15th Edition

Understanding Computers

Today and Tomorrow

Comprehensive

Chapter 14:

Databases and Database Management Systems

Deborah Morley
Charles S. Parker

Learning Objectives

- Explain what a database is, including common database terminology, and list some of the advantages and disadvantages of using databases.
- Discuss some basic concepts and characteristics of data, such as data hierarchy, entity relationships, and data definition.
- Describe the importance of data integrity, security, and privacy and how they affect database design.
- Identify some basic database classifications and discuss their differences.

Learning Objectives

4. List the most common database models and discuss how they are used today.
5. Understand how a relational database is designed, created, used, and maintained.
6. Describe some ways databases are used on the Web.

Overview

- This chapter covers:
  - What a database is, the individuals who use them, and how databases evolved
  - Important database concepts and vocabulary
  - Database classifications and models
  - The relational database
  - How databases are used on the Web

What Is a Database?

- Database
  - A collection of related data stored in a manner that enables information to be retrieved as needed
- Database Management System (DBMS)
  - Used to create, maintain, and access databases
  - Database engine
    - The part of the program that actually stores and retrieves data
  - Microsoft Access, OpenOffice Base, Corel Paradox, Oracle Database, etc.

What Is a Database?

- A database typically consists of:
  - Tables
    - Collection of related records
  - Fields (columns)
    - Single category of data to be stored in a database (name, telephone number, etc.)
  - Records (rows)
    - Collection of related fields in a database (all the fields for one customer, for example)

What Is a Database?

- A Simple Relational Database Example

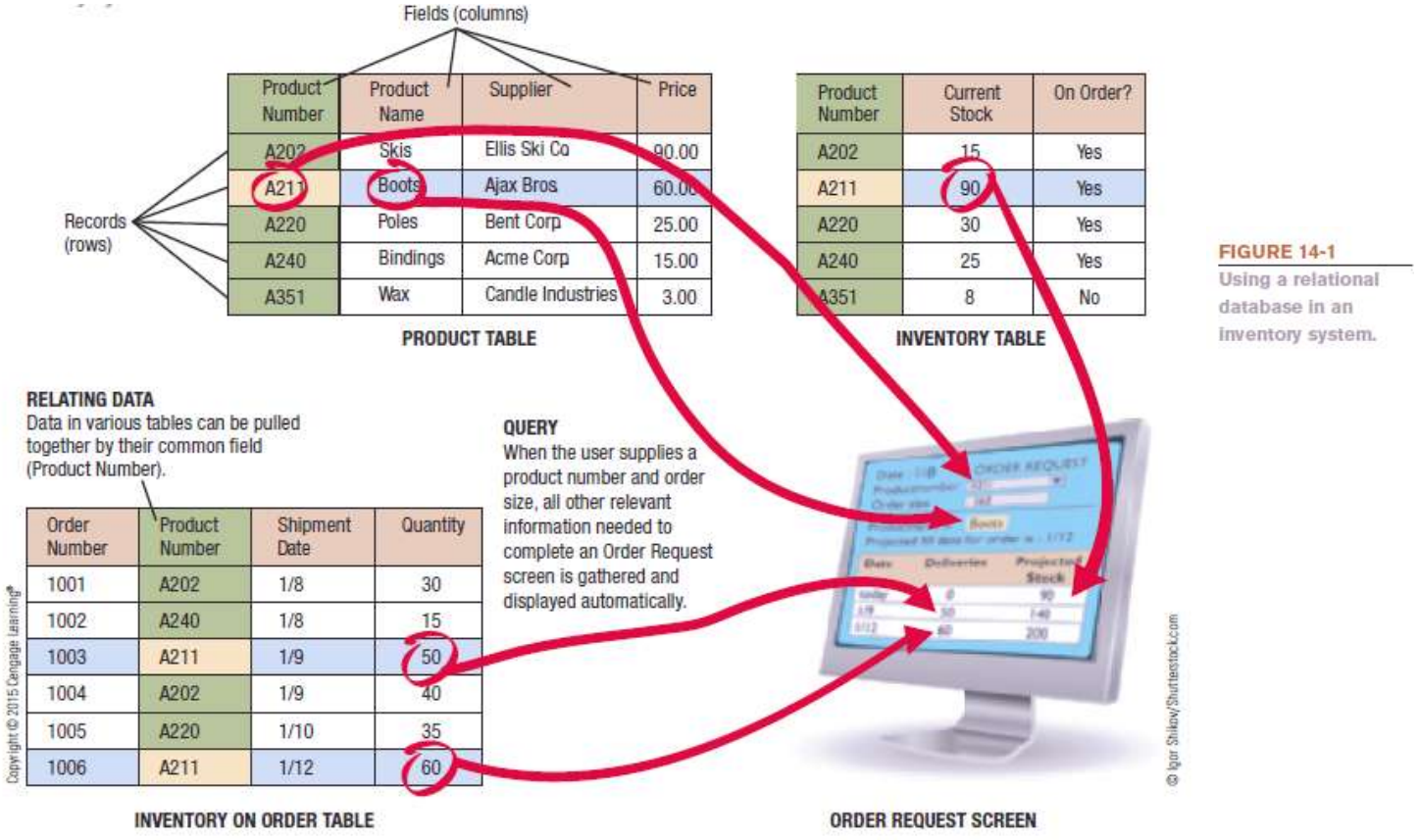


FIGURE 14-1
Using a relational database in an inventory system.

What is a Database?

- Primary Key
  - Field that uniquely identifies the records in a table
  - Field in a table that is used to relate that table to other tables

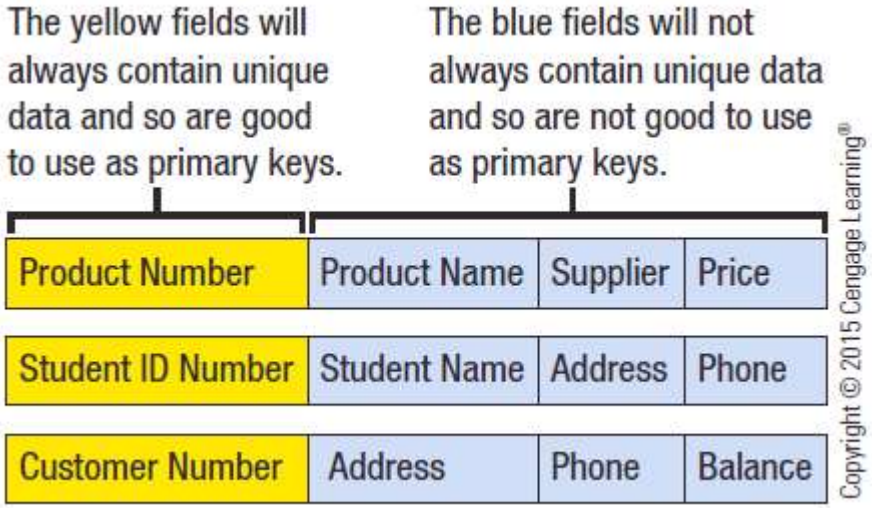


FIGURE 14-2
Primary key fields. A primary key field must contain unique data so it can be used to identify each record in the table.

What Is a Database?

- Individuals Involved with a Database Management System
  - Database Designers
    - Design the database
  - Database Developers
    - Create the database
  - Database Programmers
    - Write the programs needed to access the database or tie the database to other programs

What Is a Database?

- Database Administrators
  - Responsible for managing the databases within an organization
- Users
  - Individuals who enter data, update data, and retrieve information from the database

What Is a Database?

- The Evolution of Databases

MODEL	FLAT FILES	HIERARCHICAL	NETWORK	RELATIONAL	OBJECT-ORIENTED	MULTI-DIMENSIONAL
YEAR BEGAN	1940s	1960s	1960s	1970s	1980s	1990s
DATA ORGANIZATION	Flat files	Trees	Trees	Tables and relations	Objects	Data cubes, tables and relations, or a combination
DATA ACCESS	Low-level access	Low-level access with a standard navigational language	Low-level access with a standard navigational language	High-level, nonprocedural languages	High-level, nonprocedural, object-oriented languages	OLAP tools or programming languages
SKILL LEVEL REQUIRED TO ACCESS DATA	Programmer	Programmer	Programmer	User	User	User
ENTITY RELATIONSHIPS SUPPORTED	One-to-one	One-to-one, one-to-many	One-to-one, one-to-many, many-to-many	One-to-one, one-to-many, many-to-many	One-to-one, one-to-many, many-to-many	One-to-one, one-to-many, many-to-many
DATA AND PROGRAM INDEPENDENCE	No	No	No	Yes	Yes	Yes

FIGURE 14-3
The evolution of databases. Databases have evolved over the years, becoming more flexible, more capable, and easier to use.

What Is a Database?

