

CIS6930 midterm review
Spring 2009
Dr. Yicheng Tu

I. Introduction

1. Desirable characteristics of database systems (advantages over file systems)
2. Instances and schemas

II. Relational algebra (focus on rules/assumptions)

1. Select
2. Project
3. Union / Set difference/intersection
4. Cartesian product
5. Rename
6. Natural join, theta join
7. Division

III. SQL

1. Basic structure of SQL queries
 - a. *'select', 'from', and 'where'* clauses
 - b. Their algebraic equivalence
 - c. Tuple variables in SQL
 - d. Maintain/eliminate duplicates
2. SQL set operations
 - a. Union
 - b. Intersection
 - c. Except (Minus)
3. Aggregate functions
 - a. What are the different types of aggregates
 - b. *'group by'* and *'having'* clauses
4. Nested subqueries
 - a. Set membership and comparison
 - b. Testing for empty relations or relations without duplicates
5. How to do DIVISION in SQL?

VII. Storage and File Structure

1. Storage
 - a. The storage pyramid
 - b. Hard disk performance measurements: access time, seek time, latency
2. Sequential file organization vs. ordered files

VIII. Indexing

1. Basics
 - a. Primary/secondary index
 - b. Clustered/unclustered index
 - c. Performance of indexes vs. heap files vs. sorted files
2. B⁺-Tree Index Files
 - a. Components in a leaf node/ non-leaf node? How many of them? Fanout.
 - b. Insertion
 - c. Deletion
 - d. Bulk loading of B⁺-tree
 - e. Search a B⁺-tree
3. B-Tree Index Files

Going over the problems in the homeworks and the sample midterm will be the best way to prepare for the exam.

Midterm exam will be given at 4:35pm on March 2, 2009 (Monday) in room PHY109 (same room for our regular class meeting). It will be a 75-minute, closed books/notes/e-devices exam. Good luck!