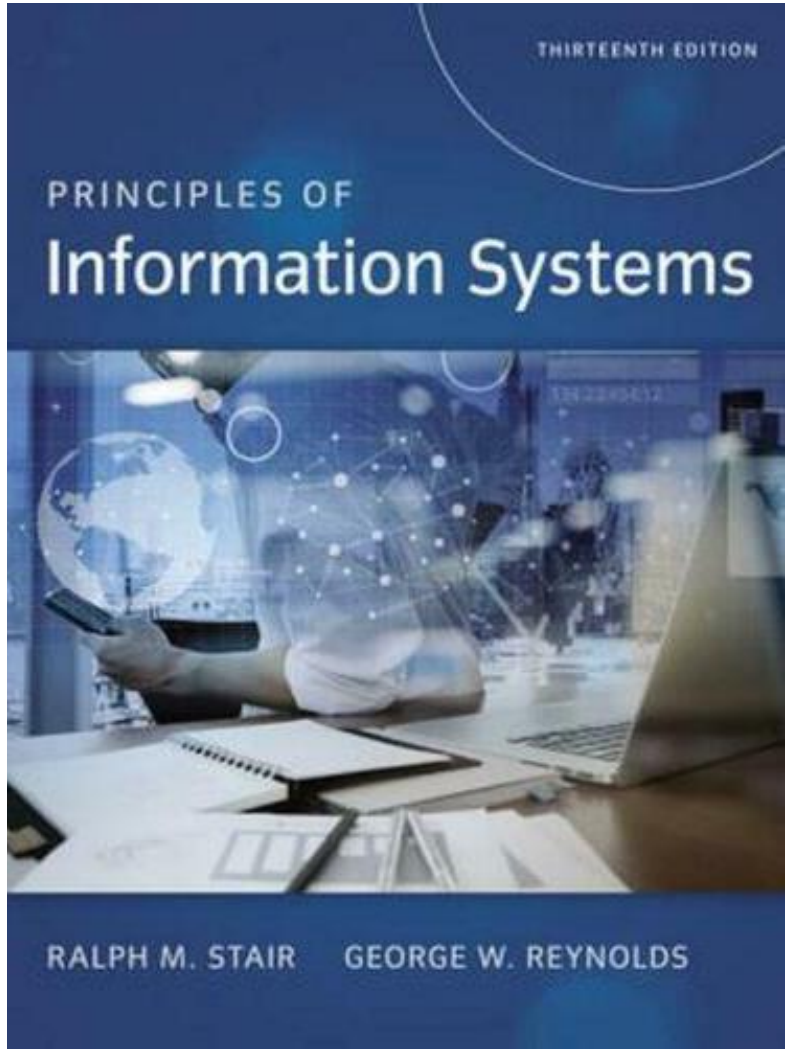
IS101 Principles of Information Systems

Database System and Big Data

Lecturer: Dr Maya Krayneva

Textbook: Stair, R., & Reynolds, G. (2016). Principles of information systems (13th ed.). Cengage Learning.

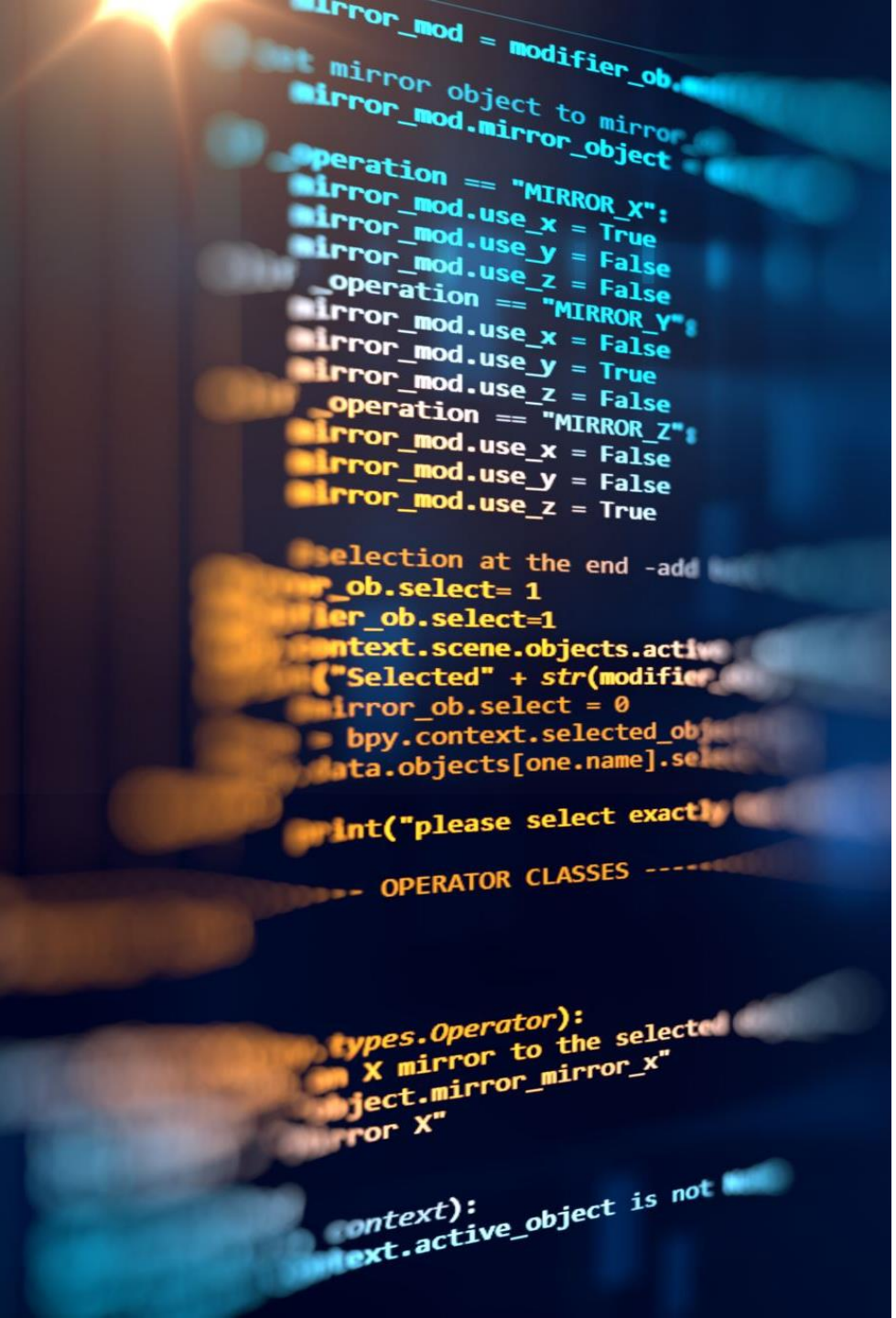
SHERIDAN
INSTITUTE OF HIGHER EDUCATION



Chapter 5: Database System and Big Data

Principles of information systems

Thirteen Edition



Introduction

- DATABASE: an organized collection of data

  Data must be organized in a meaningful way to transform it into useful information

→ DATABASE MANAGEMENT SYSTEM (DBMS) is a group of programs that:

- Manipulate the database
- Provide an interface between the database and its users and other application programs
  - FRONT-END applications interact directly with people versus BACK-END applications interact with other programs or applications

