Session 21 : SPARK

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>

**Topic:** Counting popular hashtags using Spark sql

**Solution:-**

**Tweet File - **

Starting Spark-shell

Command – spark-shell



Loading tweet in spark

**scala>** val tweets = spark.read.json("tweet.txt").registerTempTable("tweets");

**scala>** val hashtags = spark.sql("select id as id,entities.hashtags.text as words from tweets").registerTempTable("hashtags");

**scala>** val hashtag\_word = spark.sql("select id as id,hashtag from hashtags LATERAL VIEW explode(words) w as hashtag").registerTempTable("hashtag\_word");



Getting hashtag count-

**scala>** val popular\_hashtags = spark.sql("select hashtag, count(hashtag) as cnt from hashtag\_word group by hashtag order by cnt desc").show;

