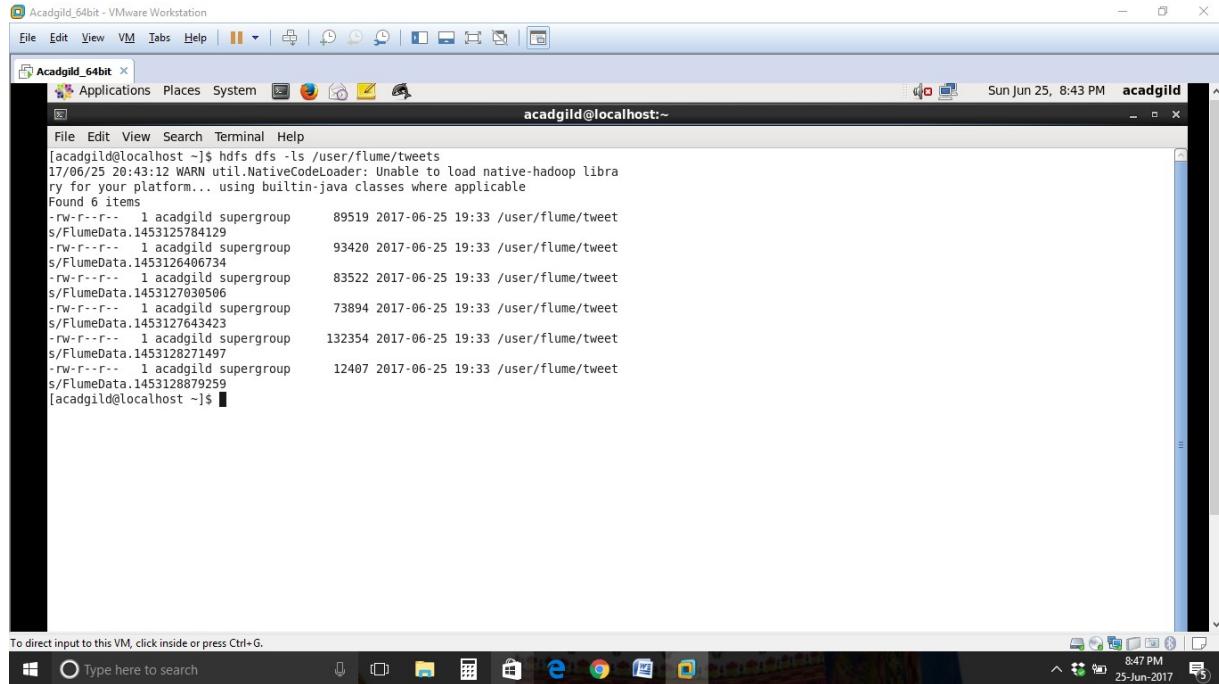


Project 1.3

Twitter Sentiment Analysis

List of files in which tweets are streamed in HDFS



```
[acadgild@localhost ~]$ hdfs dfs -ls /user/flume/tweets
17/06/25 20:43:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 6 items
-rw-r--r-- 1 acadgild supergroup 89519 2017-06-25 19:33 /user/flume/tweet
s/FlumeData.1453125784129
-rw-r--r-- 1 acadgild supergroup 93420 2017-06-25 19:33 /user/flume/tweet
s/FlumeData.1453126406734
-rw-r--r-- 1 acadgild supergroup 83522 2017-06-25 19:33 /user/flume/tweet
s/FlumeData.145312703056
-rw-r--r-- 1 acadgild supergroup 73894 2017-06-25 19:33 /user/flume/tweet
s/FlumeData.1453127643423
-rw-r--r-- 1 acadgild supergroup 132354 2017-06-25 19:33 /user/flume/tweet
s/FlumeData.1453128271497
-rw-r--r-- 1 acadgild supergroup 12407 2017-06-25 19:33 /user/flume/tweet
s/FlumeData.1453128879259
[acadgild@localhost ~]$
```

