ADVANCED DATABASES

Dual Degree in Business Administration & Data and Business Analytics BBADBA SEP-2024 ADDB-NBDA.3.M.A

Area Data Science
Number of sessions: 30
Academic year: 24-25
Degree course: THIRD
Number of credits: 6.0
Semester: 2º
Category: COMPULSORY
Language: English

Professor: **LUIS ANGEL GALINDO SANCHEZ**

E-mail: lgalindo@faculty.ie.edu

- PhD Telecommunication Engineer (2012). Polytechnic University of Madrid. ANECA accredited.
- Executive MBA (2006). IESE Business School – University of Navarra.
- Master, Service and Security in IP Networks (2005). Polytechnic University of Madrid.
- Specialist, Network and Advanced Internet Systems (1998). Carlos III University of Madrid.
- Specialist, Voice Codification (1997). Carlos III University of Madrid.
- Adjuct Professor at IE teaching ADSS, Technology Scouting, Cybercrime and IE Impact/Technology.
- Visiting Associate Professor at HEC Montreal (from 2018 till now), where I teach Enterpreneurship with my own methodology to assess the business potential of a value proposal.
- Professor in the Master of Cybercrime Master at Nebrija University (from 2018 till now).
- Responsible of the AI area applied to the detection of cybercriminals. Teaching police officers how to use Big Data and Machine Learning Technologies to be more efficient working with deep and dark web at Nebrija University.
- Associate Professor at Carlos III University of Madrid (from 2007 till now). Professor of Computer networks.

    Defined, developed and implemented a successful open innovation model, generating new relevant revenue streams results. Selective results include an increase of +200 startups working under this model with +6000 companies analyzed. Increase in revenue of +10% y-o-y leveraged on the innovative assets.

He has defined, developed and implemented a successful Digital Transformation Learning Program for more than 8 years and +2000 people trained, receiving Telefonica Excellent Teacher award along 7 consecutive years.

Principal Investigator in 7 EU funded projects ((previously in 4 more) based on the technological skills that Luis has in fields like Cybersecurity, Cyber Defense, AI, Cloud Computing or IoT. He personally researches in the application of AI to the cybersecurity and cybercrime.

Teacher in several recognised universities and business schools with excellent marks from 2007.

Luis, as High Performance Senior Consultant, has implemented +20 consultancy complex projects worldwide for large enterprises in AI, economical intelligence, cybersecurity, business development, digital transformation, risk assessment, process optimization and people management issues.

Luis is an expert at understanding customers and translating their needs into actual sales by developing new markets or new products & services and Digital mindset within everyone in the organization.

He has had a leading role in complex M&A processes with digital assess acquisition like in the first world integration of a MVNO and a social network.

Frequent speaker at Innovation and Tech international conferences. He considers himself a creovator.

Office Hours

Office hours will be on request. Please contact at:

lgalindo@faculty.ie.edu (email me in advance, if you wish to schedule a chat or a meeting).

SUBJECT DESCRIPTION

The motivation of this advanced databases course is that databases can and do realize their full potential only if they are interconnected and if they are used not only for data retrieval, but also for making their implicit knowledge explicit. In addition, the meaning of (stored, combined, and inferred) data needs to be understood well. Therefore, this course extends the foundations laid in Databases I (relational databases as a "homogenizing model") by focusing on heterogeneity and interoperability: based on an understanding of data semantics, it looks at prominent types of heterogeneous types of data, heterogeneous data bases / knowledge bases and their combination (in particular, over the Internet/Web), and the discovery of knowledge from data (bases). The course finishes by examining significant real-world applications and implications of these concepts.

The concept is to focus on breadth / overview; specializations in selected topics are probably better suited for seminar-style follow-up courses. (However, this point is open for discussion; the course can also be adjusted to focus more on depth in specific areas.)

The main topics covered in this course are the following:

- Advanced databases: Introduction
- Relational languages
- Database design
- Object-based databases
- Querying
- Transaction management

- Big Data Analytics
- Parallel and distributed databases
- Database system architecture
- Advanced topics

LEARNING OBJECTIVES

Students will understand and master relevant concepts and techniques of current databases and processing based on databases. They will understand the potentials, limitations, and risks inherentin assembling, combining, and processing huge amounts of heterogeneous data in globally interconnected environments. They will be able to design such databases and connectivity and relevant methods for combining and enriching data, and work with concrete examples of such data collection/processing.

