**Twitter Sentiment Analysis on movie Iron3**

# Load tweets data

tweetdata = LOAD '/home/acadgild/flume/twitter/tweetsp' USING PigStorage('\t') AS (tweetId:long, timestamp:chararray, text:chararray, timezone:chararray);

#split the tweet text into words and use FLATTEN to generate a un-nest the words into a new line

tweetwords = FOREACH tweetdata GENERATE tweetId, text, FLATTEN(TOKENIZE(text)) AS word, timestamp, timezone;

#Load the dictionary data where each word is rated from -5 to +5 based on its +ve or -ve meaning

dictionary = LOAD '/home/acadgild/flume/twitter/AFINN.txt' USING PigStorage('\t') AS (word: chararray, rating:int);

#Join tweet words with dictionary so that we can get rating for each word into the same relation

wordrating = JOIN tweetwords BY word LEFT OUTER, dictionary BY word USING 'replicated';

#Group the word ratings based based on tweetid(and text) so that we can group the rating for each tweet

grptweetrating = GROUP wordrating BY (tweetId, text);

#Now do AVG of rating for each tweetid(and text)

tweetrating = FOREACH grptweetrating GENERATE group, AVG(wordrating.rating) AS tweet\_rating;

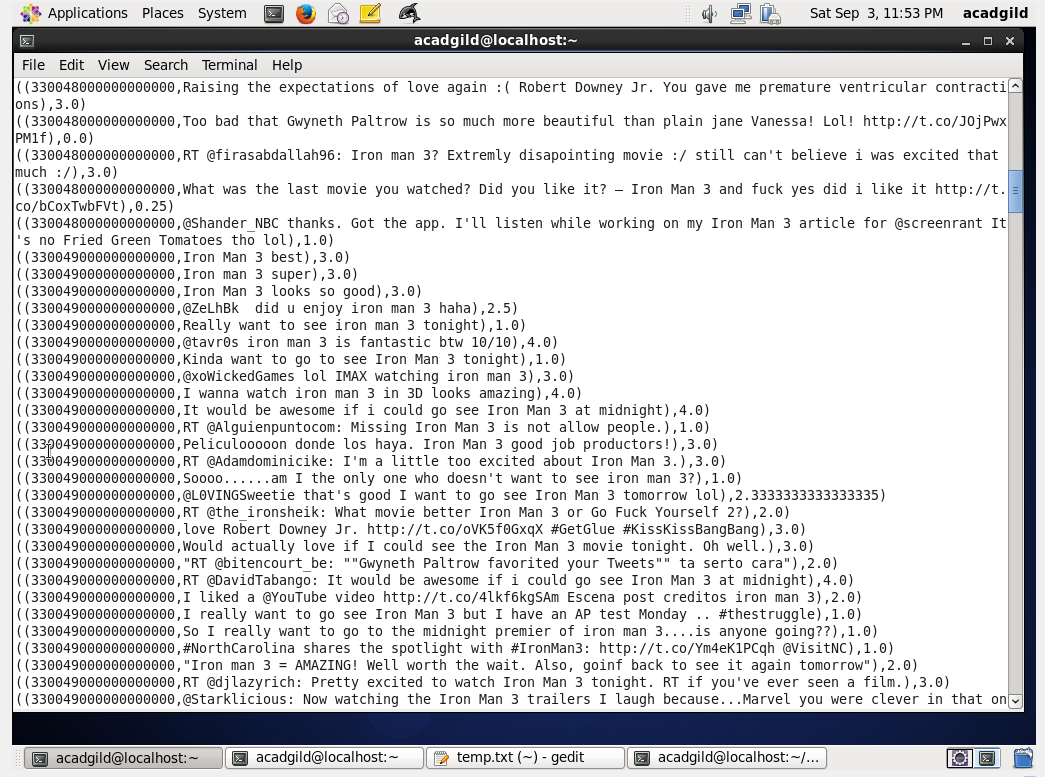
#Finally of tweet rating is >=0, then we call tweet is positive tweet

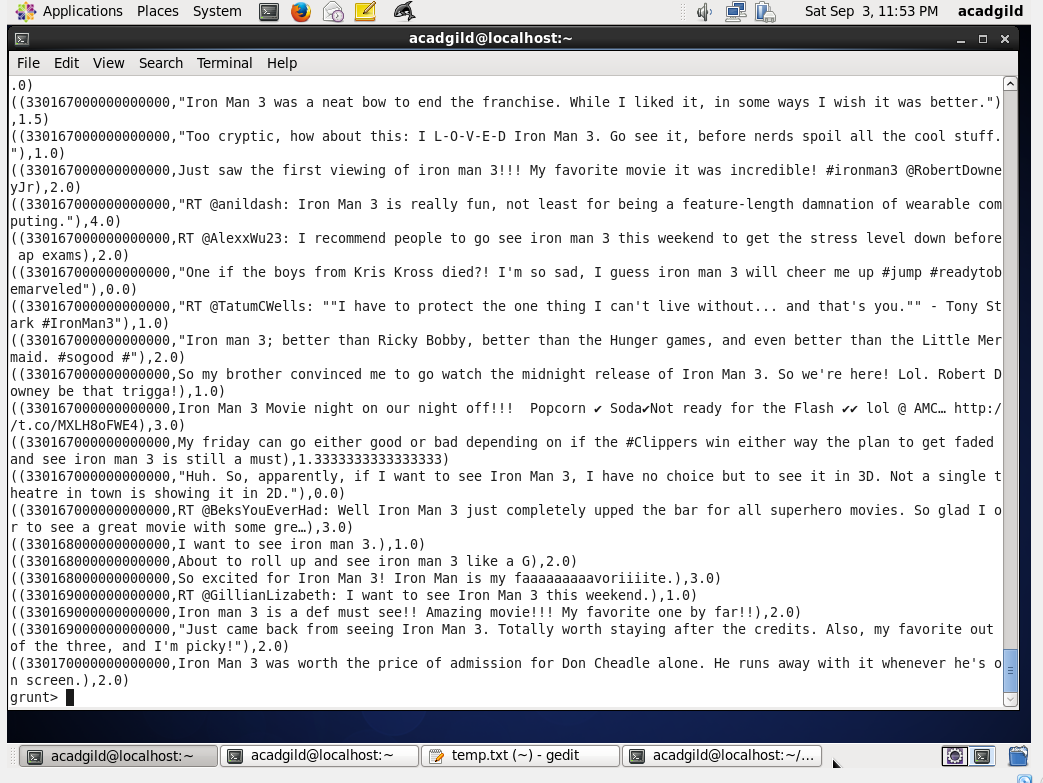
postive\_tweets = FILTER tweetrating BY tweet\_rating >=0;

#If rating < 0, then tweet is represent negative meaning

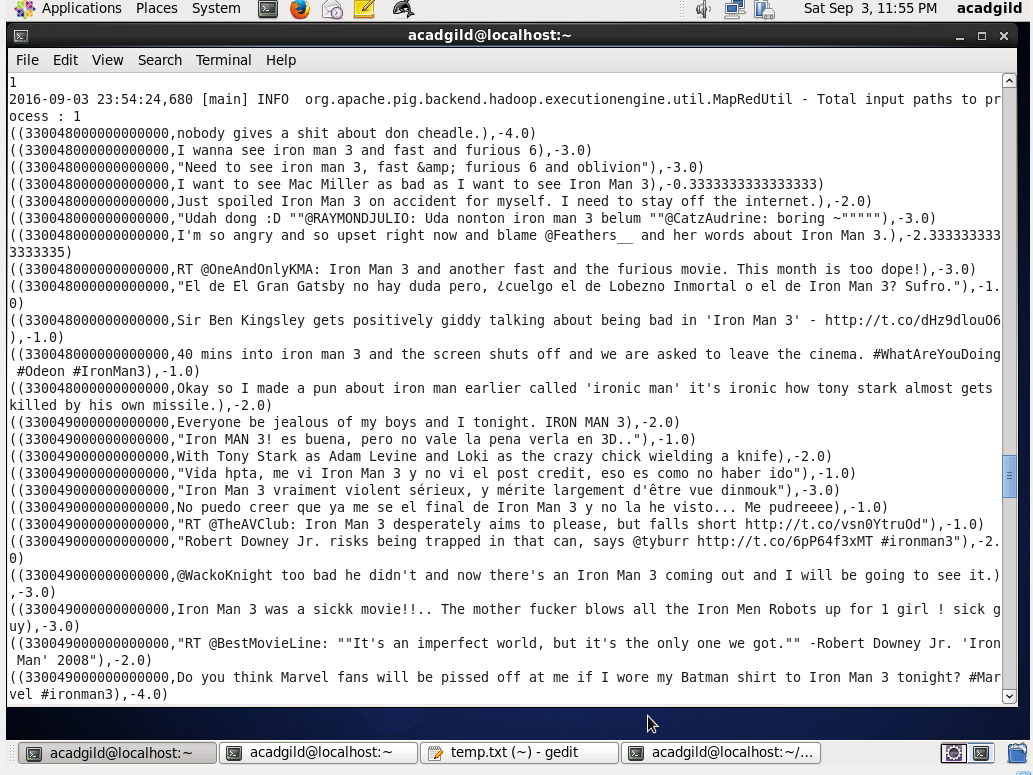
negative\_tweets = FILTER tweetrating BY tweet\_rating < 0;

