MIT 3103
ADVANCED MIS

Chapter 4

Data Resource Management

Acknowledgements

- Notes adapted from Management Information Systems by James Obrien and George Marakas

Learning Outcomes

- By the end of this chapter, the learner should be able to:
  - Describe different types of databases.
  - Differentiate data warehouse from data mart.
  - Identify applications of data mining.

Introduction

- Data are a vital organizational resource that need to be managed like other important business assets.
- Today's business enterprises cannot survive or succeed without quality data about their internal operations and external environment.

What is Data Resource Management?

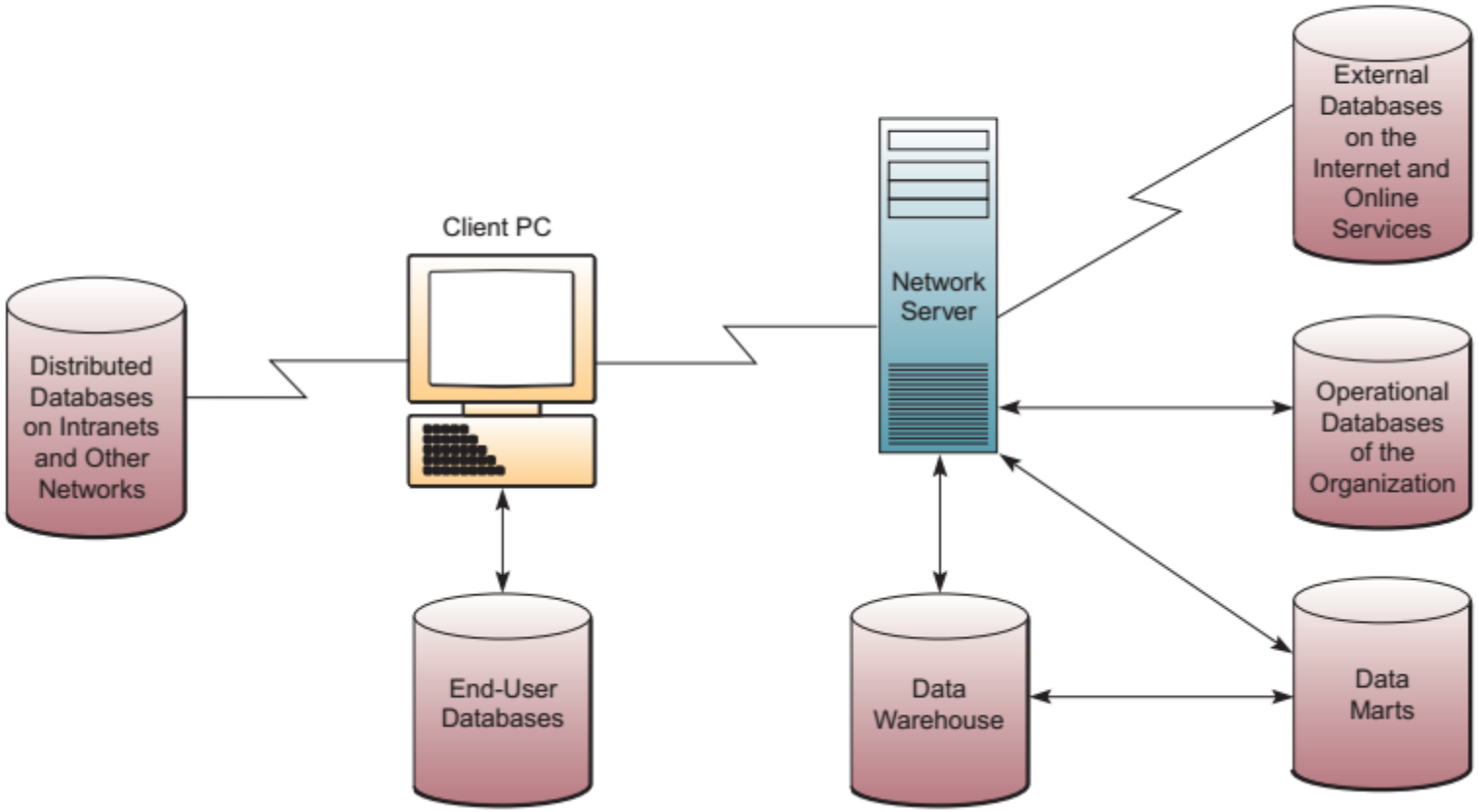
- This is why organizations and their managers need to practice **data resource management**, a managerial activity that applies information systems technologies like *database management, data warehousing,* and other data management tools to the task of managing an organization's data resources to meet the information needs of their business stakeholders.

