

**Fundamentos de la tecnología Datos Masivos/ Fundamentals of Big Data technology**

<b>Número total de créditos ECTS</b>		6
<b>Tipología</b>		Obligatoria
<b>Organización temporal</b>		1er. curso/ 1er. semestre
<b>Modalidad</b>		Presencial
<b>Idioma</b>		Castellano
<b>Contenidos</b>	<ul style="list-style-type: none"> <li>· Introducción a los Datos Masivos: procesado, análisis y visualización. Fuentes de datos en entornos Big data.</li> <li>· Estructuras de datos y tecnologías para selección de datos útiles. Criterios de calidad de datos en el análisis de Datos Masivos.</li> <li>· Técnicas de rastreo, procesamiento, indexación y recuperación de información estructurada y no estructurada. Principales estrategias de scraping y crawling.</li> <li>· Diseño y explotación de Sistemas de almacenamiento y gestión de Datos Masivos.</li> <li>· Sistemas de almacenamiento para Datos Masivos. Sistemas distribuidos. Teorema CAP</li> <li>· Paradigmas de modelado de datos más utilizados en entorno Datos Masivos: SQL y NoSQL.</li> <li>· Soluciones tecnológicas Datos Masivos disponibles.</li> <li>· Riesgos y medidas de Seguridad en el Datos Masivos Big Data. Legislación y Datos Masivos.</li> <li>· Propiedad intelectual sobre el desarrollo de proyectos de Datos Masivos. Anonimización, privacidad desde el diseño y por defecto (privacy by design) y análisis de riesgos (Evaluación de Impacto en la Privacidad Privacy Impact Assessment).</li> <li>· Privacidad y Datos Masivos Big Data. Protección de datos personales en España</li> <li>· Aseguramiento de la seguridad de la información en entornos Datos Masivos. Fugas de información.</li> </ul> <p>-----</p> <ul style="list-style-type: none"> <li>· Introduction to Big Data: processing, analysis and visualisation. Data sources in Big Data environments.</li> <li>· Data structures and technologies for selecting useful data. Data quality criteria in Big Data.</li> <li>· Techniques for crawling, processing, indexing and recovery of structured and unstructured information. Main scraping and crawling strategies.</li> <li>· Design and operation of Big Data storage and management systems.</li> <li>· Storage systems for Big Data. Distributed systems. CAP Theorem</li> <li>· Most used data modelling paradigms in Big Data environment: SQL and NoSQL.</li> <li>· Available Big Data technological solutions.</li> <li>· Risks and security measures in Big Data. Legislation and Big Data.</li> <li>· Intellectual property on the development of Big Data projects. Anonymisation, privacy by design and risk analysis (Privacy Impact Assessment).</li> <li>· Privacy and Big Data. Personal data protection in Spain.</li> <li>· Information security assurance in Big Data environments. Information leaks.</li> </ul>	
<b>Resultados de aprendizaje TÍTULO</b>	<b>Conocimientos y contenidos</b>	<p>CC1 Comprender las de técnicas y herramientas de programación especializada en modelado, gestión, procesamiento y análisis de datos en entornos de Big Data. //Understand specialized programming techniques and tools modelling, management, processing and analysis of data in Big Data environments.</p> <p>CC2 e Identificar las implicaciones legales, morales y éticas, en lo referente al uso de datos personales en el contexto de Datos Masivos// Identify the legal, moral and ethical</p>

