

PROJECT REPORT(BIG DATA AND HADOOP)

Project Name - OPINION MINING USING HADOOP ECO-SYSTEM

Source of Data - > Social Network (Twitter Data Streaming)

Data Type -> JSON - Semi Structured Data

Framework -> Hadoop

Tool -> PIG & Flume

POC -> Map-Reduce

Data Dictionary -> AFINN

- **Objective:-**

1. In this project, we will showcase how to perform Sentiment Analysis on Twitter data using Pig. To begin with, we will be collecting real-time tweets from Twitter using Flume. With the help of AFINN dictionary we can find positive as well as negative sentiments or opinion of people from the data which we have dumped into our HDFS/FLUME. We can prioritize our own customized words in AFINN dictionary as per the need of Client and give the words ratings(-5 to +5) and we can compare this words with the data we have dumped in flume using pig commands.

2. We need most advance concepts of PIG & MAP-REDUCE for getting into this project.

3. The data from Twitter is in 'Json' format, so a Pig JsonLoader is required to load the data into Pig. You need to download the required jars for the JsonLoader from the below link:-

<https://drive.google.com/drive/u/0/folders/0ByJLBTmJojjzUXdEOUpVQ1haNG8>

****NOTE-(JSON is Java Script Object Notation. It is a syntax for storing and exchanging data)**

- **Name of Pig JSON LOADER are:-**

- 1- elephant-bird-hadoop-compat-4.1.jar
- 2- elephant-bird-pig-4.1.jar
- 3- json-simple-1.1.1.jar

- **Execution Module:-**

- 1- streaming twitter data using flume(flume 1.4.0)
- 2- Invoke PIG grunt shell in map-reduce mode.
- 3- Register all the PIG JSON LOADER with PIG.
- 4- Load AFINN dictionary from HDFS to PIG bag.

STEP-1

Register the downloaded jars in pig by using the below commands:

```
Grunt> REGISTER '/home/edureka/Desktop/elephant-bird-hadoop-compat-4.1.jar';
```

```
Grunt> REGISTER '/home/edureka/Desktop/elephant-bird-pig-4.1.jar';
```

```
Grunt> REGISTER '/home/edureka/Desktop/json-simple-1.1.1.jar';
```

grunt>

```
grunt> REGISTER '/home/edureka/Downloads/elephant-bird-pig-4.1.jar';
```

```
grunt> REGISTER '/home/edureka/Downloads/elephant-bird-hadoop-compat-4.1.jar';
```

```
grunt> REGISTER '/home/edureka/Download/json-simple-1.1.1.jar';
```

2017-07-20 16:59:25 721 [main] WARN org.apache.hadoop.conf.Configuration deprecated: fo default name is deprecated. Instead, use fo defaultE

Note: You need to provide the path of the jar file accordingly.

After registering the required jars, we can now write a Pig script to perform Sentiment Analysis.

STEP-2

The tweets are in nested Json format and consists of map data types. We need to load the tweets using JsonLoader which supports maps, so we are using **elephant bird JsonLoader** to load the tweets.

Below is the first Pig statement required to load the tweets into Pig:-

```
load_tweets = LOAD '/user/flume/tweets/' USING  
com.twitter.elephantbird.pig.load.JsonLoader('-nestedLoad') AS myMap;
```

grunt>

```
grunt> load_tweets = LOAD '/user/flume/tweets/' USING com.twitter.elephantbird.pig.load.JsonLoader('-nestedLoad') AS myMap;
```

```
2017-07-20 17:22:17,884 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
```

```
2017-07-20 17:22:17,884 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
```

```
2017-07-20 17:22:17,885 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 5 time(s).
```

```
2017-07-20 17:22:17,885 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
```

