IV1351 Data Storage Paradigms 7.5 credits

Datalagring

This is a translation of the Swedish, legally binding, course syllabus.

Establishment

Grading scale

A, B, C, D, E, FX, F

Education cycle

First cycle

Main field of study

Information Technology, Technology

Specific prerequisites

Knowledge and skills in programming, 6 credits, equivalent to completed course ID1018/ID1022/DD1337/DD1310-DD1319/DD1321/DD1331/DD1333/DD100N.

Intended learning outcomes

After passing the course, the student should be able to

- explain fundamental concepts, principles and theories in databases and data storage

- design a relational database based on a requirements specification using conceptual modeling, ER diagrams and normalized relational schemas
- model data and formulate queries using the relational model and SQL
- analyze and normalize database schemas using normalization theory
- describe and compare techniques for physical storage management, indexing, transaction management and query processing and query optimization in database systems

in order to

- understand and reason about the structure and function of database systems
- be able to independently analyze information and translate it into a correct database design
- be able to design and use databases for storing and retrieving data
- be able to design well-structured and maintainable databases
- be able to assess and select appropriate technical solutions for given performance requirements.

Course contents

- Introduction to databases, data storage, and information administration
- The relational model, relational algebra and relational calculus
- The SQL query language: data definition, data manipulation and query formulation
- Normalization theory: functional dependencies, closures, Armstrong's axiom, canonical cover, Heath's theorem, third normal form (3NF) and Boyce-Codd normal form (BCNF)
- Storage management: the storage hierarchy, hard disk drives and SSDs, sequential access optimization, failure handling and RAID (levels 1–5), buffer management with page tables and replacement policies, database file organization (slotted pages, heap files), and archiving
- Transaction management: ACID properties, concurrency control, conflict and view serializability, anomalies (lost update, blind writes), exclusive and shared locks, two-phase locking (2PL, SS2PL, C2PL), wait-for graphs and deadlocks
- Index structures: primary indexes, secondary and clustering indexes, hash tables (extendible and linear hashing), B+trees
- Query processing and query optimization: query execution, blocking and non-blocking operations, index scans, selectivity estimation, external merge sort, join algorithms (nested loop, block nested loop, index-based, sort-merge), predicate and projection pushdown, query rewriting and cost model-based optimization

Examination

- TEN1 - Examination, 3.0 credits, grading scale: A, B, C, D, E, FX, F
- LAB1 - Laboratory Works and Project, 4.5 credits, grading scale: A, B, C, D, E, FX, F

Based on recommendation from KTH’s coordinator for disabilities, the examiner will decide how to adapt an examination for students with documented disability.

The examiner may apply another examination format when re-examining individual students.

If the course is discontinued, students may request to be examined during the following two academic years.

TEN1 is examined through written exam and graded quizzes. LAB1 is examined through a comprehensive database project that integrates all course topics.

Ethical approach

- All members of a group are responsible for the group's work.
- In any assessment, every student shall honestly disclose any help received and sources used.
- In an oral assessment, every student shall be able to present and answer questions about the entire assignment and solution.