CS 4604: Introduction to Database Management Systems

B. Aditya Prakash

Final Review
Final Exam

- 30% of the grade
- No books, no notes, no laptops

Allowed:
  - Only 2 letter-size pages
    - You can use both sides
    - Must be hand-written
  - And a calculator (recommended)

- Duration: 2 hours. 7:45-9:45am, May 10 2014
- Location: regular classroom
Comprehensive exam

- But main focus towards and emphasis on post-midterm stuff (= starting from lecture 11)
- Will cover all material in all lectures
- EXCEPT (i.e. things not in exam)
  1. NoSQL/MapReduce
  2. Semi-structured data/XML
  3. Data Mining/Warehousing
     (No PHP too of course)
Office Hours this week

- **By Aditya:**
  - Tuesday: 2-3:15pm
  - Thursday: 2-4pm
  - By appointment. (all at my office)

- **By Qianzhou:**
  - Monday 1:30-3:30pm McB 106
  - Wednesday 1:30-3:30pm McB 106
  - Thursday 12:30 PM - 2:00 PM McB 106
  - Friday 11:00 AM - 12:30 PM McB 106

- **By Pranav:**
  - Thursday 3:30 PM - 5:00 PM McB 106
  - Friday 1-3PM McB 106
OVERVIEW
What you learnt in the course

- **Weeks 1–4: Query/Manipulation Languages and Data Modeling**
  - Relational Algebra
  - Data definition
  - Programming with SQL
  - Entity-Relationship (E/R) approach
  - Specifying Constraints
  - Good E/R design

- **Weeks 5–8: Indexes, Processing and Optimization**
  - Storing
  - Hashing/Sorting
  - Query Optimization
  - NoSQL and Hadoop

- **Week 9-10: Relational Design**
  - Functional Dependencies
  - Normalization to avoid redundancy

- **Week 11-12: Concurrency Control**
  - Transactions
  - Logging and Recovery

- **Week 13–14: Students’ choice**
  - Practice Problems
  - XML
  - Data mining and warehousing
data

meta-data

naive

app. pgmr

casual

DBA

users

emb. DML

DML proc.

DDL int.

app. pgm(o)

query eval.

query proc.

trans. mgr

buff. mgr

file mgr

storage mgr.
SQL/RA

- Make sure you know all the operators for SQL and RA
  - Select, From, Where, Group-by, Having, Order-by
  - Set-semantics/Bag-semantics
- The base for DB
ER

- You should already have enough practice!
FDs

- Definitions of FDs, s closures, cover, normal forms, decompositions etc. etc.
  - Pay attention to multiple ways of defining the same thing!
  - E.g. ‘Key’: multiple ways of defining and understanding

- Various procedures to compute the above
Indexing and Hashing

- Know your basic structure, and definitions
- Less emphasis (as we have covered this in the midterm)
Query Processing

- Estimating costs
  - What are you estimating? = #disk accesses
  - How to estimate?
    - sorting
    - Different types of joins (NLJ, Block-NLJ, SMJ, HJ)
    - Don’t just memorize the formulae, understand how they are derived, the ‘best-case’ ‘worst-case’ scenarios
Query Optimization

- Algebraic manipulation
- Selectivity estimation
  - Many cases
Transactions

- ACID
- Problems with concurrency and Serializability concept
- Conflict-Serializability, how to detect
- 2PL, when, why, what, how
- Strict 2PL, when, why, what, how
- Know your venn diagrams!
- Deadlocks, how to detect and avoid them
Logging and Recovery: Big Picture

LogRecords

prevLSN
XID
type
pageID
length
offset
before-image
after-image

Data pages
each with a
pageLSN

master record
LSN of most
recent checkpoint

Xact Table
lastLSN
status

Dirty Page Table
recLSN

flushedLSN

LOG

DB

RAM

update
CLR

undoNextLSN

CLR

undoNextLSN

Prakash 2014

VT CS 4604
Crash Recovery: Big Picture

- Start from a **checkpoint** (found via **master** record).

- **Three phases.**
  - **Analysis** - Figure out which Xacts committed since checkpoint, which failed.
  - **REDO** all actions (repeat history)
  - **UNDO** effects of failed Xacts.
Crash Recovery: Big Picture

- Oldest log rec. of Xact active at crash
- Smallest recLSN in dirty page table after Analysis
- Last chkpt
- CRASH

• Notice: relative ordering of A, B, C may vary!
Logging and Recovery

- Make sure you know *exactly* how recovery takes place, and what is logged
  - practice
Tips

- Know your definitions!
  - Different ways of defining same thing e.g. keys

- Go through the slides
  - Checking the textbook if you are unclear

- Go through HWs, Handouts, Exams, and Practice problems
  - Textbook also has good problems! Even numbered problems have solutions on-line
  - Take advantage of our office hours

- Make use of your 2 allowed written notes!

- Bring a calculator
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
  - Fine-tuning Join algorithms
- Other powerful query languages
  - Datalog etc.
- More sophisticated locking, concurrency control
  - E.g. Hierarchical locking, time-stamped CC
- Spatial Data Management
- Distributed Databases
- More advanced data mining
- More details on NoSQL/Map Reduce etc.

..................
Course Plug

- CS 5604: Data Management Systems.
  - Graduate level course
  - Project, research papers
    • Will cover the state-of-the-art
  - Would be exciting and fun!
  - Good way to get exposed to the state-of-the-art in large data management and mining, network analysis, graph databases etc.
Good Luck!

- Especially for those of you will graduate!
- Feel free to keep in touch 😊
FINAL EXAMS

YES WE CAN