Distributed Computing

A-01. Introduction to the Course

Welcome!

- This course will be shared between the
  - 1st year of DSE MSc
  - 1st year of SSE MSc
  - 2nd year of SSE MSc
- It's organized in a rather complex way, so we'll explain how it will work for everybody

What Is a Distributed System?

- A collection of **autonomous computing elements** (**nodes**) that appear to its users as a **single coherent system** **(van Steen & Tanenbaum)**

  - Colored text are links… Please follow them to know more!

- Distributed systems are **everywhere**: pretty much anything you see on the Internet is one…

- Also, a single multi-core computer can be seen as a distributed system; distributed system techniques can be and are applied even there **(Bauman et al.)**

Do I Need A Distributed System?

- **Availability**: if one computer (or 10) break down, my website/DB/fancy Ethereum app will still work

- **Performance**: a single-machine implementation won't be able to handle the load/won't be fast enough

- **Decentralization**: I don't want my system to be controlled by a single entity

- And all these things are related in non-trivial ways... **We'll see!**

Making a System Coherent

- We've seen that distributed systems are about making disparate machines **coherent**

- To be coherent, nodes need to **collaborate**
  - We need **synchronization** (there is no **global clock**)
  - We need to **manage group membership** & **authorizations**
  - We need to deal with **node failures**

Distributed Systems: a Huge Topic

- Many courses go in **depth**: you have a (small enough) topic and they teach you everything about it

- Here, we'll go in **breadth**: we'll introduce several topics at a high level, and go deeper on a few
  - Idea: giving you starting points for learning on your own, including after university

How We See the Course

- Security Software and Engineering
  - You'll be among the ones **designing** and **securing** these systems
  - You'll be the "mechanics" who will be tinkering with these systems
- Data Science and Engineering
  - You'll be **using** these systems
  - You'll be the "pilots" who need to know their systems to use them well

Organization of the Course

- Part A (**everybody**):
  - Distributed systems in general
  - Lessons held by Matteo Dell'Amico
- Part B1 (**DSE**):
  - Big Data Engines
  - Lessons held by Giorgio Delzanno
- Part B2 (**SSE** 2nd year):
  - Blockchains and Distributed Ledgers
  - Lessons held by Marina Ribaudo

Exams: SSE

- You'll have **assignments** based on a simulator you have to complete
  - Questions like "in this scenario, which are the best design choices"?
- You'll **tinker** with the simulation and write reports answering the question
- Giacomo Benedetti will follow your assignments and give a green light to proceed to the exam
- Oral exams, where you'll present the assignments and will be asked questions about all the program

Exams: DSE

- You'll have **exercises** on Apache Spark to do during the year
- A written final examination

A Word About Theses

- More important for 2nd year students
- But consider what you're enjoying
- If you find some of these topics interesting, come talk to us!

Part A: Some Topics We'll Touch

- Making systems **consistent**
  - **Consensus** mechanisms
- What **queueing theory** tells us
  - Effects of sharing load between servers
- Handling data efficiently
  - Modeling systems with data **replication**
  - **Erasure coding**: the gifts of coding theory
- Introduction to decentralized systems
- Introduction to big data engines