Chapter 6

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

Learning Objectives

- Describe how the problems of managing data resources in a traditional file environment are solved by a database management system
- Describe the capabilities and value of a database management system
- Apply important database design principles
- Evaluate tools and technologies for accessing information from databases to improve business performance and decision making
- Assess the role of information policy, data administration, and data quality assurance in the management of a firm's data resources

© Prentice Hall 2011

RR Donnelley Tries to Master Its Data

- **Problem:** Explosive growth created information management challenges.
- **Solutions:** Use MDM to create an enterprise-wide set of data, preventing unnecessary data duplication.
- **Master data management (MDM)** enables companies like R.R. Donnelley to eliminate outdated, incomplete or incorrectly formatted data.
- Demonstrates IT's role in successful data management.
- Illustrates digital technology's role in storing and organizing data.

© Prentice Hall 2011

Organizing Data in a Traditional File Environment

- **File organization 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**
    - Describes an **entity** (person, place, thing on which we store information)
    - **Attribute:** Each characteristic, or quality, describing entity
      - E.g., Attributes Date or Grade belong to entity COURSE

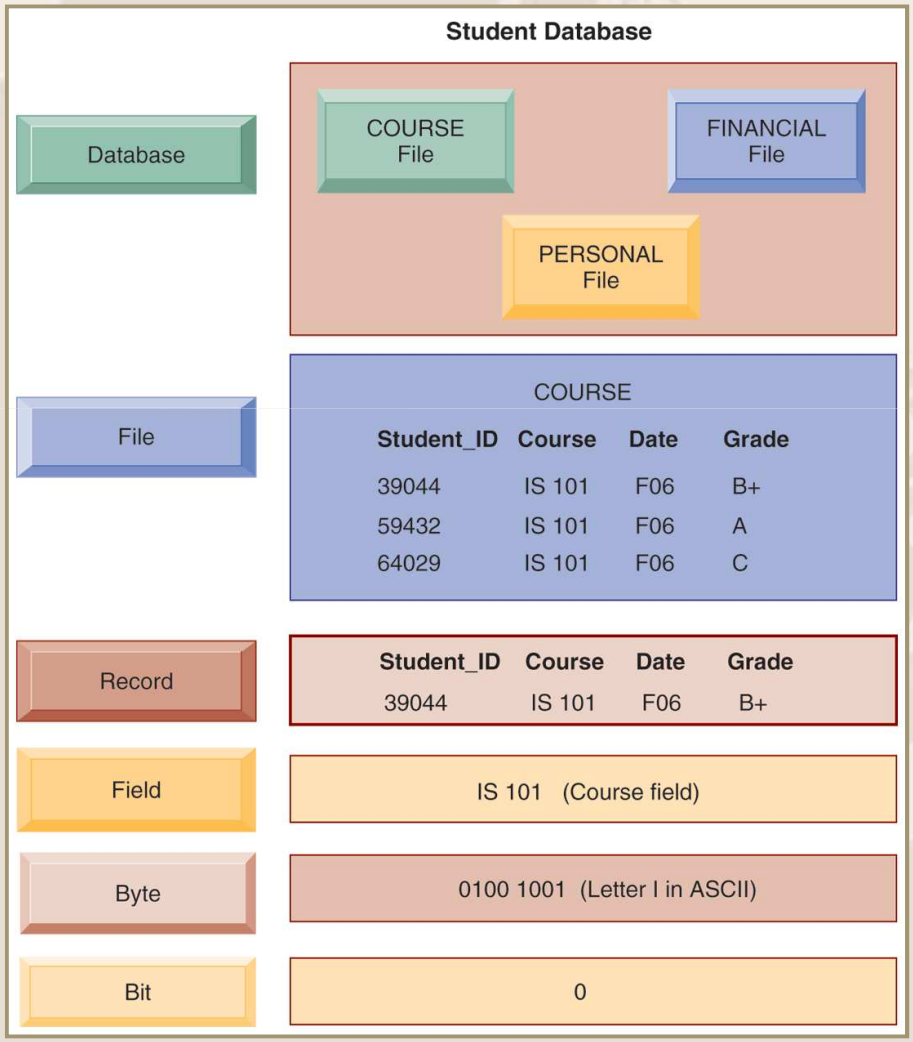
© Prentice Hall 2011

Organizing Data in a Traditional File Environment

THE DATA HIERARCHY

A computer system organizes data in a hierarchy that starts with the bit, which represents either a 0 or a 1. Bits can be grouped to form a byte to represent one character, number, or symbol. Bytes can be grouped to form a field, and related fields can be grouped to form a record. Related records can be collected to form a file, and related files can be organized into a database.

FIGURE 6-1



© Prentice Hall 2011

Organizing Data in a Traditional File Environment

- **Problems with the traditional file environment (files maintained separately by different departments)**
  - **Data redundancy:**
    - Presence of duplicate data in multiple files
  - **Data inconsistency:**
    - Same attribute has different values
  - **Program-data dependence:**
    - When changes in program requires changes to data accessed by program
  - **Lack of flexibility**
  - **Poor security**
  - **Lack of data sharing and availability**

