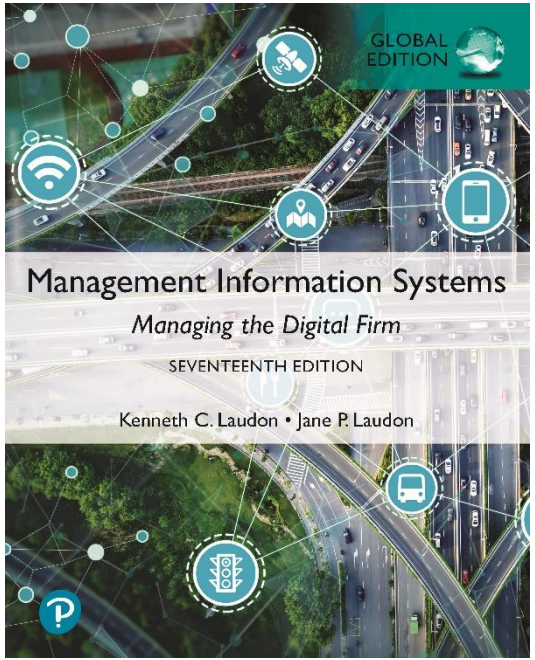
Management Information Systems: Managing the Digital Firm

Seventeenth Edition, Global Edition



Chapter 6

Foundations of Business Intelligence: Databases and Information Management

Learning Objectives

6.1 What are the problems of managing data resources in a traditional file environment?

6.2 What are the major capabilities of database management systems (DBMS), and why is a relational DBMS so powerful?

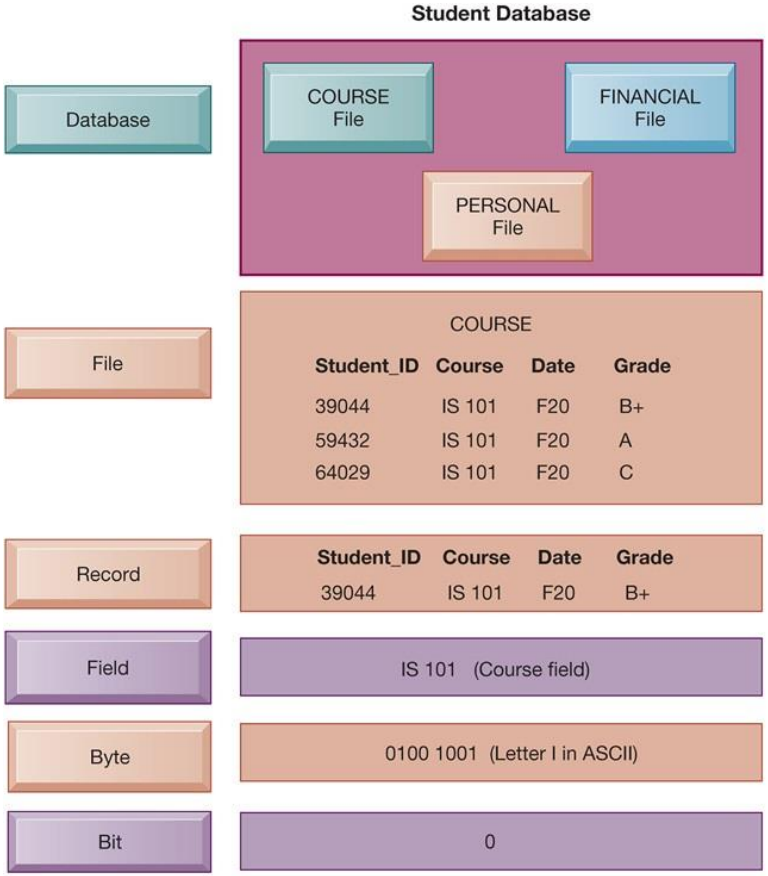
6.3 What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?

