

ANX-PR/CL/001-01

GUÍA DE APRENDIZAJE

ASIGNATURA

615001024 - Arquitecturas Para El Procesamiento Masivo De Datos

PLAN DE ESTUDIOS

61CD - 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.....	3
6. Cronograma.....	5
7. Actividades y criterios de evaluación.....	7
8. Recursos didácticos.....	10

1. Datos descriptivos

1.1. Datos de la asignatura

Nombre de la asignatura	615001024 - Arquitecturas para el Procesamiento Masivo de Datos
No de créditos	6 ECTS
Carácter	Obligatoria
Curso	Tercero curso
Semestre	Quinto semestre
Período de impartición	Septiembre-Enero
Idioma de impartición	Castellano
Titulación	61CD - Grado en Ciencia de Datos e Inteligencia Artificial
Centro responsable de la titulación	61 - Escuela Técnica Superior De Ingeniería De Sistemas Informáticos
Curso académico	2022-23

2. Profesorado

2.1. Profesorado implicado en la docencia

Nombre	Despacho	Correo electrónico	Horario de tutorías *
Bonifacio Alberto Mozo Velasco (Coordinador/a)		a.mozo@upm.es	Sin horario. Sin definir todavía
Felix Jose Fuentes Hurtado	4305	felix.fuentes@upm.es	Sin horario. Sin definir todavía

* 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

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

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

RA131 - Diseñar e implementar soluciones distribuidas/paralelas sobre grandes cantidades de datos

RA126 - Dimensionar sistemas informáticos para gestionar el volumen, velocidad y variedad de los datos

RA100 - 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

RA128 - Conocer las infraestructuras y plataformas paralelas de procesamiento de datos

RA98 - Conocer y saber utilizar las técnicas fundamentales de computación de altas prestaciones

5. Descripción de la asignatura y temario

5.1. Descripción de la asignatura

En esta asignatura se describirá la Arquitectura 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 abordarán los siguientes temas:

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

5.2. Temario de la asignatura

1. Introducción
2. Optimización de programas secuenciales
 - 2.1. Jerarquía de memoria
 - 2.2. Optimización de Python
3. SIMD, Vectorización y paralelismo de datos
4. Multiprocesadores
 - 4.1. Introducción a multiprocesadores
 - 4.2. Modelos de paralelismo
 - 4.3. Sincronización
 - 4.4. OpenMP con Python
 - 4.5. Multithreading vs multiprocessing en Python
5. Aceleradores: GPU y TPUs
 - 5.1. Arquitecturas
 - 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	Intro Duración: 02:00 LM: Actividad del tipo Lección Magistral Optimización Duración: 02:00 LM: Actividad del tipo Lección Magistral			
2	Optimización Duración: 02:00 LM: Actividad del tipo Lección Magistral	Optimización Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
3	Vectorización Duración: 02:00 LM: Actividad del tipo Lección Magistral	Optimización Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		Optimización EP: Técnica del tipo Examen de Prácticas Evaluación continua y sólo prueba final Presencial Duración: 00:00
4	Vectorización Duración: 02:00 LM: Actividad del tipo Lección Magistral	Vectorización Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
5	Multiprocesadores Duración: 02:00 LM: Actividad del tipo Lección Magistral	Vectorización Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
6	Multiprocesadores Duración: 02:00 LM: Actividad del tipo Lección Magistral Multiprocesadores Duración: 02:00 LM: Actividad del tipo Lección Magistral			
7	Multiprocesadores Duración: 02:00 LM: Actividad del tipo Lección Magistral	Multiprocesadores Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		Multiprocesadores EP: Técnica del tipo Examen de Prácticas Evaluación continua y sólo prueba final Presencial Duración: 00:00
8	Aceleradores Duración: 02:00 LM: Actividad del tipo Lección Magistral	Aceleradores Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
9	Sistemas Distribuidos Duración: 02:00 LM: Actividad del tipo Lección Magistral			Examen temas 1-5 EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 02:00