Register the downloaded jars for JSONLoader in pig by using the below commands

```
(68999149357875202,#Contest live - Follow @acadgild and participate in #AchieveMore contest to win shopping vouchers!),4.0)
((689991815019012096,@AirliftFilm @Androidheadline hey its time for you guys follow @acadgild 2 #AchieveMore & participate in contest Win Rs.500 worth vouchers!),2.0)
((689992134729830400,@Sampat @bhak_sala @sunilgulati Its time to follow @acadgild To #AchieveMore and participate in contest You can win shopping vouchers!),4.0)
((689992531188068352,Speed data management processes on #spark with SAS Data Loader for #Hadoop #newrelease. Download free trial now! https://t.co/N3ZF1f3kpq),1.0)
((689992537395601408,If you want to have a simple explanation of #HDFS works and you like cartoons, this is the place: https://t.co/B0IYRzuEKX #Hadoop #bigData),1.5)
((68999251942627328,RT @HadoopDaily: If you want to have a simple explanation of #HDFS works and you like cartoons, this is the place: https://t.co/B0IYRzuEKX ...),1.5)
((689993078083350528,RT @HadoopDaily: If you want to have a simple explanation of #HDFS works and you like cartoons, this is the place: https://t.co/B0IYRzuEKX ...),1.5)
((689993722714312705,RT @HadoopDaily: If you want to have a simple explanation of #HDFS works and you like cartoons, this is the place: https://t.co/B0IYRzuEKX ...),1.5)
((689993749666885632,#hadoop is now 10 years old. Here is a good infographic of how we got here. https://t.co/h0W5lZ6WQOA),3.0)
((689995108109701121,RT @HadoopDaily: If you want to have a simple explanation of #HDFS works and you like cartoons, this is the place: https://t.co/B0IYRzuEKX ...),1.5)
((689997211175645184,Speed data management processes on #spark with SAS Data Loader for #Hadoop #newrelease. Download free trial now! https://t.co/LrcPE1rMVL),1.0)
grunt> load tweets = LOAD 'hdfs://localhost:9000/user/flume/tweets/' USING com.twitter.elephantbird.pig.load.JsonLoader('-nestedLoad') AS myMap;
2017-06-25 20:48:01,805 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.b.counters.max
2017-06-25 20:48:01,805 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-06-25 20:48:01,806 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-25 20:48:01,898 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 1 time(s).
2017-06-25 20:48:01,898 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-06-25 20:48:01,899 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-06-25 20:48:01,899 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> REGISTER '/home/acadgild/Downloads/elephant-bird-hadoop-compat-4.1.jar';
grunt> REGISTER '/home/acadgild/Downloads/elephant-bird-pig-4.1.jar';
grunt> REGISTER '/home/acadgild/Downloads/json-simple-1.1.1.jar';
grunt>
```

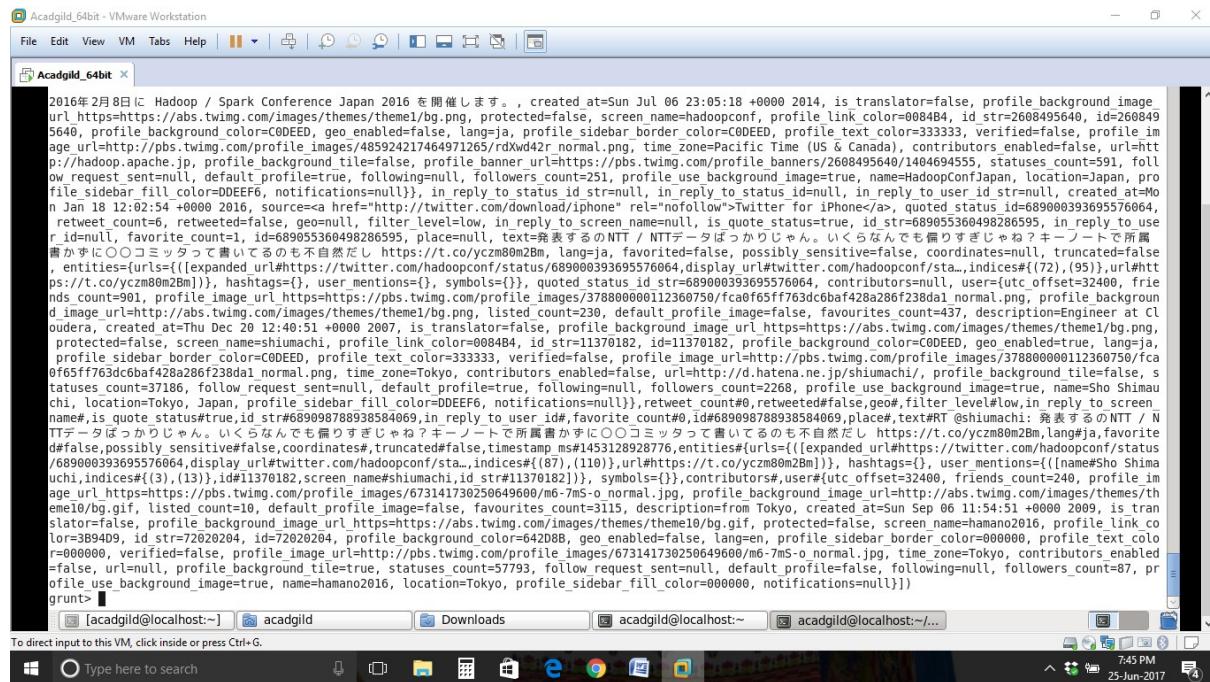
First Pig statement required to load the tweets into Pig:

