# Course guide
# 270429 - BDA - Advanced Databases

**Last modified:** 02/02/2024

**Unit in charge:** Barcelona School of Informatics
**Teaching unit:** 747 - ESSI - Department of Service and Information System Engineering.

**Degree:** BACHELOR'S DEGREE IN ARTIFICIAL INTELLIGENCE (Syllabus 2021). (Compulsory subject).

**Academic year:** 2023 **ECTS Credits:** 6.0 **Languages:** English

## LECTURER

**Coordinating lecturer:** SERGI NADAL FRANCESCH - PETAR JOVANOVIC

**Others:** Segon quadrimestre:
PETAR JOVANOVIC - 11, 12
SERGI NADAL FRANCESCH - 11, 12

## PRIOR SKILLS

Fundamental knowledge of relational data modeling.
Be able to create, consult and manipulate databases with SQL.
Foundations of knowledge representation and first-order logics
Advanced programming in Python.

## DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

**Specific:**
CE04. To design and use efficiently the most appropriate data types and structures to solve a problem.
CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.
CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.
CE10. To analyze, design, build and maintain applications in a robust, secure and efficient way, choosing the most appropriate paradigm and programming languages.
CE15. To acquire, formalize and represent human knowledge in a computable form for solving problems through a computer system in any field of application, particularly those related to aspects of computing, perception and performance in intelligent environments or environments.

**Generical:**
CG2. To use the fundamental knowledge and solid work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.
CG5. Work in multidisciplinary teams and projects related to artificial intelligence and robotics, interacting fluently with engineers and professionals from other disciplines.
CG9. To face new challenges with a broad vision of the possibilities of a professional career in the field of Artificial Intelligence. Develop the activity applying quality criteria and continuous improvement, and act rigorously in professional development. Adapt to organizational or technological changes. Work in situations of lack of information and / or with time and / or resource restrictions.

**Transversal:**
CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CT8. (ENG) Perspectiva de gènere. Conèixer i comprendre, des del propi àmbit de la titulació, les desigualtats per raó de sexe i gènere a la societat; Integrar les diferents necessitats i preferències per raó de sexe i de gènere en el disseny de solucions i resolució de problemes.

**Basic:**
CB1. That students have demonstrated to possess and understand knowledge in an area of ??study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.
CB2. That the students know how to apply their knowledge to their work or vocation in a professional way and possess the skills that are usually demonstrated through the elaboration and defense of arguments and problem solving within their area of ??study.

## TEACHING METHODOLOGY

The course has theory lectures and laboratory sessions.

Lectures: The teacher presents the topic. Students follow the lesson, take notes, and prepare additional material outside of class. They may also be asked to carry out assessment activities within these sessions.

Laboratory: Mainly, the laboratory sessions will be dedicated to the practice (with or without a computer) of the concepts introduced in the lectures. Tools relevant to the concepts introduced are presented and used in small projects in these sessions. Mini projects will also be done, in which students will work in teams. For each mini project there will be a delivery outside class time, but students will also be assessed individually in the classroom on the knowledge acquired during each of the projects.

The course has an autonomous learning component, as the students will have to work with different data management and processing tools. Apart from the support material, students should be able to resolve doubts or problems using these tools.

## LEARNING OBJECTIVES OF THE SUBJECT

1.Be able to explain and use the main mechanisms of parallel processing of queries in distributed environments, and detect bottlenecks.
2.Learn, understand and apply the fundamentals of distributed data management systems like distributed databases and distributed file systems.
3.Be able to justify and use functional-style distributed data processing environments.
4.Learn, understand and apply the fundamentals of knowledge graphs.
5.Be able to specify, design, implement and evaluate AI-oriented data management systems, including semantic databases for knowledge representation.
6.Be able to apply knowledge graphs to solve realistic problems such as data integration, graph-based data analysis, etc.
7.Be able to evaluate and select data management systems based on a certain quality criterion.
8.Be able to solve data discovery and integration problems based on available strategies, standards and technologies.
10.Be able to perform graph data query processing both.

