CSE 532: Theory of Database Systems

Himanshu Gupta
Logistics

• Web page
  – http://www.cs.sunysb.edu/~hgupta/532

• Office Hours
  – After class
What is a Database?

• Databases:
  – Collection of large amounts of data

• Examples:
  – Airline Reservation System
    • Stores flights, airlines, reservations, price information, actual departure or arrival times, etc. Historical information.
  – Banking System
    • Stores account information, transactions, bank locations, etc.
  – SOLAR System
Managing Databases

• Database Management System (DBMS)
  – Software to manage databases.

• DBMS’s job/purpose:
  – Allow data definition (schema)
  – Manage data storage (access, backup)
  – Efficient Querying of data
  – Allow concurrent access
  – Support secure, atomic access
  – Crash Recovery
  – Etc.
DBMS Architecture

• High-level Architecture:

Queries → Query Processor → Storage Manager → Data/Metadata

Data/Schema Modifications → Query Processor → Transaction Manager

Himanshu Gupta
Course Outline (Tentative)

• **Recap**: Relational Model, Relational Algebra, SQL (2 weeks)
• Datalog  (1 week)
• Data Storage and Indexes (1.5 weeks)
• Query Processing/Optimization (2 weeks)
• Recovery (1 weeks)
• Concurrency/Transactions (1.5 weeks)

• Parallel & Distributed DBs (1 week)
• Semi-structured data; XML (1 week)
• Data Mining (1 week)
Data Model

• What is a data model?
  – Structure
  – Operations
  – Constraints

• DB Data Models
  – Relational
  – Semi-structured
  – Object Oriented
Relational Model

- Based on tables. E.g.,

<table>
<thead>
<tr>
<th>Course</th>
<th>Instructor</th>
<th>Quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS 532</td>
<td>M. Kifer</td>
<td>Fall, 2004</td>
</tr>
<tr>
<td>CS 532</td>
<td>H. Gupta</td>
<td>Spring, 2005</td>
</tr>
</tbody>
</table>

- Used by *most* DBMS

- How to decide *what tables* to create?
  - Entity/Relationship = E/R Model
DB Design Process

Data Ideas ➔ E/R Model ➔ Relations/Tables ➔ Dependencies, Normal Forms ➔ Relational DBMS
Purpose of E/R Model

• An E/R diagram is a way to represent the meaning or ‘semantics’ of the data that we wish to store

• Converting E/R diagram to relational database – fairly mechanical techniques exist.

• In this class, we won’t discuss E/R modeling.
Relational Model

- **Tables** = Relations
- **Columns** = Attributes
- **Row** = Tuple

Movies

<table>
<thead>
<tr>
<th>Name</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insider</td>
<td>1999</td>
</tr>
<tr>
<td>Heat</td>
<td>1995</td>
</tr>
</tbody>
</table>

- **Why Relations?**
  - Simple; Intuitive in terms of how we think of data.
Relational Terminology

• **Relation Schema:**
  – Name(attributes) + other structure info., e.g., keys, constraints.
  – Example: **Movies**(Name, Year)

• **Relation Instance:**
  – Set of rows/tuples for the relation

• **Database Schema:**
  – Collection of relation schemas

• **Key:**
  – A set of attributes such that no two (different) rows have the same values for these attributes.
Relational Algebra

- Next set of slides.