCUMENTS	CAPPED?	SIZE
query1	5	false	8.53 KB
query2	9	false	8.97 KB
query3	1	false	8.47 KB
query4	2	false	8.20 KB
query5	3	false	8.31 KB
query6	4	false	8.42 KB
query7	2	false	8.20 KB
query8	16	false	9.73 KB

System Collections

Dashboard - IBM Bluemix x Visualization of Tweets x mLab: MongoDB-as-a-Service x Rama

ObjectLabs Corporation [US] <https://mlab.com/databases/pb/collections/query1>

{ user: "sk559486", account: "santosh kumar" }

Home: { db: "pb" }

Collection: query1

Documents Indexes Stats Tools

Documents

Delete all documents in collection Add document

Start new search

All Documents

Display mode: list table (edit table view)

records / page 10 [1 - 5 of 5]

```

{
  "_id": {
    "$oid": "5737deaa53fb43ae8ed1f50b"
  },
  "Letters": "A-E",
  "Users": 44067
}
{
  "_id": {
    "$oid": "5737deaa53fb43ae8ed1f50c"
  },
  "Letters": "F-J",
  "Users": 23398
}
{
  "_id": {
    "$oid": "5737deaa53fb43ae8ed1f50d"
  },
  "Letters": "P-T",
  "Users": 49466
}
{
  "_id": {
    "$oid": "5737deaa53fb43ae8ed1f50e"
  },
  "Letters": "U-Z",
  "Users": 12771
}

```

Documents (aka Objects)

From the "Documents" tab you can browse and search for objects in this collection. All standard query constructs are supported except for `map/reduce` queries. To use `map/reduce`, use the MongoDB shell (note that temporary result collections will be viewable in mLab).

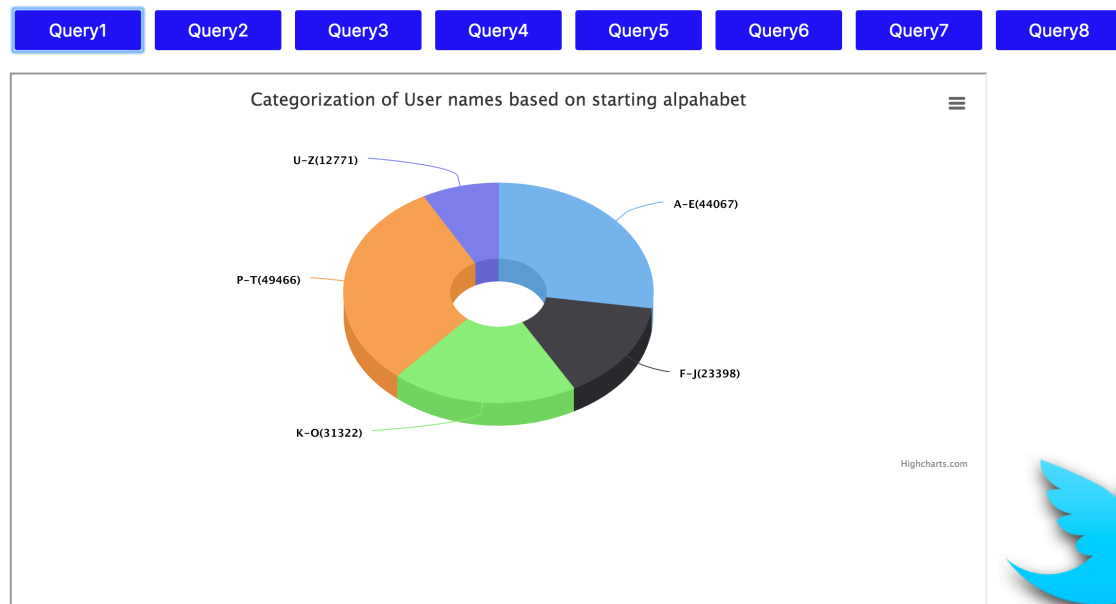
You can also add, edit, and delete individual documents from here. Bulk collection updates are not yet supported in this UI (although they are supported in the shell).

