Announcements

- DB 2 Due Tuesday Next Week
• **Database:**
  - Collection of related files containing records on people, places, or things.
  - Prior to dig. DBs, business used paper files.

• **Entity:**
  - Generalized category representing person, place, thing on which we store info.
  - E.g., SUPPLIER, PART

• **Attributes:**
  - Specific characteristics of each entity:
    - SUPPLIER name, address
    - PART description, unit price, supplier
• **Relational database:**
  - Organize data into tables
  - One table for each entity:
    - E.g., (CUSTOMER, SUPPLIER, PART, SALES)
  - **Fields** (columns) store data representing an attribute.
  - Rows store data for separate **records**.
  - **Key field:** uniquely identifies each record.
  - **Primary key:**
    - One field in each table
    - Cannot be duplicated
    - Provides unique identifier for all information in any row
A relational database organizes data in the form of two-dimensional tables. Illustrated here is a table for the entity SUPPLIER showing how it represents the entity and its attributes. Supplier_Number is the key field.

**Figure 5-1**
The PART Table

<table>
<thead>
<tr>
<th>Part_Number</th>
<th>Part_Name</th>
<th>Unit_Price</th>
<th>Supplier_Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>137</td>
<td>Door latch</td>
<td>22.00</td>
<td>8259</td>
</tr>
<tr>
<td>145</td>
<td>Side mirror</td>
<td>12.00</td>
<td>8444</td>
</tr>
<tr>
<td>150</td>
<td>Door molding</td>
<td>6.00</td>
<td>8263</td>
</tr>
<tr>
<td>152</td>
<td>Door lock</td>
<td>31.00</td>
<td>8259</td>
</tr>
<tr>
<td>155</td>
<td>Compressor</td>
<td>54.00</td>
<td>8261</td>
</tr>
<tr>
<td>178</td>
<td>Door handle</td>
<td>10.00</td>
<td>8259</td>
</tr>
</tbody>
</table>

Figure 5-2
• Establishing relationships
  • Entity-relationship diagram
    • Used to clarify table relationships in a relational database
  • Relational database tables may have:
    • One-to-one relationship
    • One-to-many relationship
    • Many-to-many relationship
      • Requires creating a table (join table, Intersection relation) that links the two tables to join information
A Simple Entity-Relationship Diagram

This diagram shows the relationship between the entities SUPPLIER and PART.

Figure 5-3
Normalization

- Process of streamlining complex groups of data to:
  - Minimize redundant data elements.
  - Minimize awkward many-to-many relationships.
  - Increase stability and flexibility.

Referential integrity rules

- Used by relational databases to ensure that relationships between coupled tables remain consistent.
Sample Order Report

Order Number: 3502
Order Date: 1/15/2008

Supplier Number: 8259
Supplier Name: CBM Inc.
Supplier Address: 74 5th Avenue, Dayton, OH 45220

<table>
<thead>
<tr>
<th>Order_Number</th>
<th>Part_Number</th>
<th>Part_Quantity</th>
<th>Part_Name</th>
<th>Unit_Price</th>
<th>Extended_Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>3502</td>
<td>137</td>
<td>10</td>
<td>Door latch</td>
<td>22.00</td>
<td>$220.00</td>
</tr>
<tr>
<td>3502</td>
<td>152</td>
<td>20</td>
<td>Door lock</td>
<td>31.00</td>
<td>620.00</td>
</tr>
<tr>
<td>3502</td>
<td>178</td>
<td>5</td>
<td>Door handle</td>
<td>10.00</td>
<td>50.00</td>
</tr>
</tbody>
</table>

Order Total: $890.00

Figure 5-4
The Final Database Design with Sample Records

Figure 5-5

<table>
<thead>
<tr>
<th>PART</th>
<th>LINE_ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part_Number</td>
<td>Order_Number</td>
</tr>
<tr>
<td>137</td>
<td>3502</td>
</tr>
<tr>
<td>145</td>
<td>3502</td>
</tr>
<tr>
<td>150</td>
<td>3502</td>
</tr>
<tr>
<td>152</td>
<td>8259</td>
</tr>
<tr>
<td>155</td>
<td>8444</td>
</tr>
<tr>
<td>178</td>
<td>8263</td>
</tr>
<tr>
<td></td>
<td>8259</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ORDER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order_Number</td>
</tr>
<tr>
<td>3502</td>
</tr>
<tr>
<td>3503</td>
</tr>
<tr>
<td>3504</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUPPLIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplier_Number</td>
</tr>
<tr>
<td>8259</td>
</tr>
<tr>
<td>8261</td>
</tr>
<tr>
<td>8263</td>
</tr>
<tr>
<td>8444</td>
</tr>
</tbody>
</table>
This diagram shows the relationship between the entities SUPPLIER, ART, LINE_ITEM, and ORDER.