grunts █

STEP-3

When we dump the above relation, we can see that all the tweets got loaded successfully.

```
((filter_level#medium,text#RT @SATAVRAJEEV: Vision of UPA 2 Govt., 15 months hardwork of Scientists, India's greatest achievement. Hats off to Indian Scientist. #Man...,contributors#,retweeted status#{filter_level=low, text=Vision of UPA 2 Govt., 15 months hardwork of Scientists, India's greatest achievement. Hats off to Indian Scientist. #Mangalyaan., contributors=null, geo=null, retweeted=false, in_reply_to_screen_name=null, possibly_sensitive=false, truncated=false, entities={trends={}, hashtags={{[text#Mangalyaan,indices#{(118),(129)}}]}, symbols={}, urls={}, user_mentions={}, lang=en, in_reply_to_status_id_str=null, id=514679859428077568, in_reply_to_user_id_str=null, source=<a href="https://mobile.twitter.com" rel="nofollow">Mobile Web (MS)</a>, favorited=false, in_reply_to_status_id=null, in_reply_to_user_id=null, created_at=Wed Sep 24 07:36:55 +0000 2014, retweet_count=61, favorite_count=36, id_str=514679859428077568, place=null, user={location=, default_profile=true, statuses_count=482, profile_background_tile=false, lang=en, profile_link_color=0084B4, profile_banner_url=https://pbs.twimg.com/profile_banners/2239518500/1411189226, id=2239518500, following=null, favourites_count=23, protected=false, profile_text_color=333333, contributors_enabled=false, description=President Indian Youth Congress and MP (Lok Sabha), verified=false, name=RAJEEV SATAV, profile_sidebar_border_color=C0DEED, profile_background_color=C0DEED, created_at=Tue Dec 10 17:29:14 +0000 2013, default_profile_image=false, followers_count=2781, geo_enabled=true, profile_image_url_https=https://pbs.twimg.com/profile_images/513187212435800064/UDZk13Ja_normal.jpeg, profile_background_image_url=http://abs.twimg.com/images/themes/theme1/bg.png, profile_background_image_url_https=https://abs.twimg.com/images/themes/theme1/bg.png, follow_request_sent=null, url=null, utc_offset=null, time_zone=null, notifications=null, friends_count=99, profile_use_background_image=true, profile_sidebar_fill_color=DDEEF6, screen_name=SATAVRAJEEV, id_str=2239518500, profile_image_url=http://pbs.twimg.com/profile_images/513187212435800064/UDZk13Ja_normal.jpeg, is_translator=false, listed_count=11, coordinates=null}, geo#, retweeted=false, in_reply_to_screen_name=, possibly_sensitive=false, truncated=false, lang=en, entities={trends={}, hashtags={{[text#Mangalyaan,indices#{(135),(140)}}]}, symbols={}, urls={}, user_mentions={{[id#2239518500,indices#{(3),(15)},screen_name=SATAVRAJEEV,id_str#2239518500,name#RAJEEV SATAV]}},in_reply_to_status_id_str#,id#514761581373444096,source=<a href="http://blackberry.com/twitter" rel="nofollow">Twitter for BlackBerry@</a>,in_reply_to_user_id_str#,favorite_count#,timestamp_ms#1411563699499,in_reply_to_status_id#,retweet_count#,in_reply_to_user_id#,created_at#Wed Sep 24 13:01:39 +0000 2014,favorite_count#,id_str#514761581373444096,place#,user#{location=Karnataka(KOLAR GOLD FIELD), default_profile=true, statuses_count=624, profile_background_tile=false, lang=en, profile_link_color=0084B4, id=920269178, following=null, favourites_count=1, protected=false, profile_text_color=333333, contributors_enabled=false, description=Gen sec Pradesh Youth Congress, verified=false, name=Karthick G.S., profile_sidebar_border_color=C0DEED, profile_background_color=C0DEED, created_at=Fri Nov 02 02:59:19 +0000 2012, default_profile_image=false, followers_count=65, geo_enabled=true, profile_image_url_https=https://pbs.twimg.com/profile_images/2795910700/ea21234e671ab50f1dcb956bb0f93bae_normal.jpeg, profile_background_image_url=http://abs.twimg.com/images/themes/theme1/bg.png, profile_background_image_url_https=https://abs.twimg.com/images/themes/theme1/bg.png, follow_request_sent=null, url=null, utc_offset=null, time_zone=null, notifications=null, friends_count=126, profile_use_background_image=true, profile_sidebar_fill_color=DDEEF6, screen_name=karthickPYC, id_str=920269178, profile_image_url=http://pbs.twimg.com/profile_images/2795910700/ea21234e671ab50f1dcb956bb0f93bae_normal.jpeg, is_translator=false, listed_count=1},coordinates#})
grunt> █
```

