

Objectives of the workshop HPC

This is an entry level workshop, aimed at reserachers with little to reasonable programming background, but no significant experience with the usage of a high performance computer.

You will learn how to run a your programs in parallel on a computer cluster. A computer cluster is a system with a lot of memory and a lot of processors which you share with hundreds of other users. A scheduler keeps track of the resources available and decides when to run your programs. You will learn how to write scripts telling the scheduler which programs to run and which resources are needed and for how long.

The operating system of a computer cluster will nearly always be Linux. The basic Linux concepts and commands will be introduced to you. But before you can learn Linux, you have to make a connection to the cluster. We will tell you how to login from your desk- or laptop on a cluster with a terminal emulation program. Editing and transferring files between your computer and the remote cluster is also on the menu of this workshop.

The programming examples are in R. So, basic understanding of R and RStudio is required. Some experience with text editing and command line interfaces is a plus.

Content of the workshop HPC

Brief Introduction into High Performance Computing

Without delving to deep in the theory of HPC the most important concepts and characteristics of cluster computers and computing are explained:

Computer cluster infrastructure

- Nodes & cores
- Memory & Storage
- Batch Scheduler
- Inlog node
- Scratch space

Parallel Execution of programs

- Pleasingly Parallel
- Parallel Programming
- Advanced Parallel Programming Techniques

This part we will end discussing for what types of computing jobs cluster computing is suited.

<hand-outs>

Example program in R: Recognizing Hand-written Digits with Machine Learning

Throughout the workshop we will use a small program to solve the challenge of recognizing handwritten digits by a computer. This is a rather famous example in Machine Learning. We will run an R script on your local computer and notice why and when we are in need of an HPC.

<course material>

Getting started with LISA

Before you can start working on LISA, you have to register your account, and install some software packages on your computer. This software is necessary to login on LISA (terminal session) and to transfer file s between your workstation and LISA. You also learn basic Linux commands. There are differnt software packages for

Mac users and Window users. On LISA you all work with the same operating system: LINUX. Linux on Lisa doesn't support a graphical user interface (GUI). We will teach you how to edit text files on a window-less system.

<course material>

Your first program on Lisa

Basic scheduler commands

Submitting an R script

Watching the output

Pleasingly parallel: finding the best hyper parameter settings

- generating a submission script
- generating batch scripts
- collecting the results

Enhancements to your

- logfiles
- email notifications
- scratch space

Start of second session

- Leftovers from previous session
- Q&A

Parallel programming in R

R packages * parallel * foreach * do parallel * batchtools

BLAS

Advanced Parallel Programming with MPI

Wrap up and Evaluation