Data Intensive Computing

Academic Year: (2022 / 2023) **Review date: 10-05-2022**

Department assigned to the subject:

Coordinating teacher: MOLINA BULLA, HAROLD YESID

Type: Electives ECTS Credits : 6.0

Year : 1 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Labs will be made using Python (mainly) and C programming languages.

OBJECTIVES

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

Data intensive computing

Parallel and distributed computing paradigms
Data-intensive computing platforms
High-Performance Computing platforms for Big Data
In-memory computing
Fault-tolerance and resilience.

LEARNING ACTIVITIES AND METHODOLOGY

AF3 Theoretical practical classes
AF4 Laboratory practices
AF5 Tutorials
AF6 Team work

AF7 Student individual work

Activity code	total hours number	presencial hours number	% Student Presence
AF3	134	134	100%
AF4	42	42	100%
AF5	24	0	0%
AF6	120	0	0%
AF7	248	0	0%
AF8	16	16	100%
SUBJECT TOTAL	600	184	30,66%

ASSESSMENT SYSTEM

CONTINUOUS EVALUATION.
Practical work done during the class: 20% * Practical work done during the class: 20%.
Massive Computing Practice on the Computer: 20% * Heterogeneous Computing Practice (Computer + GPU): 20%.
Heterogeneous Computing Practice (Computer + GPU): 20% * Special Final Practice (Distributed and Heterogeneous): 40%.
Special Final Practicum (Distributed and Heterogeneous): 40% * Special Final Practicum (Distributed and Heterogeneous): 40%.

% end-of-term-examination:	60
% of continuous assessment (assigments, 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