- Advantages and Disadvantages of the DBMS Approach
  - Advantages
    - Low level of redundancy
      - Faster response time
      - Lower storage requirements
      - Easier to secure
      - Increased data accuracy
  - Disadvantages
    - Increased vulnerability (backup is essential)

Inside the Industry Box

File Management Systems

- Tables are not related so more time-consuming and more redundancy

PRODUCT NUMBER	PRODUCT NAME	SUPPLIER	PRICE
A202	Skis	Ellis Ski Co.	90.00
A211	Boots	Ajax Bros.	60.00
A220	Poles	Bent Corp.	25.00
A240	Bindings	Acme Corp.	15.00
A351	Wax	Candle Industries	3.00

PRODUCT TABLE

PRODUCT NUMBER	PRODUCT NAME	SUPPLIER	PRICE	CURRENT STOCK	ON ORDER?
A202	Skis	Ellis Ski Co.	90.00	15	Yes
A211	Boots	Ajax Bros.	60.00	90	Yes
A220	Poles	Bent Corp.	25.00	30	Yes
A240	Bindings	Acme Corp.	15.00	25	Yes
A351	Wax	Candle Industries	3.00	8	No

INVENTORY TABLE

ORDER NUMBER	PRODUCT NUMBER	SHIPMENT DATE	PRODUCT NAME	SUPPLIER	PRICE	QUANTITY
1001	A202	1/8	Skis	Ellis Ski Co.	90.00	30
1002	A240	1/8	Bindings	Acme Corp.	15.00	15
1003	A211	1/9	Boots	Ajax Bros.	60.00	50
1004	A202	1/9	Skis	Ellis Ski Co.	90.00	40
1005	A220	1/10	Poles	Bent Corp.	25.00	35
1006	A211	1/12	Boots	Ajax Bros.	60.00	60

INVENTORY ON ORDER TABLE

REDUNDANT FIELDS

Instead of just having one field duplicated like in the databases created using a DBMS (green shaded columns), flat files created using file management systems require many more fields to be duplicated (green and blue shaded columns). Notice that the blue shaded columns shown here appear only in the Product table in Figure 14-1; when using a DBMS, these fields are not included in the Inventory and Inventory on Order tables.

Because file management systems cannot retrieve data from more than one table at a time, there is a much higher level of redundancy.

Data Concepts and Characteristics

- Data Hierarchy
  - Fields/columns
    - Hold single pieces of data
  - Records/rows
    - Groups of related fields
  - Tables
    - Collection of related records
  - Database
    - Contains a group of related tables

Data Concepts and Characteristics

- Entities and Entity Relationships
  - Entity
    - A person, object, or event of importance to the organization
    - Entities that the organization wants to store data about typically becomes a database table
  - Attributes
    - Characteristics of an entity
    - Typically become fields in the entity's database table
  - Relationship
    - An association between two or more entities

Data Concepts and Characteristics

- One to One (1:1) Entity Relationships
  - One entity is related to only one other entity of a particular type
  - Not a common type of relationship
- One to Many (O:M) Entity Relationship
  - Most common type of relationship
  - One entity can be related to more than one other entity
    - A supplier can supply more than one product to a company
- Many to Many (M:M) Entity Relationships
  - One entity can be related to more than one other entity, and those entities can be related to multiple entities of the same type as the original entity

Data Concepts and Characteristics

- Data Definition
  - The process of describing the properties of data to be included in a database table
  - During data definition, each field is assigned:
    - Name (must be unique within the table)
    - Data type (such as Text, Number, Currency, Date/Time)
    - Description (optional description of the field)
    - Properties (field size, format of the field, allowable range, if field is required, etc.)
  - Finished specifications for a table become the table structure

Data Concepts and Characteristics

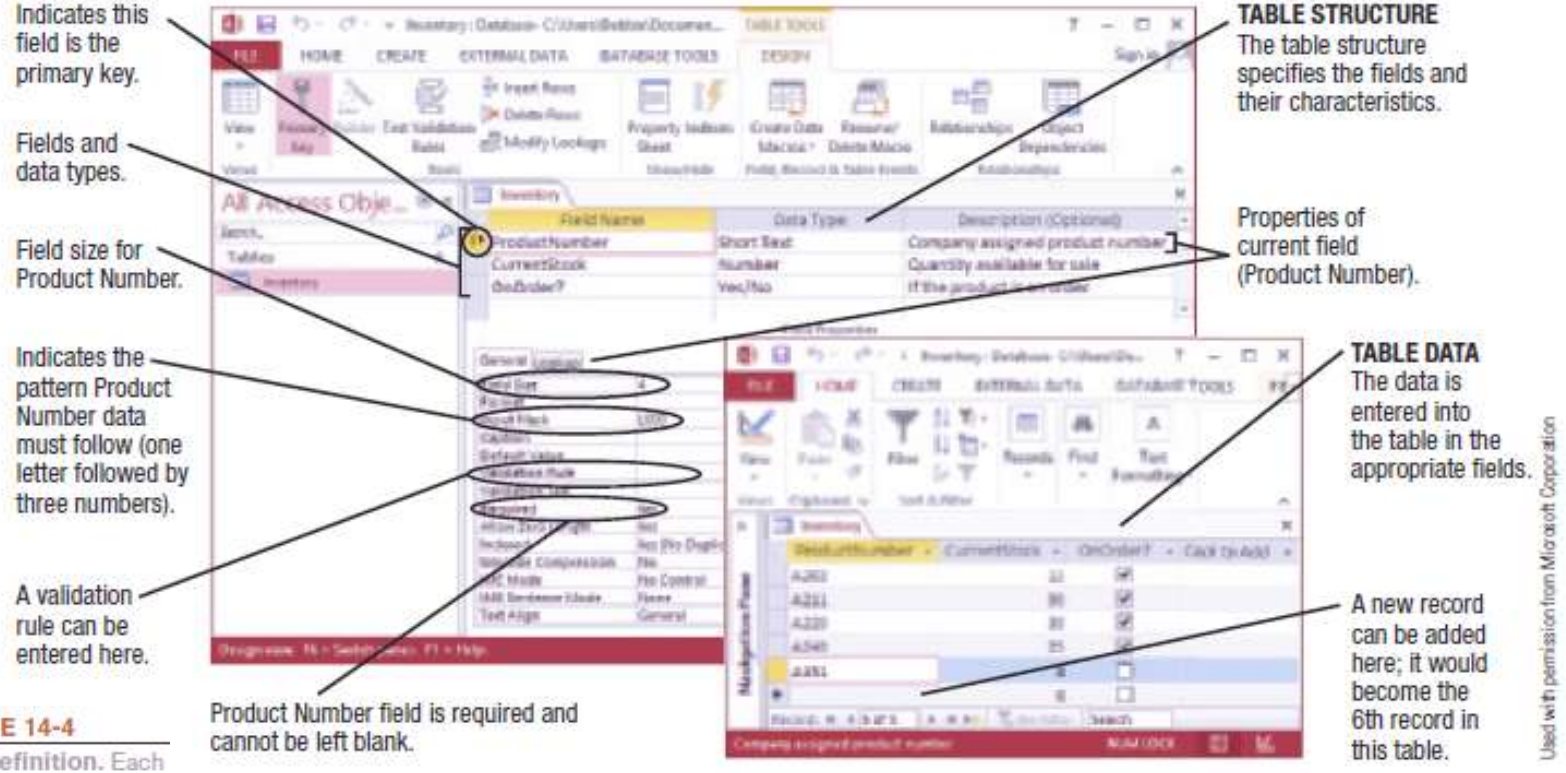


FIGURE 14-4

Data definition. Each field in a database has a defined data type and properties that can be assigned to that field.

Data Concepts and Characteristics

- The Data Dictionary
  - Contains all data definitions in a database, including:
    - Table structures
    - Security information (passwords, etc.)
    - Relationships between the tables in the database
    - Basic information about each table, such as the current number of records
  - Does not contain any of the data in the tables
  - Does contain metadata, which is information about the database tables
  - Ensures that data being entered into the database does not violate any specified criteria

Data Integrity, Security, and Privacy

- Data Integrity
  - Accuracy of Data
    - Quality of data entered determines the quality of generated information
  - Data Validation
    - Process of ensuring that data entered into the database is valid
    - Record validation rules
      - Checks all fields before changes to a record are saved
    - Can be enforced on a per transaction basis so the entire transaction will fail if one part is invalid

Data Integrity, Security, and Privacy

- Database Locking
  - Prevents two individuals from changing the same data at the same time

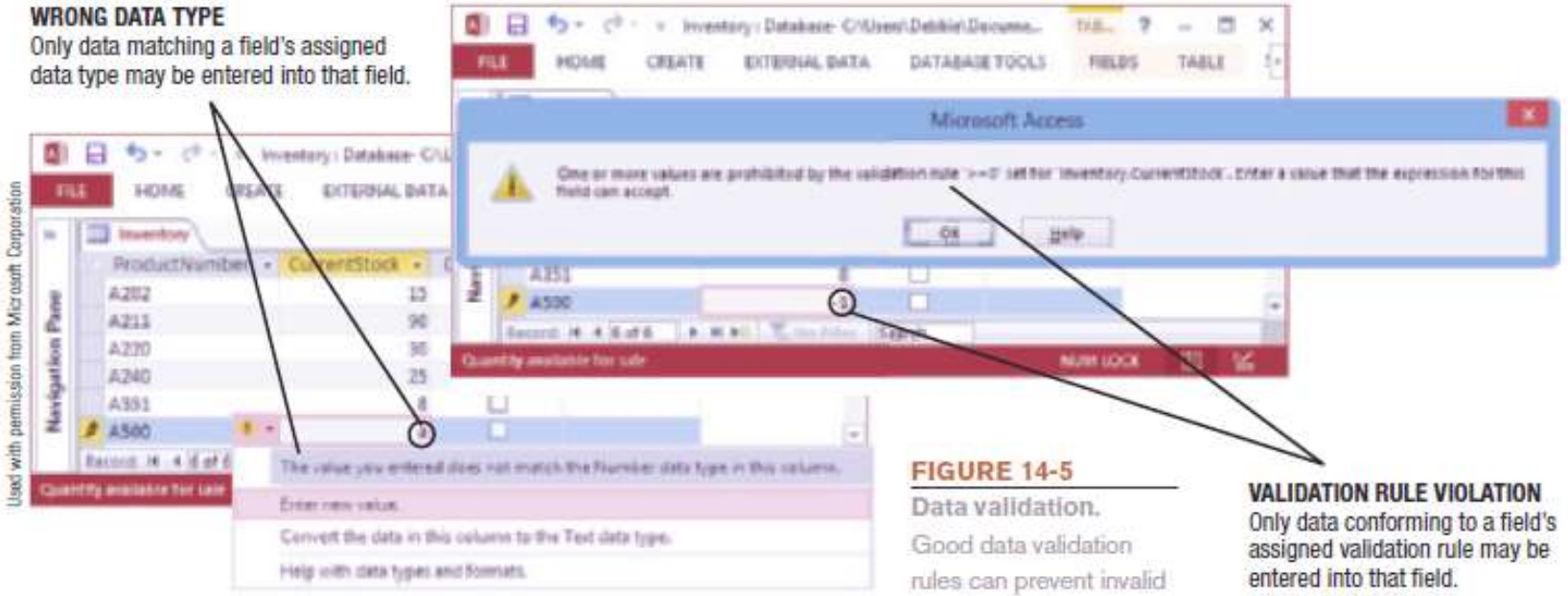


FIGURE 14-5
Data validation.
Good data validation rules can prevent invalid data from being entered into a database table.

