**Assignment #4**

|  |
| --- |
| **Twitter Tweets Sentiment Analysis**  The goal of this assignment is to leverage Spark Streaming component to consume Twitter data and perform sentiment analysis on it.   * For each tweet, break the message into tokens, then remove punctuation marks and stop words * Simple sentiment analysis   + Determine the score of whether a tweet has a positive, negative or neutral sentiment * A list of positive and negative words are provided * Display the sentiment score and tweet message to start out with   + For debugging purpose * Maintain # of positive, negative and neutral sentiment counts and print out them using window length of 10 and 30 seconds (two separate windows, not sliding window)   **Preparation:**   * See details at this [github](https://github.com/hienluu/twitter-structured-streaming" \t "_blank) project * Download the file **twitter-streaming-master.zip**by clicking  [here](https://github.com/hienluu/twitter-structured-streaming/archive/master.zip)   **Create a Twitter account:**   * + **https://dev.twitter.com/#**   + **https://apps.twitter.com/**   **Instructions for download and setup assignment:**   * Bring up "TwitterSourceExample.scala" file   + Right mouse click and select "Run As"->"Scala Application" option * Changing log level   + Bring up file log4j.properties in src/main/resources folder   + log4j.rootCategory=[ERROR|WARN|DEBUG], console   **Submission:**   * Submit the Scala file   **Resources:**   * [Real Time Streaming with Spark blog](http://zdatainc.com/2014/08/real-time-streaming-apache-spark-streaming/) * [zdata-ince - Spark Streaming Github Project](https://github.com/zdata-inc/SparkSampleProject) * [Twitter tweet schema](https://dev.twitter.com/overview/api/tweets)   **Expected Output:**  Tweet sentiments in last 10 seconds  Count=40 (NEUTRAL tweets)  Count=10 (NEGATIVE tweets)  Count=6 (POSITIVE tweets)  Tweet sentiments in last 30 seconds  Count=10 (NEGATIVE tweets)  Count=6 (POSITIVE tweets)  Count=40 (NEUTRAL tweets)  **Old stuff**:   * ***(Only for Windows Users)*** SBT - <http://www.scala-sbt.org/release/docs/Setup.html>   + Only needed for Windows laptop     - **Install** winutils.exe - (instructions are at [here](https://jaceklaskowski.gitbooks.io/mastering-apache-spark/spark-tips-and-tricks-running-spark-windows.html))     - Make sure sbt is add it to path |