

Architecture

What does this thing do?

This application:

- Reads a live Twitter feed
- Parses out each word from each Tweet
- Counts the occurrences of each word (for duration of the live stream)
- Send the list of words and their counts to a postgres database

How does it do all these things?

See technologies used below in **green**:

- Reads a live Twitter feed **with Tweepy**
- Links all tasks together **with Stream Parses' spouts and bolts**
- Parses out each word from each Tweet **with Python**
- Counts the occurrences of each word (for duration of the live stream) **with Python**
- Send the list of words and their counts to a postgres database **with psycopg2, a postgres adaptor for Python**

How do I run this thing?

Steps to get it working:

1. **Download the necessary packages to your machine:**
 1. Hadoop, postgres, etc
 2. Streamparse
 3. Tweepy
 4. PsycPG
2. **Create Twitter develop profile and get access tokens**
3. **Download the file structure to your machine**
4. **Go into Twittercredentials.py and update Twitter credentials**
5. **Go into exttweetwordcount > src > bolts > wordcount.py and update Twitter credentials**

6. To run main program enter:

1. “sparse run” (while in exttweetwordcount folder)
2. Hit Control C to stop program at any time

7. To run finalresults.py enter:

1. “python finalresults.py <optional word you’re interested in>

8. To run histogram.py enter:

1. “python histogram.py <number1>, <number2>

What is the file structure?

Here is a map of the application:

