Project – Twitter Sentiment Analysis

```
grunt> REGISTER '/home/acadgild/Flume/elephant-bird-hadoop-compat-4.1.jar';
grunt> REGISTER '/home/acadgild/Flume/elephant-bird-pig-4.1.jar';
grunt> REGISTER '/home/acadgild/Flume/json-simple-1.1.1.jar';
```

```
grunt> load_tweets = LOAD '/home/acadgild/Flume/tweets/' USING com.twitter.eleph
antbird.pig.load.JsonLoader('-nestedLoad') AS myMap;home
```

acadgild@localhost:~

File   Edit   View   Search   Terminal   Help

o_user_id#,favorite_count#0,id#689098788938584069,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
で所属書かずに〇〇コミッタって書いてるのも不自然だし https://t.co/yczm80m2Bm,lan
g#ja,favorited#false,possibly_sensitive#false,coordinates#,truncated#false,times
tamp_ms#1453128928776,entities#{urls={([expanded_url#https://twitter.com/hadoopc
onf/status/689000393695576064,display_url#twitter.com/hadoopconf/sta…,indices#{(
87),(110)},url#https://t.co/yczm80m2Bm])}, hashtags={}, user_mentions={([name#Sh
o Shimauchi,indices#{(3),(13)},id#11370182,screen_name#shiumachi,id_str#11370182
])}, symbols={}},contributors#,user#{utc_offset=32400, friends_count=240, profil
e_image_url_https=https://pbs.twimg.com/profile_images/673141730250649600/m6-7mS
-o_normal.jpg, profile_background_image_url=http://abs.twimg.com/images/themes/t
heme10/bg.gif, listed_count=10, default_profile_image=false, favourites_count=31
15, description=from Tokyo, created_at=Sun Sep 06 11:54:51 +0000 2009, is_transl
ator=false, profile_background_image_url_https=https://abs.twimg.com/images/them
es/theme10/bg.gif, protected=false, screen_name=hamano2016, profile_link_color=3
B94D9, id_str=72020204, id=72020204, profile_background_color=642D8B, geo_enable
d=false, lang=en, profile_sidebar_border_color=000000, profile_text_color=000000
, verified=false, profile_image_url=http://pbs.twimg.com/profile_images/67314173
0250649600/m6-7mS-o_normal.jpg, time_zone=Tokyo, contributors_enabled=false, url
=null, profile_background_tile=true, statuses_count=57793, follow_request_sent=n
ull, default_profile=false, following=null, followers_count=87, profile_use_back
ground_image=true, name=hamano2016, location=Tokyo, profile_sidebar_fill_color=0
00000, notifications=null}])
grunt>

