Assignment 23.1

Step1: Load the tweets

Load the tweets from tweets\_json\_data from ("/home/acadgild/assignment\_23.1/tweets\_json\_dataset using sqlContext.read.json and populate temporary table tweets

val tweets = sqlContext.read.json("/home/acadgild/assignment\_23.1/tweets\_json\_dataset").registerTempTable("tweets")

Step2: Select id, words from tweets

Select id and words from tweets table and store the result to tables hashtags

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

Step3: Explode words

Explode words and store into table hashtab\_word

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

Step4: Count hashtags

Count hashtags from hashtag\_word table group by each hashtag, sort by count hashtags

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

Screenshot is as below:

