**Finding Recent Trends and Most Popular hashtags**

Recent trends from streamed tweets can also be found using Hive queries. Since tweets collected from twitter are in JSON format, we have to use JSON input format to load the tweets into Hive. We have used Cloudera Hive JsonSerDe for this purpose. This jar file has to be present in Hive to process the data [9]. It can be added using following command. This command is used in Hive interface.

hive> add jar /usr/lib/hive/lib/jsonserde-with-dependencies.jar

Following steps are performed to find the recent trend:

## *Loading and Feature extraction*

The tweets collected from twitter are stored in HDFS. In order to work with Data stored in HDFS using HiveQL, first an external table is created which creates the table definition in the Hive Metastore. This query not only creates a schema to store the tweets, but also extracts required fields like id and entities.

## *Extracting Hashtags*

In order to extract actual hashtags from entities, we created another table which contains id and the list of hashtags. Since multiple hashtags are present in one tweet, we used UDTF (User Defined Table Generation Function) to extract each hashtag on the new row. The outcome of this phase is id and hashtag.

## *Counting hashtag*

After performing all the above steps, we have id and hashtag text. A hive query is written to count the hashtags.

## *Query for top trends*

For getting most popular hash tags or to find top trend we are use following hive query:

hive> CREATE VIEW L5 AS SELECT

ht,

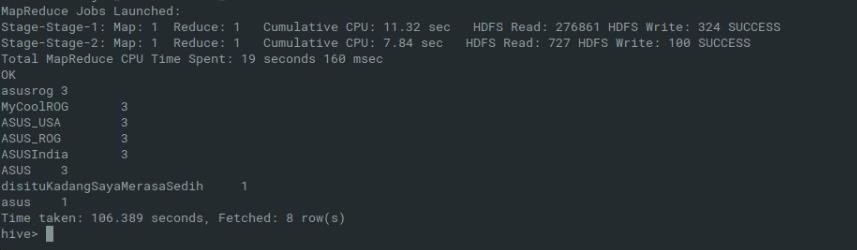
count(ht) AS countht

FROM raw\_f LATERAL VIEW

EXPLODE(entities.hasgtags.text) dummy AS ht

ORDER BY countht DESC limit 10;

**Output of the query is:**



*Fig: Most popular hashtags*

In the output we can see the popular hashtags are shown with the number of times they have occurred in our data set. For example, hashtag ASUS India has occurred 3 times in all of our tweets. The hashtags are arranged in decreasing order of their total occurrences.