Documentation of Architecture of Tweetcount Storm Application

Application Description

The Tweetcount application streams tweets from Twitter.com, passes them into a Storm architecture, which then parses and counts the words in the tweets and saves the total counts to a PostgreSQL database.

Architecture Design

The architectural design consists of one spout that received information from Twitter's Stream API and two bolts, the first of which parses the tweet text and the second that counts the words and updates a PostreSQL database.

File Structure and Description

The file structure (only pertinent files shown) is as follows:

```
exercise_2/
        finalresults.py
        histogram.py
        Tweetwordcount/
                src/
                        bolts/
                                parse.py
                                wordcount.py
                        spouts/
                                tweets.py
                topologies/
                        wordcount.clj
```

File name	Description
finalresults.py	A python script that takes a optional command line argument as a word, returns the count of that word in the Twitter stream if no argument is provided then it returns all the words in the database with their counts
histogram.py	A python script that take a range e.g. "3,8" as a command line argument and returns the words and their counts that are between the range
parse.py	Bolt that parses the tweet text
wordcount.py	Bolt that given an input counts the words in text
tweets.py	Spout that gets the tweets
wordcount.clj	Topology that brings together all of the above scripts

Running the Application

To run the application ensure that the PostgreSQL database is up and running with a user “postgres” and password “siddiqi" furthermore, the settings in /var/lib/pgsql/data/pg_hba.conf should be all set to “password” instead of “ident” after changing the settings please run /etc/init.d/postgresql reload. Finally log into the PostgreSQL database as user prosgres and execute the following commands:
Create database tcount;
\c tcount;
Create table tweetwordcount (word TEXT PRIMARY KEY NOT NULL, count INT NOT NULL);
Exit;