**Spark Streaming and Visualization**

Implemented the following framework using Apache Spark

Streaming, Kafka (optional), Elastic, and Kibana. The framework performs SENTIMENT analysis of particular hash tags in twitter data in real-time. For example, we want to do the sentiment analysis for all the tweets for #trump, #coronavirus.

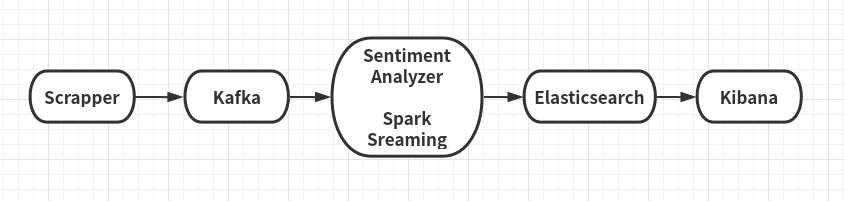


Figure: Sentiment analysis framework

The above framework has the following components:

**1. Scrapper (for python, but scala needs to produce same result)**

The scrapper will collect all tweets and sends them to Kafka for analytics. The scraper will be a standalone program written in PYTHON and should perform the followings:

a. Collected tweets in real-time with particular hash tags. For example, we

will collect all tweets with #trump, #coronavirus.

b. After filtering, we will send them to Kafka in case if you use Python.

c. Used Kafka API (producer) in the program

(https://kafka.apache.org/090/documentation.html#producerapi)

d. The scrapper program runs infinitely and takes hash tag as input parameter while running.

**2. Kafka (for Python)**

Ran Kafka Server with Zookeeper. Created a dedicated channel/topic for data transport

**3. Spark Streaming**

In Spark Streaming, created a Kafka consumer (for python, shown in the class for streaming) and periodically collected filtered tweets (required for both scala and python) from scrapper. For each hash tag, performed sentiment analysis

using Sentiment Analyzing tool (discussed below).

**3. Sentiment Analyzer**

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. It's also known as opinion mining, deriving the opinion or attitude of a speaker.

For example,

“President Donald Trump approaches his first big test this week from a

position of unusual weakness.” - has positive sentiment.

“Trump has the lowest standing in public opinion of any new president in

modern history.” - has neutral sentiment.

“Trump has displayed little interest in the policy itself, casting it as a

thankless chore to be done before getting to tax-cut legislation he values

more.” - has negative sentiment.

The above examples are taken from CNBC news:

http://www.cnbc.com/2017/03/22/trumps-first-big-test-comes-as-hes-in-an-

unusual-position-of-weakness.html

Used sentiment analyzer nltk(python) for sentiment analyzing.