Architecture

Directory and file structure

EXTWOTweetwordcount

Src
- o Bolts
  - ▪ Parse.py
  - ▪ Wordcount.py
- o Spouts
  - ▪ Tweets.py

Topologies
- ▪ EX2tweetwordcount.py

createDBAndTable.py
finalresults.py
histogram.py

Description of Architecture

Spout:
Tweepy Twitter API, it streams any tweets that filtered out special characters

Parse:
Extracts words by splitting the tweets. And each word is checked if it starts with special characters, if not, it will be added to valid words array. (#, @, rt, http, ("\"?><,\'.:;!~[]()&%$*/=+-_^$\\"). This is to avoid counting the retweets/mention/hashtag etc.

Wordcount:
Interacts with DB and table about word count.

Postgres data storage:

> **Tcount**: database

> **Tweetwordcount**: table with two columns: word (text, primary key), count (int)

File dependencies

python 2.7
Streamparse
Tweepy
psycopg2
Apache Storm, Postgres

Execution Steps (also in README.txt)

1. run: sparse quickstart EXTWOTweetwordcount and update corresponding files with the ones in this Git Repo (API credentials contained in tweets.py)

2. create db and table

       mount /dev/xvdf /data
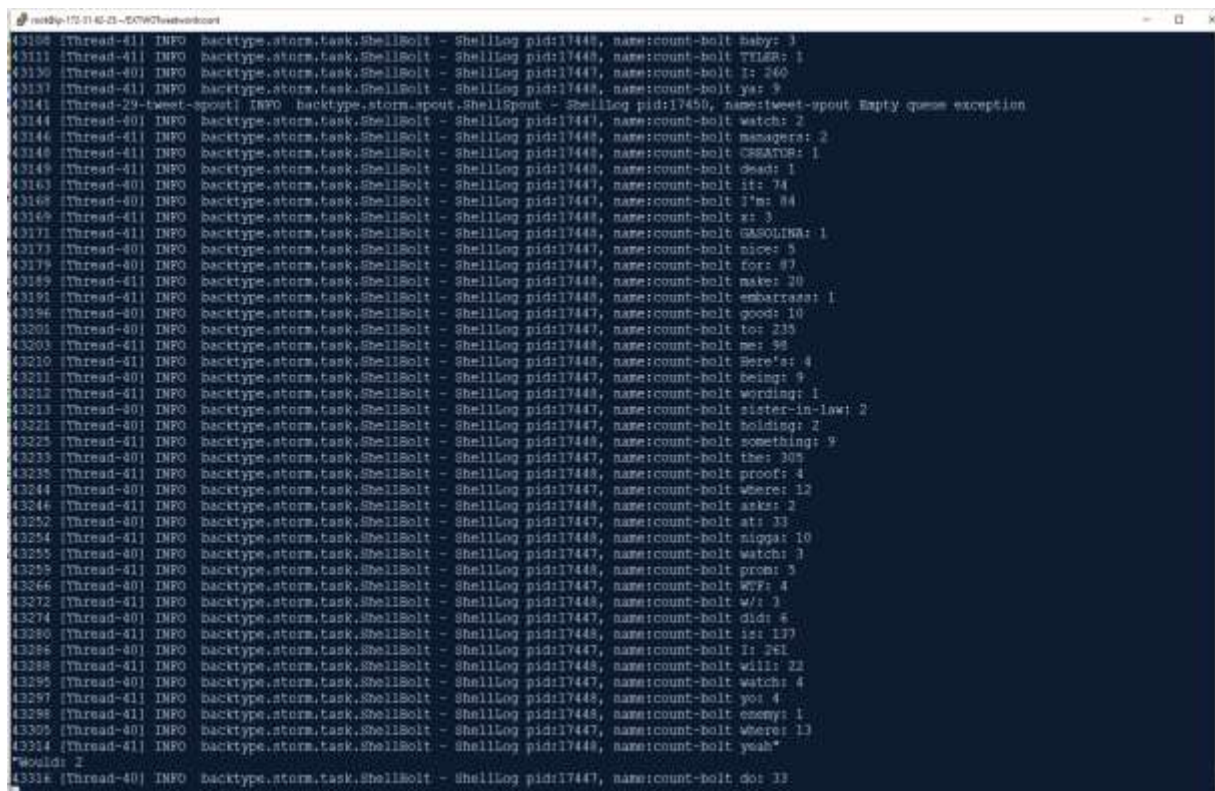
       cd /data

       ./start_postgres.sh

       cd ~

       cd EXTWOTweetwordcount

       python ./createDBAndTable.py

3. under EXTWOTweetwordcount run: sparse run

4. python finalresults.py <optional inputword>

5. python python histogram.py <lowlimit> <upperlimit>

Screenshot (picture file in Screenshots folder):

```
yahoo , 1
yall , 3
yeah , 3
yeah"
"Would , 1
year , 19
year! , 1
years , 24
yes , 11
yes! , 1
yes?"
"Fuck , 1
yesterday , 3
yesterday!!!!!!!!! , 1
yet , 15
yet!!! , 1
yet...jury , 1
yo , 10
yo!! , 1
yoga , 1
you , 732
you! , 9
you!
#cologop
#DumpTheGOP
#Trump2016 , 1
you'd , 5
you'll , 7
you're , 57
you've , 5
you.
Good , 1
young , 13
young! , 1
your , 151
yours , 3
yourself , 11
youuuu , 1
yu , 1
z , 1
zaara , 1
zach , 1
zone
Priceless , 1
zoo , 1
(@nordstrombeth) , 1
| , 12
(py27environment)[root@ip-172-31-62-25 EXTWOTweetwordcount]#
```

```
reach :  11
old :   11
stuff :   11
it! :   11
Get :   11
bring :  11
sorry :   11
final :   11
Sanders :  11
vote :   11
Trump :   11
They :   11
remember :   11
Who :   11
yes :   11
ok :   11
text :   11
tweet :   11
Wait :   11
anymore :   11
kind :   11
second :   11
win :   11
thinking :   11
seem :   11
of :   11
everyone :   11
y'all :   11
One :   11
tbh :   11
Oh :   11
tho :   11
hours :   11
read :   10
else :   10
episode :   10
fast :   10
Stop :   10
visit :   10
posted :   10
holding :   10
What's :   10
having :   10
Boy :   10
yo :   10
haha :   10
live :   10
(py27environment)[root@ip-172-31-62-25 EXTWOTweetwordcount]#
```