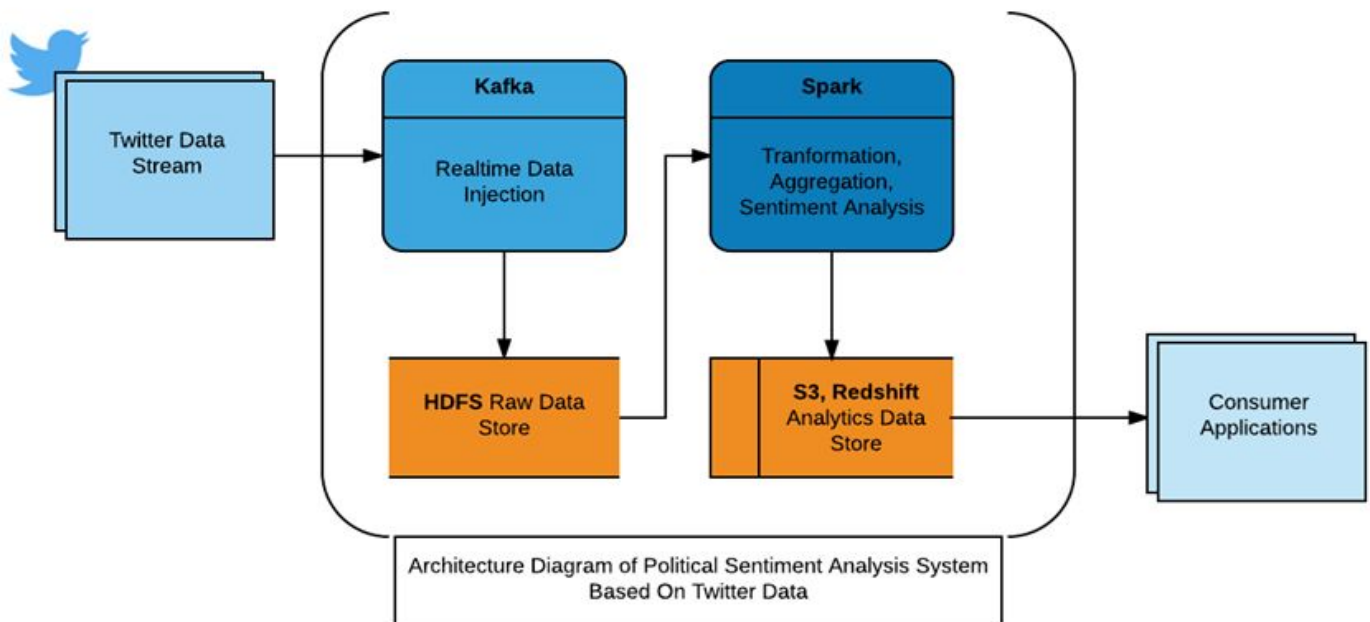


W205 Project Update
November, 2016

Shuang Chan
David Skarbrevik

Streaming Sentiment Analysis from Twitter

Review of Project:



The goal of this project is to analyze the sentiments of political twitter data in the U.S. as it relates to our current presidential election.

Here's what we've achieved so far:

- 1) Twitter Application Setup - authorizations
- 2) Kafka/Zookeeper server
- 3) HDFS schema and data transformation SQL files

Here's what is left:

- 1) Link Kafka and HDFS
- 2) Spark: cleaning/aggregating data + sentiment analysis
- 3) S3 or Redshift: storage of analytics data
- 4) Front facing "consumer" application - some way for people to access/use our data

Analysis of Project Thus Far:

In regards to the aim of the project, the 2016 U.S. Presidential Election has already passed (in an unexpected turnout). Though we can see if our analysis lines up with a Trump victory, it will be interesting to use Twitter's Streaming API to compare tweets from the time of the election to the tweets that will be pouring in over the next month about feelings for our new president (especially once he officially takes office). Perhaps we will see a shift toward sentiments uniting toward the president or perhaps sentiments will seem to become more polarized than ever.

From a technical standpoint, Kafka/Zookeeper has been a useful platform for streaming twitter data. The next step is to impose the SQL schema we have on the streaming Twitter data we're receiving. More concerning is the strategy we will use for sentiment analysis. We are still considering how this will be best achieved, but it seems there are many established ways of doing this so this should ultimately be easily implemented. We are still most excited and focused on the goal of getting streaming data from Twitter to a point of delivery for front facing internet users.

Here's a timeline of how we'll finish it in time:

Time frame	11/22/2016	11/27/2016	11/29/2016	12/4/2016	12/6/2016 (final deadline)
Goal	Stream data being stored in HDFS schema	Sentiment Analysis (spark)	Store analysis in S3	Front facing app / way for consumers to download	Presentation and final commit to github