



UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

PROCESO DE  
COORDINACIÓN DE LAS  
ENSEÑANZAS PR/CL/001



E.T.S. de Ingeniería de  
Sistemas Informáticos

# ANX-PR/CL/001-01

## GUÍA DE APRENDIZAJE

### ASIGNATURA

**613000133 - Optimización Para Grandes Volúmenes De Datos**

### PLAN DE ESTUDIOS

61AH - Máster Universitario En Aprendizaje Automático Y Datos Masivos

### CURSO ACADÉMICO Y SEMESTRE

2024/25 - Segundo semestre

## Índice

---

### Guía de Aprendizaje

1. Datos descriptivos.....	1
2. Profesorado.....	1
3. Conocimientos previos recomendados.....	2
4. Competencias y resultados de aprendizaje.....	2
5. Descripción de la asignatura y temario.....	4
6. Cronograma.....	6
7. Actividades y criterios de evaluación.....	8
8. Recursos didácticos.....	10
9. Otra información.....	11

## 1. Datos descriptivos

### 1.1. Datos de la asignatura

<b>Nombre de la asignatura</b>	613000133 - Optimización para Grandes Volúmenes de Datos
<b>No de créditos</b>	3 ECTS
<b>Carácter</b>	Obligatoria
<b>Curso</b>	Primer curso
<b>Semestre</b>	Segundo semestre
<b>Período de impartición</b>	Febrero-Junio
<b>Idioma de impartición</b>	Castellano
<b>Titulación</b>	61AH - Máster Universitario en Aprendizaje Automático y Datos Masivos
<b>Centro responsable de la titulación</b>	61 - Escuela Técnica Superior De Ingeniería De Sistemas Informáticos
<b>Curso académico</b>	2024-25

## 2. Profesorado

### 2.1. Profesorado implicado en la docencia

<b>Nombre</b>	<b>Despacho</b>	<b>Correo electrónico</b>	<b>Horario de tutorías *</b>
Bonifacio Alberto Mozo Velasco (Coordinador/a)	4313	a.mozo@upm.es	Sin horario. Sin horario. Las tutorías serán concertadas vía correo electrónico.

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

## 2.3. Profesorado externo

Nombre	Correo electrónico	Centro de procedencia
Amit Karamchandani Batra	amit.kbatra@upm.es	Colaboracion docente de Predoctoral PIF (supervisado por A. Mozo)
Luis De La Cal Garcia	l.delacal@upm.es	Colaboracion docente de Predoctoral PIF (supervisado por A. Mozo)

## 3. Conocimientos previos recomendados

---

### 3.1. Asignaturas previas que se recomienda haber cursado

El plan de estudios Máster Universitario en Aprendizaje Automático y Datos Masivos no tiene definidas asignaturas previas recomendadas para esta asignatura.

### 3.2. Otros conocimientos previos recomendados para cursar la asignatura

- Lenguaje Python

## 4. Competencias y resultados de aprendizaje

---

### 4.1. Competencias

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.

CB6 - 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

CB7 - 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

CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las

responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades

CE02 - Gestionar grandes volúmenes de información procedentes de diversas fuentes heterogéneas y efectuar su análisis en tiempo real mediante el diseño de cuadros de mando e informes para obtener conocimiento útil.

CE05 - Diseñar y desplegar modelos de aprendizaje automático de manera escalable y eficiente usando estrategias de contenerización y computación de altas prestaciones.