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

```
dump load_tweets;
```

```
@divyasaika
@VinceRugari
@contest,world,4.0)
((68999149357875202,#Contest live - Follow @acadgild and participate in #AchieveMore contest to win shopping vouchers!),4.0)
((689991815019012096,@AirliftFilm @Androidheadline hey its time for you guys follow @acadgild 2 #AchieveMore & participate in contest Win Rs.500 worth vouchers!),2.0)
((689992134729830400,@Sampat @bhak_sala @sunilgulati Its time to follow @acadgild To #AchieveMore and participate in contest You can win shopping vouchers!),4.0)
((689992531188068352,Speed data management processes on #spark with SAS Data Loader for #Hadoop #newrelease. Download free trial now! https://t.co/N3ZF1f3kpq),1.0)
((689992537395601408,If you want to have a simple explanation of #HDFS works and you like cartoons, this is the place: https://t.co/B0IYRzuEKX #Hadoop #bigData),1.5)
((68999251942627328,RT @HadoopDaily: If you want to have a simple explanation of #HDFS works and you like cartoons, this is the place: https://t.co/B0IYRzuEKX ...),1.5)
((689993078083350528,RT @HadoopDaily: If you want to have a simple explanation of #HDFS works and you like cartoons, this is the place: https://t.co/B0IYRzuEKX ...),1.5)
((689993722714312705,RT @HadoopDaily: If you want to have a simple explanation of #HDFS works and you like cartoons, this is the place: https://t.co/B0IYRzuEKX ...),1.5)
((689993749666885632,#hadoop is now 10 years old. Here is a good infographic of how we got here. https://t.co/h0W5lZ6WQOA),3.0)
((689995108109701121,RT @HadoopDaily: If you want to have a simple explanation of #HDFS works and you like cartoons, this is the place: https://t.co/B0IYRzuEKX ...),1.5)
((689997211175645184,Speed data management processes on #spark with SAS Data Loader for #Hadoop #newrelease. Download free trial now! https://t.co/LrcPE1rMVL),1.0)
grunt> load_tweets = LOAD 'hdfs://localhost:9000/user/flume/tweets/' USING com.twitter.elephantbird.pig.load.JsonLoader('-nestedLoad') AS myMap;
2017-06-25 20:48:01,806 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-06-25 20:48:01,890 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_DOUBLE 1 time(s).
2017-06-25 20:48:01,890 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-06-25 20:48:01,890 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-06-25 20:48:01,890 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> dump load_tweets;
```