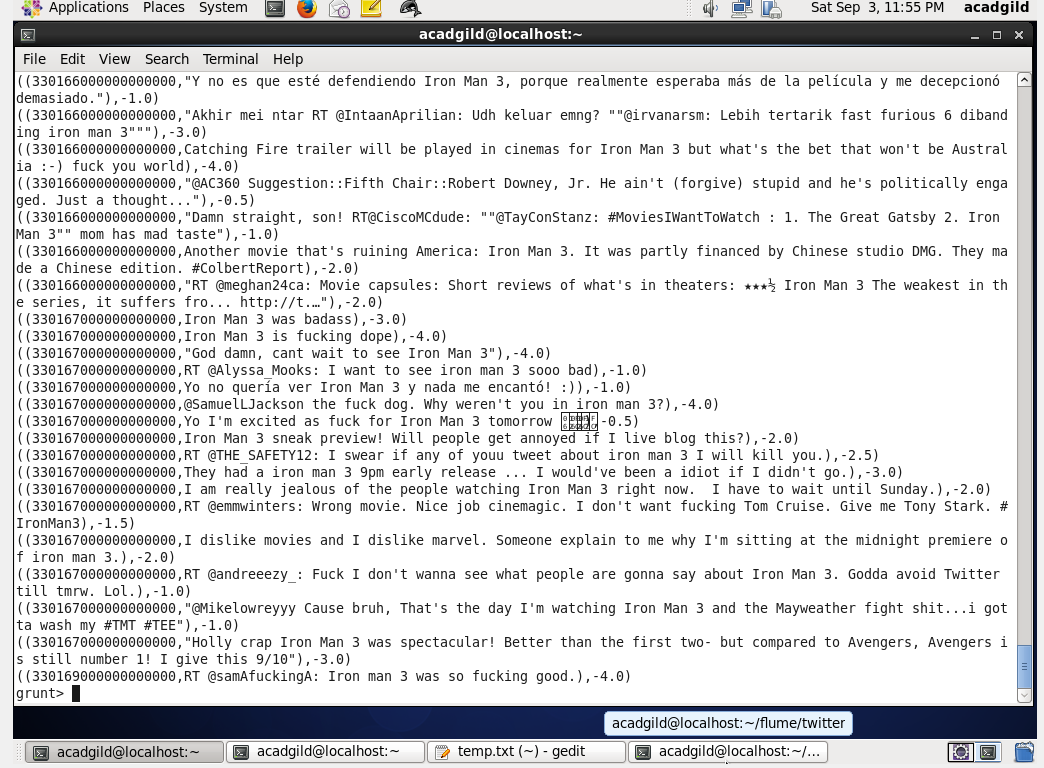
**Positive Tweets:**





**Negative Tweets:**





TimeZone based analysis:

# Load tweets data

timezonedata = LOAD '/home/acadgild/flume/twitter/time\_zone\_map.tsv' USING PigStorage('\t') AS (timezone:chararray, country:chararray);

#split the tweet text into words and use FLATTEN to generate a un-nest the words into a new line

tweetwords = FOREACH tweetdata GENERATE tweetId, FLATTEN(TOKENIZE(text)) AS word, timezone;

#Load the dictionary data where each word is rated from -5 to +5 based on its +ve or -ve meaning

dictionary = LOAD '/home/acadgild/flume/twitter/AFINN.txt' USING PigStorage('\t') AS (word: chararray, rating:int);

#Join tweet words with dictionary so that we can get rating for each word into the same relation

#Here we are doing map side join to effective use of join at Map side only

wordrating = JOIN tweetwords BY word LEFT OUTER, dictionary BY word USING 'replicated';

#Group the word ratings based based on tweetid so that we can group the rating for each tweet

grptweetrating = GROUP wordrating BY tweetId;

#Now do AVG of rating for each tweet and also

#Here we are doing the FLATTEN operation on time\_zone,

#because when grouping, time\_zone will be repeated numerous times as the time\_zone is present in every word of the tweet

tweetrating = FOREACH grptweetrating GENERATE FLATTEN(wordrating.timezone) AS timezone, AVG(wordrating.rating) AS tweet\_rating;

#Group by timezone

grptweetrating = GROUP tweetrating BY timezone;

#Finally do AVG of tweet rating by timezone

locrating = FOREACH grptweetrating GENERATE group, AVG(tweetrating.tweet\_rating) AS tweet\_rating;