STEP-4

Now, we shall extract the **id** and the **tweet text** from the above tweets. The Pig statement necessary to perform this is as shown below:

```
extract_details = FOREACH load_tweets GENERATE myMap#'id' as id,myMap#'text' as text;
```

```
grunt> extract_details = FOREACH load_tweets GENERATE myMap#'id' as id,myMap#'text' as text;
2017-07-20 17:26:31,019 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-07-20 17:26:31,021 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - dfs.permissions is deprecated. Instead, use dfs.permissions.enabled
2017-07-20 17:26:31,022 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-07-20 17:26:31,153 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-07-20 17:26:31,154 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_MAP 4 time(s).
2017-07-20 17:26:31,154 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 5 time(s).
2017-07-20 17:26:31,154 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt>
```

We can see the extracted **id** and **tweet text** from the tweets in the below screen shot.

```
(514760821079953408,Nature Communications Goes OA | The Scientist Magazine: "Nature Communications, which publishes brief sci... http://t.co/5Q7t9pJzky #OA)
(514760881356304384,#kudos to the scientist who Made it to the Mars #ISRO)
(51476090065868224,RT @liberof: Demystifying Hadoop: http://t.co/jrErDYFs7H via @YouTube)
(514760914772717568,.@HansonFSU starts a First Friday: Ask a Scientist at a weekly local fair. Sweet idea! Way to go, Ken :) http://t.co/Pg0C4Jdw4n)
(514760926357356544,The mad scientist explaining circuits in robotics lab! http://t.co/pqo5E92Kbf)
(514760966660423680,RT @consultayaz: http://t.co/ra0U4V801F Hadoop Bigdata architect I'm hiring - great people - know anyone who ...)
(514760989523206144,RT @cutypie_ayesha: #IndiaAtMars #Mangalyaan vry proud moment for every #Indians 2day. MOM successfully entered mars orbit.. big applause f...)
(514761002538532864,(A frog scientist, as you can tell by my green visage))
(514761020431429632,RT @TeradataUK: Read @craigwentworth new article "Bullet-proof SQL - #Hadoop's real battleground" here http://t.co/6vVSfxJIXM #BigData #SQL)
(51476102496080641,RT @jyotsana khatri: It has been proved that our scientist are great. It's very proud n historic moment 4 all indians. #ThankYouIsro for ma...)
(514761049803726848,RT @sree: Like standing in Delhi and aiming a baseball at a NYC window - how a scientist described India's Mars mission. #ISRO http://t.co/...)
(514761056758284290,RT @GrumpyCatsMind: I'm no scientist, but you're an idiot.)
(514761077167751168,RT @sree: Like standing in Delhi and aiming a baseball at a NYC window - how a scientist described India's Mars mission. #ISRO http://t.co/...)
(514761124726988800,@azizonomics @seanmdav @jneeley78 @davidharsanyi I preface it with, "Here's a story..." I'm not a scientist making serious claims.)
(514761181681438720,Check out September CHUG: Making business decisions with SAS & Hadoop tonight at 6:00pm http://t.co/tFDQF3xHtX #CLT)
(514761183883055104,We are hiring: Scientist, Late Stage Pharmaceutical Development (Protein Formulation) Scien http://t.co/lcVjiMsrt0 #job @Genentech)
(514761185745702912,scientists make quantum leap, teleport data from light to matter http://t.co/RyZe42XBlN)
(514761193471623169,Coming Soon...
Chemical (Se)X, an anthology

Hot-ass Scientist - http://t.co/6RTWjpPYH)
(514761242343247872,RT @IndiaFinger: "For god's sake this is the last time I'm saying, we can't send Sanjay Jha or Priyanka Chaturvedi at Mars"- Director of IS...)
(514761262165540866,Charles Spence: the food scientist changing the way we eat http://t.co/XU27FzeGWW)
(514761266729336832,The Scientist.)
(514761305798868992,ISRO scientist job very well done. To achieve this great feat in the first attempt where the likes of the US and the USSR have failed.)
(514761319195492352,http://t.co/lguRNNM1I6)
(514761352242417664,USA gov jobs - Physical Scientist http://t.co/pZuzw04dg0 #ScienceJobs)
(514761494571917313,RT @BaluMahe: This is About #Mangalyaan. Stunned About Our Scientist. Proud. @isro @MarsOrbiter http://t.co/y4fx4JemMJ)
(514761576239230976,RT @PC ArabFC: Pictures. Priyanka Chopra Honoured at the Priyadarshini Academy Awards http://t.co/IkE1M8zQko @priyankachopra)
(514761581373444096,RT @SATAVRAJEEV: Vision of UPA 2 Govt., 15 months hardwork of Scientists, India's greatest achievement. Hats off to Indian Scientist. #Man...)
```

STEP -5

We have the tweet id and the tweet text in the relation named as **extract_details**.