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



```
2016年2月8日に Hadoop / Spark Conference Japan 2016 を開催します。, created at=Sun Jul 06 23:05:18 +0000 2014, is_translator=false, profile_background_image_url=https://abs.twimg.com/images/themes/theme1/bg.png, protected=false, screen_name=hadoopconf, profile_link_color=0084B4, id_str=2608495640, id=2608495640, profile_background_color=C0DEED, geo_enabled=false, lang=ja, profile_sidebar_border_color=C0DEED, profile_text_color=333333, verified=false, profile_image_url=http://pbs.twimg.com/profile_images/485924217464971265/rDxWd42r_normal.png, time_zone=Pacific Time (US & Canada), contributors_enabled=false, url=http://hadoop.apache.jp, profile_background_tile=false, profile_banner_url=https://pbs.twimg.com/profile_banners/2608495640/1404694555, statuses_count=591, follow_request_sent=null, following=null, followers_count=251, profile_use_background_image=true, name=HadoopConfJapan, location=japan, profile_sidebar_fill_color=DDEEF6, notifications=null}, in_reply_to_status_id=null, in_reply_to_status_id_str=null, in_reply_to_user_id=null, created_at=Mon Jan 18 12:02:54 +0000 2016, source=<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>, quoted_status_id=68900039369576064, retweet_count=6,retweeted=false,geo=null,filter_level=low,in_reply_to_screen_name=null,is_quote_status=true,id_str=68900039369576064, in_reply_to_use_r_id=null,favorite_count=1,id=68900039369576064,place=null, text=発表するのNTT / NTTデータばっかりじゃん。いくらなんでも偏りすぎじゃね？キーノートで所属書かずに○○コミッタって書いてるものも不自然だし https://t.co/yCzmB0m2Bm, lang=ja, favorite_count=1, id=68900039369576064,display_url=https://twitter.com/hadoopconf/sta..., indices=[{72}, {95}], url=https://t.co/yCzmB0m2Bm)), hashtags={}, user_mentions={}, quoted_status_id_str=68900039369576064, friends_count=901, profile_image_url=https://pbs.twimg.com/profile_images/378800000112360750/fca0f65ff763dc6ba428a286f238dal_normal.png, profile_background_image_url=https://abs.twimg.com/images/themes/theme1/bg.png, listed_count=230, default_profile_image=false, favourites_count=437, description=Engineer at Cloudera, created_at=Thu Dec 20 12:40:51 +0000 2007, is_translator=false, profile_background_image_url=https://abs.twimg.com/images/themes/theme1/bg.png, protected=false, screen_name=shiumachi, profile_link_color=0084B4, id_str=11370182, id=11370182, profile_background_color=C0DEED, geo_enabled=true, lang=ja, profile_sidebar_border_color=C0DEED, profile_text_color=333333, verified=false, profile_image_url=http://pbs.twimg.com/profile_images/378800000112360750/fca0f65ff763dc6ba428a286f238dal_normal.png, filter_level=low,in_reply_to_screen_name=is_quote_status=true,id_str=68900039369576064,in_reply_to_user_id#, favorite_count=0, id=68900039369576064, place#, text#RT @shiumachi: 発表するのNTT / NTTデータばっかりじゃん。いくらなんでも偏りすぎじゃね？キーノートで所属書かずに○○コミッタって書いてるものも不自然だし https://t.co/yCzmB0m2Bm, lang#ja, favorite_count=0, possibly_sensitive=false, coordinates#, truncated=false, timestamp_ms#1453128928776, entities#[urls=[{expanded_url#https://twitter.com/hadoopconf/status/68900039369576064,display_url#https://twitter.com/hadoopconf/sta..., indices=[{87}, {110}],url#https://t.co/yCzmB0m2Bm}), hashtags={}, user_mentions={([nameSho Shimauchi,indices=[{3}, {13}],id#11370182,screen_name=shiumachi,id_str#11370182])}, symbols={}, contributors#, user#utc_offset=32400, friends_count=240, profile_image_url=https://pbs.twimg.com/profile_images/673141730259649600/m6-7m5-o_normal.jpg, profile_background_image_url=http://abs.twimg.com/images/themes/theme10/bg.gif, listed_count=10, default_profile_image=false, favourites_count=3115, description=From Tokyo, created_at=Sun Sep 06 11:54:51 +0000 2009, is_translator=false, profile_background_image_url=https://abs.twimg.com/images/themes/theme10/bg.gif, protected=false, screen_name=hamano2016, profile_link_color=3B94D9, id_str=72020204, id=72020204, profile_background_color=64208B, geo_enabled=false, lang=en, profile_sidebar_border_color=000000, profile_text_color=000000, verified=false, profile_image_url=http://pbs.twimg.com/profile_images/673141730259649600/m6-7m5-o_normal.jpg, time_zone=Tokyo, contributors_enabled=false, url=null, profile_background_tile=true, statuses_count=57793, follow_request_sent=null, default_profile=false, following=null, followers_count=87, profile_use_background_image=true, name=hamano2016, location=Tokyo, profile_sidebar_fill_color=000000, notifications=null})
```