© Prentice Hall 2011

Organizing Data in a Traditional File Environment

TRADITIONAL FILE PROCESSING

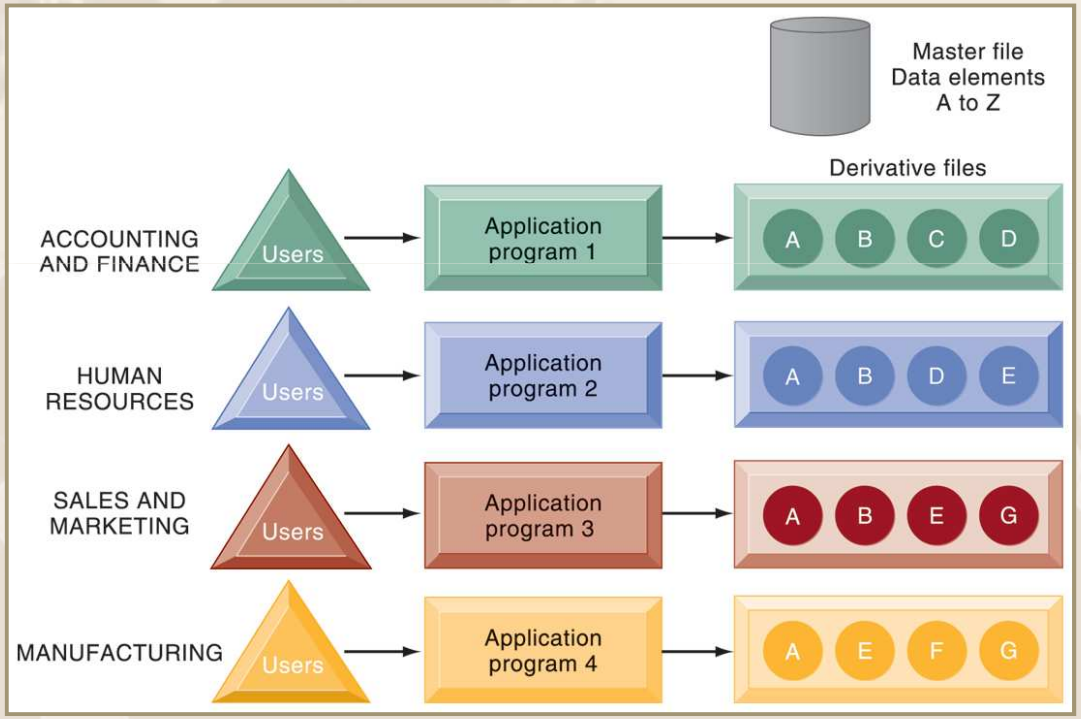


FIGURE 6-2 The use of a traditional approach to file processing encourages each functional area in a corporation to develop specialized applications. Each application requires a unique data file that is likely to be a subset of the master file. These subsets of the master file lead to data redundancy and inconsistency, processing inflexibility, and wasted storage resources.

© Prentice Hall 2011

The Database Approach to Data Management

- **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

© Prentice Hall 2011

The Database Approach to Data Management

HUMAN RESOURCES DATABASE WITH MULTIPLE VIEWS

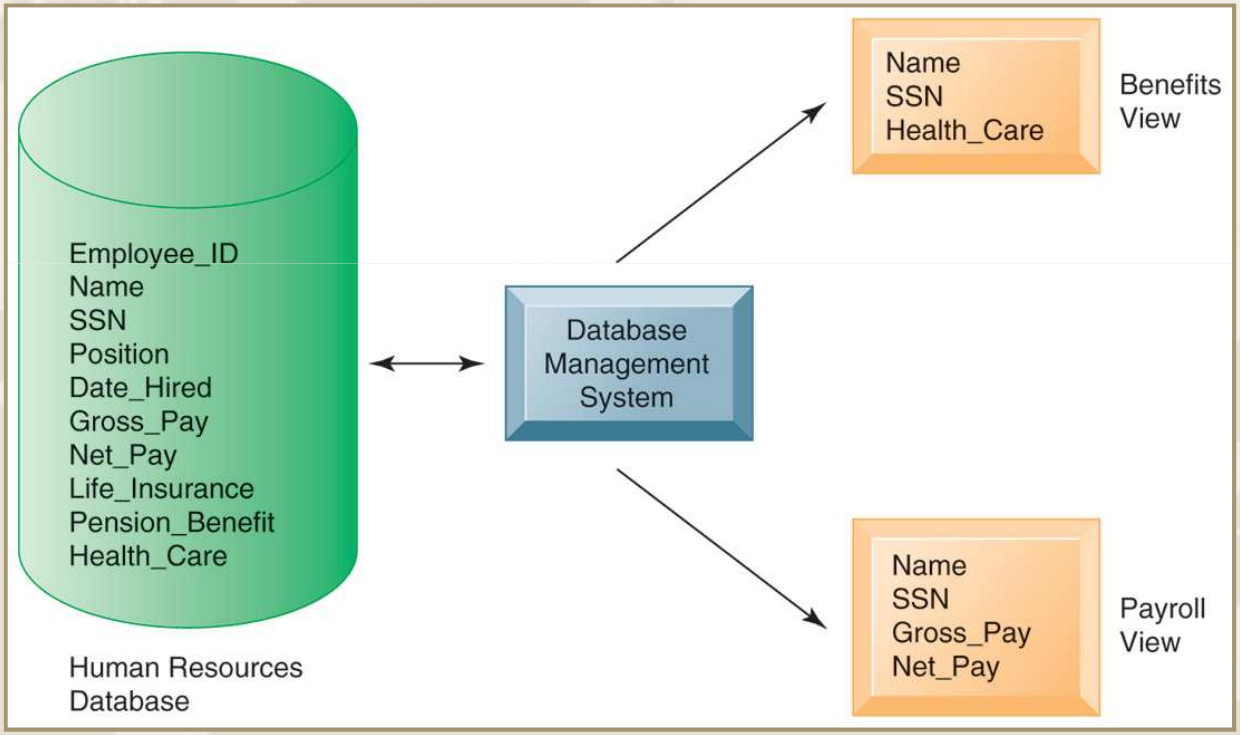


FIGURE 6-3 A single human resources database provides many different views of data, depending on the information requirements of the user. Illustrated here are two possible views, one of interest to a benefits specialist and one of interest to a member of the company's payroll department.

