Teaching Guide				
Identifying Data				2023/24
Subject (*)	Data Engineering		Code	614544002
Study programme	Máster Universitario en Intelixencia Artificial			
Descriptors				
Cycle	Period	Year	Type	Credits
Official Master's Degree	1st four-month period	First	Obligatory	3
Language	English			
Teaching method	Face-to-face			
Prerequisites				
Department	Ciencias da Computación e Tecnoloxías da Información			
Coordinador	Bernardo Roca, Guillermo de	E-mail	guillermo.debernardo@udc.es	
Lecturers	Bernardo Roca, Guillermo de	E-mail	guillermo.debernardo@udc.es	
Web				
General description	The aim of this course is to introduce the basics data engineering, notably in the scope of Big Data. The acquired skills will allow the analysis and the efficient management of heterogeneous information, both structured and non structured, within the development of AI applications, whenever traditional methods show insufficiency.			

Study programme competences / results	
Code	Study programme competences / results
A17	CE16 - Knowledge of the process and tools for processing and preparing data, from their acquisition, extraction, and cleansing to their transformation, loading, organisation and access
B2	CG02 - Successfully addressing each and every stage of an AI project
B3	CG03 - Searching and selecting that useful information required to solve complex problems, with a confident handling of bibliographical sources in the field
B4	CG04 - Suitably elaborating written essays or motivated arguments, including some point of originality, writing plans, work projects, scientific papers and formulating reasonable hypotheses in the field
B5	CG05 - Working in teams, especially of multidisciplinary nature, and being skilled in the management of time, people and decision making
B6	CB01 - Acquiring and understanding knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, frequently in a research context
B7	CB02 - The students will be able to apply the acquired knowledge and to use their capacity of solving problems in new or poorly explored environments inside wider (or multidisciplinary) contexts related to their field of study
B8	CB03 - The students will be able to integrate different pieces of knowledge, to face the complexity of formulating opinions (from information that may be incomplete or limited) and to include considerations about social and ethical responsibilities linked to the application of their knowledge and opinions
C3	CT03 - Use of the basic tools of Information and Communications Technology (ICT) required for the student's professional practice and learning along her life
C7	CT07 - Developing the ability to work in interdisciplinary or cross-disciplinary teams to provide proposal that contribute to a sustainable environmental, economic, political and social development
C8	CT08 - Appreciating the importance of research, innovation and technological development in the socioeconomic and cultural progress of society
C9	CT09 - Being able to manage time and resources: outlining plans, prioritising activities, identifying criticisms, fixing deadlines and sticking to them

Learning outcomes			
Learning outcomes	Study programme competences / results		
Develop the capacity to analyse and model data for processing in intelligent systems.	AC16	BC6 BC7	CC3 CC9

Know and understand the process of extraction, cleaning, transformation, load and preprocessing of data.	AC16	BC2 BC3 BC8	CC3 CC9
Know and learn how to use multidimensional and NoSQL databases.		BC3 BC4 BC7	CC8
Know the foundations of data lakes and data warehouses.		BC2 BC5 BC7 BC8	CC3 CC7 CC8

Contents	
Topic	**Sub-topic**
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.

Planning				
Methodologies / tests	**Competencies / Results**	**Teaching hours (in-person & virtual)**	**Student?s personal work hours**	**Total hours**
Guest lecture / keynote speech	B4 B5 C3 C9	12	0	12
Practical test:	A17 B2 B5 B7 C3	8	0	8
Problem solving	A17 B2 B4 B7 C7 C9	0	50	50
Supervised projects	A17 B2 B3 B6 B7 B8 C7 C8	5	0	5
Personalized attention		0		0
(*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.				

Methodologies	
Methodologies	**Description**
Guest lecture / keynote speech	The teacher will introduce given subjects to the students with the aim to acquire information valuable within a specific scope.
Practical test:	Problem or problems of practical character whose resolution requires the understanding and application of the theoretical and practical contents covered by the course. The students can work the solution to the proposed problems individually or in groups.
Problem solving	A project whose scope and aims require that the students work autonomously, although under the supervision of the teachers.
Supervised projects	Practical projects whose scope requires a significant fraction of the total dedication of the student to the course. Besides, the scale of these projects requires that the students apply management skills in addition to technical skills.

Personalized attention	
Methodologies	**Description**

Supervised projects Problem solving	Projects: Real or fictitious scenarios are presented to the students to introduce a given problem. The students have to apply the theoretical and practical knowledge acquired in this course to look for a solution to the question or questions posed. Usually, the study of cases will addressed in groups. The groups will present and discuss their solutions. Problem solving: The teacher will supervise to the progress of the projects via individual sessions.

Assessment			
Methodologies	Competencies / Results	Description	Qualification
Supervised projects	A17 B2 B3 B6 B7 B8 C7 C8	Defense of the solution proposed by the student or oral presentation of the developed solution.	30
Practical test:	A17 B2 B5 B7 C3	Several assessment tests will be conducted in order to evaluate the understanding of the knowledge exposed in the classes of theory and/or practical. These tests can not be repeated in the second evaluation call.	30
Problem solving	A17 B2 B4 B7 C7 C9	The evaluation of the autonomous work will include the submission of a report and a defense in which the students explain their developments and conclusions in front of the teacher and the classroom.	40

Assessment comments
FIRST AND SECOND EVALUATION CALLS [Assisting and Non-assisting students]Final grade = 0,30 * Project based learning + 0,30 * Laboratory practical tests + 0,40 * Autonomous problem solvingNon-assisting students will complete the same assignments and tests than assisting students.FINAL GRADESTo pass the course in any of the evaluation calls, the final grade must be equal or greater than 5 (from a total of 10), obtaining a minimum score of 5 (out of 10) in each of the evaluation parts.In the second opportunity the laboratory practical tests cannot be repeated, so there is no minimum score in this part.ADDITIONAL REMARKSIf plagiarism is detected in any of the works (essays or project), the final grade will be "Suspenso" (0) and the situation will be notified to the School's Board to take the appropriate disciplinary actions. If translation errors cause any contradictions between the various versions of this syllabus, the English will be the prevailing version.

Sources of information	
Basic	- Sadalage, Fowler (2012). NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley - Avi Silberschatz, Henry F. Korth, S. Sudarshan (2010). Database System Concepts. McGraw-Hill - Ihab F. Ilyas, Xu Chu, (2019). Data Cleaning. Association for Computing Machinery. ACM - Alex Gorelik (). The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science. O'Reilly
Complementary	- Matt Casters, Roland Bouman, Jos van Dongen (2013). Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration. Wiley

Recommendations
Subjects that it is recommended to have taken before
Subjects that are recommended to be taken simultaneously
Subjects that continue the syllabus
Other comments

Follow the proposed methodology, attending classes, devoting the necessary time to study and carrying out assignments and solving specific problems with the help of teachers in tutorial sessions

(*)The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.