



FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Tecnologías de procesamiento Big Data
Código	DTC-MBD-515
Título	Máster en Big Data. Tecnología y Analítica Avanzada/Master in Big Data Technologies and Advanced Analytics
Impartido en	Máster en Big Data. Tec. y Analítica Avanzada/Master in Big Data Technologies and Advanced Analytics [Primer Curso]
Nivel	Master
Cuatrimestre	Semestral
Créditos	7,5 ECTS
Carácter	Obligatoria
Departamento / Área	Departamento de Telemática y Computación

Datos del profesorado	
Profesor	
Nombre	Patricia Alcalde Sanz
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	palcalde@icai.comillas.edu
Profesor	
Nombre	Jorge Ayuso Rejas
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	jayuso@icai.comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura
Aportación al perfil profesional de la titulación
<p>El objetivo del curso es dar una visión general de las herramientas de ingesta y procesamiento del entorno big data, especialmente centrado en Spark y Hive.</p> <p>Al finalizar el curso, los alumnos:</p> <ul style="list-style-type: none">• Ser capaces de elegir cuál es la herramienta más adecuada para extraer datos de diferentes fuentes y llevarlos a un cluster Hadoop.• Tener experiencia con algunas herramientas y lenguajes de procesamiento (python, sql, etc).• Tener conocimientos profundos de spark con python y de cómo optimizar trabajos.



Prerrequisitos

Los estudiantes que deseen realizar este curso deben estar familiarizados con cualquier lenguaje de programación, preferiblemente python o SQL, y con los comandos y utilidades de Linux.

Competencias - Objetivos

Competencias

Competencias

CP1 Integrar las arquitecturas, técnicas de inteligencia artificial, análisis avanzado de datos y de visualización y de cumplimiento legal para ofrecer la solución global óptima.

CP2 Aplicar e integrar los flujos programáticos de datos masivos.

CP4 Implementar las técnicas de procesado de datos y usar las herramientas más habituales y apropiadas a las condiciones y requisitos de casos específicos

CP7 Aplicar conocimientos avanzados en Big Data y analítica de datos para desarrollar soluciones innovadoras en proyectos y en investigación, aportando y evaluando soluciones óptimas para el procesamiento y análisis de datos a gran escala

Habilidades o destrezas

HA1 Comunicar de manera oral y escrita con rigor técnico, claridad expositiva y coherencia argumentativa a todo tipo de interlocutores, técnicos y no técnicos.

HA2 Trabajar en equipos de carácter pluridisciplinar y/o internacional y organizar y liderar adecuadamente las dinámicas de grupo.

HA3 Desarrollar las habilidades interpersonales que requieren los entornos profesionales actuales (empatía, tolerancia, respeto, capacidad para aunar intereses contrapuestos).

HA4 Gestionar, organizar y planificar adecuadamente el trabajo y el tiempo, cumpliendo objetivos y estándares de calidad.

HA5 Mantener una formación y aprendizaje continuo y adaptación a los cambios tecnológicos y científicos.

Resultados de Aprendizaje

Conocimientos o contenidos

CO1 Entender los fundamentos de la analítica de datos y su aplicación en diversas áreas de la inteligencia artificial, destacando la integración en soluciones complejas y multidisciplinares para el análisis avanzado de datos masivos atendiendo a la diversidad de problemas específicos de cada área.

CO2 Comprender las técnicas de procesados de datos, las arquitecturas y herramientas más habituales y apropiadas para condiciones y requisitos de casos específicos.

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos



Contents

Theory

Unit 1. Introducción al desarrollo de software

1. IntelliJ
2. Git
3. Introduction to SQL

Unit 2. Ecosistema Hadoop

1. Hadoop Ecosystem
2. HDFS and Hadoop client
3. Hive

Unit 3. Herramientas de procesamiento y ETL

1. Kafka
2. Search Engines
3. NiFi

Unit 4. Introducción a la ciencia de datos

1. Python first steps
2. Scientific Python

Unit 5. Apache Spark para científicos de datos

1. Spark DataFrame
2. Spark ML (Machine Learning)
3. Spark packages

Laboratorio

Todas las sesiones tendrán un enfoque práctico. En el desarrollo del curso se propondrán a los alumnos prácticas que supondrán 60% de la nota final.

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Actividades de evaluación

Criterios de calificación

Peso



Prácticas	<ul style="list-style-type: none">• Media de las prácticas del alumno (0-10 puntos), todas las prácticas deben ser superadas.	60%
Examen Final	<ul style="list-style-type: none">• Comprensión de los conceptos teóricos.• Aplicación de estos conceptos a la resolución de problemas.• Análisis de los resultados de los ejercicios numéricos.	40%

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

- Apuntes y notebooks preparados por el profesor (disponibles en Moodle).
- White, T. (2015). Hadoop: The definitive guide 4th edition. " O'Reilly Media, Inc."
- Shreedharan, Hari (2014). Using Flume " O'Reilly Media, Inc."
- Karau, H., Konwinski, A., Wendell, P., & Zaharia, M. (2015). Learning spark: lightning-fast big data analysis. " O'Reilly Media, Inc."
- VanderPlas, J. (2016). Python Data Science Handbook.

En cumplimiento de la normativa vigente en materia de **protección de datos de carácter personal**, le informamos y recordamos que puede consultar los aspectos relativos a privacidad y protección de datos que ha aceptado en su matrícula entrando en esta web y pulsando "descargar"

<https://servicios.upcomillas.es/sedelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792>