Visualization screenshots

Dashboard - IBM Bluemix x Visualization of Tweets x New Tab x Rama

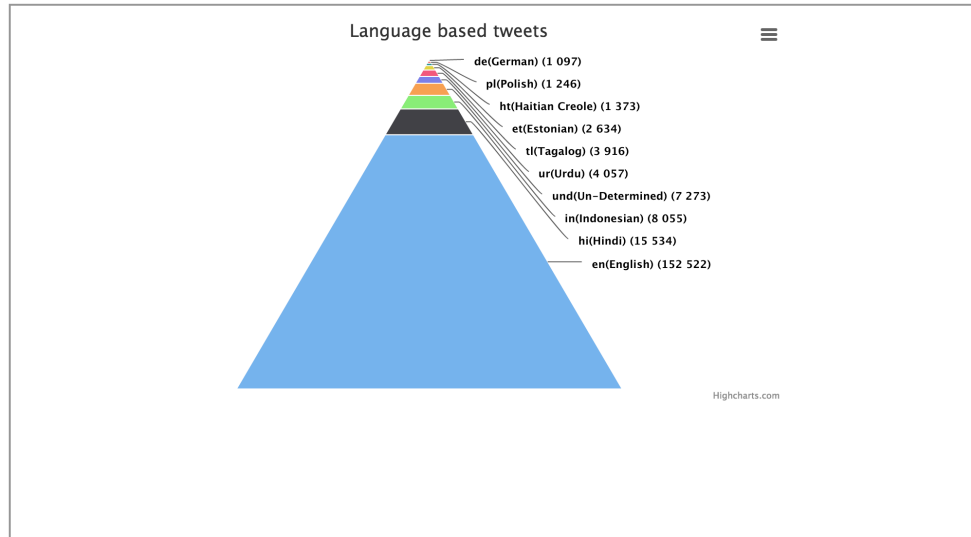
[bigdataspringdeploymentfinal.mybluemix.net](#)

