

Introdución a Data Lakes e Data Warehouses - 2ª edición

Descrición

Este módulo introdutorio e estratéxico de 5 horas de duración céntrase no deseño, gobernanza e arquitectura de infraestruturas escalables para o almacenamento de Big Data territoriais. O alumnado adquirirá os criterios metodolóxicos e institucionais necesarios para conceptualizar o ecosistema para estruturar cantidades masivas de información. Seguindo as directrices de marcos de referencia globais como TDWI e o EDM Council, o módulo analizará en detalle as principais diferenzas arquitectónicas entre un Data Lake e un Data Warehouse en contornas SIX (Sistemas de Información Xeográfica), os criterios para os datos como activo analítico e as lóxicas de inxestión e modelado de datos en capas semánticas (Bronce, Prata, Ouro). O curso permite o deseño lóxico de solucións, integrando a xestión da calidade por deseño, a liñaxe de datos e os principios europeos fundamentais de soberanía tecnolóxica e interoperabilidade federada baixo o marco Gaia-X e o estándar CDMC.

Dita actuación realizarase no marco do proxecto denominado "Aceleración de ecosistemas de emprendemento e innovación baseados en Xemelgos Dixitais". Plan de Recuperación, Transformación e Resiliencia - Financiado pola Unión Europea - NextGenerationEU.

Obxectivos

Deseñar e avaliar arquitecturas lóxicas para almacenamento masivo e Big Data, identificando o papel dos catálogos e metadatos dentro dunha estratexia corporativa.
Diferenciar, tanto cualitativa como cuantitativamente, os casos de uso óptimos para un Data Lake fronte a un Data Warehouse en proxectos de sistemas de información xeográfica (GIS).
Estruturar eficientemente as capas de datos semánticos, garantindo marcos de calidade de datos (Avaliación da Calidade de Datos) e unha liñaxe de datos completa.
Comprender e aplicar marcos europeos de gobernanza de datos e estándares globais de xestión da nube (Gaia-X e CDMC).

Dirixido a

Arquitectos de datos, xestores de TI, enxeñeiros de datos, analistas de negocios sénior e técnicos de bases de datos de empresas privadas ou administracións públicas que participen no deseño estratéxico e a gobernanza de proxectos con grandes volumes de información territorial.

Coñecementos previos recomendados:

- Coñecemento básico da estrutura e funcionamento das bases de datos relacionais (SQL) e conceptos xerais de almacenamento de información.
- Familiaridade cos fluxos de datos tradicionais ou procesos básicos de integración (coñecemento xeral do que é un proceso ETL: Extraer, Transformar, Cargar).
- Coñecemento xeral ou comprensión do ecosistema dos Sistemas de Información Xeográfica (SIG) ou dos datos espaciais (moi recomendable debido ao contexto de Territorio Intelixente do programa).
- Non se requiren coñecementos avanzados de programación en Python, Scala ou infraestrutura nativa de Big Data, xa que o módulo céntrase no deseño lóxico, a gobernanza e a arquitectura conceptual.

BENEFICIOS

Diploma de asistencia

Perfil do docente

Profesor universitario e Enxeñeiro de Datos con ampla traxectoria docente e práctica no despregamento de solucións Big Data.

PROGRAMA	Programación 2026/27
TIPO	CURSO
MATRÍCULA	Gratuíta
PERIODO INSCRICIÓN	01/07/2026 - 31/08/2026
CRITERIOS DE SELECCIÓN	Por orde de solicitude
Nº PRAZAS	20



METODOLOXÍA	Virtual
TIPO DE EDICIÓN	Edición única (desempregados/as e ocupados/as)
DURACIÓN	5 horas
DATA INICIO	04/09/2026
DATA FIN	04/09/2026
HORARIO	Venres 4 de setembro de 15:00 a 20:00 horas.
LUGAR DE DOCENCIA	Edificio localizado na r/Airas Nunes s/n, barrio de Conxo, en
CERTIFICACIÓN OFICIAL	Non
MÓDULOS TRANSVERSAIS	Igualdade de 5 horas
TECNOLOXÍA	Xemelgos dixitais/Gemelos digitais
PERIODO DOCENCIA	04/09/2026 - 04/09/2026

Temario

1. Fundamentos da arquitectura de almacenamento masivo e activos de datos (Consello TDWI e EDM): Arquitecturas estruturadas e non estruturadas no xemelgo dixital. Evolución cara ao Data Lakehouse. Identificación de compoñentes de datos corporativos, Catálogo de datos e glosario empresarial aplicado.
2. Criterios de selección arquitectónica en contornas SIX: Diferenzas clave de deseño entre Data Lake e Data Warehouse (custo, latencia e usabilidade). Clasificación de datos segundo a criticidade e a sensibilidade (Privacidade por deseño).
3. Modelado de datos en capas semánticas e xestión da calidade dos datos: Estruturación de datos territoriais en capas de medallas (bronce, prata, ouro). Implementación de regras e controis de calidade no ciclo de vida dos datos para evitar a degradación do repositorio (Data Swamp).

4. Patróns de inxestión, optimización e trazabilidade (liñaxe de datos): Procesamento por lotes fronte a transmisión en tempo real. Estratexias para optimizar fluxos e custos. Documentación da liñaxe de datos desde o sensor territorial ata o destino analítico para auditorías de fiabilidade.

5. Soberanía tecnolóxica, espazos de datos e o estándar CDMC: fundamentos e implicacións de Gaia-X. O estándar CDMC (Cloud Data Management Capabilities) aplicado a entornos híbridos. Control de uso, gobernanza de nodos federados e cumprimento normativo nos espazos de datos da UE.