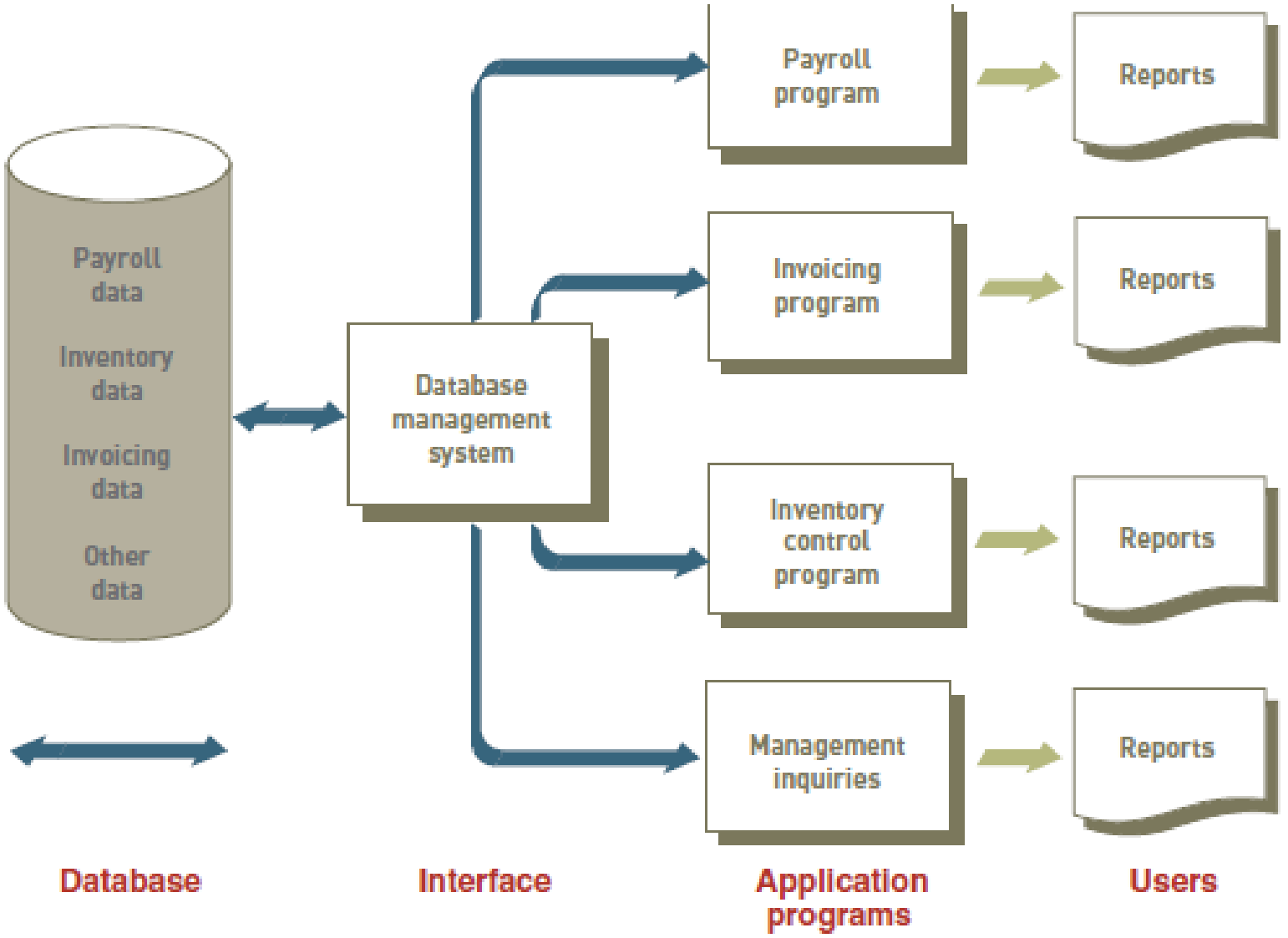


FIGURE 5.4

Database approach to data management

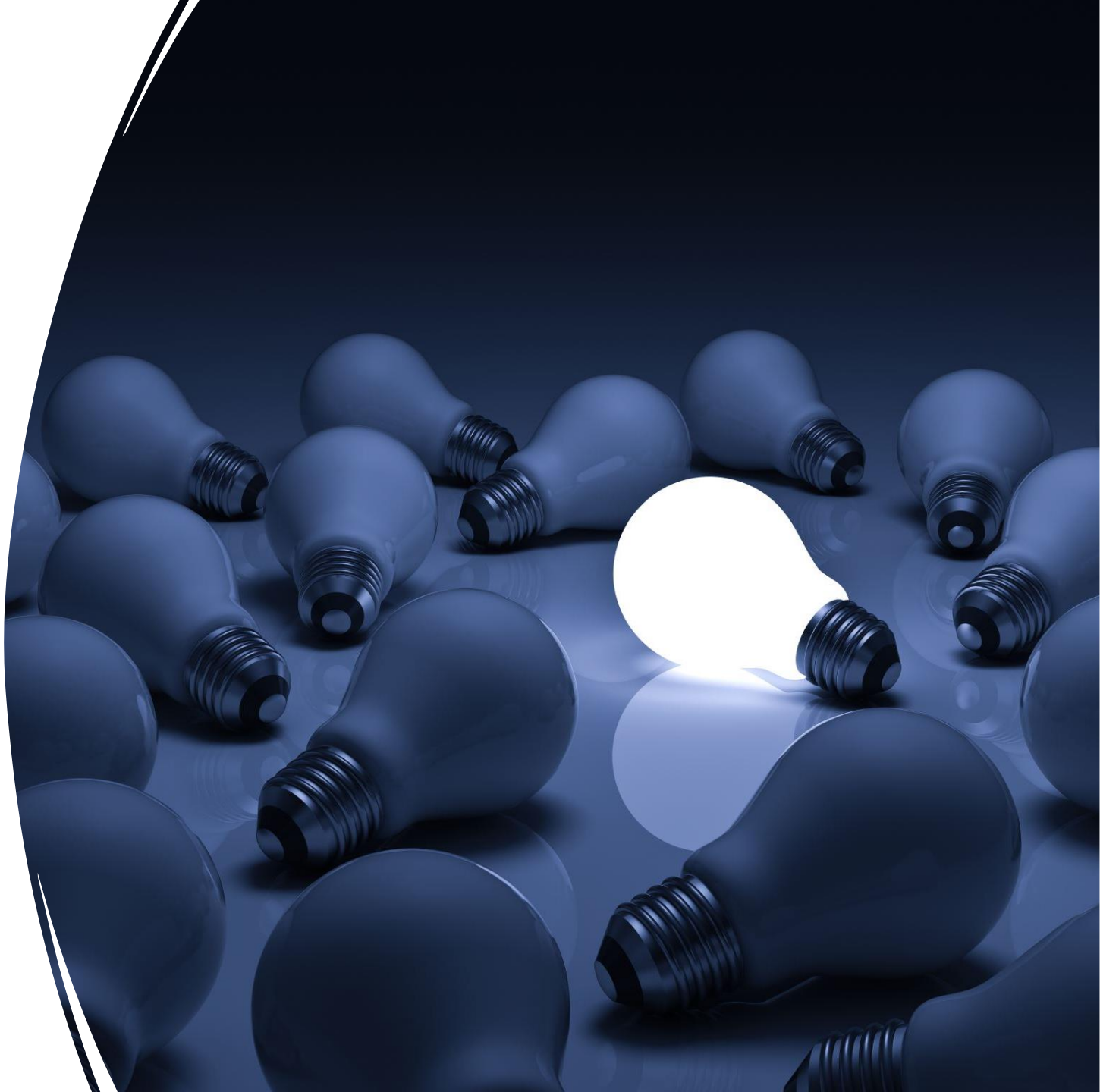
In a database approach to data management, multiple information systems share a pool of related data.

