Class Announcements

- Database Assignment 2 posted
  - Due 11/22
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

![An Access Query](Image)

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

Components of a Data Warehouse

![Components of a Data Warehouse](Image)

Figure 5-12

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
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 (OLAP)
- Data mining

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

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

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