

PigLatin Queries

Enter 'Pig -x mapreduce' in linux shell to enter in grunt environment.

```
bitcoin_tweets = LOAD 'gs://dataproc-staging-us-central1-744395099175-nyv31aa8/bitcoin-tweets/Bitcoin_tweets.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'YES_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER') as (user_name:chararray, user_location:chararray,
user_description:chararray, user_created:Datetime, user_followers:int,

not_null_bt= FILTER bitcoin_tweets BY user_name!='' AND user_created is not null
AND text!='';

three_columns_bt = FOREACH not_null_bt GENERATE user_name, user_created,
text;

trimmed_three_columns_bt = FOREACH three_columns_bt GENERATE
TRIM(user_name) as user_name, user_created, TRIM(text) as text;

url_removed_bt= FOREACH trimmed_three_columns_bt GENERATE
user_name, user_created, text_clean.remove_url(text) as text;

emoji_removed_bt= FOREACH url_removed_bt GENERATE
user_name, user_created, text_clean.remove_emoji(text) as text;

remove_new_line_bt = FOREACH emoji_removed_bt GENERATE
user_name, user_created, REPLACE(text, '\n', ' ') as text;

comma_bt = foreach remove_new_line_bt generate REPLACE(user_name, ',', '') as
user_name, user_created, REPLACE(text, ',', '') as text;

store comma_bt into 'gs://dataproc-staging-us-central1-744395099175-nyv31aa8/
bitcoin-tweets/pig_Output' using PigStorage(',');
```