HIVE Commands

//Create Database in hive

CREATE SCHEMA db_tweets;

// verify the database

SHOW DATABASES;

//If not already executed, please execute following code.

//Table to Store Twitter JSON Data.

CREATE EXTERNAL TABLE db_tweets.table1_tweets(id BIGINT,created_at STRING,source STRING,favorited BOOLEAN, retweeted_status STRUCT<text:STRING, `user` :STRUCT<screen_name:STRING,name:STRING>,retweet_count:INT>,text STRING,entities STRUCT<hashtags:ARRAY<STRUCT<text:STRING>>>,`user` STRUCT<screen_name:STRING,friends_count:INT,followers_count:INT,statuses_count:INT,verified:BOOL EAN,utc_offset:INT,time_zone:STRING>,in_reply_to_screen_name STRING) ROW FORMAT SERDE 'org.apache.hive.hcatalog.data.JsonSerDe';

// Load the data from HDFS into Table.

load data inpath '/user/hduser/projectData/DIAdata/FlumeData_1620086272028' into TABLE db_tweets.table1_tweets;

// display 1st 5 rows just for view

select * from db_tweets.table1_tweets limit 5;

//View to filter only the hashtags text.

create view hash1 as select id, entities.hashtags.text as words from db_tweets.table1_tweets;

//Split multiple hashtags into single hashtags.

create view hash2 as select id, word from hash1 lateral view explode(words) dummy as word ;

//Count and store the trends in the final table.

```
create table db_tweets.tweets_Trend_final1 stored as orc as select count(*) as count, word as hashtags
from hash2 group by word order by count desc;
```

```
select * from db_tweets.tweets_Trend_final1 where NOT hashtags = "travel" limit 15;
```

_____