Types of Databases

- Operational Databases
- Distributed Databases .
- External Databases .
- Hypermedia Database.

Types of Databases

FIGURE 5.14 Examples of some of the major types of databases used by organizations and end users.



Operational Database

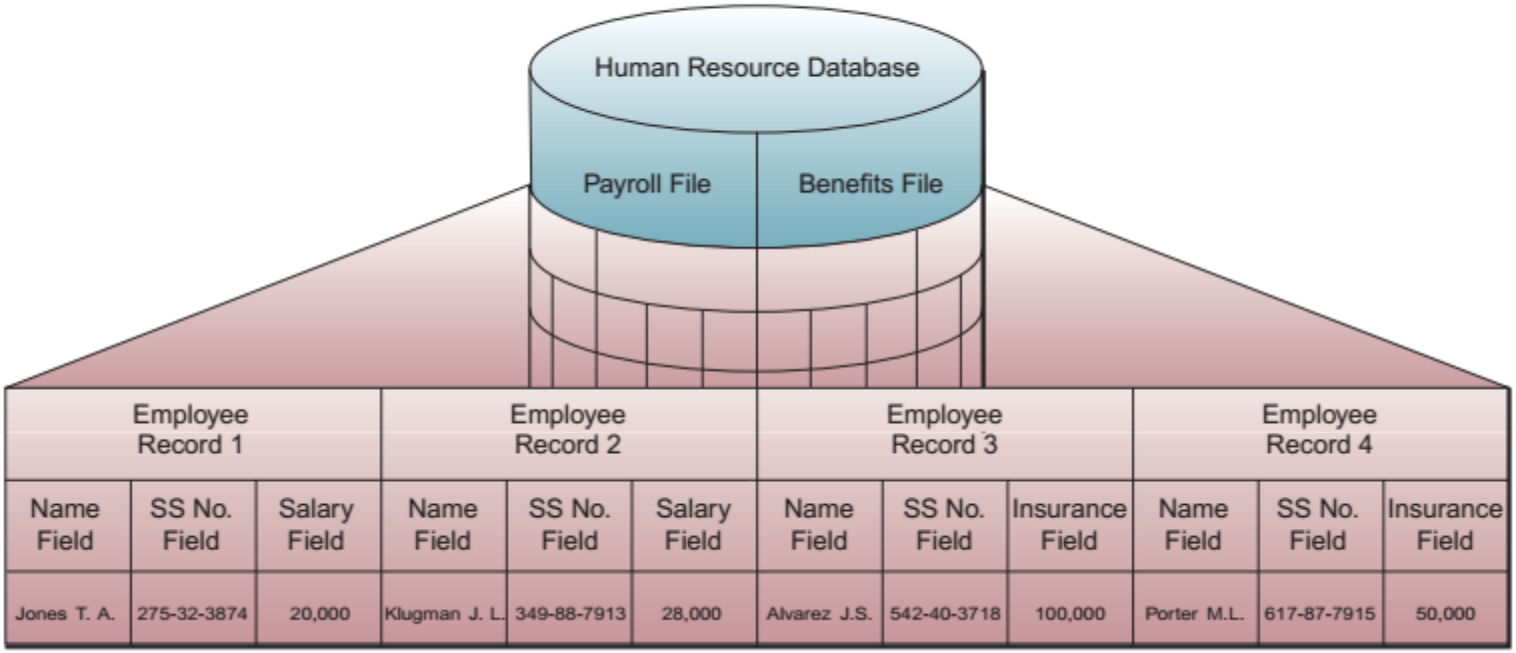
- **Operational databases** store detailed data needed to support the business processes and operations of a company.
- They are also called *subject area databases* (SADB), *transaction databases,* and *production databases.*
  - Examples are a customer database, human resource database, inventory database, and other databases containing data generated by business operations.

