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

//Table to Store Twitter JSON Data.

CREATE EXTERNAL TABLE covid\_raw\_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:BOOLEAN,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 Path to the above created Table.

load data inpath '/user/maria\_dev/covid\_data\_day (1)' into TABLE covid\_raw\_tweets;

//View to filter only the hashtags text.

create view hash1 as select id, entities.hashtags.text as words from covid\_raw\_tweet;

//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 tweets\_Trend\_final stored as orc as select count(\*) as count, word as hashtags from hash2 group by word order by count desc;

//Ignore COVID19, because that is our keyword.

Select \* from tweets\_Trend\_final where NOT hashtags = ‘COVID19’;