

Directory Name: exercise\_2

This document shows the files and folders within exercise\_2. If there is no explanation for an item, it is supporting infrastructure.

#### Tweetwordcount

- README.md

- \_resources

- Fabfile.py

- Project.clj

- Tasks.py

- Virtualenvs

- \_build

- Config.json

- Logs

- Src

- #### Bolts

- Parse.py - parses the tweets emitted by the spout and extracts words

- Wordcount.py - extracts words and keeps count in postgres database

- Tcount

- \_\_init\_\_.py

- #### Spouts

- \_\_init\_\_.py

- Tweets.py - pulls tweets from twitter streaming API

- #### Topologies

- Tweetwordcount.clj - unifies bolts and streams within the apache storm framework.

Finalresults.py - python script that should be run AFTER tweetwordcount sparse has started.

Shows the count for every word. Optional argument reduces query results to single word.

Histogram.py - python script that should be run AFTER tweetwordcount sparse has started.

Displays all the words with counts between the two integer values (which are two required arguments)

Readme.txt - Provides detailed step by step instructions

Screenshots - Screenshots at various points for the application run:

- Screenshot\_finalresults - results from finalresults.py with no arguments

- Screenshot\_runningstream - display of the stream while it is running

- Screenshot\_histogram - results from histogram.py

Plot.png - Bar chart showing the top 20 words in the twitter stream

Architecture.pdf - This document