CSE 444: Database Internals

Lectures 27
Wrap-up

Final Exam

- Thursday, June 07 at 8:30am in class
- Open book and open notes
  - Our book
  - Any other book
  - Course notes
  - Old quizzes
  - Answers to practice problems
  - Anything that is on paper!

Pre-Final Office Hours

- Monday 10am-12pm
- Tuesday 10am-12pm

Final Exam Content

- Everything that we covered in lecture, in the web quizzes, or in the labs
- Let's go over the main topics again

Relational Model

- Lecture 2
  - Relational model
  - Relational algebra
  - SQL
- No explicit questions on this content
  - Relational algebra is in questions about other topics
  - Same for SQL, it appears in various questions

DBMS Architecture and Deployments

- Lecture 3
  - Logical and physical data independence
  - Different DBMS deployments
  - Key components of a DBMS
- Expect qualitative questions on these topics

Data Storage and Indexing

- Lectures 4 through 6
  - Data storage on disk: heap files, sequential files
  - Buffer manager
  - Indexes
    - B+ trees / Hash table
    - Clustered / unclustered
    - Dense / sparse
    - Primary / secondary

Query Execution

- Lectures 7 through 9
  - Translating SQL into relational algebra
  - Physical and logical query plans
  - Pipelined query execution
  - One-pass operator algorithms
  - Index-based operator algorithms
  - Two-pass operator algorithms

Query Optimization and DBMS Tuning

- Lectures 10 through 13
  - Estimating the cost of a query plan
  - Algebraic rewrite rules
  - Bottom-up, dynamic prog. plan selection algorithm
    - Variant that operates on *logical* plans
    - Selinger-style variant
  - Index selection problem
  - Other physical tunings: why and how

Transactions

- Lectures 14 through 19
  - ACID properties
  - Concurrency control
    - Using locking
    - Using timestamps or validation
    - Different levels of isolation, including snapshot isolation
  - Recovery
    - UNDO logs
    - REDO logs
    - ARIES (undo/redo log)

Distributed and Parallel DBMSs

- Lectures 20 through 25
  - Distributed query processing
  - Distributed query optimization
  - Distributed transactions: 2PC
  - Replication: all the different types
  - Parallel DBMSs overview
  - Parallel operator algorithms
  - MapReduce implementation internals

NoSQL

- Lectures 26
  - Motivation
  - The various types of systems
  - Similarities and differences with parallel DBMSs
  - Relaxed consistency