



Guía Docente				
Datos Identificativos				2023/24
Asignatura (*)	Análise de Datos con HPC		Código	614473108
Titulación				
Descriptor				
Ciclo	Período	Curso	Tipo	Créditos
Mestrado Oficial	2º cuadrimestre	Primeiro	Optativa	6
Idioma	Inglés			
Modalidade docente	Presencial			
Prerrequisitos				
Departamento	Enxeñaría de Computadores			
Coordinación	López Taboada, Guillermo	Correo electrónico	guillermo.lopez.taboada@udc.es	
Profesorado	López Taboada, Guillermo	Correo electrónico	guillermo.lopez.taboada@udc.es	
	Rodríguez Álvarez, Gabriel		gabriel.rodriguez@udc.es	
Web	aula.cesga.es			
Descrición xeral	A cantidade cada vez maior de información accesible a través de Internet fai que o procesamento eficiente de grandes cantidades de datos sexa cada vez de maior interese. Isto levou ao desenvolvemento de novas técnicas de almacenamento e procesamento de inxentes cantidades de información, denominadas técnicas Big Data, que se adaptan de forma natural aos sistemas distribuídos.			

Competencias do título	
Código	Competencias do título

Resultados da aprendizaxe			
Resultados de aprendizaxe		Competencias do título	
O alumno será capaz de seleccionar, instalar, configurar e xestionar o software básico para o procesamento de datos masivos.	AP1	BP2	CP1
	AP2	BP6 BP8 BP10	
O alumno será capaz de implementar códigos nalguna linguaxe especializada no procesamento de datos masivos.	AP2	BP1 BP2 BP10	CP1
O alumno coñecerá e aprenderá a utilizar algunhas das ferramentas dispoñibles para Data Engineering (en particular, par Inxesta/Almacenamento/Procesado/Visualización).	AP1	BP1	CP1
	AP2	BP2	CP4
O alumno adquirirá a habilidade necesaria para a procura, selección e manexo de recursos (bibliografía, software, etc.) relacionados con Big Data.	AP1	BP1	CP1
	AP2	BP6	CP4

Contidos	
Temas	Subtemas
1. Introducción a Data Engineering	1.1 HPC vs Big Data: similitudes e diferencias no tratamento de datos 1.2 Tecnoloxías Hardware e Software para High Performance Data Engineering 1.3 Data Engineering en infraestruturas HPC vs entornos Cloud
2. Introducción a Análítica de Datos	2.1 Exploratory Data Analytics 2.2 Introducción a Machine Learning



3. Etapas de Data Engineering	3.1 Modelado (Formatos, Compresión, Deseño de Esquemas) 3.2 Inxesta (Periodicidade, Transformacións, Ferramentas) 3.3 Almacenamento (HDFS y BBDD NoSQL, HBase, MongoDB, Cassandra) 3.4 Procesado (Batch, Real-Time) 3.5 Orquestación 3.6 Análise (SQL, Machine Learning, Graphs, UI) 3.7 Gobernanza 3.8 Integración con BI (Visualización)
4. Casos de Uso	4.1 Aplicacións en Internet das cousas (entornos Smart e Industria 4.0) 4.2 Aplicacións en ciencias e enxeñarías

Planificación				
Metodoloxías / probas	Competencias	Horas presenciais	Horas non presenciais / traballo autónomo	Horas totais
Sesión maxistral	A1 A2 B1 C4	18	0	18
Prácticas de laboratorio	B1 B8 B10	20	60	80
Traballos tutelados	A1 A2 B1 B2 B8	0	45	45
Discusión dirixida	B6 C1 C4	4	2	6
Atención personalizada		1	0	1
*Os datos que aparecen na táboa de planificación son de carácter orientativo, considerando a heteroxeneidade do alumnado				

Metodoloxías	
Metodoloxías	Descrición
Sesión maxistral	Impartidas polos profesores da asignatura. Inclúen exposición de material teórico, así como de seminarios.
Prácticas de laboratorio	Resolución de problemas e casos prácticos.
Traballos tutelados	Realización de prácticas de maior entidade de forma semiautónoma, guiados polos profesores da asignatura.
Discusión dirixida	Orientación para a realización dos traballos individuais ou en grupo, resolución de dúbidas e actividades de avaliación continua.

