CS 4604: Introduction to Database Management Systems

B. Aditya Prakash

Lecture #25: Final Review
Final Exam

- 25% of the grade
- **No** books, no notes, no laptops
- **Only** 1 letter-size page
  - you can use both sides

- Duration: 2 hours. 1:05-3:05pm, May 11 2013
  Location: regular classroom
Syllabus

- **Comprehensive exam**
  - But bias towards and emphasis on post-midterm stuff
  - Will cover all material in all lectures
    - except supervised learning in data mining and PHP
**Extra office hours**

– By me: on Friday 9-11am at 122C. And by appointment at my office (KWII 2223).

– By Qianzhou: on Friday 1-3pm.
What you have learnt in the course

- Weeks 1–5, 13: Query/Manipulation Languages
  - Relational Algebra
  - Data definition
  - Programming with SQL

- Weeks 6–8: Data Modeling
  - Entity-Relationship (E/R) approach
  - Specifying Constraints
  - Good E/R design

- Weeks 9–13: Relational Design
  - The relational model
  - Converting ER to “R”
  - Normalization to avoid redundancy

- Week 14–15: Students’ choice
  - Practice Problems
  - XML
  - Query optimization
  - Transactions
  - Data mining
Mantra

- Know all the definitions!
  – Once you know what the terms really mean and imply, many details can be worked out on the fly
SQL/RA

- Make sure you know all the operators/constraints/triggers/checks etc. etc.
  - SQL and RA
  - Set-semantics/Bag-semantics
  - Extended operators
    - Outerjoins, left-outerjoins, right-outerjoins
    - Distinct, aggregators, group-by in RA
    - ....

- The base for DB
ER

- You should already have enough practice!
FDs/MDs

- Definitions of FDs, MDs, closures, cover, normal forms, decompositions etc. etc.
  - Pay attention to multiple ways of defining the same thing!

- Various procedures to compute the above

- Chase process

- LOTs of practice (understand the HWs/Handouts/Practice problems)
Indexing

- B-Trees
  - definitions
  - Insertion/deletion algorithm
- Improved versions of B-trees
  - B*-trees
  - B+-trees
  - Note the differences between them and the B-tree
Query Optimization

- Algebraic manipulation
- Selectivity estimation
  - Many cases
  - Even for joins
- Join selection
Transactions

- ACID
- Problems with concurrency and Serializability concept
- Conflict-Serializability
- 2PL
- Strict 2PL
- Crash recovery
Data Mining and Warehousing

- Skip supervised learning for exam
- Warehousing
- Mediation
- Unsupervised Learning
- ....
Essentially learnt how to create and understand XML docs

Xquery/Xpath: query languages for XML

More examples in the text...
Data Management

- Is a really exciting field (‘BIG-Data’)
- High commercial *and* academic research interest
Lots more stuff we did not cover

- Storage Manager
  - File organization
- More details about query processing
  - E.g. Join algorithms
  - Sorting
- Other powerful query languages
  - Datalog etc.
- More details about indexing
  - E.g. Hashing
- More sophisticated locking, concurrency control
  - E.g. Hierarchical locking
- Logging and Recovery
  - We did not cover ARIES: the state-of-the-art logging/recovery system
- Spatial Data Management
- Distributed Databases
- More advanced data mining
- ..................
Course Plug: Data Mining Large Networks

Facebook Network [2010]

Gene Regulatory Network [Decourty 2008]

Human Disease Network [Barabasi 2007]

The Internet [2005]
Course Plug

- CS 6604: Data Mining Large Networks in Fall 2013
  - Graduate level course
  - Project, research papers
  - Would be exciting and fun!
  - Good way to get exposed to the state-of-the-art in network analysis, graph databases etc.
Good Luck!

FINAL EXAMS

YES WE CAN