# Introduction to Linux for HPC

## Lecturer: Maria Chiara Debernardi

### Language

English

### Course description and objectives

The main objective of this short, hands-on course is to provide participants with the essential Linux and High-Performance Computing (HPC) skills needed to effectively run Python-based data science and AI workloads on Bocconi's shared HPC facility. Through guided exercises on the actual cluster, students will learn how to connect securely, navigate the Linux filesystem, manage files, submit and monitor batch jobs with SLURM, and work efficiently on the command line when working with large datasets and computationally intensive models.

At the end of the course, participants will be able to:

- access Bocconi's shared HPC facility via SSH and understand the difference between login nodes and compute nodes
- navigate the Linux filesystem on the cluster and manage files and folders (create, move, copy, remove) safely from the command line
- transfer data between their local computer and the HPC system
- submit, monitor, and cancel batch jobs with SLURM, and interpret the resulting log and output files
- use software modules and basic Python environments to run their own code reproducibly on the cluster
- search for files and filter text efficiently using Linux tools and basic regular expressions

### Audience

The course is open to all Bocconi students. In particular, it is intended for students enrolled in the Master of Science programs in the Data Science and Business Analytics (**DSBA**) or Artificial Intelligence (**AI**), as well as **third-year** students in the Bachelor's programs in **BAI** and **BEMACS**, who want to learn the Linux operating system to access and use Bocconi's HPC resources.

## Prerequisites

No prior Linux knowledge is required, but basic coding experience in Python is assumed. It is advisable to be familiar with basic computer operations, such as file management.

## Guidelines

**Registration:**
You can sign up for the course only through the yoU@B student Diary, in the "**Sign-up for various activities**" box (please note that the box appears only when registrations open. Before then it will not be visible).
You can only cancel your registration by Diary **no later** than the registration deadline for the course itself. No other methods of cancellation are allowed.
Registration will be confirmed a few days before the start of the course through a message posted in the yoU@B student Diary.

**Attendance:**

- Attendance of **75% or more** of the lessons: obtainment of the **Open Badge**
- Attendance of **less than 25%** of the lessons: placement on **Exclusion List**

## Duration

8 academic hours

## Teaching mode

The course can be attended **either in person or remotely** via live streaming of the lesson held in the classroom. The **recording** will be made available on Blackboard shortly after each class for those who wish to watch it later or review the content.

## Calendar

| Lecture | Date | Time | Room |
|---|---|---|---|
| 1 | Fri 27/03/2026 | 14.45 - 16.15 | 301 (Sarfatti) |
| 2 | Fri 03/04/2026 | 14.45 - 16.15 | 301 (Sarfatti) |
| 3 | Fri 17/04/2026 | 14.45 - 16.15 | 301 (Sarfatti) |
| 4 | Fri 17/04/2026 | 16.30 - 18.00 | 301 (Sarfatti) |

**Note**: lessons will be held in a traditional classroom, and **all the students must bring their own device**.

## Syllabus of the course

| Lecture | Topics |
|---|---|
| 1 | **Managing files and folders**<br>- The HPC resources overview<br>- Login node *vs*. compute nodes<br>- SSH login<br>- From Windows/Mac file manager to the Linux command line: parallels and analogies<br>- Linux Command Line Interface (CLI)<br>  o Up and down arrows<br>  o Tab key for completion<br>  o Interrupting commands with Ctrl+C<br>  o `nano` (to write small scripts or edit them)<br>- Help in Linux: `info` / `man` commands and `--help` option<br>- The `tree` command<br>- Linux commands:<br>  o `whoami, hostname`<br>  o `ls (-l, -a), pwd`<br>  o `mkdir, cd`<br>  o Special directory shortcuts: `~`, `.`, `..`, `/`, `-` (`$OLDPWD`)<br>  o `mv, cp (-r)`<br>  o `file, touch, cat, less, head, tail`<br>  o `rm (-r, -f), rmdir` (dangerous, not undoable!)<br>  o `clear`<br>  o `exit`, `logout`, `ctrl + d`<br>- Absolute *vs*. relative path (`realpath`)<br>- Copying files between HPC and local PC (`scp`)<br>***Exercises*** |
| 2 | **The batch system**<br>- The batch system on an HPC cluster<br>- SLURM basics<br>  o Submitting jobs: `sbatch`<br>  o Monitoring the queue and resources: `squeue`, `sinfo`<br>  o Inspecting job details: `scontrol`<br>  o Cancelling jobs: `scancel`<br>- Writing a simple SLURM job script<br>  o Structure of a batch script<br>  o Requesting resources (time, memory, CPUs)<br>  o Redirecting output and error to log files |

| Lecture | Topics |
|---|---|
| | - Practice: submitting jobs to the scheduler<br>- Storage and `lquota`<br>  o Default storage areas on the HPC (e.g., home, project space)<br>  o How quotas and storage limits affect job runs<br>- Linux commands:<br>  o `echo`<br>  o `df -h, du -sh, quota`<br>***Exercises*** |
| **3** | **Using Python and searching contents on HPC**<br>- Software environments<br>  o Using modules: `module avail/load/list/unload`<br>  o Building a minimal Python environment with *conda*<br>  o Activating the environment in a job script using Python<br>- Practice: submitting Python jobs to the scheduler<br>  o Example: simple Python scripts that read/write small files (e.g., with *open* and *pandas*)<br>  o Checking logs, understanding common errors (missing module, wrong path, exceeded time limit…)<br>- Searching for files and directories<br>- Searching text inside files<br>- Pipes and simple text-processing<br>- Linux commands:<br>  o `find (-name, -type, -mtime)`<br>  o `grep (-i, -c, -n, -v, -r, -E)`<br>  o Chaining commands with `\|`<br>  o `sort`<br>  o `uniq`<br>- Introduction to regular expressions (RegEx)<br>  o Matching characters: `., […]`<br>  o Quantifiers: *, +, ?<br>  o Anchors: ^, $<br>  o Alternation and grouping: `\|, (…)`<br>*Exercises* |
| **4** | **Recap and insights**<br>- Creating a public SSH key to the cluster<br>- Additional useful commands (optional):<br>  o `curl (-o, -s)`<br>  o `chmod (u +-rwx)`<br>  o `diff`<br>  o `history`<br>  o `wc` |

| Lecture | Topics |
|---|---|
| | - Best practices for efficient job scheduling and resource utilization<br>- Recap exercise<br>- Final self-assessment<br>*Exercises* |

## Software used

Bocconi HPC account with Linux (OS Red Hat Enterprise Linux release 9.6) and Python
Enrolled students will receive an email from Bocconi's Technology Office to activate their own HPC account before the first lesson. Instructions will be provided.
The course focuses on command-line and batch-job usage; interactive tools such as Jupyter or remote IDEs will not be covered.

## Suggested bibliography

Materials will be provided by the instructor during the course and will be accessible on Blackboard.
Recommended additional online resources:

- https://learnbyexample.github.io/cli-computing/
- https://www.freecodecamp.org/news/the-linux-commands-handbook
- https://ryanstutorials.net/linuxtutorial/
- https://learning.lpi.org/en/learning-materials/010-160/