1. OVERVIEW

Subject area	Big Data
Degree	Bachelor's Degree in Computer Engineering
School/Faculty	Architecture, Engineering and Design
Year	Fourth
ECTS	6 ECTS
Type	Compulsory
Language(s)	Spanish
Delivery mode	On campus / Online
Semester	First semester
Year	2022/2023
Coordinating professor	Enrique Puertas

2. INTRODUCTION

Big Data is a compulsory subject within the Bachelor's Degree in Computer Engineering, worth 6 ECTS credits.

This subject area belongs to the Computer Science Subject and is considered a continuation of the subject areas Artificial Intelligence, Intelligent Systems and Databases, to further analyse the solutions to the challenge posed by big data in the areas mentioned above. This subject area introduces advanced content related to the infrastructure needed to process large amounts of data in a distributed way, and Machine Learning and Visualisation algorithms to extract value from the data.

3. SKILLS AND LEARNING OUTCOMES

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

- **CB4:** Students can communicate information, ideas, problems and solutions to both specialist and non-specialist audiences.

Transversal skills (CT, by the acronym in Spanish):

- **CT10:** Initiative and entrepreneurial spirit: Ability to undertake difficult or risky actions with resolve. Ability to anticipate problems, propose improvements and persevere to ensure they are implemented. Willingness to take on and carry out tasks.

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

- **CE31:** Ability to understand and develop computational learning techniques and design and implement applications and systems that use them, including those dedicated to automatic information extraction and knowledge from large volumes of data.

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

- **RA8:** Define, assess and select storage and processing platforms of large amounts of information, as well as forming them to be used in the analysis of this information.

The following table shows how the skills developed in the subject area match up with the intended learning outcomes:

Skills	Learning outcomes
CB4, CT10, CE31	**RA8**: Define, assess and select storage and processing platforms of large amounts of information, as well as forming them to be used in the analysis of this information.

4. CONTENTS

The subject is organised into six learning units, which in turn are divided into topics:

Topic 1. Introduction to Big Data Systems.

- What is Big Data?
- The 5 Vs of Big Data
- Distributed systems and fault tolerance

Topic 2. HADOOP Ecosystems.

- HDFS archive system.
- MapReduce programming paradigm.
- YARN resource management
- Apache Hive and Apache Pig

Topic 3. Storage systems and NoSQL Databases.

- Distributed storage.
- Separating compute and storage.
- NoSQL databases

- Key-Value Databases.
- Tabular Databases.
- Graph Databases.
- Document Databases.

Topic 4. Apache Spark.

- What is Apache Spark?
- Resilient Distributed Datasets and Spark Dataframes
- Spark SQL.
- Spark Machine Learning.

Topic 5. Data processing in Streaming.

- Introduction to Stream processing.
- Publisher/Subscriber messages and model.
- Windows and real-time processing.

Topic 6. Big Data Visualisation

- Data visualisation.
- Visual Metaphors.
- Real-time dashboards.

5. TEACHING/LEARNING METHODS

The types of teaching/learning methods are as follows:

- Lectures.
- Teamwork.
- Independent working.
- Independent reading on topics and discussion.
- Tutorials, academic monitoring and assessment.

6. LEARNING ACTIVITIES

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

On campus:

Learning activity (AF, by the acronym in Spanish)	Number of hours
Lectures, reading main topics and complementary materials	50 h
Group activities.	25 h
Independent work by the student.	50 h
Tutorials, academic monitoring and assessment	25 h
TOTAL	**150**

Online:

Learning activity (AF, by the acronym in Spanish)	Number of hours
Independent work by the student.	50 h
Group activities.	50 h
Independent reading of topics and discussion	25 h
Tutorials, academic monitoring and assessment	25 h
TOTAL	**150**

7. ASSESSMENT

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

On campus:

Assessment system	Weighting
Exams and tests.	30%
Development of articles, reports or design briefs.	30%
Portfolios, peer assessment.	15%
Conferences, visits to companies and institutions.	10%
Case studies, designs, simulations and research.	15%

Online:

Assessment system	Weighting
Exams and tests.	60%
Development of articles, reports or design briefs.	10%
Portfolios, peer assessment.	10%
Case studies, designs, simulations and research.	20%

On the Campus Virtual, when you open the subject area, you will find all the details of your assessable tasks and the deadlines and assessment procedures for each task.

