COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI ICADE CIHS

Syllabus
2025 - 2026

TECHNICAL SHEET OF THE SUBJECT

Data of the subject	
Subject name	Datos Masivos/Big Data
Subject code	E000014025
Involved programs	Grado en Análisis de Negocios/Bachelor in Business Analytics [Third year]
Level	Reglada Grado Europeo
Quarter	Semestral
Credits	3,0 ECTS
Type	Obligatoria (Grado)
Department	Department of Telematics and Computer Sciencies
Coordinator	Carlos Miguel Vallez Fernández
Schedule	Check time schedule
Office hours	Arrange a meeting

Teacher Information	
Teacher	
Name	Carlos Miguel Vallez Fernández
Department	Department of Telematics and Computer Sciencies
EMail	cmvallez@icai.comillas.edu
Teacher	
Name	Gabriel Javier Maestroarena Rodas
Department	Department of Telematics and Computer Sciencies
EMail	gjmaestroarena@icai.comillas.edu
Teacher	
Name	Adriana Carla Molero Alonso
Department	Department of Telematics and Computer Sciencies
EMail	acmolero@icai.comillas.edu
Teacher	
Name	Emilio Martín Gallardo
Department	Department of Telematics and Computer Sciencies
EMail	emgallardo@icai.comillas.edu

SPECIFIC DATA OF THE SUBJECT

Contextualization of the subject

Contribution to the professional profile of the degree

Big Data technology is essential in the discipline of advanced analytics and massive data processing and a tool of increasingly widespread use in companies as an enabler of advanced analytics. Thus, the objective of the course is that the student learns the concepts related to Big Data, from its origin and history, characteristics and differences with respect to traditional systems, its parallel architecture and the differences with other parallel systems and ecosystem of the basic data cycle, from data acquisition, storage to processing. The capabilities offered by this technology will be related to the problems for which it is especially indicated, the needs it covers and the use cases that manage to provide differential value to companies and society.

At the end of the course, students will understand the usual blocks of a typical installation, how big data works, its usefulness and will know in which cases it should be applied, why and for what purpose, so that they can propose and defend specific use cases.

Prerequisites

Basic programming knowledge

Competencies - Objectives

Competences

Basic Competences

CB 02. Know how to apply and integrate their knowledge, their understanding of it, their scientific basis and their problem-solving skills in new and imprecisely defined environments, including multidisciplinary contexts, both research and highly specialized professionals.

CB 03. Know how to evaluate and select the appropriate scientific theory and the precise methodology of their fields of study to formulate judgments based on incomplete or limited information, including, when necessary and relevant, a reflection on the social or ethical responsibility linked to the solution proposed in each case.

CB 07. Be able to assume responsibility for their own professional development and specialization in one or more fields of study.

General Competences

GC. 1 Organizational and planning skills in the identification of problems in the context of massive data.

GC 02. Ability to analyze massive data from different sources: text, audio, numerical and image,

CG 03. Problem solving and decision making in a massive data environment, both quantitative and qualitative.

GC 07. Leadership and teamwork skills in the information society.

GC 8. Critical and self-critical capacity in the information society.

GC 9. Ethical commitment in the information society.

CG 11. Ability to learn and work autonomously in the information society.

Specific Competences

To understand the most important concepts associated with Big Data and its most characteristic elements.

Specific Competences of the Elective Course of Study

CEO 24. Know and use information analysis techniques to extract knowledge for business decision making.

Learning outcomes

CE25 To understand the most important concepts associated with Big Data and its most characteristic elements.

RA1. Know the concept of parallelism and the importance of the distribution of big data.

RA2. Identify the specific needs associated with the handling, acquisition and storage of massive volumes of data.

RA3. Understand and know the communications (networks) and computing technologies required for the distribution and retrieval of information, as well as its analysis.

CE24 To know and use information analysis techniques to extract knowledge for business decision making.

RA1 Knows and understands the decision support tools in the business environment.

RA2 Is able to implement a computerized decision support solution.

RA3 Understands and applies acquisition, representation and inference tools, and those techniques of automatic learning and based on intelligent systems.

THEMATIC BLOCKS AND CONTENTS

Contents - Thematic Blocks

1: Introduction to Big Data

What is Big Data and what is not Big Data?

The value of data.

Volumes and sources of data. Structured and unstructured information.

