Chapter 6

Foundations of Business Intelligence: Databases and Information Management

Management Information Systems
Chapter 6 Foundations of Business Intelligence: Databases and Information Management

Organizing Data in a Traditional File Environment

˝ **File organization concepts**
  - ˝ **Computer system organizes data in a hierarchy**
    - ˝ **Field:** Group of characters as word(s) or number
    - ˝ **Record:** Group of related fields
    - ˝ **File:** Group of records of same type
    - ˝ **Database:** Group of related files
  - ˝ **Record**: Describes an entity
  - ˝ **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**

Organizing Data in a Traditional File Environment
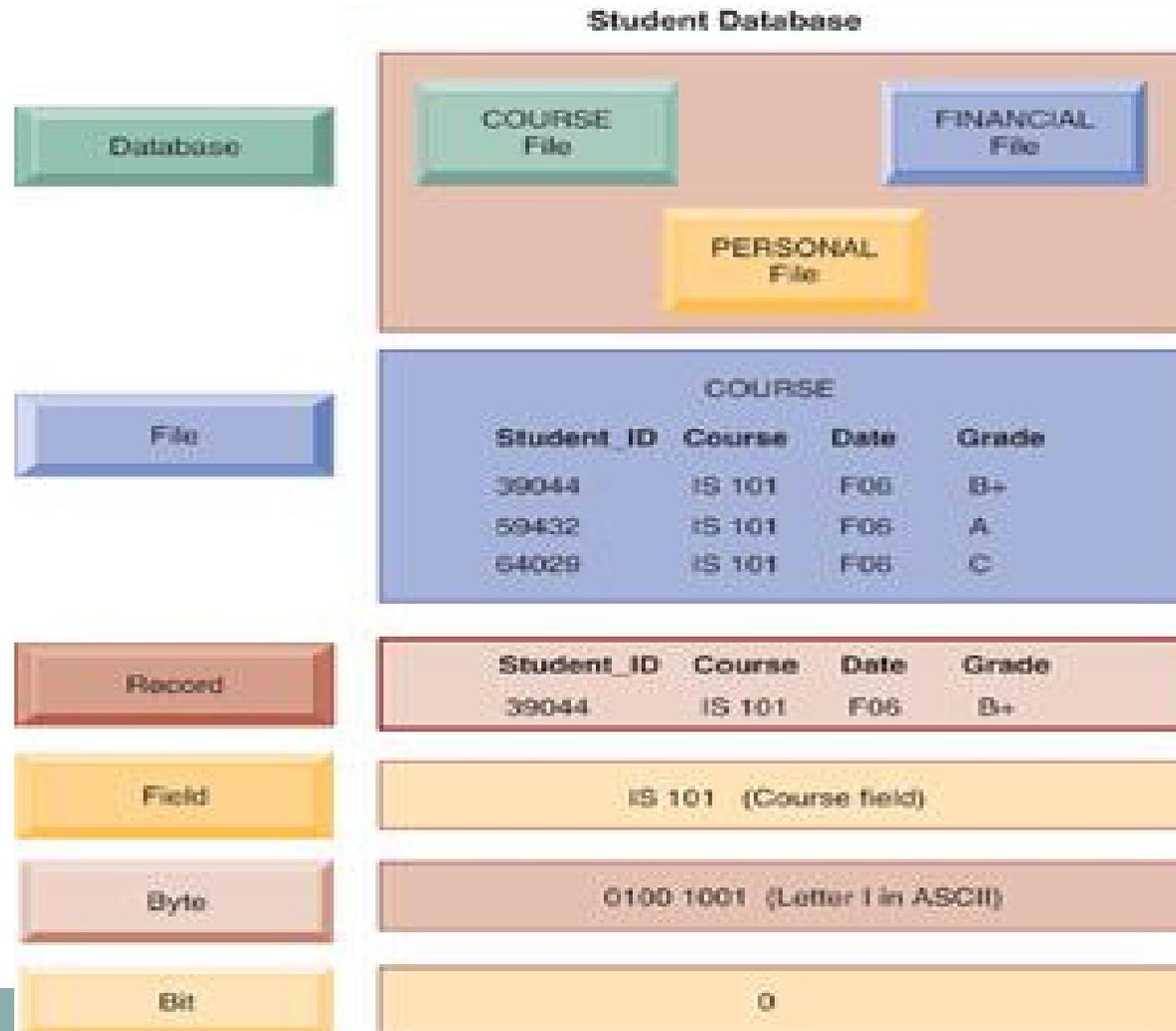
The Data Hierarchy

Figure 6-1

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.