6.4 Why are data governance and data quality assurance essential for managing the firm's data resources?

File Organization Terms and Concepts

- Database: Group of related files
- File: Group of records of same type
- Record: Group of related fields
- Field: Group of characters as word(s) or number(s)
- Entity: Person, place, thing on which we store information
- Attribute: Each characteristic, or quality, describing entity

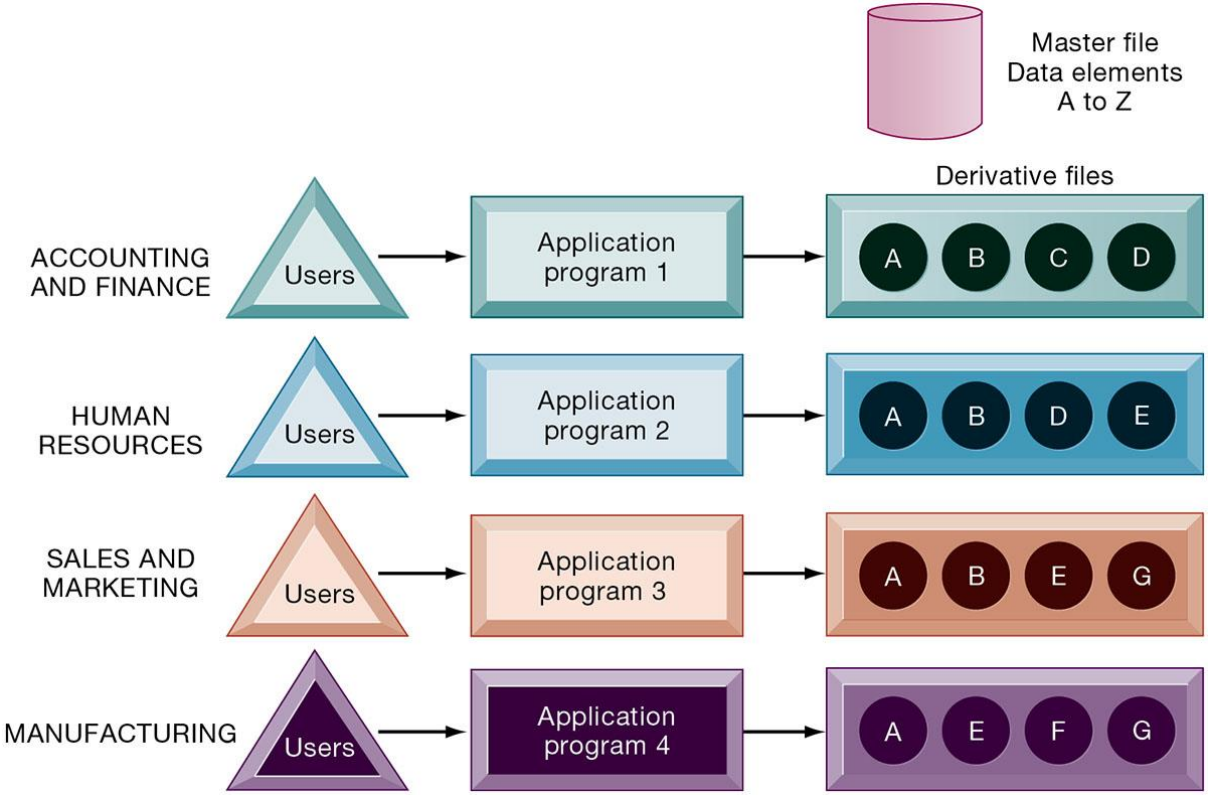
Figure 6.1 The Data Hierarchy



Problems with the Traditional File Environment

- Files maintained separately by different departments
- Data redundancy
- Data inconsistency
- Program-data dependence
- Lack of flexibility
- Poor security
- Lack of data sharing and availability

