Business Data Management (33:136:470)

Instructor: Spiros Papadimitriou

This course introduces principles and techniques for managing corporate data resources, covering the principles and functions of Relational Database Management Systems (RDBMS), and their use as a business tool. Topics covered include an overview of database management and its history, logical data models (entity-relationship models, relational schema design, and table normalization), relational database creation and administration, SQL queries, database design and normalization, ACID, and basics of database application development. The course will also survey advanced database management topics such as "big data" and NoSQL technologies, query processing and optimization, and data warehouses, as time permits. Students will also gain hands-on experience using current database technologies, and are expected to complete both written and programming homework assignments, on database infrastructure that is provided to them (MySql and Apache/PHP). Upon completion, students will have a solid understanding of database management fundamentals and how they are applied in modern practice.

List of topics

- Introduction
- Concepts and models
  - Entity-relationship (E-R) models
  - Relational model
  - Relational algebra
- SQL
  - Introduction
  - MySQL tutorial
  - Joins
  - Aggregation
  - (Advanced topics: division, skyline queries)
- File organization and indexes
  - Storage hierarchies and block-based access
  - Record formats, variable- vs. fixed-length fields
  - Tree-based indices (concepts)
  - Hash-based indices (concepts)
- Schema refinement and normalization
  - Functional dependencies
  - Database normalization (BCNF and 3NF)
  - Database design methodology
- Transactions and concurrency control
  - ACID
  - Serializability
  - Two-phase locking (2PL)
- Database application programming[*]
  - PHP and HTML tutorial

- ○ Web application development concepts
- Advanced topics (time permitting; subject to change)
  - ○ "Big data" and NoSQL
  - ○ Physical design (concepts in query optimization and tuning)
  - ○ Data warehousing and OLAP
  - ○ Extract-Transform-Load (ETL)
  - ○ Object-relational mapping (ORM)[*]
  - ○ Semi-structured data, unstructured data, and information retrieval[*]

Textbook

No required textbook – suggested reading is one of:
- "Fundamentals of Database Systems (6th ed.)", Ramez Elmasri and Shamkant B. Navathe, Addison Wesley, 2011, *or*
- "Database Management Systems", Raghu Ramakrishnan and Johannes Gehrke, McGraw-Hill, 2002.

and optionally:
- "Introduction to Database Systems", Stephane Bressan and Barbara Catania, McGraw-Hill, 2005.

Grading policies

40% Homework assignments
30% Midterm
30% Final

[*]These topics are covered in more detail in 33:136:440.