Student Database

Database

COURSE File

FINANCIAL File

PERSONAL File

File

COURSE

Student_ID	Course	Date	Grade
39044	IS 101	F06	B+
59432	IS 101	F06	A
64029	IS 101	F06	C

Record

Student_ID	Course	Date	Grade
39044	IS 101	F06	B+

Field

IS 101   (Course field)

Byte

0100 1001   (Letter I in ASCII)

Bit

0

Management Information Systems
Chapter 6 Foundations of Business Intelligence: Databases and Information Management
### Organizing Data in a Traditional File Environment

˝ **Problems with the traditional file environment (files maintained separately by different departments)**

  ˝ **Data redundancy and inconsistency**

    ˝ **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**

Organizing Data in a Traditional File Environment

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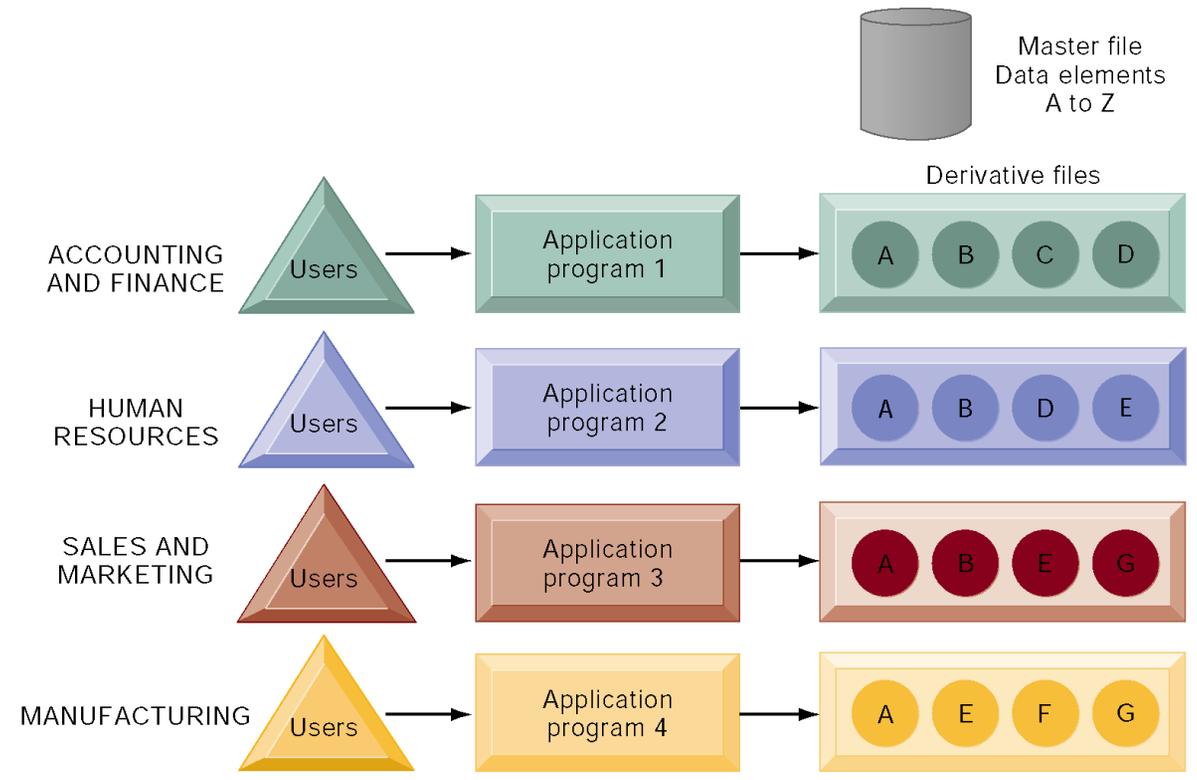
The use of a traditional approach to file processing encourages each functional area in a corporation to develop specialized applications and files. Each application requires a unique data file that is likely to be a subset of the master file. These subsets of the master file lead to data redundancy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