7.1. Ordinary exam period

To pass the subject area in the ordinary exam period, you will need a final grade of at least 5 out of 10. In addition, the following conditions must be met:

- In the practical continuous assessment activities (activities 1-5), you must have at least 70% class attendance (on-campus delivery mode only).
- Achieve a grade higher than 5 in the final test activity.

7.2. Extraordinary exam period (resits)

To pass the subject area in the ordinary exam period, you will need a final grade of at least 5.0 out of 10.0 (weighted average) for the subject area.

In any case, you will need a grade of at least 4.0 in the final test for it to be included in the weighting with the other activities.

Activities not passed in the ordinary exam period, or those not submitted, must be submitted after receiving the relevant corrections and feedback from the lecturer. Activities in the extraordinary exam period will be carried out individually and the professor may change the wording with respect to the activities in the ordinary exam period.

8. TIMELINE

The timeline with submission dates for the assessable tasks in this subject area will be indicated in this section:

Assessable tasks	Date
Activity 1. Research project on the Big Data sector.	Week 2
Activity 2. Storage practice.	Week 4
Activity 3. MapReduce programming.	Week 7
Activity 4. Apache spark programming.	Weeks 11–12
Activity 5. Data processing in Streaming	Weeks 15–16
Activity 6. Development of a Big Data end-to-end analysis project.	Week 18
Activity 7. Final test	Week 18

The timeline may be subject to change for logistical reasons related to the activities. Students will be informed of any changes in due time and course.

9. BIBLIOGRAPHY

The recommended bibliography is indicated below:

- EIJKHOUT, V. (2014). Introduction to High Performance Scientific Computing. (2ª edición).
- TANENBAUM, AS., VAN STEEN, M. (2001). Distributed Systems: Principles and Paradigms. (2ª edición). Pearson Prentice Hall.
- WHITE, T. (2015). Hadoop: The Definitive Guide. O'Reilly.
- ODERSKY, M., SPOON, L. & VENNERS, B. (2008). Programming in Scala. Ed. Artima.
- ODERSKY, M. (2010). Scala by Example. Programming Method Laboratory, Switzerland: EPFL.
- WAMPLER, D. & PAYNE, A. (2009). Programming Scala. Ed. O'Reilly.
- KARAU, H., KONWINSKI, A., WENDELL, P. & ZAHARIA, M. (2015). Learning Spark. O'Reilly.
- BIHANIC, D. (2015). New Challenges for Data Design. Springer.
- BIHANIC, D. (2015). Empowering Users through Design. Springer.
- FISHER, D., MEYER, M. (2016). Making sense of data. O'Reilly.
- GRAY. J., BOUNEGRU, L., CHAMBERS, L. (2012). The Data Journalism Handbook. O'Reilly.
- ILIINSKY, N., STEELE, J. (2011). Designing Data Visualizations. O'Reilly.
- MUNZNER, T. (2014). Visualization Analysis & Design. CRC Press.
- RYAN, L. (2016). The Visual Imperative. Creating a Visual Culture of Data Discovery.
- SIMON, P. (2014). The Visual Organization. Wiley.
- TELEA, A. (2007). Data Visualization Principles and Practice.
- WARD, M., GRINSTEIN, G., KEIM, D. (2015). Interactive Data Visualisation.
- YAU, N. (2013). Data Points. Visualization That Means Something. Wiley.

10. DIVERSITY AWARENESS UNIT

Students with special educational needs:
To ensure equal opportunities, curricular adaptations or adjustments for students with special educational needs will be outlined by the Diversity Awareness Unit (UAD, Spanish acronym).

As an essential requirement, students with special educational needs must obtain a report about the curricular adaptations/adjustments from the Diversity Awareness Unit by contacting unidad.diversidad@universidadeuropea.es at the beginning of each semester.

11. STUDENT SATISFACTION SURVEYS

Your opinion matters!

Universidad Europea encourages you to complete our satisfaction surveys to identify strengths and areas for improvement for staff, degrees and the learning process.

These surveys will be available in the survey area of your campus virtual or by email.

Your opinion is essential to improve the quality of the degree.

Many thanks for taking part.