

Assignment Instructions

Resources:

1. www.tutorialspoint.com/apache_spark/index.htm
2. www.monitorware.com/en/logsamples/apache.php
3. spark.apache.org (<http://spark.apache.org/docs/latest/quick-start.html> is very good place to start

HOW TO START THE SHELL

`/usr/bin/spark-shell`

`/etc/spark/conf`

TO USE HDFS (`hdfs:///user/cloudera/ ...`) and for local file (`file:///home/cloudera/`) in the RDD creation.

Example: `scala> var lines = sc.textFile("file:///home/cloudera/workspace/testData")`

Part 6. [5 points] Do Spark Project in Scala. See the attached pdf for project description.

Part 7. [5 Point] In class Presentation.

Part 8. [3 + 2 point] Submit your Presentation (doc/pdf/ppt only) (Sakai) and **uninstall what you had installed from University machines by 4:00 PM FRIDAY AFTER YOUR FINAL EXAM. This deadline is for uninstall alone.**