

CS-E4110 Concurrent Programming
Autumn 2016 - Tutorials
Scala Parallel Collections and Spark

28th November 2016

Task 1

To demonstrate Scala parallel collections, create a program based on the Scala parallel collections that estimates the value of π by using 100000 random samples of the unit square (with corners in origo (0,0) and (1,1)), and counting how many of them fall in the unit circle. The amount of random points sampled this way should converge to $\pi/4$, and this can be used to estimate π .

See for example “Monte Carlo method for π ” at:

<http://mathfaculty.fullerton.edu/mathews/n2003/montecarlopiomod.html>

Show your solution Scala code.

Task 2

This is a demonstration task. Doing Spark programming is not part of the exam requirements.

Port the solution of Task 1 to Spark, and show how the approach can be done in parallel using it instead of Scala parallel collections. The code should use four partitions to allow for four way parallelization of the Scala code.

Show your solution Scala code. Also show how the code is run on Spark as a standalone compiled application.

Task 3

This is a demonstration task. Doing Spark programming is not part of the exam requirements.

Using Spark command line shell, do a wordcount of the words in the Finnish national epic Kalevala, available from: <http://www.nic.funet.fi/pub/doc/literary/finnish/kalevala/kalevala.txt.gz>

Show the Scala code that is needed to do the wordcount using the Spark shell. Also show how the code is run using the Spark shell.