Chapter 5

Foundations of Business Intelligence: Databases and Information Management

Video Cases:
Case 1 Maruti Suzuki Business Intelligence and Enterprise Databases
Case 2 Data Warehousing at REI: Understanding the Customer

Student Learning Objectives

- How does a relational database organize data, and how does it differ from an object-oriented database?
- What are the principles of a database management system?
- What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?

Student Learning Objectives

- What is the role of information policy and data administration in the management of organizational data resources?
- Why is data quality assurance so important for a business?

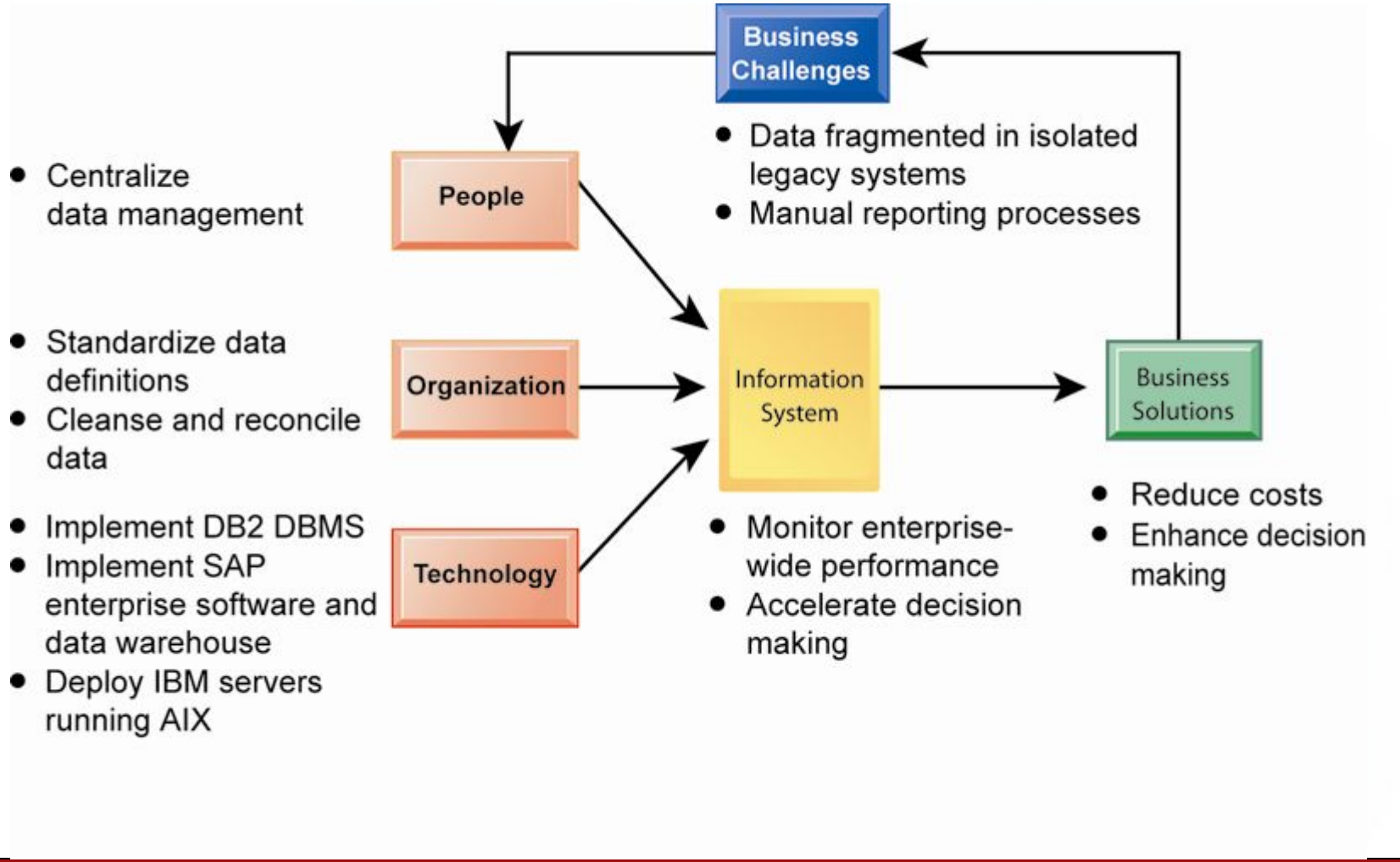
Banco de Credito Del Peru Banks on Better Data Management

- **Problem:** Multiple outdated systems, duplicate, inconsistent data
- **Solution:** Replace disparate legacy systems with single repository for business information

Banco de Credito Del Peru Banks on Better Data Management

- SAP integrated software suite included modules for enterprise resource planning and a data warehouse to support enterprise-wide, real-time tracking, reporting, and analysis
- Demonstrates IT's role in successful data management
- Illustrates digital technology's ability to lower costs while improving performance