CG1 - Capacidad para aplicar el método científico y saber organizar y planificar experimentos con rigor metodológico en el ámbito del aprendizaje automático y los datos masivos

CG2 - Participar en la aplicación de mecanismos de descripción, cuantificación, análisis, interpretación y evaluación de resultados experimentales del ámbito de los datos masivos y el aprendizaje automático

CG3 - Capacidad para reunir e interpretar datos masivos relevantes para emitir juicios que incluyan una reflexión sobre temas importantes de índole científico, social o ético en el ámbito del aprendizaje automático y los datos masivos

CG4 - Capacidad de aplicar iniciativa, integración, colaboración y potenciación de la discusión crítica en el ámbito del trabajo en equipo dentro del ámbito del aprendizaje automático y datos masivos

CG5 - Participar en la transmisión de la información generada, las ideas, los problemas y las soluciones de forma oral y escrita para un público tanto especializado como no especializado

CT1 - Creatividad

CT2 - Organización y planificación

CT3 - Gestión de la información

CT4 - Liderazgo de equipos

CT5 - Trabajo en contextos internacionales

K05 - El alumno analiza las distintas arquitecturas para el almacenamiento y procesado de datos masivos de altas prestaciones

S02 - El alumno planifica y ejecuta la gestión y el despliegue de infraestructuras de datos masivos

## 4.2. Resultados del aprendizaje

RA37 - Capacidad de analizar y diseñar arquitecturas y tecnologías de altas prestaciones para grandes volúmenes de datos

RA38 - Conocer y aplicar técnicas de diseño y desarrollo de aplicaciones de computación de altas prestaciones

RA39 - Analizar y modelar las tecnologías más adecuadas para cada problema, adecuándose a las especificaciones de los casos concretos.

## 5. Descripción de la asignatura y temario

---

### 5.1. Descripción de la asignatura

El paradigma Map-Reduce: programación de aplicaciones masivamente paralelas utilizando Map-Reduce

Frameworks para la computación intensiva en datos: Hadoop y Spark

Procesamiento de datos masivos en tiempo real (streaming): Kafka

Paralelización masiva de algoritmos de Machine Learning.

Caso de uso de paralelización de Algoritmos de Machine Learning Supervisado: Regresión Logística con Gradiente Descendente

Caso de uso de paralelización de Algoritmos de Machine Learning No Supervisado: K-Means

## 5.2. Temario de la asignatura

1. Programación de aplicaciones masivamente paralelas
  - 1.1. El paradigma Map/Reduce: programación de aplicaciones paralelas utilizando el paradigma Map-Reduce
  - 1.2. Frameworks para la computación intensiva en datos: Hadoop y Spark
  - 1.3. Procesamiento de datos masivos en tiempo real (streaming): Kafka
2. Paralelización masiva de algoritmos de Machine Learning.
  - 2.1. Caso de uso de paralelización de Algoritmos de Machine Learning Supervisado: Regresión Logística con Gradiente Descendente
  - 2.2. Caso de uso de paralelización de Algoritmos de Machine Learning No Supervisado: K-Means

## 6. Cronograma

### 6.1. Cronograma de la asignatura \*

Sem	Actividad tipo 1	Actividad tipo 2	Tele-enseñanza	Actividades de evaluación
1	<b>Clase de teoría Paralelización con Map-Reduce</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral	<b>Prácticas. Programación con Map-Reduce: Cuenta palabras, Cálculo de Medias y Varianzas en dataset</b> Duración: 03:00 PL: Actividad del tipo Prácticas de Laboratorio		
2	<b>Clase de teoría Programación paralela con Spark</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral	<b>Prácticas. Programación con Map-Reduce: Normalización de un dataset</b> Duración: 03:00 PL: Actividad del tipo Prácticas de Laboratorio		
3	<b>Clase de teoría. API básico de Spark</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral	<b>Prácticas. P1: Paralelización con Spark de un Clasificador (ML supervisado)</b> Duración: 03:00 PL: Actividad del tipo Prácticas de Laboratorio		
4	<b>Clase de teoría API extendido de Spark</b> Duración: 03:00 LM: Actividad del tipo Lección Magistral	<b>Prácticas. P1: Paralelización con Spark de un Clasificador (ML supervisado)</b> Duración: 03:00 PL: Actividad del tipo Prácticas de Laboratorio		
5	<b>Clase de teoría API extendido de Spark</b> Duración: 02:30 LM: Actividad del tipo Lección Magistral	<b>Prácticas. P2: Paralelización con Spark de K-Means (ML no supervisado)</b> Duración: 02:30 PL: Actividad del tipo Prácticas de Laboratorio		<b>Práctica 1 (RA37, RA38 y RA39)</b> TG: Técnica del tipo Trabajo en Grupo Evaluación Progresiva Presencial Duración: 00:00  <b>Práctica 2 (RA37, RA38 y RA39)</b> TG: Técnica del tipo Trabajo en Grupo Evaluación Progresiva Presencial Duración: 00:00  <b>Examen parte 1 (RA37, RA38 y RA39)</b> EX: Técnica del tipo Examen Escrito Evaluación Progresiva Presencial Duración: 01:00  <b>Práctica 1 (RA37, RA38 y RA39)</b> TG: Técnica del tipo Trabajo en Grupo Evaluación Global Presencial Duración: 00:00  <b>Práctica 2 (RA37, RA38 y RA39)</b> TG: Técnica del tipo Trabajo en Grupo Evaluación Global Presencial Duración: 00:00

