DESIGNING AND USING DATABASES

Grado en Computación e Inteligencia Artificial / Bachelor in Computer Science and Artificial Intelligence BCSAI SEP-2025 DUD-CSAI.2.M.A

Area Computer Science

Number of sessions: 30

Academic year: 25-26

Degree course: SECOND

Number of credits: 6.0

Semester: 2º

Category: COMPULSORY

Language: English

Professor: **MIKE BEATTIE**

E-mail: mbeattie@faculty.ie.edu

Mike is a technology specialist and entrepreneur in the technology sector, having worked in various start ups in the mobile payments, audio visual software and personal fitness spaces. He is also a professional coach for startups and growing tech companies. He is married, has 3 grown up children, lives in Madrid and plays lead saxophone in the Clandestina Big Band.

Office Hours

Office hours will be on request. Please contact at:

SUBJECT DESCRIPTION

Welcome to the Designing and Using Databases course. Databases are central to almost all technology projects, and this course aims to provide sufficient knowledge to be able to work with database designers and programmers in a corporate environment. It will mean better understanding when making decisions about the type of system and in designing a database solution that will fit with the needs of clients. It will introduce the different types of database that are in use in organizations today, from traditional systems that provide business intelligence to advanced systems that drive web-scale applications.
The course will lay the foundation for working with structured databases and will also look at ways organizations use databases, such as in Business Intelligence and enterprise systems. There will be a strong practical focus in the course which will involve producing the necessary documentation, creating diagrams and implementing a database using SQL, NoSQL and Big Data technologies. The course will give you, as future leaders, the skills necessary to implement new, innovative systems in the ever-changing technologically advanced world in which we live and work.

LEARNING OBJECTIVES

The primary objective of the course is to each you the essential concepts and skills to design, implement, and use database projects and systems. There will be two central themes of the course: the technical design and implementation of databases and the organizational impact of the systems.
In the design part of the course, there will be a focus on understanding the system requirements, doing cost scenarios and analyzing a system using Entity-Relationship diagrams and Data Flow Diagrams.
In the implementation part of the course, we will look at the uses of databases in corporate environments.
At the end of the course you will be able to:
- understand data structures behind database systems
- produce a database design document
- implement databases using SQL, NoSQL and Spark
- assess security and implementation issues
- understand how web developers use databases
- create simple database-back web projects
- understand the importance of Business Intelligence platforms
- choose a database system appropriate for different types of projects
- work with teams of programmers and designers and participate in the decision-making process

TEACHING METHODOLOGY

IE University teaching method is defined by its collaborative, active, and applied nature. Students actively participate in the whole process to build their knowledge and sharpen their skills. Professor's main role is to lead and guide students to achieve the learning objectives of the course. This is done by engaging in a diverse range of teaching techniques and different types of learning activities such as the following:

Learning Activity	Weighting	Estimated time a student should dedicate to prepare for and participate in
Lectures	28.0 %	42.0 hours
Discussions	17.3 %	26.0 hours
Exercises in class, Asynchronous sessions, Field Work	17.3 %	26.0 hours
Group work	25.3 %	38.0 hours
Individual studying	12.0 %	18.0 hours
TOTAL	100.0 %	150.0 hours

AI POLICY

In this course, the use of generative artificial intelligence (GenAI) is encouraged, with the goal of developing an informed critical perspective on potential uses and generated outputs.

However, be aware of the limits of GenAI in its current state of development:

·If you provide minimum effort prompts, you will get low quality results. You will need to refine your prompts to get good outcomes. This will take work.

·Don't take ChatGPT's or any GenAI's output at face value. Assume it is wrong unless you either know the answer or can cross-check it with another source. You are responsible for any errors or omissions. You will be able to validate the outputs of GenAI for topics you understand.

·AI is a tool, but one that you need to acknowledge using. Failure to do so is in violation of academic honesty policies. Acknowledging the use of AI will not impact your grade.

Suggested format to acknowledge the use of generative AI tools:

I acknowledge the use of [AI systems link] to [specify how you used generative AI]. The prompts used include [list of prompts]. The output of these prompts was used to [explain how you used the outputs in your work].

If you have chosen not to include any AI generated content in your assignment, the following disclosure is recommended:

No content generated by AI technologies has been used in this assignment.

