**Counting Popular Hashtags using Spark SQL**

**[tweets.json file is considered which contains Hashtag data]**

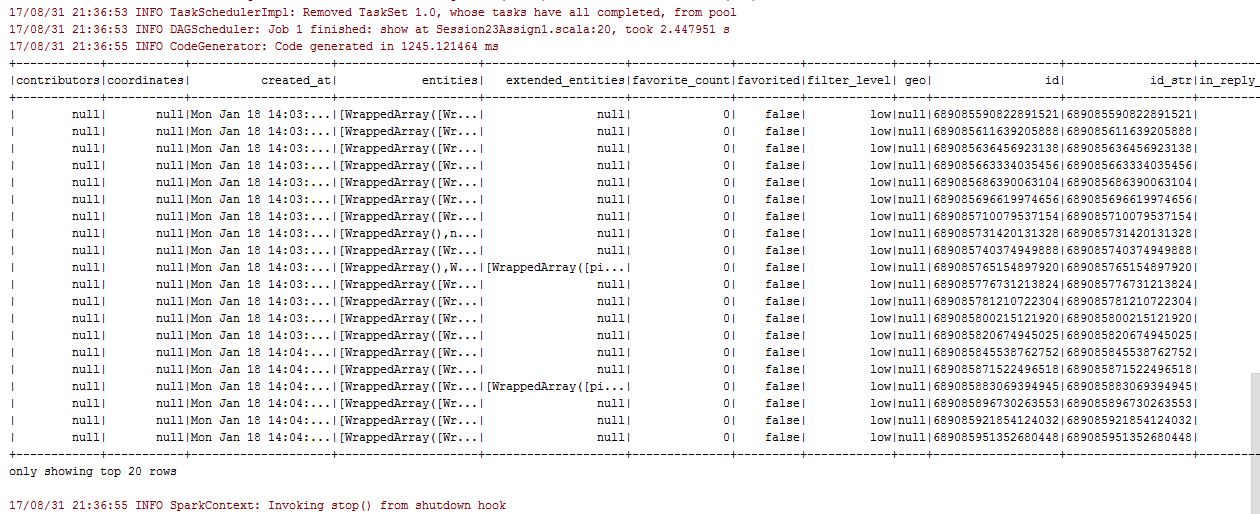
**Source Code:**

**import** org.apache.spark.sql.{Column, Row, SQLContext, SparkSession} *//Explanation is already given in Assignment18.1***object** Session23Assign1 **extends** App {  
 **val** *spark* = SparkSession.*builder*()  
 .master(**"local"**)  
 .appName(**"Session23Assign1"**)  
 .config(**"spark.sql.warehouse.dir"**, **"file:///G:/ACADGILD/course material/Hadoop/Sessions/Session 23/Assignments/Assignment1"**)  
 .getOrCreate()  
 *//Explanation is already given in Assignment 18.1  
  
 //setting path of winutils.exe* System.*setProperty*(**"hadoop.home.dir"**, **"F:/Softwares/winutils"**)  
 *//winutils.exe needs to be present inside HADOOP\_HOME directory, else below error is returned:  
 //error: java.io.IOException: Could not locate executable null\bin\winutils.exe in the Hadoop binaries.* **val** *tweets* = *spark*.*sqlContext*.read.json(**"file:///G:/ACADGILD/course material/Hadoop/Sessions/Session 23/Assignments/Assignment1/tweets.json"**)  
 *tweets*.createOrReplaceTempView(**"tweets"**)

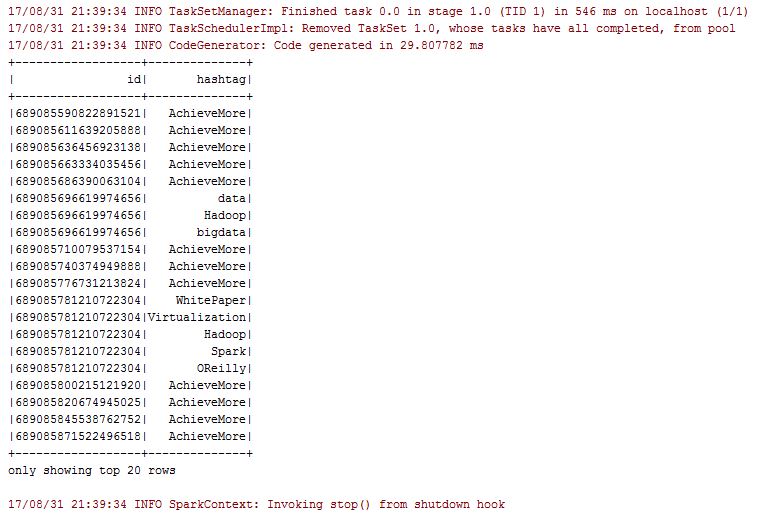
*spark*.sql(**"select \* from tweets"**).show()  
 *//REFER Screenshot 1 for OUTPUT* **val** *hashtags* = *spark*.sql(**"select id as id,entities.hashtags.text as words from tweets"**)  
 *hashtags*.createOrReplaceTempView(**"hashtags"**)

*spark*.sql(**"select \* from hashtags"**).show()  
 *//REFER Screenshot 2 for OUTPUT* **val** *hashtag\_word* = *spark*.sql(**"select id as id,hashtag from hashtags LATERAL VIEW explode(words) w as hashtag"**)  
 *hashtag\_word*.createOrReplaceTempView(**"hashtag\_word"**)  
 *spark*.sql(**"select \* from hashtag\_word"**).show()  
 *//REFER Screenshot 3 for OUTPUT* **val** *popular\_hashtags* = *spark*.sql(**"select hashtag, count(hashtag) as cnt from hashtag\_word"** +  
 **" group by hashtag"** +  
 **" order by cnt desc"**)  
 *popular\_hashtags*.show()  
 *//REFER Screenshot 4 for OUTPUT*}

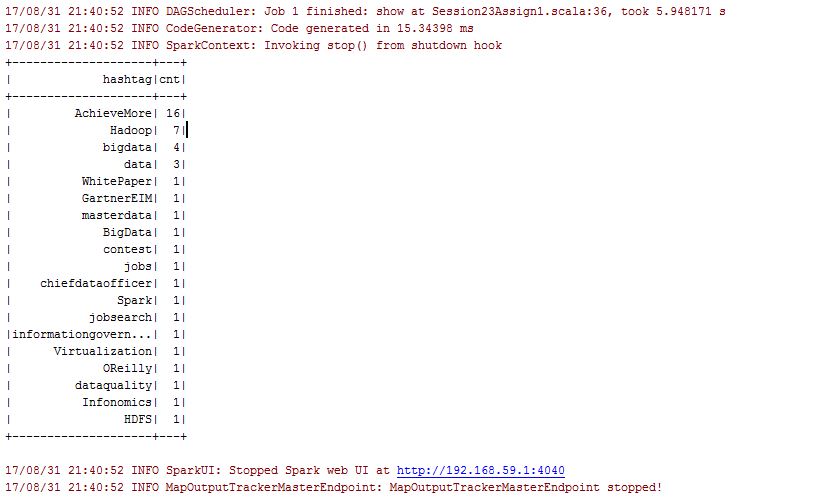
**OUTPUT**

**Screenshot 1**

**Screenshot 2**

**Screenshot 3**

**Screenshot 4**

****