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	26.7 %	40.0 hours
Discussions	13.3 %	20.0 hours
Exercises in class, Asynchronous sessions, Field Work	20.0 %	30.0 hours
Group work	20.0 %	30.0 hours
Individual studying	20.0 %	30.0 hours
TOTAL	100.0 %	150.0 hours

AI POLICY

Generative artificial intelligence (GenAI) tools may be used in this course for research, ideation, generating an outline, proofreading, grammar check, coding, image generation with appropriate acknowledgement. GenAI may not be used for exams. If a student is found to have used AI-generated content inappropriately, it will be considered academic misconduct, and the student might fail the respective assignment or the course.

If you are in doubt as to whether you are using GenAI tools appropriately in this course, I encourage you to discuss your situation with me.

In case you use GenAI for assignments, what is expected is to provide critical thinking and creativity on top of the results of the tool.

Below, a suggested format to acknowledge the use of generative AI tools. Please note that acknowledging AI will not impact your grade.

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

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

PROGRAM

SESSION 1 (LIVE IN-PERSON)

WELCOME AND APPLICATION DEVELOPMENT

Introduction to the course: logistics, evaluation system, and content at-a-glance.

Application Programs and User Interfaces, Web Fundamentals, Servlets and JSP, Application Architectures, Rapid Application Development, Application Performance, Application Security, Encryption and its Applications

SESSION 2 (LIVE IN-PERSON)

DATABASE SYSTEM ARCHITECTURES

Centralized Database Systems, Server System Architectures, Parallel Systems, Distributed Systems and Network Types.

SESSION 3 (LIVE IN-PERSON)

SQL 1

In this session we will cover intermediate SQL concepts like: Join Expressions, Views, Transactions, Integrity Constraints, SQL Data Types and Schemas, Index Definition in SQL or Authorization.

SESSION 4 (LIVE IN-PERSON)

SQL 2

In this session we will cover advanced SQL concepts like: Accessing SQL From a Programming Language, Functions and Procedures, Triggers, Recursive Queries or Advanced Aggregation Features

SESSION 5 (LIVE IN-PERSON)

CLASS EXERCISE

In this class exercise, based on the Questions that needs to be answered, we will look for different data sets available, which later will be introduced in the DB to make the queries.

SESSION 6 (LIVE IN-PERSON)

COMPLEX DATA TYPES

Semi-Structured, Object Orientation, Textual Data, Spatial Data

SESSION 7 (LIVE IN-PERSON)

DATABASE DESIGN AND E-R MODEL

The Design Process, The Entity-Relationship Model, Complex Attributes, Mapping Cardinalities, Primary Key, Removing Redundant Attributes in Entity Sets, Reducing ER Diagrams to Relational Schemas, Extended E-R Features, Entity-Relationship Design Issues, Alternative Notations for Modeling Data.

Book Chapters: Avi Silberschatz, Henry F. Korth, S. Sudarshan. Database System Concepts.

McGraw-Hill (See Bibliography)

SESSION 8 (LIVE IN-PERSON)

CLASS EXERCISE

In this class exercise, we will work in the design of our DB and the E-R model that later we will implement in a real DB

SESSION 9 (LIVE IN-PERSON)

TEAM PRESENTATION 1

10% of your grade

Team project: you will do some research, prepare and show a group presentation about a course topic of your choice (among a set of proposed options related to DB). Later, there will be Q&A.

SESSION 10 (LIVE IN-PERSON)

NORMALIZATION

Features of Good Relational Design, Functional Dependencies, Decomposition Using Functional Dependencies, Normal Forms, Functional Dependency Theory, Algorithms for Decomposition using Functional Dependencies, Decomposition Using Multivalued Dependencies, More Normal Form, Atomic Domains and First Normal Form, Modeling Temporal Data

SESSION 11 (LIVE IN-PERSON)

CLASS EXERCISES

In this class exercise, we will work with open-source relational DB. By using Postgre SQL, you will implement the previous E-R model defined and the queries to answer the requested questions.

SESSION 12 (LIVE IN-PERSON)

IN-CLASS QUIZ 1

5% of your grade

This will be a warm-up individual exam on everything we have covered so far (sessions 1 through 11) that will help you get ready for the midterm and final exams.

SESSION 13 (LIVE IN-PERSON)

QUERYING

Measures of Query Cost, Selection Operation, Sorting, Join Operation, Evaluation of Expressions, Statistical Information for Cost Estimation, Cost-based optimization.

SESSION 14 (LIVE IN-PERSON)

TRANSACTIONS

Transaction Concept, Transaction State, Concurrent Executions, Serializability, Recoverability, Implementation of Isolation

SESSION 15 (LIVE IN-PERSON)

RECAP FIRST HALF

In this session, we will run a recap discussion (Q&A) about the first half of the course to be ready for the mid-term exam

SESSION 16 (LIVE IN-PERSON)

MIDTERM EXAM

20% of your grade.