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;
```

```
dump extract_details;
```

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

The screenshot shows a VMware Workstation window with a terminal session titled "Acadgild_64bit". The terminal window displays a large number of tweets from the user "@acadgild" in a monospaced font. The tweets are primarily in English and discuss topics such as Hadoop, Big Data, and contests. The terminal has a blue background and white text. Below the terminal is a taskbar with several icons, including a file browser, a download manager, and a search bar. The status bar at the bottom shows the date and time as "25-Jun-2017 8:18 PM".

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;
```

dump tokens;

From the below screen shot, we can see that the text got divided into words.

Load the dictionary into pig by using the below statement:

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

```
[...]  
dictionary = load 'AFINN.txt' using PigStorage('\t') AS(word:chararray,rating:int);  
dump dictionary;  
[acadgild@localhost ~]#
```

We can see the contents of the AFINN dictionary in the below screen shot.

```
(worried,-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)  
(wowww,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)  
(zealots,-2)  
(zealous,2)  
[acadgild@localhost ~]#
```

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';

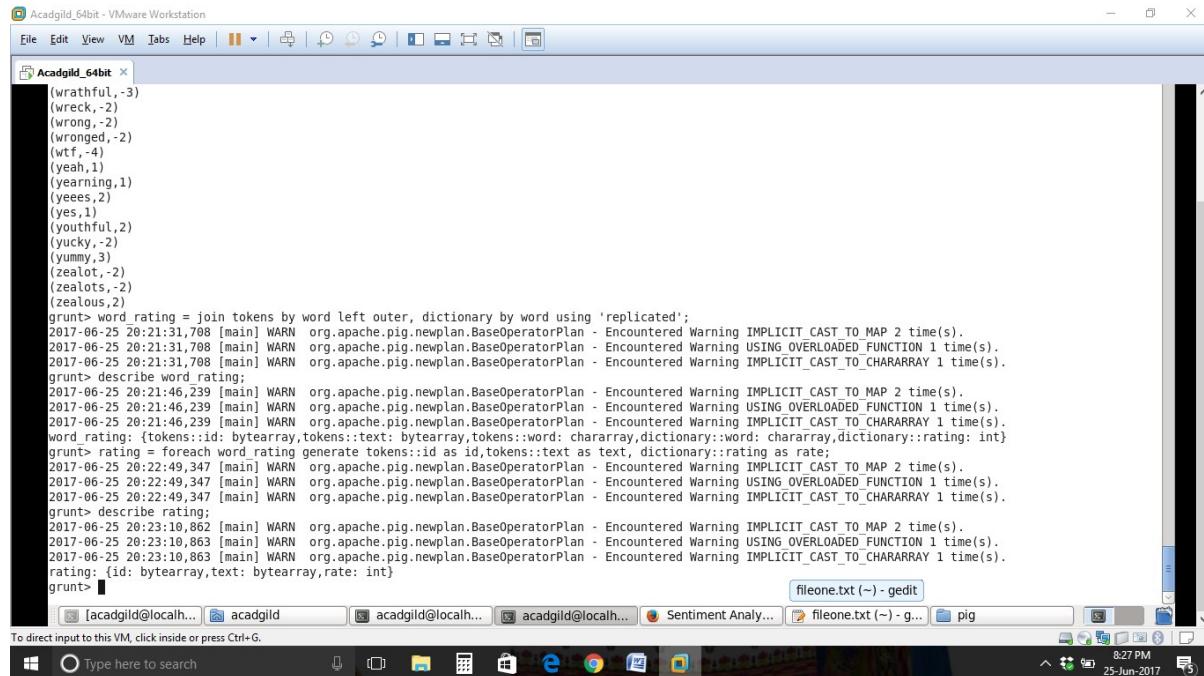
describe word_rating;

```
(worsening,-3)
(worsens,-3)
(worshiped,3)
(worst,-3)
(worth,2)
(worthless,-2)
(worthy,2)
(wow,4)
(wowwww,4)
(wrathful,-3)
(wreck,-2)
(wrong,-2)
(wronged,-2)
(wtf,-4)
(yeah,1)
(yeardining,1)
(yeet,2)
(yes,1)
(youthful,2)
(yucky,-2)
(yummy,3)
(zealot,-2)
(zealots,-2)
(zealous,2)
grunte> word_rating = join tokens by word left outer, dictionary by word using 'replicated';
2017-06-25 20:21:31,708 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-06-25 20:21:31,708 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-06-25 20:21:31,708 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> describe word_rating;
2017-06-25 20:21:46,239 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-06-25 20:21:46,239 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-06-25 20:21:46,239 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
word_rating: {tokens::id: bytearray, tokens::text: bytearray, tokens::word: chararray, dictionary::rating: int}
fileone.txt (~) - gedit
[acadgild@localhost ~]$ acadgild@localhost ~]$ acadgild@localhost ~]$ Sentiment Anal...]$ fileone.txt (~) - g...]$ pig
[acadgild@localhost ~]$ To direct input to this VM, click inside or press Ctrl+G.
[acadgild@localhost ~]$ 8:26 PM 25-Jun-2017
```