© Prentice Hall 2011

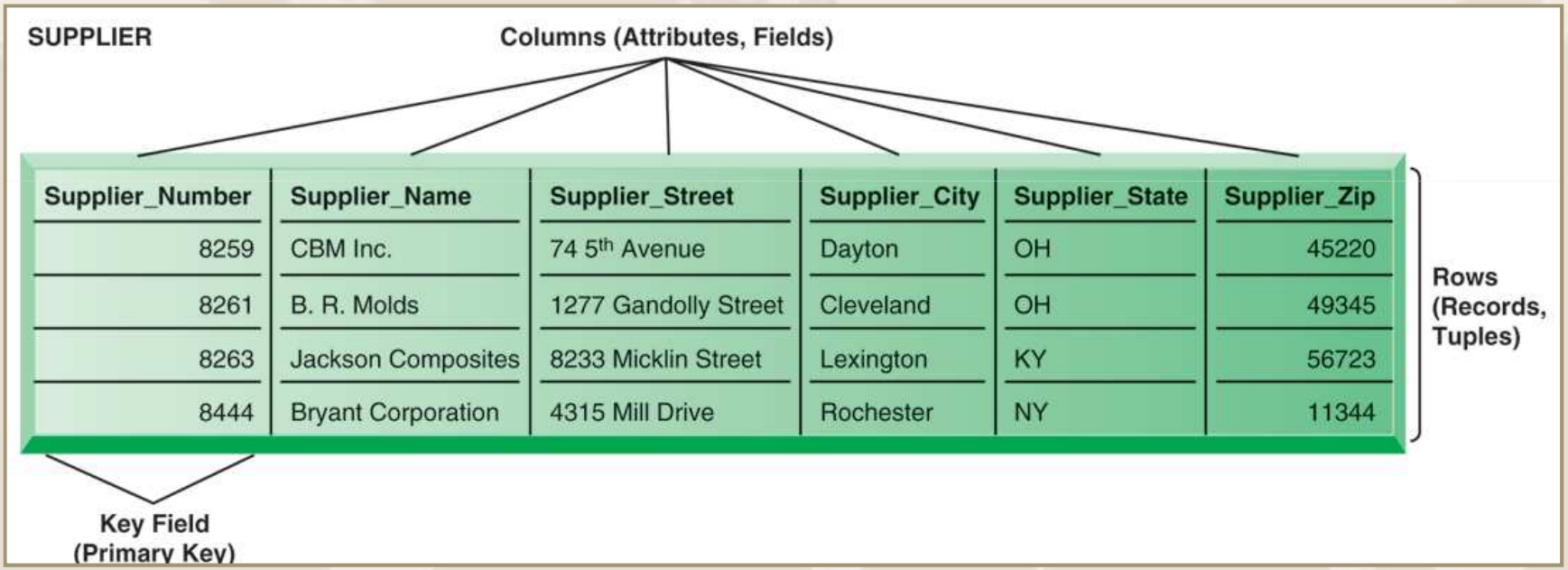
The Database Approach to Data Management

- **Relational DBMS**
  - Represent data as two-dimensional tables called relations or files
  - 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

© Prentice Hall 2011

The Database Approach to Data Management

RELATIONAL DATABASE TABLES



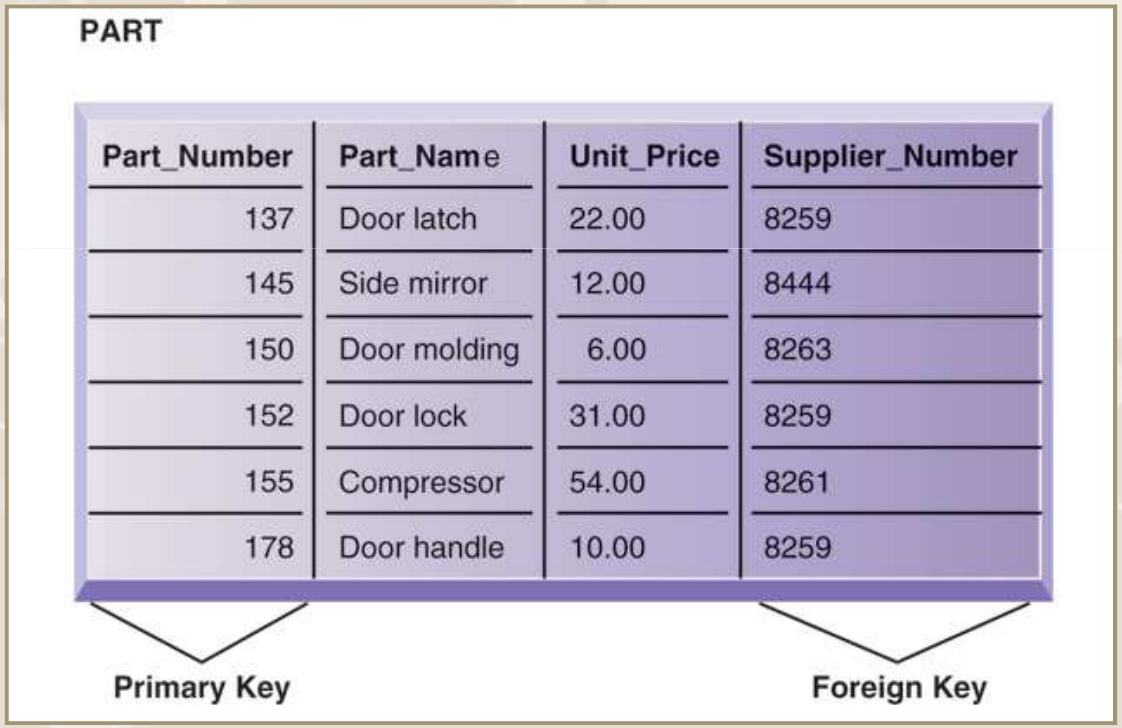
Supplier_Number	Supplier_Name	Supplier_Street	Supplier_City	Supplier_State	Supplier_Zip
8259	CBM Inc.	74 5th Avenue	Dayton	OH	45220
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345
8263	Jackson Composites	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester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

FIGURE 6-4 A relational database organizes data in the form of two-dimensional tables. Illustrated here are tables for the entities SUPPLIER and PART showing how they represent each entity and its attributes. Supplier Number is a primary key for the SUPPLIER table and a foreign key for the PART table.

© Prentice Hall 2011

The Database Approach to Data Management

RELATIONAL DATABASE TABLES (cont.)



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

FIGURE 6-4 (cont.) A relational database organizes data in the form of two-dimensional tables. Illustrated here are tables for the entities SUPPLIER and PART showing how they represent each entity and its attributes. Supplier Number is a primary key for the SUPPLIER table and a foreign key for the PART table.

