

# ANX-PR/CL/001-01

## GUÍA DE APRENDIZAJE

### ASIGNATURA

**103001003 - Big Data**

### PLAN DE ESTUDIOS

**10AZ - Master Universitario En Innovación Digital**

### CURSO ACADÉMICO Y SEMESTRE

**2023/24 - Primer semestre**

## Índice

---

### Guía de Aprendizaje

1. Datos descriptivos.....	1
2. Profesorado.....	1
3. Competencias y resultados de aprendizaje.....	2
4. Descripción de la asignatura y temario.....	2
5. Cronograma.....	4
6. Actividades y criterios de evaluación.....	6
7. Recursos didácticos.....	8
8. Otra información.....	9

## 1. Datos descriptivos

### 1.1. Datos de la asignatura

Nombre de la asignatura	103001003 - Big Data
No de créditos	3 ECTS
Carácter	Optativa
Curso	Primer curso
Semestre	Primer semestre
Período de impartición	Septiembre-Enero
Idioma de impartición	Castellano
Titulación	10AZ - Master Universitario en Innovación Digital
Centro responsable de la titulación	10 - Escuela Tecnica Superior De Ingenieros Informaticos
Curso académico	2023-24

## 2. Profesorado

### 2.1. Profesorado implicado en la docencia

Nombre	Despacho	Correo electrónico	Horario de tutorías *
Antonio Latorre De La Fuente (Coordinador/a)	4101	a.latorre@upm.es	Sin horario.
Julian Arenas Guerrero	4204	julian.arenas.guerrero@upm.es	Sin horario.
Jorge Acosta Hernandez	4201	jorge.acosta@upm.es	Sin horario.

\* Las horas de tutoría son orientativas y pueden sufrir modificaciones. Se deberá confirmar los horarios de tutorías con el profesorado.

## 3. Competencias y resultados de aprendizaje

---

### 3.1. Competencias

CB07 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

CE-CD05 - Capacidad para usar herramientas de procesamiento de big data tanto en online como en modo batch

CE-CD06 - Capacidad para extraer, integrar y consultar datos heterogéneos en diferentes escenarios

### 3.2. Resultados del aprendizaje

RA33 - Conocer técnicas de visualización y procesos de análisis de datos, y de programación, diseño y depuración de algoritmos, para computación de altas prestaciones.

RA35 - Ser capaz de procesar datos masivos.

RA34 - Conocer cómo se aplican las técnicas de computación científica en algún campo específico de ciencia o ingeniería.

## 4. Descripción de la asignatura y temario

---

### 4.1. Descripción de la asignatura

This course will allow the student to gain the fundamentals for the analysis of large volumes of data. With an eminently practical approach, the technologies and fundamentals necessary to successfully accomplish the whole data analysis process will be presented in the context of Big Data, from the raw data to the models derived from them.

## 4.2. Temario de la asignatura

1. Introduction to Big Data
  - 1.1. Architectures and applications
  - 1.2. Data types
  - 1.3. Visual analytics
2. Big Data Ecosystem
3. Big Data Technologies
  - 3.1. Technological Challenges
  - 3.2. Basic solution: gfs + MapReduce
  - 3.3. Hadoop (hdfs + yarn)
  - 3.4. Pig
  - 3.5. Hive
  - 3.6. Beyond MapReduce
    - 3.6.1. Tez
    - 3.6.2. Spark
    - 3.6.3. Flink
4. Spark
  - 4.1. Spark Basics
  - 4.2. Spark Applications
  - 4.3. Spark SQL
5. Machine Learning with Spark
  - 5.1. Brief review of Machine Learning basics
  - 5.2. Spark MLlib

## 5. Cronograma

### 5.1. Cronograma de la asignatura \*

Sem	Actividad en aula	Actividad en laboratorio	Tele-enseñanza	Actividades de evaluación
1				
2				
3				
4				
5				
6				
7				
8				
9	<b>Lesson 1</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral  <b>Lesson 2</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral			
10	<b>Lesson 2</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral  <b>Lesson 2</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas			
11	<b>Lesson 3</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral	<b>Practical Work</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
12	<b>Lesson 3</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas	<b>Practical Work</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
13	<b>Lesson 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral  <b>Lesson 4</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas	<b>Practical Work</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
14	<b>Lesson 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral  <b>Lesson 4</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas	<b>Practical Work</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		

15	<b>Lesson 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral	<b>Practical Work</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
16	<b>Lesson 5</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral  <b>Lesson 5</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas	<b>Practical Work</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		<b>Assignment Deadline</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final No presencial Duración: 00:00
17				<b>Final Exam</b> EX: Técnica del tipo Examen Escrito Evaluación continua y sólo prueba final Presencial Duración: 02:00

Para el cálculo de los valores totales, se estima que por cada crédito ECTS el alumno dedicará dependiendo del plan de estudios, entre 26 y 27 horas de trabajo presencial y no presencial.

