



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

**105001024 - Arquitecturas Para El Procesamiento Masivo De Datos**

### PLAN DE ESTUDIOS

10CD - Grado En Ciencia De Datos E Inteligencia Artificial

### CURSO ACADÉMICO Y SEMESTRE

2022/23 - Primer 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.....	9
8. Recursos didácticos.....	11

## 1. Datos descriptivos

---

### 1.1. Datos de la asignatura

<b>Nombre de la asignatura</b>	105001024 - Arquitecturas para el Procesamiento Masivo de Datos
<b>No de créditos</b>	6 ECTS
<b>Carácter</b>	Obligatoria
<b>Curso</b>	Tercero curso
<b>Semestre</b>	Quinto semestre
<b>Período de impartición</b>	Septiembre-Enero
<b>Idioma de impartición</b>	Castellano
<b>Titulación</b>	10CD - Grado en Ciencia de Datos e Inteligencia Artificial
<b>Centro responsable de la titulación</b>	10 - Escuela Tecnica Superior De Ingenieros Informaticos
<b>Curso académico</b>	2022-23

## 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 Garcia Dopico (Coordinador/a)		antonio.garcia.dopico@upm.es	- -
Jose Luis Pedraza Dominguez		joseluis.pedraza@upm.es	Sin horario.
Maria Isabel Garcia Clemente		mariaisabel.garciac@upm.es	Sin horario.

Maria De Los Santos Perez Hernandez		maria.s.perez@upm.es	Sin horario.
Fernando Perez Costoya		fernando.perez@upm.es	Sin horario.
Luis Manuel Gomez Henriquez		luismanuel.gomezh@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. Conocimientos previos recomendados

---

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

- Fundamentos De Los Computadores
- Fundamentos De La Programación
- Algoritmos Y Estructuras De Datos

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

- Poseer destrezas fundamentales de la programación
- Conocimientos básicos de concurrencia

### 4. Competencias y resultados de aprendizaje

---

#### 4.1. Competencias

CB01 - Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio

CB02 - Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio

CB03 - Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética

CB05 - Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

CE07 - Capacidad de diseñar e implementar sistemas de información (incluyendo modelos de datos y estrategias de gestión de datos) dimensionados para gestionar el volumen, velocidad y variedad de los datos, de forma adecuada para su almacenamiento, procesamiento y acceso para tratamientos posteriores.

CE08 - Poseer las destrezas para aplicar las tecnologías actuales de computación de altas prestaciones para diseñar e implementar nuevas aplicaciones de ciencia de datos.

CG02 - Capacidad para organizar y planificar tareas y proyectos, identificando objetivos, prioridades, plazos, recursos y riesgos, y controlando los procesos establecidos.

CG04 - Capacidad para innovar y encontrar soluciones creativas en situaciones complejas o de incertidumbre en el ámbito de la ingeniería.

CG06 - Identificar y utilizar las tecnologías de la información y las comunicaciones más adecuadas en el ámbito de la ingeniería.

## **4.2. Resultados del aprendizaje**

RA106 - RA-APID-7 Conocer y saber utilizar las técnicas fundamentales de computación de altas prestaciones

RA107 - RA-APID-8 Conocer las infraestructuras y plataformas paralelas de procesamiento de datos

RA110 - RA-APID-17 Evaluar y diseñar el sistema informático más apropiado para resolver los distintos problemas de Ciencia de Datos, exponiendo las dificultades técnicas y los límites de la aplicación

RA108 - RA-APID-9 Dimensionar sistemas informáticos para gestionar el volumen, velocidad y variedad de los datos

RA109 - RA-APID-15 Diseñar e implementar soluciones distribuidas/paralelas sobre grandes cantidades de datos

## 5. Descripción de la asignatura y temario

---

### 5.1. Descripción de la asignatura

En esta asignatura se describirá la infraestructura Hardware y Software que se utiliza para poder procesar el enorme volumen de datos que suele manejarse en la Ciencia de Datos y en la Inteligencia Artificial. Para ello se abordará los siguientes temas:

- Computación de alto rendimiento
  
- Sistemas Distribuidos
  
- Sistemas de almacenamiento de altas prestaciones
  
- Procesamiento batch de datos

### 5.2. Temario de la asignatura

1. Introducción
2. Optimización de programas secuenciales
  - 2.1. Optimización de programas
  - 2.2. Optimización de Python
3. SIMD, vectorización y paralelismo de datos
4. Multiprocesadores
  - 4.1. Introducción a los multiprocesadores
  - 4.2. Modelos de paralelismo
  - 4.3. Sincronización
  - 4.4. OpenMP con Python

