Comp 100
Chapter 11
Databases

Twitter Interest Groups Database -- justtweetit.com
Facebook
Amazon.com

Database

- **Transform data into information**

• Store data • Sort data • Organize data	➔	• Retrieve data • Query data • Data Mining

- **complex data sets**
- **multiple users**

- **Airline Reservations Systems**
- **Hotel Management Systems**
- **Manufacturing Systems**

Lists

- **Word**
- **Excel**

List 1			List 2	
Name	**Address**		**Name**	**Address**
James Lee	127 Adams		James Lee	349 Baker
Mary Smith	632 West Phillips		Alice Topel	19 Center Drive
Alice Topel	19 Center Drive		James Lee	349 Baker
			Evan Martin	…..

- **Data Redundancy**
- **Data Inconsistency**
- **Data Duplication**
- **Nonexistent Data …..**

Database Advantages

- **Data Centralization**
  - **Data Items e.g., Name & Address, are stored in one location only**
  - **Eliminate Data Redundancy & Inconsistency**
- **Data Integrity**
- **Data Entry Control, i.e., Eliminate Data Entry Errors**
  - **Data Duplication**
  - **Nonexistent Data …..**
- **Improve Data Flexibility, i.e., data retrieval, reports**

ID#	FirstName	LastName	Address	City	State	Zip	Telephone	Class Code
234	Li	Chan	123 Main	Tubville	NV	49874	555-5555	Eng102
453	Diane	Coyle	745 Oak	Lakeview	CO	56537	777-7777	Math112

Field – category of information (column)

- **Field Name**
- **Field Data Type, zip code, telephone, etc.**
  - **Text field**
  - **Numeric field**
  - **Computed field**
  - **Date field**
  - **Memo field**
  - **Object field**
    - ✓ **.pdf documents**
    - ✓ **Pictures**
    - ✓ **Videos**
  - **Hyperlink field**
- **Field Size**

Record – set of related fields (row)

- **set of related data items concerning an entity**
- **information regarding a specific entity**

Table, File – group of related records

Primary Key – field item that uniquely specifies a record, e.g., ID#

Relational Database – E.F. Codd

- **Information is organized into Tables (two-dimensional)**
- **Each table contains a set of discreet, related data items**
    - **Student Information Table – each record specific to a particular student**
    - **Courses Taken Table – set of courses taken by a specific student**
    - **Comp 100 Section # 12345 Table – list of students enrolled in class**
- **Tables are linked together via the primary keys into <u>Relations</u>**
- **Structured Data**
- **SQL – Structured Query Language -- Algebra**

Object-Oriented Database

- **Unstructured Data**
    - **Audio**
    - **Video**
    - **Pictures**
    - **Extremely Large Objects (Binary Large Objects, i.e., BLOB)**
        - **Rare Manuscripts**
- **OQL – Object Query Language**
- **Amazon.com**

Multidimensional Database

- **Primarly 3-Dimensional Databases**

Database Management Systems

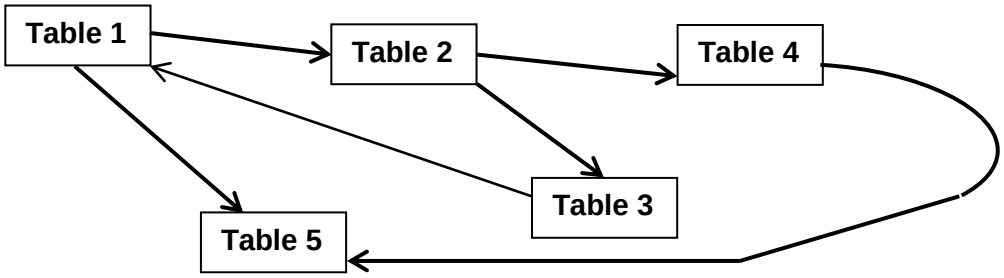
- **Create Database -- page 532 Sound Byte**
  - **Data Dictionary, i.e., Schema**
    - **Metadata – field name, field data type, field size**
    - **Access -- page 526 figure 11.10**
- **Populate it with Data**
  - **Importing – filters**
  - **Keying -- Input Form**
  - **Validation Rules**
    - **Range – 16 < required age < 140**
    - **Completeness – required fields, e.g., Last Name**
    - **Consistency – birth date vs. current age**
      current age = current date – birth date
    - **Numeric – age <> 3g**
    - **Alphabetic – name <> St3ve**
- **Viewing**
  - **Browsing a selective set of records**
- **Sorting Data**
  - **Reordering tables by selected fields**
- **Querying, i.e., Viewing -- Extracting Specific Information (SQL, Wizards)**
  - **Data Items,**
  - **Records**
  - **Tables**
  - **Relations**
- **Updating Data**
  - **Scanning**
  - **Keyboard**
  - **Import Files (electronic)**
- **Providing Information**
  - **Reports**
  - **Export Files (electronic)**

