

Data Management

Code: 106566
ECTS Credits: 6

2025/2026

Degree	Type	Year
Artificial Intelligence	OB	3

Contact

Name: Javier Panadero Martinez

Email: javier.panadero@uab.cat

Teachers

Haoyuan Li

Teaching groups languages

You can view this information at the [end](#) of this document.

Prerequisites

Although there are no formally established prerequisites, and the subject will provide students with the means to acquire the knowledge described in the syllabus. It is advisable: a good knowledge of programming, of the computer structure, of the operating system at the level of user programmer and of database systems.

Objectives and Contextualisation

The objectives of the subject are the study of the main concepts of data intensive applications regarding their reliability, scalability, and performance.

The subject will introduce data distributed systems and paradigms of data processing, programming models for batch, in-memory and streaming processing, architecture of applications like lambda. It will cover the integrity, accessibility, reliability, consistency, and security of large data processing. Also, the integration of application designs and infrastructure.

Competences

- Analyse and solve problems effectively, generating innovative and creative proposals to achieve objectives.
- Conceptualize and model alternatives of complex solutions to problems of application of artificial intelligence in different fields and create prototypes that demonstrate the validity of the proposed system.

- Introduce changes to methods and processes in the field of knowledge in order to provide innovative responses to society's needs and demands.
- Know and efficiently use techniques and tools for representation, manipulation, analysis and management of large-scale data.
- Students must have and understand knowledge of an area of study built on the basis of general secondary education, and while it relies on some advanced textbooks it also includes some aspects coming from the forefront of its field of study.
- Work cooperatively to achieve common objectives, assuming own responsibility and respecting the role of the different members of the team.

Learning Outcomes

1. Analyse and solve problems effectively, generating innovative and creative proposals to achieve objectives.
2. Conceive, design and implement data collection and annotation processes appropriate to the problem at hand that needs resolving.
3. Propose new methods or informed alternative solutions.
4. Select the most appropriate storage methods to enable efficient subsequent data retrieval and analysis.
5. Students must have and understand knowledge of an area of study built on the basis of general secondary education, and while it relies on some advanced textbooks it also includes some aspects coming from the forefront of its field of study.
6. Understand the basics of distributed databases and how to use big data processing tools.
7. Work cooperatively to achieve common objectives, assuming own responsibility and respecting the role of the different members of the team.

Content

1-Introduction to massive data applications

2-Main concepts of data management in massive data environments: reliability, scalability and maintainability. Data models and query languages.

3-Large volume data management. Data warehousing. Main principles of Data Warehousing systems, business intelligence, multidimensional modeling, OLAP operators, ETL processes

4-Introduction to in-memory databases with Redis

5-Large volume data management with Apache Spark tools. Introduction to Spark Dataframes and MLlib

Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
<hr/>			
Type: Directed			
Case studies	9	0.36	1, 2, 3, 7
Labs	12	0.48	2, 6, 4, 7
Theory	20	0.8	2, 6, 4, 5
<hr/>			
Type: Autonomous			

Autonomous study	30	1.2	1, 6, 4, 5
Case studies preparation	20	0.8	1, 2, 6, 4, 3, 7
Lab preparation	32	1.28	1, 2, 7

During the development of the subject, we may differentiate three types of teaching activities:

Theoretical classes: general description of the theoretical part of each program topic. The typical structure of a theoretical lesson will be the following: first, an introduction will be made where the objectives of the lesson and the contents to be discussed will be briefly presented. Next, the contents of the lesson will be discussed, including narrative expositions, formal developments that provide the theoretical foundations, including examples that illustrate the application of the discussed contents. Finally, the professor will present the conclusions of the lesson. Throughout the course there will be continuous assessments of groups of topics.

Laboratory classes: The practical part of the theoretical topics will be completed with sessions in the laboratory, where the student will develop a series of programs and should try to solve a specific problem that will be received at the beginning of the semester. Some of these exercises must be delivered on the specified dates. The lab sessions will be developed in groups of two or three students. The subject includes a list of sessions in the laboratory, lasting 2 hours each, where the student will carry out the exercises. The lab report will be delivered in the virtual campus to be evaluated.