Data Integrity, Security, and Privacy

- Data Security
  - Protects data against destruction and misuse
  - Protects against unauthorized access to and unauthorized use of a database
  - Database activity monitoring programs can be used to detect possible intrusions and risks
  - Prevents data loss
  - Should include strict backup and disaster-recovery procedures (disaster-recovery plan)
  - Should be used with both in-house and cloud databases

Data Integrity, Security, and Privacy

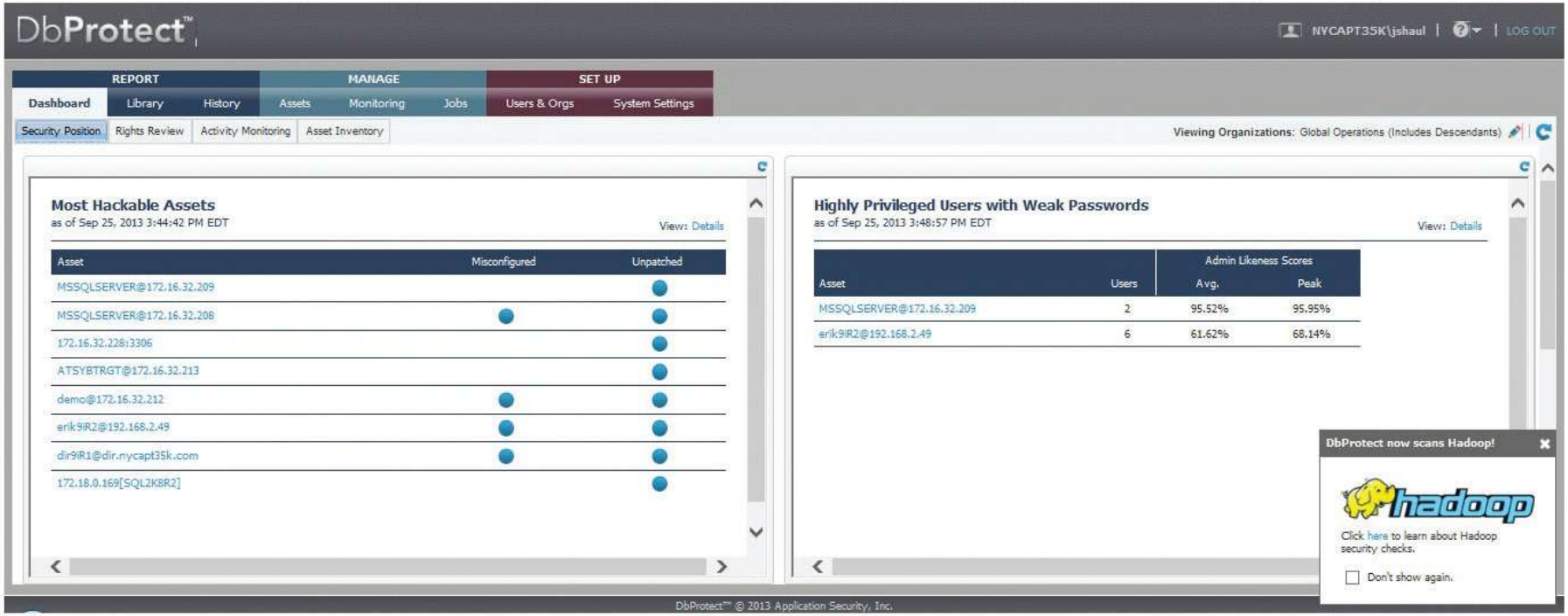


FIGURE 14-6
Database security tools. This program secures databases and displays alerts for vulnerabilities and attacks.

Data Integrity, Security, and Privacy

- Data Privacy
  - Growing concern because of the vast amounts of personal data stored in databases today
  - Many states require businesses to notify customers when their personal data has been compromised
  - Data breaches can be costly
    - One estimate is $200 per breached record

Data Organization

- Data Organization
  - Arranging data for efficient retrieval
  - Indexed organization
    - Uses an index to keep track of where data is stored in a database
- Direct Organization
  - Uses hashing algorithms to specify the exact storage location
  - Algorithms should be designed to limit collisions
  - Some systems use a combination of both indexed and direct organization

Data Organization

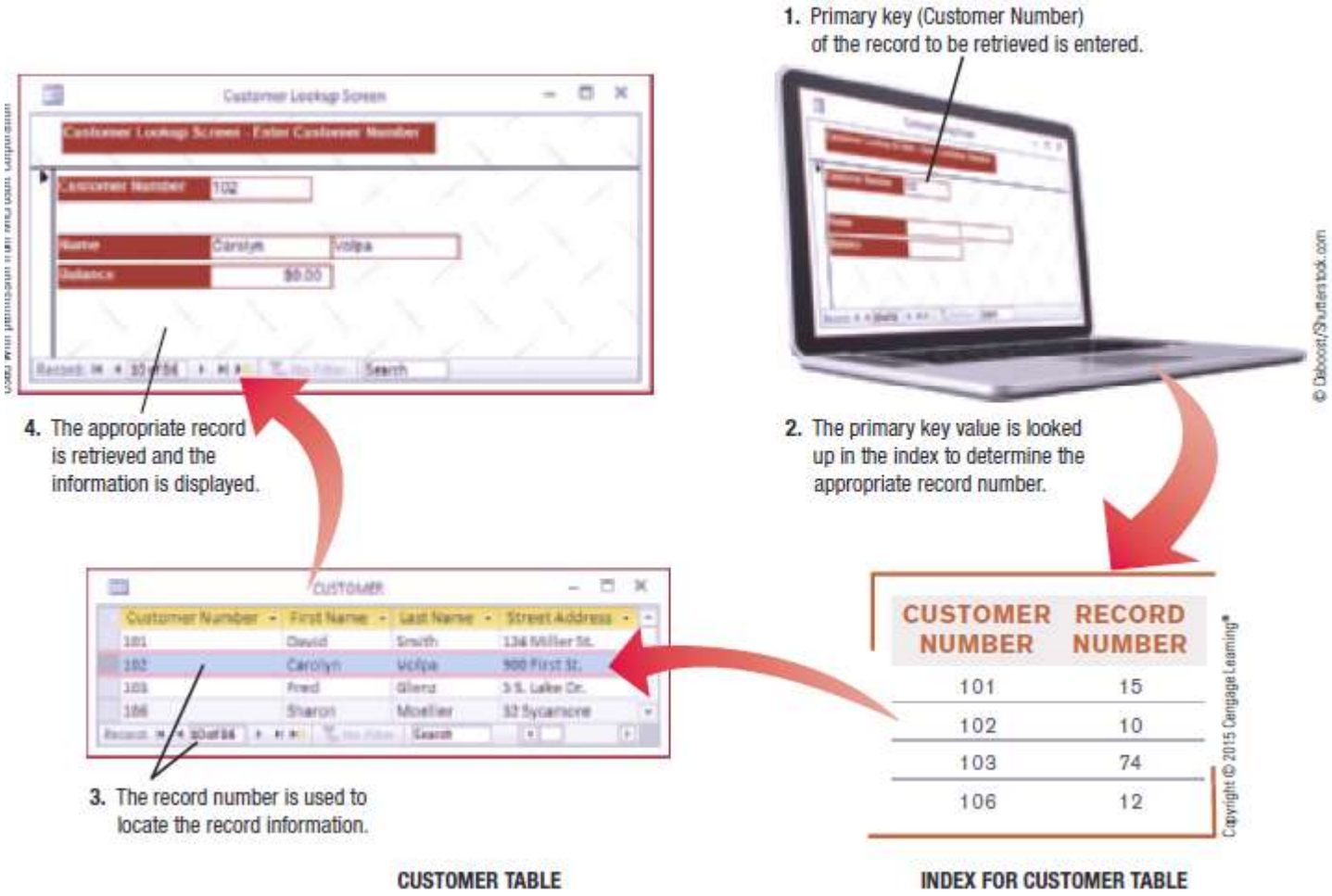


FIGURE 14-7
Indexed organization is often used for real-time transaction processing.

