Architecture of tweetwordcount

Tweetwordcount is a spout and bolt stream parse application that uses python code to grab live tweets via a spout, parse words out of the live tweet via the first bolt and track the frequency that word appears across live tweets via a second bolt while the application is run. Below is the topology of the application.
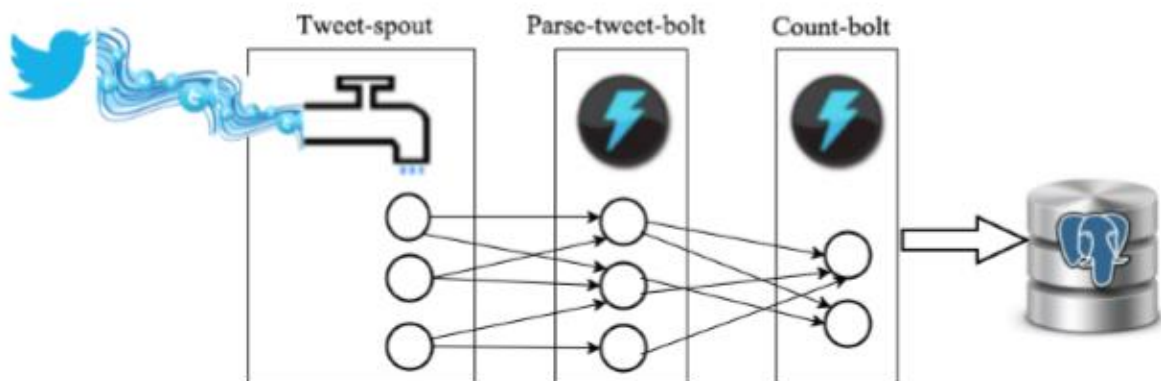

Figure 1: Application Topology

The application stores the word and its counts within the live tweets in a postgres database. This database cane be queries via finalresults.py either by passing in a word of interest or no words to get the counts of all word in the database. Below is an example of how to get this information from the database by searching for the word "wait".

>> python finalresults.py wait

Histogram.py shows words between specified counts. Below is an example of how to get all words between with frequencies between 3 and 8.

>> python histogram.py 3 8

Below is the information of the posgres database:

Database Name: tcount

Table Name: tweetwordcount

Table Columns: Word – Text; Count - Int

Below is the file structure of the application and each file's description

File	Description	Location
parse.py	Bolt that parses the words provided from the tweets.py spout	exercise_2\tweetwordcount\src\bolts
wordcount.py	Bolt that stores the word and frequency of that word within a postgres database 'tcount'	exercise_2\tweetwordcount\src\bolts
tweets.py	Spout that connects to twitter and streams live tweets	exercise_2\tweetwordcount\src\spouts
tweetwordcount.clj	Topology of the stream parse tweetwordcount application	exercise_2\tweetwordcount\topologies
tweetwordcount.txt	Virtual environment information	exercise_2\tweetwordcount\virtualenvs
config.json	Topology configuration information	exercise_2\tweetwordcount
fabfile.py	Fabric tasks	exercise_2\tweetwordcount
finalresults.py	Python file to view word frequencies by passing in a word of interest	exercise_2\tweetwordcount
histogram.py	Python file to view a histogram of words and their frequencies between a specified range of frequencies	exercise_2\tweetwordcount
project.clj	Project file	exercise_2\tweetwordcount
README.txt	Text file that has very basic instructions on how to run the application	exercise_2\tweetwordcount
tasks.py	Optional file	exercise_2\tweetwordcount