Relational Database Operations

Organize Data into
Selected Tables based on
Logical Groupings

Relation
Link between Tables
defines the data relationships

two tables in a relation must have a common field
for a group of tables to be in a relation, there must be a path connecting all the tables



One-to-One Relationship
Student ID#
Student Master Roster 1-1 (into) Comp 100 Section 123 Class Roster
Student listed in class roster only once

One-to-Many Relationship
Student ID#
Student Master Roster 1-Many Master Class Roster
Student may register for more than one class per semester

Many-to-Many Relationship
Student ID#
Student Master Roster Many-to-Many University Employment Roster
Student may work for more than one employer
Employer may hire more than one student

Normalization **(Reduce Data Redundancy)**

Database Table

- **Each table contains only data concerning a specific (see page 535 fig 11.22)**
  - **Entity**
  - **Event**
  - **Transaction**
- **Tables must be grouped using uniquely identified logical data items**

> **If a table has no logical identifiable key, it probably needs to be divided into two or more simpler tables**

- **Foreign Key**
  - **Primary Key of one table that is included as a data item in another table for the purposes of being able to establish a link between the two tables**
  - **Referential Integrity – each foreign key value must exist in the table it as a primary key**

A DATABASE is usually a set of related tables concerning
a specific operational aspect of operations
time-dependent, i.e., quarterly, information

- **Amazon.com**
  - **Order Database**
  - **Inventory Database**
  - **Suppliers Database**

Data Warehouse

- **Collection of databases about disparate topics that provides an enterprise-wide view of business operations**
- **Data is NOT operational**
- **Archival Information, i.e., Current Information + Historical Information**
- **Internal Sources – operational databases**
- **External Sources – suppliers, vendors, customers**

Information Systems

- **Office Support**
- **Transaction Processing**
  - **Batch Processing (Detail Reports, Mailing Lists, Catalog Production)**
  - **Real-Time Processing (On-Line Transaction Processing)**
- **Management Information**
  - **Summary Reports – Consolidated Information**
  - **Exception Reports**
- **Decision Support**
  - **Internal Data Sources**
    - **Operational Databases**
    - **Transaction-Processing Systems**
    - **Data Warehouse**
  - **External Data Sources**
    - **Demographic Data**
    - **Mailing Lists**
    - **Government Statistics**
  - **Model Management Systems**
    - **Analytical Tools**
    - **Models**
    - **Statistics**
  - **Knowledge-Based Systems**
    - **Expert Systems**
    - **Natural Language Processing Systems**
      - ✓ **www.nuance.com/**
      - ✓ **www.nanopac.com/**
      - ✓ **www.disaboom.com**
      - ✓ **Assistive Technology**
      - ✓ **Computer Technology**
    - **Fuzzy Logic Projection Systems**

- **Enterprise Resource Planning**
  - **Human Resource Systems**
  - **Accounting Systems**
  - **Manufacturing Operations**

- **Data Mining**
  - **Classification**
    - **define analysis categories**
  - **Estimation**
    - **measure the ability to fit the established criterion**
  - **Affinity Grouping (association rules)**
    - **determine associations between items**
    - **group items according to their associations**
  - **Clustering**
    - **group items according to similar data actions**
    - **group items without regard to the defined categories**
    - **determine whether the clusters define a category**
  - **Description & Visualization**