Banco de Credito Del Peru Banks on Better Data Management



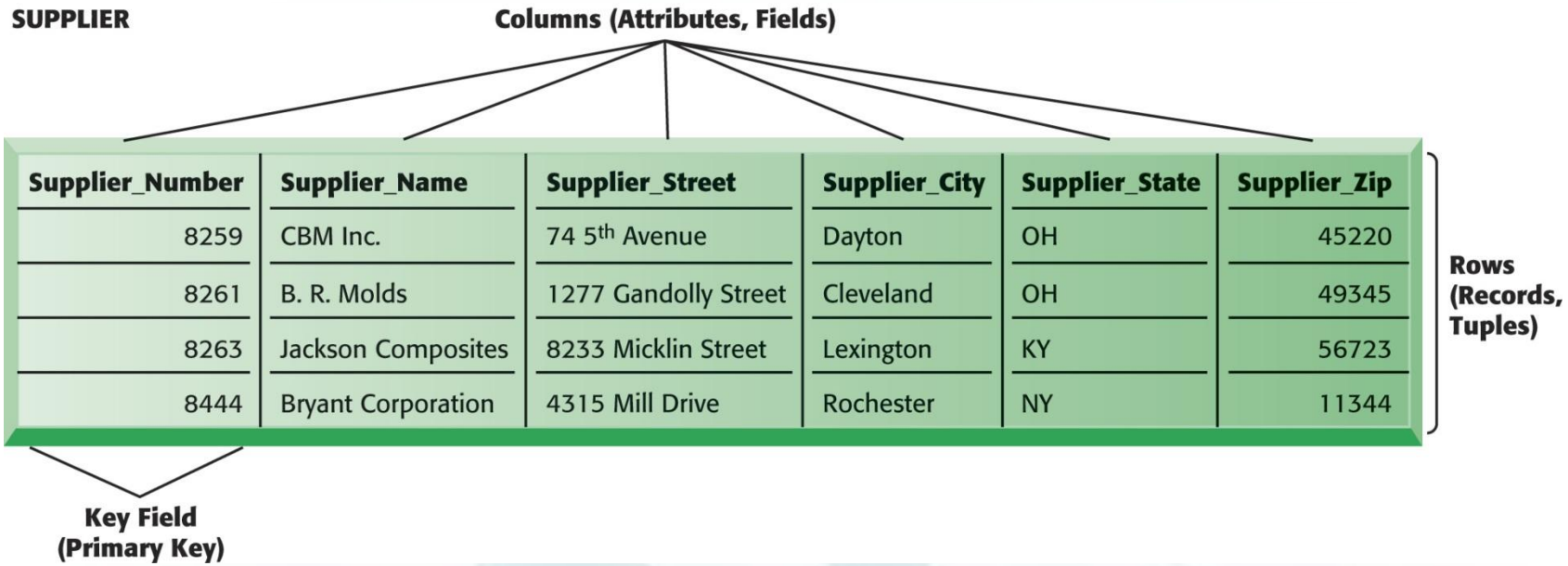
The Database Approach to Data Management

- **Database:**
  - Collection of related files containing records on people, places, or things
  - Prior to digital databases, business used file cabinets with paper files
- **Entity:**
  - Generalized category representing person, place, thing on which we store and maintain information
  - E.g., SUPPLIER, PART
- **Attributes:**
  - Specific characteristics of each entity:
    - SUPPLIER name, address
    - PART description, unit price, supplier

- **Relational database:**
  - Organize data into two-dimensional tables (relations) with columns and rows
  - One table for each entity:
    - E.g., (CUSTOMER, SUPPLIER, PART, SALES)
  - **Fields** (columns) store data representing an attribute.
  - Rows store data for separate **records**, or tuples.
  - **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

The Database Approach to Data Management

A Relational Database Table



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

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 Database Approach to Data Management

The PART Table

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

Primary Key

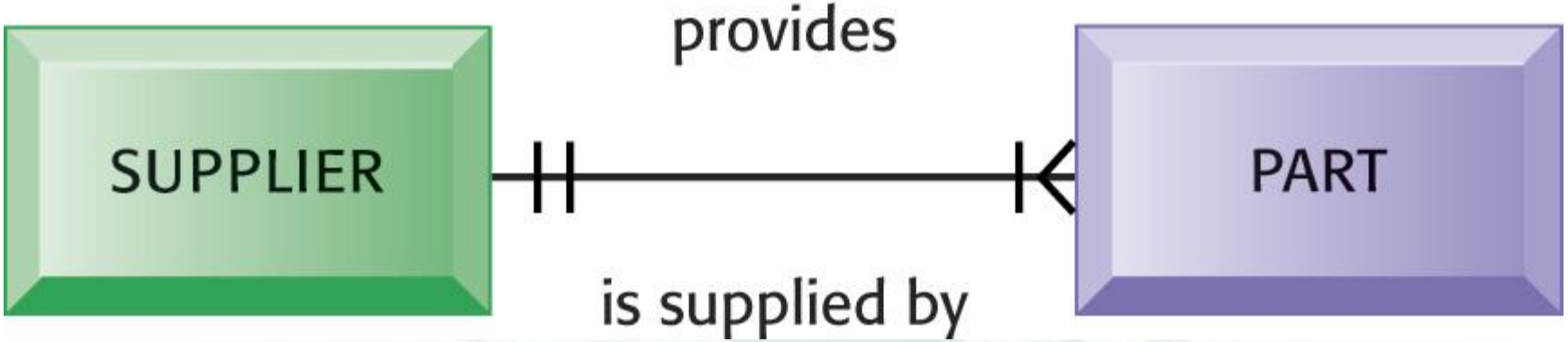
Foreign Key

Data for the entity PART have their own separate table. Part_Number is the primary key and Supplier_Number is the foreign key, enabling users to find related information from the SUPPLIER table about the supplier for each part.

