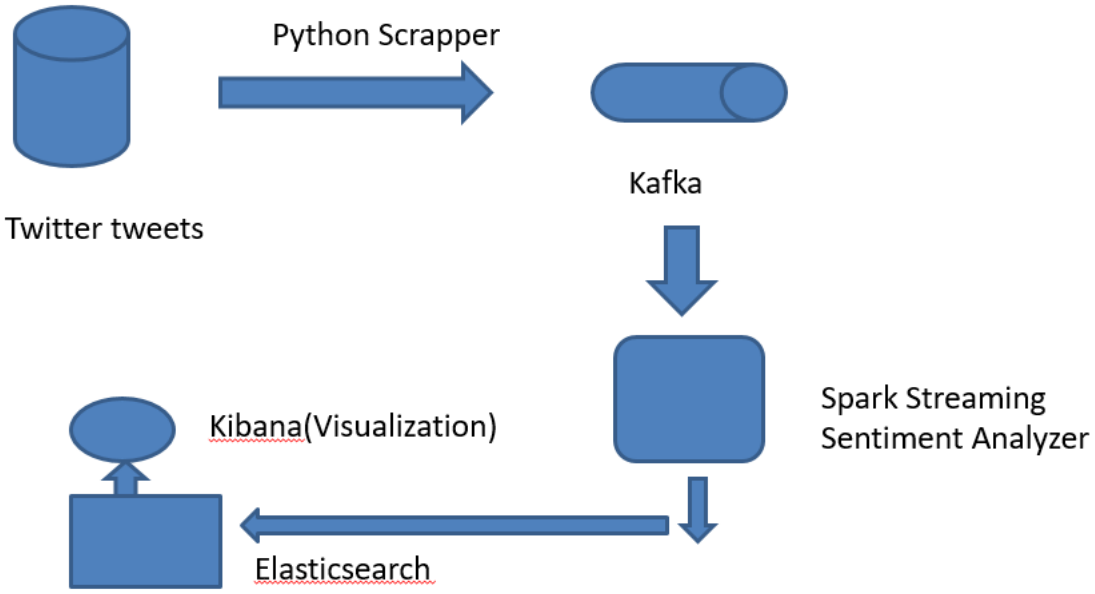
Workflow:



- First run the python scrapper to get the tweets from twitter
- Using the tweepy API to get the tweets. Filter the tweets based on location to get tweets only from USA. Also filter to get tweets having keywords trump and Obama
- Start the zookeeper server. It controls the various kafka brokers. Here we use Kafka because it is a very good fault tolerant messaging service. Even if one kafka broker fails the message is sent through another broker
- We create a kafka topic and send the messages.
- Now we create a kafka consumer for the required topic in Spark streaming. Here the tweets would be sent in real time to spark and we perform the sentiment analysis of tweets in spark, convert into json format and indexing it into elasticsearch.

For sentiment analysis I'm using textblob which is a python librabry for handling textual data. It performs the sentiment analysis and gives the result as polarity in the range -1 to +1(negative to positive)

- Before sending the data to elasticsearch, I have to first create an index, a type and then set proper mapping for the type. Since our data has a timestamp and location

coordinates it has to be specified in the mapping.

- Then we send data from spark streaming using the API provided by elasticsearch by

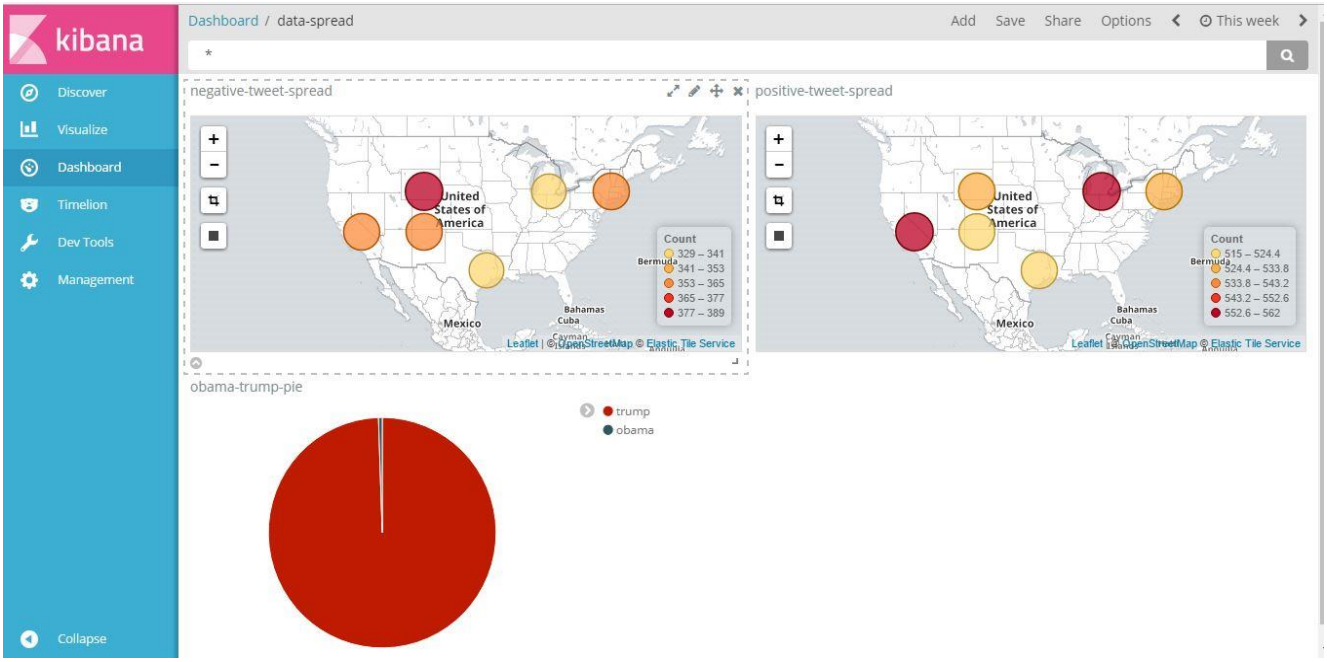
specifying the index and its type.