Figure 6-2

Database

- Collection of data organized to serve many applications by centralizing data and controlling redundant data

DataBase Management System (DBMS)

- Interfaces between application programs 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 central manage data and data security

Management Information Systems
Chapter 6 Foundations of Business Intelligence: Databases and Information Management
The Database Approach to Data Management
Human Resources Database with Multiple Views
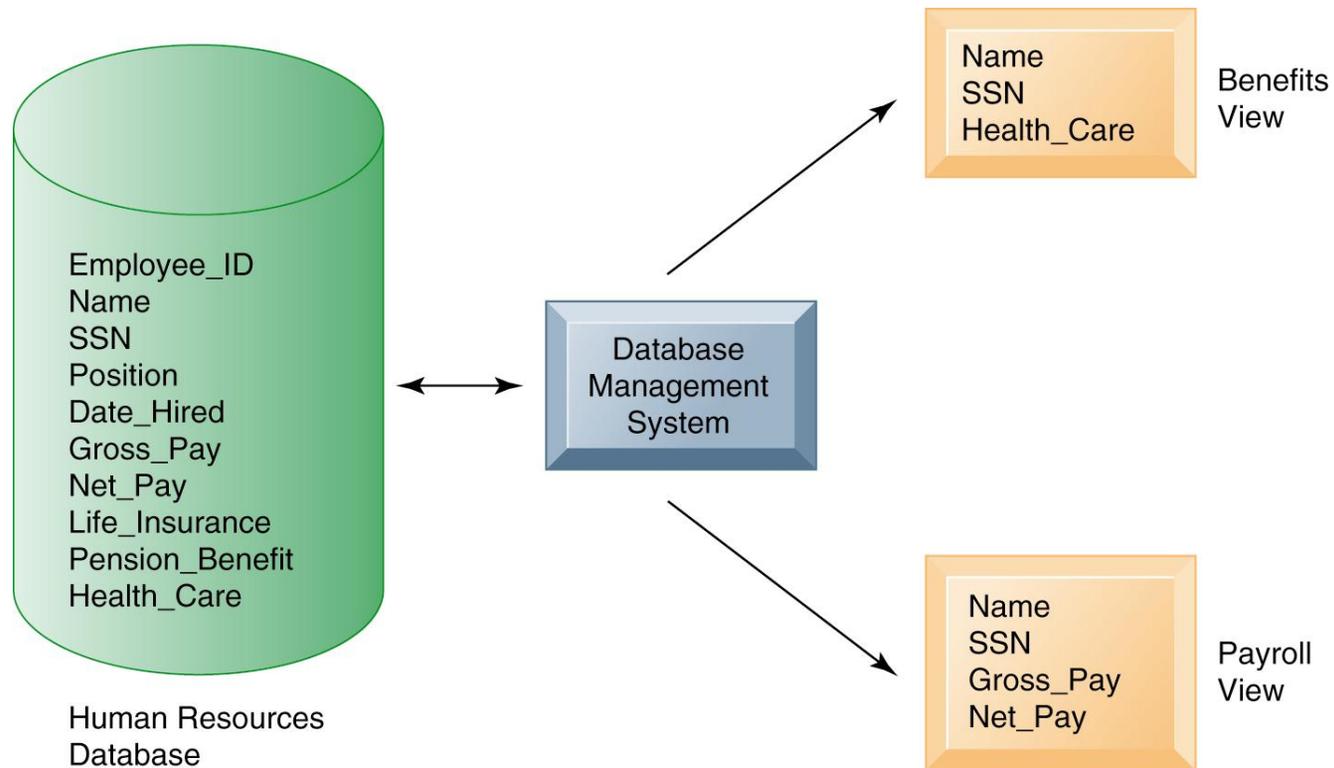
A single human resources database provides many different views of data, depending on the information requirements of the user. Illustrated here are two possible views, one of interest to a benefits specialist and one of interest to a member of the company's payroll department.

Figure 6-3

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

Management Information Systems
Chapter 6 Foundations of Business Intelligence: Databases and Information Management
The Database Approach to Data Management
Relational Database Tables

SUPPLIER   Columns (Attributes, Fields)

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

Rows (Records, Tuples)

Key Field (Primary Key)

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.