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 “Join table” or 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

The Database Approach to Data Management

- **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
  - E.g., when one table has a foreign key that points to another table, you may not add a record to the table with foreign key unless there is a corresponding record in the linked table

The Database Approach to Data Management

The shaded areas show which data came from the SUPPLIER, LINE_ITEM, and ORDER tables. The database does not maintain data on Extended Price or Order Total because they can be derived from other data in the tables.

Sample Order Report

Order Number: 3502
Order Date: 1/15/2012

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

Order_Number	Part_Number	Part_Quantity	Part_Name	Unit_Price	Extended Price
3502	137	10	Door latch	22.00	$220.00
3502	152	20	Door lock	31.00	620.00
3502	178	5	Door handle	10.00	50.00
			Order Total:		$890.00

Figure 5-4

The Final Database Design with Sample Records

Figure 5-5

The final design of the database for suppliers, parts, and orders has four tables. The LINE_ITEM table is a join table that eliminates the many-to-many relationship between ORDER and PART.

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

LINE_ITEM

Order_Number	Part_Number	Part_Quantity
3502	137	10
3502	152	20
3502	178	5

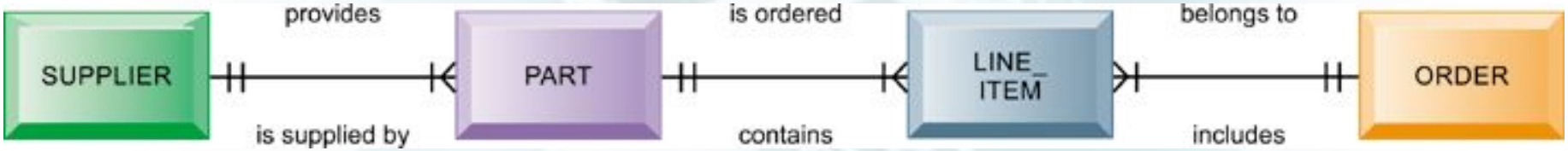
ORDER

Order_Number	Order_Date
3502	1/15/2012
3503	1/16/2012
3504	1/17/2012

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 Components	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344

Entity-Relationship Diagram for the Database with Four Tables



This diagram shows the relationship between the entities SUPPLIER, PART, LINE_ITEM, and ORDER.