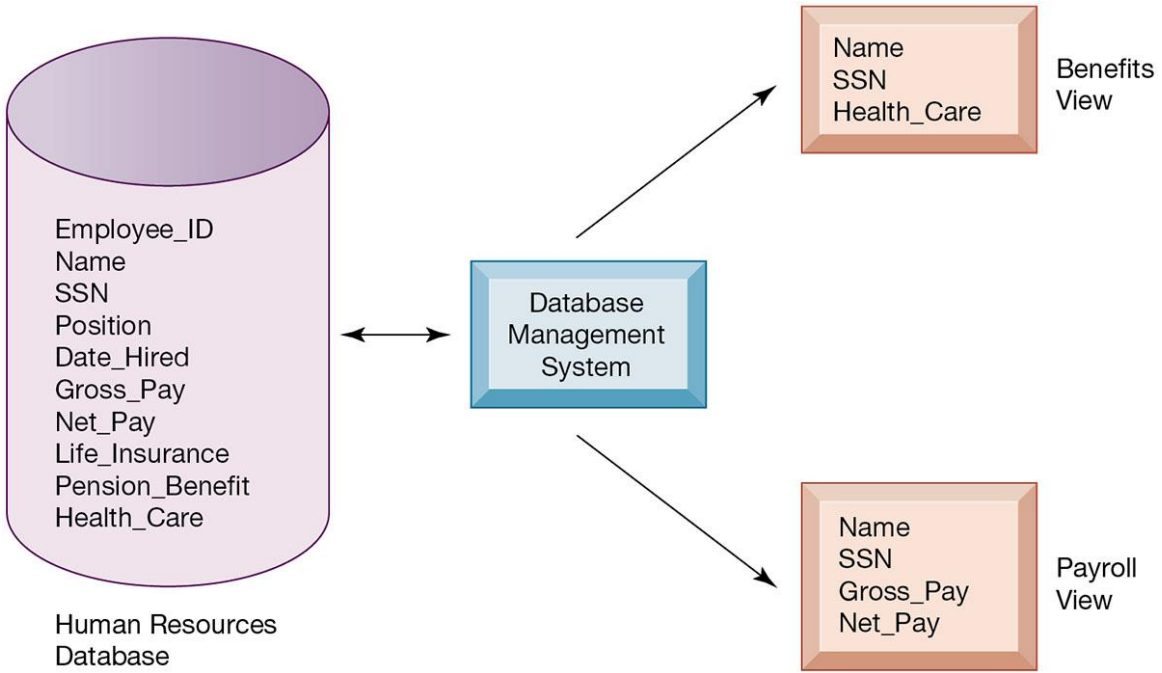
Figure 6.2 Traditional File Processing



Database Management Systems

- Database
  - Serves many applications by centralizing data and controlling redundant data
- Database management system (DBMS)
  - Interfaces between applications and physical data files
  - Separates logical and physical views of data
  - Solves problems of traditional file environment
    - Controls redundancy
    - Eliminates inconsistency
    - Uncouples programs and data
    - Enables organization to centrally manage data and data security

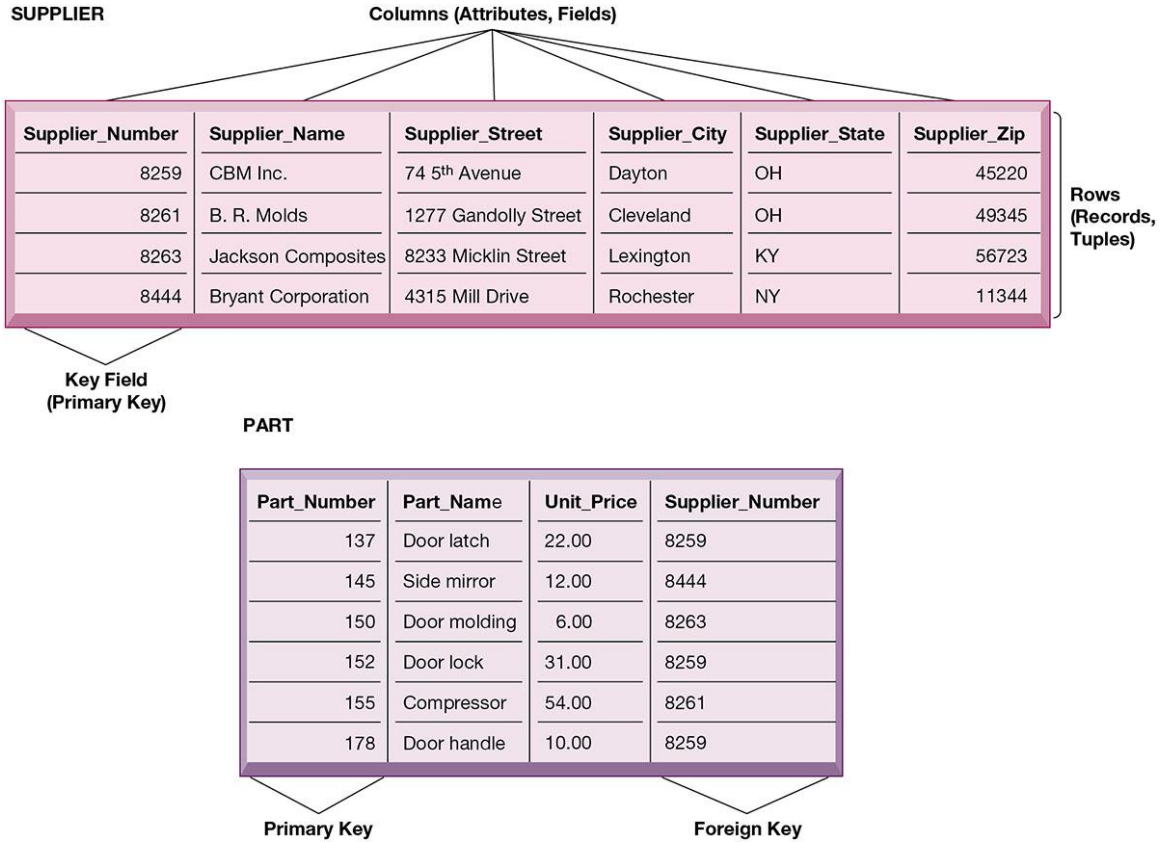
Figure 6.3 Human Resources Database with Multiple Views