Figure 6-4A

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 6-4B

Management Information Systems
**Chapter 6 Foundations of Business Intelligence: Databases
and Information Management**

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

The Database Approach to Data Management

The Three Basic Operations of a Relational DBMS

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

Select Part_Number = 137 or 150

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

Join by Supplier_Number

Part_Number	Part_Name	Supplier_Number	Supplier_Name
137	Door latch	8259	CBM Inc.
150	Door molding	8263	Jackson Components

Project selected columns

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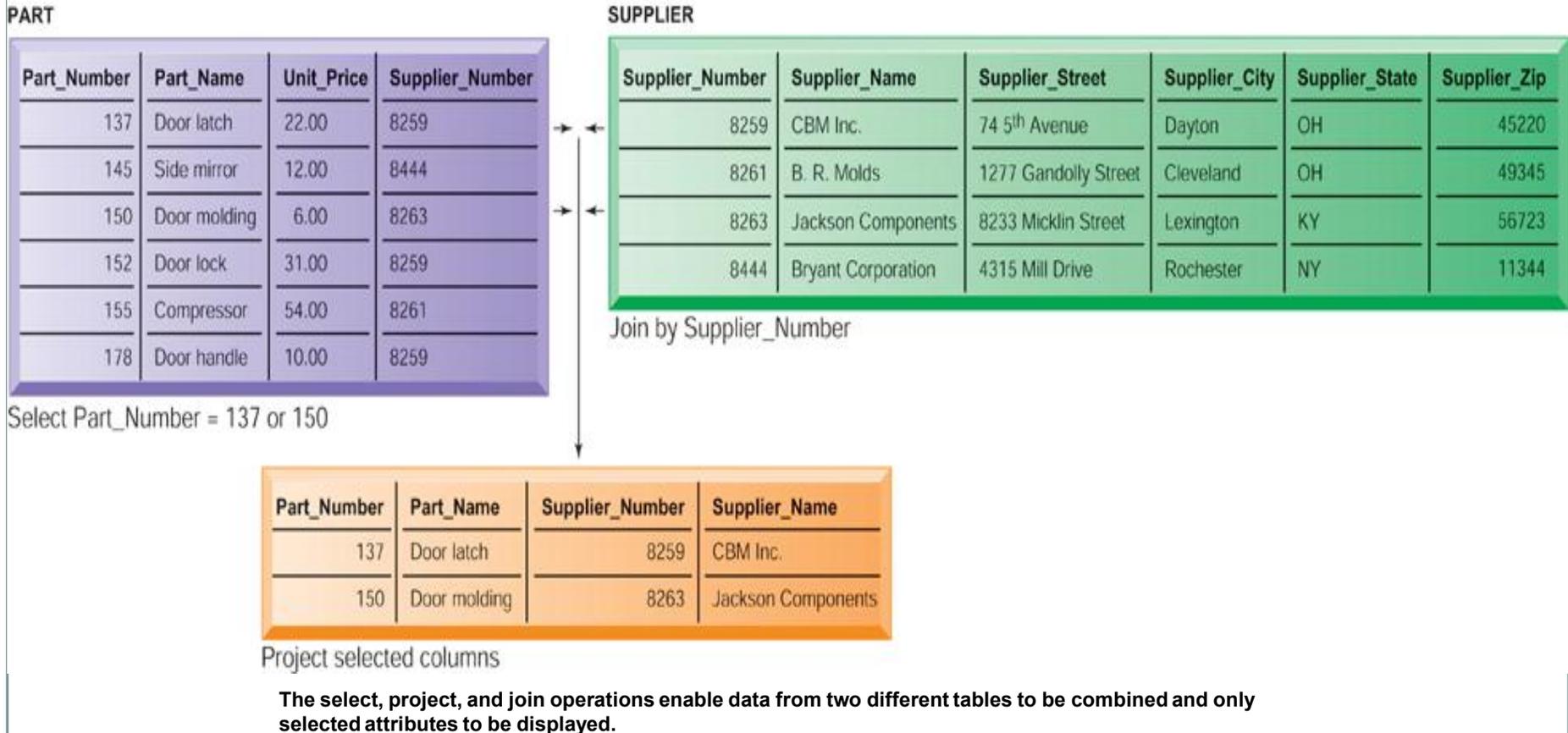
Figure 6-5

Management Information Systems
**Chapter 6 Foundations of Business Intelligence: Databases
and Information Management**