Now, we shall extract the words from the text using the TOKENIZE key word in Pig.

**tokens = foreach extract_details generate id,text,
FLATTEN(TOKENIZE(text)) As word**

```
grunt>
grunt> dictionary = load '/AFINN.txt' using PigStorage('\t') AS(word:chararray,rating:int);
2017-07-20 11:31:50,075 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-07-20 11:31:50,078 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - dfs.permissions is deprecated. Instead, use dfs.permissions.enabled
2017-07-20 11:31:50,078 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-07-20 11:31:50,125 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-07-20 11:31:50,125 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-07-20 11:31:50,128 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt>
```

STEP -6

Now, we have to analyze the Sentiment for the tweet by using the words in the text. We will rate the word as per its meaning from +5 to -5 using the dictionary AFINN. The AFINN is a dictionary which consists of 2500 words which are rated from +5 to -5 depending on their meaning. You can download the dictionary from the above given link.

We will load the dictionary into pig by using the below statement:

**dictionary = load '/AFINN.txt' using PigStorage('\t')
AS(word:chararray,rating:int);**

The output of this will be look like following image :-

```
(wooo,4)
(woow,4)
(worn,-1)
(worried,-3)
(worry,-3)
(worrying,-3)
(worse,-3)
(worsen,-3)
(worsened,-3)
(worsening,-3)
(worsens,-3)
(worshiped,3)
(worst,-3)
(worth,2)
(worthless,-2)
(worthy,2)
(wow,4)
(wowow,4)
(wowwow,4)
(wrathful,-3)
(wreck,-2)
(wrong,-2)
(wronged,-2)
(wtf,-4)
(yeah,1)
(yearning,1)
(yees,2)
(yes,1)
(youthful,2)
(yucky,-2)
(yummy,3)
(zealot,-2)
```

STEP -7

Now, let's perform a map side join by joining the **tokens** statement and the dictionary contents using this command:

```
word_rating = join tokens by word left outer, dictionary by word  
using 'replicated';
```

```

grunt> word rating = join tokens by word left outer, dictionary by word using 'replicated';
2017-07-20 11:36:21,085 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-07-20 11:36:21,088 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - dfs.permissions is deprecated. Instead, use dfs.permissions.enabled
2017-07-20 11:36:21,088 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-07-20 11:36:21,205 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-07-20 11:36:21,205 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-07-20 11:36:21,205 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt>

```

We can see the schema of the statement after performing join operation by using the below command:

```
describe word_rating;
```

```

2017-07-20 11:36:21,205 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> describe word_rating;
2017-07-20 11:37:02,267 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-07-20 11:37:02,267 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-07-20 11:37:02,267 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
word_rating: {tokens::id: bytearray,tokens::text: bytearray,tokens::word: chararray,dictionary::word:chararray,dictionary::rating: int}
grunt>

```

STEP-8

Now we will extract the **id**, **tweet text** and **word rating**(from the dictionary) by using the below relation:

```
rating = foreach word_rating generate tokens::id as id,tokens::text as text,dictionary::rating as rate;
```



```

grunt>
grunt> rating = foreach word rating generate {tokens::id as id,tokens::text as text, dictionary::rating as rate;
2017-07-20 11:44:28,101 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.
max
2017-07-20 11:44:28,104 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - dfs.permissions is deprecated. Instead, use dfs.permissions.enabled
2017-07-20 11:44:28,105 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-07-20 11:44:28,166 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-07-20 11:44:28,168 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-07-20 11:44:28,168 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt>

