COURSE DATA

Data Subject	
Code	36424
Name	Data management
Cycle	Grade
ECTS Credits	6.0
Academic year	2023 - 2024

Study (s)

Degree	Center	Acad. year	Period
1406 - Degree in Data Science	School of Engineering	2	First term

Subject-matter

Degree	Subject-matter	Character
1406 - Degree in Data Science	8 - Information Management	Obligatory

Coordination

Name	Department
CERVERON LLEO, VICENTE	240 - Computer Science
FUERTES SEDER, ARIADNA	240 - Computer Science
VES CUENCA, ESTHER DE	240 - Computer Science

SUMMARY

Data Management is a component of the Information Management track. Comprising 6 ECTS credits, it is taught in the first semester of the second year of the Degree in Data Science.

The volume of data generated, stored and processed today is growing exponentially. Although relational databases have demonstrated their ability to adapt to growing needs, traditional database architectures and models can sometimes be insufficient or inadequate for addressing the volume and type of data to be managed.

This course deals with new architectures for parallel and distributed databases, new models of databases generally known as NoSQL, and tools and techniques for data warehousing and multidimensional data models. It also presents an overview of data integration and migration processes.

Students will acquire the ability to analyze the advantages and disadvantages of various technologies in order to select the most appropriate ones in each context.

The theoretical classes will be taught in Spanish. The language for the practical and laboratory classes will be specified in the course guidelines available on the website for this degree.

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

There are no specified enrollment restrictions with other subjects of the curriculum.

Other requirements

Students are recommended to have passed the course on Fundamentals of Databases before taking this course.

OUTCOMES

1406 - Degree in Data Science

- (CG03) Capability to elaborate models, calculations, reports, to plan tasks and other works analogous to the specific field of data science.
- (CG06) Ability to access and manage information in different formats for subsequent analysis in order to obtain knowledge from data.
- (CT02) To be able to complete technical, scientific, social and human training in general, and to organise self-learning with a high degree of autonomy.
- (CT05) Ability to evaluate the advantages and disadvantages of different methodological and / or technological alternatives in different fields of application.
- (CE02) To methodologically know and apply the programming techniques and the algorithms necessary for the efficient processing of information and the computer resolution of problems that use large volumes of data.
- (CE04) To know and use the different models of data storage and database management systems using programming languages for the definition, query and handling of data.
- (CE08) Ability to understand, select and use the infrastructure and the techniques used to handle mass data, according to criteria of efficiency, scalability, security, error tolerance and adaptation to the production environment.
- (CE11) Ability to design and implement data acquisition, its integration, transformation, selection, verification of its quality and veracity from different sources, taking into account its character, heterogeneity and variability.

- (CB4) Students must be able to communicate information, ideas, problems and solutions to both expert and lay audiences.

LEARNING OUTCOMES

Understand the motivation of parallel and distributed database management systems (CT02).

Have the fundamental knowledge required to analyze different architectures and their implications for the performance, speed, scalability and security of data-partitioning strategies and to evaluate and optimize query processing on centralized and distributed data (CB4, CG03, CE08).

Understand what a data warehouse is and the reason for its importance as a support to decision making (CT05, CE02).

Know the multidimensional data model and the type of data analysis it facilitates. (CE02)

Know different database models and technologies known generically as NoSQL systems and have the fundamental knowledge required to define, design and implement information systems using these technologies. (CT05, CE02, CE04, CE08)

Know how to perform a process of data migration and integration from various sources. (CG06, CE11)

DESCRIPTION OF CONTENTS

1. Parallel and distributed databases

Database-system architectures. Parallel databases. Parallel query processing. Distributed databases. Distributed storage, query processing and recovery.

2. Query processing and optimization

Evaluation and measures of query cost: selection, join and other operations. Query optimization. Query evaluation in parallel and distributed databases.

3. Data warehousing and multidimensional model

Data warehousing: models and operations. Multidimensional model: OLAP models, SQL:1999.

4. NoSQL databases

Data models for new requirements: Big Data, semi-structured and unstructured data, distributed architectures. Column-oriented databases. Document-oriented databases. Graph-oriented databases.

5. Integration of data and data migration processes.

Integration of data from various sources. The ETL process: extraction, transformation and loading. Aspects to consider in data migration process.

WORKLOAD

ACTIVITY	Hours	% To be attended
Theory classes	30,00	100
Laboratory practices	20,00	100
Classroom practices	10,00	100
Development of individual work	15,00	0
Study and independent work	10,00	0
Readings supplementary material	15,00	0
Preparation of evaluation activities	10,00	0
Preparing lectures	10,00	0
Preparation of practical classes and problem	30,00	0
TOTAL	**150,00**	

TEACHING METHODOLOGY

MD1 - Theoretical activities. Expository development of the subject in which students will be encouraged to participate in solving specific issues by completing individual evaluation questionnaires (CG03, CT02, CT05, CE04, CE08).

MD2 - Practical activities. Students will learn by solving problems, completing exercises and analyzing case studies to acquire competences in the various aspects of the subject (CB4, CE08, CE11).

MD4 – Work in the laboratory and/or computer classroom. Students will learn by carrying out activities individually or in small groups in the laboratory and/or computer room (CG06, CE02, CE04, CE08).

The University of Valencia's e-learning platform (*Aula Virtual*) will be used to communicate with the students. Students will also have access to the learning materials used in class and the problems and exercises they need to solve via this e-learning platform.

EVALUATION

SE1 - Objective test. This will consist of one or more exams comprising theoretical and practical issues and problems. The score obtained on this component will account for 50% of the final grade for the first examination sitting. Students will need to obtain a minimum score of 5 points out of 10 on this component to pass the course (CG03, CT02, CT05, CE04, CE08)

SE2 - Evaluation of practical activities based on the student's preparation of papers/reports and/or oral presentations. These practical activities will be conducted in the computer laboratory. Attendance is compulsory unless absence is properly justified. The score obtained on this component will account for 30% of the final grade (CB4, CE08, CE11).

SE3 - Continuous assessment based on the student's participation and degree of involvement in the teaching-learning process and taking into account his/her attendance at face-to-face activities, solutions to issues and problems set periodically, and presentation and exposition of assignments. The score obtained on this component will account for 20% of the final grade (CG06, CE02, CE04, CE08).

The activities on components SE2 and SE3 cannot be retaken.

The second examination sitting will comprise an exam that will account for 70% of the final grade. A minimum of 5 points out of 10 will be needed to pass the exam, and the grade obtained during the academic year in block SE2 will account for the remaining 30%.

In all cases the evaluation system will be governed by the University of Valencia's regulations on grading and assessment for bachelor's degrees and master's degrees, which is available at:

http://www.uv.es/graus/normatives/2017_108_Reglament_avaluacio_qualificacio.pdf

REFERENCES

Basic

- Ramakrishnan, R.; .Gehrke, J. Sistemas de gestión de bases de datos / Database Management Systems. McGraw-Hill
- Elmasri, R.; Navathe S. B. Fundamentos de bases de datos / Fundamentals of Database System. Pearson
- Silberschatz, A.; Korth, H. F.; Sudarshan, S. Fundamentos de bases de datos / Database System Concepts. McGraw-Hill

Additional

- Inmon, W. H.; Linstedt, Daniel. Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault. Morgan Kaufmann Publishers Inc
- Adamson, C.; Kimball, R. Mastering data warehouse aggregates. Willey
- Strauch, Christof. NoSQL databases. Stuttgart Media University. http://www.christof-strauch.de/nosqldbs.pdf
- Robinson, Ian; Webber, Jim; Eifrem, Emil. Graph Databases. OReilly Media