Data Organization

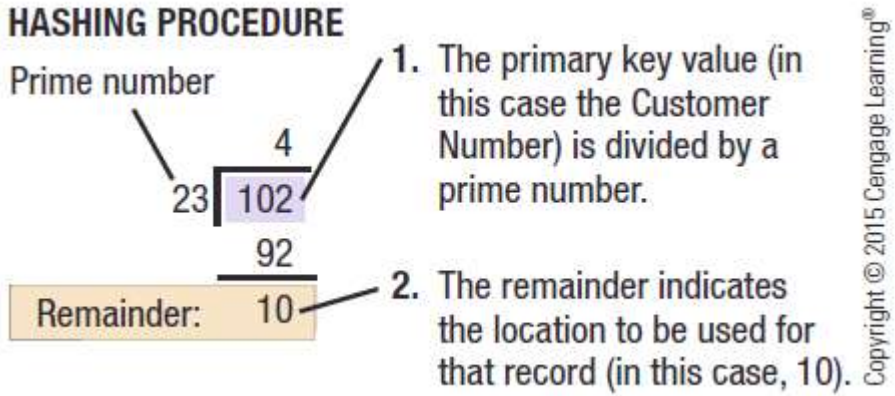
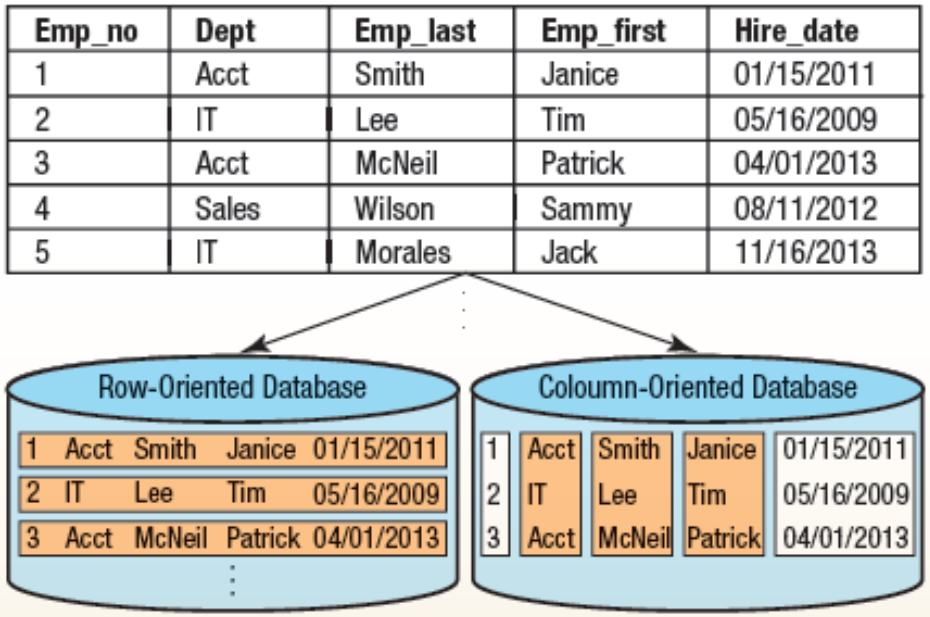


FIGURE 14-8
Direct organization is frequently used for faster real-time processing.

How It Works Box

Column Databases

- Stores data by columns instead of rows
- Improves performance by minimizing the time needed to read the disk
- Used with data warehouses and other big data applications



Example of retrieving the names of all IT employees using a row vs. a column database.

Quick Quiz

1. A column in a database in which customer names are stored would be referred to as a _____.
   a. field
   b. record
   c. table
2. True or False: Data validation procedures are used to ensure that data entered into a database matches the specified type, format, and allowable value.
3. The _____ contains metadata about the database tables in a database.

Answers:

1) a; 2) True; 3) data dictionary

Database Classifications

- Single-User vs. Multiuser Database Systems
  - Single-User Database System
    - Located on a single computer
    - Designed to be accessed by one user
    - Widely used for personal applications and very small businesses
  - Multiuser Database System
    - Designed to be accessed by multiple users (most business databases today)

Database Classifications

- Client-Server and N-Tier Database Systems
  - Client-Server Database Systems
    - Has both clients (front end) and at least one database server (back end)

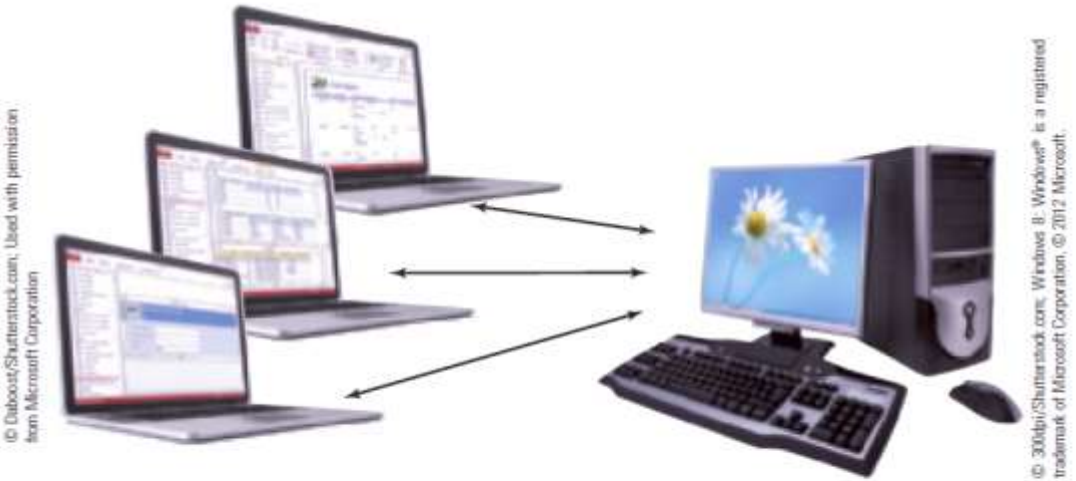
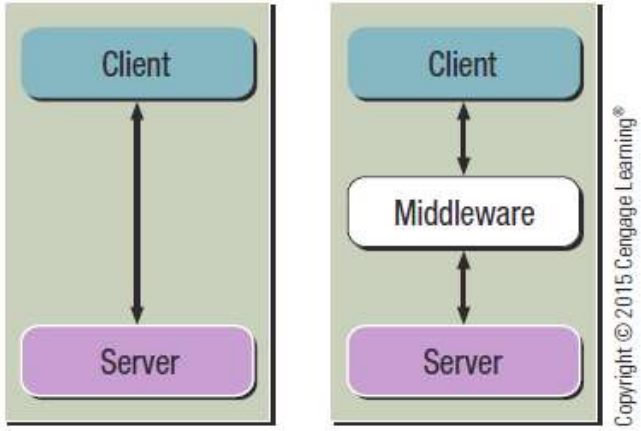


FIGURE 14-9
Client-server database systems.

Database Classifications

- N-Tier Database System
  - Has more than two tiers
  - Additional tiers typically contain software referred to as middleware
  - Allows program code to be separate from the database
  - Code can be divided into any number of logical components



2-TIER MODEL
Has just a client and a server.

N-TIER MODEL
Includes middleware, which contains additional programs used to connect the client and server.

FIGURE 14-10
A 2-tier vs. an n-tier database model.

Database Classifications

- Centralized vs. Distributed Database Systems
  - Centralized Database System
    - Database is located on a single computer, such as a server or mainframe
  - Distributed Database System
    - Data is physically divided among several computers connected by a network, but the database logically looks like it is a single database

Database Classifications

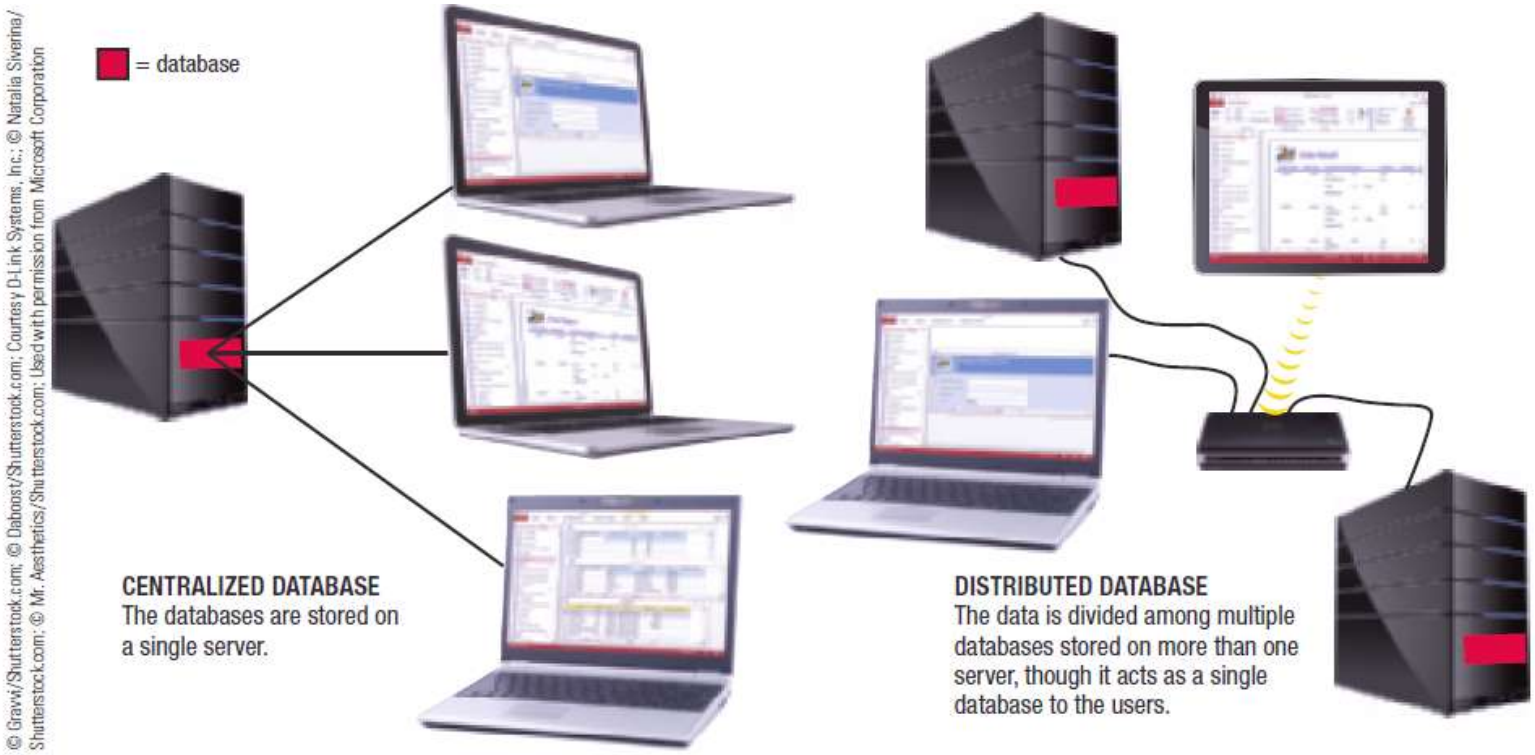


FIGURE 14-11
Centralized vs. distributed databases.

