

Academic Year: ( 2023 / 2024 )

Review date: 23-01-2024

Department assigned to the subject: Computer Science and Engineering Department

Coordinating teacher: GARCIA BLAS, FRANCISCO JAVIER

Type: Compulsory ECTS Credits : 3.0

Year : 1 Semester : 1

## OBJECTIVES

### Basic Skills

- \* Knowledge and understanding that provide a basis or opportunity for originality in developing and / or applying ideas, often in a research context
- \* That the students can apply the broader (or multidisciplinary) acquired knowledge and ability to solve problems in new or unfamiliar environments within contexts related to their field of study
- \* Students must possess the learning skills that enable them to continue studying in a way that will be largely self-directed or autonomous.

### General Competencies

- \* Apply the theoretical underpinnings of the techniques for the high-performance processing of large volumes of data as a basis for the development and adaptation of such techniques to specific problems
- \* Identify different techniques and paradigms for processing large amounts of data, and differentiate them according to their theoretical and practical features
- \* Use skills for teamwork and getting along with other independently

### Specific Skills

- \* Apply basic knowledge of big data programming techniques using advanced technologies and methods for treating large volumes of data
- \* Identify opportunities that data processing techniques can make to the improvement of the activity of enterprises and organizations
- \* Provide basic and fundamental knowledge of big data processing frameworks
- \* Identify and select suitable frameworks and software tools for the treatment of large amounts of data
- \* Making efficient use of distributed platforms for high-performance data processing

### Learning Results

- \* Manage the basics of big data processing frameworks.
- \* Ability to use high-performance architectures and technologies for large volumes of data.
- \* Knowledge of design techniques and application development of high-performance big data computing.
- \* Skills to analyze and model the most appropriate frameworks for each problem, adapting to the specifications of individual cases

## DESCRIPTION OF CONTENTS: PROGRAMME

1. Introduction to Big Data Processing
2. MapReduce Paradigm
3. Storage Systems Big Data environments
  - \* HDFS as distributed file system
  - \* Commands for managing files in HDFS
4. Frameworks for intensive computing data
  - \* Introduction to Apache Hadoop
  - \* Functional Programming in Scala
  - \* Apache Spark
  - \* Access and processing a large volume of data
  - \* Streaming Data Processing
4. Management computational resources
  - \* Introduction to Apache Yarn
  - \* Deploying applications in corporate Big Data environments

- \* Tools for monitoring Big Data applications

## LEARNING ACTIVITIES AND METHODOLOGY

Learning activities:

- \* Lectures
- \* Hands-on and lab projects
- \* Personal student work.

Teaching methodology:

- \* Presential lectures imparted in the class, using multimedia and informatics support, to develop the main concepts of the course. Reading materials will be provided to complement students knowledge.
- \* Reading of recommended texts, from papers, technical journals, manuals and reports, to extend the student knowledge of the subject topics.
- \* Solving practical jobs, problems, etc. proposed in class (individually or in groups).

## ASSESSMENT SYSTEM

1.- Continuous evaluation (40%)

- \* Class activities
- \* Individual or collective projects made along the course

2.- Final exam (60%)

It is mandatory to obtain at least 4 points over 10 in each of the evaluable parts of the subject.

<b>% end-of-term-examination:</b>	60
<b>% of continuous assessment (assignments, laboratory, practicals...):</b>	40

## BASIC BIBLIOGRAPHY

- Holden Karau, Andy Konwinski, Patrick Wendell & Matei Zaharia Learning Spark, O'Reilly, 2015
- Martin Odersky, Lex Spoon, Bill Venners Programming in Scala, Artima.