The Database Approach to Data Management

˝ **Object-Oriented DBMS (OODBMS)**

- ˝ Stores data and procedures as objects

- ˝ Capable of managing 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

Management Information Systems
**Chapter 6 Foundations of Business Intelligence: Databases
and Information Management**

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)

Management Information Systems
**Chapter 6 Foundations of Business Intelligence: Databases
and Information Management**

The Database Approach to Data Management

Microsoft Access Data Dictionary Features
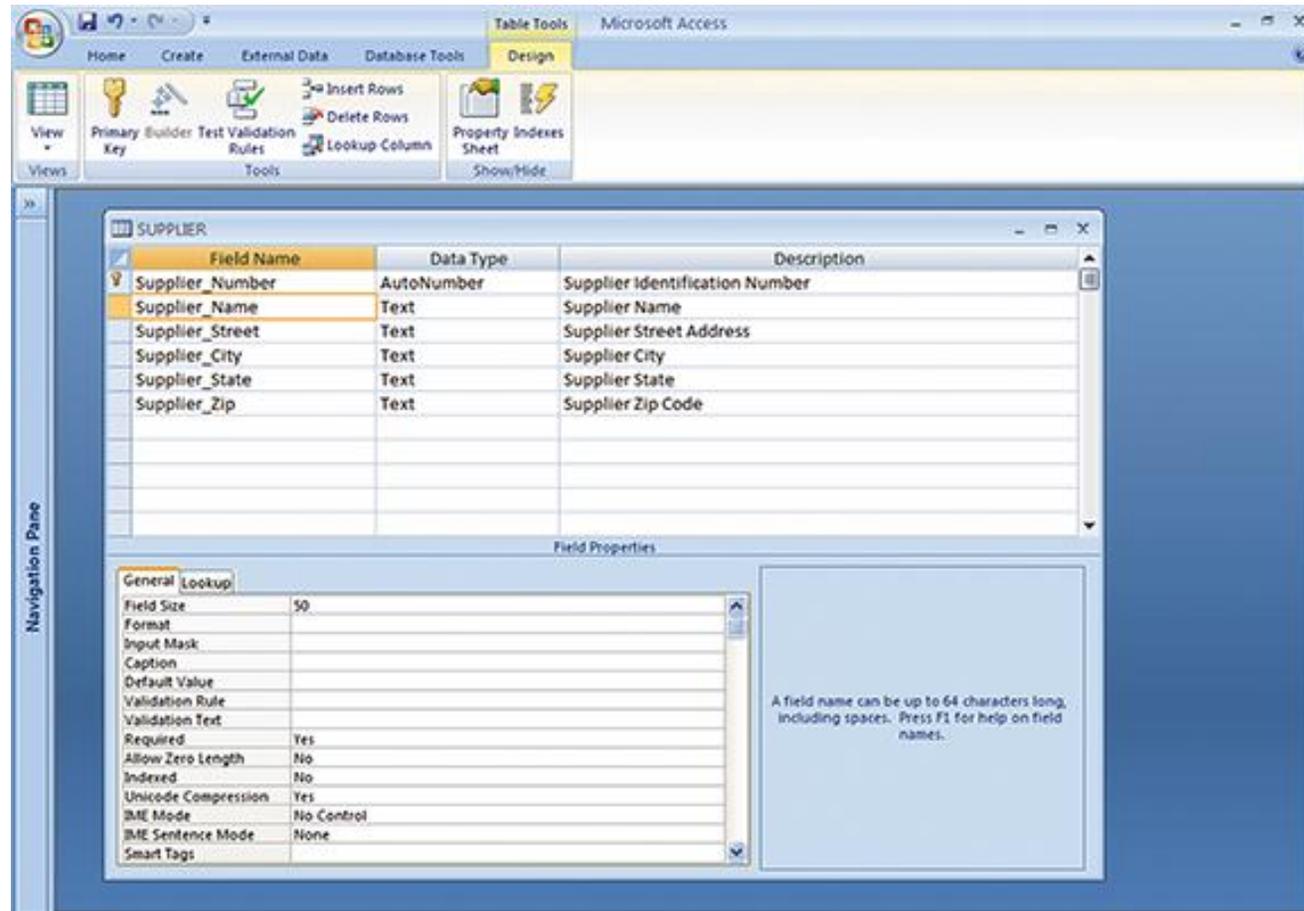
Figure 6-6

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.

