Final Review

Lecture 18

Outline

1. Final Exam

2. What We Did **Not** Cover

Format

- 4 problems, with multiple sub-parts
  - Full 2 hours (estimated at 1.5 hours)

- No notes, calculators, books, computers, phones, etc. may be used

Content (1)

1. Database Design
   a) Given a narrative, design a database (i.e. words -> logical design)
   b) Evaluate a db design in the context of FDs; justify and implement decomposition to 3NF

2. SQL (Programming)/Security
   a) Given db state + SQL, predict output
   b) Given schema + text, produce SQL
   c) SQL Injection: meaning, prevention
   d) Secure password storage

Content (2)

3. Physical Design
   a) Given a query (load) and contextual information (e.g. db structure, selectivity, frequency), diagnose performance and recommend an appropriate intervention

4. Transactions
   a) Concurrent transactions: characterize, conflict-equivalent serial schedules

+ bonus via advanced topics

The Landscape of Databases

- **Using an RDBMS**
- Conceptual/Logical design
- Physical design
- How an RDBMS works
- Non-Relational models
- Scaling
- Data analytics

Using an RDBMS

- DBMS diversity
  - DB2, Oracle, SQL Server
- Advanced database programming
  - Views
  - Stored procedures, triggers
  - Cursors, ORM
  - Data formats: CSV, XML, JSON
  - [Web] APIs
- Reporting
  - Analytics

Conceptual/Logical Design

- ERDs
  - Multi-way relationships
  - Other dialects
  - UML
- Normal forms
  - BCNF, 4NF, 5NF, 6NF

Physical Design

- Storage, details of …
  - Cost models
  - File organization
  - Buffering
- Indexing, details of …
  - Hashing
  - Trees
  - Specialized

How an RDBMS Works

- Recovery
  - Storage hierarchy, logging
- Query evaluation/optimization
  - Join processing, sorting
- Distributed databases

Non-Relational Models

- XML, Object-Relational
- Spatial
- Time series
- Probabilistic

Scaling

- Distributed Databases
  - OLAP
- Cloud computing
  - MapReduce
- Crowdsourcing
  - Mechanical Turk
  - Human computation (e.g. reCAPTCHA, Duolingo)
- Security
  - Differential privacy
  - Regulation (e.g. HIPAA), provenance

Data Analytics

- Ranking (e.g. PageRank)
- Data Warehouses
  - Data cube
- Natural Language Processing (NLP)
  - Entity matching
- Human-Computer Interaction (HCI)
  - Visualization
  - [Facilitated] exploration
  - Explainable queries

Summary

- You have learned the basic flow of designing a database (for correctness, performance, and security), using an existing database, and developing a database-enabled application – well done!
  - Best of luck on Exam 3 :)

- If you are interested in topics we did not cover, consider taking a relevant class (e.g. Data Science, Machine Learning, Web Development, DB Seminar)
  - Always feel free to come by my office and chat/work on a cool project :)