Figure 5-6

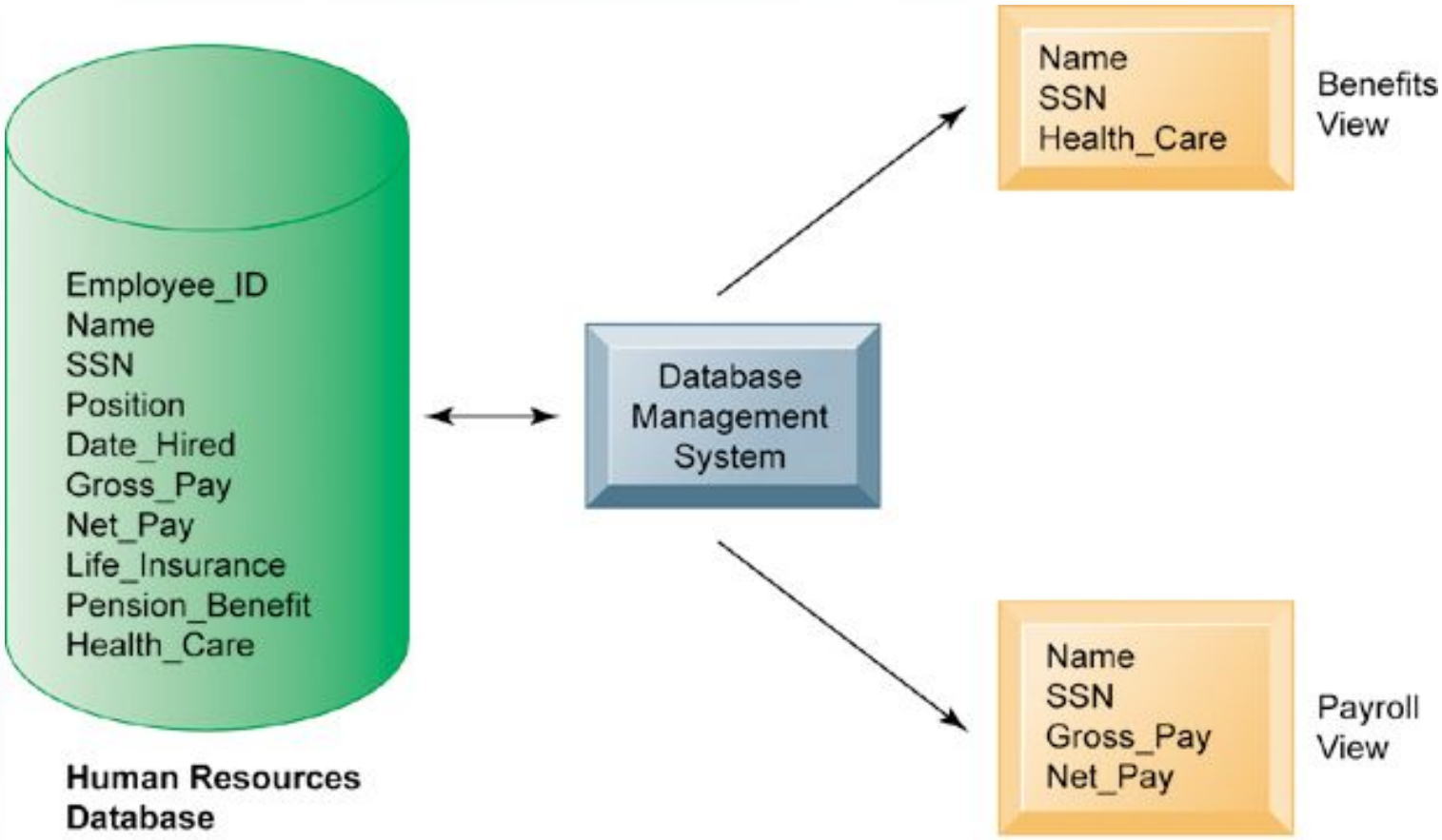
DBMS

- **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

A single human resources database provides many different views of data, depending on the information requirements of the user. Illustrated here are two possible views, one of interest to a benefits specialist and one of interest to a member of the company's payroll department.

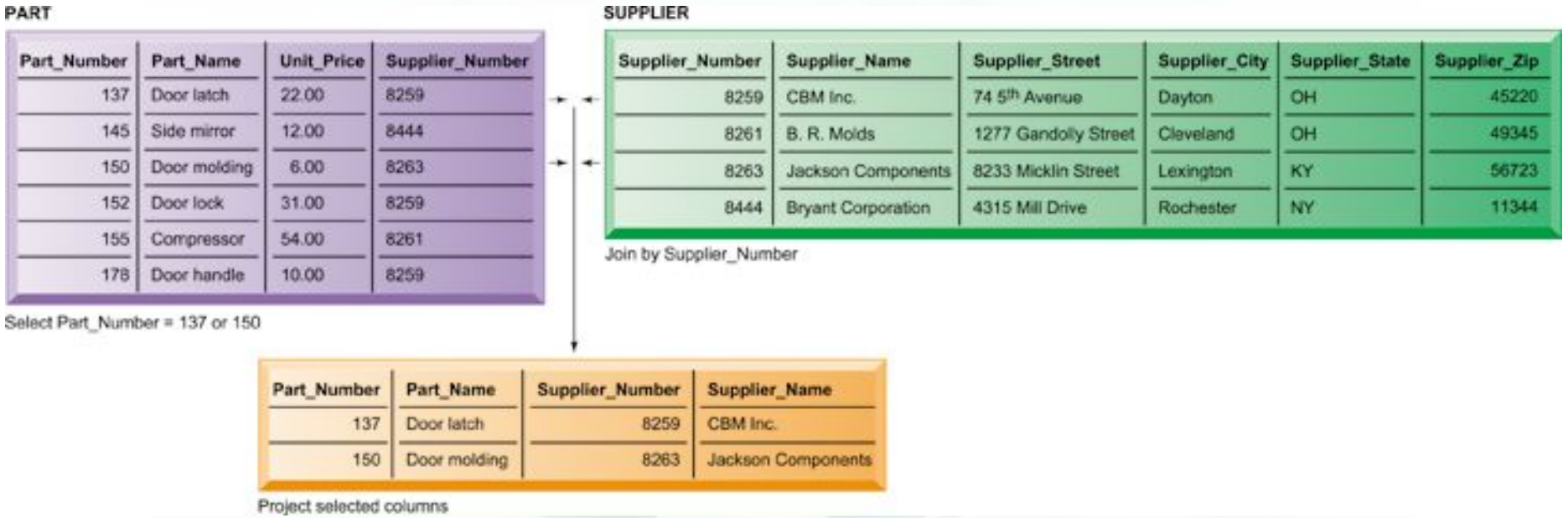


Operations of a Relational DBMS

- **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

Database Management Systems

The Three Basic Operations of a Relational DBMS



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
  - **Report generation, e.g., Crystal Reports**

Access Data Dictionary Features

Microsoft Access has a rudimentary data dictionary capability that displays information about the size, format, and other characteristics of each field in a database. Displayed here is the information maintained in the SUPPLIER table. The small key icon to the left of Supplier_Number indicates that it is a key field.