© Prentice Hall 2011

The Database Approach to Data Management

- **Operations of a Relational DBMS**
  - **Three basic operations used to develop useful sets of data**
    - **SELECT:** Creates subset of data of all records that meet stated criteria
    - **JOIN:** Combines relational tables to provide user with more information than available in individual tables
    - **PROJECT:** Creates subset of columns in table, creating tables with only the information specified

© Prentice Hall 2011

The Database Approach to Data Management

THE THREE BASIC OPERATIONS OF A RELATIONAL DBMS

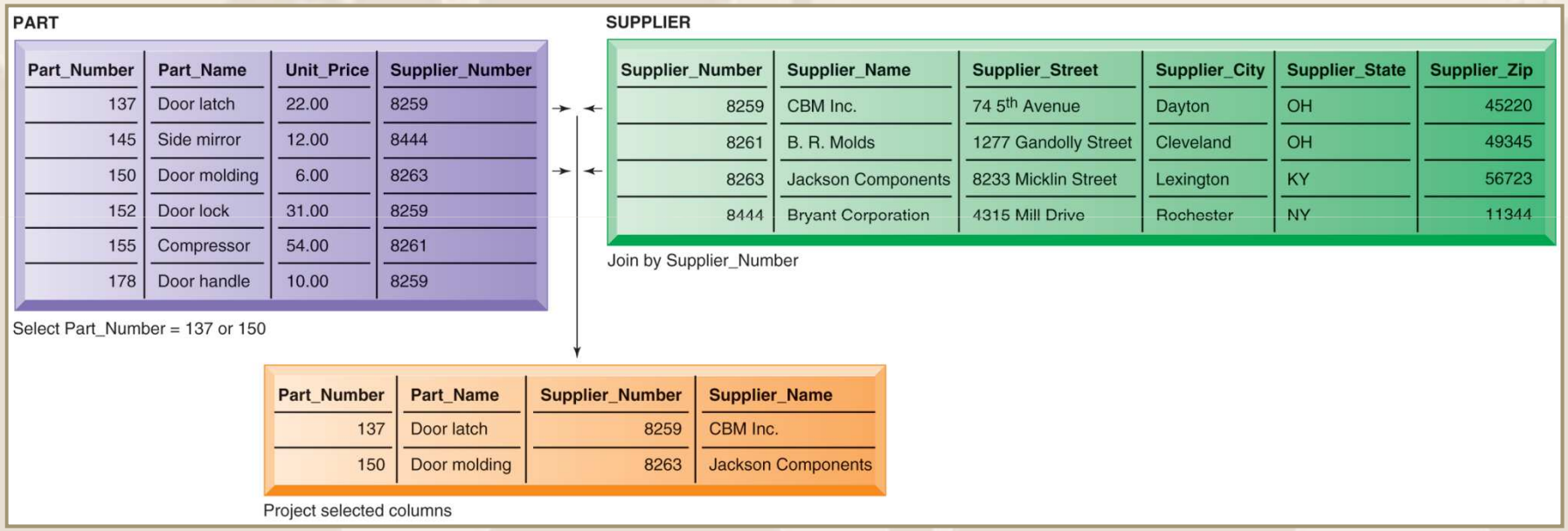


FIGURE 6-5 The select, join, and project operations enable data from two different tables to be combined and only selected attributes to be displayed.

© Prentice Hall 2011

The Database Approach to Data Management

- Object-Oriented DBMS (OODBMS)
  - Stores data and procedures as objects
  - Objects can be graphics, multimedia, Java applets
  - Relatively slow compared with relational DBMS for processing large numbers of transactions
  - Hybrid object-relational DBMS: Provide capabilities of both OODBMS and relational DBMS
- Databases in the cloud
  - Typically less functionality than on-premises DBs
  - Amazon Web Services, Microsoft SQL Azure

© Prentice Hall 2011

The Database Approach to Data Management

- **Capabilities of Database Management Systems**
  - **Data definition capability: Specifies structure of database content, used to create tables and define characteristics of fields**
  - **Data dictionary: Automated or manual file storing definitions of data elements and their characteristics**
  - **Data manipulation language: Used to add, change, delete, retrieve data from database**
    - Structured Query Language (SQL)
    - Microsoft Access user tools for generation SQL
  - **Many DBMS have report generation capabilities for creating polished reports (Crystal Reports)**

