

Big Data

Academic Year: (2022 / 2023)

Review date: 20-05-2022

Department assigned to the subject: Computer Science and Engineering Department

Coordinating teacher: CALLE GOMEZ, FRANCISCO JAVIER

Type: Compulsory ECTS Credits : 6.0

Year : 1 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

- Structured Databases
- Algebraic Data Languages
- SQL
- OLAP Databases and Data Warehouse

Programming skills
(desirable basics of Javascript)

OBJECTIVES

Basic and general competences

- The students should apply their knowledge and their ability to solve problems in new or unfamiliar environments within broader (or multidisciplinary) contexts related to their field of study
- Students should communicate their conclusions and the knowledge and rationale, to specialists and non-specialists in a clear and unambiguous
- Students must possess the learning skills that enable them to continue studying in a self-directed or autonomous way
- Ability to understand and apply methods and techniques in the field of Computer Engineering in financial markets
- Ability to conceive, design or create, implement and adopt a substantial process of developing and creating software for financial markets

Specific skills

- Analyze and evaluate the feasibility of implementing a data management system according to the needs
- Analyze and understand the main tools for managing large amounts of data storage, access and review

Learning outcomes are determined by both the contents of the subject as the framework main of this matter which is Systems decision support in the financial sector.

DESCRIPTION OF CONTENTS: PROGRAMME

Block I: Theoretical Foundation.

Item 1: Introduction: Social and technological framework

- The IT Society
- Current role of information and data
- Storage paradigms
- Characterization of the Big Data concept

Item 2: Approach to Big Data

- Transactional vs. Analytical databases
- Physical organizations suited to the process
- Architectures: distributed systems and CAP.
- ROLAP warehouses. Analytical operation in SQL.

Item 3: Integration, transformation and Cleaning

- Integration of sources
- Transformation and Cleaning

- Google Refine
- SPARQL

Block II: Implementing Big Data

Item 4: BigData operability

- The Map-Reduce paradigm
- Legal and ethical aspects: Privacy and Security

Item 5: Back-End for BigData I: MongoDB

- Introduction to Mongo DB.
- Basic operability in MongoDB
- Aggregation in MongoDB. Pipeline and Map-Reduce.
- Replication and Distribution in MongoDB

Item 6: Back-End for BigData II: Cassandra

- Cassandra's Basics
- Design on Cassandra

Item 7: Back-End for BigData III: Hadoop

- The HADOOP ecosystem and its installation
- SandBox
- HADOOP functionality
- Map-Reduce in HADOOP

LEARNING ACTIVITIES AND METHODOLOGY

Theoretical classes: presentations accompanied by digital supporting materials.

Theoretical practical classes: Combination of theoretical classes accompanied by the resolution of practical exercises.

Laboratory practices: Practices to be developed in specific laboratories for the different subjects.

- Total number of hours: 23 Face-to-face: 100%

Tutorials: Face-to-face and / or distance tutorials (videoconference): Total number of hours: 7 Face-to-face: 100%

E-learning activities: tutorials, recommended reading, documentation: Total number of hours: 50 Face-to-face: 0%

Individual student work: Individual student activities that complement the rest of the activities (both face-to-face and non-face-to-face), as well as exam preparation: No. Total hours: 70 Face-to-face: 0%

Teaching methodologies

- Exhibitions in class by the teacher with the support of computer and audiovisual media, in which the main concepts of the subject are developed and the bibliography is provided to complement the students' learning.
- Critical reading of texts recommended by the professor of the subject: press articles, reports, manuals and / or academic articles, either for later discussion in class, or to expand and consolidate knowledge of the subject.
- Resolution of practical cases, problems, etc. raised by the teacher
- Exhibition and discussion in class, under the moderation of the teacher of topics related to the content of the subject, as well as practical cases
- Preparation of work and reports individually or in groups
- Specific e-learning activities, related to the semi-face-to-face nature of the degree, self-correction activities, participation in forums, and any other online teaching mechanism

ASSESSMENT SYSTEM

Two workload will be proposed that covering the course contents.

Worklab: 80%

Exam: 20%

% end-of-term-examination: 20

% of continuous assessment (assignments, laboratory, practicals...): 80

BASIC BIBLIOGRAPHY

- Apache¿ Hadoop® <http://hadoop.apache.org/>, Apache¿ Hadoop®, 2016

- MongoDB <http://www.mongodb.org>, MongoDB, 2016

BASIC ELECTRONIC RESOURCES

- Google Refine . Tutorial: [href="https://www.irekia.euskadi.net/assets/attachments/2421/Tutorial4_IntroduccionGoogleR](https://www.irekia.euskadi.net/assets/attachments/2421/Tutorial4_IntroduccionGoogleR)
- MongoDB . Documentación MongoDB: <http://docs.mongodb.org/manual/>
- Sandbox . Documentación Sandbox: <http://hortonworks.com/products/hortonworks-sandbox/>