10	Sistemas Distribuidos Duración: 02:00 LM: Actividad del tipo Lección Magistral	Sistemas Distribuidos Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
11	Sistemas Distribuidos Duración: 02:00 LM: Actividad del tipo Lección Magistral	Sistemas Distribuidos Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		Sistemas Distribuidos EP: Técnica del tipo Examen de Prácticas Evaluación continua y sólo prueba final Presencial Duración: 00:00
12	Almacenamiento altas prestaciones Duración: 02:00 LM: Actividad del tipo Lección Magistral Almacenamiento altas prestaciones Duración: 02:00 LM: Actividad del tipo Lección Magistral			
13	Procesamiento batch de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral	Procesamiento batch de datos Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		Procesamiento batch de datos EP: Técnica del tipo Examen de Prácticas Evaluación continua y sólo prueba final Presencial Duración: 00:00
14		Procesamiento batch de datos Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio Procesamiento batch de datos Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
15		Procesamiento batch de datos Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio Procesamiento batch de datos Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		Procesamiento batch de datos EP: Técnica del tipo Examen de Prácticas Evaluación continua y sólo prueba final Presencial Duración: 00:00
16				
17				Examen temas 6-8 EX: Técnica del tipo Examen Escrito Evaluación continua y sólo prueba final Presencial Duración: 01:00 Examen Recuperación Temes1-5 (Prueba final) EX: Técnica del tipo Examen Escrito Evaluación sólo prueba final Presencial Duración: 01:00

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.

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
3	Optimizacion	EP: Técnica del tipo Examen de Prácticas	Presencial	00:00	5%	3 / 10	CE07 CE08 CG02 CG04 CG06 CB05
7	Multiprocesadores	EP: Técnica del tipo Examen de Prácticas	Presencial	00:00	5%	3 / 10	CE07 CE08 CB05 CG02 CG04 CG06
9	Examen temas 1-5	EX: Técnica del tipo Examen Escrito	Presencial	02:00	35%	3 / 10	CB03 CB01 CB02
11	Sistemas Distribuidos	EP: Técnica del tipo Examen de Prácticas	Presencial	00:00	5%	3 / 10	CB05 CE07 CE08 CG02 CG04 CG06
13	Procesamiento batch de datos	EP: Técnica del tipo Examen de Prácticas	Presencial	00:00	5%	3 / 10	CB05 CE07 CE08 CG02 CG04 CG06
15	Procesamiento batch de datos	EP: Técnica del tipo Examen de Prácticas	Presencial	00:00	10%	3 / 10	CB05 CE07 CE08 CG02 CG04 CG06

17	Examen temas 6-8	EX: Técnica del tipo Examen Escrito	Presencial	01:00	35%	3 / 10	CB01 CB02 CB03
----	------------------	-------------------------------------	------------	-------	-----	--------	----------------------

7.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
3	Optimizacion	EP: Técnica del tipo Examen de Prácticas	Presencial	00:00	5%	3 / 10	CE07 CE08 CG02 CG04 CG06 CB05
7	Multiprocesadores	EP: Técnica del tipo Examen de Prácticas	Presencial	00:00	5%	3 / 10	CE07 CE08 CB05 CG02 CG04 CG06
11	Sistemas Distribuidos	EP: Técnica del tipo Examen de Prácticas	Presencial	00:00	5%	3 / 10	CB05 CE07 CE08 CG02 CG04 CG06
13	Procesamiento batch de datos	EP: Técnica del tipo Examen de Prácticas	Presencial	00:00	5%	3 / 10	CB05 CE07 CE08 CG02 CG04 CG06
15	Procesamiento batch de datos	EP: Técnica del tipo Examen de Prácticas	Presencial	00:00	10%	3 / 10	CB05 CE07 CE08 CG02 CG04 CG06
17	Examen temas 6-8	EX: Técnica del tipo Examen Escrito	Presencial	01:00	35%	3 / 10	CB01 CB02 CB03
17	Examen Recuperación Tems1-5 (Prueba final)	EX: Técnica del tipo Examen Escrito	Presencial	01:00	35%	3 / 10	CB02 CB03 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 los temas y prácticas 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	CB01 CB02 CB03
Para la evaluación de las prácticas suspensas, el alumno entregará para cada una de ellas, y como muy tarde el día de la convocatoria extraordinaria, el mismo material pedido para cada práctica en la convocatoria ordinaria.	EP: Técnica del tipo Examen de Prácticas	Presencial	01:00	30%	3 / 10	CB05 CE07 CE08 CG02 CG04 CG06

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 en ambas convocatorias

8. Recursos didácticos

8.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
Página web de la asignatura	Recursos web	En esta página (moodle.upm.es) se irá volcando la documentación a utilizar