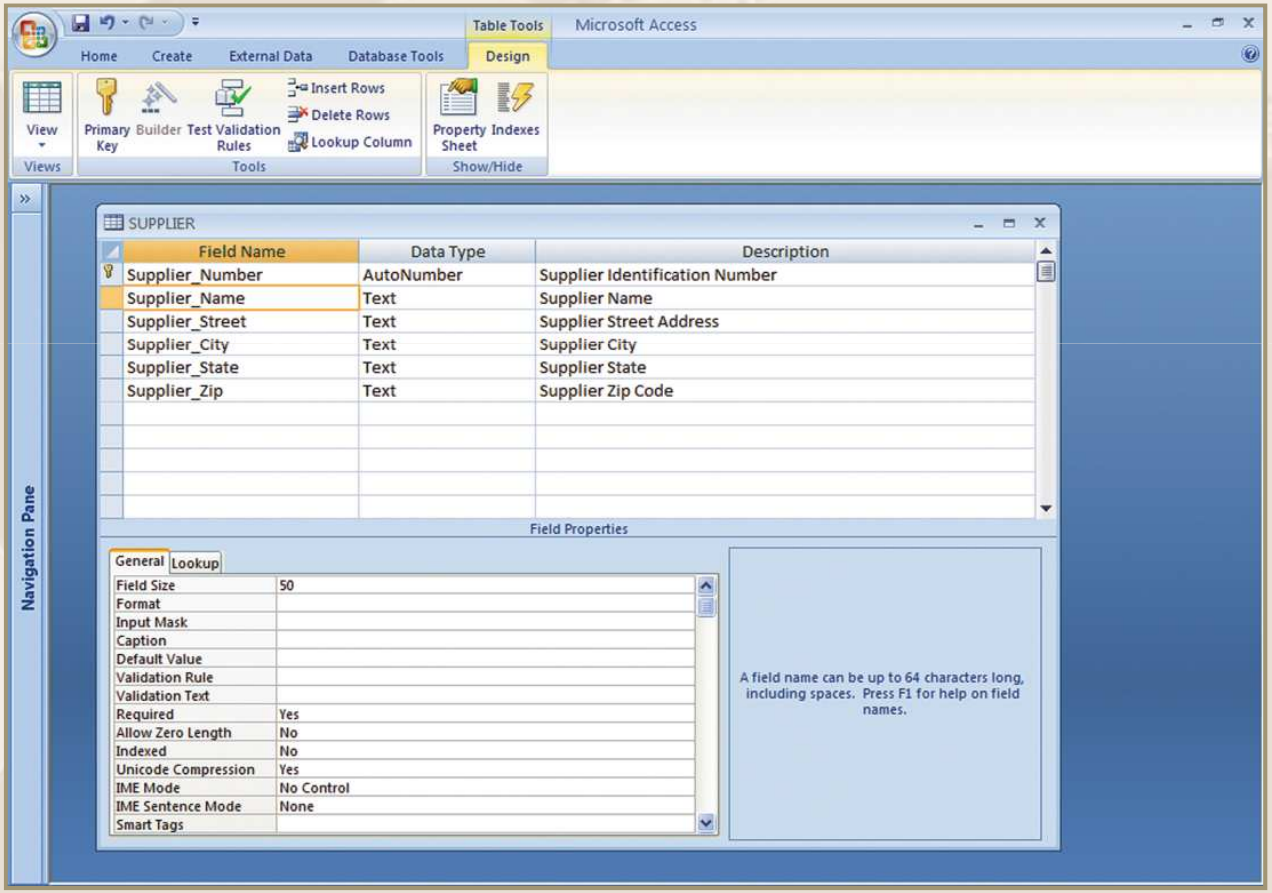
© Prentice Hall 2011

The Database Approach to Data Management

MICROSOFT 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 6-6



© Prentice Hall 2011

The Database Approach to Data Management

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;
```

FIGURE 6-7 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 6-5.

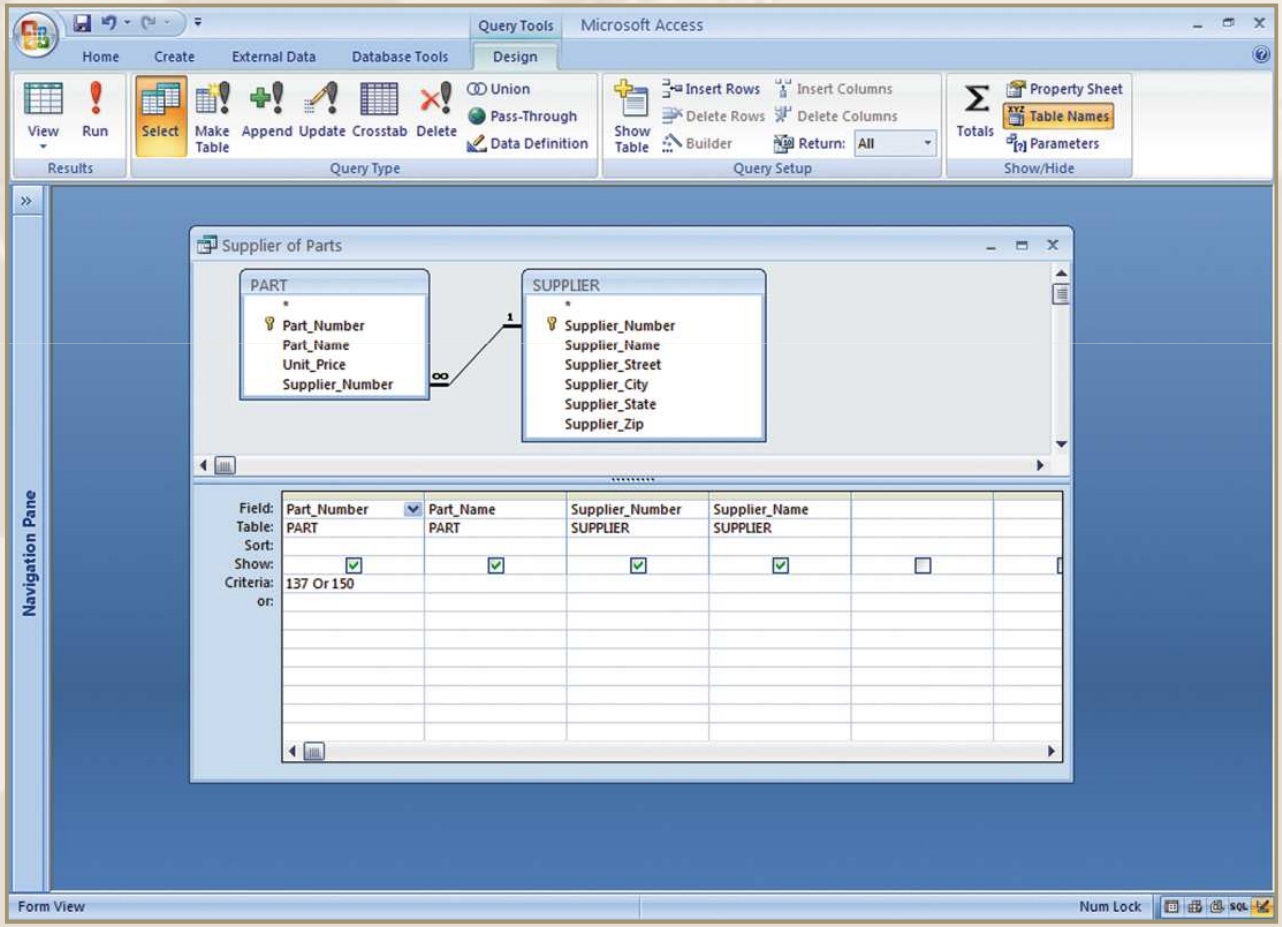
© Prentice Hall 2011

The Database Approach to Data Management

AN ACCESS QUERY

Illustrated here is how the query in Figure 6-7 would be constructed using Microsoft Access query building tools. It shows the tables, fields, and selection criteria used for the query.

FIGURE 6-8



© Prentice Hall 2011

The Database Approach to Data Management

- **Designing Databases**
  - Conceptual (logical) design: Abstract model from business perspective
  - Physical design: How database is arranged on direct-access storage devices
- **Design process identifies**
  - Relationships among data elements, redundant database elements
  - Most efficient way to group data elements to meet business requirements, needs of application programs
- **Normalization**
  - Streamlining complex groupings of data to minimize redundant data elements and awkward many-to-many relationships

© Prentice Hall 2011

The Database Approach to Data Management

AN UNNORMALIZED RELATION FOR ORDER

FIGURE 6-9 An unnormalized relation contains repeating groups. For example, there can be many parts and suppliers for each order. There is only a one-to-one correspondence between Order_Number and Order_Date.

© Prentice Hall 2011

The Database Approach to Data Management

NORMALIZED TABLES CREATED FROM ORDER

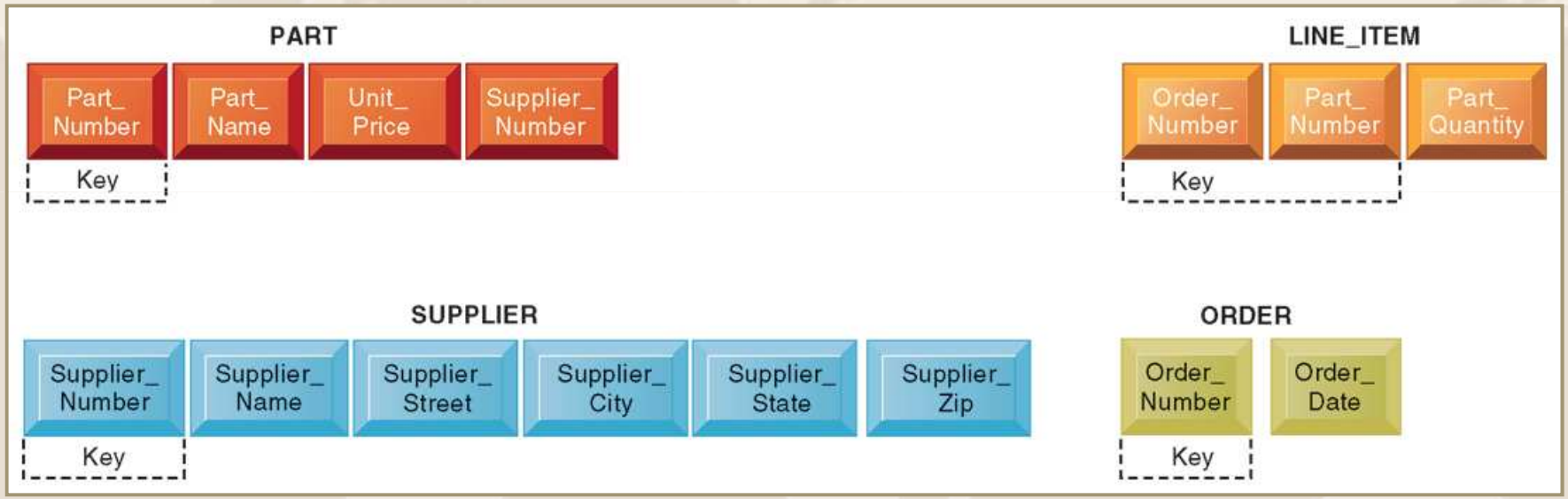


FIGURE 6-10 An unnormalized relation contains repeating groups. For example, there can be many parts and suppliers for each order. There is only a one-to-one correspondence between Order_Number and Order_Date.

© Prentice Hall 2011

The Database Approach to Data Management

- Entity-relationship diagram
  - Used by database designers to document the data model
  - Illustrates relationships between entities
- Distributing databases: Storing database in more than one place
  - Partitioned: Separate locations store different parts of database
  - Replicated: Central database duplicated in entirety at different locations

© Prentice Hall 2011

The Database Approach to Data Management

AN ENTITY-RELATIONSHIP DIAGRAM

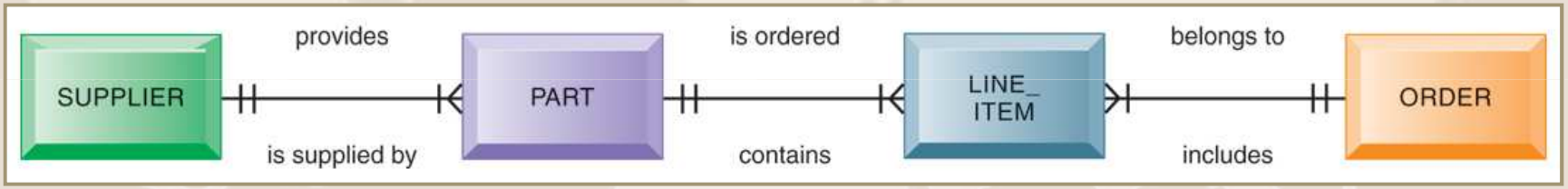


