

# Sistemas de computación para datos masivos

**MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN  
DATA SCIENCE**

***UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO***

Este documento puede utilizarse como documentación de referencia de esta asignatura para la solicitud de reconocimiento de créditos en otros estudios. Para su plena validez debe estar sellado por la Secretaría de Estudiantes UIMP.



# DATOS GENERALES

## Breve descripción

Esta asignatura tiene como objetivo proporcionar al estudiante un contacto directo con las tecnologías, principalmente de ingeniería informática, que permiten el despliegue de las herramientas de análisis de datos y el desarrollo e implementación de nuevas soluciones. El alumno obtendrá una visión general del funcionamiento y la aplicación de grandes sistemas computadores para el tratamiento y manejo de datos.

## Título asignatura

Sistemas de computación para datos masivos

## Código asignatura

102271

## Curso académico

2022-23

## Planes donde se imparte

[MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN DATA SCIENCE](#)

## Créditos ECTS

4

## Carácter de la asignatura

OPTATIVA

## Duración

Cuatrimestral

## Idioma

Castellano e Inglés

# CONTENIDOS

## Contenidos

1. Arquitectura de una e-Infraestructura.
2. Computación HPC y HTC: servidores, clusters, supercomputadores.
3. Redes de interconexión de sistemas.
4. Sistemas de almacenamiento
5. Computación distribuida. Computación en paralelo. Introducción a MPI.
6. Gestión clásica de un cluster de computación. Sistemas de colas. Benchmarking. Monitorización.
7. Transmisión de datos en Internet y Wans.
8. Principios de gestión como servicio.
9. Evaluación.

## COMPETENCIAS

### Generales

CG1 - Integrarse eficazmente en un grupo de trabajo y trabajar en equipo, compartir la información disponible e integrar su actividad en la actividad del grupo colaborando de forma activa en la consecución de objetivos comunes

CG2 - Capacidad de estudio, síntesis y autonomía suficientes para desarrollar de forma autónoma proyectos básicos de investigación

CG6 - Buscar, obtener, procesar, comunicar información y transformarla en conocimiento

CG7 - Conocer las herramientas metodológicas necesarias para desarrollar proyectos avanzados

CG8 - Capacidad de actualización de los conocimientos expuestos en el ámbito de la comunidad científica

### Transversales

CT1 - Analizar y combinar información utilizando diferentes fuentes

CT2 - Conocer la problemática ética y legal relacionada con el análisis de datos y entender su importancia para una sociedad basada en los valores de la libertad, la justicia, la igualdad y el pluralismo

CT3 - Dominio de la gestión del tiempo

CT4 - Afrontar tareas y situaciones críticas

CT6 - Capacidades asociadas al trabajo en equipo: cooperación, liderazgo, saber escuchar

### Específicas

DSDM02 - Desarrollar e implementar modelos de datos, incluidos los metadatos

DSENG01 - Aplicar los principios de ingeniería a la investigación, diseño y desarrollo de un prototipo de aplicaciones de análisis de datos, o al desarrollo de estructuras, instrumentos, máquinas, experimentos, procesos, sistemas requeridos para ello

DSENG02 - Desarrollar y aplicar soluciones computacionales para problemas en un cierto dominio de aplicación, usando una amplia gama de plataformas de análisis de datos

# PLAN DE APRENDIZAJE

## Actividades formativas

AF1 - Participación y asistencia a lecciones magistrales y seminarios (30 horas)

AF2 - Realización de prácticas de computación y análisis de datos (30 horas)

AF6 - Tutorías (presenciales o por medio de recursos telemáticos) (5 horas)

AF7 - Elaboración de informes de laboratorio y trabajos (20 horas)

AF8 - Estudio individual de contenidos de la asignatura (20 horas)

AF9 - Trabajo en grupo (10 horas)

A10 - Pruebas de evaluación (5 horas)

## Metodologías docentes

La asignatura comenzará por una exposición de los conceptos básicos, incluyendo ejemplos sencillos pero relevantes, que serán analizados individualmente y discutidos en común.

Se revisarán los diferentes componentes de una solución, y los actores que participan en el desarrollo de la misma.

Los estudiantes, organizados en grupos, realizarán un análisis detallado de un caso de estudio empleando las diferentes tecnologías discutidas, por ejemplo supercomputación, recursos cloud, o plataformas de desarrollo de proyectos.

## Resultados de aprendizaje

- Identificar los recursos de computación adecuados, tanto locales como distribuidos, para abordar un problema de Data Science.
- Definir los requerimientos técnicos de nuevas aplicaciones en analítica de datos a partir de un diseño de alto nivel.
- Implementar algoritmos paralelos sencillos pero útiles en Data Science para su uso en un cluster o en un supercomputador.