Database Classifications

- Disk-Based vs. In-Memory Database Systems
  - Disk-Based Systems
    - Data is stored on hard drives
  - In-Memory Databases (IMDBs)
    - Data is stored in main memory
    - Dramatically faster than disk-based databases
    - Good backup procedures are essential
    - Used both in high-end systems where performance is crucial and in small-footprint, embedded applications

Quick Quiz

1. Which type of database system is beginning to be used in high-end systems where performance is crucial?
   a. In-memory databases
   b. Disk-based databases
   c. Single-user databases
2. True or False: With the n-tier database model, there is at least one middle piece of software between the client and the server.
3. With a(n) _____ database system, the databases used by the system are all located on a single computer.

Answers:

1) a; 2) True; 3) centralized

Database Models

- The Hierarchical and Network Database Models
  - Hierarchical Databases
    - Organizes data in a tree structure
    - Typically a one-to-many relationship between data entities
  - Network Databases
    - Allow both one-to-many and many-to-many relationships between data elements
  - Most databases today are neither hierarchical or network models

The Relational Database Model (RDBMS)

- The Relational Database Model (RDBMS)
  - Data is organized in tables related by common fields
  - Most widely used database model today
  - Designing a Relational Database
    - Identify the purpose of the database
    - Determine the tables and fields
    - Assign the fields to a table and reorganize as needed to minimize redundancy (normalization – most databases stop at 3NF)
    - Finalize the structure (primary keys, field properties, etc.)

The Relational Database Model (RDBMS)

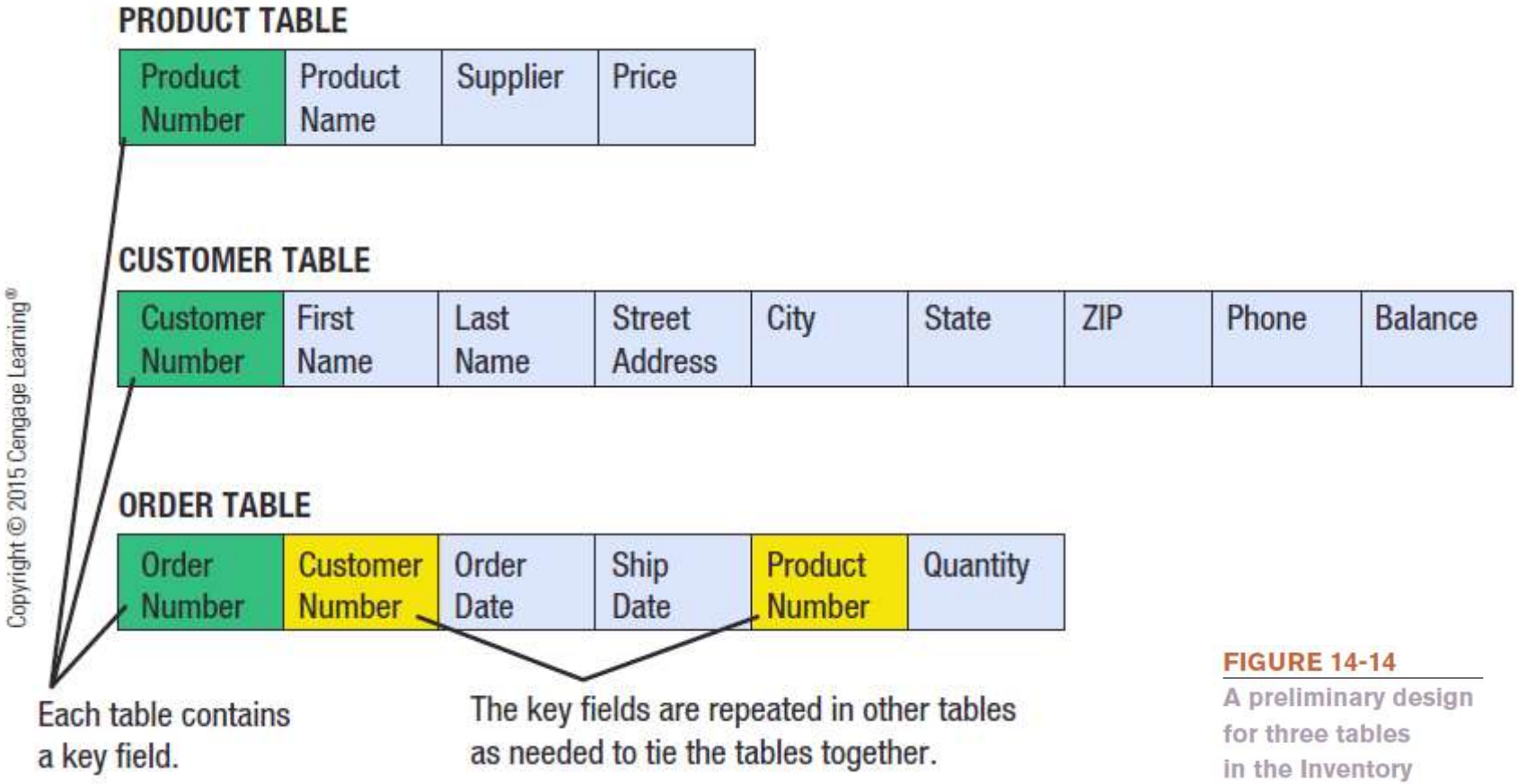


FIGURE 14-14
A preliminary design for three tables in the Inventory database.

The Relational Database Model (RDBMS)

- Creating a Relational Database
  - Creating the Tables
    - Each table is created using the table structure developed during the database design process
      - In Access, can use Design view or Datasheet view
  - Entering and Editing Data
    - Existing data can be migrated to the new database
    - New data can be added via a form or Datasheet view
      - In either case, the same data is being manipulated

Database Models

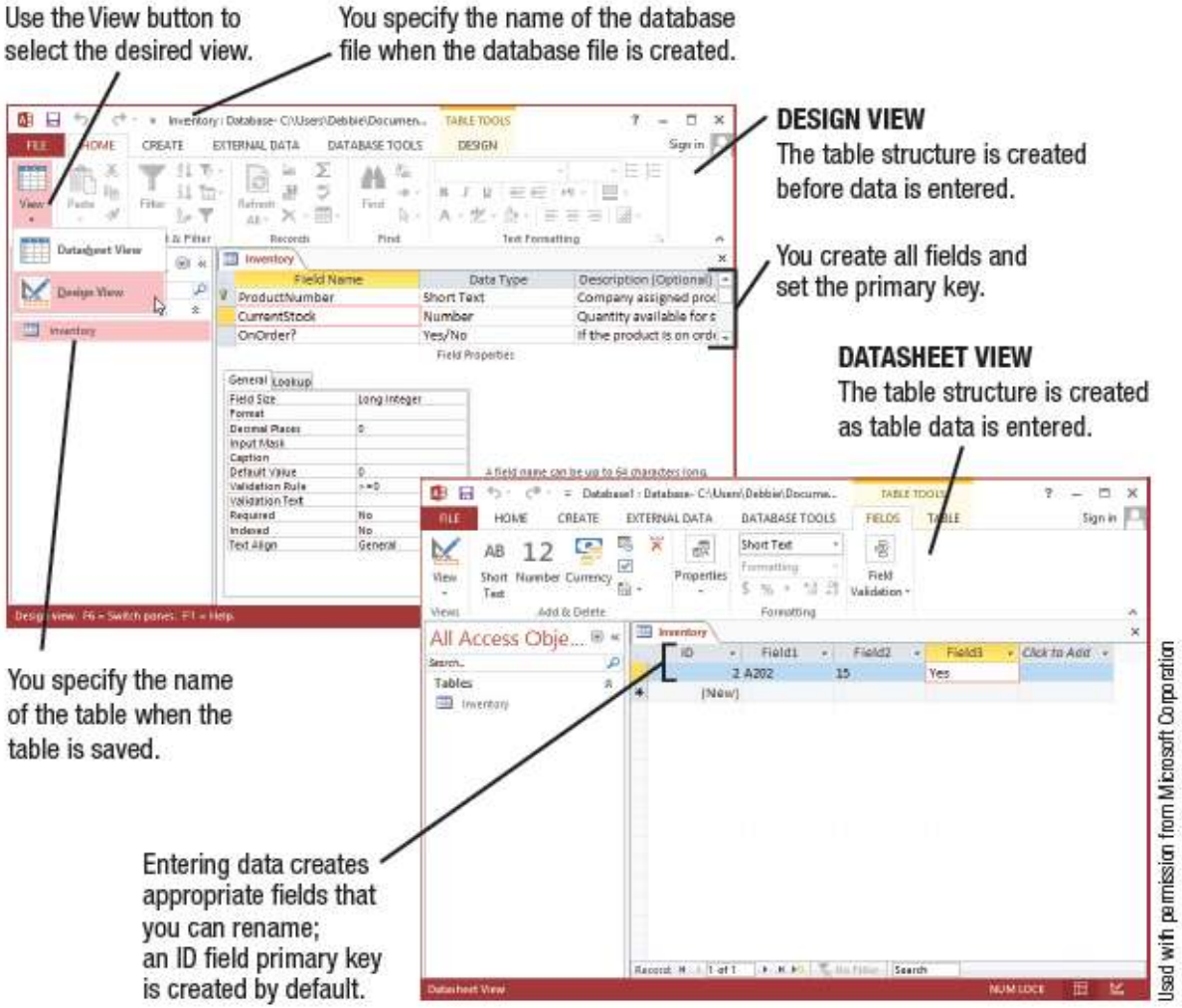


FIGURE 14-15
Tables can be created using Design view or Datasheet view.