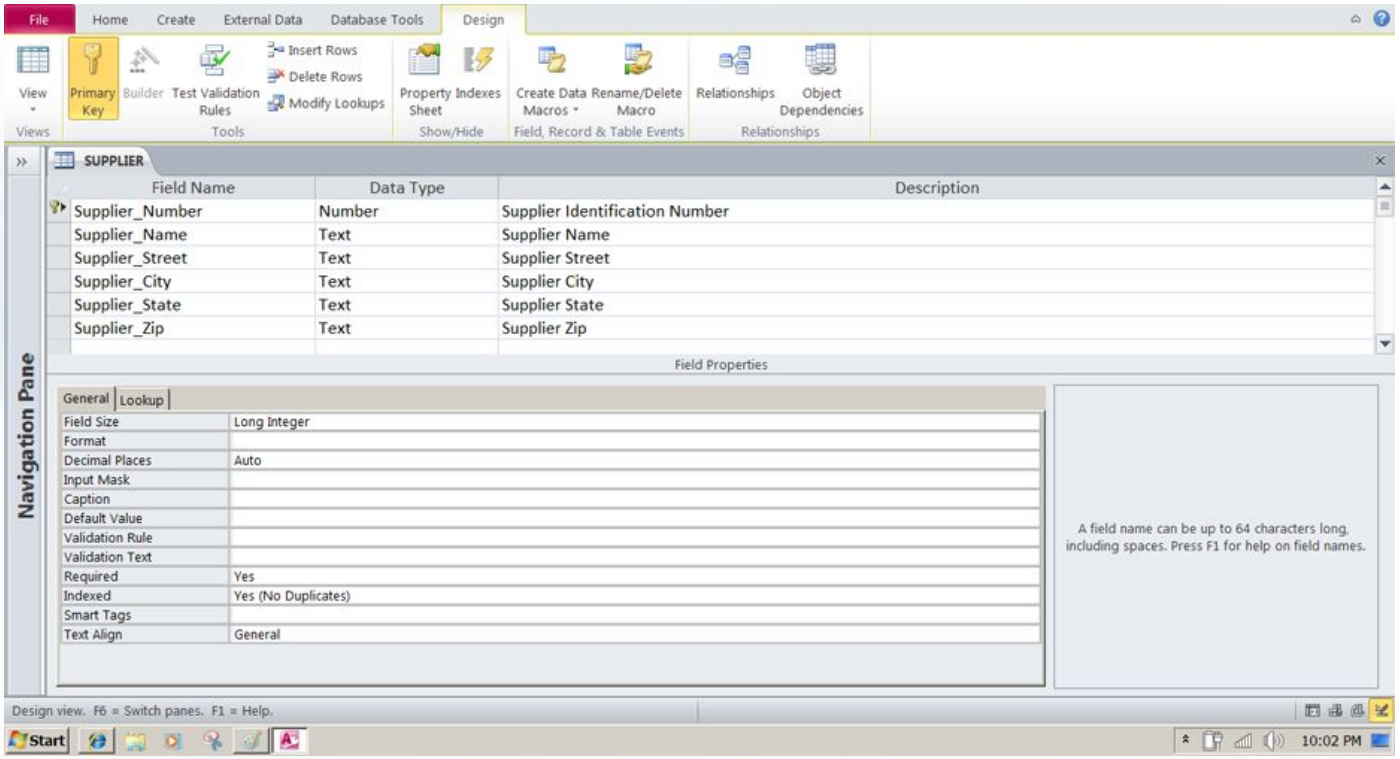


Figure 5-9

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

Illustrated here is how the query in Figure 5-10 would be constructed using Microsoft Access query-building tools. It shows the tables, fields, and selection criteria used for the query.

