

Academic Year: (2023 / 2024)

Review date: 14-05-2023

Department assigned to the subject: Computer Science and Engineering Department

Coordinating teacher: CALLE GOMEZ, FRANCISCO JAVIER

Type: Compulsory ECTS Credits : 3.0

Year : 1 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

- Programming skills
- Experience in Windows environment

Yet not required, it is also desirable to meet:

- basics of structured databases (at least, Relational Model)
- notions of relational algebra (or at least, set theory)
- basics of data languages (specially SQL)
- basics of Javascript

All of these will be introduced in class, and further materials provided so the student can catch up (or at least, get introduced) at home.

OBJECTIVES

This course aims are

- To distinguish different approaches to secondary storage, depending upon the needs, but focusing on storage with analytic purposes.
- To learn the wide diversity of solutions, and specifically to get introduced to some of the most widespread tools supporting Big Data implantations.

Agenda will cover from the information acquisition and preparation to the manipulation on some DBMS
Therefore, it is largely characterized by a practical focus

To achieve these goals, the student must acquire a set of generic capabilities, knowledge, skills and attitudes.

Cross/Generic Capabilities

- o Analysis and synthesis abilities
- o Organize and plan abilities
- o Troubleshooting
- o Ability to apply knowledge in practice

Specific Capabilities

Cognitive (Knowledge): Storage paradigms, Information Lifecycle, back-end solutions for Big Data

Procedural/Instrumental (Know how)

- Information acquisition and preparation
- Data manipulation (by different languages) on diverse DBMS:
 - Structured DBMS
 - Document oriented DBMS
 - Column oriented DBMS

Attitudinal (To be):

ability to design queries (creativity), concerns about the effectiveness and the efficiency, and ability to discuss and clarify the diverse solutions to each specific problem

DESCRIPTION OF CONTENTS: PROGRAMME

ITEM 1. Storage Paradigms

- Introduction to Storage: archives and files
- Databases and DBMS
- Evolution of Storage and DBMS: OLTP vs. OLAP

- Massive storage: ROLAP vs. RTOLAP

ITEM 2. Structured Storages

- Relational databases
- Data Warehousing
- Analytical Processing on Relational DB

ITEM 3. Information Acquisition and preparation

- Acquisition and extraction
- Transformation, Cleaning & Integration

ITEM 4. DBMS Supporting Big Data: MongoDB (document-oriented)

ITEM 5. Introduction to other Back-End DBMS: Cassandra and Neo4J

LEARNING ACTIVITIES AND METHODOLOGY

Lectures:

Highly practical lectures, most will take place at a computer room.

Most tools involved in the course are free, so the student is encouraged to practice with them (after classes).

It is an ambitious (and large) syllabus: some further details will be left for the student to review at home (complementary lectures).

Assignments:

Except from first item (introductory and of theoretical nature) every block will end up with an assignment, for the student to solve at home.

Practical exercises (similar to those in the assignment) will be proposed in the classroom.

Tutoring sessions:

Individual tutoring sessions will be provided for solving eventual doubts, so the student can catch up with the group in case some parts are not found to be completely understood.

ASSESSMENT SYSTEM

During the course, different practical (out of the classroom) assignments will be carried out, both individually and in groups.

Each Item in the program (from 2nd to 5th) will end with an assignment. Each has a weight of 15% in the global assessment (except from item 4, whose weight will be double that).

Final test (8th week) to assess global learning.

% end-of-term-examination:	25
% of continuous assessment (assignments, laboratory, practicals...):	75

BASIC BIBLIOGRAPHY

- J. Calle Course Teaching Materials (provided via Aula Global webpage) Each item will have specific references (mostly, links to webpages where documentation on tools usage, syntax, etc. can be freely accessed), Aula Global, 2018

ADDITIONAL BIBLIOGRAPHY

- Elmasri, R. y Navathe, SB Fundamentals of Database Systems. , Pearson .
- Hurwitz, J, Nugent, A, Halper, F, Kaufman, M. Big data for dummies , Wiley, 2013
- Ramakrishnan, R. y Gehrke, J. Database management systems. , McGraw Hill. .
- Warden P. Big Data glossary. A guide to the New Generation of data tools. , O'Reilly , 2011
- null Express Learning: Database Management Systems. , ITL Education Solutions Lt. Pearson India Pubs, 2012
- Rijmenam, M.V. Think Bigger (ISBN-13: 978-0-8144-3415-4), Amacom, 2014

BASIC ELECTRONIC RESOURCES

- Oracle Corp . SQL Language Quick Reference: <https://docs.oracle.com/en/database/oracle/oracle-database/12.2/sqlqr/index.html>
- Oracle Corp. . Oracle SQL*Plus Quick Reference: <https://docs.oracle.com/en/database/oracle/oracle-database/12.2/sqlqr/index.html#SQPQR101>
- © 2022 DataStax Inc. . CQL 3.0 Reference for Apache Cassandra: <https://docs.datastax.com>.

com/en/archived/cql/3.3/cql/cql_reference/cqlReferenceTOC.html

- © 2022 MongoDB, Inc. . Mongo DB Manual and Documentation:
<https://www.mongodb.com/docs/manual/introduction/>

- © 2022 MongoDB, Inc. . Mongo DB Glossary of terms and concepts:
<https://www.mongodb.com/docs/manual/reference/glossary/>