UNIVERSIDAD DE MURCIA

1. Identificación

1.1. De la asignatura

Curso Académico	2024/2025
Titulación	GRADO EN CIENCIA E INGENIERÍA DE DATOS
Nombre de la asignatura	PROCESAMIENTO PARALELO DE DATOS
Código	6599
Curso	TERCERO
Carácter	OBLIGATORIA
Número de grupos	1
Créditos ECTS	6.0
Estimación del volumen de trabajo	150.0
Organización temporal	2º Cuatrimestre
Idiomas en que se imparte	Español

1.2. Del profesorado: Equipo docente

CUENCA MUÑOZ, ANTONIO JAVIER

Coordinación de los grupos: **GRUPO 1**

Coordinador de la asignatura

Categoría
PROFESORES TITULARES DE UNIVERSIDAD

Área
ARQUITECTURA Y TECNOLOGÍA DE COMPUTADORES

Departamento
INGENIERÍA Y TECNOLOGÍA DE COMPUTADORES

Correo electrónico / Página web / Tutoría electrónica
jcuenca@um.es http://www.ditec.um.es/~javiercm/ Tutoría electrónica: **Sí**

Teléfono, horario y lugar de atención al alumnado

Duración:	Día:	Horario:	Lugar:
A	Miércoles	11:00-14:00	868884821, Facultad de Informática B1.3.031

Observaciones:
Despacho 3.33. Avisar previamente por mensaje privado del Aula Virtual

CEBRIAN GONZALEZ, JUAN MANUEL

Docente: **GRUPO 1**

Coordinación de los grupos:

Categoría
PROFESOR AYUDANTE DOCTOR

Área
ARQUITECTURA Y TECNOLOGÍA DE COMPUTADORES

Departamento
INGENIERÍA Y TECNOLOGÍA DE COMPUTADORES

Correo electrónico / Página web / Tutoría electrónica
jcebrian@um.es https://orcid.org/0000-0002-3731-9301 Tutoría electrónica: **Sí**

Teléfono, horario y lugar de atención al alumnado

Duración:	Día:	Horario:	Lugar:
A	Miércoles	10:00-13:00	868884609, Facultad de Informática B1.3.023

Observaciones:
Tutorías online a través del aula virtual o correo electrónico (jcebrian@um.es)

2. Presentación

En esta asignatura el alumno estudiará el papel que el uso del paralelismo tiene a la hora de acelerar el procesamiento de datos.

El conocimiento teórico partirá de los conceptos técnicos más generales de paralelismo y de sistemas distribuidos. Tras ello, en el núcleo de la asignatura, se estudiarán las herramientas y tecnologías actuales para el procesamiento de datos masivos.

En el aspecto práctico, el alumno realizará diversas sesiones con un enfoque incremental a fin de adquirir el conocimiento y la habilidad de programar y/o desplegar soluciones de procesamiento de datos en paralelo.

Los contenidos generales que se ven son:

- Introducción a la computación paralela y a los sistemas distribuidos.
- Computación paralela intensiva en datos (paralelismo de datos, modelos y abstracciones).
- Software y entornos de ejecución específicos para análisis avanzado de datos.

3. Condiciones de acceso a la asignatura

3. Condiciones de acceso a la asignatura

3.1. Incompatibilidades

No constan

3.2. Requisitos

No constan

3.3. Recomendaciones

Se recomienda haber superado las asignaturas «Fundamentos de Programación», «Estructuras de Datos», «Análisis y Diseño de Algoritmos», «Fundamentos de Computadores» y «Programación para Ciencia de Datos».

4. Competencias

4.1. Competencias básicas

- CB2: 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

4.2. Competencias de la titulación

- CG2: Conocer y aplicar los fundamentos de ingeniería y tecnologías informáticas actuales para diseñar e implementar nuevas aplicaciones de análisis de datos.
- CG3: Capacidad para desarrollar experimentos y para implementar sistemas, infraestructuras, procesos y herramientas con el fin de soportar la manipulación de los datos durante todo el ciclo de vida de estos.
- CE7: Conocer la arquitectura y el funcionamiento de los computadores, la interconexión de los componentes que los forman y su software de sistema básico.
- CE9: Utilizar sistemas distribuidos, sistemas de cálculo paralelo, sistemas de procesamiento por lotes y plataformas de procesamiento de flujos de datos, incluyendo soluciones online y basadas en la nube para la provisión de servicios bajo demanda y escalables.

