instruction to run for ex2

dan wald

1. on aws, mount drive if necessary
   1. mount -t ext4 /dev/xvdf /data
2. make sure packages are installed
   1. pip install psycopg2
   2. pip install tweepy
3. login to postgres
   1. /data/start\_postgres.sh
4. change user to w205 for best practice
   1. su – w205
5. start a folder directory for postgres in your desired location
   1. sparse quickstart extweetwordcount
6. clone the git directory
   1. git clone <https://github.com/dswald/w205_wald.git>
   2. git checkout -b hw\_dev
   3. git pull origin hw\_dev
7. navigate to the git repository ./w205/ex2/ex2tweetwordcount/topologies
   1. remove file “ex2tweetwordcount.clj” if it exists
8. navigate to git repository ./w205/ex2/ex2tweetwordcount/
   1. sparse run #streaming data
   2. wait for a while to populate data
9. tools exploration:
   1. all tools are at to top level of ./w205/ex2/ex2tweetwordcount/
   2. you need to look at the following files in addition to the twitter storm applications:
      1. finalresults.py
      2. histogram.py
      3. barchart.py
         1. # note that plotting within python 2 is disabled, so i copied this code over to a jupyter notebook running python 3, populating the data with most\_used = records[-1:-21:-1]