

Databases and Big Data

Prerequisites II.1101

Organization: Lectures + Tutorials + labs + presentations

Assessment: Written examination (50%) + Presentation (20%) + Reports (30%)

Credits: 5 ECTS

Background

Complex systems implementation often requires creation, use or consolidation of structured data in order to save, search, or interact with other systems. This module provides a set of skills that complement those acquired during the first-year and aims to design and manage relational and non-relational databases. These are designed to process massive data that are generated by machines, users, sensors, etc. Phenomenon known by Big Data, which emerged as the unavoidable subject of the moment

Objectives

Skills

The module aims to prepare future engineers to design and administrate relational database.

Emphasis is placed on:

- The use of a methodology for designing database.
- Control of physical and logical architecture of a relational database.
- The process of optimizing a database.
- The administration of a database.
- Managing concurrency.

The module discusses the limitations of the relational model and introduces the post-relational models, in particular NoSQL databases

Knowledge

Concepts

- Relational model
 - Model entity/relationship (E/R)
 - Relational Algebra
 - Normal form
 - SQL Language
- Integrity and Transaction
- Big Data : The 4Vs (Volume, Velocity, Variety and Veracity)
- Massively Parallel Processing: MapReduce
- NoSQL/NewSQL Ecosystems

Know-how

- Conceptual Data Modeling: Model E/R
- Implementation rules for passage from E/R Model to relational model
- Design a normalized relational schema
- Formulation of queries in relational algebra
- Implementation of an Oracle database and using SQL and PL/SQL: development, management, concurrency, etc.
- Administration and optimization of a database
- Installation and deployment of NoSQL database
- Write queries following the MapReduce paradigm
- Modeling a NoSQL database

- Querying massive data
- Handling massive data

Teaching approach

The module will run in 14 lectures (3h/lecture). Tutorials start from the second session of lectures and are used to practice modeling and designing databases. Labs sessions are used to implement and administer an Oracle database and also to practice on different types of NoSQL databases. Students will conduct research on a topic related to advanced databases and will make a presentation in group during the last week of the module.

References

- Bases de données objet & relationnel. Paris: Eyrolles. Gardarin, G.(1999).
- Conception des bases de données relationnelles - En pratique (Vuibert, 2001), AKOKA, COMYN-WATTIAU
- Gestion et administration des bases de données (Dunod, 2003), BOUJLIDA
- Date, CJ (2000). An Introduction to Database Systems (7th ed.). Reading, MA: Addison-Wesley.
- http://georges.gardarin.free.fr/Cours_Total/IndexTotal.html
- Big Data:Principles and best practices of scalable realtime data systems. Nathan Marz, James Warren