

Session 23 Assignment 1

To count popular hashtags.

Solution:

Kept the sample tweet under `/home/bigdata/deepak/docs/Acadgild/Session23Assignment1/tweets`

Read the tweet file

```
scala> val tweets = spark.read.json("file:///home/bigdata/deepak/docs/Acadgild/Session23Assignment1/tweets").registerTempTable("tweets")
warning: there was one deprecation warning; re-run with -deprecation for details
2018-06-05 06:25:47 WARN ObjectStore:568 - Failed to get database global_temp, returning NoSuchObjectException
2018-06-05 06:25:53 WARN Utils:66 - Truncated the string representation of a plan since it was too large. This behavior can be adjusted by setting 'spark.debug.maxToStringFields' in SparkEnv.conf.
tweets: Unit = ()
```

Execute the below commands to get the count of the 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 = ()

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| 1|
+-----+

popular_hashtags: Unit = ()
```