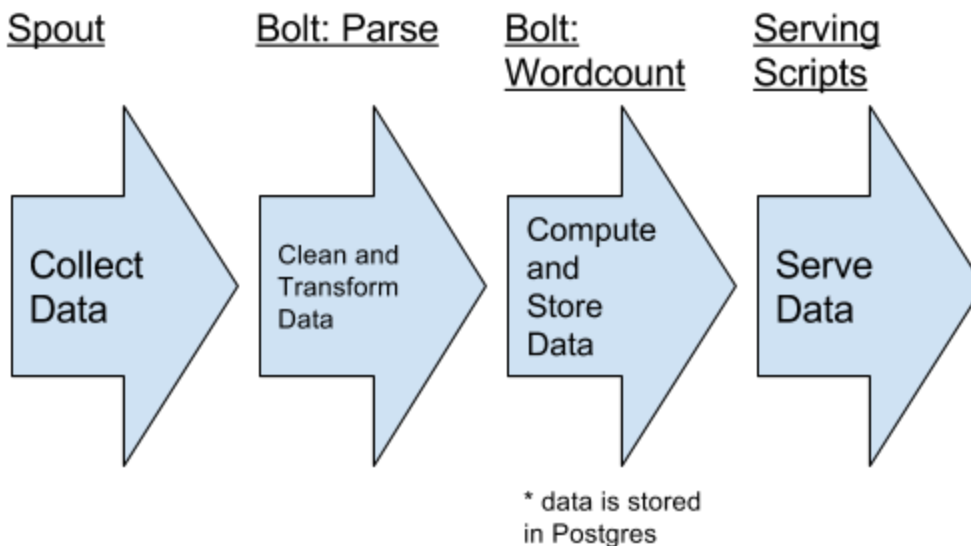**Application idea**

Collect Twitter tweet data, parse tweets into words, and count the words to understand what topics users are discussing.

Directory and file structure
- w205_2017_fall/exercise_2/
  - Topology (extweetwordcount/topologies/)
    - Spouts and Bolts (extweetwordcount/src/)
      - Spout: (spouts/tweets.py)
        - Collect Tweet data
        - Contains Twitter credentials
      - Bolts: (bolts/)
        - Parse (parse.py)
          - Splits Tweet data into words, filters out the hash tags, "RT", "@" and "urls"
        - Count (wordcount.py)
          - Updates word count in Postgres table
  - Serving scripts:
    - Display count of all words, or of a specific word (finalresults.py)
    - Display all words within a count range (histogram.py)
  - Instructions to run the application (README.txt)

Architecture

Spout — Collect Data

Bolt: Parse — Clean and Transform Data

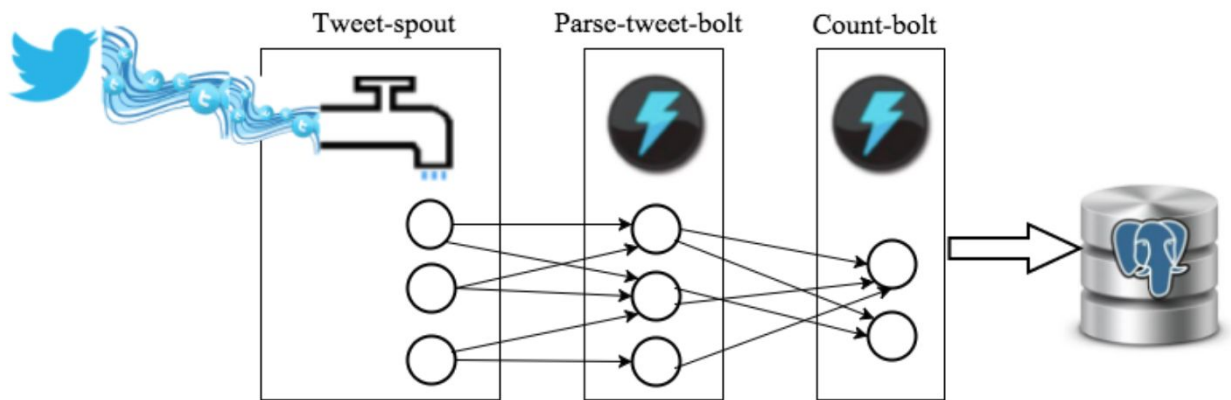Bolt: Wordcount — Compute and Store Data

Serving Scripts — Serve Data

* data is stored in Postgres

Topology

Figure 1: Application Topology