		implications regarding the use of personal data in the context of Big Data.
	<b>Habilidades y destrezas</b>	<p>HD1 - Desarrollar arquitecturas, eficaces y eficientes, aplicando los diferentes modelos de almacenamiento de datos masivos, así como sistemas de bases de datos, para su procesamiento en infraestructuras distribuidas. // Develop effective and efficient architectures by applying different mass data storage models as well as database systems for processing in distributed infrastructures.</p> <p>HD2 Resolver problemas reales de clasificación, modelización e interpretación de un conjunto de datos en el contexto de Datos Masivos// Solve real problems of classification, modelling and interpretation of a dataset in Big Data context.</p> <p>HD3 Utilizar eficientemente las arquitecturas y servicios en Nube disponibles para el análisis y procesamiento de datos masivos. //Efficiently use the available Cloud architectures and services for big data análisis and processing.</p> <p>HD5 Sintetizar los resultados de los modelos de análisis de datos en un contexto de toma de decisiones. // Synthesise the results of data analysis models in a decision-making context.</p>
	<b>Competencias</b>	<p>CP1 Gestionar modelos de Ciencia de Datos en su análisis, diseño y despliegue en el contexto de Datos Masivos con infraestructuras y servicios de Computación en Nube. // Manage Data Science models in their analysis, design, and deployment in a Big Data context with Cloud Computing infrastructures and services.</p> <p>CP6 Identificar la solución más óptima para un caso propuesto que utilice modelos y técnicas de Datos Masivos, en términos de eficiencia, eficacia e interpretación de resultados. // Identify the most optimal Big Data solution for a Big Data issue, in terms of efficiency, effectiveness and interpretation of results.</p>
<b>Resultados de aprendizaje ASIGNATURA</b>		
<p>RA.1.- Identificar posibles fuentes de datos y evaluar la aplicabilidad de los datos utilizando parámetros de calidad, que representen de forma óptima una situación física o virtual.</p> <p>RA.2.- Modelar datos siguiendo los paradigmas más utilizados dentro del entorno de Datos Masivos Big Data: SQL y NoSQL.</p> <p>RA.3.- Aplicar técnicas de procesamiento de grandes cantidades de información que deriven en un trato eficiente de los datos.</p> <p>RA.4.- Elaborar, en base a los requisitos extraídos de un conjunto de datos, un sistema de almacenamiento acorde a las necesidades detectadas.</p> <p>RA.5.- Diseñar y poner en práctica un sistema que combine técnicas de búsqueda, captura, almacenamiento, análisis, explotación y visualización de datos, para modelizar eficientemente la realidad subyacente.</p> <p>RA6.- Identificar de forma proactiva los beneficios y los riesgos inherentes a la aplicación de técnicas de procesamiento masivo de datos en una situación o problema concreto.</p> <p>RA.7.- Saber identificar los principales riesgos asociados al procesamiento de datos, y aplicar medidas de protección que garanticen la seguridad y privacidad del usuario, satisfaciendo siempre tanto la legislación correspondiente como los principios éticos.</p> <p>RA.8.- Ser capaz de instalar, configurar y gestionar códigos en diferentes lenguajes especializados en el procesamiento de datos masivos en la nube.</p> <p>RA.9.- Diseñar y desarrollar sistemas de virtualización de infraestructuras en la nube.</p> <p>RA.10.- Ser capaz de desarrollar y desplegar aplicaciones para la nube.</p> <p>-----</p>		

RA.1.- Identify possible data sources and evaluate the applicability of the data using quality parameters, which optimally represent a physical or virtual situation.

RA.2.- Model data following the most used paradigms within the Big Data environment: SQL and NoSQL.

RA.3.- Apply techniques for processing large amounts of information that result in an efficient treatment of data.

RA.4.- Develop, on the basis of the requirements extracted from a set of data, a storage system in accordance with the needs detected.

RA.5.- Design and implement a system that combines data search, capture, storage, analysis, exploitation and visualisation techniques in order to efficiently model the underlying reality.

RA6.- Proactively identify the benefits and risks inherent to the application of massive data processing techniques in a specific situation or problem.

RA.7.- Know how to identify the main risks associated with data processing, and apply protection measures that guarantee the security and privacy of the user, always satisfying both the corresponding legislation and ethical principles.

RA.8.- Be able to install, configure and manage codes in different languages specialised in the processing of massive data in the cloud.

RA.9.- Design and develop virtualisation systems for cloud infrastructures.

RA.10.- Be able to develop and deploy applications for the cloud.

Actividades formativas	Horas totales
AF1. Lección magistral participativa/ Participative master class	20
AF2. Prácticas de laboratorio/ Laboratory practicals	14
AF3. Resolución de problemas/ Problem solving	14
AF8. Trabajo autónomo/Autonomous work	98
AF7. Evaluación/Evaluation	4
<b>Total</b>	<b>150</b>

* Sistemas de evaluación	MÍNIMO	MÁXIMO
La Evaluación del Portafolio. Informe sobre resolución de problemas o estudio de casos / Portfolio assessment. Report on problem solving or case studies	20	30
Prueba final / Final test.	40	60
Participación en grupos de debate /Participation in discussion groups	10	10
Evaluación del portafolio. Informe sobre Diseño de Proyectos / Portfolio evaluation. Report on Project Design	10	20

\* El peso de cada sistema de evaluación quedará definido en la guía docente de la asignatura.