UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Technological Infrastructures For Data Science

2122-2-F9101Q052

Aims

The course aims at providing a solid understanding of the technological platforms (Cloud and Containers) data collection and management, as well as of the computing platforms (architectures, algorithms, and infrastructures) that can be used to analyze those data.

The exercises will provide the student with the basic capabilities necessary to interact with such platforms.

Contents

Virtualized platforms for collecting and handling data characterized by volume and velocity. Data processing architectures, processing infrastructure, Big Data platforms for Data Science, examples of platforms.

Detailed program

- Data processing architectures
- Infrastructure Management
- Virtualization and containerization
- Cloud environments, models, and costs
- Big Data Platforms for Data Science
- Workload Types
- MapReduce: the Hadoop/Spark ecosystem

- Event and stream processing
- Platform Examples
    - Apache Spark
    - Apache Kafka
    - Google Cloud Platform
- DevOps - DataOps - MLOp

Prerequisites

Basic knowledge of:

- a programming language (e.g., Python)
- the architecture of a computer (CPU, Memory, Disk...).

Teaching form

Classroom lectures, classroom exercises. The course will be held in English

Textbook and teaching resource

Lecture notes and slides provided by the lecturers.

Semester

Second year, first semester

Assessment method

The exam will consist of two parts. The two parts will have to be **done in the same session**.

The first part will consist of a set of closed and open questions to be taken in one hour (approximately there will be 9 closed and 4 open questions, however, there may be small variations in the structure of the exam). The first part of the exam will be conducted in paper form and the weight of each question for the calculation of the grade will be explicitly indicated at the top of the sheet containing the questions. The answers to the questions will be written on the sheet itself, but it is up to the student to use additional sheets to provide a more extensive response to the open-ended questions.

The second part will consist of a short project to be carried out in groups of 2 and its oral presentation.

Examples of project types:

- Analysis and testing of a particular technology platform (advantages, disadvantages, costs, learning curve, building an application using that technology)
- Design of a cloud application (choice of provider, type of virtual machines, services, cost estimation, quality of service estimation, supported data size estimation, corrective activities needed in case of unforeseen events)

Once the student has taken both tests, the exam will be considered passed if both of these conditions are met:

1. For both parts, the student will have scored more than half of the 15 points available to them
2. The sum of the points of the two parts is greater than or equal to 18.

In this case, the student will be able to record a grade consisting of the sum of the points.

Office hours

Tuesday 12:30-14:30 ask for email confirmation