



High Performance Computing in Bioinformatics

| | |
|-----------------|--|
| Course code: | INF9380 |
| Time and place: | April 18-29 2016, University of Oslo, Norway |
| Lecturers: | Torbjørn Rognes, Abdulrahman Azab, Merete Molton Worren ++ |
| ECTS: | 5 |

Course description

This course focuses on the application of high performance computing (HPC) to bioinformatics analysis. The main target is to provide a background on how to effectively use HPC clusters for running computationally- or data-intensive bioinformatics applications. The course will mainly include teaching students selected bioinformatics tools and workflows, and how to use HPC platforms to speed up and maximize the overall throughput of intensive bioinformatics analysis. This will include how to optimize the use of available compute nodes, and how to adapt the application to the available resources on each compute node. The course will cover both how to efficiently use parallelism when writing your own programs, as well as how to adapt and wrap existing tools in a manner that efficiently exploits resources available on parallel architectures.

Course program

This is an intensive two-week course with lectures and hands-on exercises 7 hours a day, Monday to Friday, 9-17, with 1 hour lunch break. About half the time with lectures and the other half with exercises. In total, lectures and exercises are estimated to 70 hours.

- Self-study / reading of curriculum is estimated to 30 hours.
- Preparation time for written report (home exam) is estimated to 40 hours.
- The total workload for the course is estimated to 140 hours.

Prerequisites

Basic unix competence and basic knowledge of bioinformatics applications is required. Basic programming skills, preferably in Python.

Learning outcome and competence

After finishing the course the students should know:

- Resource intensive bioinformatics tools for, e.g. assembly, mapping/alignment, and multiple alignment. This would include the use of command-line tools and portal based tools, e.g. Galaxy.
- How those tools work, how this would influence the runtime, and the possibility of parallelizing the computation.
- When to use parallelization and distribution.
- The basic structure of HPC clusters, and how to run jobs on a cluster.
- How to evaluate the use of resources on a cluster, and how to optimize the use of memory and CPUs

Curriculum

- Lecture slides
- Parts of the book “Introduction to High-Performance Scientific Computing” by Victor Eijkhout (CC BY 3.0) (<http://pages.tacc.utexas.edu/~eijkhout/istc/istc.html>)

Evaluation

Practical student project (home exam) with hand-in of written report.

Note: Traveling and accommodation expenses shall be covered by NORBIS for the PhD students affiliated to our school.