Utility, necessity and examples of Big Data use cases.

2: Life Cycle of a Big Data project.

Data Pipeline of a Big Data project (ingestion, storage, processing/computing and presentation/visualization).

Professional profiles.

Basic example of ingestion and associated problems due to available resources.

Introduction to governance, privacy and data protection. DPO profile, RGPD and ARCO rights.

3: Basic concepts about systems

Linux based operating systems. Introduction and basic commands.

Distributed Systems. Introduction and related concepts (Parallelization, Fault Tolerance, Balancing, Availability, Redundancy)

Virtualization concept. Introduction and platforms. Virtual machines vs. containers.

Datacenters: high speed networks and high performance computing.

4: Hadoop Ecosystem

Introduction to Hadoop. Components

Hardware and Software Architecture.

5: Mass Storage

HDFS file system.

Services and roles.

HUE. Data retrieval and visualization.

6: Introduction to Mass Processing.

YARN.

MapReduce.

Spark.

7: Cloud & Big Data

On-premise vs Cloud infrastructure.

Concept "as a service".

Cloud providers and platforms. Services and products offered.

8: Big Data technology in the enterprise

Impact of Big Data on Analytics and Integration.

TEACHING METHODOLOGY

General methodological aspects of the subject

In-class Methodology: Activities

1.Master class and general presentations. Exposure of the main concepts and procedures through the explanation by the professor. It will include dynamic presentations, small practical examples and spontaneous or formal participation of the students. In addition, the professor will propose the reading of topics prepared by the professor, scientific and informative articles that will allow to introduce or deepen in a topic, with the objective of contributing ideas and generating debate (24 hours).CE 25,CG 03, CG 02 CG 01, CB 03, CB 02

2.Resolution in class of practical problems. Resolution of some first problems to put the student in context. The resolution will be in charge of the teacher and the students in a cooperative way (2 hours). CE 25,CG 03, CG 02 CG 01, CB 03, CB 02

3.Practical. Each student will carry out in isolation or in group a series of regulated practices. The practices will end with the writing of a report or the writing of the different experiences (4 hours). CE 25,CG 03, CG 02 CG 01, CB 03, CB 02

.

Non-Presential Methodology: Activities

Individual study of the material. Activity carried out individually by the student to prepare the topic to be discussed in class. (60 hours). CE 25,CG 03, CG 02 CG 01, CB 03, CB 02

Resolution of practical problems to be solved outside class time by the student. The student must use and internalize the knowledge provided in the subject. The correction to the class will be made by some of the students or the teacher according to the cases. The individualized correction of each exercise will be carried out by the student or by a colleague according to the cases (exchange method) (10 hours).CE 25,CG 03, CG 02 CG 01, CB 03, CB 02

Practices of the subject. Two practical exercises that reflect the knowledge acquired in the course (10 hours).CE 25,CG 03, CG 02 CG 01, CB 03, CB 02

.

SUMMARY STUDENT WORKING HOURS

SUMMARY OF STUDENT WORK HOURS

CLASSROOM HOURS

Lecture 24

Discussions and problem solving 2

Practical exercises 4

Evaluation 2

NON-FACE-TO-FACE HOURS

Independent work on theoretical contents 9

Autonomous work on practical contents 9

Final practice of the course 10

Study 30

ECTS CREDITS:

3 (90 hours)

.

EVALUATION AND CRITERIA

Evaluation activities

Evaluation criteria	**WEIGHT**
Performance of exams:	

- Final exam.
- Comprehension of concepts by means of open tests, focused on practical cases.
- Application of acquired concepts
- Presentation and written communication.

60%

Evaluation of continuous theoretical performance:

- Performance of follow-up tests, discussions and exercises.
- Understanding of concepts through objective tests and one minute papers.
- Discussion and analysis of technical/scientific papers.
- Analysis and interpretation of the results obtained in problem solving through debates.

30%

Evaluation of continuous practical performance:

- Cases to be developed in the practical class.
- Understanding of concepts.
- Application of concepts to the development of cases on the new concepts raised.
- Application of all the knowledge acquired during the course.

10%

Ratings

The use of AI to create complete works or relevant parts, without citing the source or the tool (and the rest of the elements requested and detailed below) or without being expressly allowed in the description of the work, will be considered plagiarism and regulated according to the General Regulations of the University.

