
CSE 444 Database Systems Internals

Credits

4.0 (3 hrs lecture, 1 hr section)

Lead Instructor

Magdalena Balazinska

Textbook

- *Database Systems*, Garcia-Molina

Course Description

Internals of a relational database management systems (DBMS). Architecture of a DBMS. Data storage and indexing. Query execution, buffer management, operator algorithms. Query optimization. Transactions: recovery and concurrency control. Parallel query evaluation. Distributed transactions. Data replication. NoSQL and NewSQL systems. Students build a medium size DBMS in Java.

Prerequisites

CSE 332; CSE 344.

CE Major Status

Selected Elective

Course Objectives

This course teaches the internals of relational database management systems and how to build such systems. The course covers the full stack. Starting from a quick review of the relational model and SQL, the course dives into the architecture of a DBMS and then goes in depth into each component: storage manager, query executor, query optimizer, transactions, parallel processing, distributed transactions, data replication, architecture of NoSQL and NewSQL systems, etc. Students get to implement each of the components in their own Java DBMS. At the end of the quarter, successful students have a working DBMS that can execute SQL queries, update data, run concurrent transactions, recover from failures, and either execute in parallel or optimize queries.

ABET Outcomes

- (1) an ability to identify, formulate, and solve complex engineering problems by applying principles of engineering, science, and mathematics
- (3) an ability to communicate effectively with a range of audiences

(6) an ability to develop and conduct appropriate experimentation, analyze and interpret data, and use engineering judgment to draw conclusions

(7) an ability to acquire and apply new knowledge as needed, using appropriate learning strategies

Course Topics

- data models
- conceptual design
- query languages
- system components
- data storage
- query optimization
- transaction processing.
- parallel, shared-nothing processing
- distributed transactions
- data replication