Extweetwordcount Application

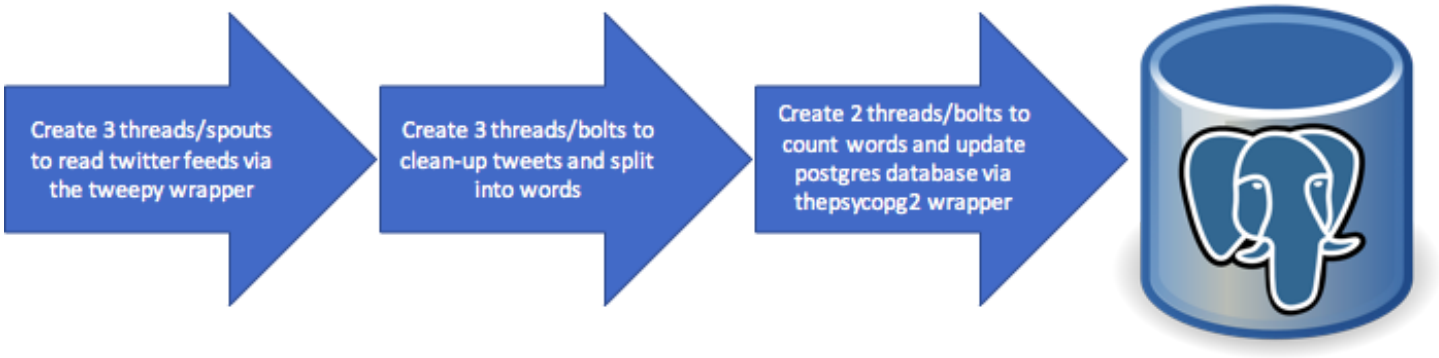
Directory / file structure:

File / directory name	Description
Plot.png	Barchart for top 20 words by count
README.md	Read me file for github
Readme.txt	Text version of the Read me file (per exercise requirements)
Twittercredentials.py	Python file for application / user credentials for twitter
Twittercredentials.pyc	Byte code compiled version of the Twitter credentials python file
_build	Internal directory for streamparse
_resources	Internal directory for streamparse
config.json	Configuration file for streamparse
fabfile.py	Pre and post topology submission python callback functions
finalresults.py	Final results count (per exercise requirements)
hello-stream-twitter.py	Sample twitter streaming application
histogram.py	Histogram (per exercise requirements)
logs	Log directory for streamparse output
project.clj	Clojure file for the streamparse project
psycopg-sample.py	Sample python - postgres database query program
screenshots	Screenshots directory (per exercise requirements)
src	Source code directory for spouts and bolts
spouts	Source code directory for spouts
tweets.py	Python code reading tweets and pushing to bolts
bolts	Source code directory for bolts
parse.py	Python code for parsing tweets and stripping individual words
wordcount.py	Python code for counting words and updating the database
tasks.py	Pre and post topology submission python callback functions
topologies	Directory for the bolt / spout topology file
virtualenvs	Internal directory for streamparse

Application idea

The application idea is fairly regimented by the exercise. 3 separate (thread) instances of the tweet-spout read data from twitter via the Twitter API and the tweepy Python wrapper. These 3 spouts are connected to 3 bolts called parse-tweet-bolt, which parse, clean up and split the tweets into individual words. These words are passed to 2 instances of the word-count bolt, which creates and updates the postgres database.

Description of the architecture



File dependencies

Access to twitter application depends on proper credentials for the Twitter account. These credentials reside in the Twittercredentials.py file.

How to run the application

1. Modify the Twittercredentials.py file to update Twitter API credentials. (the credentials will be sent via email).
2. type: `cd extweetwordcount`
3. type: `sparse run`
4. Either after you stop after a while or in a separate window, type: `python histgram.py 10,50`
5. Either after you stop after a while or in a separate window, type: `python finalresults.py`