

# Storm Twitter Streaming Architecture

## Directory Tree

```
| -tweetwordcount
| ---analysis
| -----finalresults.py
| -----histogram.py
| ---src
| -----bolts
| -----parse.py
| -----wordcount.py
| -----spouts
| -----credentials.py
| -----tweets.py
| ---topologies
| -----tweetwordcount.clj
```

## Application

This application can be used to track in real-time Twitter mentions for a marketing campaign. It can be extended to include tweet geolocation data and perhaps connected to a Tableau dashboard for real-time monitoring.

## Architecture and Dependencies

The `tweets` spout uses the Twitter API to extract tweets in English. This is then passed along to the `parse` bolt which splits the tweets into words, filtering out hash tags, RT, @, urls, leading and lagging punctuation, and any non-ascii characters. This is then passed to the `wordcount` bolt which keeps a counter that tracks how many times each word is seen. This is persisted in a postgresql database by the `wordcount` bolt.

This application requires Python 2.7.10 or more recent. Module requirements are described in the README.