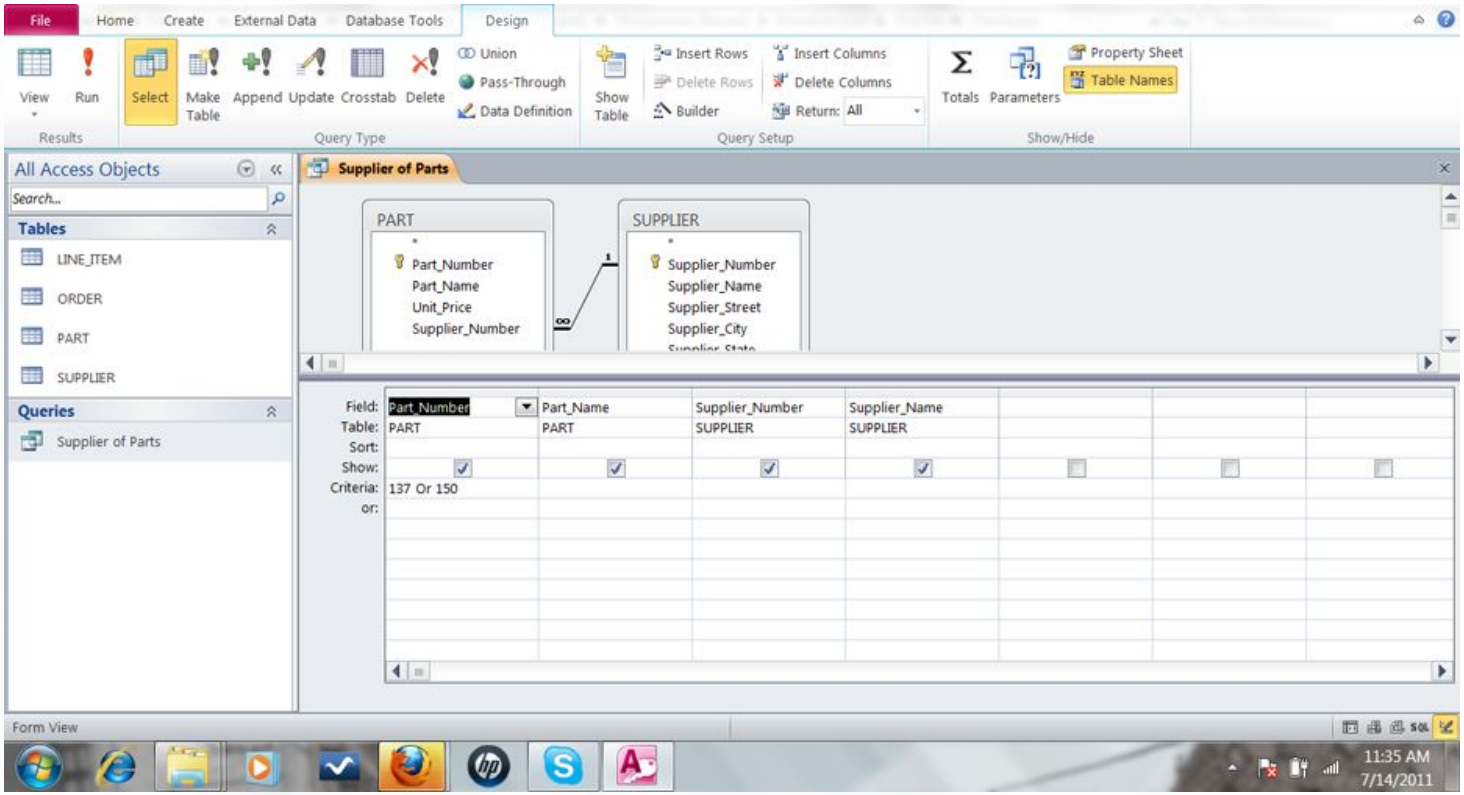


Figure 5-11

Object-Oriented DBMS (OODBMS)

- **Stores data and procedures that act on those data as objects to be retrieved and shared**
- **Used to manage multimedia components or Java applets in Web applications**
- **Relatively slow compared to relational DBMS**
- **Hybrid object-relational DBMS:** provide capabilities of both types

Databases in the Cloud

- Typically have less functionality than on-premises database services.

Using Databases to Improve Business Performance and Decision Making

- Databases provide information to help the company run the business more efficiently, and help managers and employees make better decisions
- Tools for analyzing, accessing vast quantities of data:
  - Data warehousing
  - Multidimensional data analysis
  - Data mining
  - Utilizing Web interfaces to databases

Data Warehouses

- **Data warehouse:**
  - Database that stores current and historical data that may be of interest to decision makers
  - Consolidates and standardizes data from many systems, operational and transactional databases
  - 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

The data warehouse extracts current and historical data from multiple operational systems inside the organization. These data are combined with data from external sources and reorganized into a central database designed for management reporting and analysis. The information directory provides users with information about the data available in the warehouse.

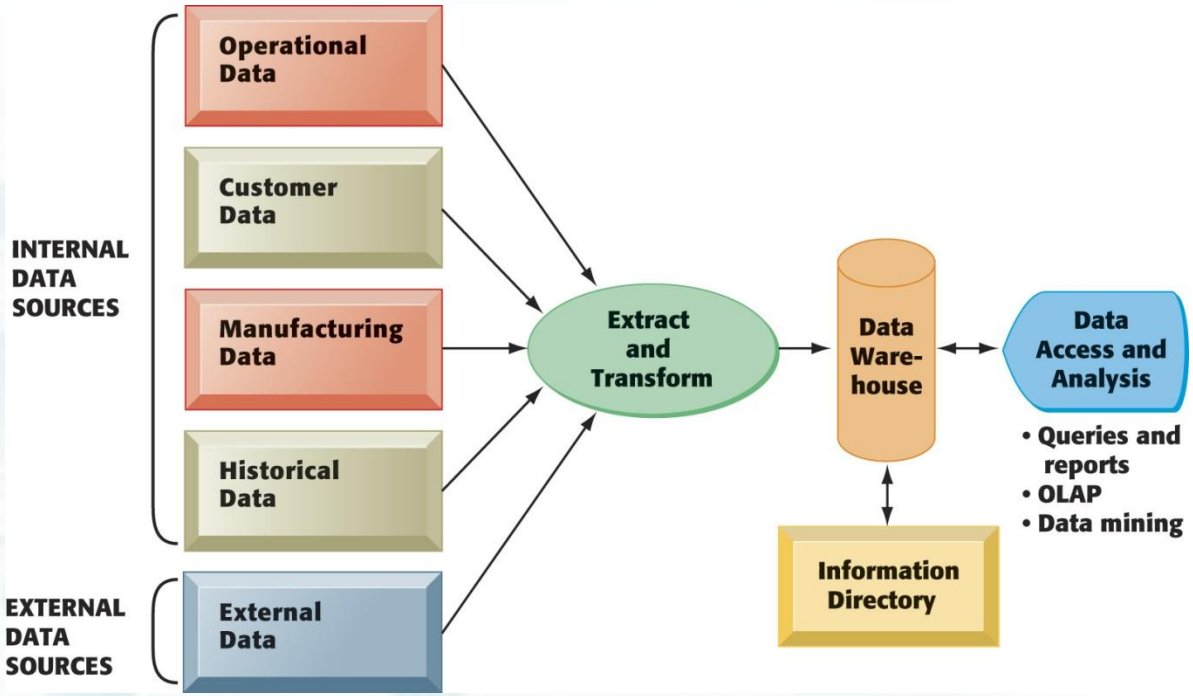


Figure 5-12

Business Intelligence, Multidimensional Data Analysis, and Data Mining

- **Business intelligence:** tools for consolidating, analyzing, and providing access to large amounts of data to improve decision making
  - Software for database reporting and querying
  - Tools for multidimensional data analysis (online analytical processing)
  - Data mining
- E.g., Harrah's Entertainment gathers and analyzes customer data to create gambling profile and identify most profitable customers

Online Analytical Processing (OLAP)

- **Supports multidimensional data analysis, enabling users to view the same data in different ways using multiple dimensions**
  - Each aspect of information—product, pricing, cost, region, or time period—represents a different dimension
  - E.g., comparing sales in East in June versus May and July
- **Enables users to obtain online answers to ad hoc questions such as these in a fairly rapid amount of time**

Multidimensional Data Model

The view that is showing is product versus region. If you rotate the cube 90 degrees, the face that will show is product versus actual and projected sales. If you rotate the cube 90 degrees again, you will see region versus actual and projected sales. Other views are possible.

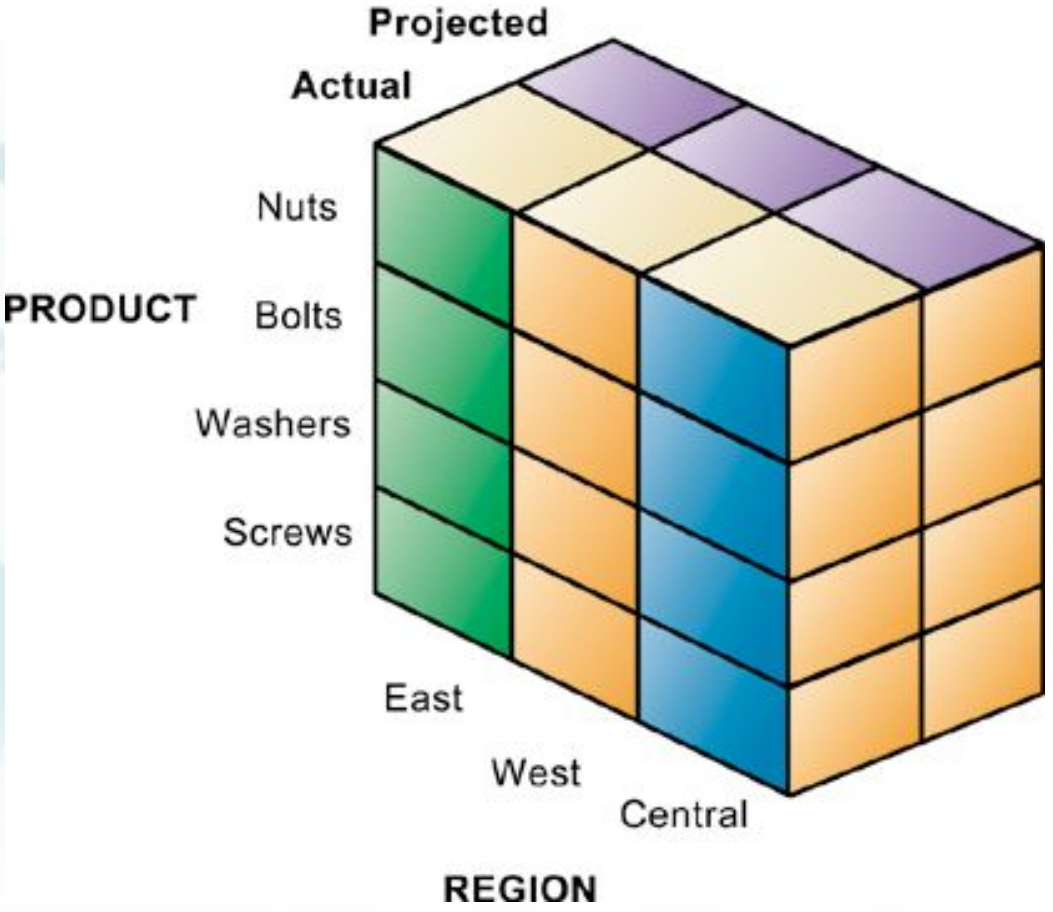


Figure 5-13

Data Mining

- **Finds hidden patterns and relationships in large databases and infers rules from them to predict future behavior**
- **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
  - **Clustering:** discovering as yet unclassified groupings
  - **Forecasting:** uses series of values to forecast future values

Interactive Session: People
Asking the Customer by Asking the Database

- **Read the Interactive Session and then discuss the following questions:**
- **Why would a customer database be so useful for a company such as Forbes or Kodak? What would happen if these companies had not kept their customer data in databases?**
- **List and describe two entities and several of their attributes that might be found in Kodak's's marketing database.**
- **How did better data management improve each company's business performance? Give examples of two decisions that were improved by mining these customer databases.**

Data Mining

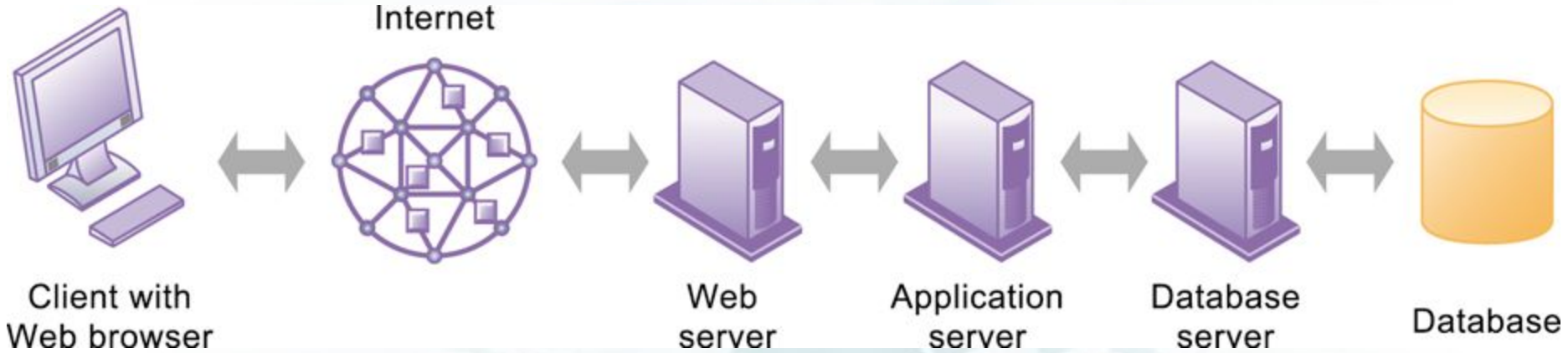
- **One popular use of data mining:** analyzing patterns in customer data for one-to-one marketing campaigns or for identifying profitable customers
- **Predictive analysis:**
  - Uses data mining techniques, historical data, and assumptions about future conditions to predict outcomes of events, such as the probability a customer will respond to an offer or purchase a specific product

- **Text Mining**
  - Unstructured data (mostly text files) accounts for 80% of an organization's useful information
  - Text mining allows businesses to 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
  - Content mining, structure mining, usage mining

Databases and the Web

- **Firms use the Web to make information from their internal databases available to customers and partners.**
- **Middleware and other software make this possible**
  - Web server
  - Application servers or CGI
  - Database server
- **Web interfaces provide familiarity to users and savings over redesigning and rebuilding legacy systems**

Linking Internal Databases to the Web



Users access an organization's internal database through the Web using their desktop PCs and Web browser software.

Figure 5-15

Establishing an Information Policy

- **Information policy**
  - States organization's rules for organizing, managing, storing, sharing information
- **Data administration**
  - Responsible for specific policies and procedures through which data can be managed as a resource
- **Database administration**
  - Database design and management group responsible for defining and organizing the structure and content of the database, and maintaining the database.

Ensuring Data Quality

- **Poor data quality:** major obstacle to successful customer relationship management
- **Data quality problems** caused by:
  - Redundant and inconsistent data produced by multiple systems
  - Data input errors
- **Data quality audit:** structured survey of the accuracy and completeness of data
- **Data cleansing:** detects and corrects incorrect, incomplete, improperly formatted, and redundant data

Interactive Session: Organizations
Controversy Whirls Around the CPSC Database

- **Read the Interactive Session and then discuss the following questions:**
  - **What is the value of the CPSC database to consumers, businesses, and the U.S. government?**
  - **What problems are raised by this database? Why is it so controversial? Why is data quality an issue?**
  - **Name two entities in the CPSC database and describe some of their attributes.**
  - **When buying a crib or other product, would you use this database?**