DATA WAREHOUSES & DATA MINING.

- Databases – Recap
- Data Warehouses
- OLAP/OLTP

Management Information Systems
MANAGING THE DIGITAL FIRM, 12TH EDITION GLOBAL EDITION

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

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

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

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

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- Master data management (MDM) enables companies like R.R. Donnelley to eliminate outdated, incomplete or incorrectly formatted data.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.

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

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

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

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.

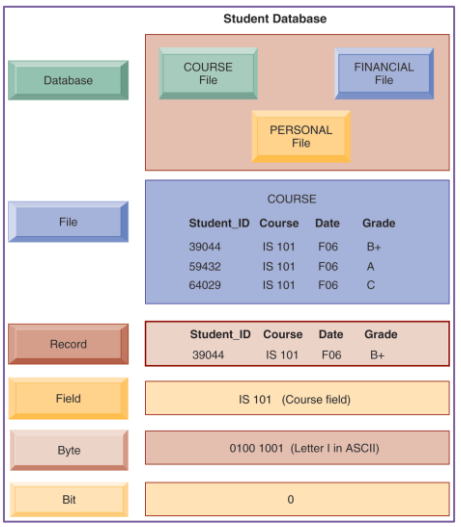


FIGURE 6-1

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:**
    - 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**

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

Organizing Data in a Traditional File Environment

TRADITIONAL FILE PROCESSING

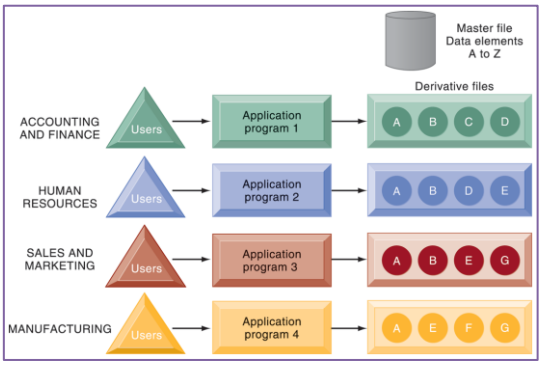


FIGURE 6-2 The use of a traditional approach to file processing encourages each functional area in a corporation to develop specialized applications. Each application requires a unique data file that is likely to be a subset of the master file. These subsets of the master file lead to data redundancy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

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

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

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

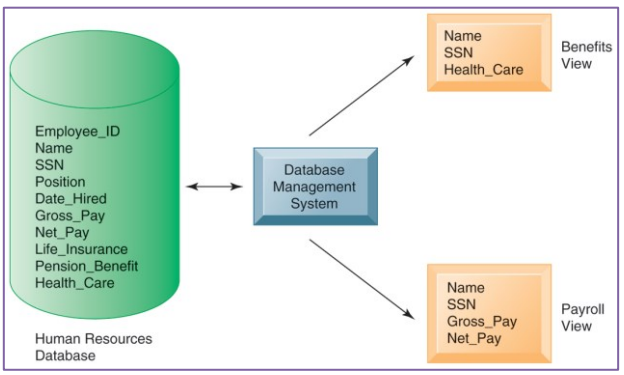


FIGURE 6-3 A single human resources database provides many different views of data, depending on the information requirements of the user. Illustrated here are two possible views, one of interest to a benefits specialist and one of interest to a member of the company's payroll department.

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

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

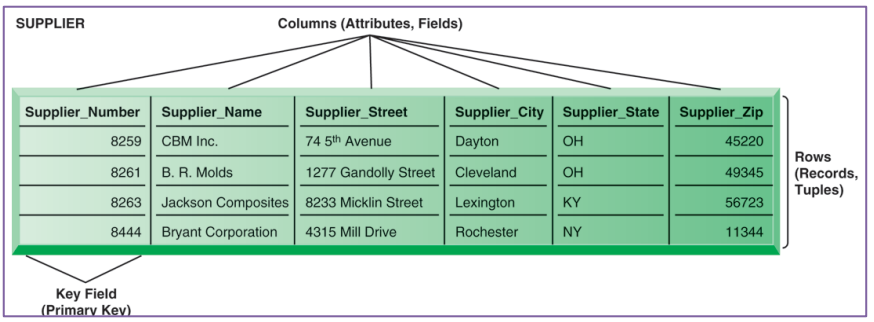


FIGURE 6-4 A relational database organizes data in the form of two-dimensional tables. Illustrated here are tables for the entities SUPPLIER and PART showing how they represent each entity and its attributes. Supplier Number is a primary key for the SUPPLIER table and a foreign key for the PART table.

