Twitter wordcount using Apache Storm

This exercise is about using Apache Storm to parse a continuous stream of twitter feeds and count the word occurrences, and record them in a postgres database. It uses streamparse, a python library that simplifies building storm applications. The application has the following folder structure as required by streamparse:

- Topologies: this folder contains the tolpology. Here we define that we want 3 spouts (Data sources), 3 parse-tweet bolts (to break up the twitter steam into individual words and a count of 1), and 2 WordCounter bolts (to count the occurrences of each word). The parse-tweet bolts are akin to mappers and the WordCounter bolts are akin to reducers.
- Src/spouts: contains source code for the spouts. In this application we use a single spout class. The spout connects to a twitter feed and outputs the contents of a feed.
- Src/bolts: contains source code for the bolts. In this application we use 2 bolt classes. The parse-tweet bolts read the output of the spouts, and pass the words on to the WordCounter bolt. The WordCounter bolt counts words and outputs them to the log (so they are visible on the screen) and also updates word counts in the postgres db.