Assignment 21.1

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

Solution:-

Below is the sample tweets:-

Source Code:-

- val tweets = spark.read.json("file:///home/acadgild/hadoop/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

Screenshots:-