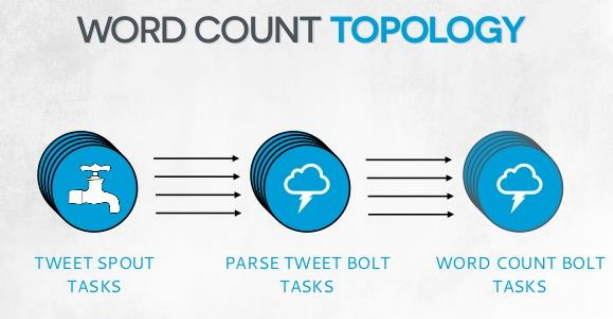
Twitter word count Storm application



The tweetwordcount application consists of a single spout that sends in a stream of tweets.
There are 2 bolts in the topology, the first one parses the tweet stream into words and the second one counts the words. The second bolt uses a postgres table Tweetwordcount to store the words and word frequency. It checks the database first if the word exists. If the word exists in the database, then the count associated with the word is incremented, else the word is added to the database with a count of 1. It also emits the word and its count in a live manner on the screen.

One area of improvement that I can think of is to do with opening and closing of connection with postgres database with each bolt operation. That is one area that could be optimized.

The analysis.py uses the table Tweetwordcount to get a list of top 20 words and creates a histogram of that in plots.png as shown below.

I am using matplotlib to generate the plot and that should be installed in the env.

The plot is shown below as well.