SESSION 17 (LIVE IN-PERSON)

CONCURRENCY CONTROL

Lock-Based Protocols, Timestamp-Based Protocols, Validation-Based Protocols, Multiple Granularity, Multiversion Schemes, Insert and Delete Operations, Concurrency in Index Structures

SESSION 18 (LIVE IN-PERSON)

CLASS EXERCISES

In this class exercise, we will work with NoSQL DBs by using MongoDB. You will have to search some data files and load in your MongoDB database.

SESSION 19 (LIVE IN-PERSON)

RECOVERY SYSTEM

Failure Classification, Storage Structure, Recovery and Atomicity, Log-Based Recovery, Remote Backup Systems

SESSION 20 (LIVE IN-PERSON)

CLASS EXERCISES

In this class exercise, you will have to search some data files, run an ETL process to load them in your MongoDB database to get some insights from them.

SESSION 21 (LIVE IN-PERSON)

PARALELL AND DISTRIBUTED DATABASES

Introduction to Paralell processing, Parallel Sort, Parallel Join, Parallel Evaluation of Query Plans, Distributed Query Processing, Parallel and Distributed Transaction Processing

SESSION 22 (LIVE IN-PERSON)

CLASS EXERCISES

In this class exercise, you will have to search some data files, run an ETL process to load them in your MongoDB database and run a query optimization process to improve the performance of the queries when queryieng large datasets.

SESSION 23 (LIVE IN-PERSON)

BLOCKCHAIN DATABASES

Blockchain Properties, Achieving Blockchain Properties via Cryptographic Hash Functions, Consensus, Data Management in a Blockchain, Smart Contracts, Performance Enhancement, Emerging Applications.

SESSION 24 (LIVE IN-PERSON)

ADVANCED INDEXING

Advanced Indexing techniques.

SESSION 25 (LIVE IN-PERSON)

TEAM PRESENTATION 2

10% of your grade

Team project: you will do some research, prepare and show a group presentation about a course topic of your choice (among a set of proposed options on Blockchain databases). Later, there will be Q&A.

SESSION 26 (LIVE IN-PERSON)

IN-CLASS QUIZ 2

5% of your grade

This will be a small individual exam on everything we have covered since the midterm (sessions 17 through 25) that will help you get ready for the final exam.

SESSION 27 (LIVE IN-PERSON)

A VIEW TO THE WHOLE PROCESS OF DB AND FUTURE TRENDS

In this session, we will review the whole process of designing, deploying and Mantaining a DB and we will take a look to the future trends in Data.

SESSION 28 (LIVE IN-PERSON)

RECAP SECOND HALF

In this session, we will run a recap discussion (Q&A) about the second half of the course to be ready for the final exam

SESSION 29 (LIVE IN-PERSON)

MOCKUP EXAM

In this session we will cover a session for solving doubts in a practical way of Q&A

SESSION 30 (LIVE IN-PERSON)

FINAL EXAM

30% of your grade

EVALUATION CRITERIA

Class Participation (20%)

This includes optional exercises, voluntary participation on the whiteboard, discussion board (forum) activity, class attendance, presentations in recap session, and active participation in in-class discussions, with the goal of ensuring a continued learning process, good teamworking, and ability to apply class concepts in real-world problems. Participation is based on the quality, rather on the quantity, of your contributions.

Workgroups (20%)

Throughout the course, a number of mandatory assignments will be given that are to be completed in groups of 4-5 people.

In-class quizzes (10%)

Students will have to individually solve two in-class quizzes in sessions 13 and 26, each of which will cover the previously covered material.

Midterm Exam (20%)

The midterm exam will comprise all material given in class, up to sessions 16.

Final Exam (30%)

The final exam will include all the material explained in the course, and will be held during session 30.

Minimal Marks:

A minimum passing grade in the final exam (3.5) 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.

All the presentations/videos/exams will be submitted via Campus Online. No other option will be accepted.

criteria	percentage	Learning Objectives	Comments
Final Exam	30 %		
Midterm Exam	20 %		
Group Work	20 %		
Class Participation	20 %		This includes optional exercises, discussion board activity, and class attendance and participation.
In-class Quizzes	10 %		

RE-SIT / RE-TAKE POLICY

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 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

 - Avi Silberschatz, Henry F. Korth, S. Sudarshan. *Database System Concepts.* McGraw-Hill. ISBN 9780078022159 (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.