![uc3m | Universidad Carlos III de Madrid]

Massive computing

Academic Year: (2023 / 2024)   **Review date: 19-04-2023**

Department assigned to the subject: Signal and Communications Theory Department

Coordinating teacher: MOLINA BULLA, HAROLD YESID

Type: Compulsory ECTS Credits : 6.0

Year : 3 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Labs will be made using C and Python programming languages.

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 common 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.

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 traffic lights 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

AF1: THEORETICAL-PRACTICAL CLASSES. They will present the knowledge that should be acquired. They will receive the class notes and will have basic texts of reference to facilitate the follow-up of the classes and the development of the subsequent work. Exercises, practical problems on the part of the student will be solved and workshops and evaluation test will be held to acquire the necessary skills.
AF2: Updated to allegation
AF3: INDIVIDUAL OR GROUP WORK OF THE STUDENT.

MD1: THEORETICAL LESSONS. Exhibitions in class with both support of computer and audiovisual media, in which the main concepts of the subject are developed and the materials and bibliography are provided to complement the students' learning.
MD2: ASSIGNMENTS. Resolution of practical use cases, problems, etc. raised by the teacher, individually or in groups.
MD3: TUTORIALS. Individualized assistance (individual tutorials) or group (collective tutorials) to

students by the teacher.

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

- Holden Karau, Andy Konwinski, Patrick Wendell & Matei Zaharia Programming in Scala, Artima.

- Mike Frampton Mastering Apache Spark, Packt Publishing, 2015