PROGRAM

SESSION 1 (LIVE IN-PERSON)

- Course Introduction
- Learning and having fun
- Why Databases are so important
- Where we see databases today
- Trade offs

SESSION 2 (LIVE IN-PERSON)

Data Models and Query Languages

- Relational Model vs Document Model
- Query Languages for Data
- Graph-like Data Models
- Data Definition Language (DDL)

SESSION 3 (LIVE IN-PERSON)

Database design I

Key elements of data structures for designers :

- Hash Indexes
- SSTables
- LSM-Trees
- B-Trees

SESSION 4 (LIVE IN-PERSON)

Database Design II

Next level efficiencies :

- Normalisation
- Many-to-many Relationships

- Database Engines
- Partitioning and Indexes
- Replication

SESSION 5 (LIVE IN-PERSON)

Database Design III

- The final steps
- Logical and Physical Database Design
- Data Flow Diagrams (DFDs)
- Entity Relationship (ER) Diagrams

SQL Tutorial I: The SELECT Statement

SESSION 6 (LIVE IN-PERSON)

SQL Tutorial II

Filtering and Categorization:

- WHERE and
- CASE clauses

SESSION 7 (LIVE IN-PERSON)

Relational Data Models

- Normalisation
- Many-to-many Relationships
- Database Engines
- Partitioning and Indexes

SESSION 8 (LIVE IN-PERSON)

SQL Joins and Set Theory

- SQL Joins
- Review of Set Theory and De Morgan's Laws
- Logical Equivalence

SESSION 9 (LIVE IN-PERSON)

SQL Tutorial III

- Aggregating results
- Window Functions
- Common Table Expressions

SESSION 10 (LIVE IN-PERSON)

SQL Tutorial IV

- Date And Time Functions

- Time-series analysis in SQL
- Timeseries Databases

SESSION 11 (LIVE IN-PERSON)

Advanced SQL Topics

- Exploratory Data Analysis
-

  Machine Learning datasets

  Loading from a Database to an application

SESSION 12 (LIVE IN-PERSON)

Data outside the database

- Storing and Modifying data
- Bulk Data Load operations and tools
- SQLDump, data file types and encodings

SESSION 13 (LIVE IN-PERSON)

Data Warehouse and Data Mart Design

- Facts and Dimensions
- Star vs Snowflake schema design
- Review of Enterprise Data Architectures and Patterns
- Role of Databases in Data Engineering pipelines

SESSION 14 (LIVE IN-PERSON)

Database Applications I : Web Applications

- Building dynamic, database-driven web sites
- Review of common stacks: LAMP and Flask-based

SESSION 15 (LIVE IN-PERSON)

MIDTERM EXAM

SESSION 16 (LIVE IN-PERSON)

Database Applications II: Business Intelligence

- The role of Business Intelligence in Enterprise IT
- Demo of Business Intelligence platforms
- Building a simple data application with Dash

SESSION 17 (LIVE IN-PERSON)

Database Applications III: Business Intelligence (continued)

- Building a simple data application with Dash

- Designing data : top down and bottom up

SESSION 18 (LIVE IN-PERSON)

Introduction to NoSQL

Sample implementation of most relevant NoSQL data structures.

SESSION 19 (LIVE IN-PERSON)

Document Databases

- Operations on document collections
- Partitioning and sharding
- Query processing using Javascript
- Basics of MongoDB and PyMongo

SESSION 20 (LIVE IN-PERSON)

MongoDB Lab 1

Answering questions on real-world datasets using MongoDB Query Language (MQL)

SESSION 21 (LIVE IN-PERSON)

Designing Web-Apps with MongoDB

- Prototyping Web Applications using Node.js and MongoDB
- What is the MEAN stack?