FIGURE 6-11 This diagram shows the relationships between the entities SUPPLIER, PART, LINE_ITEM, and ORDER that might be used to model the database in Figure 6-10.

© Prentice Hall 2011

Using Databases to Improve Business Performance and Decision Making

- Very large databases and systems require special capabilities, tools
  - To analyze large quantities of data
  - To access data from multiple systems
- Three key techniques
  1. Data warehousing
  2. Data mining
  3. Tools for accessing internal databases through the Web

© Prentice Hall 2011

Using Databases to Improve Business Performance and Decision Making

- **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
  - Data warehouse system will provide query, analysis, and reporting tools
- **Data marts:**
  - Subset of data warehouse
  - Summarized or highly focused portion of firm's data for use by specific population of users
  - Typically focuses on single subject or line of business

© Prentice Hall 2011

The Database Approach to Data Management

COMPONENTS OF A DATA WAREHOUSE

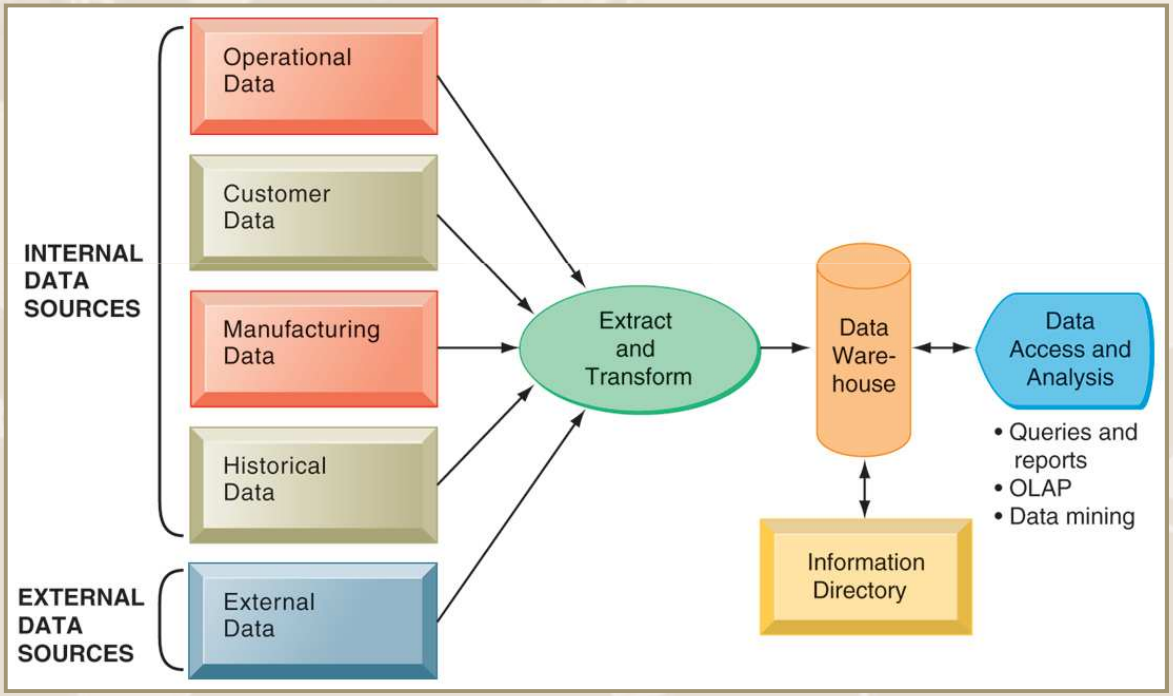


FIGURE 6-12 The data warehouse extracts current and historical data from multiple operational systems inside the organization. These data are combined with data from external sources and reorganized into a central database designed for management reporting and analysis. The information directory provides users with information about the data available in the warehouse.

© Prentice Hall 2011

Using Databases to Improve Business Performance and Decision Making

- **Business Intelligence:**
  - **Tools for consolidating, analyzing, and providing access to vast amounts of data to help users make better business decisions**
  - **E.g., Harrah's Entertainment analyzes customers to develop gambling profiles and identify most profitable customers**
  - **Principle tools include:**
    - Software for database query and reporting
    - Online analytical processing (OLAP)
    - Data mining

© Prentice Hall 2011

Using Databases to Improve Business Performance and Decision Making

- **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
    - E.g., how many washers sold in the East in June compared with other regions?
  - **OLAP enables rapid, online answers to ad hoc queries**

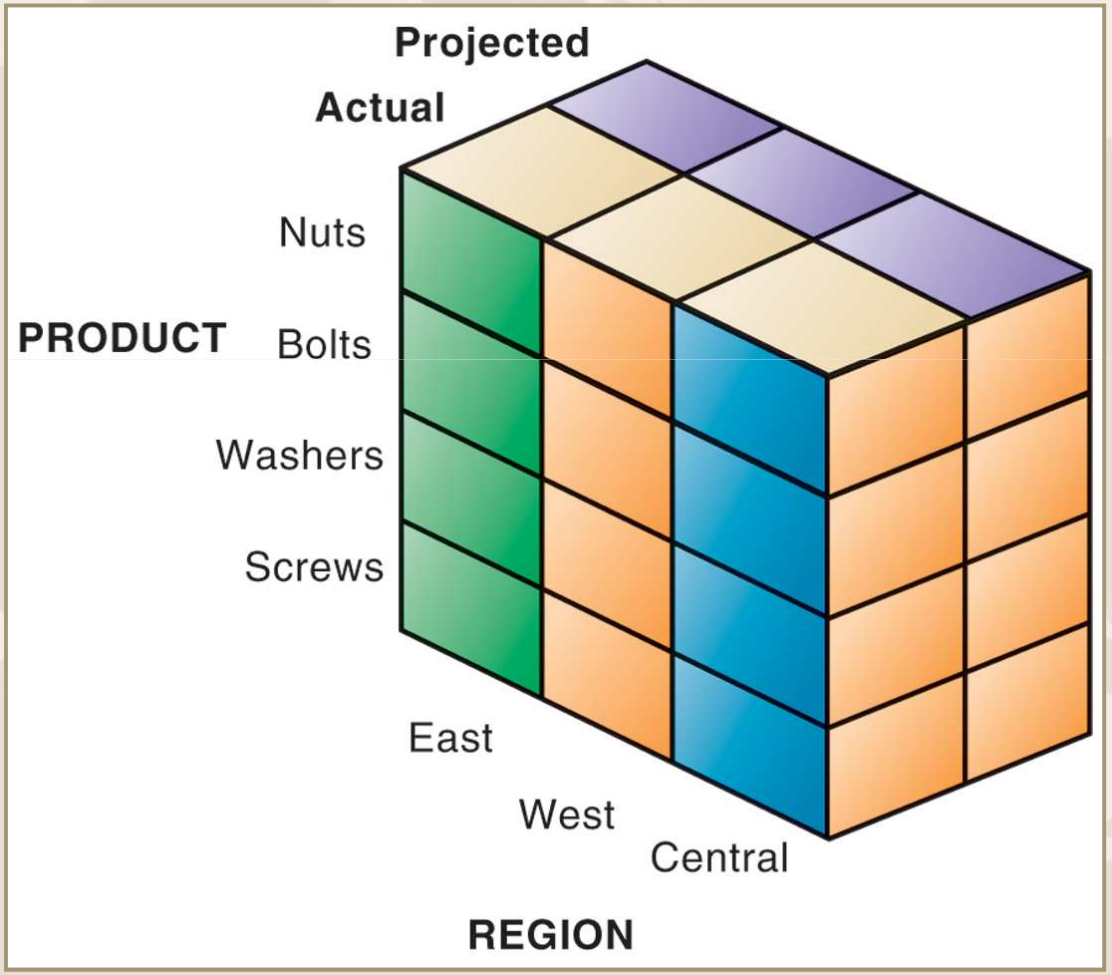
© Prentice Hall 2011

The Database Approach to Data Management

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 6-13



© Prentice Hall 2011

