The homework questions are due at the 23:59 on Saturday 5 November. Please turn in source codes, compilation, submission scripts used and also output files. Please cite any references you use.

1 Parallel R and clustering

Homework question 1

a) Install PBD R on Rocket (you can follow [https://github.com/Digitaalhumanitaaria/biohpc](https://github.com/Digitaalhumanitaaria/biohpc) and try the example clustering program. How long does it take to run on 1, 2, 4, 8 and 16 nodes? (You may need to get average run times for randomly generated data sets).

b) Write your own parallel k-means clustering program or use the one provided as a basis for parallelization. Compare the speeds to those of PBD R.

Homework question 2

Summarize and review [1]. What parallel frameworks do they also introduce and how do they differ from MPI?

2 OpenCL

Homework question 3

a) Review the short introduction to OpenCL at [https://github.com/andreyrozumnyi/intro-to-javacl-and-pyopencl](https://github.com/andreyrozumnyi/intro-to-javacl-and-pyopencl) and do the three exercises.

b) What fraction of peak performance does the matrix multiplication code get on the hardware you use?
c) Give suggestions for what you might do to get higher performance.

**Homework question 4**
Write your own OpenCL k-means program and run it on two different devices (e.g., your phone, laptop CPU, laptop GPU, desktop CPU, desktop GPU, Rocket Xeon or Rocket Xeon Phi). How does the performance compare between the two devices? (if you cannot write your own OpenCL k-means program, partial credit will be given for describing and running a program you have found online).

**Homework question 5**
With reference to some of the available scientific literature, explain why autotuning may be useful for OpenCL applications running on a variety of devices.

**References**