Big Data

Academic Year: (2020 / 2021)   **Review date:** 16/09/2020 22:55:03

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

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 MongoBD
- Aggregation in MongoBD. Pipeline and Map-Reduce.
- Replication and Distribution in MongoBD

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

Theory class: presentation with digital support of the basic content of the course. Hours: 22
Worklabs + problems: Application of theory to several big data software. Hours: 28 + 18
Tutorial: both face to face and video-conference
E-learning activities: participation in the activities that the teacher proposes.

ASSESSMENT SYSTEM

% end-of-term-examination/test:	20
% of continuous assessment (assigments, laboratory, practicals…):	80

Two workload will be proposed that covering the course contents.
Worklab: 80%
Exam: 20%

BASIC BIBLIOGRAPHY

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

- MongoBD http://www.mongodb.org, MongoBD, 2016

BASIC ELECTRONIC RESOURCES

- Google Refine . Tutorial: <a
href="https://www.irekia.euskadi.net/assets/attachments/2421/Tutorial4_IntroduccionGoogleR

- MongoBD . Documentación MongBD: http://docs.mongodb.org/manual/

- Sandbox . Documentanción Sandbox: http://hortonworks.com/products/hortonworks-sandbox/