Syllabus
2024 - 2025

GENERAL INFORMATION

Data of the subject	
Subject name	Big Data Architecture
Subject code	DTC-IMAT-312
Mainprogram	Bachelor's Degree in Mathematical Engineering and Artificial Intelligence
Involved programs	Grado en Ingeniería Matemática e Inteligencia Artificial [Third year]
Credits	3,0 ECTS
Type	Obligatoria (Grado)
Department	Department of Telematics and Computer Sciencies

Teacher Information	
Teacher	
Name	Antonio Javier Samaniego Jurado
EMail	ajsamaniego@icai.comillas.edu
Teacher	
Name	Ignacio Pérez Torres
EMail	iptorres@icai.comillas.edu
Teacher	
Name	Guillermo Gallego Reina
Department	Department of Telematics and Computer Sciencies
EMail	ggallego@icai.comillas.edu

DESCRIPTION OF THE SUBJECT

Contextualization of the subject

Prerequisites

You must know the programming techniques acquired in previous courses.

Course contents

Contents

1. Introduction to Big Data
    - Data governance.
    - Life cycle / Roles.
    - Public data pools.
    - Governance tools.

- DataOps.

2. Introduction to distributed systems in Big Data environments
    - Networking.
    - Processing Units.
    - Parallelization.
    - Benchmarking.
    - Servers.
3. Hadoop ecosystem
    - Hadoop Introduction.
    - Cluster.
    - HUE.
    - Cloudera.
    - Docker.
4. HDFS Distributed Storage.
    - HDFS introduction.
    - Features.
    - Commands to use.
5. Distributed processing
    - YARN.
    - Introduction to Spark.
    - Introduction to MapReduce.
    - Benchmarking.
6. Infrastructures for the deployment of Big Data solutions
    - Introduction to Big Data solutions.
    - On-premise vs Cloud.
    - Environments:
        - Databricks, Google Colab, ...

EVALUATION AND CRITERIA

Evaluation activities	Evaluation criteria	Weight
Exams: • Intersemester Test. • Final exam.	• **Intersemester Test (20%):** Comprehension of the concepts in the Introduction of Big Data. • **Final Exam (50%):** The knowledge acquired in relation to Big Data Architecture will be evaluated.	70
Practical sessions: • Collaborative Challenges. • Non-face-to-face jobs. • Practices.	The attitude, **participation** and completion of the **practices** and **challenges** posed in collaborative and individual sessions.	10
	Final project of the subject that the student will	

Final project	deliver at the end of the course. This project will consist of the end-to-end design of a Big Data project following a use case proposed by the student/teacher.	20

Grading

The final grade in the ordinary and extraordinary call for the subject will depend on the evaluation of the following activities:

Final Grade = 20% Intersemester_Test + 50% Final_Exam + 10% Weekly Practices + 20% Final Project

To pass the subject, students must obtain at least 5 points out of 10 in the final exam of the subject and in the final practice, both in the ordinary and extraordinary calls.

Failure to attend 15% or more of the contact hours for this subject may result in the impossibility of taking the ordinary and extraordinary exams.

BIBLIOGRAPHY AND RESOURCES

Basic References

1. The Cloud Data Lake: A Guide to Building Robust Cloud Data Architecture
2. The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science
3. DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition
4. Ramcharan Kakarla, Sundar Krishnan, Sridhar Alla - Applied Data Science Using PySpark_ Learn the End-to-End Predictive Model-Building Cycle (2021, Apress)