```

STEP-9

Now, we will group the **rating of all the words in a tweet** by using the below relation:

```
word_group = group rating by (id,text);
```

```

grunt>
grunt> word_group = group rating by (id,text);
2017-07-20 11:45:57,098 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-07-20 11:45:57,098 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-07-20 11:45:57,098 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt>

```

STEP-10

Now, let's perform the **Average** operation on the **rating of the words per each tweet**.

```
avg_rate = foreach word_group generate group, AVG(rating.rate) as tweet_rating;
```

```

grunt> avg_rate = foreach word_group generate group, AVG(rating.rate) as tweet_rating;
2017-07-20 11:47:29,359 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-07-20 11:47:29,359 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-07-20 11:47:29,360 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt>

```

STEP-11

From the above relation, we will get all the tweets i.e., both positive and negative.

Here, we can classify the positive tweets by taking the rating of the tweet which can be from **0-5**. We can classify the negative tweets by taking the rating of the tweet from **-5 to -1**.

We have now successfully performed the Sentiment Analysis on Twitter data using Pig. We now have the tweets and its rating, so let's perform an operation to filter out the positive tweets.

Now we will filter the positive tweets using the below statement:

```
positive_tweets = filter avg_rate by tweet_rating >= 0;
```

```

grunt>
grunt> positive_tweets = filter avg_rate by tweet_rating >= 0;
2017-07-20 11:48:39,103 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-07-20 11:48:39,103 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-07-20 11:48:39,103 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 1 time(s).
2017-07-20 11:48:39,103 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt>

```

We can see the positive tweets and its rating in the below screen shot.

```

oMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1 SECONDS)
2017-07-20 11:55:14,597 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 8 time(s); retry policy is RetryUp
oMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1 SECONDS)
2017-07-20 11:55:15,599 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 9 time(s); retry policy is RetryUp
oMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1 SECONDS)
2017-07-20 11:55:15,701 [main] ERROR org.apache.hadoop.security.UserGroupInformation - PrivilegedActionException as:edureka (auth:SIMPLE) cause:java.io.IOException: ja
va.net.ConnectException: Call From localhost.localdomain/127.0.0.1 to 0.0.0.0:10020 failed on connection exception: java.net.ConnectException: Connection refused; For m
ore details see: http://wiki.apache.org/hadoop/ConnectionRefused
2017-07-20 11:55:15,702 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregat
ion.
2017-07-20 11:55:15,702 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-07-20 11:55:15,704 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-07-20 11:55:15,704 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-07-20 11:55:15,726 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-07-20 11:55:15,726 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
((514760652024344576,PM never taken the credit for Mangalayan,he said dedicated scientist, services,citizens have worked day & night to take the country forward),2.
3)
((514760766478508032,If you think like a scientist it reduces the sting of failure http://t.co/mryYCuosn4 #act #work #entrepreneur #business #freedom #you),0.0)
((51476096660423680,RT @consultayaz: http://t.co/ra0U4V801F Hadoop Bigdata architect I'm hiring - great people - know anyone who ...),3.0)
((514760989523206144,RT @cutypie_ayasha: #IndiaAtMars #Mangalyaan vry proud moment for every #Indians 2day. MOM successfully entered mars orbit.. big applause f...),1.666
566666666667)
((514761024960880641,RT @jyotsana khatri: It has been proved that our scientist are great. It's very proud n historic moment 4 all indians. #ThankYouIsro for ma...),2.0)
((514761185745702912,scientists make quantum leap, teleport data from light to matter http://t.co/RyZe42XBln),1.0)
((514761305798868992,ISRO scientist job very well done. To achieve this great feat in the first attempt where the likes of the US and the USSR have failed.),2.5)
((514761581373444096,RT @SATAVRAJEEV: Vision of UPA 2 Govt., 15 months hardwork of Scientists, India's greatest achievement. Hats off to Indian Scientist. #Man...),3.0)
((887286061247979521,There's room for many models, not just one. Publishing evolved around a community of interest model, consider the s... https://t.co/jMhVlabdA0),1.0)
((887286100028309504,RT @PcOurHeartbeat: very few people have done what priyanka did , i'm very proud of her - sushmita sen
#HappyBirthdayPriyankaChopra https://t.co/...),2.0)
((887286104184979456,RT @redesign: There's room for many models, not just one. Publishing evolved around a community of interest model, consider the s...),1.0)
((887286230789914624,RT @quietlibrarian5: Identifying Predatory Publishers. Some good comment from @andypf and @GinnyBarbour https://t.co/pQHimNik00),3.0)
grunt>

