



## Guia docent

# 270951 - ADSDB - Algorismes, Estructures de Dades i Bases de Dades

Última modificació: 23/11/2023

**Unitat responsable:** Facultat d'Informàtica de Barcelona

**Unitat que imparteix:** 723 - CS - Departament de Ciències de la Computació.

**Titulació:** MÀSTER UNIVERSITARI EN CIÈNCIA DE DADES (Pla 2021). (Assignatura obligatòria).

**Curs:** 2023

**Crèdits ECTS:** 6.0

**Idiomes:** Anglès

### PROFESSORAT

---

**Professorat responsable:** MARIA JOSEFINA SIERRA SANTIBAÑEZ - ANNA QUERALT CALAFAT

**Altres:**

Primer quadrimestre:

ANNA QUERALT CALAFAT - 11, 12

OSCAR ROMERO MORAL - 11, 12

MARIA JOSEFINA SIERRA SANTIBAÑEZ - 11, 12

### CAPACITATS PRÈVIES

---

This course assumes basic competences in algorithms, data structures and databases. The course is structured to cope with different backgrounds and learning needs but basic knowledge on Computer Science principles is assumed: notions of computer architecture, basic programming constructs and data structures.

### COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

---

**Específiques:**

CE1. Desenvolupar algorismes eficients fonamentats en el coneixement i comprensió de la teoria de la complexitat computacional i les principals estructures de dades, dins de l'àmbit de ciència de dades

CE2. Aplicar els fonaments de la gestió i processament de dades en un problema de ciència de dades

**Genèriques:**

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

**Transversals:**

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

**Bàsiques:**

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

## METODOLOGIES DOCENTS

The students are divided in two tracks: one for students with a minor in computer science (track 1) and another one for students with a major in computer science (track 2).

Students in track 1 will study fundamental concepts in algorithms, data structures and databases. First, additional material to read, study and understand is provided. Lectures focus on the main concepts and those that require some additional explanation to guarantee a proper understanding. Students will have a large bank of self-assessing exercises to practice their understanding on their own. During the course, they will have to solve some mandatory exercises to guarantee a smooth learning process. Additionally, in the face-to-face lectures, the lecturer will solve doubts, go through representative exercises to guarantee a solid understanding and discuss exercises (to be solved during the lecture) with the students.

Students in track 2 will carry out the DS-EtE project and, for them, this course is a project course. There, students must create an end-to-end system architecture to ingest, store, process, learn models and deploy such system for a realistic project with realistic data. The students must develop good practices developing such architecture (what nowadays is known as DataOps / MLOps).

This course has a strong self-learning component and shares the same objectives in both tracks. The lecturers will supervise the students progress during the semester to guarantee a proper progress.

## OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

- 1.To analyse the cost of iterative and recursive algorithms
- 2.To review some simple data structures: stacks, queues, lists, and trees
- 3.To know, explain, design, analyse, compare and implement the main data structures and algorithms that can be used to implement priority queues
- 4.To know, explain, design, analyse, compare and implement the main data structures and algorithms that can be used to implement dictionaries
- 5.To know, explain, design, analyse, compare and implement the main data structures and algorithms that can be used to represent graphs and solve classic graph problems such as traversals, topological ordering and shortest paths
- 6.To know, understand, explain, analyse and compare some algorithm design techniques: greedy, divide and conquer, and dynamic programming
- 7.To be aware of the limits of computation: to understand the definitions of the P and NP classes, the concept of Polynomial-Time reduction, the notion of NP-Completeness, and to know some classic NP-complete problems
- 8.Describe what is a database and a database management system
- 9.Effectively use the standard Structured Query Language (SQL) to query relational databases
- 10.Explain the relational data model, including its data structures, the relational algebra and integrity constraints
- 11.Given a set of informational requirements, model the logic schema of a relational database
- 12.Identify the main objectives of a database management system query optimizer
- 13.Apply data structures, algorithms and database queries to solve a problem in a realistic situation

## HORES TOTS DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores grup gran	27,0	18.00
Hores aprenentatge autònom	96,0	64.00
Hores grup petit	27,0	18.00

**Dedicació total:** 150 h

## CONTINGUTS

### Basics of Analysis of Algorithms

#### Descripció:

Worst case, best case and average case cost analysis. Asymptotic order of growth notations: Big-O, Omega and Theta. Analysis of the correctness and cost of iterative and recursive algorithms.



### Simple Data Structures: Review

**Descripció:**

Stacks, queues, lists and trees.

### Priority Queues

**Descripció:**

Operations of priority queues. Implementations with heaps. Heapsort.

### Dictionaries

**Descripció:**

Operations of dictionaries. Basic implementations: tables and lists. Advanced implementations: hash tables, binary search trees, and AVL trees.

### Graphs

**Descripció:**

Representations: adjacency matrices, adjacency lists and implicit representations. Depth-first search (DFS). Breadth-first search (BFS). Topological sort. Algorithms for shortest paths. Algorithm for minimum spanning trees.

### Algorithm design techniques

**Descripció:**

Greedy, divide and conquer, and dynamic programming.

### Introduction to NP and Computational Intractability

**Descripció:**

Basic introduction to P and NP classes, Polynomial-Time reduction, and NP-completeness. Examples of classic NP-complete problems.

### Introduction to databases and database management systems

**Descripció:**

Main concepts on databases and database management systems. Relational database management systems.

### SQL: Data-definition language and data-manipulation language

**Descripció:**

Introduction to the SQL language.



### The relational model

**Descripció:**

Data structures and integrity constraints.