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

The Database Approach to Data Management

RELATIONAL DATABASE TABLES (cont.)

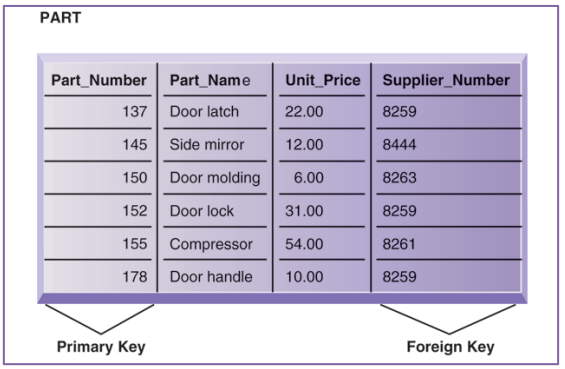


FIGURE 6-4 (cont.)

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.

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

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

The Database Approach to Data Management

THE THREE BASIC OPERATIONS OF A RELATIONAL DBMS

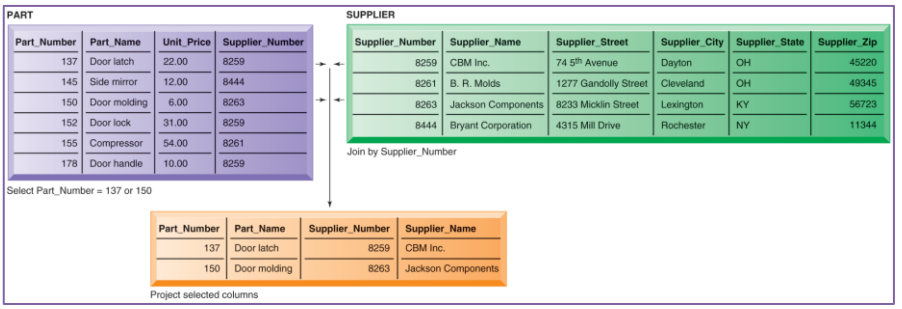


FIGURE 6-5 The select, join, and project operations enable data from two different tables to be combined and only selected attributes to be displayed.

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**
  - **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**

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

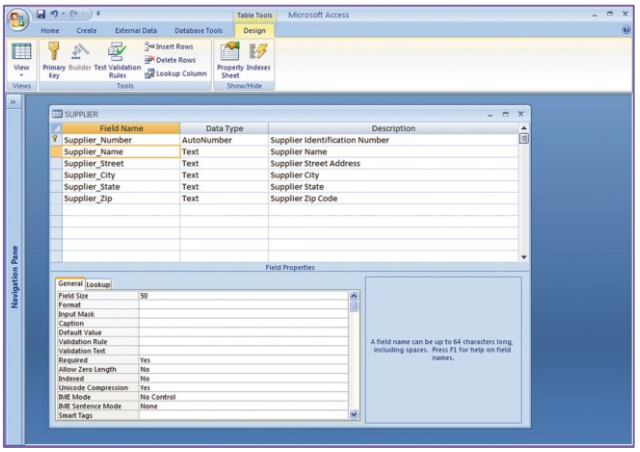
CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

The Database Approach to Data Management

MICROSOFT ACCESS DATA DICTIONARY FEATURES

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



Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE:
DATABASES AND INFORMATION MANAGEMENT

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.

