Computing Foundations of Data Science

Lecturer
Dimitrios SACHARIDIS (Coordinator)

Course mnemonic
INFO-H600

ECTS credits
5 credits

Language(s) of instruction
English

Course period
First term

Campus
Solbosch

Course content

The course has three parts:

- **Introduction to Python**. The basic programming concepts are introduced using the Python languages.
- **Data Science with Python**. An introduction to data science concepts is made using Python modules. The focus is on manipulating and analyzing numerical and tabular data using numpy and pandas modules. The introduction also covers handling missing values, dealing with hierarchical data like json, and basic data visualization methods.
- **Introduction to Big Data Management**. This part introduces the challenges of big data, discusses computational architectures and programming abstractions like distributed/ parallel computing and map-reduce. These concepts are discussed in the context of the python framework Dask, and Apache Spark using python as the interface (pySpark).

Objectives (and/or specific learning outcomes)

Upon completion of the course, the student is expected to:

- Obtain a good level of programming with Python.
- Understand the basic data science concepts, and be able to manage and analyze structured data.
- Be able to perform data science tasks using Python.
- Understand the challenges of working with Big Data, and understand the techniques that can be used to address these challenges.
- Be able to perform big data management in Python using the Dask and Apache Spark frameworks.

Pre-requisits and co-requisits

Required knowledge and skills

A basic understanding of programming is assumed. Familiarity with Python is preferable, but the course will provide all necessary material.

Teaching method and learning activities

The teaching of each part consists of a brief theoretical introduction and is followed by practical sessions.

There are two assignments that correspond to the topics of the first two parts.

There is a project, to be done in groups, that corresponds to the topic of the third part.

Course notes

Université virtuelle

Other information

Place(s) of teaching

Solbosch

Contact(s)

Dimitris Sacharidis dimitris.sacharidis@ulb.be

Evaluation method(s)

Other, Project and Practice work

Evaluation method(s) (additional information)

Students are evaluated based on two assignments and one project.

Each assignment contributes 25% of the total mark (i.e., 5 points).

The project contributes 50% of the total mark (i.e., 10 points).

There is no final exam.

Main language(s) of evaluation

English

Programmes

Programmes proposing this course at the Brussels School of Engineering

MS-BGDA | Specialized Master in data science, Big data | unit U

Programmes proposing this course at the Solvay Brussels School of Economics and Management

MS-BGDA | Specialized Master in data science, Big data | unit U

Programmes proposing this course at the faculty of Sciences

MS-BGDA | Specialized Master in data science, Big data | unit U