Relational DBMS

- Represent data as two-dimensional tables
- Each table contains data on entity and attributes
- Table: grid of columns and rows
  - Rows (tuples): Records for different entities
  - Fields (columns): Represents attribute for entity
  - Key field: Field used to uniquely identify each record
  - Primary key: Field in table used for key fields
  - Foreign key: Primary key used in second table as look-up field to identify records from original table

Figure 6.4 Relational Database Tables



SUPPLIER

Supplier_Number	Supplier_Name	Supplier_Street	Supplier_City	Supplier_State	Supplier_Zip
8259	CBM Inc.	74 5th Avenue	Dayton	OH	45220
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345
8263	Jackson Composites	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344

PART

Part_Number	Part_Name	Unit_Price	Supplier_Number
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

Capabilities of Database Management Systems

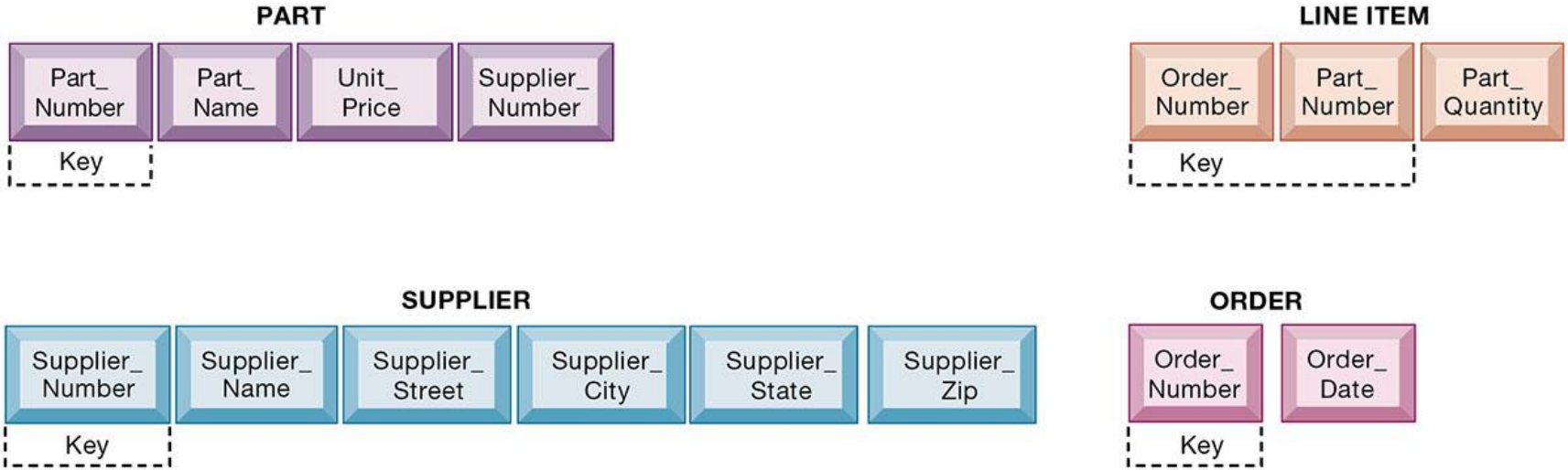
- Data definition
- Data dictionary
- Querying and reporting
  - Data manipulation language
    - Structured Query Language (SQL)
- Many DBMS have report generation capabilities for creating polished reports (Microsoft Access)