Management Information Systems

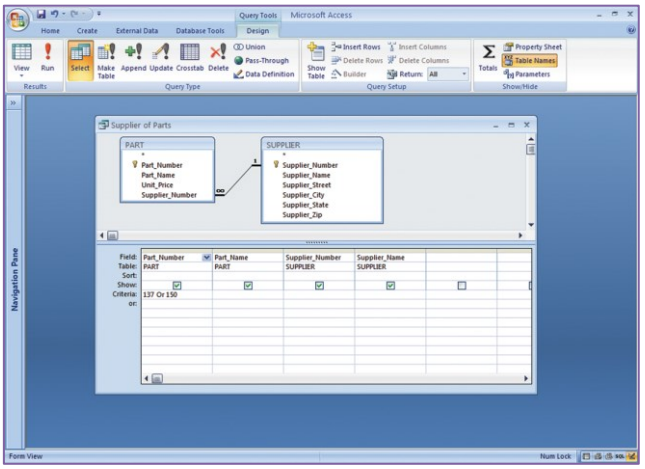
CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE:
DATABASES AND INFORMATION MANAGEMENT

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



Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

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

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

The Database Approach to Data Management

AN UNNORMALIZED RELATION FOR ORDER

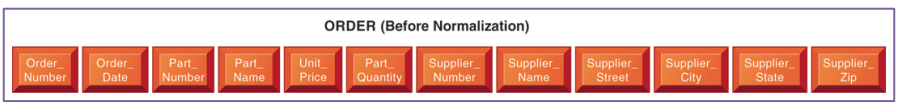


FIGURE 6-9 An unnormalized relation contains repeating groups. For example, there can be many parts and suppliers for each order. There is only a one-to-one correspondence between Order_Number and Order_Date.

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

The Database Approach to Data Management

NORMALIZED TABLES CREATED FROM ORDER

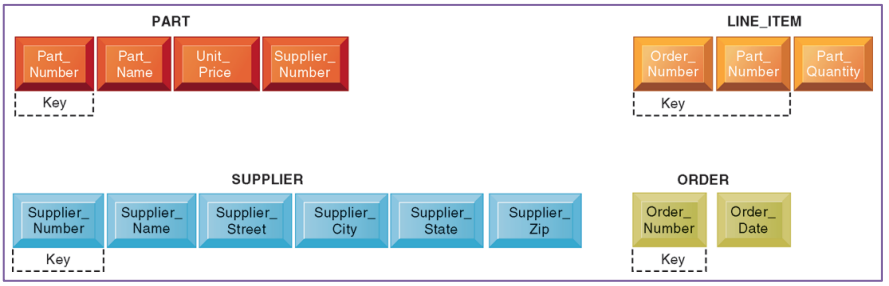


FIGURE 6-10 An unnormalized relation contains repeating groups. For example, there can be many parts and suppliers for each order. There is only a one-to-one correspondence between Order_Number and Order_Date.

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

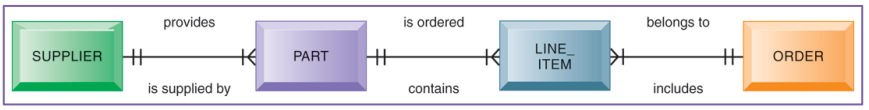
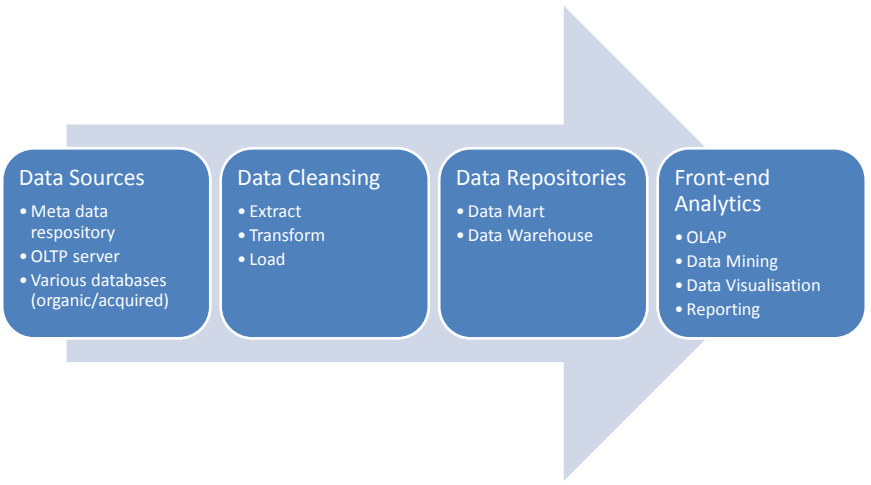


FIGURE 6-11 This diagram shows the relationships between the entities SUPPLIER, PART, LINE_ITEM, and ORDER that might be used to model the database in Figure 6-10.