The use of AI is allowed only and exclusively for individual and/or group practices. It can be used to reformulate or refine texts as

well as to obtain code or help in the resolution of problems. However, it is necessary and mandatory to reference the sections where it has been used, reference which AI has been used, include the promp used and the answer obtained and finally include an explanation with your own words that demonstrate the understanding of the answer received (what the command does, the meaning of the paragraph, etc.).

t is expressly forbidden to use any AI and/or software to communicate with a person or AI during the completion of the class test/quiz and during the final exam. Failure to comply with this rule will be considered plagiarism and regulated according to the General Regulations of the University.

Grading Criteria

The grade in the **ordinary c**all of the course will be obtained as follows:

- Final exam (between 50.0 % and 60.0%).
  - **Final exam with a weight of 50%**
- Short exams of theoretical or practical content, developmental or test type (between 10.0 % and 30.0%)
  - **15% Short quiz style exams during classes**
- Evaluation of individual work (between 10.0 % and 10.0%)
  - **10% Individual practice reports**
- Evaluation of group work (between 10.0 % and 20.0%)
  - **15% Group practice reports**
- Active student participation in the classroom (between 10.0 % and 20.0%)
  - **10% attendance and participation in class (between 10.0% and 20.0%)**

The grade in the **extraordinary** call of the course will be obtained as follows:

- Final exam (between 50.0 % and 60.0 %).
-
- Final exam with a weight of 60%
- Short exams of theoretical or practical content, developmental or test type (between 10.0% and 30.0%).
-
- 10% Short quiz-style exams during the classes
- Evaluation of individual work (between 10.0 % and 10.0%)
-
- 10% Individual practice reports
- Evaluation of group work (between 10.0 % and 20.0%)
-

- 10% Group practice reports
- Active student participation in the classroom (between 10.0 % and 20.0%)
-
- 10% attendance and participation in class (between 10.0% and 20.0%)

In order to pass the course, students must have at least 5 points out of 10 in both the final exam and the practicals. Likewise, all the practicals must have been handed in. This applies both in the ordinary and extraordinary exams.

WORK PLAN AND SCHEDULE

Activities	Date of realization	Delivery date
Chapters 1,2,3,4 Quiz 1	Weeks 1,2,3	
Chapters 5 & 6	Weeks 4 & 5	
Group Practice 1 Quiz 2	Weeks 6 & 7	
Chapters 7 & 8 individual practice 1	Weeks 8 & 9	
Group Practice 2 Quiz 3	Week 10	
Chapters 9 & 10	Weeks 11 & 12	
Individual practice 2 Group practice 3 Prepare examn and doubts Quiz 4	Weeks 13 & 14	

BIBLIOGRAPHY AND RESOURCES

Basic Bibliography
Powerpoints available in Moodle

Complementary Bibliography

- Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.
- P. Zikopoulos, C. Eaton, D. DeRoos, T. Deutsch,&G. Lapis,(2012) Understanding Big Data, USA:McGraw-Hill Books
- White, T. (2009). Hadoop. The Definitive Guide. Vol. 1, OReilly Media. Inc, San Jose, CA, USA
- L. Joyanes (2013), Big Data: Análisis de grandes volúmenes de datos en organizaciones, Editorial Alfaomega
- Chambers, B., & Zaharia, M. (2018). Spark: The definitive guide: Big data processing made simple. " O'Reilly Media
- Holmes, A. (2012). Hadoop in practice. Manning Publications Co..
- Shotts, W. (2019). The Linux command line: a complete introduction. No Starch Press.
- Karau, H., Konwinski, A., Wendell, P., & Zaharia, M. (2015). Learning spark: lightning-fast big data analysis. " O'Reilly Media, Inc.".
- M. van Steen and A.S. Tanenbaum, Distributed Systems, 3rd ed., distributed-systems.net, 2017.
- Noah Iliinsky ad Julie (2011) Steele Data Visualizations O'Reilly
- Nathan Yau (2011) Visualize This: The FlowingData Guide to Design, Visualization, and Statistics
- Lutz, M. (2009). Python Pocket Reference: Python in Your Pocket" O'Reilly Media, Inc.".
- Lutz, M. (2013). Learning python: Powerful object-oriented programming. " O'Reilly Media, Inc.".