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;

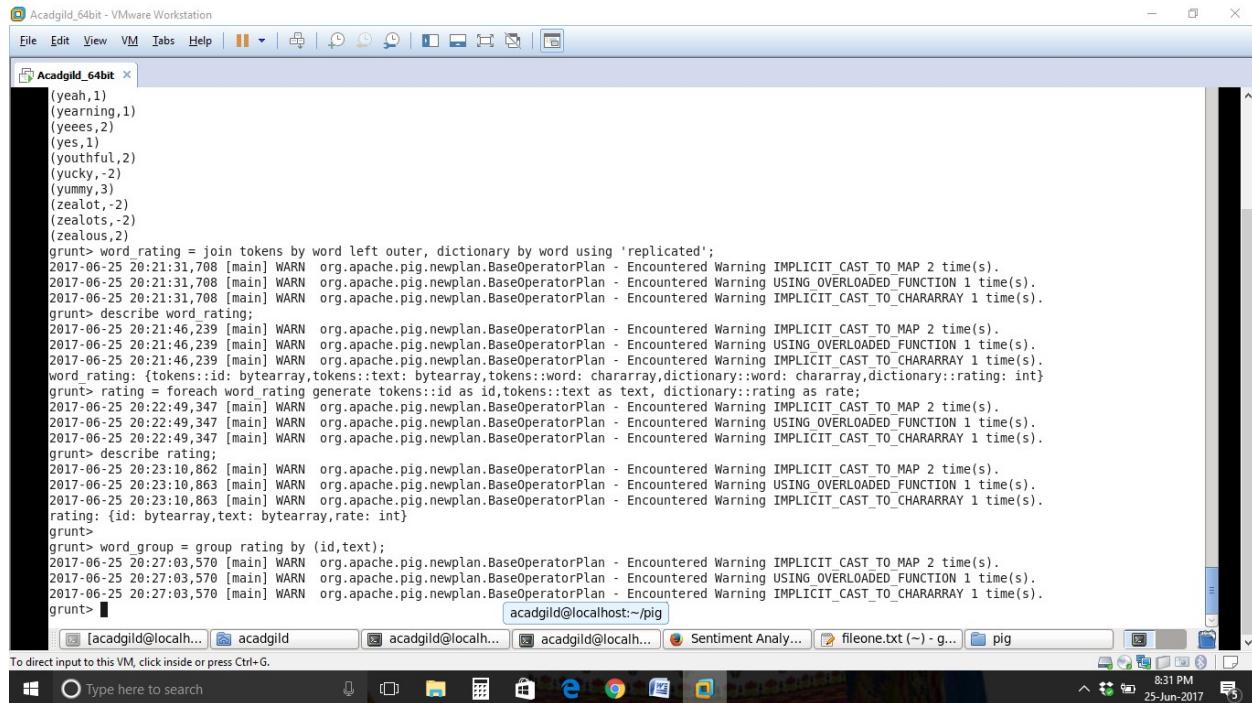
describe rating;



```
(wrathful,-3)
(wreck,-2)
(wrong,-2)
(wrongo,-2)
(wtf,-4)
(yeah,1)
(yearning,1)
(yees,2)
(yes,1)
(youthful,2)
(yucky,-2)
(yummy,3)
(zealot,-2)
(zealots,-2)
(zealous,2)
grunt> word_rating = join tokens by word left outer, dictionary by word using 'replicated';
2017-06-25 20:21:31,708 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-06-25 20:21:31,708 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-06-25 20:21:31,708 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> describe word rating;
2017-06-25 20:21:46,239 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-06-25 20:21:46,239 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-06-25 20:21:46,239 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
word_rating: {tokens::id: bytearray, tokens::text: bytearray, tokens::word: chararray, dictionary::chararray, dictionary::rating: int}
grunt> rating = foreach word_rating generate tokens::id as id, tokens::text as text, dictionary::rating as rate;
2017-06-25 20:22:49,347 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-06-25 20:22:49,347 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-06-25 20:22:49,347 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
rating: {id: bytearray, text: bytearray, rate: int}
grunt> 
```

