

Introduction to Big Data Processing Infrastructures

Introduction

- This course is not about Big Data, it is about infrastructures
- We should have visited a datacenter, but... Corona
- We need SSH, a little bit of C/C++, vim, Python
- We will use AWS and Google Cloud computing!
- Of course we will make a project and an oral exam
- Prof of course is a physicist of the IIFN
- The IIFN-CNAF hosts the Tier-1 datacenter in Bologna

Setting up AWS

- It is a pay for use model
- You have at the beginning a 50\$ credit

Big data

- They tend to be non structured
- They are characterized by 4 Vs: Volume, Variety, Veracity (can I trust them?), Velocity
- Another important point is Value, what I want to extract from the data