Designing Databases

- Conceptual design vs. physical design
- Normalization
  - Streamlining complex groupings of data to minimize redundant data elements and awkward many-to-many relationships
- Referential integrity
  - Rules used by RDBMS to ensure relationships between tables remain consistent
- Entity-relationship diagram
- A correct data model is essential for a system serving the business well

Figure 6.9 An Unnormalized Relation for Order

Figure 6.10 Normalized Tables Created from Order



The Challenge of Big Data

- Big data
  - Massive sets of unstructured/semi-structured data from web traffic, social media, sensors, and so on
- Volumes too great for typical DBMS
  - Petabytes, exabytes of data
- Can reveal more patterns, relationships and anomalies
- Requires new tools and technologies to manage and analyze

Business Intelligence Infrastructure (1 of 4)

- Array of tools for obtaining information from separate systems and from big data
  - Data warehouse
  - Data mart
  - Hadoop
  - In-memory computing
  - Analytical platforms

Business Intelligence Infrastructure (2 of 4)

- Data warehouse
  - Stores current and historical data from many core operational transaction systems
  - Consolidates and standardizes information for use across enterprise, but data cannot be altered
  - Provides analysis and reporting tools
- Data marts
  - Subset of data warehouse
  - Typically focus on single subject or line of business

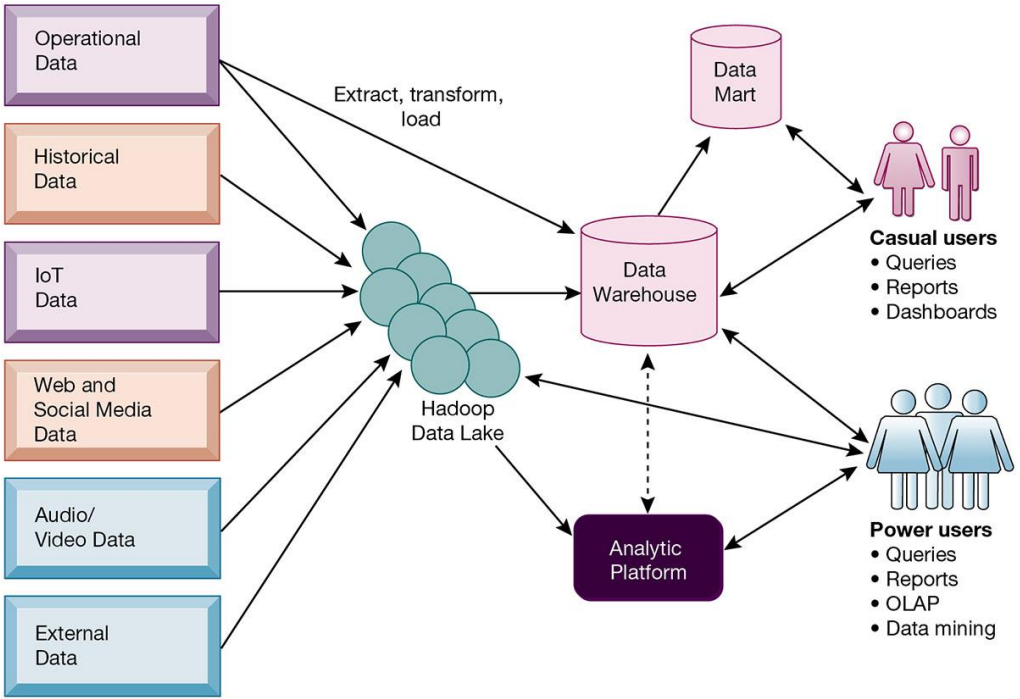
Business Intelligence Infrastructure (3 of 4)

- Hadoop
  - Enables distributed parallel processing of big data across inexpensive computers
  - Key services
    - Hadoop Distributed File System (HDFS): data storage
    - MapReduce: breaks data into clusters for work
    - Hbase: No SQL database
  - Used by Yahoo, NextBio

Business Intelligence Infrastructure (4 of 4)