Database Models

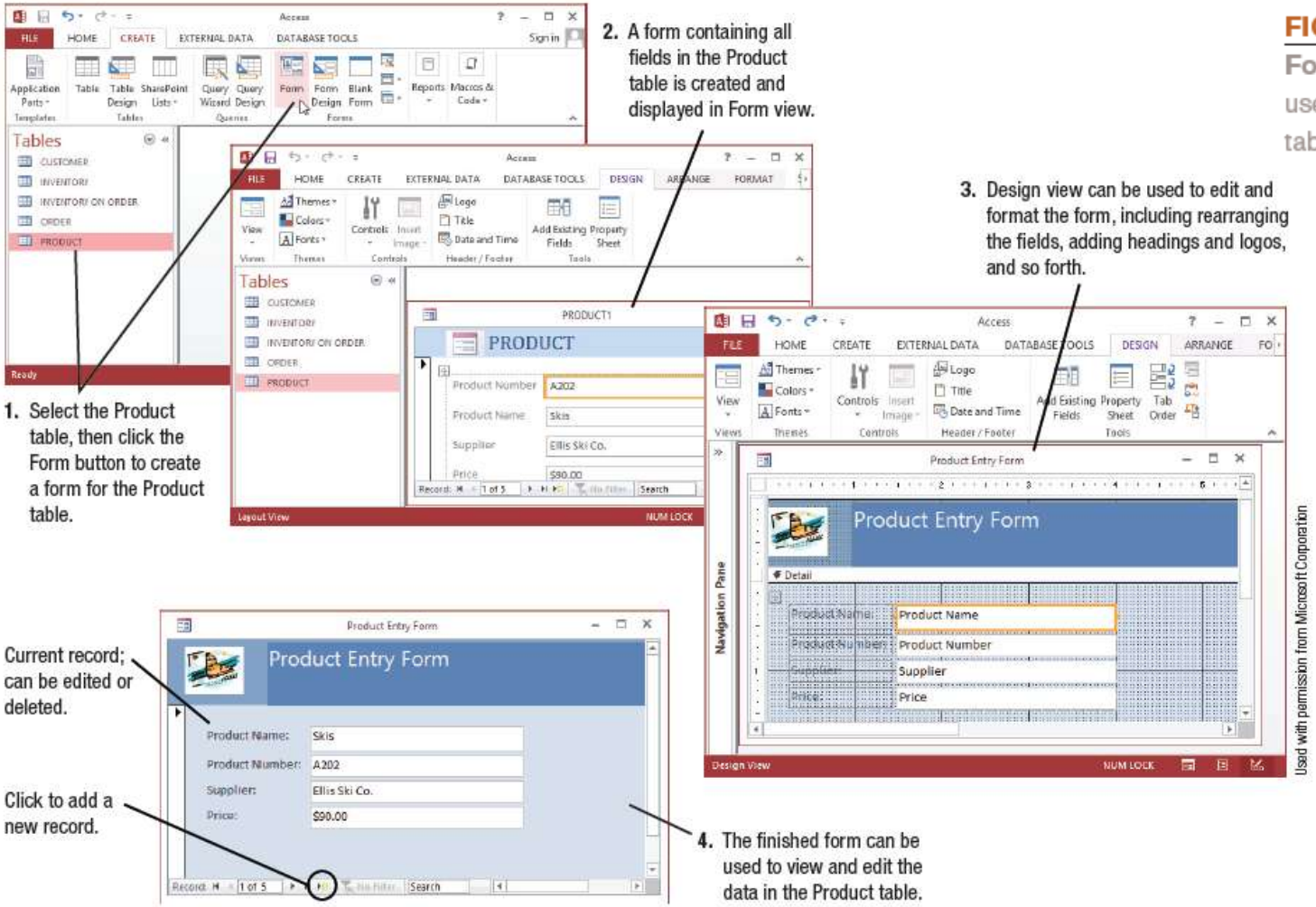


FIGURE 14-16

Forms. Forms can be used to view and edit table data.

The Relational Database Model (RDBMS)

- Relating Tables
  - Once all tables have been created, they can be related to one another using their primary keys

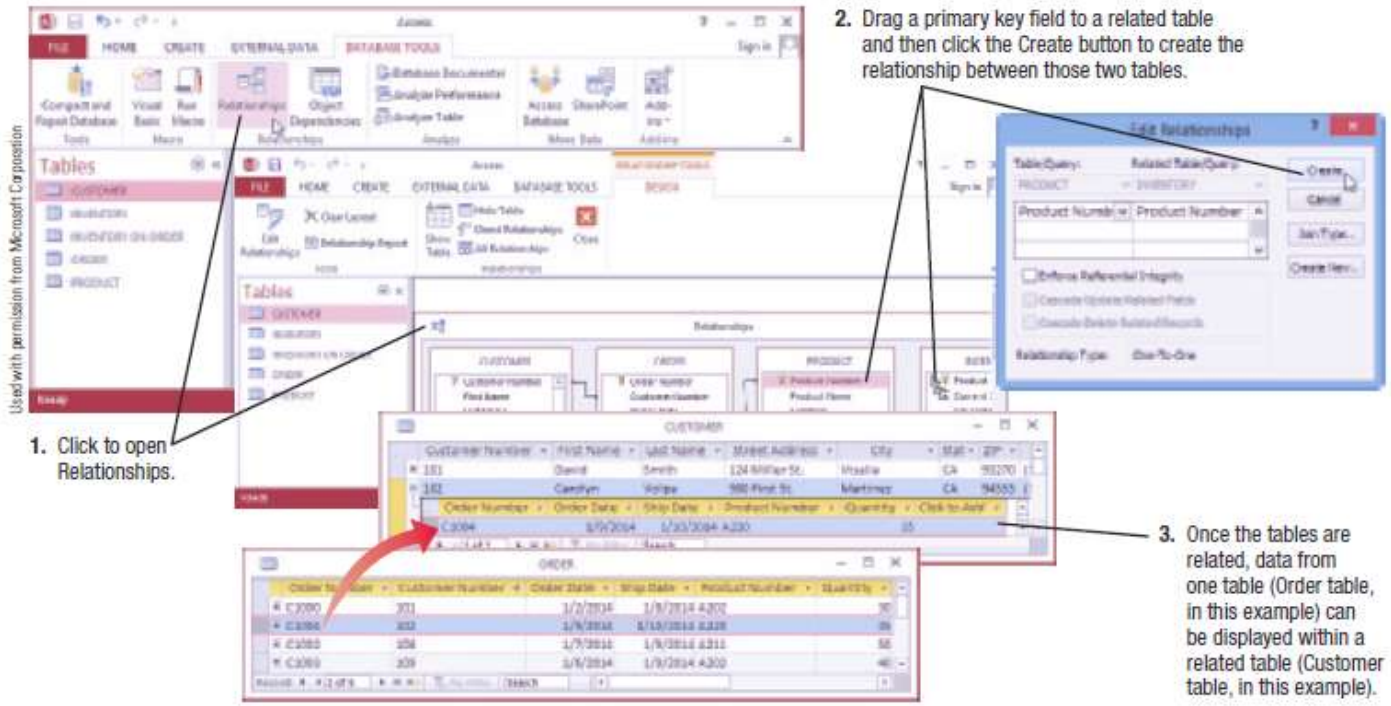


FIGURE 14-17
Relating tables.