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;
```

Illustrated here are the SQL statements for a query to select suppliers for parts 137 or 150. They produce a
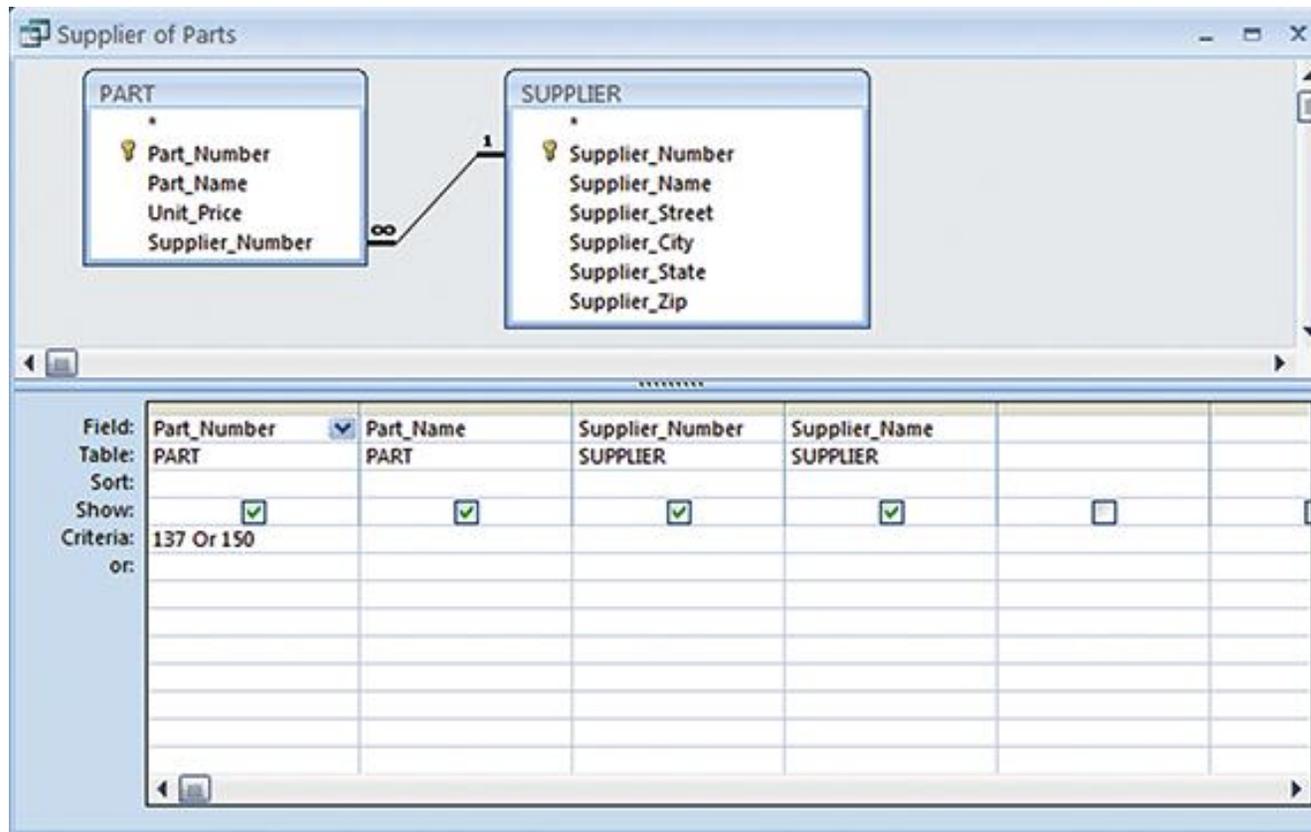list with the same results as Figure 6-5.

Figure 6-7

The Database Approach to Data Management

An Access Query

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

Figure 6-8

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

The Database Approach to Data Management

An Unnormalized Relation for Order

ORDER (Before Normalization)

Order_ Number	Order_ Date	Part_ Number	Part_ Name	Unit_ Price	Part_ Quantity	Supplier_ Number	Supplier_ Name	Supplier_ Street	Supplier_ City	Supplier_ State	Supplier_ Zip

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.
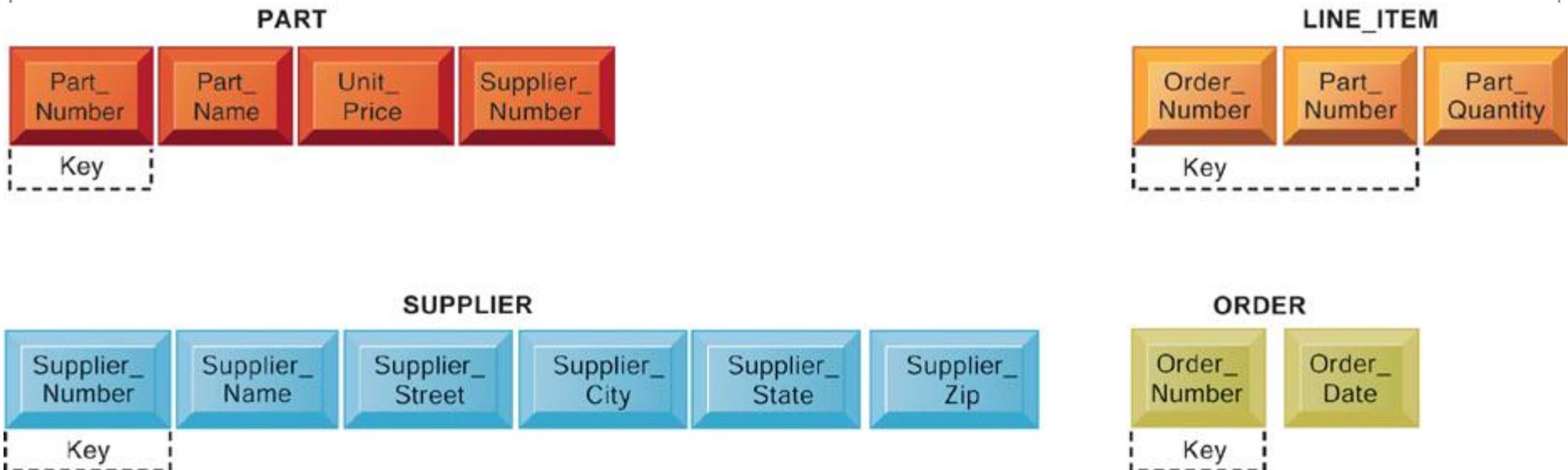
Figure 6-9

Normalized Tables Created from Order

After normalization, the original relation ORDER has been broken down into four smaller relations. The relation ORDER is left with only two attributes and the relation LINE_ITEM has a combined, or concatenated, key consisting of Order_Number and Part_Number.