Atención personalizada	
Metodoloxías	Descrición
Discusión dirixida Prácticas de laboratorio Traballos tutelados	Durante as prácticas de laboratorio, traballos tutelados, e discusións dirixidas, os estudantes poderán presentar preguntas, dúbidas, etc. O profesor, atendendo ás súas solicitudes, repasará conceptos, resolverá novos problemas ou utilizará calquera actividade que considere adecuada para resolver as cuestións expostas.

Avaliación			
Metodoloxías	Competencias	Descrición	Cualificación
Prácticas de laboratorio	B1 B8 B10	Evaluación de las prácticas llevadas a cabo por los estudiantes.	50
Traballos tutelados	A1 A2 B1 B2 B8	Evaluación de los trabajos tutelados desarrollados por los estudiantes.	50

Observacións avaliación



Non presentado: Considerarase non presentado @ alumn@ que non entregue ningunha práctica nin traballo academicamente dirixido.

Segunda oportunidade (extraordinaria - xuño/xullo): Volver a realizar aquelas prácticas e traballos tutelados que non se entregaran ou versións melloradas dos xa entregados.

Para os casos de realización fraudulenta de exercicios ou probas será de aplicación o recollido na Normativa de avaliación do rendemento académico dos estudantes e de revisión de cualificacións.

En concreto, a realización fraudulenta das probas ou actividades de avaliación, unha vez comprobada, implicará directamente a cualificación de suspenso na convocatoria en que se cometa: o/a estudante será cualificado con ?suspenso? (nota numérica 0) na convocatoria correspondente do curso académico, tanto se a comisión da falta se produce na primeira oportunidade como na segunda. Para isto, procederase a modificar a súa cualificación na acta de primeira oportunidade, se fose necesario.

Fontes de información

Bibliografía básica	- Tom White (2015). Hadoop: The Definitive Guide. O'Reilly (4ª ed.) - Wes McKinney (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly (2ª ed.)
Bibliografía complementaria	- Alex Holmes (2014). Hadoop in practice. Manning (2ª ed.)

Recomendacións

Materias que se recomenda ter cursado previamente

Materias que se recomenda cursar simultaneamente

Materias que continúan o temario

Observacións

Recomendacións para o estudo da materia Debido ao forte compoñente práctico é recomendable ir facendo as actividades prácticas e traballos academicamente dirixidos de forma regular ao longo do cuadrimestre. O coñecemento do inglés tanto falado como escrito é imprescindible dado que a bibliografía e as conferencias externas poden desenvolverse en inglés.

Observacións Farase un uso intensivo de ferramentas de comunicación online: videoconferencia, chat, etc. As sesións presenciais serán gravadas para ou revisión posterior. Ademais, farase uso da ferramenta Aula CESGA para a distribución de contidos, creación de foros de discusión, etc...

As ferramentas software utilizadas nesta materia son xeralmente open-source ou teñen licenza gratuita para estudantes.

Perspectiva de xénero-Segundo se recolle nas distintas normativas de aplicación para a docencia universitaria incorpórase a perspectiva de xénero nesta materia (usarase linguaxe non sexista, utilizarase bibliografía de autores/as de ambos sexos, propiciárase a intervención en clase de alumnos e alumnas...)-Traballárase para identificar e modificar prexuízos e actitudes sexistas e influirase na contorna para modificalos e fomentar valores de respecto e igualdade.-Deberanse detectar situacións de discriminación por razón de xénero e proporáanse accións e medidas para corrixilas.

(*)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