Assignment 8

Use Spark to calculate a variance of a collection of 100 numbers, randomly generated in range 1-100.
It must be in one-pass, ie. without means, see lecture note Spark Part 2, slide #5
You can use the code below as the starting point

```
class RunningVar {
 // declare var here
 var …
 // Compute initial variance for numbers
 def this(numbers: Iterator[Double]) {
    numbers.foreach(this.add(_))
 }
 // Update variance for a single value
 def add(value: Double) {
        …
 }
 // Merge another RunningVar object and update variance
 def merge(other: RunningVar) = {
    …
 }
}

val doubleRDD = //a collection of 100 numbers, randomly generated from 1-100
doubleRDD
.mapPartitions(v => Iterator(new RunningVar(v)))
.reduce((a, b) => a.merge(b))
```

The spark executable is at /opt/spark/bin/

Submit

1. Your code
2. The console output after you run it
3. Send an email to Long our TA – (LHN4@pitt.edu) and cc me (chatree@pitt.edu), with the following:
   a. #2 - output.txt
   b. The location of your output directory
   c. Put "CS1699: Assignment 8" in the subject line