Figure 6-10

Management Information Systems
**Chapter 6 Foundations of Business Intelligence: Databases
and Information Management**

The Database Approach to Data Management

˝ **Entity-relationship diagram**

   ˝   Used by database designers to document the data model
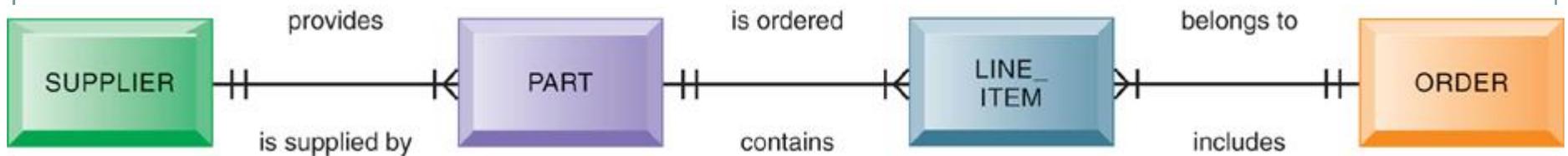
   ˝   Illustrates relationships between entities

˝ **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

Management Information Systems
Chapter 6 Foundations of Business Intelligence: Databases and Information Management
The Database Approach to Data Management
An Entity-Relationship Diagram

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

Figure 6-11

Management Information Systems
**Chapter 6 Foundations of Business Intelligence: Databases
and Information Management**

The Database Approach to Data Management

˝ **Distributing databases**

  ˝ Two main methods of distributing a database

    ˝ **Partitioned**: Separate locations store different parts of database

