CSC 355 Database Systems

Marcus Schaefer
Databases?
Database

1. DB models aspects of the real world
   \textit{(miniworld, universe of discourse)}

2. Collection of data
   \begin{itemize}
   \item logically coherent \quad \rightarrow \quad \textit{Information}
   \item Meaningful
   \end{itemize}

3.Designed for specific purpose
Uses of Database

- Traditional (Employee, student, product database)
- Online Shopping
- Search Engines
- Data Warehousing (OLAP)
- Data Mining
- Genetic Databases
- Geographical Information Systems
Types of Database

- Traditional (Postgres, Oracle, MySQL)
- Deductive Databases
- Multimedia Databases
- Distributed Databases
- Spatial Databases
- Object-Oriented Databases
- No-SQL Databases
No-SQL Databases

- Key-Value (Riak)
- Columnar (Cassandra, HBase)
- Document (MongoDB)
- Graph (Neo4J)
Sizes of Database

- Personal (1 User), Megabytes
- Workgroup (<25 Users), Megabytes
- Department (25-100 Users), Gigabytes
- Enterprise (100-1000s), Gigabytes
- Internet (> 1000s), Terabytes - Petabytes

How long does it take to find a piece of data in petabytes of data?
Database Management System (DBMS)

Software to

1. Define a database
   (data types, relations, constraints)

2. Construct a database
   (populate database with data)

3. Manipulate database
   (query and update data in database)
Database People

- Database designers
- Application developers
- Database administrators
- Users
Explore University Sample Database

<table>
<thead>
<tr>
<th>LastName</th>
<th>FirstName</th>
<th>SID</th>
<th>SSN</th>
<th>Career</th>
<th>Program</th>
<th>City</th>
<th>Started</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snowdon</td>
<td>Jonathan</td>
<td>8871</td>
<td>1231231231</td>
<td>GRD</td>
<td>INFO-SYS</td>
<td>Springfield</td>
<td>2005</td>
</tr>
<tr>
<td>Winter</td>
<td>Abigail</td>
<td>11035</td>
<td>111111111111</td>
<td>GRD</td>
<td>PHD</td>
<td>Chicago</td>
<td>2003</td>
</tr>
<tr>
<td>Patel</td>
<td>Deepa</td>
<td>14662</td>
<td></td>
<td>GRD</td>
<td>COMP-SCI</td>
<td>Evanston</td>
<td>2003</td>
</tr>
<tr>
<td>Starck</td>
<td>Jason</td>
<td>19992</td>
<td>789789789</td>
<td>UGRD</td>
<td>INFO-SYS</td>
<td>Springfield</td>
<td>2003</td>
</tr>
<tr>
<td>Johnson</td>
<td>Peter</td>
<td>32105</td>
<td>123456789</td>
<td>UGRD</td>
<td>COMP-SCI</td>
<td>Chicago</td>
<td>2004</td>
</tr>
<tr>
<td>Patel</td>
<td>Prakash</td>
<td>75234</td>
<td></td>
<td>UGRD</td>
<td>COMP-SCI</td>
<td>Chicago</td>
<td>2001</td>
</tr>
<tr>
<td>Brennigan</td>
<td>Marcus</td>
<td>90421</td>
<td>987654321</td>
<td>UGRD</td>
<td>COMP-GPH</td>
<td>Evanston</td>
<td>2001</td>
</tr>
<tr>
<td>Snowdon</td>
<td>Jennifer</td>
<td>93321</td>
<td>321321321</td>
<td>GRD</td>
<td>COMP-SCI</td>
<td>Springfield</td>
<td>2004</td>
</tr>
</tbody>
</table>
Explore University Sample Database
University Sample Database

User Data
- Records, Fields (Columns)
- Data elements

Meta Data
- Data Types
- Relationships
- Constraints
- Indexes

Other Data
- Log Records
- Statistics

Relational Databases
File Processing

File system is backbone of operating system

File system for data storage:

Adapted from http://blogs.netapp.com/databases/WindowsLiveWriter/image_29.png
Disadvantages of File Processing

- Program-Data Dependence
- Redundancy (Duplication of Data)
- Limitation on data sharing
- Development time
- Maintenance
Advantages of Databases

• Program-Data Independence
• Control of Data Redundancy
• Data Consistency
• Data Quality (constraints)
• Data Sharing (customized access through views)
• Improved Data Access
• Program Maintenance
DBMS Architecture
## DBMS

<table>
<thead>
<tr>
<th>Open Source</th>
<th>Proprietary</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2</td>
<td>Access (Microsoft)</td>
</tr>
<tr>
<td>MySQL,</td>
<td>DB2 (IBM)</td>
</tr>
<tr>
<td>Postgres</td>
<td>Oracle</td>
</tr>
<tr>
<td></td>
<td>SQL Server (Microsoft)</td>
</tr>
<tr>
<td></td>
<td>Sybase (SAP)</td>
</tr>
</tbody>
</table>
Data Modeling

Describe structure of data (relationships, behavior) at different levels of abstraction.

Conceptual/External
  high-level user view

Internal
  Logical: structure of data for DBMS
  Physical: storage details (indexes) for DBMS
Data Models

Conceptual/External
- ER-model (Entity-Relationship)
- UML

Logical
- Relational data model
- Object data model
- Network data model
- Hierarchical data model
ER-modeling

Describes **entities**, their **relationships**, and **attributes**

Used for designing and analyzing a database

Ullman, Widom, *A First Course*, p. 149
Relational Data Model

• Data in tables (extensional representation of relation)

• Models relationship between data in tables

Ullman, Widom, *A First Course*, p. 26

Figure 2.5: Example database schema about movies
Database Languages

DDL: Data definition language
defines data types, tables
includes DSL (Data storage language)

DML: Data Manipulation Language
language for retrieving and manipulating data

Types:
- high-level (nonprocedural, declarative): SQL
- low-level (procedural)
Transactional Processing

Transaction: A group of database operations that should appear as a unit to the user.

Example:

Transfer $100 from account A to account B.

Requirements on transactions:

Atomicity
Consistency
Isolation
Durability
## Class Outline

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intro to Database systems, Relational Model (Chapters 1/2)</td>
</tr>
<tr>
<td>2-4</td>
<td>SQL (with transactions) (Chapter 6)</td>
</tr>
<tr>
<td>5</td>
<td>Relational Design: Functional Dependencies and Normalization (Chapter 3)</td>
</tr>
<tr>
<td>6</td>
<td>Constraints &amp; Triggers (Chapter 7)</td>
</tr>
<tr>
<td>7</td>
<td>Views &amp; Indices (Chapter 8)</td>
</tr>
<tr>
<td>8-9</td>
<td>Database Programming (Chapter 9)</td>
</tr>
<tr>
<td>10</td>
<td>Advanced Topics: recursive SQL, ORL, semi-structured data, No-SQL</td>
</tr>
</tbody>
</table>