IST 210: Organization of Data

Todd S. Bacastow

Dhiraj Joshi
Class Administration

- Instructor and Assistant
  - Todd S. Bacastow
  - Dhiraj Joshi

- Course logistics
  - On-line material/Text
  - Syllabus
  - Office Hours
    - Bacastow
      - Monday - Friday, 224 EES Building, By Appointment
Texts

- **Required Texts:**
  - Online IST Student Roadmap

- **Optional Resources:**
  - David Buser, Jon Duckett, Brian Francis, John Kauffman, Juan T Llibre, David Sussman, Chris Ullman, Beginning Active Server Pages 3.0. WROX Press. ISBN: 1861003382.
Grading

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Number of Course Assignments</th>
<th>Point Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 0</td>
<td></td>
<td>20 Points</td>
</tr>
<tr>
<td>Problem Documents</td>
<td>2 @ 100 points each</td>
<td>200 Points</td>
</tr>
<tr>
<td>Problem Presentations</td>
<td>2 @ 50 points each</td>
<td>100 Points</td>
</tr>
<tr>
<td>Self and Team Evaluations</td>
<td>2 @ 25 points each</td>
<td>50 Points</td>
</tr>
<tr>
<td>Module 1 and 2 Quizzes</td>
<td>2 @ 100 points each</td>
<td>200 Points</td>
</tr>
<tr>
<td>Discussion Activities</td>
<td>12 @ 5 points each</td>
<td>60 Points</td>
</tr>
<tr>
<td>Spot Quizzes</td>
<td>10 @ 5 points each</td>
<td>50 Points</td>
</tr>
<tr>
<td>Lab Activities</td>
<td>6 @ 25 points each</td>
<td>150 Points</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>830 Points</strong></td>
</tr>
</tbody>
</table>
Course Goals

- Provide an introduction to databases and database technologies
- Demonstrate an understanding of database modeling techniques and the design process
- Demonstrate the fundamentals of SQL
- Present issues relating to database management and administration
- Understand the implications and future directions of databases and database technologies
What is a Database

- General
  - collection of information that exists over a long period of time

- IST 210
  - a collection of data that is managed by a database management system (DBMS)
Definition (cont’d)

- A very large, integrated collection of data.
- A shared collection of logically related data designed to meet the information needs of an organization.
- Models real-world **enterprise**
  - Entities (e.g., students, courses)
  - Relationships (e.g., Madonna is taking IST 210)
Important Elements

- Three key elements of database definition:
  - Shared
  - Interrelated
  - Predefined applications

- Side notes:
  - Database is an abstraction of the real world
  - Data becomes information only when they are used to provide answers to questions
Why Study Databases??

- Shift from *computation* to *information*
  - Focus on the way to structure information
- Datasets increasing in diversity and volume.
  - Digital libraries, interactive video, e-commerce
  - ... need for DBMS exploding
- DBMS encompasses most of the information technology
  - OS, languages, theory, multimedia, logic, web
3 Aspects to Studying DBMS’s

- Database modeling and design
- Programming queries and DB operations
- DBMS implementation
Why Database Design?
Database Management System (DBMS)

- **DBMS:**
  - A software system that enables users to define, create, and maintain the database and which provides controlled access to this database.

- Provide a layer between user application programs and the data
  - Data Definition Language
  - Data Manipulation Language
What does a DBMS do?

- Database management systems:
  - Provide efficient and secure access to large amounts of data.

- Address problems such as:
  - How to store the data
  - How to query data efficiently
  - How to update the data securely (by multiple users)

- Contrast with using file systems for the same task
Typical DBMS Functions

- Storage management
- High level query and data manipulation language
- Efficient query processing
- Transaction processing
- Resiliency: recovery from crashes
- Interface with programming languages
File-based Processing

- Sales
  - Data entry and reports
  - File handling routines
  - File definition
  - Sales files

- Contracts
  - Data entry and reports
  - File handling routines
  - File definition
  - Contracts files

Sales application programs
Contracts application programs
Problems with File-based Systems

- Same data is stored in multiple places. *Inconsistencies!*
- We need to write special programs for each user question
- Data can be corrupted due to system crash while change is made.
- User programs are not easy to share data or evolve
Database Management System (DBMS)
DBMS Advantages

- Control of data redundancy
  - Have a central depository of all data and their descriptions
  - Same information stored only once
- Data Integrity
- Controlled access to database
- Data independence
- Concurrent Access
- Crash recovery
DBMS Disadvantages

- Complexity
- Cost of DBMS software, hardware and data conversion
- Performance
- Higher impact of a failure
DBMS Applications

- Bank systems
- Airline reservation systems
- Corporate (or student) records
- Store sales records
- Patient electronic medical records
- Internet search engines
- Multimedia search engines
- DNA sequence search engines