



UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

PROCESO DE  
COORDINACIÓN DE LAS  
ENSEÑANZAS PR/CL/001



E.T.S. de Ingenieros  
Informaticos

# 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

2021/22 - 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.....	3
5. Cronograma.....	5
6. Actividades y criterios de evaluación.....	8
7. Recursos didácticos.....	10
8. Otra información.....	11

## 1. Datos descriptivos

---

### 1.1. Datos de la asignatura

<b>Nombre de la asignatura</b>	103001003 - Big Data
<b>No de créditos</b>	3 ECTS
<b>Carácter</b>	Optativa
<b>Curso</b>	Primer curso
<b>Semestre</b>	Primer semestre
<b>Período de impartición</b>	Septiembre-Enero
<b>Idioma de impartición</b>	Castellano
<b>Titulación</b>	10AZ - Master Universitario en Innovación Digital
<b>Centro responsable de la titulación</b>	10 - Escuela Tecnica Superior De Ingenieros Informaticos
<b>Curso académico</b>	2021-22

## 2. Profesorado

---

### 2.1. Profesorado implicado en la docencia

<b>Nombre</b>	<b>Despacho</b>	<b>Correo electrónico</b>	<b>Horario de tutorías</b> *
Antonio Latorre De La Fuente (Coordinador/a)	4202	a.latorre@upm.es	Sin horario.
Pablo Toharia Rabasco	4102	pablo.toharia@upm.es	Sin horario.
Jesus Montes Sanchez	4204	jesus.montes@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

CB06 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación

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-CD02 - Capacidad para aplicar técnicas para la generación de visualizaciones adecuadas para el análisis y la exploración de datos en un contexto determinado, y para la correcta comunicación de los resultados del análisis

CE-CD03 - Capacidad para seleccionar las técnicas y herramientas para visualización de grandes cantidades de datos más adecuadas para resolver un determinado problema de ciencia de datos

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

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

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.

## 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. Brief Introduction to Scala
  - 4.3. Spark Applications

#### 4.4. 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 presencial en aula	Actividad presencial en laboratorio	Tele-enseñanza	Actividades de evaluación
1	<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		<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	
2	<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		<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	
3	<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	<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	
4	<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	<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	
5	<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	<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	

6	<p><b>Lesson 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Lesson 4</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>	<p><b>Practical Work</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio</p>	<p><b>Lesson 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Lesson 4</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p> <p><b>Practical Work</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio</p>	
7	<p><b>Lesson 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p>	<p><b>Practical Work</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>	<p><b>Lesson 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Practical Work</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>	
8	<p><b>Lesson 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p>	<p><b>Practical Work</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>	<p><b>Lesson 4</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Practical Work</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>	
9	<p><b>Lesson 5</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Lesson 5</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>	<p><b>Practical Work</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio</p>	<p><b>Lesson 5</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Lesson 5</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p> <p><b>Practical Work</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio</p>	
10				<p><b>Assignment Deadline</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final No presencial Duración: 00:00</p>
11				
12				
13				
14				
15				
16				
17				<p><b>Final Exam</b> EX: Técnica del tipo Examen Escrito Evaluación continua y sólo prueba final Presencial Duración: 02:00</p>

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 continua

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
10	Assignment Deadline	TG: Técnica del tipo Trabajo en Grupo	No Presencial	00:00	80%	4 / 10	CE-CD03 CE-CD05 CB10 CE-CD02 CE-CD06 CB07 CB06
17	Final Exam	EX: Técnica del tipo Examen Escrito	Presencial	02:00	20%	4 / 10	CE-CD03 CE-CD05 CB10 CE-CD02 CE-CD06 CB07 CB06

#### 6.1.2. Evaluación sólo prueba final

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
10	Assignment Deadline	TG: Técnica del tipo Trabajo en Grupo	No Presencial	00:00	80%	4 / 10	CE-CD03 CE-CD05 CB10 CE-CD02 CE-CD06 CB07 CB06
17	Final Exam	EX: Técnica del tipo Examen Escrito	Presencial	02:00	20%	4 / 10	CE-CD03 CE-CD05 CB10 CE-CD02 CE-CD06 CB07 CB06

### 6.1.3. Evaluación convocatoria extraordinaria

No se ha definido la evaluación extraordinaria.

## 6.2. Criterios de evaluación

### Schedule

This schedule is preliminary and subject to changes that will be communicated on the website of the course.

### Continuous evaluation

This section covers the evaluation criteria for this course. All the students enrolled in this course will be subject, by default, to the continuous evaluation scheme. For this reason, this learning guide will be focused on this approach and details all the evaluation activities in the timeline of the course. Those students interested in the final examination evaluation scheme are referred to the next section of this document.

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 schedule of the course. There will be some classes devoted to this assignment, where the students will count with the support of the instructor, that should be, in general, complemented with autonomous work by the student. No late assignments will be accepted for evaluation.

The **final grade** for this 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.

### Final exam evaluation

In general, the regulations for this evaluation scheme will be the same as for the continuous evaluation option. In particular:

- The students will to conduct the same practical work without the in-class support of the instructors.
- The deadlines for the assignments will be the same as for the continuous evaluation scheme.

## Extraordinary evaluation in July

If the student does not succeed in this course, she will have to repeat those parts 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 of the course.

## 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 the EIT-Digital Master in Data Science and lectures are delivered in English.