The Relational Database Model (RDBMS)

- Retrieving Information from a Relational Database
  - Query
    - A request to see information from a database that matches specific criteria
    - Every DBMS provides tools users can use to query the database for information
    - Can also write in structured query language (SQL)
    - Must be designed to extract information as efficiently as possible
    - Poorly written queries can impact the overall performance of the system

The Relational Database Model (RDBMS)

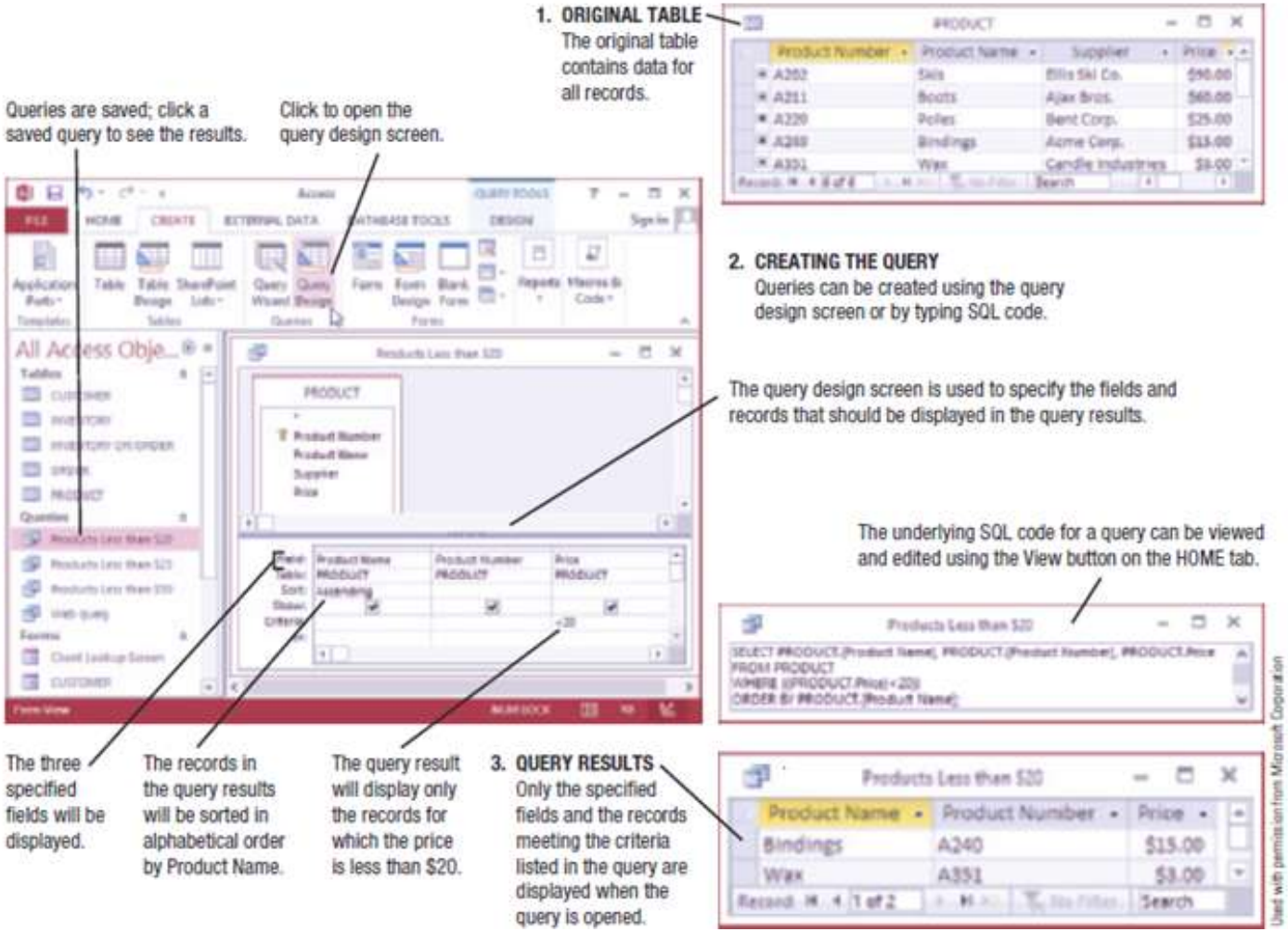


FIGURE 14-18
Querying a database. This example pulls information from the Product table in the Inventory database.

The Relational Database Model (RDBMS)

- Reports
  - Formatted way of looking at a database table or the results of a query
  - Can pull data from more than one table
  - Many programs have wizards or other tools to make it easy to create a report
  - Can be modified and customized using the Design view
  - Reports in Microsoft Access are saved as objects in the database file

The Relational Database Model (RDBMS)

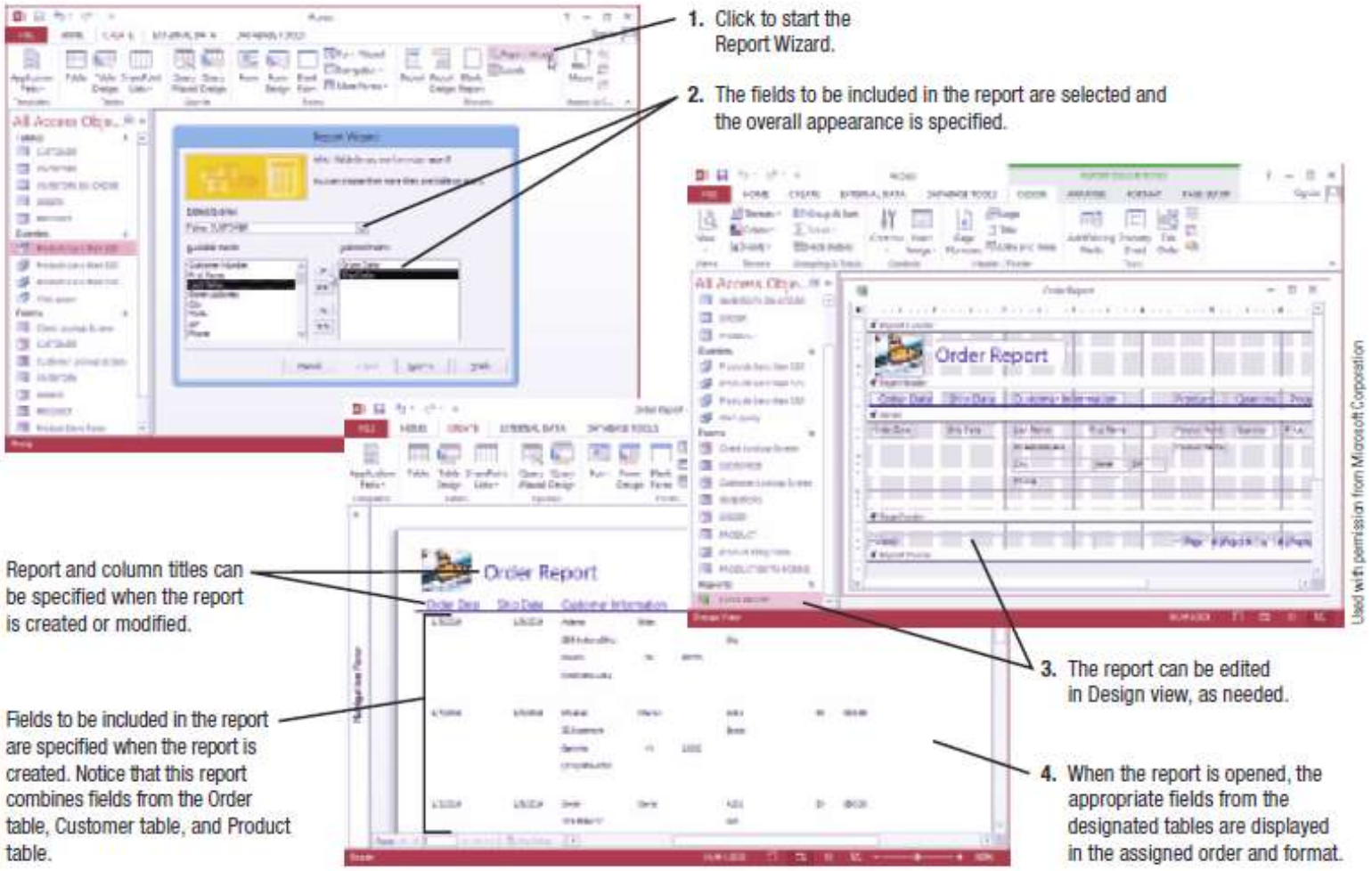


FIGURE 14-19
Reports. Display table information with a more formal, businesslike appearance.

The Relational Database Model (RDBMS)

- Maintaining a Relational Database
  - Table structures can be modified when needed
  - Other possible modifications:
    - Adding new indexes to speed up queries
    - Deleting obsolete data
    - Upgrading database software, installing patches
    - Repairing/restoring data that has become corrupt
    - Continuing to evaluate and improve security

The Object-Oriented Database Model

- The Object-Oriented Database Model
  - Object-Oriented Database Management System (OODBMS)
    - Database system in which multiple types of data are stored as objects along with their related code
    - Can contain virtually any type of data (video clip, text with music, etc.) along with the methods to be used with that data
    - Objects can be retrieved using queries (object query language or OQL)
    - Objects can be reused in other applications to create new applications quickly

Trend Box

Law Enforcement Databases

- Have been used for years but new database are now emerging that hold non-traditional data like photos and biometric data
  - Next Generation Identification (NGI)
    - Includes AFIT to store and match fingerprints
    - Includes support for photos and face-matching
    - Future improvements include support for voice, iris, DNA, palm prints, etc.

Hybrid Database Models

- Hybrid Database Models
  - A combination of two or more database types or models
    - Hybrid XML/Relational Database
      - Can store and retrieve both XML data and relational data

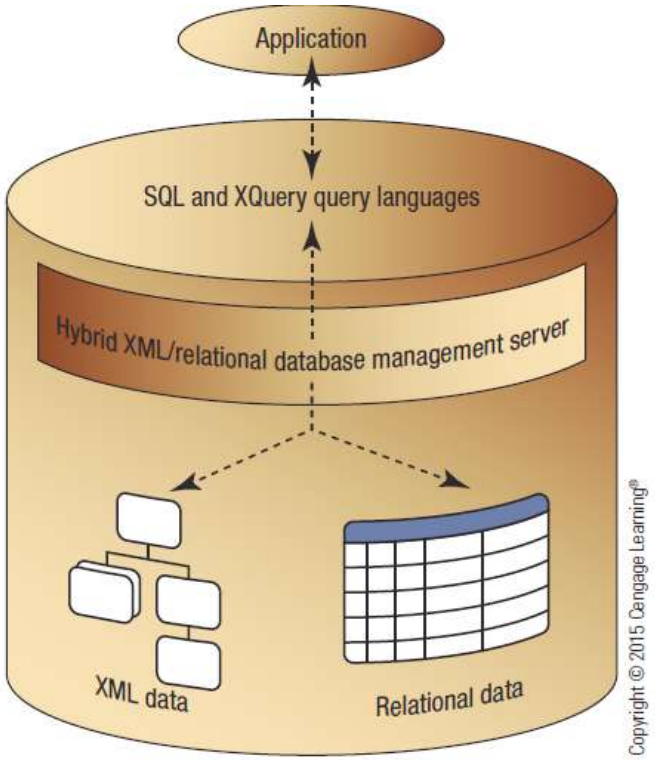


FIGURE 14-21
Hybrid XML/relational databases.

