II.2314 / II.2414 – Databases and Big Data

General information

Title: Databases and Big Data
Module ID: II.2314 / II.2414
Module leader: Maurras TOGBE
ECTS: 4
Average amount of work per student: from 100 to 150 hours, including 42 hours supervised
Teamwork: yes
Keywords: Relational databases, SQL, PL/SQL, Oracle, JDBC, Big Data, NoSQL, MongoDB, Neo4j, Elasticsearch

Presentation

Implementing complex systems often requires the creation, use, or consolidation of structured data for backup, search, or interaction with other systems. Beyond information systems, all fields potentially have such a need. This module provides a set of skills that complement those acquired in the first year with the aim of designing and administering relational and non-relational databases. These are designed to be able to process the masses of data that are generated by machines, users, sensors, a phenomenon known by Big Data and which is becoming the essential subject of the moment.

Educational objectives

- Design a database
- Master the logical and physical architecture elements of a relational database.
- Make optimized and adapted technical choices according to the application cases
- Technology and bibliography survey

Prerequisite

Relational database and Java basics for the JDBC part

Content/Program

- Relational Model (Relational Algebra, Normal Form, PL/SQL Procedural Language)
- Transaction integrity and management
- Indexing
- JDBC
- Massively parallel processing
- Big Data: The 4Vs (Volume, Velocity, Variety, and Veracity)
- The NoSQL ecosystem

Tools used

- Java Development Environment
- Oracle: SQL and PL/SQL
- Connecting to Oracle via JDBC
- Docker
- MongoDB, Neo4J, Cassandra, Elasticsearch, and Kiabana

Pedagogical methods

Learning methods

Each week will be organized as follows: classes or practical work. Part of the practical work will be used to implement and administer an Oracle database. A practical session will be devoted to the use of JDBC to connect a Java program to the database. The rest of the practical sessions will be devoted to NoSQL comics of different types. Students will be asked to conduct research in teams on a topic related to advanced databases that they will present in a final report.

Evaluation methods

The evaluation of this module is composed of:

- An individual evaluation during the 4th grade (10%)
- Participation and finalization of practical work (20%)
- A practical exercise in pairs during the 6th session (10%)
- A practical exercise in pairs during the 14th session (10%)
- An individual book-closed evaluation during the exam week (50%)

Language of work

Materials, courses and assessments are presented in English. Deliverables and evaluations can be written in French or English.

Bibliography, Webography, Other sources

- Books
  - Raghu Ramakrishnan, Johannes Gehrke – Database Management Systems – Third edition
  - Georges Gardarin - Databases - Eyrolles (Paperback)
  - Andreas Meier - A Practical Introduction to Relational Databases - IRIS Collection
  - Big Data: Principles and best practices of scalable Realtime data systems, Nathan Marz and James Warren, Manning, 2014
  - Principles of Distributed Databases (3rd edition), M. T. Oszu and P. Valduriez, Addison-Wesley Longman, 2011
  - Seven databases in seven weeks, Eric Redmond and Jim R. Wilson, Pragmatic Bookshelf, 2012
- Webography:
  - http://infolab.stanford.edu/~ullman/fcdb/oracle/or-plsql.html (PL/SQL)
  - http://docs.oracle.com/cd/B28359_01/appdev.111/b28370/toc.htm
  - http://georges.gardarin.free.fr/
  - http://codex.cs.yale.edu/avi/db-book/db4/index.html