Course guide
270667 - DW - Data Warehousing

Last modified: 12/07/2021

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

Degree: MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Optional subject).

Academic year: 2021 **ECTS Credits:** 6.0 **Languages:** English

LECTURER

Coordinating lecturer: ALBERTO ABELLO GAMAZO

Others: Primer quadrimestre:
ALBERTO ABELLO GAMAZO - 11, 12
PETAR JOVANOVIC - 11, 12

PRIOR SKILLS

Basic knowledge on relational databases and SQL.

Specifically, it will be assumed knowledge on:
- UML class diagrams
- Relational algebra
- SQL queries
- Relational views
- B-tree operations (i.e., insertion and splits)
- Basic concepts on physical query optimization

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.
CEE5.1. Capability to participate in improvement projects or to create service systems, providing in particular: a) innovation and research proposals based on new uses and developments of information technologies, b) application of the most appropriate software engineering and databases principles when developing information systems, c) definition, installation and management of infrastructure / platform necessary for the efficient running of service systems.
CEE5.3. Capability to work in interdisciplinary engineering services teams and, provided the necessary domain experience, capability to work autonomously in specific service systems.

Generical:
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.

Transversal:
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.

Basic:
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

TEACHING METHODOLOGY

The course comprises theory, and lab sessions.

Theory: Inverted class techniques will be used, which require that the student work on the provided multimedia materials before the class. Then, theory lectures comprise the teacher's complementary explanations and problem solving.

Laboratory: Some representative tools will be used for the application of theoretical concepts (e.g., Indyco Builder, PotgreSQL, Oracle, Pentaho Data Integration, Tableau). The course includes continuous hands-on through a course project, divided into three logical blocks: data warehouse modelling, data integration and migration (ETL), and descriptive visualisation, in which the students will work in teams. There will be three project deliverables outside the class hours, while in the class the students will be as well individually assessed about the knowledge acquired during each project block.

LEARNING OBJECTIVES OF THE SUBJECT

1.Be able to model multidimensional data warehouses and visually analyze their data
2.Be able to apply specific physical design techniques for decisional systems
3.Be able to design and implement data migration processes (i.e., ETL)

STUDY LOAD

Type	Hours	Percentage
Hours small group	25,5	17.00
Hours large group	25,5	17.00
Self study	96,0	64.00
Guided activities	3,0	2.00

Total learning time: 150 h

CONTENTS

Introduction

Description:
Comparison of operational and decisional systems; Metadata

Data warehousing architectures

Description:
Corporate Information Factory; DW 2.0

Multidimensional modeling, OLAP tools

Description:
Structure; Integrity constraints; Operations; Advanced concepts

Database physical desing for analytical queries

Description:
Star-join and join indexes; Bitmaps; Materialized views; Spatio-temporal data

Extraction, Transformation and Load

Description:
Data quality; Schema and Data Integration; ETL management

Visualization and descriptive analytics

Description:
Key Performance Indicators; Dashboarding

ACTIVITIES

Theoretical lectures

Description:
In these activities, the lecturer will introduce the main theoretical concepts of the subject. Besides lecturing, cooperative learning techniques will be used. These demand the active participation of the students, and consequently will be evaluated.

Specific objectives:
1, 2, 3

Related competencies :
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.
CEE5.3. Capability to work in interdisciplinary engineering services teams and, provided the necessary domain experience, capability to work autonomously in specific service systems.
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.
CEE5.1. Capability to participate in improvement projects or to create service systems, providing in particular: a) innovation and research proposals based on new uses and developments of information technologies, b) application of the most appropriate software engineering and databases principles when developing information systems, c) definition, installation and management of infrastructure / platform necessary for the efficient running of service systems.
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

Full-or-part-time: 50h
Theory classes: 25h
Self study: 25h

Hands-on sessions

Description:
The student will be asked to practice the different concepts introduced in the theoretical lectures. This includes problem solving either on the computer or on paper.

Specific objectives:
1, 2, 3

Related competencies :
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.
CEE5.3. Capability to work in interdisciplinary engineering services teams and, provided the necessary domain experience, capability to work autonomously in specific service systems.
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.
CEE5.1. Capability to participate in improvement projects or to create service systems, providing in particular: a) innovation and research proposals based on new uses and developments of information technologies, b) application of the most appropriate software engineering and databases principles when developing information systems, c) definition, installation and management of infrastructure / platform necessary for the efficient running of service systems.
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

Full-or-part-time: 81h
Laboratory classes: 27h
Self study: 54h

Exam

Description:
Written exam of the theoretical concepts introduced along the course.

Specific objectives:
1, 2, 3

Related competencies :
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.
CEE5.3. Capability to work in interdisciplinary engineering services teams and, provided the necessary domain experience, capability to work autonomously in specific service systems.
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.
CEE5.1. Capability to participate in improvement projects or to create service systems, providing in particular: a) innovation and research proposals based on new uses and developments of information technologies, b) application of the most appropriate software engineering and databases principles when developing information systems, c) definition, installation and management of infrastructure / platform necessary for the efficient running of service systems.
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

Full-or-part-time: 19h
Theory classes: 2h
Self study: 17h

GRADING SYSTEM

Final Mark = min(10 ; 60%E + 40%L + 10%P)

L = Weighted average of the marks of the three lab deliverables
E = Final exam
P = Participation in the class

BIBLIOGRAPHY

Basic:
- Golfarelli, M.; Rizzi, S. Data warehouse design: modern principles and methodologies. McGraw Hill, 2009. ISBN 9780071610391.
- Jensen, C.S.; Pedersen, T.B.; Thomsen, C.W. Multidimensional databases and data warehousing. Morgan & Claypool, 2010. ISBN 9781608455379.
- Inmon, W.H.; Imhoff, C.; Sousa, R. Corporate information factory. 2nd ed. John Wiley, 2001. ISBN 0471399612.
- Kimball, R. [et al.]. The data warehouse lifecycle toolkit. 2nd ed. Wiley publishing, 2008. ISBN 9780470149775.
- Garcia-Molina, H.; Ullman, J.D.; Widom, J. Database systems: the complete book [on line]. 2nd ed. Pearson Education Limited, 2013 [Consultation: 15/02/2022]. Available on: https://ebookcentral.proquest.com/lib/upcatalunya.ebooks/detail.action?pq.origsite=primo&docID=5174436. ISBN 9781292024479.
- Vaisman, A.; Zimanyi, E. Data warehouse systems: design and implentation. Springer, 2014. ISBN 9783642546549.

Complementary:
- Teorey, T.J.; Nadeau, T.; Lightstone, S. Database modeling and design: logical design. 5th ed. Morgan Kaufmann Publishers/Elsevier, 2011. ISBN 9780123820204.
- Lightstone, S.; Teorey, T.J.; Nadeau, T. Physical database design: the database professional's guide to exploiting indexes, views, storage, and more. Morgan Kaufmann Publishers, 2007. ISBN 9780123693891.
- Ramakrishnan, R.; Gehrke, J. Database management systems. 3rd ed. McGraw-Hill, 2003. ISBN 0071151109.
- Lewis, J. Cost-based oracle fundamentals. Apress, 2006. ISBN 9781590596364.

RESOURCES

Hyperlink:
- http://cs.ulb.ac.be/conferences/ebiss.html
- http://tdwi.org
- https://deds.ulb.ac.be