Using Databases to Improve Business Performance and Decision Making

- **Data mining:**
  - **More discovery driven than OLAP**
  - **Finds hidden patterns, relationships in large databases and infers rules to predict future behavior**
  - **E.g., Finding patterns in customer data for one-to-one marketing campaigns or to identify profitable customers.**
  - **Types of information obtainable from data mining**
    - Associations
    - Sequences
    - Classification
    - Clustering
    - Forecasting

© Prentice Hall 2011

Using Databases to Improve Business Performance and Decision Making

- **Predictive analysis**
  - Uses data mining techniques, historical data, and assumptions about future conditions to predict outcomes of events
  - E.g., Probability a customer will respond to an offer
- **Text mining**
  - Extracts key elements from large unstructured data sets (e.g., stored e-mails)

© Prentice Hall 2011

Using Databases to Improve Business Performance and Decision Making

WHAT CAN BUSINESSES LEARN FROM TEXT MINING?

Read the Interactive Session and discuss the following questions

- **What challenges does the increase in unstructured data present for businesses?**
- **How does text-mining improve decision-making?**
- **What kinds of companies are most likely to benefit from text mining software? Explain your answer.**
- **In what ways could text mining potentially lead to the erosion of personal information privacy? Explain.**

© Prentice Hall 2011

Using Databases to Improve Business Performance and Decision Making

- **Web mining**
  - **Discovery and analysis of useful patterns and information from WWW**
    - E.g., to understand customer behavior, evaluate effectiveness of Web site, etc.
  - **Web content mining**
    - Knowledge extracted from content of Web pages
  - **Web structure mining**
    - E.g., links to and from Web page
  - **Web usage mining**
    - User interaction data recorded by Web server

© Prentice Hall 2011

Using Databases to Improve Business Performance and Decision Making

- **Databases and the Web**
  - **Many companies use Web to make some internal databases available to customers or partners**
  - **Typical configuration includes:**
    - Web server
    - Application server/middleware/CGI scripts
    - Database server (hosting DBM)
  - **Advantages of using 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

© Prentice Hall 2011

The Database Approach to Data Management

LINKING INTERNAL DATABASES TO THE WEB

