

Reference:

<http://docs.aws.amazon.com/gettingstarted/latest/emr/getting-started-emr-sentiment-create-twitter-account.html>

<https://gist.github.com/BhavdeepSethi/50f41c6db7b02e2ca94f>

1. Get the collector program following steps from reference. Simply a listen and save program.

Change the parameter of count(originally it is 500), set the key words to be love. Next put the data to my bucket.

2. Run EMR, the reason it fails is





1.because I don't have access the the file in tutorial. So I use "\$aws s3 cp "to retrieve files.

2. The mapper is not corrected.







3. Use the mapper program found on github, s3 copy the files to my own bucket. It success. We run the EMR on four collections. With more than 20000 entries, the key word is love.

```
[ec2-user@ip-172-31-52-231 sentiment]$ ls
collector.py      tweets.May14-1239.txt  tweets.May14-1807.txt  Creation time
deleted_tweets.txt tweets.May14-1753.txt  twitterparams.py
twaiter.py        tweets.May14-1759.txt  twitterparams.pyc
twaiter.pyc       tweets.May14-1801.txt  2015-05-14 12:
[ec2-user@ip-172-31-52-231 sentiment]$ s3cmd put tweets.May14-1753.txt s3://el27
56/input/
tweets.May14-1753.txt -> s3://el2756/input/tweets.May14-1753.txt [1 of 1]
68603 of 68603 100% in 0s 821.70 kB/s done
[ec2-user@ip-172-31-52-231 sentiment]$ s3cmd put tweets.May14-1759.txt s3://el27
56/input/
tweets.May14-1759.txt -> s3://el2756/input/tweets.May14-1759.txt [1 of 1]
55166 of 55166 100% in 0s 585.42 kB/s done
[ec2-user@ip-172-31-52-231 sentiment]$ s3cmd put tweets.May14-1801.txt s3://el27
56/input/
tweets.May14-1801.txt -> s3://el2756/input/tweets.May14-1801.txt [1 of 1]
69340 of 69340 100% in 0s 664.55 kB/s done
[ec2-user@ip-172-31-52-231 sentiment]$
```

**AWS** ▾**Services** ▾**Edit** ▾**Upload****Create Folder****Actions** ▾[All Buckets](#) / [el2756](#) / [input](#)

		Name
<input type="checkbox"/>		tweets.May14-1753.txt
<input type="checkbox"/>		tweets.May14-1759.txt
<input type="checkbox"/>		tweets.May14-1801.txt
<input type="checkbox"/>		tweets.May14-1807.txt

**AWS** ▾**Services** ▾**Edit** ▾**Upload****Create Folder****Actions** ▾**All Buckets** / **el2756**

		Name
<input type="checkbox"/>		classifier.p
<input type="checkbox"/>		config-nltk.sh
<input type="checkbox"/>		elasticmapreduce
<input type="checkbox"/>		input
<input type="checkbox"/>		mapper
<input type="checkbox"/>		output

AWS

Services

Edit

Elastic MapReduce

Cluster List

Create cluster

View details

Clone

Terminate

Filter:









All clusters

Filter clusters ...

8 clusters (all loaded)

		Name	ID	Status	Created
<input type="checkbox"/>	▶	My cluster	j-3RIBNNJJ3FPEO	Terminated All steps completed	2015
<input type="checkbox"/>	▶	My cluster	j-NC0RMJ5NNU92	Terminated	2015
<input type="checkbox"/>	▶	My cluster	j-21JUWC2CRX8VE	Terminated All steps completed	2015
<input type="checkbox"/>	▶	My cluster	j-117IJD7KQQNOY	Terminated User request	2015
<input type="checkbox"/>	▶	● My cluster	j-3UQRSY3L3LPHK	Terminated with errors Bootstrap failure	2015
<input type="checkbox"/>	▶	● My cluster	j-3UFRN4G1QCOAA	Terminated with errors Bootstrap failure	2015
<input type="checkbox"/>	▶	● sentiment-tutorial	j-1UAR0LI2Y6NF5	Terminated with errors Bootstrap failure	2015
<input type="checkbox"/>	▶	Senti-tutorial	j-3SG9Y6W60QIP7	Terminated	2015

**AWS** ▾**Services** ▾**Edit** ▾**Upload****Create Folder****Actions** ▾**All Buckets** / **el2756** / **output**

	Name	Storage Class	Size
<input type="checkbox"/>	 _SUCCESS	Standard	0 bytes
<input type="checkbox"/>	 part-00000	Standard	0 bytes
<input type="checkbox"/>	 part-00001	Standard	18 bytes
<input type="checkbox"/>	 part-00002	Standard	37 bytes
<input type="checkbox"/>	 part-00003	Standard	38 bytes
<input type="checkbox"/>	 part-00004	Standard	60 bytes
<input type="checkbox"/>	 part-00005	Standard	42 bytes
<input type="checkbox"/>	 part-00006	Standard	20 bytes

```
Ryan:~ ryan$ cd ssh -i ~/.aws/Enrui.pem ec2-user@52.4.201.0
-bash: cd: ssh: No such file or directory
Ryan:~ ryan$ ssh -i ~/.aws/Enrui.pem ec2-user@52.4.201.0
Last login: Thu May 14 15:48:36 2015 from cpe-67-243-152-204.nyc.res.rr.com

Welcome to Amazon Linux AMI
Kernel: 3.10.0-123.el7.x86_64
Architecture: x86_64
Release: 2015.03
Distro: Amazon Linux
System: Amazon Linux AMI
System Arch: x86_64
System Arch: x86_64

https://aws.amazon.com/amazon-linux-ami/2013.09-release-notes/
Amazon Linux version 2015.03 is available.
[ec2-user@ip-172-31-52-231 ~]$ ls
classifier.p#classifier.p sentiment
[ec2-user@ip-172-31-52-231 ~]$ mv classifier.p#classifier.p classifier.p
[ec2-user@ip-172-31-52-231 ~]$ ls
classifier.p sentiment
[ec2-user@ip-172-31-52-231 ~]$ vi classifier.p
[ec2-user@ip-172-31-52-231 ~]$
[ec2-user@ip-172-31-52-231 ~]$
[ec2-user@ip-172-31-52-231 ~]$
[ec2-user@ip-172-31-52-231 ~]$
[ec2-user@ip-172-31-52-231 ~]$
[ec2-user@ip-172-31-52-231 ~]$ ls
classifier.p sentiment
[ec2-user@ip-172-31-52-231 ~]$
```