

## FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Optativa Complementaria. Tecnologías de procesamiento Big Data
Código	DTC-MBD-515
Título	Máster en Big Data. Tecnología y Analítica Avanzada/Master in Big Data Technologies and Advanced Analytics
Impartido en	Máster en Big Data. Tec. y Analítica Avanzada/Master in Big Data Technologies and Advanced Analytics [Primer Curso]
Nivel	Master
Cuatrimestre	Semestral
Créditos	7,5 ECTS
Carácter	Obligatoria
Departamento / Área	Departamento de Telemática y Computación

Datos del profesorado	
<b>Profesor</b>	
Nombre	Jorge Ayuso Rejas
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	jayuso@icai.comillas.edu
<b>Profesor</b>	
Nombre	Patricia Alcalde Sanz
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	palcalde@icai.comillas.edu

## DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura
<p><b>Aportación al perfil profesional de la titulación</b></p> <p>The purpose of the course is to give an overview of the ingestion and processing tools of the big data environment, especially focused on Spark and Hive.</p> <p>By the end of the course, students will:</p> <ul style="list-style-type: none"> <li>• Be able to choose which is the most appropriate tool to extract data from different sources and take it to a Hadoop cluster.</li> <li>• Have experience with some processing tools and languages (python, sql, etc).</li> <li>• Have deep knowledge of spark with python and how to optimize jobs.</li> </ul>
<p><b>Prerequisitos</b></p> <p>Students willing to take this course should be familiar with any programming language, preferably python or SQL and with Linux</p>



commands and utilities.

## Competencias - Objetivos

## BLOQUES TEMÁTICOS Y CONTENIDOS

### Contenidos – Bloques Temáticos

#### Contents

#### Theory

#### Unit 1. Introduction to software development

1. IntelliJ
2. Git
3. Introduction to SQL

#### Unit 2. Hadoop Ecosystem

1. Hadoop Ecosystem
2. HDFS and Hadoop client
3. Hive

#### Unit 3. Processing tools and ETLs

1. Kafka
2. Search Engines
3. NiFi

#### Unit 3. Introduction to Data scientist

1. Python first steps
2. Scientific Python

#### Unit 4. Apache Spark for Data scientist

1. Spark DataFrame
2. Spark ML (Machine Learning)
3. Spark packages

#### Laboratory

All sessions will have a hands-on approach. In the development of the course will be proposed to students practices that will be 60% of the final grade.

## METODOLOGÍA DOCENTE



## Aspectos metodológicos generales de la asignatura

### EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Assessment activities	Grading criteria	Weight
Practices	<ul style="list-style-type: none"><li>Mean of student's practices (0-10 points), all practices must be passed</li></ul>	60%
Final exam	<ul style="list-style-type: none"><li>Understanding of the theoretical concepts.</li><li>Application of these concepts to problem-solving.</li><li>Critical analysis of numerical exercises' results.</li></ul>	40%

### BIBLIOGRAFÍA Y RECURSOS

#### Bibliografía Básica

- Notes and notebooks prepared by the lecturer (available in Moodle).
- White, T. (2015). Hadoop: The definitive guide 4th edition. " O'Reilly Media, Inc."
- Shreedharan, Hari (2014). Using Flume " O'Reilly Media, Inc."
- Karau, H., Konwinski, A., Wendell, P., & Zaharia, M. (2015). Learning spark: lightning-fast big data analysis. " O'Reilly Media, Inc."
- VanderPlas, J. (2016). Python Data Science Handbook.

En cumplimiento de la normativa vigente en materia de **protección de datos de carácter personal**, le informamos y recordamos que puede consultar los aspectos relativos a privacidad y protección de datos [que ha aceptado en su matrícula](#) entrando en esta web y pulsando "descargar"

<https://servicios.upcomillas.es/sedelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792>