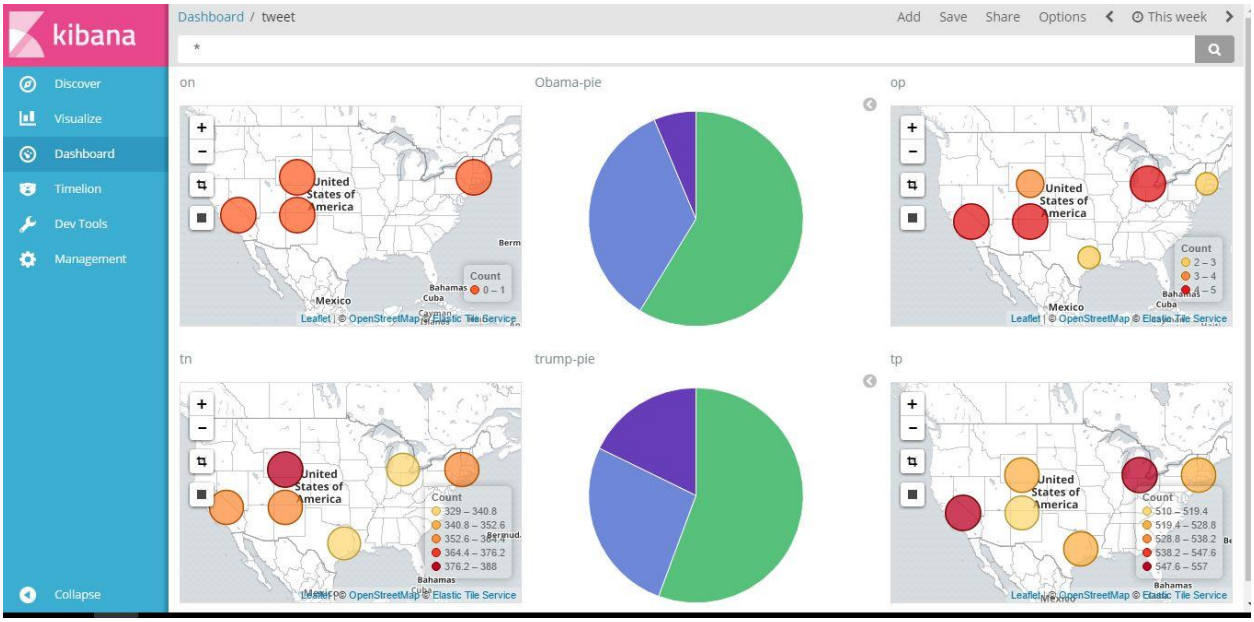
- Once we have the data in Elasticsearch we can analyze it by time and

create visualizations.

- I have created a dashboard having different maps for positive, negative and neutral tweets. I can even filter it on whether the data was on Obama or trump.
- Before creating visualizations I can define a search and store it(equivalent to views in a database) and create visualizations for these views.
- I created different Tile maps for different category of tweets and later added these visualizations to the dashboard. In a dashboard all the components are connected meaning, if we apply a filter to single component then it would be applied to all the visualizations in the dashboard.

Visualizations:




kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Collapse
Dashboard / tweet
Add
Save
Share
Options
This week
*
on
Obama-pie
op
tn
trump-pie
tp
United States of America
Mexico
Bahamas
Cuba
Count
0 – 1
2 – 3
3 – 4
4 – 5
329 – 340.8
340.8 – 352.6
352.6 – 364.4
364.4 – 376.2
376.2 – 388
510 – 519.4
519.4 – 528.8
528.8 – 538.2
538.2 – 547.6
547.6 – 557
Leaflet | © OpenStreetMap © Elastic Tile Service