UAB
Universitat Autònoma de Barcelona

High Performance Computing and Big Data Analytics

Code: 43917
Credits: 12

2026/2027

Degree programme	Type	Course
Bioinformatics	OP	1

Contact lecturer

Name : Miquel Àngel Senar Rosell

Email : miquelangel.senar@uab.cat

Teaching staff

Oscar Lao Grueso

Emanuele Raineri

Teaching staff (external to UAB)

Sandra Méndez

Group languages

You can consult this information at the end of the document.

Prerequisites

To carry out this module is necessary to have passed previously both compulsory modules: Programming in Bioinformatics and Core Bioinformatics.

It is recommended you have a Level B2 of English or equivalent.

Objectives

This module aims to provide students with the necessary knowledge and skills (1) to develop robust and scalable software applications to solve bioinformatics problems through the combined use of general middleware tools and specific bioinformatics tools; (2) to apply advanced Python programming methods within the bioinformatics field and (3) to perform advanced statistical analyses of Big Data.

Learning outcomes

- CA11 (Design efficient computational solutions to bioinformatics problems through the use of algorithms, programming tools, and advanced execution tools.) Design efficient computational solutions to bioinformatics problems through the use of algorithms, programming tools, and advanced execution tools.

- CA12 (Integrate multidisciplinary resources working collaboratively for the development of biomolecular data analysis workflows.) Integrate multidisciplinary resources working collaboratively for the development of biomolecular data analysis workflows.
- CA13 (Communicate clearly, orally and in writing, in technical English, complex results of bioinformatics analysis based on big data.) Communicate clearly, orally and in writing, in technical English, complex results of bioinformatics analysis based on big data.
- KA13 (Describe the main characteristics of modern computing architectures and their impact on the performance of bioinformatics applications.) Describe the main characteristics of modern computing architectures and their impact on the performance of bioinformatics applications.
- KA14 (Describe the statistical and computational underpinnings, including graph theory and machine learning principles, applicable to the analysis of massive biomolecular data.) Describe the statistical and computational underpinnings, including graph theory and machine learning principles, applicable to the analysis of massive biomolecular data.
- KA15 (Explain the principles of performance engineering and the tools that allow optimising the execution and reproducibility of bioinformatic analysis processes.) Explain the principles of performance engineering and the tools that allow optimising the execution and reproducibility of bioinformatic analysis processes.
- SA14 (Use advanced programming models, including parallelisation techniques, to solve complex bioinformatics problems.) Use advanced programming models, including parallelisation techniques, to solve complex bioinformatics problems.
- SA15 (Use computational tools and platforms to develop and document reproducible and scalable bioinformatics workflows.) Use computational tools and platforms to develop and document reproducible and scalable bioinformatics workflows.
- SA16 (Apply statistical analysis and machine learning techniques to the classification and grouping of biomolecular data.) Apply statistical analysis and machine learning techniques to the classification and grouping of biomolecular data.

Contents

Modern Computer Architecture

- Cluster systems
- System Middleware and Programming Frameworks

Advanced Programming Models

- Advanced shell scripting
- Using system tools for bioinformatics analysis
- High Performance Computing with Python

Big Data Analytics

- Theory and tools of advanced statistics in Big Data analytics (dimensionality reduction, variable selection and Spark)
- Machine learning theory and algorithms. Applications in Bioinformatics
- Predictive modelling: data mining, model evaluation and validation
- Data classification: naïve Bayes and decision trees learning
- Association rule learning
- Clustering analysis: k-means algorithm
- Graph Theory for Big Data

Learning activities and methodology

Title	Hours	ECTS	Learning outcomes
Solving problems in class and work in the biocomputing lab	32	1.28	
Regular study	226	9.04	
Theoretical classes	38	1.52	

By following a problem-oriented approach, students will get insight about efficient computational algorithms, methods and platforms and the statistical methods to be applied to challenging bioinformatics problems dealing with Big Data.

Attendance at class sessions is mandatory.

In this course, the use of Artificial Intelligence (AI) technologies is permitted as an integral part of the assignment development, provided that the final result reflects a significant contribution from the student in personal analysis and reflection. The student must clearly identify which parts were generated using this technology, specify the tools used, and include a critical reflection on how they influenced the process and final outcome of the assignment. Lack of transparency in the use of AI will be considered a breach of academic honesty and may result in a penalty on the assignment grade or greater sanctions in serious cases.

Annotation: within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Assessment

Continuous assessment activities

Title	Weight	Hours	ECTS	Learning outcomes
Works done and presented by the student (student's portfolio)	100%	4	0.16	CA11, CA12, CA13, KA13, KA14, KA15, SA14, SA15, SA16

The assessment system is organized around the student portfolio.

Main evaluation activities

- Student's portfolio (60%): works done and presented by the student all along the course. No individual assessment activity will account for more than 50% of the final grade. To be eligible for the assessment of these activities, regular attendance at all class sessions is required.

Should the teaching staff deem it appropriate, an oral examination may be conducted to verify the authorship of any of the assessment tasks.

Not valuable

The student will be graded as \"Not Valuable\" if the weight of the evaluation is less than 25% of the final score.

Honors grade (MH)

Honors are decided by the faculty responsible for the course. UAB regulations specify that an honors grade can only be granted to students with a final grade equal or greater than 90%, and only to 5% of enrolled students. If the number of students with a final grade above 9 is above 5% of the total enrollment, the faculty in charge will prioritize based on the students' overall performance in all parts of the course.

Unique assessment

This subject/module does not provide for the single assessment system.

Irregularities

Any irregularity committed during an assessment activity (academic misconduct, plagiarism, or improper use of AI, unless such use is expressly authorized in the course syllabus) that may lead to a significant alteration of the grade will result in that activity being graded as 0. If the course syllabus stipulates that obtaining a minimum mark in this assessment is an essential requirement to pass the course, or if multiple irregularities occur in the assessment activities of the same course, the final grade for the course will be 0. Furthermore, disciplinary proceedings may be initiated against any student who incurs any of these irregularities.

Bibliography

Updated bibliography will be recommended in each session of this module by the professor, and links will be made available on the Student's Area of the MSc Bioinformatics official website

Software

Linux + SLURM and other tools from Linux enviroments

Python and other tools from its ecosystem

R and other tools from its ecosystem

Course groups and languages

The information provided is provisional until November 30. After this date, you will be able to consult the language of each group through this link. To access the information, you will need to enter the course CODE

Type of teaching	Group	Language	Semester	Shift
(TEm) Theory (master)	1	English	first semester	morning-mixed

(PLABm) Practical laboratories (master)	1	English	first semester	morning-mixed