The Database Approach to Data Management

• 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 table:

<table>
<thead>
<tr>
<th>Name</th>
<th>Address</th>
<th>City</th>
<th>State</th>
<th>Zip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplier 1</td>
<td>123 Pine St.</td>
<td>Toronto</td>
<td>ON</td>
<td>M1C</td>
</tr>
<tr>
<td>Supplier 2</td>
<td>456 Maple Dr.</td>
<td>Calgary</td>
<td>AB</td>
<td>T2A</td>
</tr>
<tr>
<td>Supplier 3</td>
<td>789 Oak Ave.</td>
<td>Vancouver</td>
<td>BC</td>
<td>V6H</td>
</tr>
</tbody>
</table>

Key field (Primary Key)

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>101</td>
<td>Door latch</td>
<td>22.00</td>
<td>8209</td>
</tr>
<tr>
<td>145</td>
<td>Side mirror</td>
<td>12.00</td>
<td>8444</td>
</tr>
<tr>
<td>160</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>8298</td>
</tr>
<tr>
<td>155</td>
<td>Compressor</td>
<td>64.00</td>
<td>8291</td>
</tr>
<tr>
<td>178</td>
<td>Door handle</td>
<td>10.00</td>
<td>8299</td>
</tr>
</tbody>
</table>

Primary Key

Foreign Key

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

• 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.
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

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

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
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

- Operations data
- Customer data
- Manufacturing data
- Historical data
- Data access and analysis

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

- OLAP data mining
- Querying and reporting
- Decision to respond to the data and patterns.
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

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

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

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