# **SISTEMA DE EVALUACIÓN**

## **Descripción del sistema de evaluación**

SE1 - Examen (escrito, oral y/o práctico en el aula de computación) (40%)

SE2 - Valoración de informes y trabajos escritos (60%)

## PROFESORADO

### Profesor responsable

**Beivide Palacio, Julio Ramón**

*Catedrático de Arquitectura y Tecnología de Computadores  
Universidad de Cantabria*

### Profesorado

**Lloret Iglesias, Lara**

*Científica Titular del CSIC  
Instituto de Física de Cantabria (IFCA), Consejo Superior de Investigaciones Científicas*

**Palacio Hoz, Aída**

*IT Support Administrator  
Instituto de Física de Cantabria (IFCA), CSIC-UC*

**Bosque Orero, Jose Luis**

*Profesor Titular de Universidad del área de Arquitectura y Tecnología de Computadores  
Universidad de Cantabria*

**Cabrillo Bartolomé, José Ibán**

*Responsable de Sistemas de Computación Avanzada  
Instituto de Física de Cantabria (IFCA), CSIC-UC*

# HORARIO

## Horario

01/02/2023

15:30 - 17:30

Arquitecturas de datacenters y supercomputers

Julio Ramón Beivide Palacio

Catedrático de Arquitectura y Tecnología de Computadores  
Universidad de Cantabria

06/02/2023

15:30 - 17:30

Computación paralela HPC y HTC

Julio Ramón Beivide Palacio

Catedrático de Arquitectura y Tecnología de Computadores  
Universidad de Cantabria

08/02/2023

15:30 - 17:30

Servidores, clusters y supercomputadores

Julio Ramón Beivide Palacio

Catedrático de Arquitectura y Tecnología de Computadores  
Universidad de Cantabria



13/02/2023

17:30 - 19:30

Servidores, clusters y supercomputadores

Julio Ramón Bevide Palacio

Catedrático de Arquitectura y Tecnología de Computadores  
Universidad de Cantabria

15/02/2023

15:30 - 17:30

Servidores, clusters y supercomputadores

Julio Ramón Bevide Palacio

Catedrático de Arquitectura y Tecnología de Computadores  
Universidad de Cantabria

20/02/2023

15:30 - 17:30

Redes de Interconexión

Julio Ramón Bevide Palacio

Catedrático de Arquitectura y Tecnología de Computadores  
Universidad de Cantabria

23/02/2023

15:30 - 17:30

Redes de Interconexión

Julio Ramón Beivide Palacio

Catedrático de Arquitectura y Tecnología de Computadores  
Universidad de Cantabria

27/02/2023

15:30 - 17:30

Introducción a Slurm. Programación MPI: generación de procesos y comunicación punto a punto

Jose Luis Bosque Orero

Profesor Titular de Universidad del área de Arquitectura y Tecnología de Computadores  
Universidad de Cantabria

01/03/2023

15:30 - 17:30

Programación MPI: Comunicación colectiva

Jose Luis Bosque Orero

Profesor Titular de Universidad del área de Arquitectura y Tecnología de Computadores  
Universidad de Cantabria

09/03/2023

15:30 - 17:30

Gestión Clúster

José Ibán Cabrillo Bartolomé

Responsable de Sistemas de Computación Avanzada  
Instituto de Física de Cantabria (IFCA), CSIC-UC

14/03/2023

15:30 - 17:30

Gestión Clúster

Aída Palacio Hoz

IT Support Administrator  
Instituto de Física de Cantabria (IFCA), CSIC-UC

16/03/2023

15:30 - 17:30

Gestión Clúster

Aída Palacio Hoz

IT Support Administrator  
Instituto de Física de Cantabria (IFCA), CSIC-UC

21/03/2023

15:30 - 17:30

Gestión Clúster

Aída Palacio Hoz

IT Support Administrator  
Instituto de Física de Cantabria (IFCA), CSIC-UC

23/03/2023

15:30 - 17:30

Gestión Clúster

Aída Palacio Hoz

IT Support Administrator  
Instituto de Física de Cantabria (IFCA), CSIC-UC

09/05/2023

15:30 - 17:30

Examen

Julio Ramón Beivide Palacio

Catedrático de Arquitectura y Tecnología de Computadores  
Universidad de Cantabria

# BIBLIOGRAFÍA Y ENLACES RELACIONADOS

## Bibliografía

### Bibliografía básica

*Computer Architecture: A Quantitative Approach* (Fifth Edition). J.L. Hennessy & D.A. Patterson. The Morgan Kaufmann Series in Computer Architecture and Design (2011)

*The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines* (Second Edition). L.A. Barroso, J. Clidaras & U. Hözlze. Morgan & Claypool Pub. Synthesis Lectures on Computer (2013)