4.3. Competencias transversales y de materia

No constan

5. Contenidos

5.1. Teoría

Tema 1: Introducción

- Introducción a los SistemasDistribuidos
- Arquitecturas para Big Data
- Introducción a los Sistemas Paralelos

Tema 2: Tecnologías de Big Data

- Procesamiento MapReduce
- Hadoop
- Apache Spark
- Framework Dask
- Infraestructura en la nube para Big Data
- Otros frameworks y herramientas

Tema 3: Paralelismo en Big Data

- Implementación general de MapReduce
- Implementación de MapReduce paralelizada con hilos C++

5.2. Prácticas

- **Práctica 1: Introducción (Relacionada con el Tema 1).**
  - Programas paralelos: pruebas y medición de prestaciones (speedup, escalabilidad, ...)
  - Servicio distribuido mediante microservicios
  - Despliegue de un servicio REST dockerizado.
  - Ejecución de llamadas y cálculo de tiempos de respuesta.

- **Práctica 2: Uso básico de Big Data(Relacionada con el Tema 2).**
  - Instalación y configuración de Apache Spark
  - Procesamiento de datos con Apache Spark
  - Instalación y configuración de Dark
  - Procesamiento de datos con Dask
  - Big Data en Amazon Web Service (AWS)

- **Práctica 3: Implementación paralela (Relacionada con el Tema 3).**

  Implementar un ejemplo de MapReduce en C++ y paralelizarlo tanto con Parallel Patterns Library (PPL) como con threads.

6. Actividades Formativas

Actividad Formativa	Metodología	Horas	Presencialidad
AF1: Exposición teórica: Exposición de los contenidos teóricos de la asignatura por parte del profesor. También se contemplan las sesiones informativas sobre el desarrollo del trabajo de fin de grado o prácticas externas.		20.0	100.0
AF2: Seminarios y actividades de aula: Exposición, análisis y debate dentro del contexto de aplicaciones específicas de contenidos teóricos, así como planteamiento y resolución de ejercicios y casos prácticos en el aula, tanto al grupo completo como en grupos reducidos. También se contemplan conferencias, debates y seminarios temáticos.		5.0	100.0
AF3: Prácticas de laboratorio: Ejercicios y resolución de problemas, aprendizaje orientado a proyectos, estudio de casos, exposición y discusión de trabajos, simulaciones y/o prácticas con ordenadores, generalmente desarrolladas en grupos reducidos.		28.0	100.0
AF4: Trabajo autónomo del alumno: Estudio y preparación de contenidos teórico-prácticos, trabajo individual consistente en lecturas, búsquedas de información, sistematización de contenidos, elaboración de informes o estudio para la elaboración de casos entre otras actividades.		90.0	0.0
AF5: Tutorías formativas y resolución de dudas: Asistencia individualizada -tutorías individuales- o en grupo -tutorías colectivas- a los estudiantes por parte del profesor.		3.0	100.0
AF6: Evaluación: Pruebas individuales, ya sean escritas, orales o con medios informáticos, donde el estudiante demostrará los conocimientos teóricos y prácticos adquiridos durante las actividades formativas asociadas a la enseñanza de la materia.		4.0	100.0
	Totales	150,00	