Visualization of Tweets



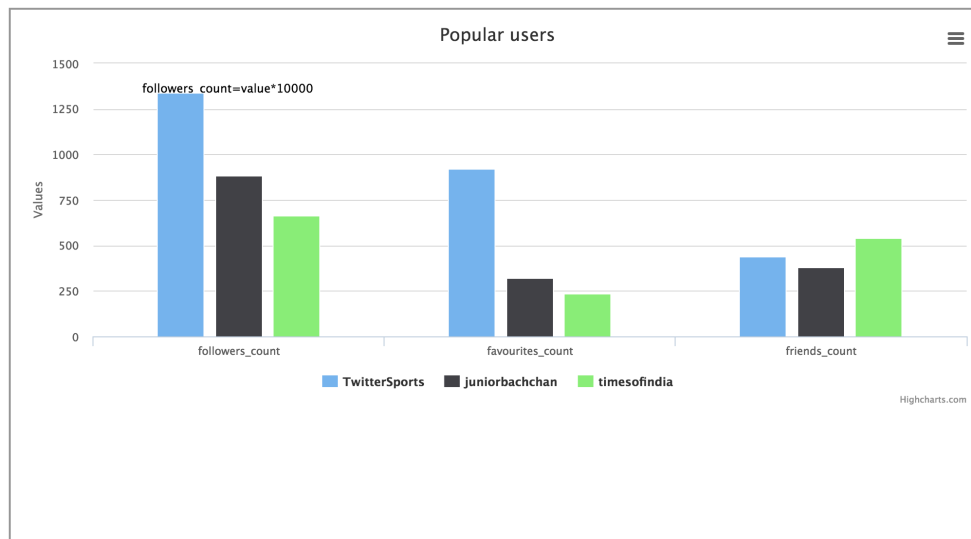
Visualization of Tweets

Query1 Query2 Query3 Query4 Query5 Query6 Query7 Query8



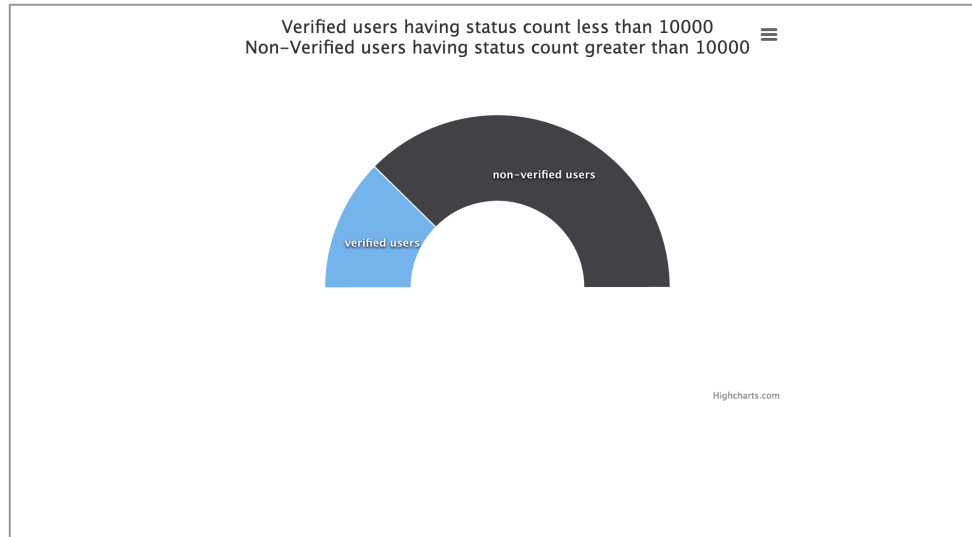
Visualization of Tweets

Query1 Query2 Query3 Query4 Query5 Query6 Query7 Query8



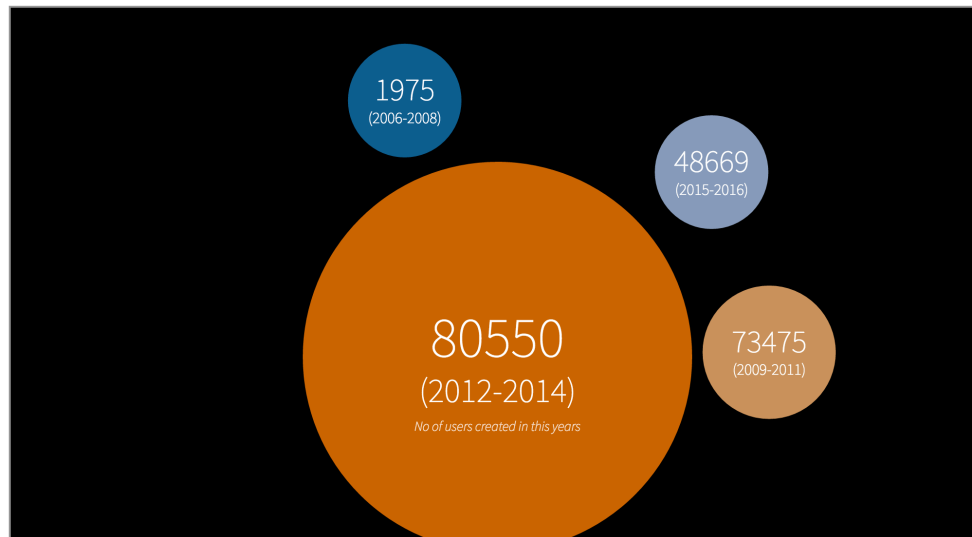
Visualization of Tweets

Query1 Query2 Query3 Query4 Query5 Query6 Query7 Query8



Visualization of Tweets

Query1 Query2 Query3 Query4 Query5 Query6 Query7 Query8



Visualization of Tweets

Query1

Query2

Query3

Query4

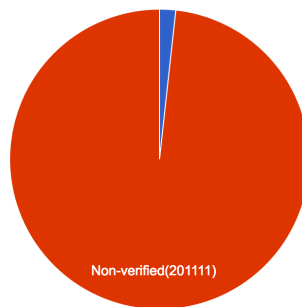
Query5

Query6

Query7

Query8

Verified vs non-verified users



Visualization of Tweets

Query1

Query2

Query3

Query4

Query5

Query6

Query7

Query8

