VNIVERSITAT DE VALÈNCIA

Course Guide
42591 Programming and advanced computational techniques in bioinformatics

COURSE DATA

Data Subject	
Code	42591
Name	Programming and advanced computational techniques in bioinformatics
Cycle	Master's degree
ECTS Credits	3.0
Academic year	2022 - 2023

Study (s)

Degree	Center	Acad. year	Period
2116 - Master's Degree in Bioinformatics	School of Engineering	2	First term

Subject-matter

Degree	Subject-matter	Character
2116 - Master's Degree in Bioinformatics	9 - Programming and advanced computational techniques in bioinformatics	Obligatory

Coordination

Name	Department
ARNAU LLOMBART, VICENTE	240 - Computer Science

SUMMARY

This course explores the potential parallel programming can bring us to solve large problems bioinformatics.
Additionally, applications and libraries using the most widely used bioinformatics programming languages seen in the master.
Knowing of the main tools available in the scientific community for data storage and processing bioinformatics. You will see that is Hadoop, NoSQL, Big Data and Cloud Computing.

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

There are no specified enrollment restrictions with other subjects of the curriculum.

Other requirements

None.

COMPETENCES (RD 1393/2007) // LEARNING OUTCOMES (RD 822/2021)

2116 - Master's Degree in Bioinformatics

- Students should apply acquired knowledge to solve problems in unfamiliar contexts within their field of study, including multidisciplinary scenarios.
- Students should be able to integrate knowledge and address the complexity of making informed judgments based on incomplete or limited information, including reflections on the social and ethical responsibilities associated with the application of their knowledge and judgments.
- Students should communicate conclusions and underlying knowledge clearly and unambiguously to both specialized and non-specialized audiences.
- Students should demonstrate self-directed learning skills for continued academic growth.
- Be able to access the information required (databases, scientific articles, etc.) and to interpret and use it sensibly.
- Students should possess and understand foundational knowledge that enables original thinking and research in the field.
- Be able to access to information tools in other areas of knowledge and use them properly.
- To be able to assess the need to complete the scientific, historical, language, informatics, literature, ethics, social and human background in general, attending conferences, courses or doing complementary activities, self-assessing the contribution of these activities towards a comprehensive development.
- Desarrollar la iniciativa personal y ser capaces de realizar una toma rápida y eficaz de decisiones en su labor profesional y/o investigadora.
- Trabajar en equipo con eficiencia en su labor profesional y/o investigadora y con personas de diferente procedencia.
- Conocer y emplear las principales aplicaciones bioinformáticas y las librerías existentes para los lenguajes de programación vistos en el Máster.
- Comprender en qué tipo de aplicaciones la programación paralela y los grandes sistemas de computación son requeridos para la resolución de problemas bioinformáticos y analizar sus prestaciones.

LEARNING OUTCOMES (RD 1393/2007) // NO CONTENT (RD 822/2021)

Understand the potential of parallel programming can bring us to solve large problems bioinformatics
Using bioinformatics applications and libraries most used programming languages in the master watched.
Date of the main tools available in the scientific community for data storage and processing bioinformatics.
Understand the data processing environments in "pipeline".

DESCRIPTION OF CONTENTS

1. Introduction to parallel programming.

Will present the basics of the subject. Will present the basics of HPC, as OpenMP, Cuda, SSE. It uses the Python language for parallel programming.
Finally, we discuss the new technologies related to Big Dat: Hadoop, Cloud Computing, NoSQL, ...

2. Hadoop

Apache Hadoop (High-availability distributed object-oriented platform) is an open-source software framework that supports data-intensive distributed applications, licensed under the Apache v2 license. It supports the running of applications on large clusters of commodity hardware. Hadoop was derived from Google's MapReduce and Google File System (GFS) papers.

3. Big Data

Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis and visualization.
Bioinformatics is one of the disciplines considered is within the Big Data.

4. Cloud Computing

Cloud computing is a colloquial expression used to describe a variety of different types of computing concepts that involve a large number of computers that are connected through a real-time communication network (typically the Internet).[1] Cloud computing is a jargon term without a commonly accepted non-ambiguous scientific or technical definition. In science, cloud computing is a synonym for distributed computing over a network and means the ability to run a program on many connected computers at the same time. The popularity of the term can be attributed to its use in marketing to sell hosted services in the sense of application service provisioning that run client server software on a remote location.

5. NoSQL

A NoSQL database provides a mechanism for storage and retrieval of data that uses looser consistency models than traditional relational databases.

WORKLOAD

ACTIVITY	Hours	% To be attended
Theory classes	10,00	100
Laboratory practices	5,00	100
Development of group work	5,00	0
Development of individual work	10,00	0
Study and independent work	15,00	0
Readings supplementary material	10,00	0
Preparation of evaluation activities	10,00	0
Preparing lectures	5,00	0
Preparation of practical classes and problem	5,00	0
TOTAL	**75,00**	

TEACHING METHODOLOGY

MD1 - Task training of the teaching-learning environment interaction in the classroom through expository sessions. Previous assignments include preparation (information search, reading texts supplied by teachers), teaching sessions themselves and the later work of deepening.

MD2 - Learning through problem solving and case studies, through which it is acquiring skills on different aspects of materials and subjects.

MD3 - Actividades de laboratorio driving range. INCLUDE preparations realization of the con las seguimiento driving range and I support the teacher, freestanding trabajo y elaborations on-line reports las driving range.

MD4 - Cross-disciplinary skills. Include attendance at courses, conferences or round tables organized by the CEC of the Master and / or conduct of a bibliographic work on issues that contribute to the integral. It produces a report of activities.

EVALUATION

In the two calls:

SE1 Continuous assessment: minimum 5 and maximum 15.
SE2 Activities: minimum 10 and maximum 40.
SE3 Laboratory: minimum 25 and maximum 50.
SE4 Exams: minimum 0 and maximum 50.

REFERENCES

Basic

- Parallel Processing via MPI & OpenMP, M. Firuziaan, O. Nommensen. Linux Enterprise, 10/2002
- Big Data: A Revolution That Will Transform How We Live, Work, and Think. Viktor Mayer-Schonberger, Kenneth Cukier.
- INTRODUCCION A LA PROGRAMACION PARALELA. FRANCISCO ALMEIDA , EDICIONES PARANINFO, S.A., 2008.