```

STEP-12

Now we will filter the negative tweets using the below statement:

```
negative_tweets = filter avg_rate by tweet_rating<=0;
```

```

grunt>
grunt> negative_tweets = filter avg_rate by tweet_rating<=0;
2017-07-20 14:30:33,709 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.
max
2017-07-20 14:30:33,712 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - dfs.permissions is deprecated. Instead, use dfs.permissions.enabled
2017-07-20 14:30:33,713 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-07-20 14:30:33,826 [main] WARN org.apache.pig.PigServer - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-07-20 14:30:33,826 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-07-20 14:30:33,826 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 4 time(s).
2017-07-20 14:30:33,826 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt>

```

We can see the positive tweets and its rating in the below screen shot.

```

oMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1 SECONDS)
2017-07-20 14:35:56,456 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 6 time(s); retry policy is RetryUpT
oMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1 SECONDS)
2017-07-20 14:35:57,457 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 7 time(s); retry policy is RetryUpT
oMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1 SECONDS)
2017-07-20 14:35:58,459 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 8 time(s); retry policy is RetryUpT
oMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1 SECONDS)
2017-07-20 14:35:59,461 [main] INFO org.apache.hadoop.ipc.Client - Retrying connect to server: 0.0.0.0/0.0.0.0:10020. Already tried 9 time(s); retry policy is RetryUpT
oMaximumCountWithFixedSleep(maxRetries=10, sleepTime=1 SECONDS)
2017-07-20 14:35:59,563 [main] ERROR org.apache.hadoop.security.UserGroupInformation - PrivilegedActionException as:edureka (auth:SIMPLE) cause:java.io.IOException: ja
va.net.ConnectException: Call From localhost.localdomain/127.0.0.1 to 0.0.0.0:10020 failed on connection exception: java.net.ConnectException: Connection refused; For m
ore details see: http://wiki.apache.org/hadoop/ConnectionRefused
2017-07-20 14:35:59,563 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Unable to retrieve job to compute warning aggregat
ion.
2017-07-20 14:35:59,564 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-07-20 14:35:59,565 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-07-20 14:35:59,566 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-07-20 14:35:59,594 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-07-20 14:35:59,594 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
((514760716159827968,Scientist who has an autoimmune hair loss disease researches it and finds a cure in a drug already approved by FDA ... http://t.co/3ozPHZmAad),-0.5)
((514760766478508032,If you think like a scientist it reduces the sting of failure http://t.co/mryYCuosn4 #act #work #entrepreneur #business #freedom #you),0.0)
((514760816818540546,no body said it was easy, and no one ever said it would be so hard :( ♫ The Scientist by @coldplay – https://t.co/oIpUufJDfw),-0.5)
((514760926357356544,The mad scientist explaining circuits in robotics lab! http://t.co/pqo5E92Kbf),-3.0)
((514761056758284290,RT @GrumpyCatsMind: I'm no scientist, but you're an idiot.),-1.0)
((887286139173969922,RT @realscientists: I believe that the current system of scholarly publishing is one reason for a widespread lack of trust in science... ),-0.5)
((887286164775940096,RT @PurnL: Dr. David Kelly CMG (59), Welsh Scientist, died this day in 2003, in suspicious circumstances. RIP. Never forgotten. https://t.c...),-2.5)
((887286238369136640,RT @DaniallHafiz: 37. Quantico
Ingat macam csi punya cerita, tapi lain sikit. Pasal fbi agents kena betrayed dgn fbi sendiri. Ade Pri... ),-3.0)
((887286239476449281,RT @LeaBlackMiami: &amp; got rid of Isis in one month, repealed &amp; replaced ACA in 1 day, &amp; scientist proved climate hoax, &amp; we're all b
uy!... ),-2.0)

```

- **Conclusion :-**

AS above we can extract the data from various social networking site (eg- Facebook,Instagram etc.) and can analysis various comment basis on the data. But for that we need the API of that application software and we need to download relevant jar files.

- **References:-**

- * Google.com
- * Youtube
- *wikipedia
- *Apache lucid Network