

W205 Exercise 2

Twitter Stream Architecture

Linh Tran

Twitter Streaming API Intro

Objectives

Understand distribution of tweeted words in real time by:

1. Streaming live tweets from Twitter API
2. Parse tweets into words
3. Count occurrences of words
4. Store words and counts into a database for query

Getting Started

1. Review README.txt for step by step instructions to set up application
2. Key Files
 - a. `exercise_2/psycopg-setup.py` - *script to set up database and table*
 - b. `exercise_2/histogram.py` - *show word and counts by bin endpoints*
 - c. `exercise_2/finalresults.py` - *show word count for specific word or all words*
 - d. `exercise_2/exttweetwordcount/topologies/exttweetwordcount.clj` - *topology file*
 - e. `exercise_2/exttweetwordcount/src/spouts/tweets.py` - *spout*
 - f. `exercise_2/exttweetwordcount/src/bolt/parse.py` - *bolt*
 - g. `exercise_2/exttweetwordcount/src/bolt/wordcount.py` - *bolt*

Architecture