\* El cronograma sigue una planificación teórica de la asignatura y puede sufrir modificaciones durante el curso derivadas de la situación creada por la COVID-19.

## 6. Actividades y criterios de evaluación

### 6.1. Actividades de evaluación de la asignatura

#### 6.1.1. Evaluación (progresiva)

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
16	Assignment Deadline	TG: Técnica del tipo Trabajo en Grupo	No Presencial	00:00	80%	4 / 10	CB07 CB10 CE-CD05 CE-CD06
17	Final Exam	EX: Técnica del tipo Examen Escrito	Presencial	02:00	20%	4 / 10	CB07 CB10 CE-CD05 CE-CD06

#### 6.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
16	Assignment Deadline	TG: Técnica del tipo Trabajo en Grupo	No Presencial	00:00	80%	4 / 10	CB07 CB10 CE-CD05 CE-CD06
17	Final Exam	EX: Técnica del tipo Examen Escrito	Presencial	02:00	20%	4 / 10	CB07 CB10 CE-CD05 CE-CD06

#### 6.1.3. Evaluación convocatoria extraordinaria

No se ha definido la evaluación extraordinaria.



## 6.2. Criterios de evaluación

### Regular call

This section covers the evaluation criteria for this course. All the students enrolled in this course will be subject, by default, to the progressive evaluation scheme. For this reason, this learning guide will be focused on this evaluation approach and it details all the evaluation activities in the timeline of the course.

The evaluation of the course will take into account both the theoretical and the practical knowledge acquired in the lectures and in the practical work carried out during the course, respectively.

This course will be evaluated in two ways:

- **Final exam.** At the end of the course, there will be a final exam covering all the contents presented during the course.
- **Practical work.** This assignment will be presented during the course, at class, in the date detailed in the timeline of the course. There will be some classes devoted to this assignment, in which the students will count with the support of the instructor, which should be, in general, complemented with autonomous work by the student. The deadline for the assignment will be fixed at the end of the term, as shown in the timeline of the course. No late assignments will be accepted for evaluation. The delivery of the assignment is considered a mandatory activity in order to pass the course.

The **final grade** for this course will be computed as follows: 20% for the final exam and 80% for the assignment. To pass the course, **a minimum score of 4** is required for each of these parts and a **grand mean** of 5 is needed combining these two items of evaluation.

### Extraordinary call

If the students do not succeed in this course, they will have to repeat the evaluation activities not passed in the ordinary evaluation. There will be a new call for the final exam as well as a new deadline for the assignment with the same requirements as in the regular call.

## 7. Recursos didácticos

### 7.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
Book 1	Bibliografía	Jiawei Han, Micheline Kamber, Data Mining : Concepts and Techniques, 2nd edition, Morgan Kaufmann, ISBN 1558609016, 2006.
Book 2	Bibliografía	Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson Addison Wesley, ISBN: 0321321367, 2005
Book 3	Bibliografía	Ian Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition, Morgan Kaufmann, ISBN: 0120884070, 2005.
Book 4	Bibliografía	Ian Witten, Eibe Frank, Mark Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition, Morgan Kaufmann, ISBN: 978-0-12-374856-0, 2011.
Book 5	Bibliografía	Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia. Learning Spark: Lightning-Fast Big Data Analysis. O'Reilly Media. 2015.
Book 6	Bibliografía	Sandy Ryza, Uri Laserson, Sean Owen, Josh Wills. Advanced Analytics with Spark: Patterns for Learning from Data at Scale. O'Reilly Media. 2015.
Spark documentation	Recursos web	<a href="http://spark.apache.org/docs/latest/">http://spark.apache.org/docs/latest/</a>

Web site of the course	Recursos web	UPM Moodle
Hive documentation	Recursos web	<a href="https://cwiki.apache.org/confluence/display/Hive/Home">https://cwiki.apache.org/confluence/display/Hive/Home</a>

## 8. Otra información

---

### 8.1. Otra información sobre la asignatura

This course is jointly offered with other Master Programmes and lectures are delivered in English.