



FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Tecnología de Datos Masivos / Big Data Technology
Código	DTC-BA-517
Título	Grado en Análisis de Negocios / Business Analytics por la Universidad Pontificia Comillas
Impartido en	Grado en Ingeniería en Tecnologías de Telecom. y Grado en Análisis de Negocios/Business Analytics [Quinto Curso]
Nivel	Reglada Grado Europeo
Cuatrimestre	Semestral
Créditos	6,0 ECTS
Carácter	Obligatoria (Grado)
Departamento / Área	Departamento de Telemática y Computación
Responsable	Jorge Lopez-Malla Matute

Datos del profesorado	
Profesor	
Nombre	Jorge López-Malla Matute
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	jlopezmalla@icai.comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura
Aportación al perfil profesional de la titulación
<p>El alumno obtendrá a lo largo del curso los conocimientos sobre computación distribuida, ecosistema Hadoop, bases de datos NOSQL y tecnologías de computación masiva tanto Batch como Streaming, centrándose principalmente en Apache Spark y Apache Kafka.</p> <p>Una vez adquiridos estos conceptos el alumno deberá poder identificar las necesidades de una arquitectura distribuida dependiendo de las necesidades, de resolver las distintas problemáticas de datos masivos usando tanto Map & Reduce como apache Spark para aquellas casuísticas Batch y Apache Spark y Apache Kafka para las casuísticas Streaming. Así mismo deberá ser capaz de identificar la mejor solución de almacenamiento NOSQL para cada caso de uso.</p>
Prerequisitos
Big Data / Datos Masivos

Competencias - Objetivos
Competencias



GENERALES		
CG01	Capacidad de organización y planificación en la identificación de problemas en el contexto de datos masivos	
	RA1	Identificar y organizar temporalmente las tareas necesarias para la realización sus actividades de aprendizaje, cumpliendo con los plazos establecidos de entrega de actividades.
	RA2	Ser capaz de integrarse y participar en el desarrollo organizado de un trabajo en grupo.
CG02	Capacidad de análisis de datos masivos procedentes de diversas fuentes: texto, audio, numérica e imagen	
	RA1	Describir, relacionar e interpretar situaciones y planteamientos sencillos
	RA2	Seleccionar las fuentes de información digital más significativas
	RA3	Identificar las fuentes de información más relevantes para cada problema
	RA4	Utilizar las herramientas digitales idóneas para la adquisición y gestión de la información y es capaz de integrar esa información de una manera racional
CG03	Resolución de problemas y toma de decisiones en un entorno de datos masivos tanto cuantitativos como cualitativos	
	RA1	Resolver problemas y realizar trabajos aplicando una diversidad metodológica de análisis
	RA2	Identificar correctamente los conocimientos aplicables a cada situación
	RA3	Identificar la metodología más apropiada para la resolución de los problemas planteados, conociendo las herramientas software más relevantes y aportando una visión crítica.
CG06	Habilidades interpersonales en la sociedad de la información: escuchar, argumentar y debatir	
	RA1	Utiliza el diálogo para colaborar y generar buenas relaciones, escuchando las opiniones de los demás y estableciendo diálogos constructivos
	RA2	Comunica sus ideas de manera efectiva y argumentada
	RA3	Conoce la técnica del debate y la oratoria y sabe emplearla en cuestiones profesionales
CG07	Capacidad de liderazgo y trabajo en equipo, en la sociedad de la información	
	RA1	Participar de forma activa en el trabajo de grupo compartiendo información, conocimientos y experiencias
	RA2	Orientarse a la consecución de acuerdos y objetivos comunes
	RA3	Ser capaz de organizar y/o coordinar equipos de trabajo
CG08	Capacidad crítica y autocrítica en la sociedad de la información	
	RA1	Identificar los supuestos y las limitaciones de los modelos y metodologías a aplicar en cada situación