Case studies: during the final sessions of the subject, a list of practical cases will be presented to the students. These cases will contain challenges to solve with data sets and business objectives. Students will work in groups to describe a list of conclusions of their work in an oral presentation and a final report.

This approach to work is oriented to promote an active learning and developing competencies of organization and planning skills, oral and written communication, teamwork and critical reasoning. The quality of the exercises carried out, their presentation and their functioning will be especially valued

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

Continous Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Individual exam 1	30%	2	0.08	1, 6, 5
Individual exam 2	30%	2	0.08	1, 6, 5
Lab work	40%	23	0.92	1, 2, 6, 4, 3, 7

This course does not offer a single-assessment system. Repeating students must complete all planned activities, both theoretical and practical; that is, there will be no differentiated treatment for repeating students.

The course consists of three parts: Theory, Problem Solving, and Laboratory Work. The Theory and Problem Solving components account for 60% of the final grade, while Laboratory Work makes up the remaining 40%.

The dates for continuous assessment tests and lab sessions will be published on the virtual campus at the beginning of the course and may be subject to rescheduling due to possible contingencies. All changes will be

communicated through the virtual campus, which is understood to be the standard communication channel between faculty and students.

Honors distinctions ("matrícula de honor") will be awarded based on five percent (or fraction thereof) of the total number of students enrolled across all course groups. Only students with a final grade equal to or greater than 9 will be eligible.

The assessment method for each part of the course (Theory, Problem Solving, and Laboratory Work) is detailed below:

Theory and Problem Solving

The course will follow a continuous assessment methodology that allows students to progressively eliminate material as they advance. Two written continuous assessment tests are scheduled:

- The first test (P1) will take place during the midterm exam week.
- The second test (P2) will take place during the final exam week.

Exact dates will be published at the beginning of the course and may change due to contingencies. All updates will be announced via the virtual campus, as this is considered the standard information exchange tool between faculty and students.

Each test will account for 30% of the final course grade.

To be eligible to take the second continuous assessment test (P2), students must achieve a minimum score of 3.5 in the first test (P1). Otherwise, they must take the resit exam (ER), which will cover the entire course content. Additionally, if the average score of P1 and P2 is below 5, the student must also take the resit exam in order to pass the course.

For each test, the time, date, and location for review sessions will be provided. Students may review their test with the instructor during this session. If a student does not attend the scheduled review, no subsequent review will be allowed.

Students wishing to attend the review session must notify their theory instructor by email at least 24 hours in advance. If no notification is given within this timeframe, the test will not be reviewed.

During the review session, exercises will not be explained or solved. The test will be shown solely for the student to identify errors and understand the rationale behind the grade received.

Exam solutions will not be published on the virtual campus. Students wishing to see the solution to a particular question must request a tutorial session after the review process has concluded.

Retaken Exam

Only students who have not passed the continuous assessment-either because they did not reach the minimum score of 5 out of 10, or because they did not follow it-are eligible to take the resit exam.

This exam will cover the full syllabus and will have a maximum score of 7 points. A minimum score of 5 is required for the theoretical part to be averaged with the lab grade. A score below 5 will result in failing the course.

Any attempt to cheat during an assessment activity-either during the activity or in the grading process-will result in a final course grade of 3, and a disciplinary case will be opened and recorded in the student's academic file.

The teaching staff reserves the right to modify the format of midterm and final exams as deemed appropriate, regardless of formats used in previous years.

Laboratory Sessions

Lab work will be assessed based on the work completed during lab sessions and on the reports written for each session. Lab work must be done in groups of three students.

Attendance to lab sessions is mandatory. Missing a session will result in failure of the practical component and, consequently, failure of the course. In the case of a justified absence, it must be reported in advance to the instructor and an official signed justification must be provided within the established timeframe. Notification must always occur prior to the session.

