

# Architecture

## Directory and file structure

### EXTWOTweetwordcount

#### Src

- Bolts
  - Parse.py
  - Wordcount.py
- 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

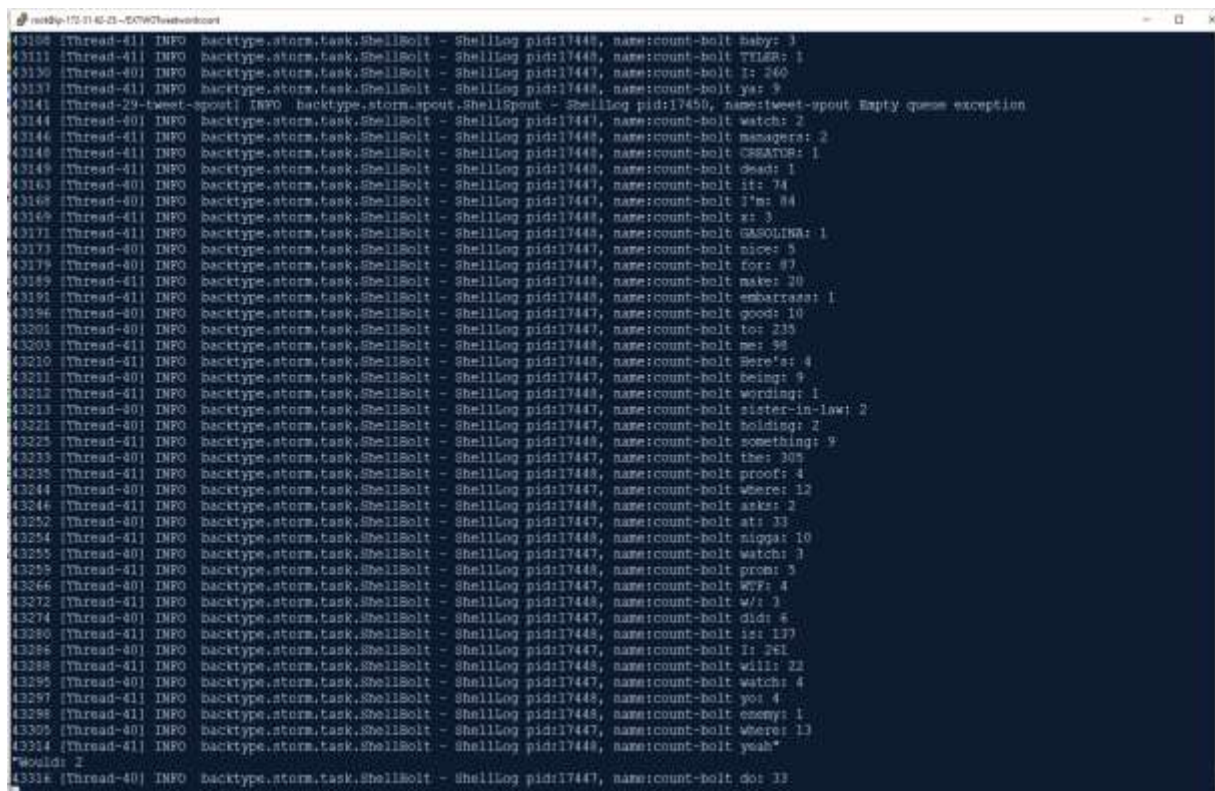
./start_postgres.sh

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):



```
root@ip-172-31-42-23:~# EXTWOTweetwordcount
43108 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt baby: 3
43111 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt TILER: 1
43113 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt I: 240
43117 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt ya: 9
43141 [Thread-29] INFO backtype.storm.task.ShellBolt - ShellLog pid:17450, name:tweet-spout Empty queue exception
43144 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt watch: 2
43146 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt managers: 2
43148 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt CREATOR: 1
43149 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt dead: 1
43163 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt it: 74
43168 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt m: 84
43169 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt s: 3
43171 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt CASOLINA: 1
43173 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt nice: 5
43179 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt for: 87
43189 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt make: 20
43191 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt embarrass: 1
43196 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt good: 10
43201 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt to: 235
43203 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt me: 58
43210 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt Here's: 4
43211 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt being: 9
43212 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt wording: 1
43213 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt sister-in-law: 2
43221 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt holding: 2
43225 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt something: 9
43233 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt the: 305
43235 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt proof: 4
43244 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt where: 12
43246 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt asks: 2
43252 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt at: 33
43254 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt nigga: 10
43255 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt watch: 3
43259 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt prom: 3
43266 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt WTF: 4
43272 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt w: 3
43274 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt did: 6
43280 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt is: 137
43286 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt I: 261
43288 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt will: 22
43295 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt watch: 4
43297 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt you: 4
43298 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt enemy: 1
43305 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt where: 13
43314 [Thread-41] INFO backtype.storm.task.ShellBolt - ShellLog pid:17448, name:count-bolt yeah*
*would: 2
43318 [Thread-40] INFO backtype.storm.task.ShellBolt - ShellLog pid:17447, name:count-bolt do: 33
```

root@ip-172-31-62-25:~/EXTWOTweetwordcount

```
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!
#collogop
#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]#
```

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
thh : 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]#
```