SESSION 21 ASSIGNMENT 1

**Problem Statement:**

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

<https://drive.google.com/file/d/0B_Qjau8wv1KobUlaOEtfNEtQNkU/view?usp=sharing>

**Input:**

Tweet.json

**Solution:**

import org.apache.log4j.{Level, Logger}

import org.apache.spark.sql.SparkSession

object Assignment\_21\_1 extends App {

Logger.getLogger("org").setLevel(Level.OFF)

Logger.getLogger("akka").setLevel(Level.OFF)

val spark = SparkSession.builder()

.master("local")

.appName("example")

.config("spark.sql.warehouse.dir", "file:///C:")

.getOrCreate()

val sqlContext = spark.sqlContext

// Create the DataFrame

val tweets = sqlContext.read.json("C:/ACADGILD/Big Data/tweet.json").registerTempTable("tweets")

// SQL statements can be run by using the sql methods provided by sqlContext.

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

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

// Displays the content of the DataFrame to stdout

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

}

**Output:**