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

```
grunt> dump extract_details
```

```
acadgild@localhost:~                                          _ □ ✕

File  Edit  View  Search  Terminal  Help

(689095108109701121,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 …)
(689095143094358016,RT @sysfore: #BigData on #AmazonWebServices #AWS #cloud  #IoT #Hadoop
https://t.co/zsLdfE4Ylv https://t.co/CJKdDtJqHG)
(689095191861526528,RT @mloughlin: 4 Steps to a Smart #DataLake https://t.co/GhC7s2Z22B… #BigData
 #FinTech #NoSQL #Hadoop)
(689095215769108480,RT @cathy_burkey: Got #bigdata? Manage it on your own terms with #Hadoop and
@SASDataMGMT https://t.co/HtUwR0Nm7V https://t.co/xCpwKwWy2t)
(689095275445637120,Why @Syncsort introduced the #mainframe to #Hadoop, ... Q&amp;A
https://t.co/3DJw5LUdRi)
(689095406173745152,Who needs #Hadoop? Your data probably fits in RAM and that's easier, faster,
cheaper. https://t.co/EDzsC6vSr3)
(689095542455033856,RT @EMCECS: Faster insights than ever before with Elastic Cloud #Storage (#EC
S) and #Hadoop: https://t.co/zrepnZBupc https://t.co/C9rTls2Jg7)
(689095813549678593,RT @shimtoru: 参加登録始まりました！2月8日、参加費無料です ｜
Hadoop / Spark Conference Japan 2016 Tickets, Mon, Feb 8, 2016 at 10:00 AM ｜ Eventbrite https://t
.…)
(689098580037095424,#CustomerService #Job in #SãoPaulo: Customer Operations Engineer - Global at
Cloudera https://t.co/8yI45tXivr #Hadoop #BigData #data #Jobs)
(689098719489339392,One of my Hadoop consultant is looking for new project let me know if any one
 any open requirement for hadoop please…https://t.co/bhacfbWlFs)
(689098788938584069,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のも不自然だし https://t.co/yczm80m2Bm)
grunt> ▮
```

```
grunt> tokens = foreach extract_details generate id,text, FLATTEN(TOKENIZE(text)) As word;
```

```
grunt> dump tokens▮
```

```
acadgild@localhost:~                                          _ □ ✕

File  Edit  View  Search  Terminal  Help

(689098719489339392,One of my Hadoop consultant is looking for new project let me know if any one
 any open requirement for hadoop please…https://t.co/bhacfbWlFs,any)
(689098719489339392,One of my Hadoop consultant is looking for new project let me know if any one
 any open requirement for hadoop please…https://t.co/bhacfbWlFs,one)
(689098719489339392,One of my Hadoop consultant is looking for new project let me know if any one
 any open requirement for hadoop please…https://t.co/bhacfbWlFs,any)
(689098719489339392,One of my Hadoop consultant is looking for new project let me know if any one
 any open requirement for hadoop please…https://t.co/bhacfbWlFs,open)
(689098719489339392,One of my Hadoop consultant is looking for new project let me know if any one
 any open requirement for hadoop please…https://t.co/bhacfbWlFs,requirement)
(689098719489339392,One of my Hadoop consultant is looking for new project let me know if any one
 any open requirement for hadoop please…https://t.co/bhacfbWlFs,for)
(689098719489339392,One of my Hadoop consultant is looking for new project let me know if any one
 any open requirement for hadoop please…https://t.co/bhacfbWlFs,hadoop)
(689098719489339392,One of my Hadoop consultant is looking for new project let me know if any one
 any open requirement for hadoop please…https://t.co/bhacfbWlFs,please…https://t.co/bhacfbWlFs)
(689098788938584069,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のも不自然だし https://t.co/yczm80m2Bm,
RT)
(689098788938584069,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のも不自然だし https://t.co/yczm80m2Bm,
@shiumachi:)
(689098788938584069,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のも不自然だし https://t.co/yczm80m2Bm,
```

```
grunt> dictionary = load '/home/acadgild/Flume/AFINN.txt' using PigStorage('\t') AS(word:chararra
y,rating:int);
```

```
grunt> dump dictionary
```

File   Edit   View   Search   Terminal   Help

```
(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)
(yeees,2)
(yes,1)
(youthful,2)
(yucky,-2)
(yummy,3)
(zealot,-2)
(zealots,-2)
(zealous,2)
grunt>
```