    ˝ **Replicated**: Central database duplicated in entirety at different locations

  ˝ Advantages

    ˝ Reduced vulnerability

    ˝ Increased responsiveness

  ˝ Drawbacks
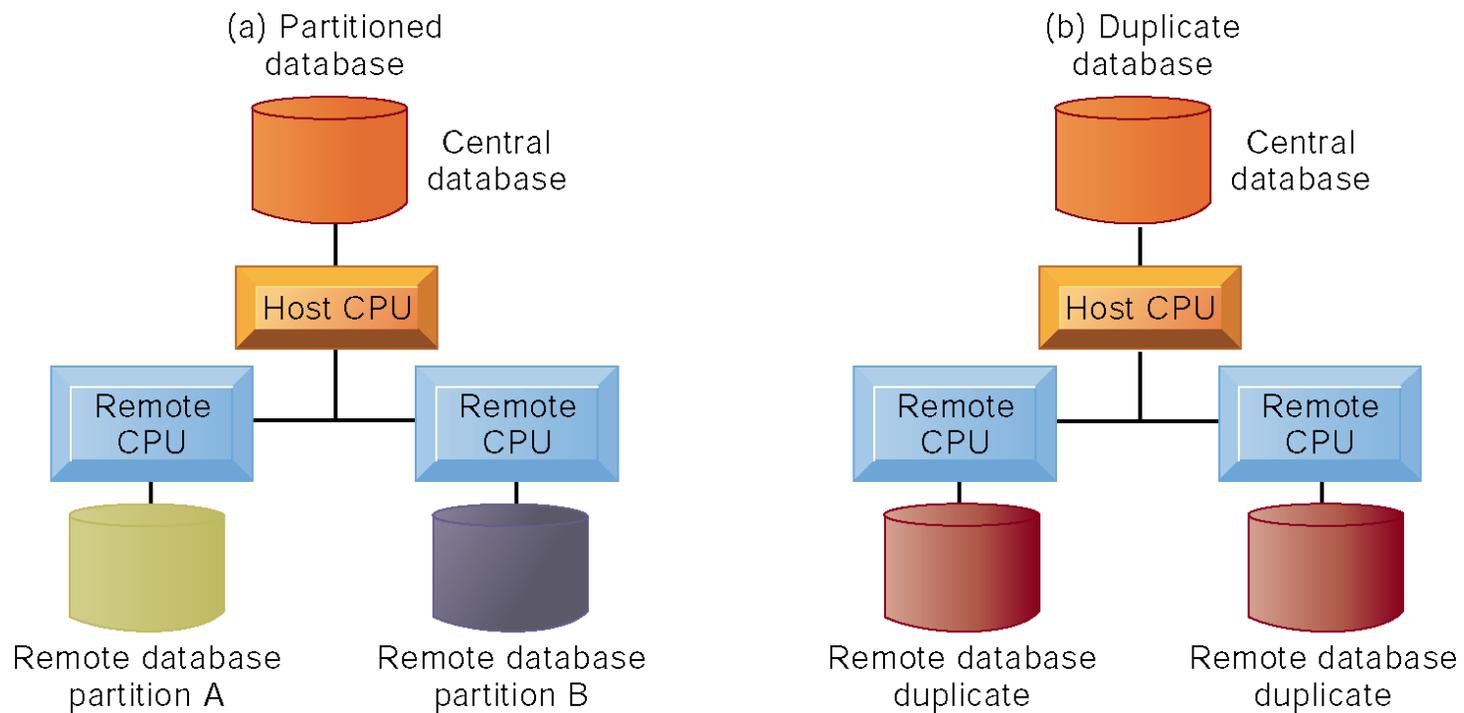
    ˝ Departures from using standard definitions

    ˝ Security problems

Distributed Databases

(a) Partitioned database

Central database

Host CPU

Remote CPU

Remote CPU

Remote database partition A

Remote database partition B

(b) Duplicate database

Central database

Host CPU

Remote CPU

Remote CPU

Remote database duplicate

Remote database duplicate

There are alternative ways of distributing a database. The central database can be partitioned (a) so that each remote processor has the necessary data to serve its own local needs. The central database also can be replicated (b) at all remote locations.

Figure 6-12

Management Information Systems
Chapter 6 Foundations of Business Intelligence: Databases and Information Management
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

  ˝ Data warehousing

  ˝ Data mining

  ˝ Tools for accessing internal databases through the Web