



## Guía Docente

Datos Identificativos					2024/25
Asignatura (*)	Enxeñaría de Datos	Código	614544002		
Titulación					
Descritores					
Ciclo	Período	Curso	Tipo	Créditos	
Mestrado Oficial	1º cuatrimestre	Primeiro	Obrigatoria	3	
Idioma	Inglés				
Modalidade docente	Presencial				
Prerrequisitos					
Departamento	Ciencias da Computación e Tecnoloxías da Información				
Coordinación	Bernardo Roca, Guillermo de	Correo electrónico	guillermo.debernardo@udc.es		
Profesorado	Bernardo Roca, Guillermo de	Correo electrónico	guillermo.debernardo@udc.es		
Web					
Descrición xeral	O obxectivo da materia é a introdución dos aspectos básicos da enxeñaría de datos, fundamentalmente no ámbito do Big Data. As competencias adquiridas permitirán a análise e a xestión eficiente de información heteroxénea, tanto estruturada como non estruturada, dentro do desenvolvemento de aplicacións de IA, alí onde os métodos tradicionais mostren a súa insuficiencia.				

## Competencias / Resultados do título

Código	Competencias / Resultados do título
--------	-------------------------------------

## Resultados da aprendizaxe

Resultados de aprendizaxe	Competencias / Resultados do título		
RA1 - Desenvolver a capacidade de analizar e modelar datos para o seu procesado en sistemas intelixentes.	AM16	BM6 BM7	CM3 CM9
RA2 - Coñecer e comprender o proceso de extracción, limpeza, transformación, carga e procesado de datos.	AM16	BM2 BM3 BM8	CM3 CM7 CM9
RA3 - Coñecer e saber utilizar bases de datos multidimensionais e de tipo NoSQL.		BM3 BM4 BM7	CM8
RA4 - Coñecer os fundamentos de data lakes e data warehouses.		BM2 BM5 BM7 BM8	CM3 CM7 CM8

## Contidos

Temas	Subtemas
Conceptos e fundamentos de Enxeñaría de datos	Conceptos e definicións básicas, problemas de carga eficiente en escenarios Big Data, almacenamento de datos masivos e acceso aos mesmos.
Técnicas de limpeza e preparación de datos.	Técnicas máis comúns. Definición de fluxos de procesamento. Medidas de calidade.
Estruturas avanzadas e almacéns de datos eficientes para Big Data	Data warehouses e BD multidimensionais, Data lakes, Bases de Datos NoSQL.

## Planificación



Metodoloxías / probas	Competencias / Resultados	Horas lectivas (presenciais e virtuais)	Horas traballo autónomo	Horas totais
Sesión maxistral	B4 B5 C3 C9	12	0	12
Prácticas de laboratorio	A17 B2 B5 B7 C3	10	30	40
Proba mixta	A17 B2 B3 B6 B7 B8 C7 C8	3	20	23
Atención personalizada		0	0	0

\*Os datos que aparecen na táboa de planificación son de carácter orientativo, considerando a heteroxeneidade do alumnado

Metodoloxías	
Metodoloxías	Descrición
Sesión maxistral	O profesorado presenta un tema ao alumnado co obxectivo de facilitar un conxunto de información con alcance concreto. AVALIACIÓN CONTINUA Carácter obrigatorio asistencia non obrigatoria AVALIACIÓN GLOBAL Carácter obrigatorio
Prácticas de laboratorio	O profesorado da materia expón ao alumnado un problema ou problemas de carácter práctico cuxa resolución require a comprensión e aplicación dos contidos teórico-prácticos incluídos nos contidos da materia. O alumnado pode traballar a solución aos problemas expostos de forma individual ou en grupos. AVALIACIÓN CONTINUA Carácter obrigatorio Asistencia obrigatoria (mín. 75% das prácticas) AVALIACIÓN GLOBAL Carácter obrigatorio
Proba mixta	O exame abarca todos os temas do curso. O alumnado debe desenvolver, relacionar, organizar e presentar os coñecementos que posúe sobre cada tema dado nunha resposta razoada e ben articulada.