7. Horario de la asignatura

https://www.um.es/web/estudios/grados/ciencia-ingenieria-datos/2024-25#horarios

8. Sistemas de Evaluación

Identificador	Denominación del instrumento de evaluación	Criterios de Valoración	Ponderación
SE1	Exámenes individuales: Pruebas objetivas, de desarrollo, de respuesta corta, de ejecución de tareas o de escala de actitudes, realizadas por los estudiantes para mostrar los conocimientos teóricos y prácticos adquiridos ya sea de forma oral, escrita o utilizando un ordenador.		50.0
SE3	Procedimientos de observación del trabajo del estudiante: Registros de participación, de realización de actividades, cumplimiento de plazos, participación en foros, informes de seguimiento del trabajo fin de grado y registros sobre el desarrollo de las prácticas externas.		0.0
SE4	Informe técnico. En este instrumento incluimos los resultados de actividades prácticas, o de laboratorio, junto con sus memorias descriptivas y posibles resúmenes del estado del arte sobre temas concretos. La opción de realizar entrevistas personales o presentaciones de los trabajos realizados también entran en esta categoría.		50.0

9. Fechas de exámenes

https://www.um.es/web/estudios/grados/ciencia-ingenieria-datos/2024-25#examenes

10. Resultados del Aprendizaje

- Conocer los modelos de programación usados en el desarrollo de aplicaciones paralelas.
- Aprender técnicas de análisis y diseño de algoritmos paralelos intensivos en datos.
- Ser capaz de identificar y resolver los problemas fundamentales de las aplicaciones paralelas.
- Aprender a usar herramientas e infraestructuras de soporte para la computación paralela intensiva, incluyendo soluciones online y basadas en la nube.

11. Bibliografía

Bibliografía básica

- George Coulouris, Jean Dollmore, Tim Kindberg, “Sistemas distribuidos. Conceptos y diseño”
- Martin Kleppmann, “Designing Data Intensive Application”
- Nathan Marz y otros, "Big Data: Principles and Best Practices of Scalable Realtime Data Systems"
- Holden Karau, Mika Kimmins - “Scaling Python with Dask. From Data Science to Machine Learning”. O'Reilly Media.

Bibliografía complementaria

- Francisco Almeida y otros, “Introducción a la programación paralela”
- John L. Hennessy, David A. Patterson, “Computer architecture: a quantitative approach”
- Tom White, "Hadoop: The Definitive Guide" (O’Reilly)
- Holden Karau y otros, "Learning Spark: Lightning-Fast Big Data Analysis" (O’Reilly)
- Fabian Hueske y VasilikiKalavri, "Stream Processing with Apache Flink: Fundamentals, Implementation, and Operation of Streaming Applications" (O’Reilly)
- Maurice Herlihy, NirShavit, Victor Luchangco, Michael Spear, “The Art of Multiprocessor Programming” 2nd Edition.
- Fabio Nelli, “Parallel and High Performance Programming with Python. Unlock parallel and concurrent programming in Python using multithreading, CUDA, Pytorch and Dask”. Orange Education Pvt Ltd, AVA.

12. Observaciones

Esta asignatura no se encuentra vinculada de forma directa con los Objetivos de Desarrollo Sostenible.

NECESIDADES EDUCATIVAS ESPECIALES

Aquellos estudiantes con discapacidad o necesidades educativas especiales podrán dirigirse al Servicio de Atención a la Diversidad y Voluntariado (ADYV - https://www.um.es/adyv) para recibir orientación sobre un mejor aprovechamiento de su proceso formativo y, en su caso, la adopción de medidas de equiparación y de mejora para la inclusión, en virtud de la Resolución Rectoral R-358/2016. El tratamiento de la información sobre este alumnado, en cumplimiento con la LOPD, es de estricta confidencialidad.

REGLAMENTO DE EVALUACIÓN DE ESTUDIANTES

El artículo 8.6 del Reglamento de Evaluación de Estudiantes (REVA) prevé que "salvo en el caso de actividades definidas como obligatorias en la guía docente, si el o la estudiante no puede seguir el proceso de evaluación continua por circunstancias sobrevenidas debidamente justificadas, tendrá derecho a realizar una prueba global".

Se recuerda asimismo que el artículo 22.1 del Reglamento de Evaluación de Estudiantes (REVA) estipula que "el o la estudiante que se valga de conductas fraudulentas, incluida la indebida atribución de identidad o autoría, o esté en posesión de medios o instrumentos que faciliten dichas conductas, obtendrá la calificación de cero en el procedimiento de evaluación y, en su caso, podrá ser objeto de sanción, previa apertura de expediente disciplinario".