Group the rating of all the words in a tweet by using the below relation:

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



```
(yeah,1)
(yeeres,2)
(yes,1)
(youthful,2)
(yucky,-2)
(yummy,3)
(zealot,-2)
(zealots,-2)
(zealous,2)
grunt> word_rating = join tokens by word left outer, dictionary by word using 'replicated';
2017-06-25 20:21:31,708 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-06-25 20:21:31,708 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-06-25 20:21:31,708 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> describe word_rating;
2017-06-25 20:21:46,239 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-06-25 20:21:46,239 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-06-25 20:21:46,239 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - 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> rating = foreach word_rating generate tokens::id as id,tokens::text as text,dictionary::word as rate;
2017-06-25 20:22:49,347 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-06-25 20:22:49,347 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-06-25 20:22:49,347 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> describe rating;
2017-06-25 20:23:10,862 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-06-25 20:23:10,863 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-06-25 20:23:10,863 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
rating: {id: bytearray,text: bytearray,rate: int}
grunt>
grunt> word_group = group rating by (id,text);
2017-06-25 20:27:03,570 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_MAP 2 time(s).
2017-06-25 20:27:03,570 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING_OVERLOADED_FUNCTION 1 time(s).
2017-06-25 20:27:03,570 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt>
```

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;
```

```
dump avg_rate;
```

Acadgild_64bit - VMware Workstation

File Edit View VM Tabs Help || |

Acadgild_64bit

```
(youthful,2)
(yucky,-2)
(yummy,3)
(zealot,-2)
(zealots,-2)
(zealous,2)
grunt> word rating = join tokens by word left outer, dictionary by word using 'replicated';
2017-06-25 20:21:31,708 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT CAST TO MAP 2 time(s).
2017-06-25 20:21:31,708 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING OVERLOADED FUNCTION 1 time(s).
2017-06-25 20:21:31,708 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> describe word_rating;
2017-06-25 20:21:46,239 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT CAST TO MAP 2 time(s).
2017-06-25 20:21:46,239 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING OVERLOADED FUNCTION 1 time(s).
2017-06-25 20:21:46,239 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - 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> rating = foreach word_rating generate tokens::id as id,tokens::text as text,dictionary::rating as rate;
2017-06-25 20:22:49,347 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT CAST TO MAP 2 time(s).
2017-06-25 20:22:49,347 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING OVERLOADED FUNCTION 1 time(s).
2017-06-25 20:22:49,347 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> describe rating;
2017-06-25 20:23:10,862 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT CAST TO MAP 2 time(s).
2017-06-25 20:23:10,863 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING OVERLOADED FUNCTION 1 time(s).
2017-06-25 20:23:10,863 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
rating: {id: bytearray,text: bytearray,rate: int}
grunt> word_group = group rating by (id,text);
2017-06-25 20:27:03,570 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT CAST TO MAP 2 time(s).
2017-06-25 20:27:03,570 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING OVERLOADED FUNCTION 1 time(s).
2017-06-25 20:27:03,570 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> avg_rate = foreach word_group generate group, AVG(rate) as tweet_rating;
2017-06-25 20:28:49,542 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT CAST TO MAP 2 time(s).
2017-06-25 20:28:49,542 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning USING OVERLOADED FUNCTION 1 time(s).
2017-06-25 20:28:49,542 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
grunt> fileone.txt (~) -> edit
```

Calculated the Average rating of the tweet using the rating of the each word

Filter the positive tweets using the below statement:

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

```
dump positive_tweets;
```

The positive tweets and its rating in the below screen shot