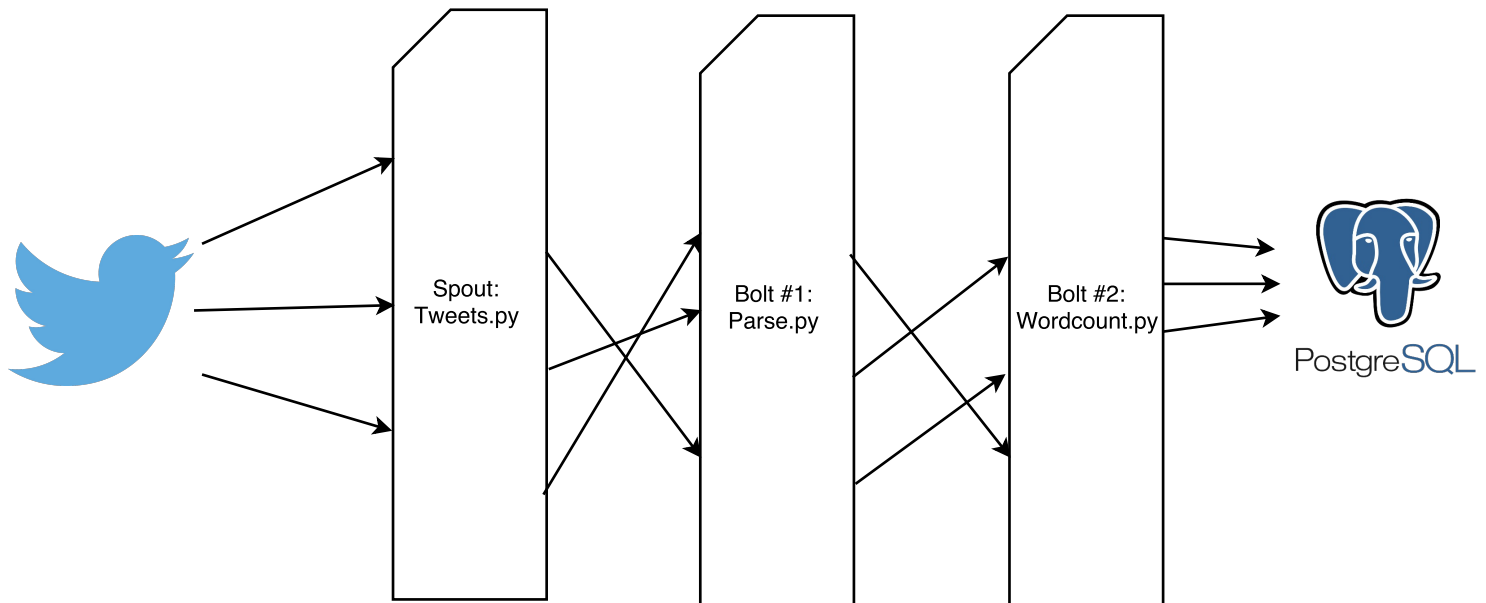


Twitter Stream Architecture



Directory/File System Structure

| FOLDERS |
|---------------------|
| ✓ exercise2 |
| ✓ EX2Tweetwordcount |
| > _build |
| > _resources |
| > dev-resources |
| > logs |
| ✓ src |
| ✓ bolts |
| __init__.py |
| parse.py |
| wordcount.py |
| ✓ spouts |
| __init__.py |
| tweets.py |
| > test |
| ✓ topologies |
| tweetwordcount.clj |
| > virtualenvs |
| .gitignore |
| config.json |
| fabfile.py |
| finalresults.py |
| histogram.py |
| project.clj |
| README.md |
| tasks.py |
| tcount.sql |
| > screenshots |
| .gitignore |
| Plot.png |
| plot.R |
| README.md |

| Notable Files and Descriptions | |
|--------------------------------|---|
| File | Description |
| tweets.py | Listens to twitter feed and then emits the feed to parse bolt. Uses tweepy python library and user api credentials to access the twitter stream. |
| parse.py | Takes in emitted tweets from tweets.py and splits them into separate words, then parses out any special characters that wouldn't be counted as words. |
| wordcount.py | Takes in emitted words from parse.py and counts the frequency of each word; counting is caps-sensitive. Then either inserts or updates words and counts into postgres db. |
| finalresults.py | If no arguments inputted when run, outputs a list of all words and counts in the db; if words are inputted, displays the counts of that word in the db. |
| histogram.py | Specify two numbers; running this will produce a list of words that have a count of at least the first number and less than the second number. |

For more information, see README.md in github repository.