## STUDY LOAD

| Type | Hours | Percentage |
|---|---|---|
| Hours large group | 30,0 | 20.00 |
| Hours small group | 30,0 | 20.00 |
| Self study | 90,0 | 60.00 |

**Total learning time:** 150 h

## CONTENTS

### Introduction to data systems for Artificial Intelligence.

**Description:**
The complete AI lifecycle with DevOps and DataOps. Data acquisition, cleaning, and preparation. Model selection and management. Model debugging and serving.

### Large-scale data management and processing

**Description:**
Distributed databases. Overview of distributed data management and processing. Distributed files system. Distributed data processing frameworks (MapReduce/Spark). Dataflow processing models. Declarative dataflow programs.

### Semantic data management

**Description:**
Foundations of graph data management. Knowledge graph representations with RDF, RDFS, OWL and their relationship with first-order logics. Pattern matching and the SPARQL query language. Languages for describing and validating knowledge graphs.

### Data integration

**Description:**
Data discovery. Data quality evaluation. Schema and data integration.

### Architectures for data-centric AI systems and their governance

**Description:**
Centralized and Distributed functional architectures of data management systems for AI. Data governance.

## ACTIVITIES

**introduction to data systems for AI**

**Description:**
Introduction of the subject, motivation and overview of the data lifecycle for AI.

**Specific objectives:**
5

**Related competencies :**
CG2. To use the fundamental knowledge and solid work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG5. Work in multidisciplinary teams and projects related to artificial intelligence and robotics, interacting fluently with engineers and professionals from other disciplines.
CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.
CE15. To acquire, formalize and represent human knowledge in a computable form for solving problems through a computer system in any field of application, particularly those related to aspects of computing, perception and performance in intelligent environments or environments.
CE04. To design and use efficiently the most appropriate data types and structures to solve a problem.
CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CB1. That students have demonstrated to possess and understand knowledge in an area of ??study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.

**Full-or-part-time:** 4h
Theory classes: 2h
Self study: 2h

**Study of large-scale data management and processing**

**Specific objectives:**
1, 2, 3, 5, 7

**Related competencies :**
CG9. To face new challenges with a broad vision of the possibilities of a professional career in the field of Artificial Intelligence. Develop the activity applying quality criteria and continuous improvement, and act rigorously in professional development. Adapt to organizational or technological changes. Work in situations of lack of information and / or with time and / or resource restrictions.
CG2. To use the fundamental knowledge and solid work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.
CG5. Work in multidisciplinary teams and projects related to artificial intelligence and robotics, interacting fluently with engineers and professionals from other disciplines.
CE10. To analyze, design, build and maintain applications in a robust, secure and efficient way, choosing the most appropriate paradigm and programming languages.
CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.
CE15. To acquire, formalize and represent human knowledge in a computable form for solving problems through a computer system in any field of application, particularly those related to aspects of computing, perception and performance in intelligent environments or environments.
CE04. To design and use efficiently the most appropriate data types and structures to solve a problem.
CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.
CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CB1. That students have demonstrated to possess and understand knowledge in an area of ??study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.
CB2. That the students know how to apply their knowledge to their work or vocation in a professional way and possess the skills that are usually demonstrated through the elaboration and defense of arguments and problem solving within their area of ??study.

**Full-or-part-time:** 38h
Theory classes: 8h
Laboratory classes: 10h
Self study: 20h

**Study of semantic data management**

**Specific objectives:**
4, 5, 6, 10

**Related competencies :**
CG9. To face new challenges with a broad vision of the possibilities of a professional career in the field of Artificial Intelligence. Develop the activity applying quality criteria and continuous improvement, and act rigorously in professional development. Adapt to organizational or technological changes. Work in situations of lack of information and / or with time and / or resource restrictions.
CG2. To use the fundamental knowledge and solid work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.
CG5. Work in multidisciplinary teams and projects related to artificial intelligence and robotics, interacting fluently with engineers and professionals from other disciplines.
CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.
CE15. To acquire, formalize and represent human knowledge in a computable form for solving problems through a computer system in any field of application, particularly those related to aspects of computing, perception and performance in intelligent environments or environments.
CE04. To design and use efficiently the most appropriate data types and structures to solve a problem.
CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.
CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.
CT8. (ENG) Perspectiva de gènere. Conèixer i comprendre, des del propi àmbit de la titulació, les desigualtats per raó de sexe i gènere a la societat; Integrar les diferents necessitats i preferències per raó de sexe i de gènere en el disseny de solucions i resolució de problemes.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CB1. That students have demonstrated to possess and understand knowledge in an area of ??study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.
CB2. That the students know how to apply their knowledge to their work or vocation in a professional way and possess the skills that are usually demonstrated through the elaboration and defense of arguments and problem solving within their area of ??study.