- In-memory computing
  - Used in big data analysis
  - Uses computers main memory (RAM) for data storage to avoid delays in retrieving data from disk storage
  - Can reduce hours/days of processing to seconds
  - Requires optimized hardware
- Analytic platforms
  - High-speed platforms using both relational and non-relational tools optimized for large datasets

Figure 6.13 Contemporary Business Intelligence Infrastructure



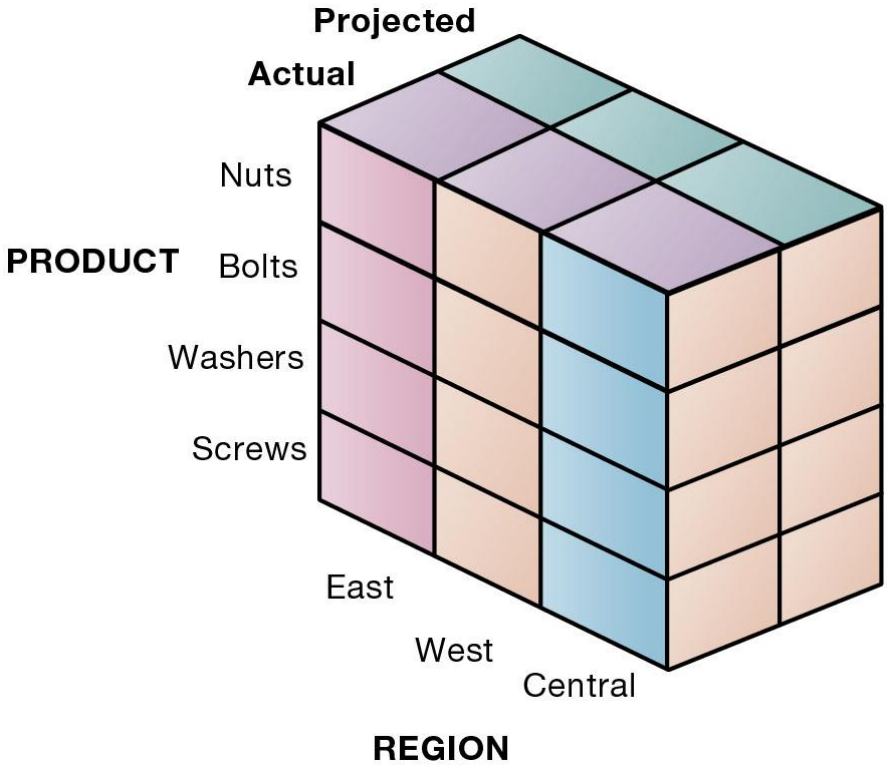
Analytical Tools: Relationships, Patterns, Trends

- Tools for consolidating, analyzing, and providing access to vast amounts of data to help users make better business decisions
  - Multidimensional data analysis (OLAP)
  - Data mining
  - Text mining
  - Web mining

Online Analytical Processing (OLAP)

- Supports multidimensional data analysis
  - Viewing data using multiple dimensions
  - Each aspect of information (product, pricing, cost, region, time period) is different dimension
  - Example: How many washers sold in the East in June compared to the sales forecast?
- OLAP enables rapid, online answers to ad hoc queries

Figure 6.14 Multidimensional Data Model



Data Mining

- Finds hidden patterns, relationships in datasets
  - Example: customer buying patterns
- Infers rules to predict future behavior
- Types of information obtainable from data mining:
  - Associations analyzing
  - Sequences
  - Classification
  - Clustering
  - Forecasting

Text Mining and Web Mining

- Text mining
  - Extracts key elements from large unstructured text data sets
  - Sentiment analysis software
- Web mining
  - Discovery and analysis of useful patterns and information from web
  - Web content mining
  - Web structure mining
  - Web usage mining

Databases and the Web

- Many companies use the web to make some internal databases available to customers or partners
- Advantages of using the web for database access:
  - Ease of use of browser software
  - Web interface requires few or no changes to database
  - Inexpensive to add web interface to system

Copyright

This work is protected by United States copyright laws and is provided solely for the use of instructors in teaching their courses and assessing student learning. Dissemination or sale of any part of this work (including on the World Wide Web) will destroy the integrity of the work and is not permitted. The work and materials from it should never be made available to students except by instructors using the accompanying text in their classes. All recipients of this work are expected to abide by these restrictions and to honor the intended pedagogical purposes and the needs of other instructors who rely on these materials.