#### 4.5. Multithreading vs multiprocessing en Python

### 5. Aceleradores: GPUs y TPUs

#### 5.1. Arquitectura de las GPUs

#### 5.2. Cómo se programan (PyCUDA)

### 6. Sistemas Distribuidos

#### 6.1. Arquitecturas y modelos

#### 6.2. Tolerancia a fallos, fiabilidad, escalabilidad

#### 6.3. Mecanismos de comunicación

#### 6.4. Programación con MPI (mpi4py)

### 7. Sistemas de almacenamiento de altas prestaciones

#### 7.1. HDD/SSD/RAID/DAS/SAN/NAS

#### 7.2. SFP

### 8. Procesamiento batch de datos

#### 8.1. Map-Reduce

#### 8.2. RDDs y Spark

## 6. Cronograma

### 6.1. Cronograma de la asignatura \*

Sem	Actividad en aula	Actividad en laboratorio	Tele-enseñanza	Actividades de evaluación
1	<b>Intro</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral  <b>Optimización: Memoria</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral  <b>Optimización: Memoria</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
2	<b>Optimización: Memoria 2h</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas  <b>Optimización: memoria 1h ILP 1h</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
3	<b>Optimización: ILP</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas	<b>Practica Optimización</b> Duración: 01:00 PL: Actividad del tipo Prácticas de Laboratorio		
4	<b>Vectorización</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral	<b>Practica Optimización</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		<b>Práctica optimización</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Presencial Duración: 00:00
5	<b>Vectorización</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas  <b>Multiprocesadores</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
6	<b>Multiprocesadores</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
7	<b>Aceleradores: GPUs</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral	<b>Practica Multiprocesadores OpenMP</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		<b>Practica multiprocesadores</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Presencial Duración: 00:00



8	<p><b>Aceleradores: GPUs</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p> <p><b>Sistemas Distribuidos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>			
9	<p><b>Sistemas Distribuidos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>			<p><b>Examen temas 1-5</b> EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 01:00</p>
10	<p><b>Sistemas Distribuidos: MPI</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>			
11	<p><b>Almacenamiento altas prestaciones</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>	<p><b>Practica MPI</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		<p><b>Práctica MPI</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Presencial Duración: 00:00</p>
12	<p><b>Almacenamiento altas prestaciones</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p><b>Procesamiento batch de datos: MapReduce</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>			
13	<p><b>Procesamiento batch de datos: MapReduce</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>	<p><b>Procesamiento batch de datos: Map Reduce</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		<p><b>Práctica MapReduce</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Presencial Duración: 00:00</p>
14	<p><b>Procesamiento batch de datos: Spark</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>			
15	<p><b>Procesamiento batch de datos: Spark</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>	<p><b>Procesamiento batch de datos: Map Reduce</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
16		<p><b>Procesamiento batch de datos: Map Reduce</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p> <p><b>Procesamiento batch de datos: Map Reduce</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		<p><b>Práctica Spark</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua y sólo prueba final Presencial Duración: 00:00</p> <p><b>Examen temas6-8</b> EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 01:00</p>
17				<p><b>Examen Recuperación (Prueba global)</b> EX: Técnica del tipo Examen Escrito Evaluación 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.

## 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
4	Práctica optimización	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	5%	3 / 10	CG04 CG06 CE08 CB05 CG02 CE07
7	Practica multiprocesadores	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	5%	3 / 10	CG04 CE07 CG06 CE08 CB05 CG02
9	Examen temas 1-5	EX: Técnica del tipo Examen Escrito	Presencial	01:00	35%	3 / 10	CB02 CB01 CB03
11	Práctica MPI	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	5%	3 / 10	CE07 CG04 CG06 CE08 CB05 CG02
13	Práctica MapReduce	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	5%	3 / 10	CE07 CG04 CG06 CE08 CB05 CG02
16	Práctica Spark	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	10%	3 / 10	CE07 CG04 CG06 CE08 CB05 CG02

16	Examen temas6-8	EX: Técnica del tipo Examen Escrito	Presencial	01:00	35%	3 / 10	CB03 CB02 CB01
----	-----------------	-------------------------------------	------------	-------	-----	--------	----------------------

### 7.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
4	Práctica optimización	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	5%	3 / 10	CG04 CG06 CE08 CB05 CG02 CE07
7	Practica multiprocesadores	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	5%	3 / 10	CG04 CE07 CG06 CE08 CB05 CG02
11	Práctica MPI	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	5%	3 / 10	CE07 CG04 CG06 CE08 CB05 CG02
13	Práctica MapReduce	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	5%	3 / 10	CE07 CG04 CG06 CE08 CB05 CG02
16	Práctica Spark	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	10%	3 / 10	CE07 CG04 CG06 CE08 CB05 CG02
17	Examen Recuperación (Prueba global)	EX: Técnica del tipo Examen Escrito	Presencial	02:00	70%	3 / 10	CB03 CB02 CB01

### 7.1.3. Evaluación convocatoria extraordinaria

Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
El alumno se examinará el día de la convocatoria extraordinaria de las partes que no haya aprobado en la convocatoria ordinaria. El examen constará de dos bloques: temas 1-5 (bloque I) y temas 6-8 (bloque II)	EX: Técnica del tipo Examen Escrito	Presencial	02:00	70%	3 / 10	CB03 CB02 CB01
Para la recuperación de las prácticas suspendidas, el alumno reentregará cada práctica que haya suspendido en la convocatoria ordinaria, como muy tarde el día de la convocatoria extraordinaria.	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:01	30%	3 / 10	CE07 CG04 CG06 CE08 CB03 CB05 CG02

## 7.2. Criterios de evaluación

Habrán dos exámenes, cada uno con un peso del 35% de la nota, así como cuatro prácticas cortas, cada una con un 5% de la nota y un proyecto más grande con el 10% de la nota. Las prácticas pueden recuperarse en la convocatoria ordinaria y la extraordinaria (volviendo a entregarlas) y los exámenes se podrán recuperar también.

## 8. Recursos didácticos

### 8.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
Página web de la asignatura	Recursos web	En esta página se irá volcando la documentación a utilizar