SESSION 22 (LIVE IN-PERSON)

Introduction to Big Data & Hadoop Ecosystem

- Introduction to Big Data Technologies and Frameworks
- Distributed Data Processing and Massive Parallel Processing Paradigms
- Hadoop and Map-Reduce

SESSION 23 (LIVE IN-PERSON)

Spark Basics

- Spark Architecture and concept
- Spark Cluster configuration parameters and tuning
- A demonstration of basic Spark usage to manipulate data

SESSION 24 (LIVE IN-PERSON)

Spark Batch Processing Functions and Lab I

SESSION 25 (LIVE IN-PERSON)

Spark Batch Processing Functions and Lab II

SESSION 26 (LIVE IN-PERSON)

SparkSQL and Spark ML

Latest trends
Top tips

SESSION 27 (LIVE IN-PERSON)

Distributed Databases: Cassandra

Cassandra Data Model
Cassandra Distributed Architecture
Nodes, Rings, Tokens, Replication
CQL (Cassadra Query Language)
Data Types and Collections
Secondary Indexes
Materialised Probabilistic Data Structures
Bloom filters

SESSION 28 (LIVE IN-PERSON)

Cassandra Lab

SESSION 29 (LIVE IN-PERSON)

Final Project Presentations

SESSION 30 (LIVE IN-PERSON)

Final Exam

EVALUATION CRITERIA

criteria	percentage	Learning Objectives	Comments
Final Exam	30 %		
Group Presentation	10 %		
Individual work	20 %		
Group Work	10 %		
Class Participation	10 %		
Intermediate tests	20 %		

RE-SIT / RE-TAKE POLICY

A minimum passing grade of 3.5 (35%) in the Midterm and Final exam is required to pass the subject. If a student scores lower than this minimum, he will have to go to June retake, irrespective of their overall course grade. The overall passing course grade is 5.0.

Each student has four chances to pass any given course distributed over two consecutive academic years: ordinary call exams and extraordinary call exams (re-sits) in June/July.

Students who do not comply with the 80% attendance rule during the semester will fail both calls for this Academic Year (ordinary and extraordinary) and have to re-take the course (i.e., re-enroll) in the next Academic Year.

Evaluation criteria:

•Students failing the course in the ordinary call (during the semester) will have to re-sit the exam in June / July (except those not complying with the attendance rule, who will not have that opportunity and must directly re-enroll in the course on the next Academic Year).

• The extraordinary call exams in June / July (re-sits) require your physical presence at the campus you are enrolled in (Segovia or Madrid). There is no possibility to change the date, location or format of any exam, under any circumstances. Dates and location of the June / July re-sit exams will be posted in advance. Please take this into consideration when planning your summer.

• The June / July re-sit exam will consist of a comprehensive exam. Your final grade for the course will depend on the performance in this exam only; continuous evaluation over the semester will not be taken into consideration. Students will have to achieve the minimum passing grade of 5 and can obtain a maximum grade of 8.0 (out of 10.0) – i.e., “notable” in the in the re-sit exam.

• Retakers: Students who failed the subject on a previous Academic Year and are now re-enrolled as re-takers in a course will be needed to check the syllabus of the assigned professor, as well as contact the professor individually, regarding the specific evaluation criteria for them as retakers in the course during that semester (ordinary call of that Academic Year). The maximum grade that may be obtained in the retake exam (3rd call) is 10.0.

After ordinary and extraordinary call exams are graded by the professor, you will have a possibility to attend a review session for that exam and course grade. Please be available to attend the session in order to clarify any concerns you might have regarding your exam. Your professor will inform you about the time and place of the review session. Any grade appeals require that the student attended the review session prior to appealing.

Students failing more than 18 ECTS credits in the academic year after the June-July re-sits will be asked to leave the Program. Please, make sure to prepare yourself well for the exams in order to pass your failed subjects.

In case you decide to skip the opportunity to re-sit for an exam during the June / July extraordinary call, you will need to enroll in that course again for the next Academic Year as a re-taker and pay the corresponding extra cost. As you know, students have a total of four allowed calls to pass a given subject or course, in order to remain in the program.

BIBLIOGRAPHY

Compulsory

- Alex Petrov. *Database Internals.* O'Reilly. ISBN 978149204034 (Digital)
- Guy Harrison. *Next Generation Databases: NoSQL, NewSQL, and Big Data.* Appress. ISBN 9781484213292 (Digital)

BEHAVIOR RULES

Please, check the University's Code of Conduct here. The Program Director may provide further indications.

ATTENDANCE POLICY

Please, check the University's Attendance Policy here. The Program Director may provide further indications.

ETHICAL POLICY

Please, check the University's Ethics Code here. The Program Director may provide further indications.