MB

Data Journey



DATA WAREHOUSES

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
  1. Data warehousing
  2. Data mining
  3. Tools for accessing internal databases through the Web

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

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

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

The Database Approach to Data Management

COMPONENTS OF A DATA WAREHOUSE

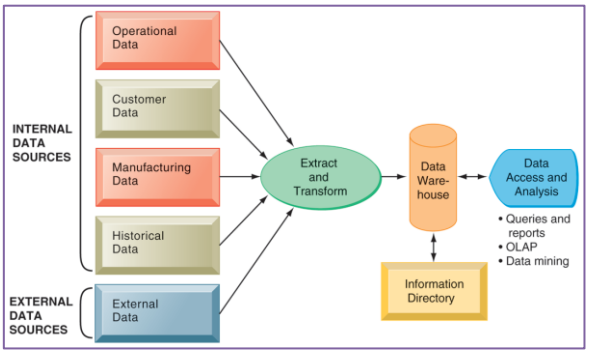


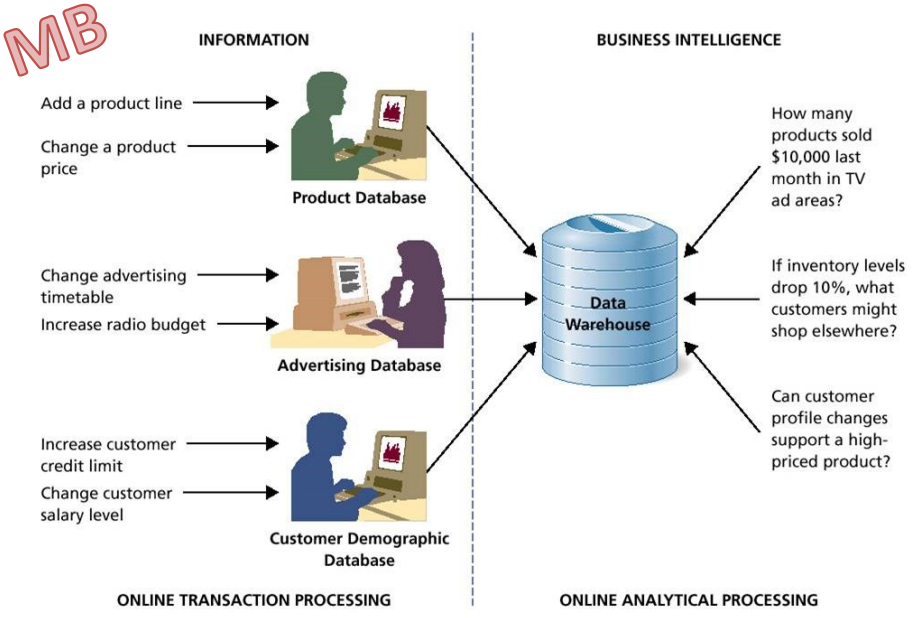
FIGURE 6-12 The data warehouse extracts current and historical data from multiple operational systems inside the organization. These data are combined with data from external sources and reorganized into a central database designed for management reporting and analysis. The information directory provides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

MB

Online Transactions Processing (OLTP)

- OLTP – Online Transactions Processing is the gathering of input data, processing that data and updating existing information to reflect the gathered & processed information.
- Databases which support these functions are often called operational databases (e.g. customer db, product db).

MB



MB

Online Analytical Processing (OLAP)

- This is the manipulation of information to support decision making.

- Used in conjunction with data warehouses.
- Data warehouse is special form of a databases that contains information gathered from operational databases for the purpose of supporting decision making.
- Data warehouses do not support OLTP.

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE: DATABASES AND INFORMATION MANAGEMENT

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**

Management Information Systems

CHAPTER 6: FOUNDATIONS OF BUSINESS INTELLIGENCE:
DATABASES AND INFORMATION MANAGEMENT

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