Counting popular hashtags using Spark sql

In this assignment we are analyzing some popular hashtags in twitter using Spark Sql.
We have collected the data in a json file and will be using same as source.

Below is the code used to find the popular hashtags-

- val tweets = spark.read.json("/home/acadgild/Assignment-21/tweets.json").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

Now we will see each line of code one by one-

First we are reading the json file that we have kept in local file system and registering it as temporary table named as tweets-

```
scala> val tweets = spark.read.json("/home/acadgild/Assignment-21/tweets.json").registerTempTable("tweets")
warning: there was one deprecation warning; re-run with -deprecation for details
18/01/10 20:07:43 WARN Utils: Truncated the string representation of a plan since it was too large. This behavi
Fields' in SparkEnv.conf.
tweets: Unit = ()

scala>
```

Now from above temporary table we are selecting id, and hashtags and again registering it as another temporary table named as hashtags-

```
scala> val hashtags = spark.sql("select id as id,entities.hashtags.text as words from tweets").registerTempTable("hashtags")
warning: there was one deprecation warning; re-run with -deprecation for details
hashtags: Unit = ()

scala> val hashtag_word = spark.sql("select id as id,hashtag from hashtags LATERAL VIEW explode(words) w as hashtag").registerTempTable("hashtag_word")
warning: there was one deprecation warning; re-run with -deprecation for details
hashtag_word: Unit = ()
```

Finally we are selecting the hashtag and count of it from above created temporary table and displaying it-

```
scala> val popular_hashtags = spark.sql("select hashtag, count(hashtag) as cnt from hashtag_word group by hashtag order by cnt desc").show
+--------------------+---+
|             hashtag|cnt|
+--------------------+---+
|         AchieveMore| 15|
|              Hadoop|  7|
|             bigdata|  4|
|                data|  3|
|          masterdata|  1|
|          GartnerEIM|  1|
|           jobsearch|  1|
|          WhitePaper|  1|
|             contest|  1|
|                jobs|  1|
|         dataquality|  1|
|    chiefdataofficer|  1|
|             BigData|  1|
|informationgovern...|  1|
|             OReilly|  1|
|      Virtualization|  1|
|          Infonomics|  1|
|               Spark|  1|
```