Considerations when building a database

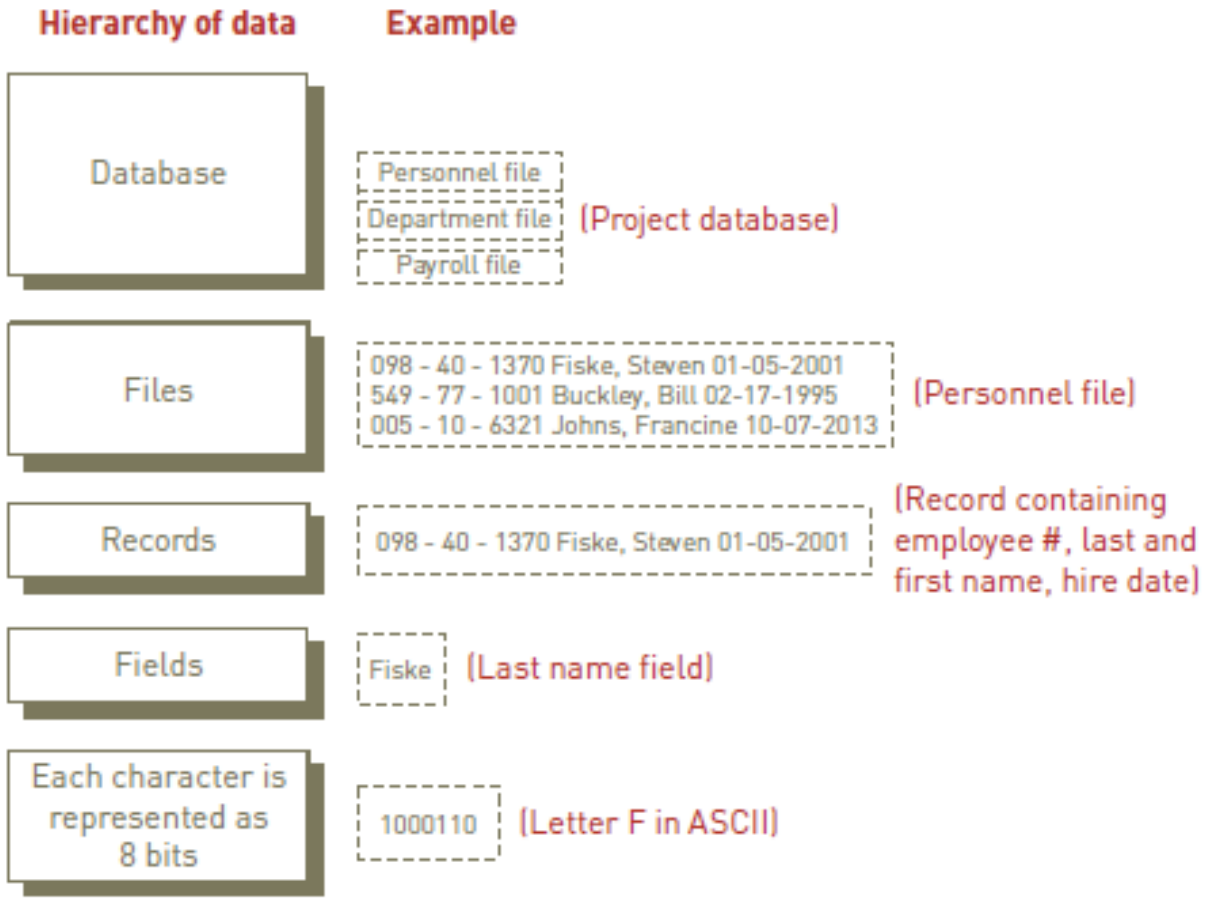
- CONTENT: what data should be **collected**? What is the cost?
- ACCESS: what data should be provided to which **users** and when?
- LOGICAL STRUCTURE: how should data be **arranged** so that it makes sense?
- PHYSICAL ORGANIZATION: where should data be physically **located**?
- ARCHIVING: how long to **store**?
- SECURITY: how can data be **protected**?

Describe a database that makes part of your IS solution.

- What data will be stored in this database?
- How the data will be organized and related?
- How the data will be used?
- What kind of decisions will be supported through the data analysis?

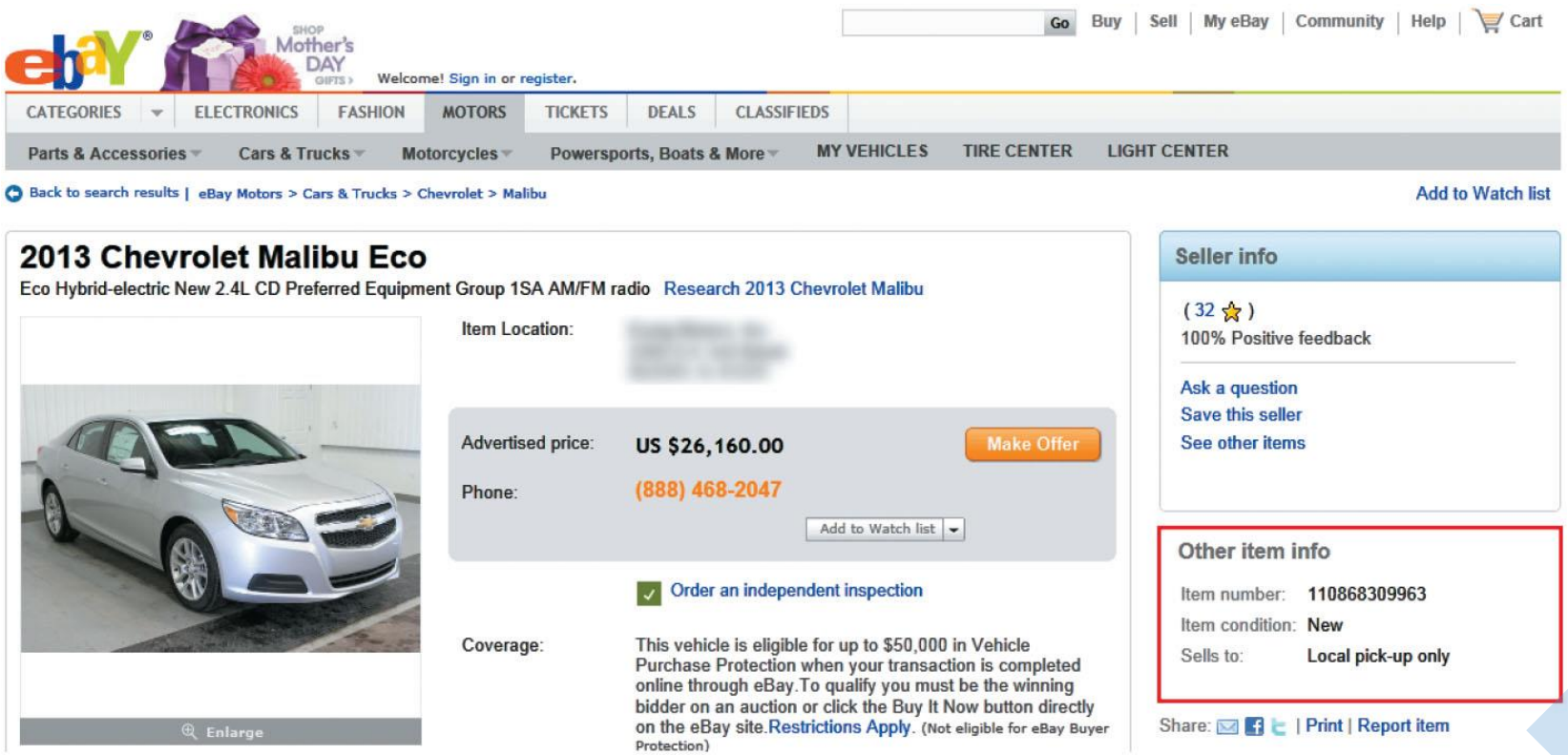


Hierarchy of Data



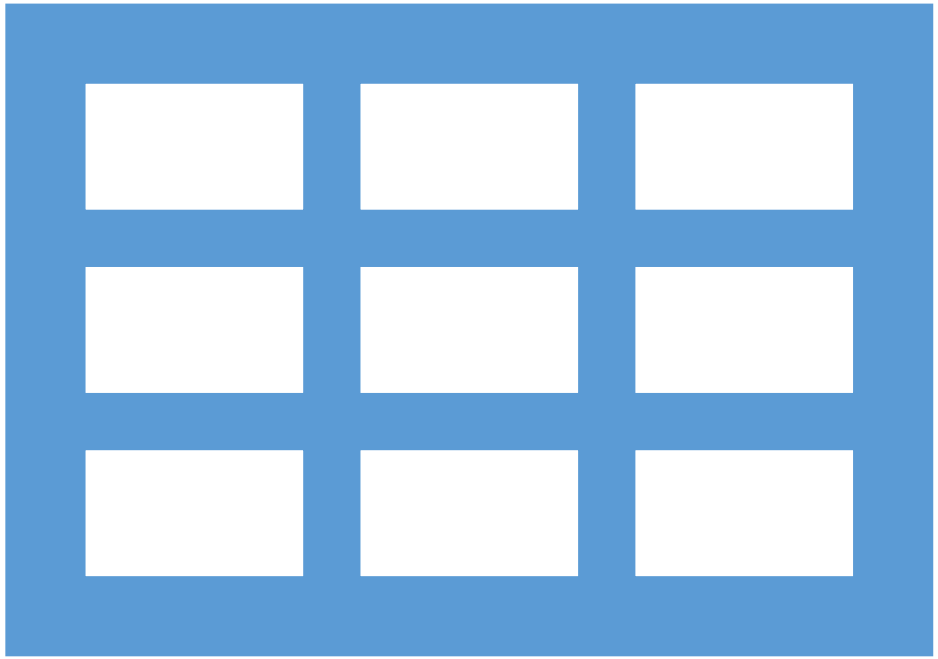
- FILE: a collection of related records
- RECORD: a collection of related data fields
- FIELD: a name, number, or combination of characters that describes an aspect of a business object or activity
- A BYTE is made up of eight bits: Each byte represents a character
- A BIT (binary digit) represents a circuit that is either on or off

- PRIMARY KEY: a field or set of fields that uniquely identifies the record
- FOREIGN KEY: Attribute in one table that refers to the primary key in another table



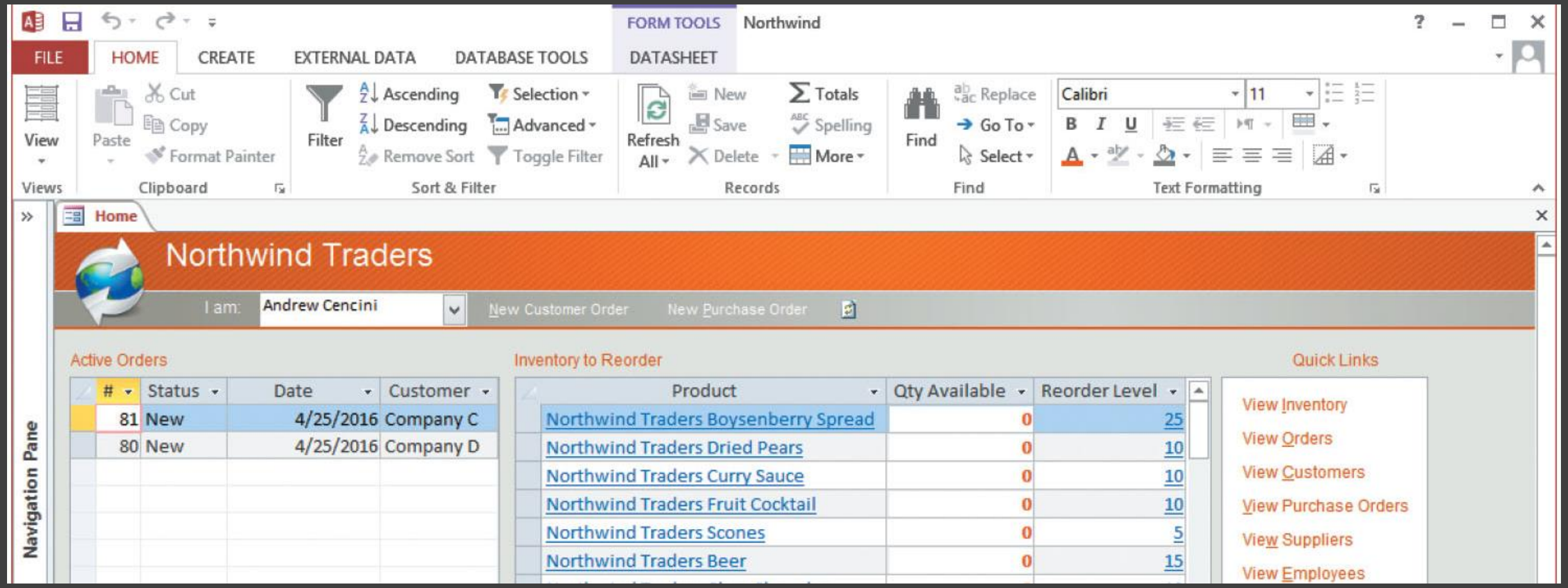
Example Primary key: eBay assigns an item number as a primary key to keep track of each item in its database.

Relational Database Model



Relational model: a simple but highly useful way to organize data into collections of **two-dimensional tables** called relations

- Each row in the table represents an **entity**
- Each column represents an **attribute** of that entity



- Screenshot of Microsoft Access, a relational database management system that has tools, tips, and shortcuts which simplifies the process of creating and modifying a database.

Data Entities and Attributes

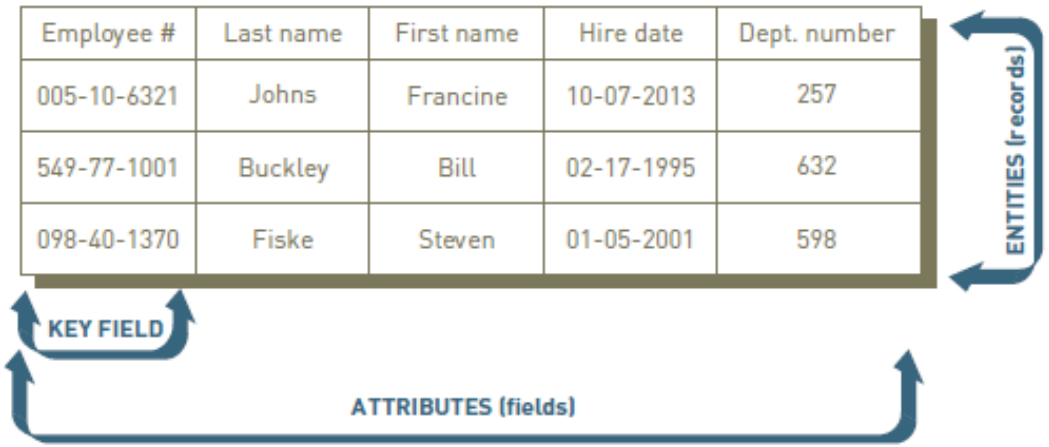


FIGURE 5.2
Keys and attributes
The key field is the employee number. The attributes include last name, first name, hire date, and department number.

- ENTITY: a person, place, or thing for which data is collected, stored, and maintained
- ATTRIBUTE: a characteristic of an entity
- DATA ITEM: the specific value of an attribute
- DOMAIN: Range of allowable values for a data attribute

Data Table 1: Project Table

Project	Description	Dept. number
155	Payroll	257
498	Widgets	632
226	Sales manual	598

Data Table 2: Department Table

Dept.	Dept. name	Manager SSN
257	Accounting	005-10-6321
632	Manufacturing	549-77-1001
598	Marketing	098-40-1370

Data Table 3: Manager Table

SSN	Last name	First name	Hire date	Dept. number
005-10-6321	Johns	Francine	10-07-2013	257
549-77-1001	Buckley	Bill	02-17-1995	632
098-40-1370	Fiske	Steven	01-05-2001	598

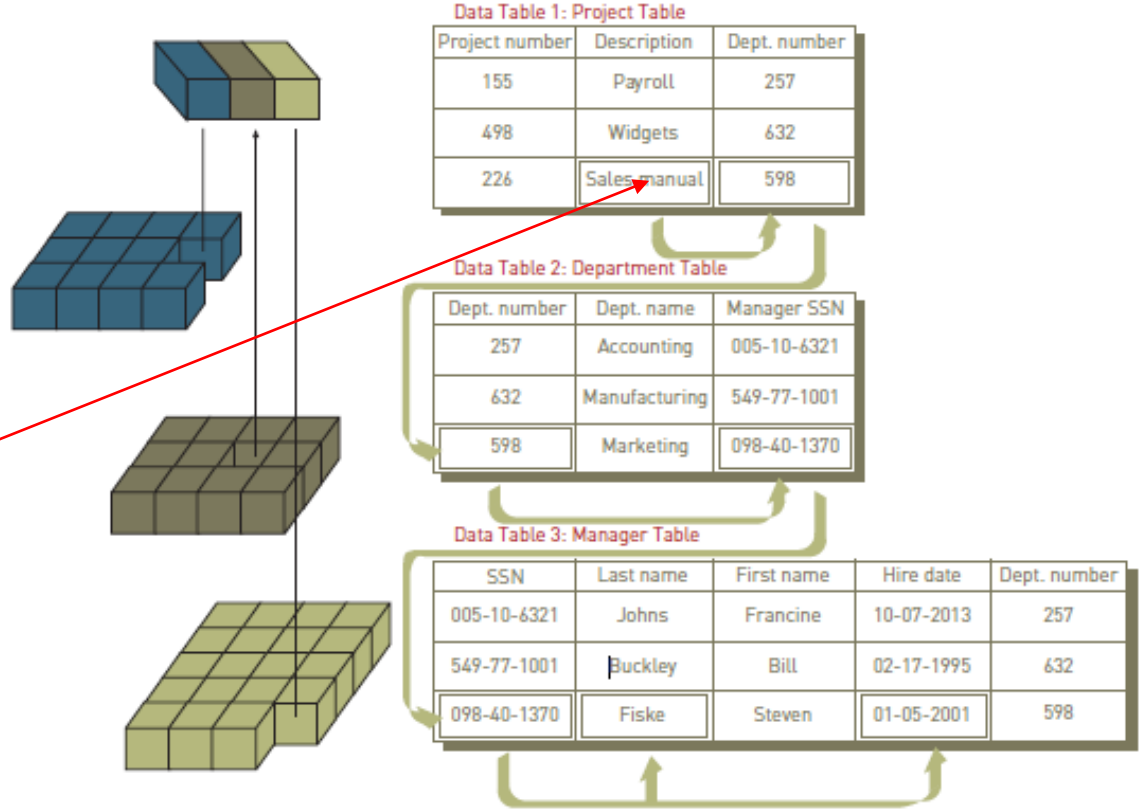
FIGURE 5.7

Relational database model

In the relational model, data is placed in two-dimensional tables, or relations. As long as they share at least one common attribute, these relations can be linked to provide output useful information. In this example, all three tables include the Dept. number attribute.

Manipulating Data

- SELECTING: **eliminating** rows according to certain criteria
- PROJECTING: eliminating columns in a table
- JOINING: **combining** two or more tables
- LINKING: combining two or more tables through common data attributes to form a new table with only the unique data attributes



Data Table 1: Project Table

Project number	Description	Dept. number
155	Payroll	257
498	Widgets	632
226	Sales manual	598

Data Table 2: Department Table

Dept. number	Dept. name	Manager SSN
257	Accounting	005-10-6321
632	Manufacturing	549-77-1001
598	Marketing	098-40-1370

Data Table 3: Manager Table

SSN	Last name	First name	Hire date	Dept. number
005-10-6321	Johns	Francine	10-07-2013	257
549-77-1001	Buckley	Bill	02-17-1995	632
098-40-1370	Fiske	Steven	01-05-2001	598

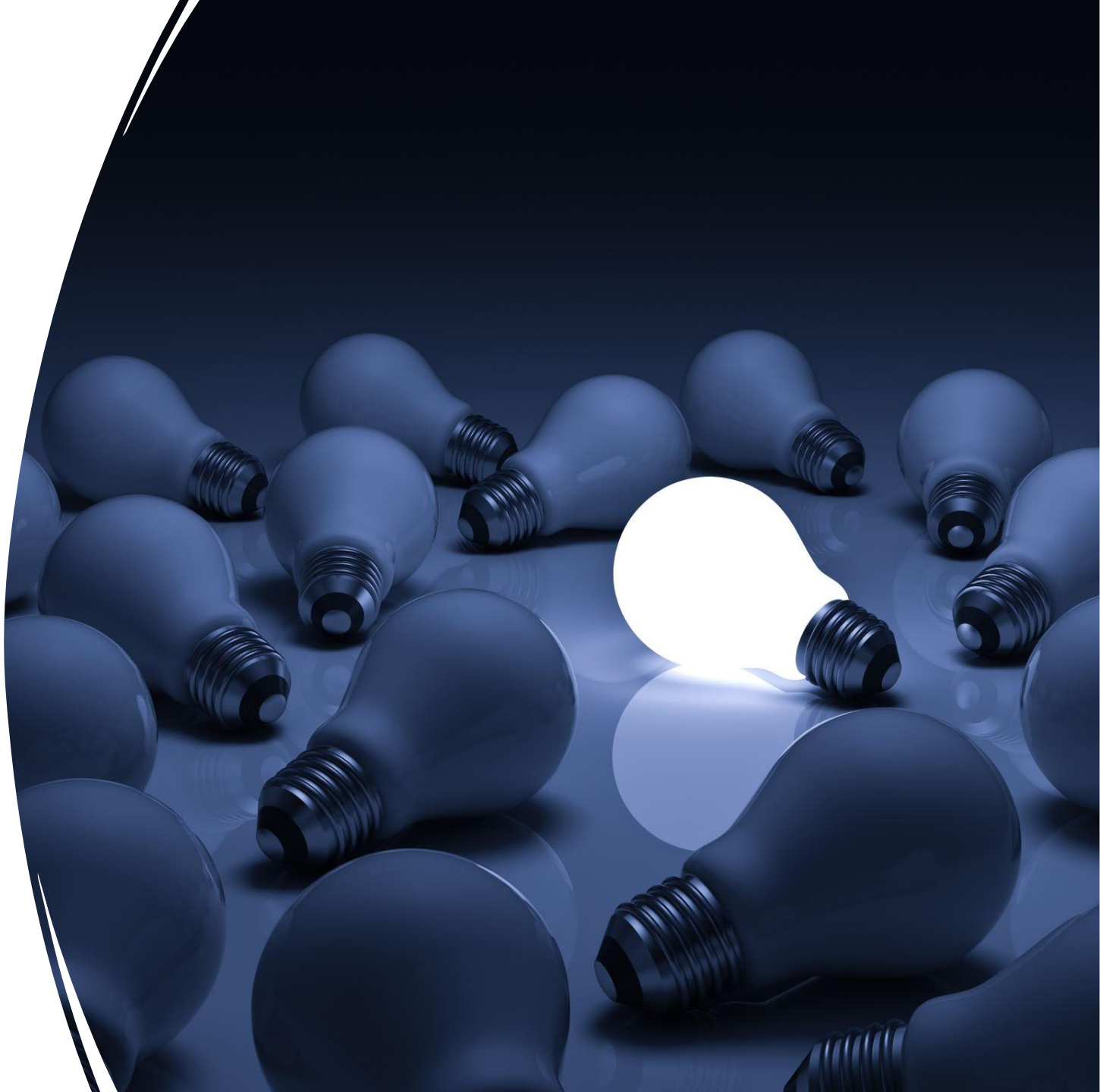
FIGURE 5.9

Linking data tables to answer an inquiry

To find the name and hire date of the manager working on the sales manual project, the president needs three tables: Project, Department, and Manager. The project description (Sales manual) leads to the department number (598) in the Project table, which leads to the manager's Social Security number (098-40-1370) in the Department table, which leads to the manager's last name (Fiske) and hire date (01-05-2001) in the Manager table.

Give an example of a table contained in your database.

Illustrate entities, attributes, data items and primary keys in this table (similar to Fig 5.2).



Data Modeling

- Data model: a diagram of data entities and their relationships

→ ENTERPRISE DATA MODELING: data modeling done at the level of the **entire enterprise**

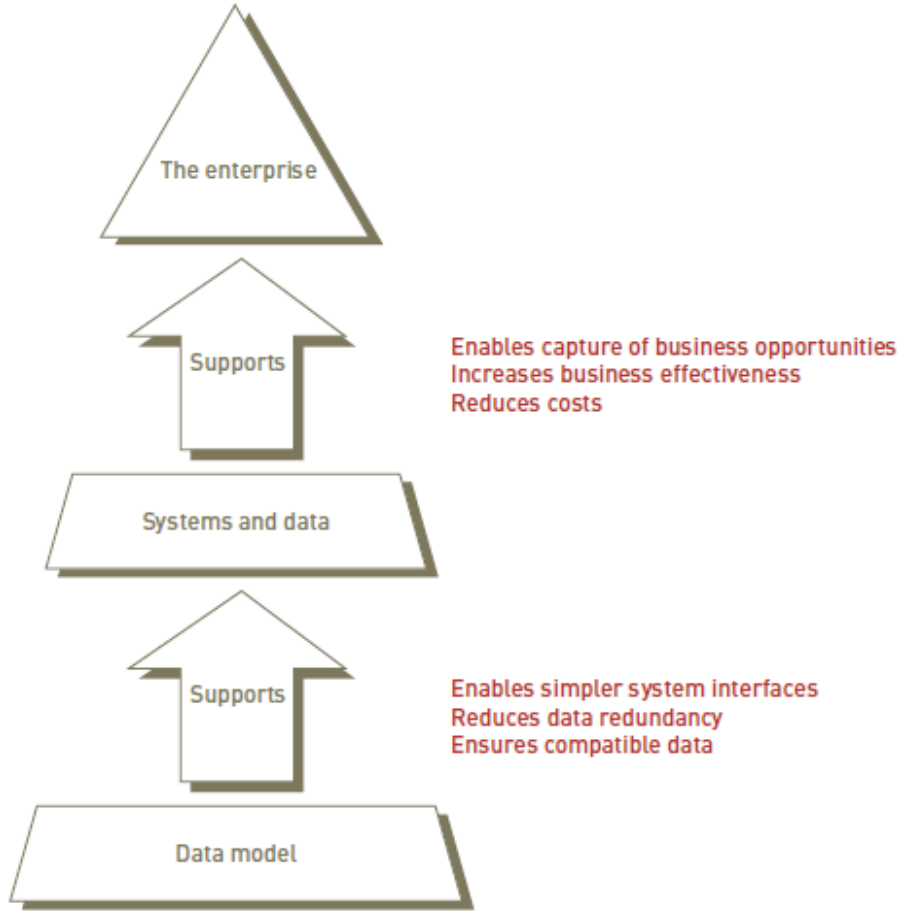


FIGURE 5.5
Enterprise data model
The enterprise data model provides a roadmap for building database and information systems.

Data Modeling: Entity-relationship (ER) diagrams

- Use of basic **graphical symbols**
- WATCH VIDEO

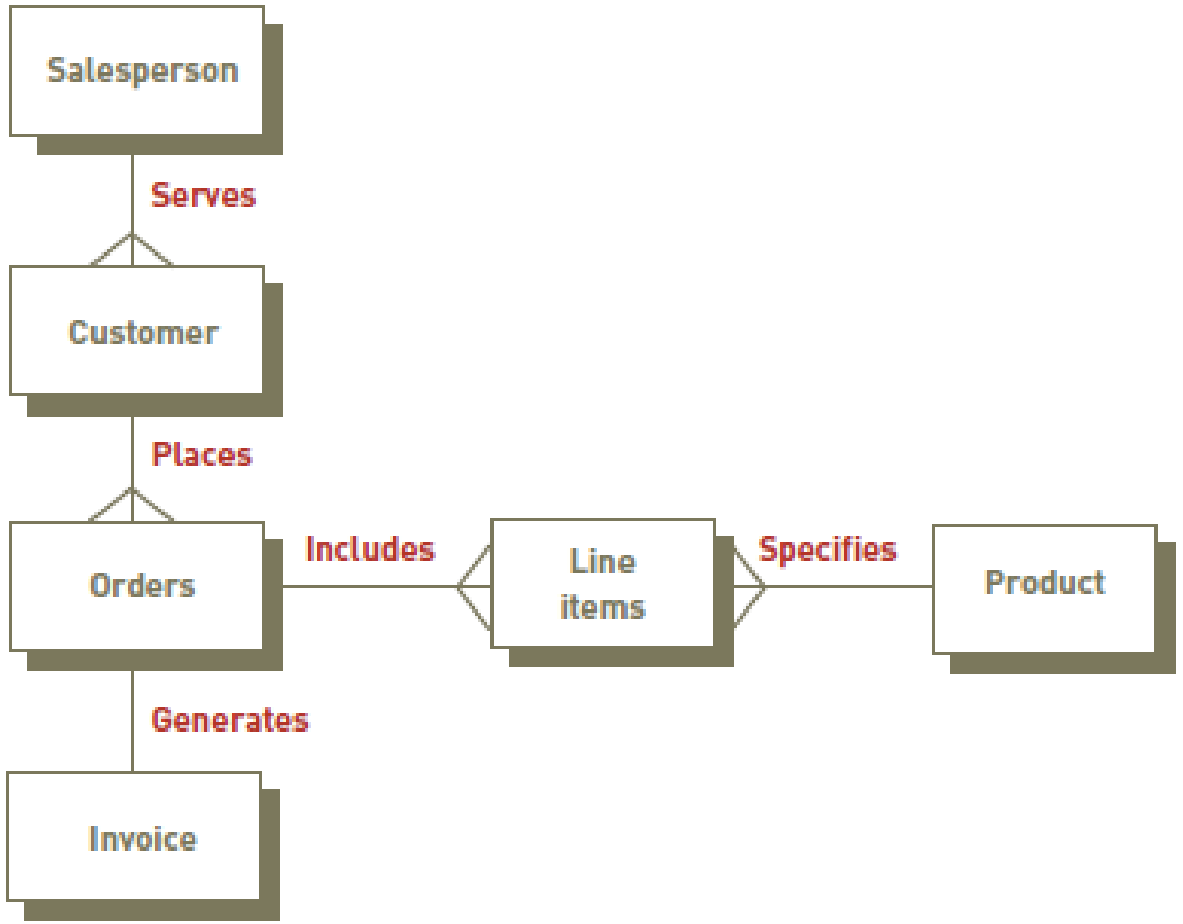


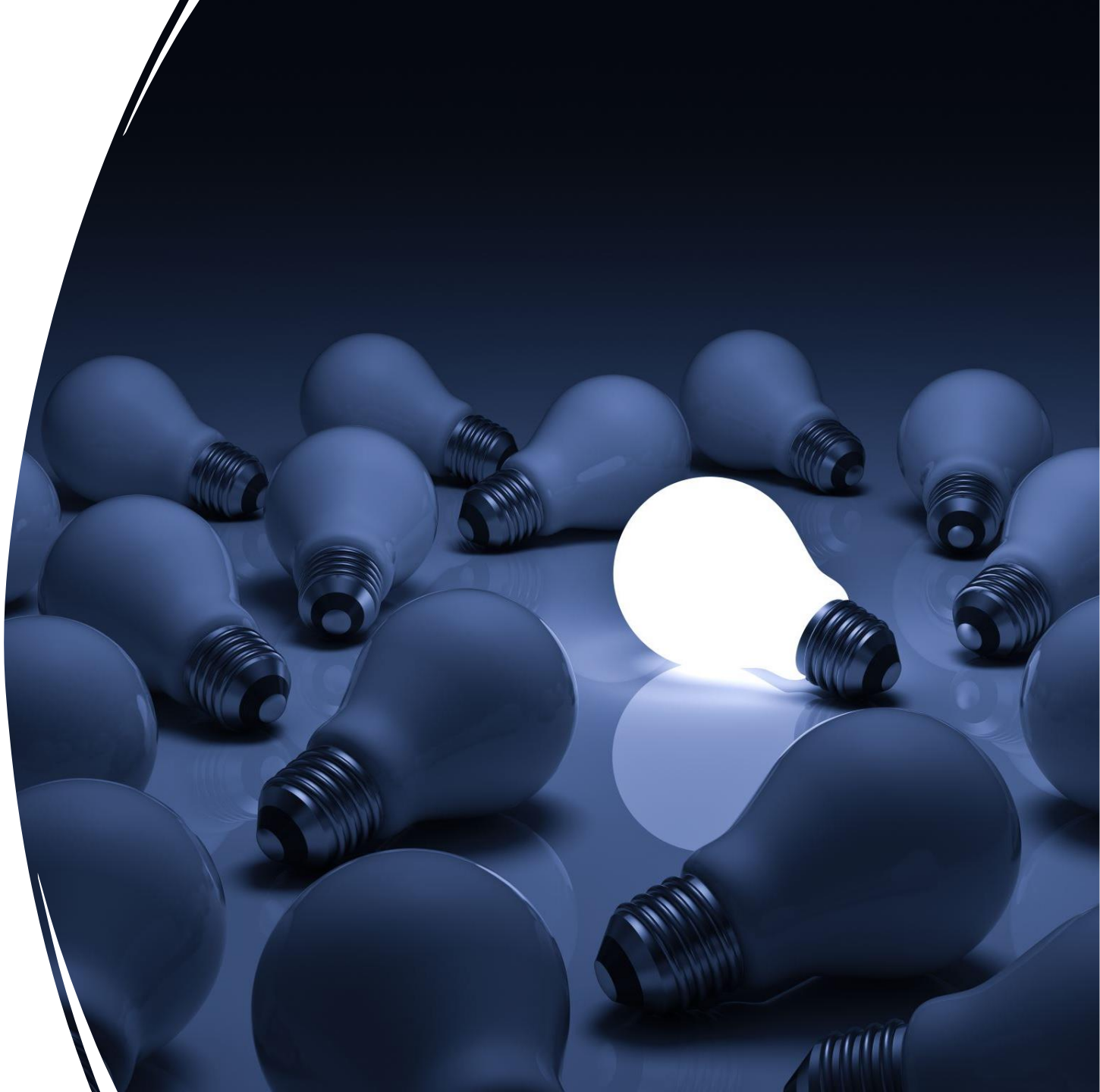
FIGURE 5.6

Entity-relationship (ER) diagram for a customer order database

Development of ER diagrams helps ensure that the logical structure of application programs is consistent with the data relationships in the database.

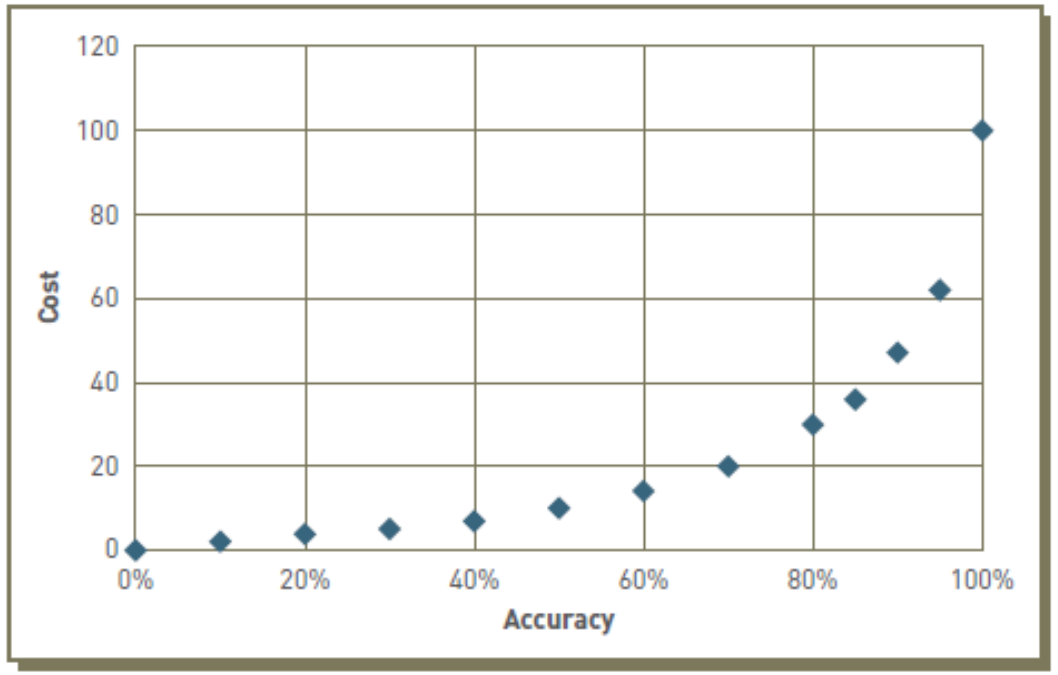
Draw an entity-relationship diagram of your database.

Include all entities and relationships between them (similar to Fig 5.6).



Data Cleansing

FIGURE 5.11
Tradeoff of cost versus accuracy
The cost of performing data cleansing to achieve 100 percent database accuracy can be prohibitively expensive.



DATA CLEANSING (data cleaning or data scrubbing): the process of detecting and then correcting or deleting incomplete, incorrect, inaccurate, irrelevant records that reside in a database

SQL Databases

SQL: a special-purpose programming language for accessing and manipulating data stored in a relational database

1986: SQL was adopted by ANSI as the standard query language for relational databases

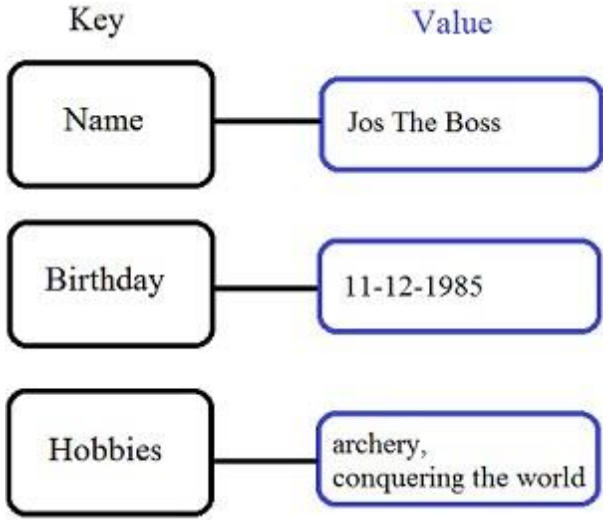
TABLE 5.1 Examples of SQL commands

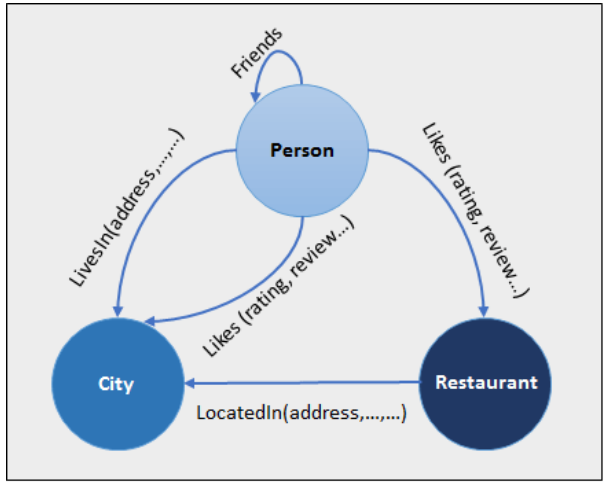
SQL Command	Description
SELECT ClientName, Debt FROM Client WHERE Debt > 1000	This query displays clients (ClientName) and the amount they owe the company (Debt) from a database table called Client; the query would only display clients who owe the company more than $1,000 (WHERE Debt > 1000).
SELECT ClientName, ClientNum, OrderNum FROM Client, Order WHERE Client.Client-Num=Order.ClientNum	This command is an example of a join command that combines data from two tables: the Client table and the Order table (FROM Client, Order). The command creates a new table with the client name, client number, and order number (SELECT ClientName, ClientNum, OrderNum). Both tables include the client number, which allows them to be joined. This ability is indicated in the WHERE clause, which states that the client number in the Client table is the same as (equal to) the client number in the Order table (WHERE Client. ClientNum=Order.ClientNum).
GRANT INSERT ON Client to Guthrie	This command is an example of a security command. It allows Bob Guthrie to insert new values or rows into the Client table.

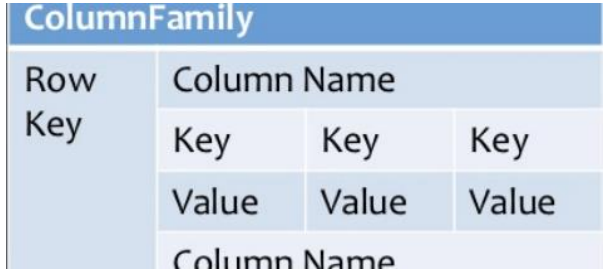
NoSQL database

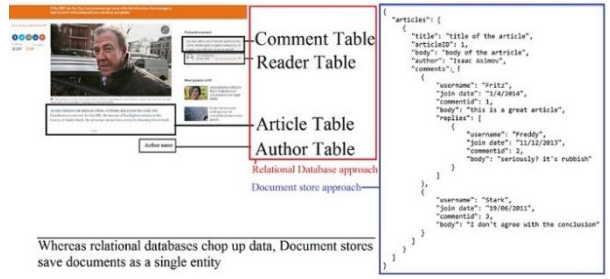
Provides a means to store and retrieve data that is modeled using some means other than the simple two-dimensional tabular relations used in relational databases:

- key-value stores
- graph database
- column-oriented database
- document database









Data Warehouses, Data Marts, and Data Lakes

Data warehouse: **a large database** that collects business information from many sources in the enterprise in support of management decision making

ETL process: Extract, Transform, and Load

- Data mart: a **subset of a data warehouse**
- Data lake: takes a "store everything" approach to big data, saving **all the data in its raw and unaltered form**

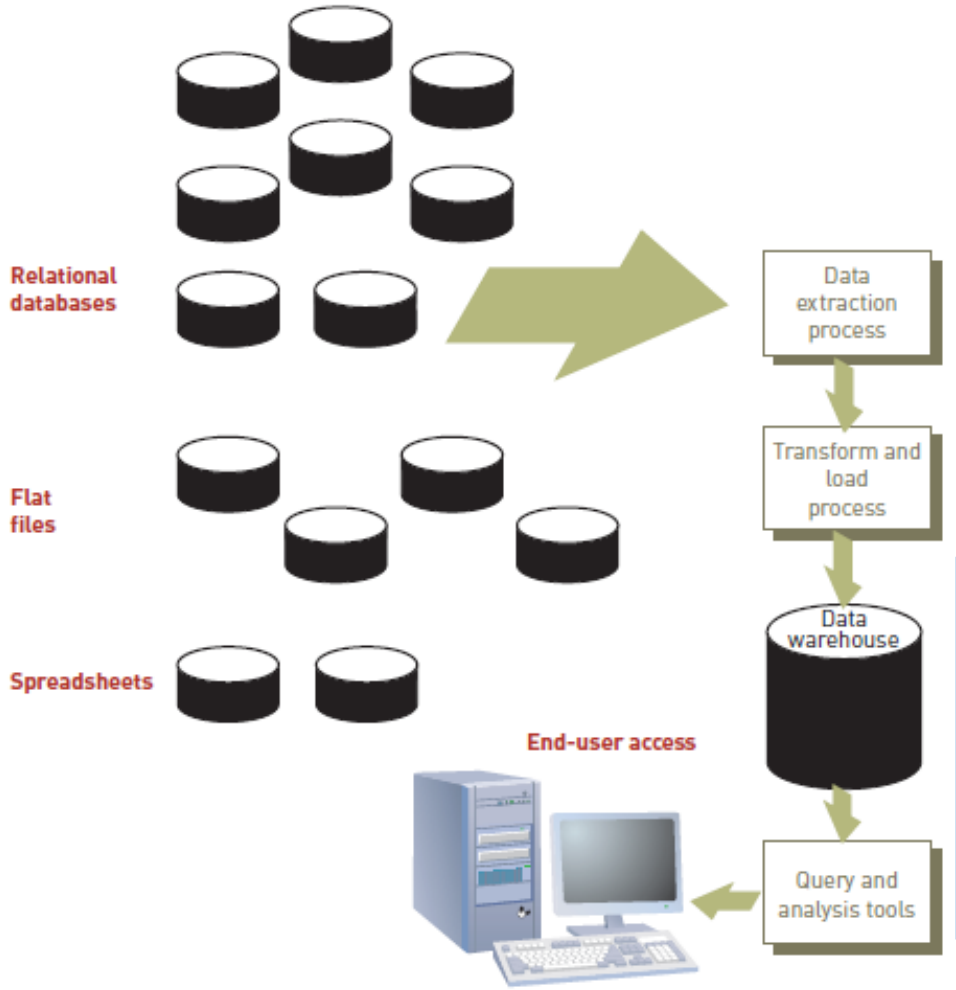


FIGURE 5.23
Elements of a data warehouse
A data warehouse can help managers and executives relate information in innovative ways to make better decisions.

Database as a Service (DaaS)

The database is **stored** on a **service provider**'s servers (database administration is handled by the service provider)

The database is accessed by the client over a network, typically the Internet

Example: Amazon Relational Database Service (Amazon RDS)

Group exercises

Form groups and let every group focus on resolving one case

- Vehicle Theft Database (p. 198)
- Cleansing Weather Data (p. 205)
- Walgreens Data Assimilation (p. 221)
- Telefonica Brasil (p. 228)

Homework

READ TEXTBOOK CHAPTER 6

TEST YOUR KNOWLEDGE FROM PREVIOUS LECTURES

ON A RANDOM CASE STUDY FROM THE TEXTBOOK