Atención personalizada	
Metodoloxías	Descrición
Sesión maxistral	Sesión maxistral: atenderanse as dúbidas relacionadas coas metodoloxías e os casos de estudo analizados nas leccións.
Prácticas de laboratorio	Prácticas de laboratorio: atenderanse as dúbidas relacionadas cos casos de estudo a analizar.

Avaliación			
Metodoloxías	Competencias / Resultados	Descrición	Cualificación
Proba mixta	A17 B2 B3 B6 B7 B8 C7 C8	O exame abarca todos os temas do curso. O alumnado debe desenvolver, relacionar, organizar e presentar os coñecementos que posúe sobre cada tema dado nunha resposta razoada e ben articulada. Resultados de aprendizaxe avaliados: RA1, RA3, RA4	40
Prácticas de laboratorio	A17 B2 B5 B7 C3	Entrega de prácticas de laboratorio expostas ao longo do curso nas datas estipuladas previamente. Resultados de aprendizaxe avaliados: RA3, RA4	60



## Observacións avaliación

**SISTEMA DE AVALIACIÓN CONTINUAPROBA 1:** Prácticas de laboratorio% Cualificación: 60%.% mínimo: Para a liberación desta parte da materia, o alumno deberá obter unha cualificación igual ou superior a 5 puntos (sobre 10).

**PROBA 2:** Exame% Cualificación: 40%. % mínimo: Para a liberación desta parte da materia, o alumno deberá obter unha cualificación igual ou superior a 5 puntos(sobre 10).

**SISTEMA DE AVALIACIÓN GLOBAL**Procedemento de elección da modalidade de avaliación global: considérase que o alumnado optou polo sistema de avaliación global se non realiza a proba 1 do sistema de avaliación continua.

**PROBA 1:** Prácticas de laboratorio% Cualificación: 60%.% mínimo: Para a liberación desta parte da materia, o alumno deberá obter unha cualificación igual ou superior a 5 puntos (sobre 10).

**PROBA 2:** Exame% Cualificación: 40%. % mínimo: Para a liberación desta parte da materia, o alumno deberá obter unha cualificación igual ou superior a 5 puntos(sobre 10).

**CRITERIOS DE AVALIACIÓN DA CONVOCATORIA EXTRAORDINARIA (2ª OPORTUNIDADE) E FIN DE CARREIRA** Utilizaranse os sistemas de avaliación continua e global descritos anteriormente.

**REXISTRO PROCESO DE CUALIFICACIÓN**

Independentemente do sistema de avaliación e da convocatoria, se non se supera algunha parte da avaliación, pero a puntuación global é superior a 4 (sobre 10), a cualificación en acta será de 4.

**OUTRAS CONSIDERACIÓNS** No caso de producirse algunha contradición entre as distintas versións da guía, por algún erro de tradución, a versión que prevalecerá é a versión en inglés.Todos os aspectos relacionados con ?dispensa académica?, ?dedicación ao estudo?, ?permanencia? e ?fraude académica? rexeranse de acordo coa normativa académica vixente da UDC.

## Fontes de información

<b>Bibliografía básica</b>	<ul style="list-style-type: none"> <li>- Sadalage, Fowler (2012). NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley</li> <li>- Avi Silberschatz, Henry F. Korth, S. Sudarshan (2010). Database System Concepts. McGraw-Hill</li> <li>- Ihab F. Ilyas, Xu Chu, (2019). Data Cleaning. Association for Computing Machinery. ACM</li> <li>- Alex Gorelik (). The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science. O'Reilly</li> </ul>
<b>Bibliografía complementaria</b>	<ul style="list-style-type: none"> <li>- Matt Casters, Roland Bouman, Jos van Dongen (2013). Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration. Wiley</li> </ul>

## Recomendacións

**Materias que se recomenda ter cursado previamente**

**Materias que se recomenda cursar simultaneamente**

**Materias que continúan o temario**

## Observacións

(\*A Guía docente é o documento onde se visualiza a proposta académica da UDC. Este documento é público e non se pode modificar, salvo casos excepcionais baixo a revisión do órgano competente dacordo coa normativa vixente que establece o proceso de elaboración de guías