

High Performance Computing and Data Infrastructure

Introduction to the course



DATA SCIENCE &
ARTIFICIAL INTELLIGENCE



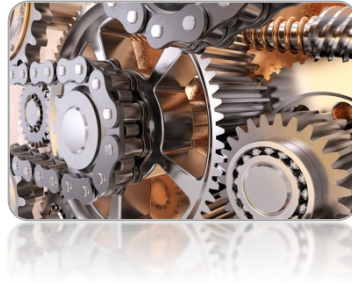
SCIENTIFIC &
DATA-INTENSIVE COMPUTING

2023-2024 @ Università di Trieste

Outline



What you
will learn



How it
works



Some
common
rules & principles



What you will learn

The general information from the syllabus

This course consists of an introduction to computational infrastructure for high performance computing and data management.

It will give some details on how to plan, install and manage computational infrastructures both for providing HPC resource and data infrastructures. The two of them very often are strictly intercorrelated.

The course will present tools and methods also to provide access to such infrastructures and to guarantee data resource interoperability.



The real structure of the course

HPC S.Cozzini / N. Tosato (20h)	Cloud R.Lot 12h	Data M. Prenassi (16h)
common part: infrastructure management: TCO / energy / people...		



What you will learn - HPC

HPC computational nodes and network

Hardware for HPC: CPU, GPU, memory, Top500, Green500

Network details and models.

HPC Storage infrastructure

Storage devices technologies, NAS, SAN, Raid, Performance evaluation

Distributed FS and Parallel FS, CEPH in detail

Managing HPC infrastructure

Total cost of ownership

Energy/Software /People



What you will learn - Cloud Infrastructure

Advanced networking for Cloud

In cloud infrastructure the Container Network Interface (CNI) is one of the key element in performance tuning. We will cover the differences between interfaces (ebpf e BGP), their behaviours and performances under different workloads.

We will then cover multiple CNIs via Multus and the mellanox plugin

Advanced Storage for Cloud

Connecting CEPH to a Kubernetes cluster and the differences between different provisionings



What you will learn – Data Infrastructure

introduction on Big Data, Open Data and FAIR principles concepts applied to data infrastructure

data and metadata models and structures

data resource interoperability and access (ontologies and standardised vocabulary)

Data Resource Interoperability and Access and PIDs

Basic UML (class, activity and deployment diagram) and dataflow diagram

File formats CSV, XSD, JSON and HDF5 with practical parsing examples (Python)

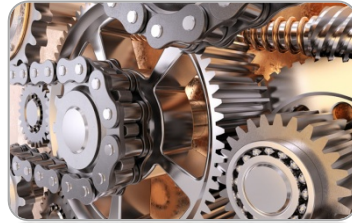
Practical use of metadata models, creation and population of relational databases with python ORM 4h

Integration of metadata in a digital infrastructure at the application level (Flask/Django web framework)

Outline



What you
will learn



How it
works



Some
common
rules & principles



How it works

The baseline is:

1) **lectures:** we prepare slides and examples and we expose them.

Some of the lectures are actually hands-on sessions and tutorials

The lectures will be recorded and you'll find the related files both in the Teams and on a Google Drive in the long-term.

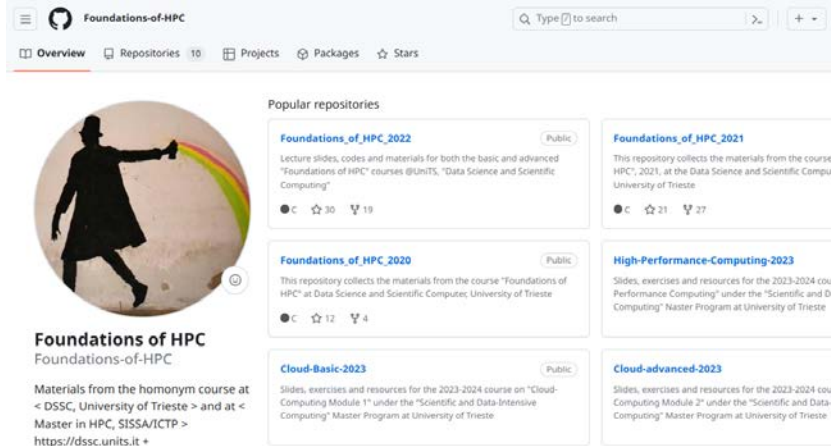
The pdfs that we will use and the example codes will be put in a GitHub repository.



How it works

The baseline is:

1) **lectures**: we prepare slides and examples and we expose them.



`github.com/Foundations-of-HPC`

`/HPC-Data-Infrastructure-2024`
`/`

The repos are online and public; in the landing page you find the detailed list of the lectures.



How it works

The baseline is:

- 1) **lectures:** we prepare slides and examples and we expose them.
- 2) **questions:** you ask what you want whenever you want; if we say something wrong or stupid, you notice that and you raise your hand.



How it works

The baseline is:

- 1) **lectures:** we prepare slides and examples
- 2) **questions:** you ask what you want whenever you want
- 3) **discussions:** you ask to discuss much deeper some aspects or topics from previous/future lectures; you read other materials and bring them into the discussions.



How it works

The baseline is:

- 1) **lectures:** we prepare slides and examples
- 2) **questions:** you ask what you want whenever you want
- 3) **discussions:** you ask to discuss much deeper
- 4) **Office hours:** each of us will be available weekly an hour (we'll communicate which one). However, we are ready for chatting, Q&A or alike upon appointment



How it works

Tools:

ORFEO hpc/cloud/data infrastructure

Please be sure to have an account on it and on its services

Starting point:

[Welcome to Orfeo documentation! - ORFEO documentation \(areasciencepark.it\)](#)

Outline



What you
will learn



How it
works



Some
common
rules & principles



Common rules & principles

- 1) **Don't be shy**, every question is legitimate and useful; stupid questions do not exist, stupid answers do exist.
- 2) Our main focus is that you **learn**, not to grade you.
- 3) Learning is a **process**, not a result; we're interested in both your learning process *and* in your final level.
- 4) We (you and us) will be **honest** in all elements: ChatGPT help should be declared 😊 by both sides
- 5) Be ready for changes and rearrangements: we are pretty busy people 😊



Exam / 1

Eventually, this journey ends with an exam and an evaluation of your level of comprehension and acquired skills.

The evaluation consists in **2 steps**

1. You complete a final assignment that will be given before the end of the lecture;
2. you attend the exam session
 - a) every candidate will discuss individually its own assignment and the answers to the questions given at the begin of the exam



Exam / 2

The final assignment

We will prepare a set of exercises, and you will have to pick **1 for HPC**, **1 for CI** and **1 for DI**.

You must solve the problems, write a report about your work and send us the report 1 week before the exam session.

You must also upload the code that you have written on a git.

No working group is allowed (i.e. everyone has to write its own report).



Exam / 3

You will be allowed to **repeat** the oral exam **without any constraint. However**, please

- (i) do not try multiple times in a short time span (save our time and take yours to digest better)
- (ii) self-organize with your peers so to have groups of you that ask for an exam
- (iii) there will be regular exam sessions, however usually we do prefer to agree with small groups of students to have exams whenever it fits at best.

Questions, comments, doubts, fears... ?

stefano.cozzini@areasciencepark.it

ruggero.lot @areasciencepark.it

marco.prenassi@areasciencepark.it