60949 - Management of Large-Scale Data

1.General information

1.1.Aims of the course

1.2.Context and importance of this course in the degree

1.3.Recommendations to take this course

2.Learning goals

2.1.Competences

2.2.Learning goals

2.3.Importance of learning goals

3.Assessment (1st and 2nd call)

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

4.Methodology, learning tasks, syllabus and resources

4.1.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.

4.2.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)

4.3.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.

4.4.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.

4.5.Bibliography and recommended resources

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