

Academic Year: (2023 / 2024)

Review date: 19-12-2023

Department assigned to the subject: Signal and Communications Theory Department

Coordinating teacher: MOLINA BULLA, HAROLD YESID

Type: Electives ECTS Credits : 6.0

Year : 1 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Basic programming skills in Python

OBJECTIVES

The main objective of this course is to train analysts in data science, either for research or the labour market, using the latest techniques applied in machine learning.

We will learn how to get the most out of the computational resources that we can have at our disposal, from our own computer, resources in the cloud for programming with GPUs and programming for Big Data; knowledge in high demand in various environments.

The basic concepts of parallel programming will be explained:

- 1- using the resources of general purpose computers,
- 2- hybrid programming: using the resources of a normal computer and specific hardware such as graphics processing units (GPUs).
- 3- distributed and cloud programming, for big cases, such as Big Data.

Basic competences

CB6 Having and understanding the knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context

CB7 Students know how to apply their acquired knowledge and problem-solving skills in new or unfamiliar settings within broader (or multidisciplinary) contexts related to their field of study.

CB8 Students are able to integrate knowledge and to face the complexity of making judgments based on information that, being incomplete or limited, includes reflections on the social and ethical responsibilities linked to the application of their knowledge and judgments.

CB9 Students know how to communicate their conclusions and the knowledge and ultimate reasons behind them to specialised and non-specialised audiences in a clear and unambiguous way.

CB10 Students have the learning skills that will enable them to continue studying in a way that will be largely self-directed or autonomous.

General competences

CG1 Ability to maintain continuous education after his/her graduation, enabling him/her to cope with new technologies.

CG2 Ability to apply the knowledge of skills and research methods related to engineering.

CG3 Ability to apply the knowledge of research skills and methods related to Life Sciences.

CG4 Ability to contribute to the widening of the frontiers of knowledge through an original research, part of which merits publication referenced at an international level.

Specific competences

CE4 Ability to use techniques for processing massive amounts of medical data and images.

CE5 Ability to implement medical imaging and data processing methods.

DESCRIPTION OF CONTENTS: PROGRAMME

Parallel Programming:

- * Multiprocessor/Multicore Programming.
- * Parallel Programming with Shared Memory (and the dangers for data)
- * Parallel Programming with Shared Memory with semaphores and locking

Hybrid Programming

- * Use of GPUs and how they differ from ordinary computer processors.
- * How GPUs are programmed: when and how they can be used.
- * Use of advanced GPU programming techniques, GPU resource management.
- * How to use GPUs in Machine Learning.

Distributed Programming

- * What is distributed computing and how can we take advantage of it?
- * Use of distributed programming platforms for BigData and Machine Learning.
- * Use cases of Apache Spark for machine learning.

LEARNING ACTIVITIES AND METHODOLOGY

It is a practice-oriented course. The basic theoretical concepts of intensive computing will be taught, in order to carry out guided practice. Use cases of the technologies will be implemented, oriented to form an advanced researcher in data science.

Training activities:

- AF3 Theoretical and practical classes
- AF4 Laboratory practicals
- AF5 Tutorials
- AF6 Group work
- AF7 Individual student work
- AF8 Partial and final exams

METHODOLOGY

MD1: Class lectures by the teacher with the support of computer and audiovisual media, in which the main concepts of the subject are developed and the bibliography is provided to complement the students' learning.

MD2: Critical reading of texts recommended by the subject teacher.

MD3: Resolution of practical cases, problems, etc. posed by the teacher individually or in groups.

MD4: Presentation and discussion in class, under the moderation of the lecturer.

There will be 2 hours a week of tutorials for students where the teacher will be available in his or her office.

ASSESSMENT SYSTEM

CONTINUOUS EVALUATION.

Practical work done during the class: 20%

Massive Computing Practice on the Computer: 20%

Heterogeneous Computing Practice (Computer + GPU): 20%.

Special Final Practice (Distributed and Heterogeneous): 40%.

EXTRAORDINARY CALL

- * In-person exam with questions about the practical works carried out.

% end-of-term-examination: 60

% of continuous assessment (assignments, laboratory, practicals...): 40

BASIC BIBLIOGRAPHY

- Benjamin Bengfort ; Jenny Kim Interactive Spark using PySpark, O'Reilly Media, 2016
- Holden Karau ; Rachel Warren High Performance Spark, O'Reilly Media, 2017
- Ian Gorton, Deborah K. Gracio Data-Intensive Computing: Architectures, Algorithms, and Applications, Cambridge University Press New York, 2012