

Write Up

Assignment 8

- The main output of the program is as follows:
 - The variance calculated by my algorithm is:
 - myVar: org.apache.spark.rdd.RDD[Double] = MapPartitionsRDD[1] at map at <console>:29
 - res2: org.apache.spark.rdd.RDD[RunningVar] = MapPartitionsRDD[2] at mapPartitions at <console>:35
 - res3: Double = **873.74040000000001** (The variance my program calculated)
 - The variance calculated by the built in function is:
 - realVar: Double = 873.7403999999999
- The program is run using:
 - spark-shell --master yarn -i A8.scala
- The directory that the program and output from console is in (Which professor Chatree said could just be the output of running the program) Output8 in my hdfs.