Outline

• Overview of databases and DBMS

• Course topics

• Course requirements
Database: a large, integrated collection of data.
Database Management System (DBMS): a software package designed to store and manage a large amount of data.
Case Study: The Internet Shop*

- **DBDudes Inc.**: a well-known database consulting firm
- **Barns and Nobble (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.
    • Each order contains the ISBN of a book and a quantity.
  – **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**

```
CREATE TABLE Books
(isbn CHAR(10),
title CHAR(80),
author CHAR(80),
qty_in_stock INTEGER,
price REAL,
year INTEGER,
PRIMARY KEY(isbn))

CREATE TABLE Customers
(cid INTEGER,
cname CHAR(80),
address CHAR(200),
PRIMARY KEY(cid))

CREATE TABLE Orders
(ordernum INTEGER,
isbn CHAR(10),
cid INTEGER,
cardnum CHAR(16),
qty INTEGER,
order_date DATE,
ship_date DATE,
PRIMARY KEY(ordernum, isbn),
FOREIGN KEY (isbn) REFERENCES Books,
FOREIGN KEY (cid) REFERENCES Customers)

CREATE VIEW OrderInfo
(isbn, cid, qty, order_date, ship_date)
AS SELECT O.isbn, O.cid, O.qty, O.order_date, O.ship_date
FROM Orders O
```

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

Orders

<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 26, 2006</td>
</tr>
</tbody>
</table>

Orderlists

<table>
<thead>
<tr>
<th>ordernum</th>
<th>isbn</th>
<th>qty</th>
<th>ship_date</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>0-07-11</td>
<td>2</td>
<td>Jan 6, 2006</td>
</tr>
<tr>
<td>120</td>
<td>1-12-23</td>
<td>1</td>
<td>Jan 11, 2006</td>
</tr>
<tr>
<td>120</td>
<td>0-07-24</td>
<td>3</td>
<td>Jan 26, 2006</td>
</tr>
</tbody>
</table>

Redundant Storage!
Step 5: Internet Application Development

Presentation tier

Client Program
(Web Browser)

HTTP, Javascript, Cookies

Application logic tier

Application Server
(Apache Tomcat…)

JSP, PHP
Servlets, XSLT

Data management tier

Database System
(MySQL, DB2…)

XML, stored procedures

B&N Data:
• Books
• Customers
  (User login)
• Orders
• Orderlists

B&N Business logic:
• Home page
• Login page
• Search page
• Cart page
• Confirm page

B&N Client:
• User input
• Display of output

Yanlei Diao, University of Massachusetts Amherst

1/29/2008
An Example Internet Store

Welcome to The Internet BookShop. We have just recently opened for business!!

We've spent a great deal of time in creating a website really suited to the desires of today's book buyers. We offer fast and convenient service with operators available at all hours of the day to process your requests.

And of course, membership is free of charge! Sign up today!

You may log into your account by clicking here, or create a new account by clicking here. If you want to add items to your shopping cart without logging in, feel free to.
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

**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>Legacies of the Turf</td>
<td>Edward L. Bowen</td>
<td>29.95</td>
<td>2003</td>
<td>10</td>
</tr>
<tr>
<td>1-12-23</td>
<td>Seattle Slew</td>
<td>Dan Mearns</td>
<td>24.95</td>
<td>2000</td>
<td>0</td>
</tr>
<tr>
<td>0-07-24</td>
<td>Spectacular Bid</td>
<td>Timothy Capps</td>
<td>16.95</td>
<td>2001</td>
<td>3</td>
</tr>
</tbody>
</table>

**Hash Index on Books.isbn**

isbn number

Yanlei Diao, University of Massachusetts Amherst
DBMS Architecture

Query Processor

- Query Parser
- Query Rewriter
- Query Optimizer
- Query Executor

Transactional Storage Manager

- Lock Manager
- Access Methods
- Buffer Manager
- Log Manager

Disk Manager

- Disk Space Manager

DB
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
Prerequisites

• CMPSCI 287: Programming Language Paradigms

• Or consent of the instructor
  – Data structures and algorithms
  – Sufficient programming experience
Course Goals and Topics

Database systems are at the core of large-scale information management, providing the most efficient mechanism for storing, updating, and retrieving structured data. This course will provide an introduction to the design and use of database systems, as well as the key issues in building such systems. The World Wide Web is the largest distributed information management system. In addition to database topics, this course will also provide an introduction to key technologies for managing and exchanging data on the World Wide Web.

In presenting the fundamental principles of databases we will cover the relational model, conceptual design, and query languages. We will also cover core database implementation issues including storage and indexing, query processing and optimization, transaction management and recovery. In presenting modern Internet-based data management we will cover a set of technologies for XML data management and web application development.

Time and Place:  
Tuesday and Thursday 2:30pm-3:45pm, LGRT 161

Professor:  
Yanlei Diao
Contact:  
yanlei@cs.umass.edu
Office Hours:  
Tuesday 1:00pm-2:00pm or by appointment, CMPS 232

Teaching Assistant:
Ravishankar Rajamony
Contact:  
ravi ~at~ cs.umass.edu
Office Hour:  
Tuesday 5pm-6pm, Common Room, Computer Science Department

Contact instructor+TA:  
cs445-help@edlab-mail.cs.umass.edu

Broadcast Mailing List:
We have created a course broadcast email address, cs445@edlab-mail.cs.umass.edu, for use by course participants. The instructor and TA will send all course related announcements (e.g., clarifications of homework or programming assignments, questions/answers of general interest, etc.) to this mailing list.

By default, email is sent to your edlab account. It is your responsibility to forward the mail to some other account if you want it to go somewhere else. To forward mail, create a file called "forward" in your home directory on the edlab machines. The only line in the file should contain the alternate email address. All mail to your edlab account will then be forwarded to the alternate address.

Please note that whenever you send email to cs445@edlab-mail.cs.umass.edu, your email message will be broadcast to the entire class, so use the address carefully. A question/comment or some information you want to share with the class would be an appropriate email to send to cs445@edlab-mail.cs.umass.edu. A request for a due date extension would not be appropriate.
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 midnight 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 pop quizzes and in-class exercises
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/dean_students/code_conduct/acad_honesty.htm](http://www.umass.edu/dean_students/code_conduct/acad_honesty.htm)
  – The staff of CS 445 will be vigorous in enforcing them.
Contact Information

• Instructor: Yanlei Diao
  – Email: yanlei at cs.umass.edu
  – Office hours: Tue 1:00-2:20 pm, or by appointment, Room 232 in CS building

• TA: Ravishankar Rajamony
  – Email: rravi at cs.umass.edu
  – Office hour: Tue 5-6 pm, common room, CS building

• Mailing lists
  – cs445-help at edlab-mail.cs.umass.edu
  – cs445 at edlab-mail.cs.umass.edu

• All available on the home page