Universidad Zaragoza

Academic Year/course: 2024/25

62225 - Management of Large-Scale Data

Syllabus Information

Academic year: 2024/25
Subject: 62225 - Management of Large-Scale Data
Faculty / School: 110 - Escuela de Ingeniería y Arquitectura
Degree: 534 - Master's Degree in Informatics Engineering
ECTS: 6.0
Year: 1
Semester: Second semester
Subject type: Compulsory
Module:

1. General information

This subject aims to achieve the following objectives:

- Help the student to analyse the requirements for the management of a problem that involves large volumes of data (storage, transfer, processing, visualization and interaction).
- Help the student to develop the necessary elements to integrate heterogeneous data sources, using classic techniques for representation, processing, analysis, visualization and interaction with heterogeneous data repositories.
- Help the student to develop an application for a given context, where the management of large volumes of data is necessary, taking into account criteria of scalability, usability and regulations.

2. Learning results

Upon completion of this subject, the student will be able to:

1. Understand and specify the necessary requirements for the interaction, storage, transfer, and processing of large volumes of data.
2. Know, understand and apply the most common techniques for the representation, treatment, analysis of and interaction with heterogeneous data repositories.
3. Design, develop and evaluate an application that facilitates the creation and management of large volumes of data, in accordance with scalability criteria and existing regulations.

3. Syllabus

- Introduction to and motivation of the problem of large volumes of data (*Big Data*).
- Storage of large volumes of data:
    - *Data warehouses* Star design.
    - NoSQL databases.
- Management of large volumes of data:
    - Data distribution.
    - Integration of information with heterogeneous data sources.
    - Use of knowledge representation techniques (ontologies) for the representation of data sources and their access and integration.
    - Parallel processing techniques: MapReduce (Hadoop).
    - Data *stream* management systems.
    - Other techniques: mobile agents.
- Interaction with large volumes of data:
    - Visualization techniques.
    - Design of appropriate user interfaces.
    - Usability.
- Analysis of large volumes of data:
    - Data mining.
    - Text mining, sentiment analysis.
- Use cases and applications, such as:
    - Data coming from sensors.
    - Unstructured data on the Web.
    - Recommendation systems.

- Analysis of blogs and social networks.
- *Smart cities*
- Intelligent transportation systems.

4. Academic activities

The subject consists of 6 ECTS which represent around 150 hours of student´s dedication, distributed as follows:

- Activities directed by the subject's teaching staff: 60 h (master class, problem and case solving, laboratory practices, special practices)
- Completion of practical application or research work: 65 h
- Theory study: 20 h
- Assessment tests: 5 h

5. Assessment system

The student must demonstrate that they has achieved the expected learning results through the following assessment activities:

- **Production and presentation of work** . Study of a topic related to the subject, preparation of a report on the same, and presentation in class. [20%]. Learning results: 1, 2 and 3
- **Project** . A group project with computers, in which the knowledge and skills acquired in the subject will be put into practice. [50%]. Learning results: 1, 2 and 3
- **Final written test including short answer and long answer questions** . [30%]. Learning results: 1, 2 and 3

In order to average all the grades, it will be necessary to individually pass each of the evaluation activities.

The student who does not choose the evaluation procedure described above, does not pass these tests during the teaching period, or who would like to improve their grade, will be entitled to a comprehensive test, which will be scheduled within the exam period corresponding to the first or second call, as appropriate. This comprehensive test will consist of the following evaluation activities: 1) completion and presentation of the previously indicated work (study of a subject) [20%], 2) completion and defence of the project (subject practices) [50%], and 3) completion of a final written test including short answer and long answer questions [30%]. In order to average all the grades, it will be necessary to individually pass each of the evaluation activities.

6. Sustainable Development Goals

10 - Reduction of Inequalities