Outline

- Overview of databases and DBMS's
- Course topics and requirements

Databases and DBMS's

- A **database** is a large, integrated collection of data
- A **database management system** (DBMS) is a software system designed to store and manage a large amount of data
  - Declarative interface to define data stored, add data, update data, and query data
  - Efficient querying
  - Concurrent users
  - Reliable storage and crash recovery
  - Access control...

Commercial DBMS's

- INGRES
- Informix
- Postgres
- Sybase
- MS SQL Server
- System R
- IBM DB2
- Oracle
- MySQL

Earlier Database Applications

- **OnLine Transaction Processing (OLTP)**
  - Data with many small items, many queries, many updates
  - E.g., banking, airline reservations
  - E.g., university database

Recent Database Applications

- **OnLine Analytical Processing (OLAP)**, also known as Data Warehousing
  - Large amounts of data over years, complex queries, designed for analysis and reporting
  - Sales data analysis, e.g., Walmart, Target, ...
  - Fraud analysis, e.g., credit card use, insurance
More Recent DB Applications

- **Electronic commerce**
  - E.g., amazon.com, ebay.com
  - Integrating thousands of catalogs

- **Social networking**
  - E.g., facebook.com, myspace.com, with 100 million users at a popular site
  - Supporting many users
  - Real-time analysis of user behavior

How does one build a database?

Example: The Internet Shop*

- **DBDudes Inc.**: a well-known database consulting firm
- **Barns and Noble (B&N)**: a large bookstore specializing in books on horse racing
- B&N decides to go online but needs help

**Step 0**: DBDudes makes B&N agree to
- pay steep fees and
- schedule a lunch meeting for requirements analysis

* The example and all related material was taken from "Database Management Systems" Edition 3.

### Step 1: Requirements Analysis

- "I’d like my customers to be able to browse my catalog of books and place orders online."
  - **Books**: For each book, B&N’s catalog contains its ISBN number, title, author, price, year of publication, ...
  - **Customers**: Most customers are regulars with names and addresses registered with B&N.
    - New customers must first call and establish an account.
  - **On the new website**: Customers identify themselves before browsing and ordering.
  - **Shipping**: For each order, B&N ships all copies of a book together once they become available.

### Step 2: Conceptual Design

- A high level description of the data in terms of the **Entity-Relationship (ER) model**.
- Design review:
  - What if a customer places two orders of the same book in one day? 
  - Modification: add "ordernum" to Orders

### Step 3: Logical Design

- Mapping the ER diagram to the relational model
- **Access control**: use views to restrict the access of certain employees to customer sensitive information
Step 4: Schema Refinement

<table>
<thead>
<tr>
<th>ordernum</th>
<th>isbn</th>
<th>cid</th>
<th>cardnum</th>
<th>qty</th>
<th>order_date</th>
<th>ship_date</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>0-07-11</td>
<td>123</td>
<td>40241160</td>
<td>2</td>
<td>Jan 3, 2006</td>
<td>Jan 6, 2006</td>
</tr>
<tr>
<td>120</td>
<td>1-12-23</td>
<td>123</td>
<td>40241160</td>
<td>1</td>
<td>Jan 3, 2006</td>
<td>Jan 11, 2006</td>
</tr>
<tr>
<td>120</td>
<td>0-07-24</td>
<td>123</td>
<td>40241160</td>
<td>3</td>
<td>Jan 3, 2006</td>
<td>Jan 24, 2006</td>
</tr>
</tbody>
</table>

Step 5: Internet Application Development

Presentation tier
- Client Program
  - (Web Browser)
  - HTML, Javascript, Cookies

Application logic tier
- Application Server
  - (Apache Tomcat, JSP, Servlets)

Data management tier
- Database System
  - (MySQL, DB2, ...)

Example SQL Queries

- **Search Page**
  ```sql```
  ```
  SELECT isbn, title, author, price
  FROM Books
  WHERE author = '<SearchString>'
  ORDER BY title
  ```

- **Login Page**
  ```sql```
  ```
  SELECT cid, username, password
  FROM Customers
  WHERE username = '<SpecifiedUsername>'
  ```

Step 6: Physical Design

- Auxiliary data structures, indexes, to speed up searches, e.g., hash index, B-tree, R-tree

Books

<table>
<thead>
<tr>
<th>isbn</th>
<th>title</th>
<th>author</th>
<th>price</th>
<th>year</th>
<th>qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-07-11</td>
<td>Spectacular Bid</td>
<td>Timothy Capps</td>
<td>29.95</td>
<td>2003</td>
<td>10</td>
</tr>
<tr>
<td>1-12-23</td>
<td>Legacies of the Turf</td>
<td>Edward L. Bowen, Dan Mearns</td>
<td>24.95</td>
<td>2000</td>
<td>8</td>
</tr>
<tr>
<td>0-07-24</td>
<td>Seattle Slew</td>
<td>...</td>
<td>16.95</td>
<td>2001</td>
<td>3</td>
</tr>
</tbody>
</table>

Hash Index on Books.author

What is inside DBMS?
Course Topics

- Fundamentals
  - Data modeling
  - Relational design
  - Query languages, SQL
- Database implementations
  - Storage and indexing
  - Query processing and optimization
  - Transaction management
- Internet technologies
  - XML, XML query languages
  - Web application development

Course Web Site

or
http://www.cs.umass.edu/~yanlei/
→ Teaching
→ 445, Fall 2009

Textbook

Database Management Systems
3rd Edition
Ramakrishnan and Gehrke

Lecture notes will be posted on the schedule page before class.

Grading

- Homework: 25%
- Course Project: 20%
- Midterm: 20%
- Final: 25%
- Attendance, Participation: 10%
Homework: 25%

- 5 assignments throughout the semester
  - Written problem sets
  - Programming exercises with query languages including SQL and XQuery
- Dates that each assignment is out and due: see the schedule page
- Assignments: posted on the assignments page
- Submission: hardcopy before class on due date
- Policy on late submissions

Project: 20%

- General theme: build a web application using MySQL backend
- Groups of 2-3
- Project work will include:
  - Schema design
  - DB implementation
  - Web site design
- Multiple milestones & deliverables
  - See the schedule page
  - See the projects page for details
- Submission: via email, before class on due date

Exams

- Midterm (20%)
  - In-class, closed-book exam
  - At the beginning of the 9th week
- Final (25%)
  - Closed-book exam
  - Waiting to be scheduled in the final exam period

Attendance & Participation (10%)

- Attend every class
- Ask questions, contribute to answers
- Participate in in-class exercises and discussions

Academic Honesty

- All submitted work must be your own!
  - Although students are encouraged to study together, each student is expected to produce his or her own solutions to the homework problems.
  - Copying or using sections of someone else’s program or assignment (even if it has been modified by you), or copying a solution from an external source, is not acceptable.
  - The University guidelines for academic misconduct: http://www.umass.edu/col干细胞/codeofconduct/acadhonesty/
  - The staff of CS 445 will be vigorous in enforcing them.

Contact Information

- Instructor: Yanlei Diao
  - Email: yanlei at cs.umass.edu
- TA: Hong Yuan
  - Email: hongyuan at cs.umass.edu
- Mailing lists
  - cs445-help at edlab-mail.cs.umass.edu
  - cs445 at edlab-mail.cs.umass.edu
- All available on the home page