*/\*Sample Pig Queries for filtering of data.*

*In this part we have segregated the data on the basis of the tweets and retweets based on the geographic locations.*

*We first load the .csv file into HDFS in Hadoop as "/tweet2" input file\*/*

**tweet = LOAD '/tweet2' USING PigStorage(',') as (text:chararray, id:int, screenName:chararray, retweetCount:int, isRetweet:boolean, Countr1:chararray);***//reading the file into a relation "tweet" as tuples using the LOAD operator with "," as delimiter*

**tweet\_data = GROUP tweet BY Countr1;***//grouping the tuples according to the countries*

**count1= foreach tweet\_data generate group, COUNT($1) as cnt,$1;***//for all the tuple we again group them according to the tweets and retweets and keep a counter each for the number of tweets and retweets received in the form of bags*

**m1= foreach count1{**

**row= FILTER $2 BY(text=='1');**

**generate flatten(row);**

**};**// un-nesting of bags and converting into tuples using the "flatten" option

**countr = GROUP m1 BY Countr1;***//the counts of twwets and retweets obtained are further grouped according to the country names*

**count\_final= foreach countr generate group, COUNT($1) as cnt,$1;***//for each country we store the counts of tweets and retweets in the form of tuples*

**STORE count\_final INTO '/op1' using PigStorage(',');***//the tuples are stored in output file "/op1" and sent to HDFS from which the data is extracted and converted into .csv file*