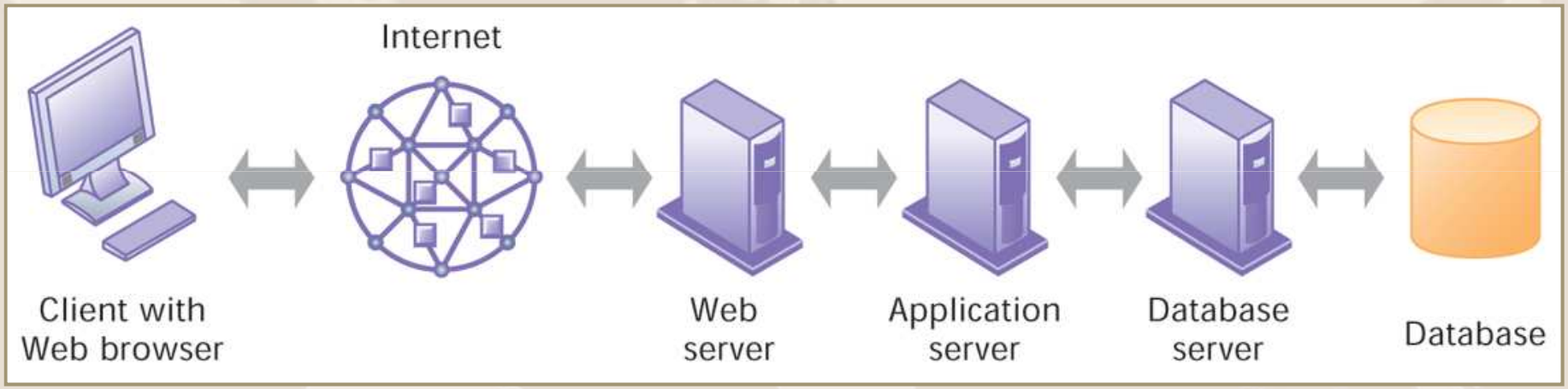


FIGURE 6-14 Users access an organization's internal database through the Web using their desktop PCs and Web browser software.

© Prentice Hall 2011

Managing Data Resources

- **Establishing an information policy**
  - **Firm's rules, procedures, roles for sharing, managing, standardizing data**
  - **Data administration:**
    - Firm function responsible for specific policies and procedures to manage data
  - **Data governance:**
    - Policies and processes for managing availability, usability, integrity, and security of enterprise data, especially as it relates to government regulations
  - **Database administration:**
    - Defining, organizing, implementing, maintaining database; performed by database design and management group

© Prentice Hall 2011

Managing Data Resources

- **Ensuring data quality**
  - **More than 25% of critical data in Fortune 1000 company databases are inaccurate or incomplete**
  - **Most data quality problems stem from faulty input**
  - **Before new database in place, need to:**
    - Identify and correct faulty data
    - Establish better routines for editing data once database in operation

© Prentice Hall 2011

Managing Data Resources

- **Data quality audit:**
  - **Structured survey of the accuracy and level of completeness of the data in an information system**
    - Survey samples from data files, or
    - Survey end users for perceptions of quality
- **Data cleansing**
  - **Software to detect and correct data that are incorrect, incomplete, improperly formatted, or redundant**
  - **Enforces consistency among different sets of data from separate information systems**

© Prentice Hall 2011

Managing Data Resources

CREDIT BUREAU ERRORS—BIG PEOPLE PROBLEMS

Read the Interactive Session and discuss the following questions

- **Assess the business impact of credit bureaus' data quality problems for the credit bureaus, for lenders, for individuals.**
- **Are any ethical issues raised by credit bureaus' data quality problems? Explain your answer.**
- **Analyze the people, organization, and technology factors responsible for credit bureaus' data quality problems.**
- **What can be done to solve these problems?**

© Prentice Hall 2011

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.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

Printed in the United States of America.

Copyright © 2011 Pearson Education, Inc.

Publishing as Prentice Hall