

# Module 5 Introduction Data processing infrastructures for scientific applications

Andrea Chierici – INFN-CNAF  
chierici<at>cnafe.infn.it

# What this module is about

- This module is about **Computing Infrastructures**
  - to store and analyze (Big) Data
- We will describe small and big infrastructures
- We will define **HTC**, **HPC** and **Cloud** infrastructures
  - Both local and distributed
- We will Introduce **Containers** and containerized applications
- We will discuss about Computing Models for scientific applications
- And if we have time, we will define Fog, Edge and IoT infrastructures

# Course pre-requisites

- A basic knowledge of the **Linux Operating System** and its command line interface
  - A nice introductory tutorial:
    - <https://ryanstutorials.net/linuxtutorial/#welcome>
- Course will be in English (talks, slides, exercises, exams)
- Students should use their personal **laptop** connected to the ALMAWIFI
- You should be confident with a non-graphical text editor
  - e.g., vim, nano, emacs
- **If you miss one or more of them, just inform me – today or later via email**

# Exam

- Project + Oral examination
- During the exam, students (in groups) will present a project that will include also work **produced in module 4**
- The project will run on google cloud within a docker container
  - You will learn more about this during the course

# Course dates / location

- Calendar
  - Start date: 19 Sept. 2024
  - End date: 19 Dec. 2024
  - <https://www.unibo.it/en/teaching/course-unit-catalogue/course-unit/2023/467739/orariolezioni#468135>
- Lessons will be in (check [calendar](#)):
  - **Laboratorio Informatico (most of the lessons)**  
Via Irnerio 46
  - **Bodoniana (some lessons in November)**  
Via San Donato 19/2
- **Some swaps with module 4** will happen
- Support hours and contacts
  - We can organize appointments via email

# About me

- I'm a **computer scientist**
- I work for the Italian Institute of Nuclear Physics ([www.infn.it](http://www.infn.it)) at CNAF
- I'm the coordinator of the CNAF Tier1 datacenter farming group, the main computing facility of INFN
- I'm involved in several initiatives and projects related to computing
  - High Throughput Computing
  - High Performance Computing
  - Cloud