### The relational algebra

**Descripció:**

The relational algebra operators and how to build data pipes with them. Notion of semantic and syntactic optimization.

### Logical design of relational databases

**Descripció:**

Design the logical schema of a database.

### Notions of physical design and physical database optimization

**Descripció:**

Notions of query optimizer, access plan and cost model.

### Data Science-related advanced topics

**Descripció:**

Students with a major in Computer Science will investigate on advanced topics specific for data science projects. For example, data quality, entity resolution, data integration, etc.

## ACTIVITATS

### Basic concepts on Algorithms, Data Structures, and Databases

**Descripció:**

Motivation and main concepts on algorithmics, data structures, and database management systems.

**Objectius específics:**

1, 2, 8

**Competències relacionades:**

CE2. Aplicar els fonaments de la gestió i processament de dades en un problema de ciència de dades

CE1. Desenvolupar algoritmes eficients fonamentats en el coneixement i comprensió de la teoria de la complexitat computacional i les principals estructures de dades, dins de l'àmbit de ciència de dades

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

**Dedicació:** 7h 30m

Grup gran/Teoria: 7h 30m



## Partial Exam

### Descripció:

For students with a minor in Computer Science, this exam evaluates their knowledge on fundamental concepts of algorithms, data structures and databases.

### Objectius específics:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

### Competències relacionades:

CE2. Aplicar els fonaments de la gestió i processament de dades en un problema de ciència de dades

CE1. Desenvolupar algoritmes eficients fonamentats en el coneixement i comprensió de la teoria de la complexitat computacional i les principals estructures de dades, dins de l'àmbit de ciència de dades

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

### Dedicació: 9h

Activitats dirigides: 3h

Aprenentatge autònom: 6h

## Computer Science for Data Science

### Descripció:

Students in the minor track will put in practice all the contents of the course by solving exercises related to algorithms, data structures, and databases. Students in the major track will undertake a project spanning all main phases of a data science. As result, they are asked to develop a quality realistic end-to-end system architecture for a data science project.

### Objectius específics:

3, 4, 5, 6, 7, 9, 10, 11, 12, 13

### Competències relacionades:

CG1. Identificar i aplicar els mètodes i processos de gestió de dades més adequats per gestionar el cicle de vida de les dades, incloent-hi dades estructurades i no estructurades

CE2. Aplicar els fonaments de la gestió i processament de dades en un problema de ciència de dades

CE1. Desenvolupar algoritmes eficients fonamentats en el coneixement i comprensió de la teoria de la complexitat computacional i les principals estructures de dades, dins de l'àmbit de ciència de dades

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

### Dedicació: 124h 30m

Grup petit/Laboratori: 40h 30m

Aprenentatge autònom: 84h



## Final Exam

### Descripció:

For students with a minor in Computer Science, this exam evaluates their knowledge on fundamental concepts of algorithms, data structures and databases

### Objectius específics:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

### Competències relacionades:

CE2. Aplicar els fonaments de la gestió i processament de dades en un problema de ciència de dades

CE1. Desenvolupar algorismes eficients fonamentats en el coneixement i comprensió de la teoria de la complexitat computacional i les principals estructures de dades, dins de l'àmbit de ciència de dades

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CT4. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ: Gestionar l'adquisició, l'estructuració, l'anàlisi i la visualització de dades i informació de l'àmbit d'especialitat, i valorar de forma crítica els resultats d'aquesta gestió.

### Dedicació: 9h

Activitats dirigides: 3h

Aprenentatge autònom: 6h

## SISTEMA DE QUALIFICACIÓ

Let DBM = Databases midterm exam grade,

DBF = Databases final exam grade,

DBEx = Database exercises to be solved during the course,

ADSM = Algorithms and Data Structures midterm exam grade,

ADSF = Algorithms and Data Structures final exam grade and

ADSEx = Algorithms and Data Structures exercises to be solved during the course,

DS-EtE = DS-EtE project grade

Then,

1) If the student followed track 1 (see methodology) the mark is calculated as follows:

$BD = \text{MAX}(0.2 \cdot \text{DBEx} + 0.8 \cdot \text{DBM}, \text{DBF})$ ,

$\text{ADS} = \text{MAX}(0.2 \cdot \text{ADSEx} + 0.4 \cdot \text{ADSM} + 0.4 \cdot \text{ADSF}, \text{ADSF})$ .

$\text{ADSDB (final course mark)} = 0.5 \cdot \text{BD} + 0.5 \cdot \text{ADS}$

2) If the student followed track 2 (see methodology) the mark is calculated as follows:

$\text{ADSDB (final course mark)} = \text{DS-EtE}$

## BIBLIOGRAFIA

### Bàsica:

- Cormen, T.H. [et al.]. Introduction to algorithms [en línia]. 4rd ed. Cambridge: MIT Press, 2022 [Consulta: 10/01/2024]. Disponible a :

<https://search-ebshost-com.recursos.biblioteca.upc.edu/login.aspx?direct=true&AuthType=ip,uid&db=nlebk&AN=2932690&site=ehost-live&ebv=EK&ppid=Page-1>. ISBN 9780262046305.

- Goodrich, M.T.; Tamassia, R.; Goldwasser, M.H. Data structures and algorithms in Python. New York: John Wiley & Sons, 2013. ISBN 9781118290279.

- Savitch, Walter J; Mock, Kenrick. Problem solving with C++. Tenth edition. New York, NY: Pearson, [2018]. ISBN 9780134448282.



## RECURSOS

---

**Enllaç web:**

- <http://learnsql.fib.upc.edu>. LEARN-SQL
- <https://jutge.org>. Jutge