

Herramientas en la nube para la Ciencia de Datos

**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 en entorno de computación *Cloud*, que permiten el despliegue de las herramientas de análisis de datos y el desarrollo e implementación de nuevas soluciones.

Título asignatura

Herramientas en la nube para la Ciencia de Datos

Código asignatura

102272

Curso académico

2021-22

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

Principios básicos de computación Cloud:

- Fundamentos del modelo de computación Cloud.
- Modelos de servicio: Infraestructura como Servicio (IaaS), Plataforma como servicio (PaaS), Software como servicio (SaaS).
- Clouds híbridos. Acceso a recursos comerciales.
- Composición de servicios. Herramientas básicas de orquestación.
- Contenedores. Uso de Contenedores y Docker. Orquestación de contenedores.

Sistemas de almacenamiento escalable de nueva generación:

- Almacenamiento en entornos Cloud. Almacenamiento bloque. Almacenamiento de objeto.
- Sistemas de almacenamiento para procesamiento paralelo.
- Sistemas de analítica de datos avanzados para grandes volúmenes de datos.
- Herramientas de almacenamiento.

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

AF2 - Realización de prácticas de computación y análisis de datos

AF3 - Desarrollo de proyectos guiados

AF4 - Participación en casos prácticos en empresas o centros de investigación

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

AF7 - Elaboración de informes de laboratorio y trabajos

AF8 - Estudio individual de contenidos de la asignatura

AF9 - Trabajo en grupo

A10 - Pruebas de evaluación

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 el despliegue de una plataforma de orquestación de contenedores sobre un proveedor de infraestructura como servicio.

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.
- Saber acceder a sistemas en entorno Cloud en formato IaaS, PaaS o SaaS, para

implementar una solución en Big Data.

- Utilizar una solución basada en contenedores (mediante la herramienta Docker).
- Cómo diseñar y gestionar un proyecto de computación Cloud.

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

López García, Álvaro

Científico Titular

Instituto de Física de Cantabria (IFCA), CSIC-UC

Profesorado

San Miguel Gutiérrez, Marcos

Grado en Física

Investigador Predoctoral

Instituto de Física de Partículas (IFCA-CSIC).

Palacio Hoz, Aída

IT Support Administrator

Instituto de Física de Cantabria (IFCA), CSIC-UC

Cabrillo Bartolomé, José Ibán

Responsable de Sistemas de Computación Avanzada

Instituto de Física de Cantabria (IFCA), CSIC-UC

HORARIO

Horario

15/02/2022

15:30 - 17:30

Introducción / Cloud Computing

Álvaro López García

Científico Titular
Instituto de Física de Cantabria (IFCA), CSIC-UC

17/02/2022

15:30 - 17:30

Cloud Computing

Álvaro López García

Científico Titular
Instituto de Física de Cantabria (IFCA), CSIC-UC

24/02/2022

15:30 - 17:30

Cloud Computing

Aída Palacio Hoz

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

01/03/2022

15:30 - 17:30

Sistemas de almacenamiento para procesamiento paralelo

José Ibán Cabrillo Bartolomé

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

03/03/2022

15:30 - 17:30

Sistemas de almacenamiento para procesamiento paralelo

José Ibán Cabrillo Bartolomé

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

08/03/2022

15:30 - 17:30

Sistemas de almacenamiento para procesamiento paralelo

José Ibán Cabrillo Bartolomé

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

10/03/2022

15:30 - 17:30

Sistemas de analítica de datos avanzados para grandes volúmenes de datos

Aída Palacio Hoz

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

15/03/2022

15:30 - 16:30

Sistemas de analítica de datos avanzados para grandes volúmenes de datos

Aída Palacio Hoz

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

16:30 - 17:30

Sistemas de analítica de datos avanzados para grandes volúmenes de datos

Aída Palacio Hoz

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

17/03/2022

15:30 - 17:30

Sistemas de analítica de datos avanzados para grandes volúmenes de datos

Aída Palacio Hoz

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

22/03/2022

15:30 - 16:30

Docker

Aída Palacio Hoz

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

16:30 - 17:30

Docker

Aída Palacio Hoz

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

24/03/2022

15:30 - 16:30

Workflows: Prefect, Airflow, etc

Marcos San Miguel Gutiérrez

Grado en Física
Investigador Predoctoral
Instituto de Física de Partículas (IFCA-CSIC).

16:30 - 17:30

Workflows: Prefect, Airflow, etc

Marcos San Miguel Gutiérrez

Grado en Física
Investigador Predoctoral
Instituto de Física de Partículas (IFCA-CSIC).

22/04/2022

15:30 - 17:30

Examen

Aída Palacio Hoz

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

BIBLIOGRAFÍA Y ENLACES RELACIONADOS

Bibliografía

Cloud Computing: Automating the Virtualized Data Center, 1st Ed. V. Josyula, M. Orr, G. Page.
Cisco Press, 2012