	RA2	Ejercer una valoración crítica de los modelos y metodologías a aplicar en cada situación
CG09	Compromiso ético en la sociedad de la información	
	RA1	Ser honesto en el desarrollo de las actividades académicas y en otros aspectos de la vida y no ser pasivo ante la deshonestidad de otros.
	RA2	Conocer y desarrollar las bases de la ética profesional
CG11	Capacidad para aprender y trabajar autónomamente en la sociedad de la información	
	RA1	Organizar los materiales de la asignatura con el fin de articular su aprendizaje
	RA2	Poseer criterio para decidir la utilización de los medios humanos y otros recursos puestos a su disposición para el aprendizaje
	RA3	Realizar sus trabajos y su actividad necesitando sólo unas indicaciones iniciales y un seguimiento básico
ESPECÍFICAS		
CE25	Comprender los conceptos más importantes asociados al Big Data y sus elementos característicos	
	RA1	Conocer el concepto de paralelismo y la importancia de la distribución de datos voluminosos
	RA2	Identificar las necesidades específicas asociadas al manejo, adquisición y almacenamiento de volúmenes masivos de datos.
	RA3	Comprender y conocer las tecnologías de comunicaciones (redes) y computación precisas para la distribución y recuperación de información, así como de su análisis.
CEO34	Entender la computación y las bases de datos distribuidas, con énfasis en la arquitectura Hadoop y las herramientas actuales más extendidas en el contexto de Big Data.	
	RA01	Conoce y comprende la problemática del Big Data y las técnicas más avanzadas de computación y almacenamiento usadas en la actualidad.
	RA02	Comprende y es capaz de administrar una arquitectura Hadoop y otras arquitecturas NoSQL.
	RA03	Comprende y aplica las técnicas de MapReduce y otras técnicas avanzadas de computación específicas para el análisis de datos voluminosos.

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

1. Conceptos básicos de tecnologías de datos masivos
2. Ecosistema Hadoop

3. Soluciones de almacenamiento distribuido
4. Apache Spark Arquitectura y uso para la resolución de casuísticas de datos masivos
5. Machine Learning y Apache Spark
6. Conceptos de Streaming
7. Resolución de problemas Streaming usando Spark Streaming
8. Resolución de problemas Streaming usando Apache Kafka
9. Datawarehouse y Data lake
10. Soluciones Cloud para procesamientos de datos masivos

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

RESUMEN HORAS DE TRABAJO DEL ALUMNO

HORAS PRESENCIALES
HORAS NO PRESENCIALES
CRÉDITOS ECTS: 6,0 (0 horas)

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Actividades de evaluación	Criterios de evaluación	Peso
Exámenes final	Se realizará una prueba teórico-práctica la final de la asignatura en el cual el alumno demostró los resultados de aprendizaje adquiridos.	50
Desarrollo de prácticas individuales o en grupo	Se realizarán prácticas sobre las herramientas o tecnologías vistas en clase.	10
Desarrollo de prácticas individuales o en grupo	Se realizarán prácticas sobre las herramientas o tecnologías vistas en clase.	20
Prueba intersemestral	Se realizará una prueba a mitad del semestre evaluando los conocimientos adquiridos de una forma teórica práctica.	10
Participación activa del alumno en el aula.	Participación, actitud y desarrollo de trabajos realizados durante la asignatura.	10



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

GUÍA DOCENTE

2022 - 2023

Calificaciones

La nota final (NF) se calculará de la siguiente forma, para la convocatoria ordinaria como para la extraordinaria.

$NF = 50\% \text{ Examen final} + 10\% \text{ Prueba intersemestral} + 30\% \text{ Prácticas} + 10\% \text{ Participación y actitud}$

Para aprobar la asignatura las notas mínimas deberán ser las siguiente:

- Prueba intersemestral: 4.
- Examen final: 5.
- Prácticas: 6.

BIBLIOGRAFÍA Y RECURSOS

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>