

62225 - Management of Large-Scale Data

Información del Plan Docente

Academic Year	2017/18
Faculty / School	110 - Escuela de Ingeniería y Arquitectura
Degree	534 - Master's in IT Engineering
ECTS	6.0
Year	1
Semester	Second semester
Subject Type	Compulsory
Module	---

1.General information

1.1.Introduction

1.2.Recommendations to take this course

1.3.Context and importance of this course in the degree

1.4.Activities and key dates

2.Learning goals

2.1.Learning goals

2.2.Importance of learning goals

3.Aims of the course and competences

3.1.Aims of the course

3.2.Competences

4.Assessment (1st and 2nd call)

4.1.Assessment tasks (description of tasks, marking system and assessment criteria)

5.Methodology, learning tasks, syllabus and resources

5.1.Methodological overview

The methodology followed in this course is oriented towards achievement of the learning objectives. A wide range of teaching and learning tasks are implemented, such as classroom activities and autonomous work.

Classroom activities

- Theory sessions. Teacher's presentation of the course contents, including illustrative examples.

62225 - Management of Large-Scale Data

- Talks by experts. When possible, external experts to the university will explain some contents.
- Seminars. Sessions based on oral or written contributions by the students.
- Problem-based learning. Educative approach oriented towards a teaching and learning methodology where students tackle real problems in small groups under the supervision of a teacher.
- Practice sessions. Any practical or collaborative activity done in class.
- Laboratory sessions. Activities developed with specialized equipment (in labs, computer labs).
- Tutorials. Students can review or discuss with the teacher the materials and topics presented in class.
- Assessment. A set of written/oral tests, lab assignments, projects, other assignments, etc.

Autonomous work

- Theoretical assignments. Preparing seminars, readings, research, assignments or write reports, etc. to be presented or submitted to the teacher in theory sessions.
- Practical assignments. Preparing activities to be presented or submitted to the teacher in practice sessions.
- Theoretical study. Study of contents related to the theory sessions: it includes any study activity not considered previously (study for exams, work in the library, complementary readings, solve problems and exercises, etc.).
- Practical study. Related to the practice sessions.
- Complementary activities. Formative activities related to the course, but not to the preparation of exams or included within the assessment activities: readings, seminars, videos, etc.

5.2.Learning tasks

The course (6 ECTS: 150 hours) includes the following learning tasks:

- Classroom activities (40 hours). Theory sessions, expert talks, seminars, problem-solving and cases, and lab assignments.
- Practice and research assignments (80 hours).
- Tutorials (5 hours).
- Autonomous work and study (20 hours).
- Assessment (5 hours). Exam and defense of the course project.

5.3.Syllabus

The course will address the following topics:

1. Introduction and motivation to the problem of large volumes of data (Big Data).
2. Storage of large amounts of data:
 - Data warehouses. Star schema design.
 - NoSQL databases.
3. Management of large amounts of data:
 - Data distribution.
 - Information integration considering heterogeneous data sources.
 - Use of knowledge representation techniques (ontologies) to represent data sources and their access and integration.
 - Parallel processing techniques: MapReduce (Hadoop).
 - Data Stream Management Systems.
 - Other techniques: mobile agents.
4. Interaction with large amounts of data:
 - Visualization techniques.
 - Design of appropriate user interfaces.
 - Usability.
5. Analysis of large amounts of data:
 - Data mining.
 - Sentiment analysis.
 - Text mining.
6. Use cases and applications, such as:
 - Data provided by sensors.
 - Unstructured data on the Web.

62225 - Management of Large-Scale Data

- Recommendation Systems.
- Analysis of blogs and social networks.
- Smart cities.
- Intelligent Transportation Systems.

5.4. Course planning and calendar

Further information concerning the timetable, classroom, office hours, assessment dates and other details regarding this course, will be provided on the first day of class or please refer to the Center website.

5.5. Bibliography and recommended resources

- Adamson, Christopher. Star schema : the complete reference / Christopher Adamson . New York : McGraw-Hill, 2010
- Joyanes Aguilar, Luis. Big Data : análisis de grandes volúmenes de datos en organizaciones / Luis Joyanes Aguilar . - 1ª ed. [Barcelona] : Marcombo, 2014.
- Kimball, Ralph. The data warehouse toolkit : the definitive guide to dimensional modeling / Ralph Kimball, Margy Ross . 3rd ed. Indianapolis : John Wiley & sons, cop.2013
- Marz, Nathan. Big Data: Principles and Best Practices of Scalable Realtime Data Systems / Nathan Marz, James Warren Manning Publications, 2014.
- Meirelles, Isabel. Design for information : An introduction to the histories, theories and best practices behind effective information visualizations / Isabel Meirelles Rockport Publishers, 2013.
- Ward, Matthew O.. Interactive data visualization : Foundations, techniques and applications / Matthew O. Ward...[et al.] CRC Press, 2010.
- Jensen, Christian S. Multidimensional databases and data warehousing / Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen . [San Rafael (California)] : Morgan & Claypool Publishers, cop. 2010
- Kimball, Ralph. The data warehouse lifecycle toolkit / Ralph Kimball. 2nd ed. John Wiley & Sons, 2008
- Krishnan, Krish. Data warehousing in the age of Big Data / Krish Krishnan Amsterdam : Morgan Kaufmann is an imprint of Elsevier, cop. 2013.
- Liu, Bing. Sentiment Analysis and Opinion Mining : Synthesis Lectures on Human Language Technologies / Bing Liu Morgan & Claypool Publishers, 2012.
- Liu, Bing. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data / Bing Liu Springer, 2011.
- Loshin, David. Big data analytics : from strategic planning to enterprise integration with tools, techniques, NoSQL, and graph / David Loshin Amsterdam : Elsevier, cop. 2013.
- Malinowski, Elzbieta. Advanced data warehouse design : from conventional to spatial and temporal applications / Elzbieta Malinowski, Esteban Zimányi . [1st ed.], 2nd corr. print. Berlin : Springer, cop. 2009
- Sumathi, S.. Introduction to Data Mining and its Applications / S. Sumathi, S. N. Sivanandam Springer, 2006..