create table tweet(json string);

load data local inpath 'Desktop/tweet\_cleaned.json' overwrite into table tweet;

create table tweetdetail(hashtag string, location string, count int, id string, uname string, text string, date string);

insert into table tweetdetail select get\_json\_object(tweet.json,'$.entities.hashtags.text') as hashtag, get\_json\_object(tweet.json,'$.user.location') as location,

get\_json\_object(tweet.json,'$.user.followers\_count') as count,

get\_json\_object(tweet.json,'$.id\_str') as id,

get\_json\_object(tweet.json,'$.user.name') as uname,

get\_json\_object(tweet.json,'$.text') as text,

to\_date(from\_unixtime(unix\_timestamp(get\_json\_object(tweet.json,'$.created\_at'),"EEE MMM dd HH:mm:ss zzzzz yyyy")))as date

from tweet;

a. What are the hashtags used and how many times each hashtag is used?

create table hashtag1 as

select usedht

from tweetdetail lateral view explode(split(substr(hashtag,2,length(hashtag)-2),',')) ht as usedht;

select usedht,count(\*) c from hashtag1 group by usedht order by c desc;

b. Which State have the most active users and how many tweets are posted by State?

select location,count(\*) c from tweetdetail group by location order by c desc;

select location,count(\*) c from tweetdetail group by location order by c desc limit 1;

c. Based on the user’s followers count, who are the top ten users who have tweeted?

select uname,max(count) maxc from tweetdetail

group by uname

order by maxc desc

limit 10;

d. What is the polarity score for each tweet that was posted? Does the tweet have a positive or negative sentiment?

create table dictionary(word string, scores int)

row format delimited

fields terminated by '\t';

load data local inpath 'Desktop/Dictionary.txt' overwrite into table dictionary;

Create table content as

Select date, id, uname, tword

From tweetdetail lateral view explode(split(REGEXP\_REPLACE(text, '[^a-zA-Z]+', ' '),' ')) a as tword;

create table text\_score as

select c.date, c.id, c.uname, sum(d.scores) as tweetscore

from content c join dictionary d on

lower(c.tword) = d.word

where d.scores IS NOT NULL

group by c.date,c.id,c.uname;

select \* from text\_score;