OVERVIEW

Subject area	Principles of Big Data
Degree	Bachelor's Degree in Data Science
School/Faculty	*Faculty of Science, Engineering and Design*
Year	Third-party
ECTS	6 ECTS
Type	Core
Language(s)	Spanish
Delivery Mode	On campus
Semester	Semester 5

1. INTRODUCTION

It is essential nowadays to understand the principles of Big Data in our environment. The volume of data generated every day has to be managed and processed to increase the value of data within the business value chain. Principles of Big Data will introduce students to the world of mass and scalable data processing, as well as how to manage data on different levels - files to databases, on both a local level and in the Cloud.

This subject teaches students about design, development, monitoring and testing, as well as how to solve data extraction, transformation and load problems on a business level. It will also teach how to identify and apply optimisation methods for Big Data infrastructures in keeping with the resources available. This subject will give students an overall view of the most common infrastructures in Big Data, which will serve as a base for other subjects in the Bachelor's Degree in Data Science.

2. SKILLS AND LEARNING OUTCOMES

Basic skills (CB, by the acronym in Spanish):

- CB1: Students have shown their knowledge and understanding of a study area originating from general secondary school education, and are usually at the level where, with the support of more advanced textbooks, they may also demonstrate awareness of the latest developments in their field of study.
- CB2: Students can apply their knowledge to their work or vocation in a professional manner and possess the skills which are usually evident through the forming and defending of opinions and resolving problems within their study area
- CB3 - Students must have the ability to gather and interpret relevant data (usually within their study area) to form opinions which include reflecting on relevant social, scientific or ethical matters.

Cross-curricular skills (CT, by the acronym in Spanish):

- CT03 - Teamwork: ability to integrate and collaborate actively with other people, areas and/or organisations to reach common goals.
- CT05: Analysis and problem-solving: be able to critically assess information, break down complex situations, identify patterns and consider different alternatives, approaches and perspectives in order to find the best solutions and effective negotiations.
- CT08. Entrepreneurial spirit: ability to take on and carry out activities that generate new opportunities, foresee problems or lead to improvements.

Specific skills (CE, by the acronym in Spanish):

- CE10. Ability to apply Big Data methods, architecture and techniques to manage data effectively.
- CE11. Ability to apply computational learning techniques in order to design and implement applications and systems that use them, including those dedicated to the automatic extraction of information and knowledge from large volumes of data.
- CE12. Understand system interoperability techniques and data integration and aggregation.
- CE13. Ability to design efficient interfaces in the context of Big Data that guarantee accessibility and usability, using graphic and analytical representation techniques.

Learning outcomes (RA, by the acronym in Spanish):

- RA1: Use methods, architecture and techniques to store and manage large volume databases to solve practical cases.
- RA2: Interpret and apply models and standards in the field of big data systems to practical cases.
- RA3: Design, develop and evaluate graphical interfaces to visualise data making use of languages and specific environments. Propose alternative solutions and decide on the most suitable from a usability and user experience point of view.
- RA4: Implement computer applications which use large-volume databases. This includes the use of machine learning technology to obtain models, visualise data and how to interpret them.

Skills	Learning outcomes
CB1	RA2: Interpret and apply models and standards in the field of big data systems to practical cases.
CB2	RA1: Use methods, architecture and techniques to store and manage large volume databases to solve practical cases. RA2: Interpret and apply models and standards in the field of big data systems to practical cases.

CB3	RA2: Interpret and apply models and standards in the field of big data systems to practical cases. RA3: Design, develop and evaluate graphical interfaces to visualise data making use of languages and specific environments. Propose alternative solutions and decide on the most suitable from a usability and user experience point of view.
CT03	RA1: Use methods, architecture and techniques to store and manage large volume databases to solve practical cases.
CT05	RA2: Interpret and apply models and standards in the field of big data systems to practical cases. RA3: Design, develop and evaluate graphical interfaces to visualise data making use of languages and specific environments. Propose alternative solutions and decide on the most suitable from a usability and user experience point of view.
CT08	RA2: Interpret and apply models and standards in the field of big data systems to practical cases.
CE10	RA1: Use methods, architecture and techniques to store and manage large volume databases to solve practical cases.
CE11	RA3: Design, develop and evaluate graphical interfaces to visualise data making use of languages and specific environments. Propose alternative solutions and decide on the most suitable from a usability and user experience point of view.
CE12	RA2: Interpret and apply models and standards in the field of big data systems to practical cases. RA4: Implement computer applications which use large-volume databases. This includes the use of machine learning technology to obtain models, visualise data and how to interpret them.
CE13	RA4: Implement computer applications which use large-volume databases. This includes the use of machine learning technology to obtain models, visualise data and how to interpret them.

3. CONTENTS

Principles of Big Data

1. Introduction to Big Data (architecture and design)
2. Managing large volumes of data Non-conventional NoSQL databases BD query processing and trends in data processing

3. Cloud Databases
4. Data Parallelism
5. Interoperability and data aggregation

4. TEACHING/LEARNING METHODS

The types of teaching/learning methods are as follows:

- Collaborative learning: students learn to work with other people (colleagues and professors) to find creative, comprehensive and constructive solutions to questions and problems that arise from the given case studies, using relevant knowledge and available resources in relation to each subject.
- Problem-based learning: students face problems they must solve either working as a team or independently.
- Master Lecture: presentations by the professor using the appropriate technological tools to facilitate understanding of the subject matter.
- Directed academic activities: more independent tasks (individual or in groups), involving search for information, written summaries, debates and public defence of work.

5. LEARNING ACTIVITIES

The types of learning activities, plus the amount of time spent on each activity, are as follows:

On campus:

Learning activity	Number of hours
Master classes	44
Laboratory work, problem solving and case studies (exercises directed by the teacher)	10
Non-simultaneous activities (Digital Blocks)	8
Knowledge test.	4
Independent working	60
Tutorials	6
Practical work at home	21
TOTAL	**150**

6. ASSESSMENT

The assessment methods, plus their weighting in the final grade for the subject area, are as follows:

On campus:

Assessment system	Weighting
Final global knowledge test.	40%
Activities assigned in the Digital Blocks	10%
Classwork (on campus practical work)	20%
Project (presentation and final submission)	25%
Attendance and participation	5%
Total	100%

On the Virtual Campus, when you open the subject area, you can see all the details of your assessment activities and the deadlines and assessment procedures for each activity.

7. BIBLIOGRAPHY

The reference publication to accompany this subject area is:

- Hans Weber, Big Data: A Complete Guide to the Basic Concepts in Data Science, Cyber Security, Analytics and Metrics: 2 (Big Data and Artificial Intelligence)

The recommended bibliography is indicated below:

- Rafael Caballero, Enrique Martin (2018) BIG DATA con PYTHON. Recolección, almacenamiento y proceso
- Bernard Marr (2016) BIG DATA: Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance
- Judith Hiwitz (2013) Big Data for Dummies
- Jules Damji (2020) Learning Spark: Lightning-fast Data Analytics
- Maryann Kisamore (2021) Basics Of Apache Kafka: Open-Source Distributed Event Streaming Platform: Apache Kafka Platform
- Alberto Artasanchez (2021) AWS for Solutions Architects: Design your cloud infrastructure by implementing DevOps, containers, and Amazon Web Services