Multidimensional Databases (MDDB)

- Multidimensional Databases (MDDB)
  - Designed to be used with data warehousing
  - Often used in conjunction with Online Analytical Processing (OLAP)
    - MOLAP (Multidimensional OLAP)
      - Data is stored in single structures called data cubes
    - ROLAP (Relational OLAP)
      - Data is stored in an existing relational database using tables to store the summary information
    - HOLAP (Hybrid OLAP)
      - Combination of MOLAP and ROLAP technologies

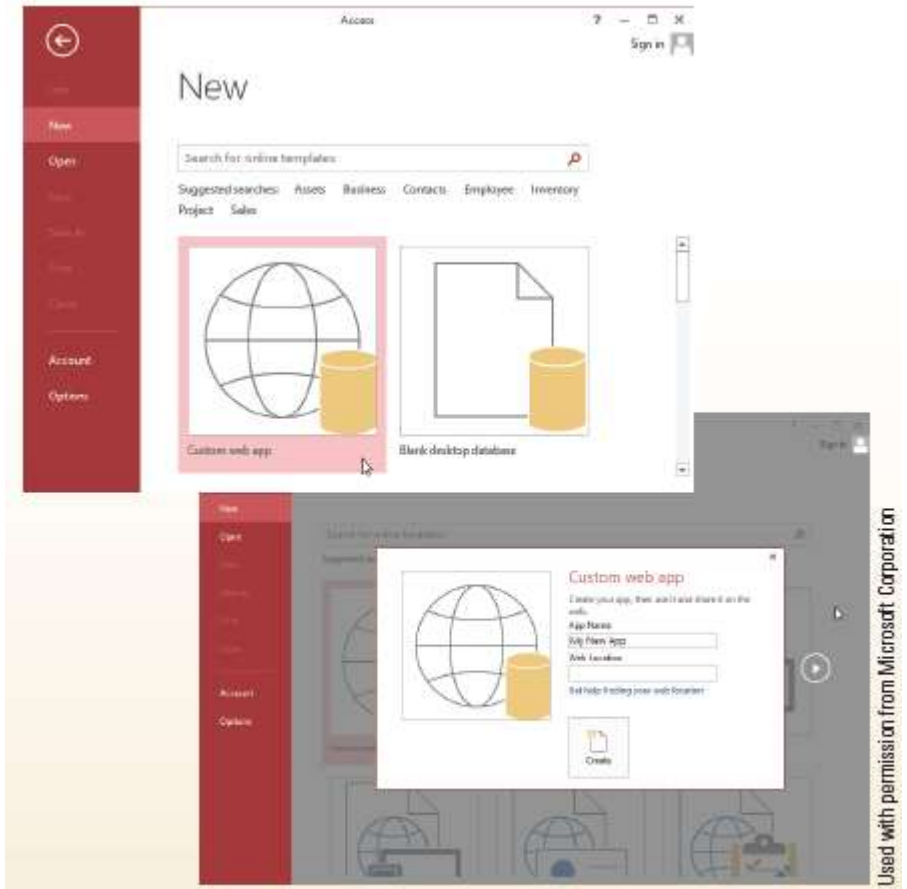
Cloud Databases

- Typically hosted on a cloud database provider's servers that is accessible to users via the Web
- Examples of Cloud Databases in Use
  - Information retrieval
    - Data to be accessed and displayed on a Web page is often stored in a database, i.e, Search sites
  - Support and facilitate e-commerce
    - Display product information, pricing, customer information, shopping cart content, etc.
  - Cloud databases allow Web pages to be dynamic Web pages

Technology and You Box

Cloud Databases

- Use growing rapidly
- Typically built using a cloud provider (Windows Azure, Amazon SimpleDB, or Google Cloud SQL)
- Requires less in-house hardware and maintenance
- Individuals can create via Microsoft Access web apps



With Microsoft Access, you can create custom cloud databases.

Cloud Databases

- How Cloud Databases Work
  - Visitor makes request by
    - Filling out a Web page form
    - Selecting an option from a menu displayed on a Web page form
  - Web server converts the request into a database query, passes it onto the database server, and then sends the results back to the visitor

Cloud Databases

- Middleware
  - Software used to connect two otherwise separate applications, such as a Web server and a database management system
  - Commonly written as scripts
  - Common languages include
    - JavaScript
    - VBScript
    - CGI Scripts
    - Active Server Pages (ASPs)
    - PHP Scripts

Cloud Databases

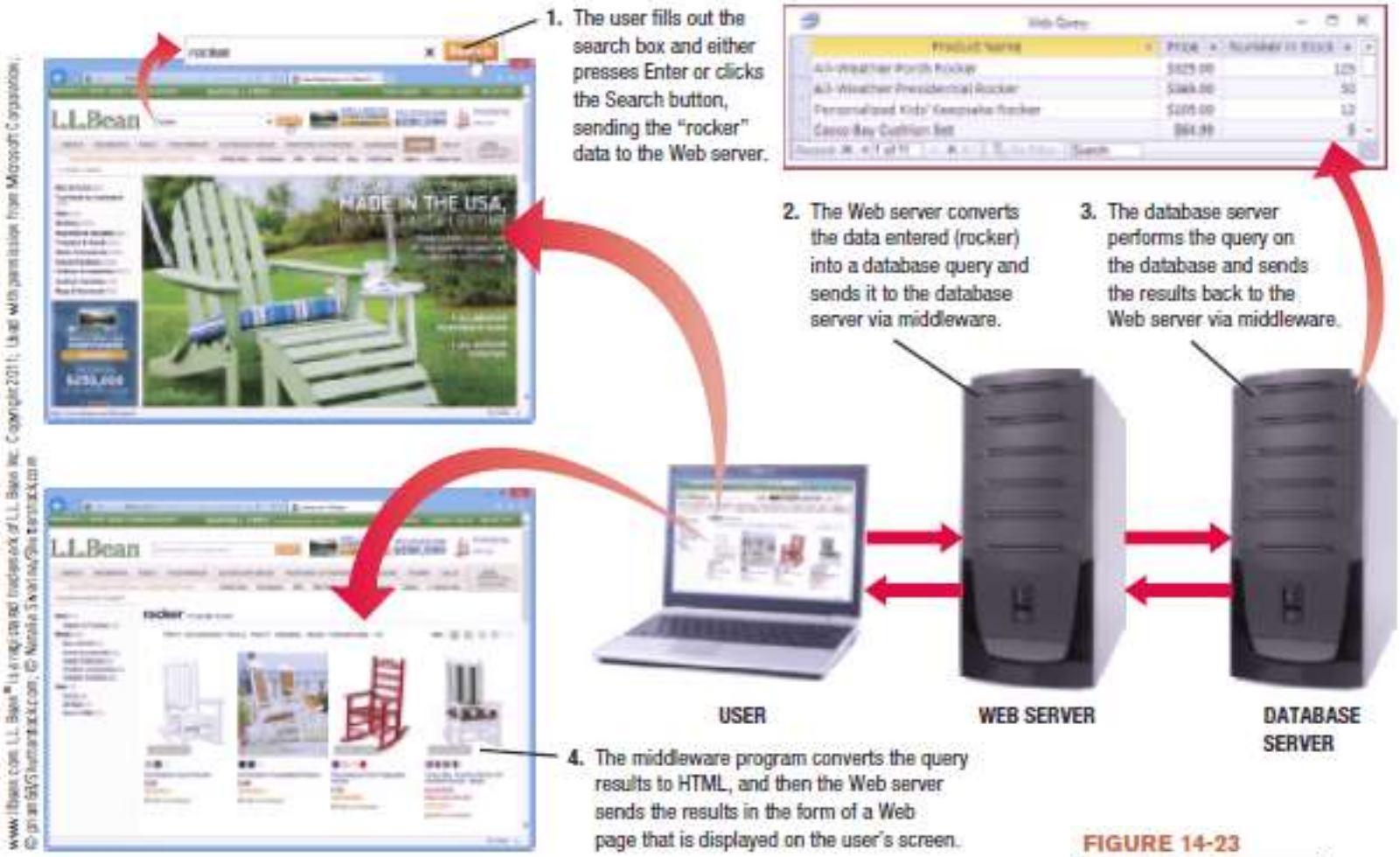


FIGURE 14-23
A cloud database in action.

Quick Quiz

1. Which of the following is the most widely used type of database today?
   a. Network
   b. Relational
   c. Object-oriented
2. True or False: Databases are often used in conjunction with dynamic Web pages.
3. A(n) _____ is used to extract specific information from a database by specifying particular conditions about the data to be retrieved.

Answers:

1) b; 2) True; 3) query

Summary

- What Is a Database?
- Data Concepts and Characteristics
- Database Classifications
- Database Models
- Cloud Databases