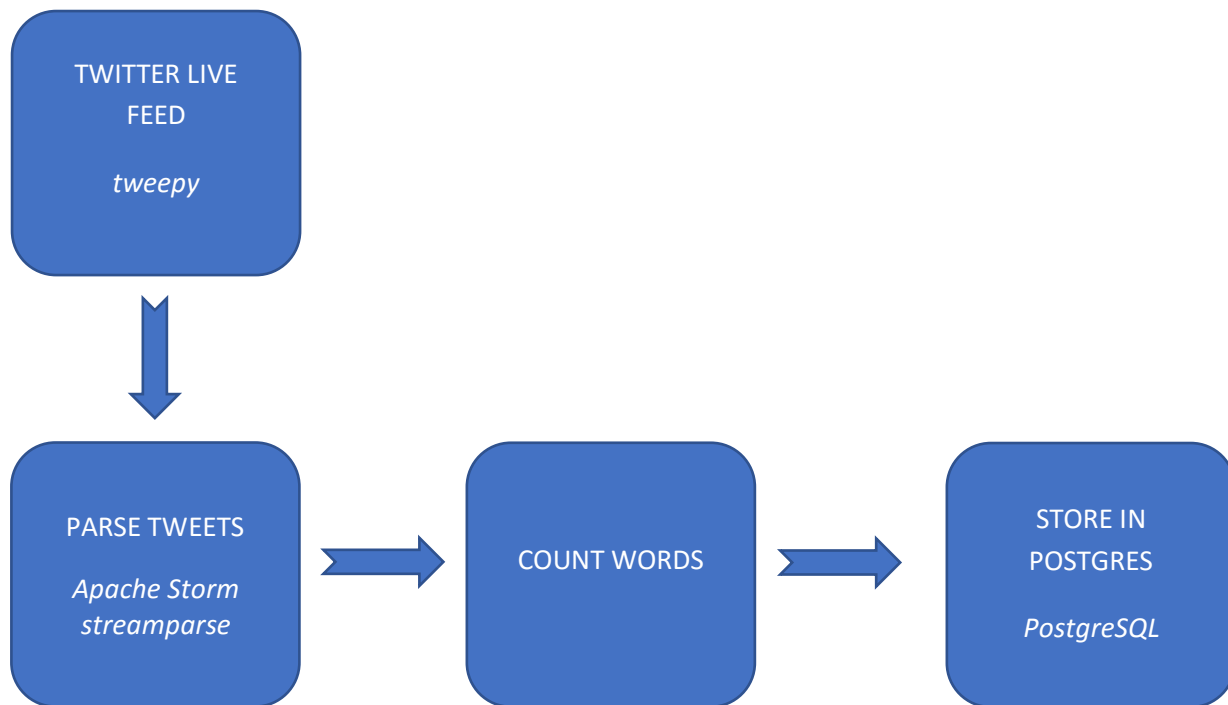


## Overview of Architecture



*italicized are the instrumental packages used at each stage*

## Application Summary

Our application parses tweets from Twitter and stores each individual word from every tweet into our Postgres database. We then increment a counter for each word as newly arrived tweets contain them.

## Directory Layout:

- `exttweetwordcount` contains the bulk of our application
  - `topologies` folder contains our clojure file with the layout for apache storm to run
  - `src` contains our bolts and spouts
    - `spouts` contains file to stream live tweets
    - `bolts` parses the tweets and counts the words
- `finalresults.py` is an executable python script that displays all words and their count frequency
  - if passed an argument of any word, it will return the count of that word in our database
- `histogram.py` is an executable python script that returns all words in a count range inbetween two integers

### **File Dependencies**

Out of the Ordinary Python Packages Required

- Tweepy, Apache Storm, psycopg2

### **Instructions on Running Application**

Clone entire folder 'Exercise 2' from github repo

[https://github.com/hxu23/w205\\_2017\\_fall](https://github.com/hxu23/w205_2017_fall)

in the extweetwordcount folder, type sparse run