Figure 5-6
DBMS

- Specific type of software for creating, storing, organizing, and accessing data from a database
- Separates the logical and physical views of the data
  - **Logical view:** how end users view data
  - **Physical view:** how data are actually structured and organized
- **Examples of DBMS:** Microsoft Access, DB2, Oracle Database, Microsoft SQL Server, MySQL
Human Resources Database with Multiple Views

Figure 5-7
Operations of a Relational DBMS

• Select:
  • Creates a subset of all records meeting stated criteria

• Join:
  • Combines relational tables to present the server with more information than is available from individual tables

• Project:
  • Creates a subset consisting of columns in a table
  • Permits user to create new tables containing only desired information
The select, project, and join operations enable data from two different tables to be combined and only selected attributes to be displayed.
Capabilities of Database Management Systems

- **Data definition capabilities:**
  - Specify structure of content of database.

- **Data dictionary:**
  - Automated or manual file storing definitions of data elements and their characteristics.

- **Querying and reporting:**
  - **Data manipulation language**
    - Structured query language (SQL)
    - Microsoft Access query-building tools
Example of an SQL Query

SELECT PART.Part_Number, PART.Part_Name, SUPPLIER.Supplier_Number, SUPPLIER.Supplier_Name
FROM PART, SUPPLIER
WHERE PART.Supplier_Number = SUPPLIER.Supplier_Number AND Part_Number = 137 OR Part_Number = 150;

Illustrated here are the SQL statements for a query to select suppliers for parts 137 or 150. They produce a list with the same results as Figure 5-8.

Figure 5-10
An Access Query

Figure 5-11
Object-Oriented DBMS (OODBMS)

- Stores data and procedures that act on those data as objects to be retrieved and shared
- Better suited for storing graphic objects, drawings, video, than DBMS designed for structuring data only
- Used to manage multimedia components or Java applets in Web applications
- Relatively slow compared to relational DBMS
- Object-relational DBMS: provide capabilities of both types
Data Warehouses

• **Data warehouse:**
  - Database that stores current and historical data for decision makers
  - Consolidates and standardizes data from many systems
  - Data can be accessed but not altered

• **Data mart:**
  - Subset of data warehouses that is highly focused and isolated for a specific population of users
Components of a Data Warehouse

Operational Data
Customer Data
Manufacturing Data
Historical Data

INTERNAL DATA SOURCES

External Data

EXTERNAL DATA SOURCES

Extract and Transform

Data Warehouse

Data Access and Analysis
- Queries and reports
- OLAP
- Data mining

Information Directory

Figure 5-12
Business Intelligence, Multidimensional Data Analysis, and Data Mining

- **Business intelligence**: tools for consolidating, analyzing, and providing access to data to improve decision making
  - Software for database reporting and querying
  - Tools for multidimensional data analysis (online analytical processing --OLAP)
  - Data mining
Business Intelligence

Databases

Data Warehouse

Business Intelligence

Keep track of transactions

Find patterns and insights

Decision to respond to the data and patterns

OLAP
Data Mining

Querying
Reporting

Figure 5-13
Online Analytical Processing (OLAP)

- Supports multidimensional data analysis
  - Enable users to view same data in different ways using multiple dimensions
  - Dimension can be — product, pricing, cost, region, or time period
  - E.g., comparing sales in East in June versus May and July
Using Databases to Improve Business Performance and Decision Making

Figure 5-14
Multidimensional Data Model

Essentials of Management Information Systems
Chapter 5 Foundations of Business Intelligence: Databases and Information Management
Data Mining

- Finds hidden patterns and relationships in large databases
- Types of information obtainable from data mining
  - **Associations**: occurrences linked to single event
  - **Sequences**: events linked over time
  - **Classifications**: patterns describing a group an item belongs to
  - **Clusters**: discovering as yet unclassified groupings
  - **Forecasting**: uses series of values to forecast future values
Data Mining

• One popular use of data mining: identifying profitable customers

• Predictive analysis:
  • Uses historical data, and assumptions about future conditions to predict outcomes of events
  • E.g. such the probability a customer will respond to an offer or purchase a specific product
• Text Mining
  • Unstructured data (mostly text files) accounts for 80 percent of an organization’s useful information.
  • Text mining -- extract key elements from, discover patterns in, and summarize large unstructured data sets.

• Web Mining
  • Discovery and analysis of useful patterns and information from the Web