Implement the below blog at your end and send the complete documentation.

https://drive.google.com/file/d/0B\_Qjau8wv1KobUlaOEtfNEtQNkU/view?usp=sharing

3. Output

Documentation containing code and screenshots

Counting the popular hashtags.

Code

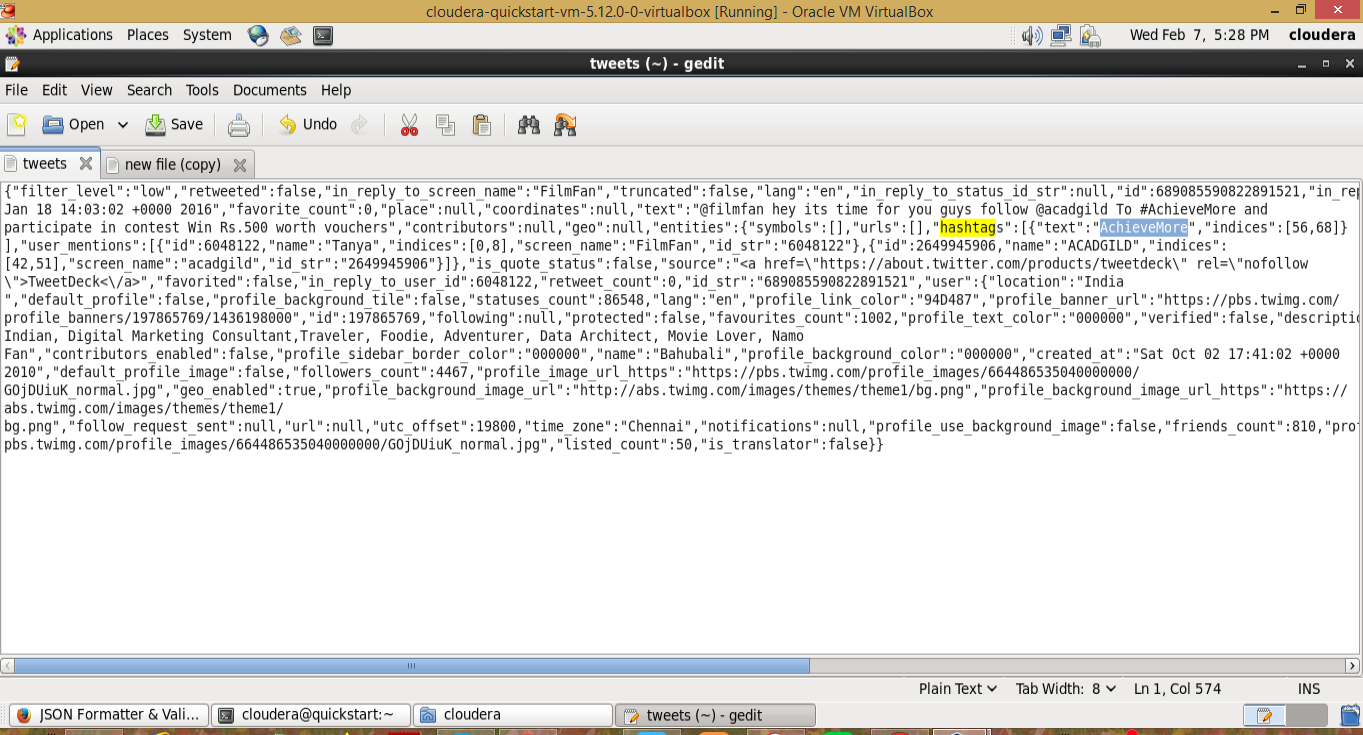
*val tweets = spark.read.json("[file:///home/cloudeaa/tweets](../../datasets/tweets)").registerTempTable("tweets")*

*val hashtags = spark.sql("select id as id,entities.hashtags.text as words from tweets").registerTempTable("hashtags")*

*val hashtag\_word = spark.sql("select id as id,hashtag from hashtags LATERAL VIEW explode(words) w as hashtag").registerTempTable("hashtag\_word")*

*val popular\_hashtags = spark.sql("select hashtag, count(hashtag) as cnt from hashtag\_word group by hashtag order by cnt desc").show*

Input file



Output