**Full-or-part-time:** 34h
Theory classes: 8h
Laboratory classes: 10h
Self study: 16h

**Study of data integration**

**Specific objectives:**
3, 5, 6, 8

**Related competencies :**
CG9. To face new challenges with a broad vision of the possibilities of a professional career in the field of Artificial Intelligence. Develop the activity applying quality criteria and continuous improvement, and act rigorously in professional development. Adapt to organizational or technological changes. Work in situations of lack of information and / or with time and / or resource restrictions.
CG2. To use the fundamental knowledge and solid work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.
CG5. Work in multidisciplinary teams and projects related to artificial intelligence and robotics, interacting fluently with engineers and professionals from other disciplines.
CE10. To analyze, design, build and maintain applications in a robust, secure and efficient way, choosing the most appropriate paradigm and programming languages.
CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.
CE15. To acquire, formalize and represent human knowledge in a computable form for solving problems through a computer system in any field of application, particularly those related to aspects of computing, perception and performance in intelligent environments or environments.
CE04. To design and use efficiently the most appropriate data types and structures to solve a problem.
CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.
CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.
CT8. (ENG) Perspectiva de gènere. Conèixer i comprendre, des del propi àmbit de la titulació, les desigualtats per raó de sexe i gènere a la societat; Integrar les diferents necessitats i preferències per raó de sexe i de gènere en el disseny de solucions i resolució de problemes.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CB1. That students have demonstrated to possess and understand knowledge in an area of ??study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.
CB2. That the students know how to apply their knowledge to their work or vocation in a professional way and possess the skills that are usually demonstrated through the elaboration and defense of arguments and problem solving within their area of ??study.

**Full-or-part-time:** 20h
Theory classes: 4h
Laboratory classes: 6h
Self study: 10h

**Study of architectures for data-centric AI systems**

**Specific objectives:**
2, 5, 7

**Related competencies :**
CG2. To use the fundamental knowledge and solid work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.
CG5. Work in multidisciplinary teams and projects related to artificial intelligence and robotics, interacting fluently with engineers and professionals from other disciplines.
CE10. To analyze, design, build and maintain applications in a robust, secure and efficient way, choosing the most appropriate paradigm and programming languages.
CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.
CE15. To acquire, formalize and represent human knowledge in a computable form for solving problems through a computer system in any field of application, particularly those related to aspects of computing, perception and performance in intelligent environments or environments.
CE04. To design and use efficiently the most appropriate data types and structures to solve a problem.
CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CB1. That students have demonstrated to possess and understand knowledge in an area of ??study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.
CB2. That the students know how to apply their knowledge to their work or vocation in a professional way and possess the skills that are usually demonstrated through the elaboration and defense of arguments and problem solving within their area of ??study.

**Full-or-part-time:** 18h
Theory classes: 4h
Laboratory classes: 4h
Self study: 10h

**Midterm exam**

**Specific objectives:**
1, 2, 3, 5

**Related competencies :**
CG9. To face new challenges with a broad vision of the possibilities of a professional career in the field of Artificial Intelligence. Develop the activity applying quality criteria and continuous improvement, and act rigorously in professional development. Adapt to organizational or technological changes. Work in situations of lack of information and / or with time and / or resource restrictions.
CG2. To use the fundamental knowledge and solid work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.
CG5. Work in multidisciplinary teams and projects related to artificial intelligence and robotics, interacting fluently with engineers and professionals from other disciplines.
CE10. To analyze, design, build and maintain applications in a robust, secure and efficient way, choosing the most appropriate paradigm and programming languages.
CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.
CE15. To acquire, formalize and represent human knowledge in a computable form for solving problems through a computer system in any field of application, particularly those related to aspects of computing, perception and performance in intelligent environments or environments.
CE04. To design and use efficiently the most appropriate data types and structures to solve a problem.
CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.
CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CB1. That students have demonstrated to possess and understand knowledge in an area of ??study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.
CB2. That the students know how to apply their knowledge to their work or vocation in a professional way and possess the skills that are usually demonstrated through the elaboration and defense of arguments and problem solving within their area of ??study.

