

Guía docente

270218 - PSD - Paralelismo y Sistemas Distribuidos

Última modificación: 31/01/2025

Unidad responsable: Facultad de Informática de Barcelona
Unidad que imparte: 701 - DAC - Departamento de Arquitectura de Computadores.
Titulación: GRADO EN CIENCIA E INGENIERÍA DE DATOS (Plan 2017). (Asignatura obligatoria).
Curso: 2024 **Créditos ECTS:** 6.0 **Idiomas:** Catalán

PROFESORADO

Profesorado responsable:

Otros:

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:

CE4. Utilizar los sistemas de computación actuales, incluidos sistemas de alto rendimiento, para el proceso de grandes volúmenes de datos desde el conocimiento de su estructura, funcionamiento y particularidades.

Genéricas:

CG1. Concebir sistemas computacionales que integren datos de procedencias y formas muy diversas, creen con ellos modelos matemáticos, razonen sobre dichos modelos y actúen en consecuencia, aprendiendo de la experiencia.

CG2. Elegir y aplicar los métodos y técnicas más adecuados a un problema definido por datos que representen un reto por su volumen, velocidad, variedad o heterogeneidad, incluidos métodos informáticos, matemáticos, estadísticos y de procesado de la señal.

CG4. Identificar oportunidades para aplicaciones innovadoras orientadas a datos en entornos tecnológicos en continua evolución.

Transversales:

CT4. Trabajo en equipo. Ser capaz de trabajar como miembro de un equipo interdisciplinar, ya sea como un miembro más o realizando tareas de dirección, con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

CT5. Uso solvente de los recursos de información. Gestionar la adquisición, la estructuración, el análisis y la visualización de datos e información en el ámbito de especialidad y valorar de forma crítica los resultados de dicha gestión.

CT6. Aprendizaje autónomo. Detectar deficiencias en el propio conocimiento y superarlas mediante la reflexión crítica y la elección de la mejor actuación para ampliar dicho conocimiento.

CT7. Tercera lengua. Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

Básicas:

CB1. 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.

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.

CB5. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

METODOLOGÍAS DOCENTES

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTE

Tipo	Horas	Porcentaje
Horas aprendizaje autónomo	90,0	60.00
Horas grupo pequeño	30,0	20.00
Horas grupo grande	30,0	20.00

Dedicación total: 150 h

CONTENIDOS

Fundamentos de la supercomputación paralela y distribuida

Descripción:

En este tema, los alumnos aprenderán conceptos básicos de la computación paralela así como métricas que les ayudarán a evaluar tanto el rendimiento de sus programas como los límites derivados de la propia estructura de la aplicación.

Arquitecturas paralelas y distribuidas

Descripción:

En este tema, los alumnos aprenderán las principales características de las arquitecturas paralelas y distribuidas que más les pueden influir a la hora de diseñar sus programas de análisis de datos o de entender el rendimiento (o pérdida de rendimiento) de los mismos.

Entornos de ejecución de computación paralela y análisis de datos

Descripción:

En este tema los alumnos aprenderán los diferentes entornos que principalmente pueden encontrarse al ejecutar tanto las aplicaciones que generan datos como las que los almacenan o analizan. Se hará énfasis en las diferencias entre los tres entornos y en impacto que puede tener en la eficiencia de sus aplicaciones.

Modelos de programación para supercomputadores

Descripción:

En este tema los alumnos verán los principios básicos de los modelos de programación más utilizados en los entornos HPC: MPI, OpenMP y modelos híbridos MPI+OpenMP. Se darán las herramientas para detectar y gestionar los principales detalles que puedan afectar tanto a la robustez de sus programas como a la eficiencia.

También se introducirán modelos orientados a co-procesadores que ofrecen una buena eficiencia rendimiento vs. consumo energético y muy utilizados en el análisis de datos.

Software y entorno de ejecución específico para análisis avanzados

Descripción:

En este tema los alumnos verán en más detalle las características de los modelos de programación y entornos de ejecución para almacenamiento y análisis de datos. Se utilizará como referencia el modelo Apache Spark/Hadoop, como referencia de almacenamiento de datos Cassandra y como herramientas de análisis TensorFlow/keras.

Machine Learning en Supercomputadores: Caso de ejemplo basado en Spark/Cassandra/TensorFlow

Descripción:

En este tema profundizaremos en un entorno de Machine learning usando como referencia el modelo Apache Spark, como DB key/value Cassandra y como herramienta de análisis TensorFlow. Se explicarán los elementos más importantes de estos tres componentes que puedan afectar en mayor medida tanto al diseño de aplicaciones de machine learning como al almacenamiento de datos y análisis.

Sesiones de laboratorio

Descripción:

Las sesiones de laboratorio se agruparán en dos proyectos que se irán realizando tanto en las sesiones de laboratorio como en trabajo autónomo. Los dos proyectos estarán relacionados con la programación, análisis y optimización de un caso lo más realista posible en dos entornos: entornos de ejecución paralelo (mpi+openmp, sistemas de colas, etc), utilizados para generar y post-procesar datos, y entornos específicos de gestión y análisis de datos como es Apache Spark + Cassandra + TensorFlow.

SISTEMA DE CALIFICACIÓN

BIBLIOGRAFÍA

Básica:

- TORRES, Jordi. Hand-on sessions at GitHub.
- Torres, J. Slides of the course. UPC,
- Torres, J. Understanding supercomputing: with Marenostrium Supercomputer in Barcelona. Universitat Politècnica de Catalunya, Barcelona Supercomputing Center, 2016. ISBN 9781365376825.
- Torres, J. Hello world en TensorFlow. Universitat Politècnica de Catalunya, Barcelona Supercomputing Center, 2016. ISBN 9781326532383.
- Macias, M.; Gómez, M.; Tous, R.; Torres, J. Introducción a Apache Spark: para empezar a programar el big data. UOC, 2015. ISBN 9788491160373.
- Articles from Technical Journals in the area.

Complementaria:

- Torres, J. Empresas en la nube: ventajas y retos del cloud computing. Libros de Cabecera, 2011. ISBN 9788493908225.