

GUÍA DOCENTE

ECOSISTEMAS BIG
DATA Y
COMPUTACIÓN EN LA
NUBE / BIG DATA
ECOSYSTEMS AND
CLOUD COMPUTING

Ecosistemas Big Data Y Computación En La Nube / Big Data Ecosystems And Cloud Computing

Número total de créditos ECTS		6
Tipología		Obligatoria
Organización temporal		Curso 1; Semestre 1
Idioma		Castellano e inglés
Contenidos		<p>1. Arquitecturas de Procesamiento Distribuido: Fundamentos y ecosistema Hadoop. Procesamiento en memoria con Spark (RDDs y DataFrames).</p> <p>2. Almacenamiento en la Nube para Data Science: Servicios de almacenamiento de objetos (S3/Blob/GCS) y su integración con clústeres de cálculo.</p> <p>3. Despliegue de Infraestructuras Cloud: Plataformas gestionadas (AWS EMR/Dataproc) vs Infraestructura como Servicio.</p> <p>4. Introducción a la Contenerización en Flujos de Datos: Uso de contenedores (Docker) para la portabilidad de modelos y orquestación básica en entornos productivos.</p> <p>-----</p> <p>1. Distributed Processing Architectures: Fundamentals and the Hadoop ecosystem. In-memory processing with Spark (RDDs and DataFrames).</p> <p>2. Cloud Storage for Data Science: Object storage services (S3/Blob/GCS) and their integration with compute clusters.</p> <p>3. Cloud Infrastructure Deployment: Managed platforms (AWS EMR/Dataproc) vs. Infrastructure as a Service.</p> <p>4. Introduction to Containerization in Data Pipelines: Use of containers (Docker) for model portability and basic orchestration in production environments.</p>
Resultados de aprendizaje TÍTULO	Conocimientos y contenidos	<p>CC01 Comprender estrategias avanzadas de preprocesamiento, imputación de valores ausentes y codificación de variables para garantizar la calidad del dato en modelos estadísticos complejos. / Understand advanced preprocessing strategies, missing-value imputation techniques, and variable encoding methods to ensure data quality in complex statistical models.</p> <p>CC05 Analizar arquitecturas distribuidas de Big Data y servicios de computación en la nube, optimizando el rendimiento y la escalabilidad de los flujos de datos masivos. / Analyze distributed Big Data architectures and cloud computing services, optimizing the performance and scalability of massive data workflows.</p> <p>CC09 Fundamentar soluciones de ingeniería de datos para la ingesta y almacenamiento persistente en entornos distribuidos, asegurando la integridad y disponibilidad en sistemas de alta concurrencia. / Provide data engineering solutions for ingestion and persistent storage in distributed environments, ensuring integrity and availability in high-concurrency systems.</p>
	Habilidades y destrezas	<p>HD02 Desarrollar scripts y flujos de trabajo avanzados utilizando lenguajes de programación específicos (como Python o R) para la ingesta, manipulación y modelado analítico de datos, optimizando el código para entornos de producción. / Develop advanced scripts and workflows using specific programming languages (such as Python or R) for data ingestion, manipulation, and analytical modeling, optimizing the code for production environments.</p> <p>HD05 Evaluar y seleccionar la arquitectura analítica y los modelos algorítmicos más idóneos para resolver problemas complejos de negocio, justificando la elección en términos de eficiencia computacional, precisión y escalabilidad. / Evaluate and select the most suitable analytical architecture and algorithmic models to address</p>

		<p>complex business problems, justifying the choice in terms of computational efficiency, accuracy, and scalability.</p> <p>HD08 Confeccionar aplicaciones de software analítico, integrando algoritmos predictivos en entornos de producción mediante buenas prácticas de ingeniería de software (MLOps/ Develop analytical software applications, integrating predictive algorithms in production environments following software engineering best practices (MLOps).</p> <p>HD09 Aplicar modelos de segmentación, agrupación de observaciones y reducción de variables / Apply segmentation models, clustering of observations, and variable reduction.</p>
	Competencias	<p>CP01 Desarrollar e implementar soluciones de Ciencia de Datos en proyectos con incertidumbre técnica, asegurando que los resultados respondan a los objetivos de la organización. / Develop and implement Data Science solutions within projects characterized by technical uncertainty, ensuring that outcomes align with organizational objectives.</p>

Resultados de aprendizaje ASIGNATURA

1. Instalar y preparar infraestructuras de Big Data en la nube (clústeres Hadoop y Spark).
2. Organizar el almacenamiento distribuido en servicios cloud (S3, ADLS, HDFS).
3. Gestionar recursos de computación escalable en entornos AWS, Azure y GCP.
-
4. Install and set up Big Data infrastructures in the cloud (Hadoop and Spark clusters).
5. Organise distributed storage on cloud services (S3, ADLS, HDFS).
6. Manage scalable computing resources in AWS, Azure, and GCP environments.

Modalidad Presencial	Actividades formativas	Horas totales	Horas presenciales	Horas virtuales síncronas	Horas virtuales asíncronas
	Clases Expositivas / Participatory learning	22	22	0	0
	Seminarios / Seminars	2	2	0	0
	Clases prácticas / Practical lessons	10	10	0	0
	Prácticas de Laboratorio / Laboratory practicals	12	12	0	0
	Trabajo autónomo / Autonomous work	102	0	0	102
	Prueba de evaluación final / Final evaluation test	2	2	0	0
	Total	150	48	0	102
	Sistemas de evaluación			MÍNIMO	MÁXIMO
	Evaluación Final: prueba o examen / Final evaluation test			50	50
Resolución de problemas / Problem resolution			10	30	
Estudio de casos / Proyectos / Case studies / projects			10	30	
Total			70	110	