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

```
grunt> describe word_rating;
```

```
word_rating: {tokens::id: bytearray,tokens::text: bytearray,tokens::word: chararray,dictionary::w
ord: chararray,dictionary::rating: int}
grunt>
```

```
grunt> rating = foreach word_rating generate tokens::id as id,tokens::text as text, dictionary::r
ating as rate;
```
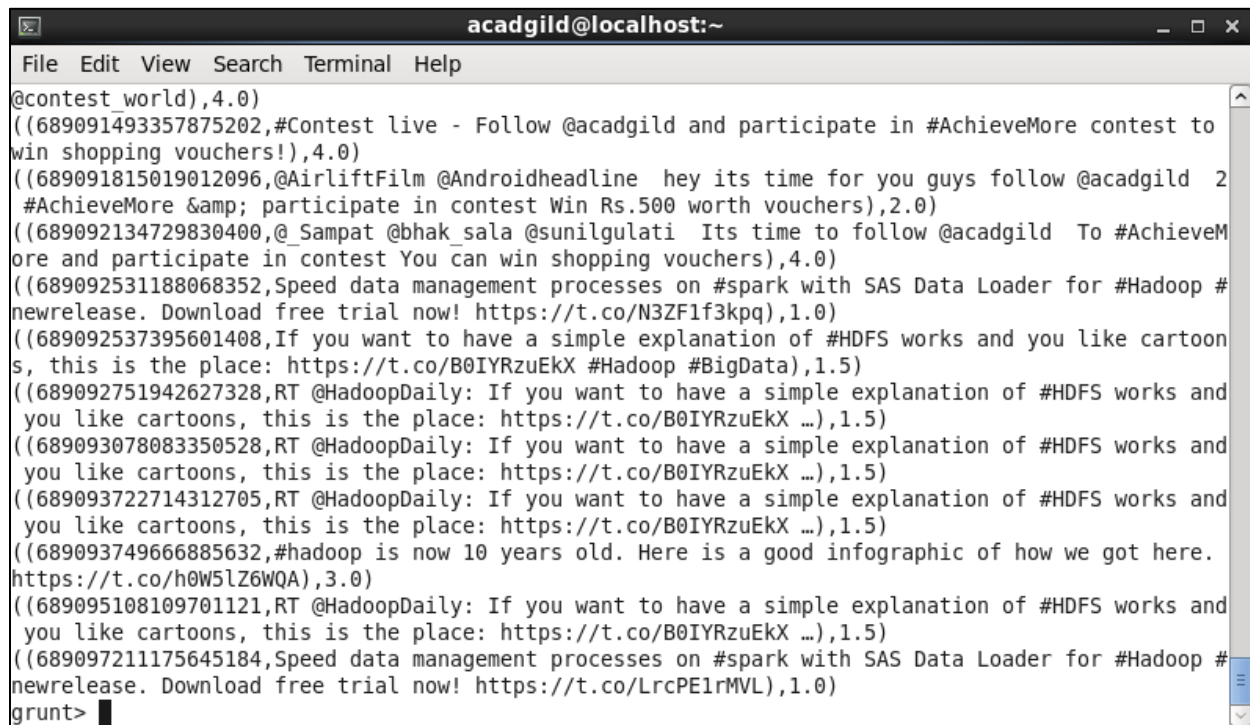
```
g IMPLICIT_CAST_TO_CHARARR
grunt> describe rating;
```

```
g USING_OVERLOADED_FUNCTION 1 time(s).
2016-07-16 14:28:44,560 [main] WARN  org.apache.pig.newplan.BaseOperatorPlan - Encountered Warnin
g IMPLICIT_CAST_TO_CHARARRAY 1 time(s).
rating: {id: bytearray,text: bytearray,rate: int}
```

```
grunt> word_group = group rating by (id,text);
2016-07-16-21:26:40,587 [main] WARN  org.apache.pig.ne
```

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

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

File   Edit   View   Search   Terminal   Help

```
@contest_world),4.0)
((689091493357875202,#Contest live - Follow @acadgild and participate in #AchieveMore contest to
win shopping vouchers!),4.0)
((689091815019012096,@AirliftFilm @Androidheadline  hey its time for you guys follow @acadgild  2
 #AchieveMore &amp; participate in contest Win Rs.500 worth vouchers),2.0)
((689092134729830400,@_Sampat @bhak_sala @sunilgulati  Its time to follow @acadgild  To #AchieveM
ore and participate in contest You can win shopping vouchers),4.0)
((689092531188068352,Speed data management processes on #spark with SAS Data Loader for #Hadoop #
newrelease. Download free trial now! https://t.co/N3ZF1f3kpq),1.0)
((689092537395601408,If you want to have a simple explanation of #HDFS works and you like cartoon
s, this is the place: https://t.co/B0IYRzuEkX #Hadoop #BigData),1.5)
((689092751942627328,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)
((689093078083350528,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)
((689093722714312705,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)
((689093749666885632,#hadoop is now 10 years old. Here is a good infographic of how we got here.
https://t.co/h0W5lZ6WQA),3.0)
((689095108109701121,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)
((689097211175645184,Speed data management processes on #spark with SAS Data Loader for #Hadoop #
newrelease. Download free trial now! https://t.co/LrcPE1rMVL),1.0)
grunt>
```