				<b>Examen parte 1 (RA37, RA38 y RA39)</b> EX: Técnica del tipo Examen Escrito Evaluación Global Presencial Duración: 01:00
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				

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.

## 7. Actividades y criterios de evaluación

### 7.1. Actividades de evaluación de la asignatura

#### 7.1.1. Evaluación (progresiva)

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
5	Practica 1 (RA37, RA38 y RA39)	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	35%	4 / 10	S02 CT1 CT2 CE02 CE05
5	Practica 2 (RA37, RA38 y RA39)	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	35%	4 / 10	S02 CG1 CG4 CG3 CG5 CT3 CT4 CB10
5	Examen parte 1 (RA37, RA38 y RA39)	EX: Técnica del tipo Examen Escrito	Presencial	01:00	30%	4 / 10	CT1 CB6 CB7 CE05 K05

#### 7.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
5	Practica 1 (RA37, RA38 y RA39)	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	35%	4 / 10	CG1 CG4 CG3 CG5 CT3 CT4 CB10
5	Practica 2 (RA37, RA38 y RA39)	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	35%	4 / 10	S02 CT1 CT2 CE02 CE05

5	Examen parte 1 (RA37, RA38 y RA39)	EX: Técnica del tipo Examen Escrito	Presencial	01:00	30%	4 / 10	CT1 CB6 CB7 CE05 K05
---	------------------------------------	-------------------------------------	------------	-------	-----	--------	----------------------------------

### 7.1.3. Evaluación convocatoria extraordinaria

Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
Extraordinaria. Practica 1 (RA37, RA38 y RA39)	TI: Técnica del tipo Trabajo Individual	Presencial	00:00	35%	4 / 10	CT2 CE02 CE05 S02 CT1
Extraordinaria. Examen	EX: Técnica del tipo Examen Escrito	Presencial	01:30	30%	4 / 10	CB6 CB7 CE05 K05 CT1
Extraordinaria. Practica 2	TI: Técnica del tipo Trabajo Individual	Presencial	00:00	35%	4 / 10	CG1 CG4 CG3 CG5 CT3 CT4 CB10

## 7.2. Criterios de evaluación

Los estudiantes desarrollarán dos actividades prácticas P1 y P2 a lo largo del curso y un examen individual escrito de los conceptos del temario.

Estas 3 actividades se realizarán en evaluación continua, en la prueba de evaluación global y en la convocatoria extraordinaria.

## 8. Recursos didácticos

---

### 8.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
moodle	Recursos web	Toda la documentación necesaria para el desarrollo del curso se suministrará a través de la plataforma moodle de la UPM
Bibliografía	Bibliografía	Lee, D., Damji, J. S., Wenig, B., Das, T. (2020). Learning Spark. Estados Unidos: O'Reilly Media.

## 9. Otra información

---

### 9.1. Otra información sobre la asignatura

La asignatura se imparte de modo condensado en 5 semanas, a 6 horas por semana, por lo que semana 1 significa la primera semana desde que comience la impartición, no desde el inicio del curso.