**Full-or-part-time:** 16h
Guided activities: 2h
Self study: 14h

**Final exam**

**Specific objectives:**
4, 6, 7, 8, 10

**Related competencies :**
CG9. To face new challenges with a broad vision of the possibilities of a professional career in the field of Artificial Intelligence. Develop the activity applying quality criteria and continuous improvement, and act rigorously in professional development. Adapt to organizational or technological changes. Work in situations of lack of information and / or with time and / or resource restrictions.
CG2. To use the fundamental knowledge and solid work methodologies acquired during the studies to adapt to the new technological scenarios of the future.
CG3. To define, evaluate and select hardware and software platforms for the development and execution of computer systems, services and applications in the field of artificial intelligence.
CG5. Work in multidisciplinary teams and projects related to artificial intelligence and robotics, interacting fluently with engineers and professionals from other disciplines.
CE10. To analyze, design, build and maintain applications in a robust, secure and efficient way, choosing the most appropriate paradigm and programming languages.
CE08. To detect the characteristics, functionalities and components of data managers, which allow the adequate use of them in information flows, and the design, analysis and implementation of applications based on them.
CE15. To acquire, formalize and represent human knowledge in a computable form for solving problems through a computer system in any field of application, particularly those related to aspects of computing, perception and performance in intelligent environments or environments.
CE04. To design and use efficiently the most appropriate data types and structures to solve a problem.
CE09. To ideate, design and integrate intelligent data analysis systems with their application in production and service environments.
CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.
CT8. (ENG) Perspectiva de gènere. Conèixer i comprendre, des del propi àmbit de la titulació, les desigualtats per raó de sexe i gènere a la societat; Integrar les diferents necessitats i preferències per raó de sexe i de gènere en el disseny de solucions i resolució de problemes.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CB1. That students have demonstrated to possess and understand knowledge in an area of ??study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.
CB2. That the students know how to apply their knowledge to their work or vocation in a professional way and possess the skills that are usually demonstrated through the elaboration and defense of arguments and problem solving within their area of ??study.

**Full-or-part-time:** 20h
Guided activities: 2h
Self study: 18h

## GRADING SYSTEM

The qualification of technical skills is based on:

- NPR: Project grade, as a weighted average of the mini-projects of the course

- NEP: Grade of the partial exam.

- NEF: Grade of the final exam.

Final grade = NPR*0.40+NEP*0.20+NEF*0.40

For students who can participate in the re-evaluation, the re-evaluation exam mark will replace NEF and NEP. In any case, the final mark will be the maximum between the ordinary mark and the re-evaluation mark.

## BIBLIOGRAPHY

**Basic:**
- Badia, Antonio. SQL for data science : data cleaning, wrangling and analytics with relational databases. Springer, [2020]. ISBN 9783030575915.
- Özsu, M. Tamer; Valduriez, Patrick. Principles of distributed database systems. Fourth edition. Springer, [2020]. ISBN 9783030262525.
- Hogan, Aidan. The Web of Data. Springer. 2020. ISBN 978-3-030-51580-5.
- Sadalage, Pramod J; Fowler, Martin. NoSQL distilled : a brief guide to the emerging world of polygot persistence. Addison-Wesley, 2013. ISBN 9780321826626.
- Groppe, Sven. Data management and query processing in semantic web databases. Springer, 2011. ISBN 9783642193569.
- Abiteboul, S. Web data management. Cambridge University Press, 2012. ISBN 9781107012431.

**Complementary:**
- Özsu, M. Tamer; Liu, Ling. Encyclopedia of database systems [Recurs electrònic]. Springer, 2009. ISBN 9780387399409.
- Aggarwal, Charu C; Wang, Haixun. Managing and mining graph data. Springer, cop. 2010. ISBN 9781441960443.
- Lenzerini, Maurizio. Data Integration: A Theoretical Perspective. PODS '02: Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, 2002. ISBN 1-58113-507-6.
- Özsu, M. Tamer. A Survey of RDF Data Management Systems. Cornell University Library, 2016.

## RESOURCES

**Hyperlink:**
- https://learnsql2.fib.upc.edu/moodle/course/view.php?id=83