It is important to clarify that personal travel and work-related reasons are not considered valid justifications, since the practical session calendar is available from the beginning of the course.

Justified, non-medical absences must be rescheduled for another session within the same week. Only students who justify their absence due to illness will be exempt from this rescheduling. In any case, missing the assigned session and thereby preventing group work will require the student to complete the lab individually.

Full attendance for the entire duration of each lab session is required. Attendance will be recorded at the beginning of the session, and again at the end, when the instructor will inquire about the work completed. Each group member will be assessed individually. The grading breakdown for lab components will be specified in the course's detailed guidelines.

All lab sessions carry the same weight. The specific grading criteria for each session will be included in the corresponding assignment. It is the student's responsibility to read this information carefully and to ensure they sign the attendance sheet for each session.

Arriving more than 15 minutes late will be recorded as a "no-show," and the session cannot be made up. This condition will not apply to students who provide an official justification for the delay (e.g., a medical attendance certificate).

Lab sessions cannot be retaken. A minimum average score of 5 is required to pass this component. There is no minimum score required for individual labs in order to calculate the overall average.

Plagiarism and Cheating

Without prejudice to other disciplinary measures that may apply, and in accordance with current academic regulations, any irregularities committed by a student that may affect the grading of an assessment activity will result in a score of zero (0). Activities graded in this manner will not be eligible for retake. If passing one of these activities is necessary to pass the course, the course will be failed without the possibility of passing it within the same academic year.

Such irregularities include, but are not limited to:

- total or partial copying of a lab, report, or any other graded activity; allowing others to copy;
- unauthorized use of AI tools (e.g., Copilot, ChatGPT, or similar) in any graded activity will result in a score of zero;
- submitting a group assignment not fully completed by the group members (this applies to all members, not just those who didn't contribute);
- submitting materials created by third parties, including translations or adaptations, or any work that is not original and exclusively the student's own;
- having communication devices (e.g., mobile phones, smartwatches, camera pens, etc.) accessible during individual theoretical-practical assessment sessions (exams);
- talking to peers during individual theoretical-practical assessment sessions (exams);
- copying or attempting to copy from other students during theoretical-practical assessments (exams);
- using or attempting to use materials related to the subject during theoretical-practical assessments (exams), unless explicitly allowed.

In summary: copying, allowing others to copy, or committing plagiarism (or attempting to) in any assessment activity will result in a FAIL, with no compensation or recognition of partial components in future academic years.

If a student fails the course because they did not meet the minimum required grade in any assessment activity, the final grade will be the lower of either 4.5 or the weighted average of all grades. Exceptions: students who do not participate in any assessment activities will receive a grade of "Not Assessed," and students who commit irregularities in any assessment activity will receive a final grade equal to the lower of 3.0 or the weighted average of their grades (and therefore cannot pass the course by compensation).

Bibliography

Designing Data intensive applications - Martin Kleppmann, O'Reilly, 2017

The Data warehouse ETL toolkit - Ralph Kimball, Joe Caserta. Wiley, 2004

Spark, the definitive guide, Big data processing made simple. Bill Chambers and Matei Zaharia, O'Reilly, 2018

Learning Spark - Lightning fast data analysis - Holden Karau, Andi Konwinski, Patrick Wendell, Matei Zaharia, O'Reilly, 2015

Beginning Scala - Layka, Vishal. Apress; 2nd ed. 2015.

Redis in Action - Josiah L. Carlson. Manning, 2013.

Software

OpenNebula private cloud services will be used during the course

Groups and Languages

Please note that this information is provisional until 30 November 2025. You can check it through this [link](#). To consult the language you will need to enter the CODE of the subject.

Name	Group	Language	Semester	Turn
(PAUL) Classroom practices	711	English	first semester	afternoon
(PLAB) Practical laboratories	711	English	first semester	afternoon
(TE) Theory	71	English	first semester	afternoon