60949 - Management of Large-Scale Data

Academic Year	2017/18
Subject	60949 - Management of Large-Scale Data
Faculty / School	110 - Escuela de Ingeniería y Arquitectura
Degree	533 - Master's Degree in Telecommunications Engineering
ECTS	6.0
Year	2
Semester	Second semester
Subject Type	Optional
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

- Lectures. The teacher presents and explains the course contents with illustrative examples.
- Talks by experts. When possible, external experts to the university will expose or explain some contents.
- Seminars. Period of instruction based on oral or written contributions by the students.
- Problem-based learning. Educative approach where students tackle real problems in small groups under the supervision of a teacher.
- Practice sessions. Any practical or collaborative activity in class.
- Lab sessions. Activities developed in special spaces with specialized equipment (labs, computer labs).
- Tutorials. Time for students to review and discuss materials and topics presented in class with the teacher.
- Assessment. A set of written/oral tests, lab assignments, projects, other assignments, etc., used to evaluate the progress of students.

Autonomous work

- Assignments. Preparation of seminars, readings, research, assignments or written reports, etc., to be presented or submitted to the professor in lectures and practice sessions.
- Autonomous work and study. Study of contents related to the lectures and practice sessions: it includes any study activity not considered previously (study for exams, work in the library, complementary readings, solve problems and exercises, etc.).
- Complementary activities. Formative activities related to the course (exam preparation or assessment) like readings, seminars, conferences, videos, etc.

5.2.Learning tasks

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

- Teaching sessions (40 hours). Lectures, practice sessions, laboratory sessions, special talks, etc.
- Project and assignments (80 hours)
- Tutorials (5 hours)
- Study (20 hours)
- Assessment (5 hours)

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
   1. Data mining.
   2. Sentiment analysis.
   3. Text mining.
6. Use cases and applications
   - Data provided by sensors.
   - Unstructured data on the Web.

- 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 EINA website.

5.5.Bibliography and recommended resources

BB	Adamson, Christopher. Star schema : the complete reference / Christopher Adamson . New York : McGraw-Hill, 2010
BB	Joyanes Aguilar, Luis. Big Data : análisis de grandes volúmenes de datos en organizaciones / Luis Joyanes Aguilar . - 1ª ed. [Barcelona] : Marcombo, 2014.
BB	Kimball, Ralph. The data warehouse toolkit : the definitive guide to dimensional modeling / Ralph Kimball, Margy Ross . 3rd ed. Indianapolis : John Wiley & sons, cop. 2013
BC	Krishnan, Krish. Data warehousing in the age of Big Data / Krish Krishnan Amsterdam : Morgan Kaufmann is an imprint of Elsevier, cop. 2013.
BC	Loshin, David. Big data analytics : from strategic planning to enterprise integration with tools, techniques, NoSQL, and graph / David Loshin Amsterdam : Elsevier, cop. 2013.
BB	Marz, Nathan. Big Data: Principles and Best Practices of Scalable Realtime Data Systems / Nathan Marz, James Warren Manning Publications, 2014.
BB	Meirelles, Isabel. Design for information : An introduction to the histories, theories and best practices behind effective information visualizations / Isabel Meirelles Rockport Publishers, 2013.
BB	Ward, Matthew O.. Interative data visualization : Foundations, techniques and applications / Matthew O. Ward...[et al.] CRC Press, 2010.
BC	Jensen, Christian S. Multidimensional databases and data warehousing / Christian S. Jensen, Torben Bach Pedersen, Christian Thomsen . [San Rafael (California)] : Morgan & Claypool Publishers, cop. 2010
BC	Kimball, Ralph. The data warehouse lifecycle toolkit / Ralph Kimball. 2nd ed. John Wiley & Sons, 2008
BC	Liu, Bing. Sentiment Analysis and Opinion

60949 - Management of Large-Scale Data

BC	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.
BC	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
BC	Sumathi, S.. Introduction to Data Mining and its Applications / S. Sumathi, S. N. Sivanandam Springer, 2006..

LIST OF URLs:

Slides, problem descriptions, case studies and instructions of practice sessions that the teachers of the course make available through the plaform Moodle.
https://moodle2.unizar.es/add/