Twitter Application

Goal: The goal of this Twitter application is to stream live tweets into a Postgres database that saves each word and the occurrence of each word within 120 seconds.

Architecture description: With Apache storm at the core, streamparse is utilized to retrieve the data from Twitter's API using python scripting as a means of communication.

Main Directory: ~/exercise_2/EX2Tweetwordcount

File Dependencies: Before running the first time, the spout and bolt files are key in this application. After running using the "sparse run" command, the files are then copied into the _resources folder and the program will always look there for the corresponding spout and bolt files. In addition to the spout and bolt files, the topology file is the file that connects the spouts and bolts together and makes the application run.

Important files:

Topology file: ~ /exercise_2/EX2Tweetwordcount/topologies/tweetwordcount.clj

Spout file: ~ /exercise_2/EX2Tweetwordcount/src/spouts/tweets.py

Bolt file: ~ /exercise_2/EX2Tweetwordcount/src/bolts/parse.py

~ /exercise_2/EX2Tweetwordcount/src/bolts/wordcount.py

*****Important Running Information*****

It is very important to install streamparse 2.0.1 or lower! If a newer version is installed then this will cause issues with the programs that streamparse is dependent upon.

Use this command to install streamparse 2.0.1:

```
pip install https://pypi.python.org/packages/source/s/streamparse/streamparse-2.0.1.tar.gz
```

Running instructions

1. sudo sparse run (This will fail saying the program could not find a log file)

2. sudo lein run -m streamparse.commands.run/-main topologies/tweetwordcount.clj -t 120 --option 'topology.workers=2' --option 'topology.acker.executors=2' --option 'streamparse.log.path=/<pathtosource>/EX2Tweetwordcount/logs' --option 'streamparse.log.level="debug"'

For some reason, following the instructions given does not produce a working program. "sparse run" is needed to generate the _resource folder but for some reason it will fail due to it not finding a log file. The second command is actually the command that pops up after "sparse run" is entered. The only difference is that the "**streamparse.log.path=/<pathtosource>/EX2Tweetwordcount/logs**" has quotes around it, so it cannot be found. After removing the added quotes and entering the command manually, the program works!