Operational Database

- For example, a human resource database like that shown in Figure 5.2 would include data identifying each employee and his or her time worked, compensation, benefits, performance appraisals, training and development status, and other related human resource data.
- Figure 5.15 illustrates some of the common operational databases that can be created and managed for a small business using Microsoft Access database management software.

Operational Database

FIGURE 5.2 Examples of the logical data elements in information systems. Note especially the examples of how data fields, records, files, and databases relate.



Employee Record 1			Employee Record 2			Employee Record 3			Employee Record 4		
Name Field	SS No. Field	Salary Field	Name Field	SS No. Field	Salary Field	Name Field	SS No. Field	Insurance Field	Name Field	SS No. Field	Insurance Field
Jones T. A.	275-32-3874	20,000	Klugman J. L.	349-88-7913	28,000	Alvarez J.S.	542-40-3718	100,000	Porter M.L.	617-87-7915	50,000

Operational Database

FIGURE 5.15
Examples of operational databases that can be created and managed for a small business by microcomputer database management software like Microsoft Access.

Source: Courtesy of Microsoft®.

Distributed Database

- Many organizations replicate and distribute copies or parts of databases to network servers at a variety of sites.
- These **distributed databases** can reside on network servers on the World Wide Web, on corporate intranets or extranets, or on other company networks.

Distributed Database

- Distributed databases may be copies of operational or analytical databases, hypermedia or discussion databases, or any other type of database.
- Replication and distribution of databases improve database performance at end-user worksites.

Distributed Database

- Ensuring that the data in an organization's distributed databases are consistently and concurrently updated is a major challenge of distributed database management.
- Distributed databases have both advantages and disadvantages.

Distributed Database

- One primary advantage of a distributed database lies with the protection of valuable data.
- If all of an organization's data reside in a single physical location, any catastrophic event like a fire or damage to the media holding the data would result in an equally catastrophic loss of use of that data.

Distributed Database

- By having databases distributed in multiple locations, the negative impact of such an event can be minimized.

Distributed Database

- Another advantage of distributed databases is found in their storage requirements.
  - Often, a large database system may be distributed into smaller databases based on some logical relationship between the data and the location.

Distributed Database

- E.g., a company with several branch operations may distribute its data so that each branch operation location is also the location of its branch database.
  - Because multiple databases in a distributed system can be joined together, each location has control of its local data while all other locations can access any database in the company if so desired.

Distributed Database

- The primary challenge is the maintenance of data accuracy.
  - If a company distributes its database to multiple locations, any change to the data in one location must somehow be updated in all other locations.
  - This updating can be accomplished in one of two ways:
    - *Replication* or
    - *Duplication*.

Distributed Database: Replication

- Updating a distributed database using **replication** involves using a specialized software application that looks at each distributed database and then finds the changes made to it.
  - Once these changes have been identified, the replication process makes all of the distributed databases look the same by making the appropriate changes to each one.

Distributed Database: Replication

- The replication process is very complex and, depending on the number and size of the distributed databases, can consume a lot of time and computer resources.

Distributed Database: Duplication

- The **duplication** process, in contrast, is much less complicated.
  - It basically identifies one database as a master and then duplicates that database at a prescribed time after hours so that each distributed location has the same data.

Distributed Database: Duplication

- One drawback to the duplication process is that no changes can ever be made to any database other than the master to avoid having local changes overwritten during the duplication process.
- Nonetheless, properly used, duplication and replication can keep all distributed locations current with the latest data.

Distributed Database

- One additional challenge associated with distributed databases is the extra computing power and bandwidth necessary to access multiple databases in multiple locations.

External Databases

- Access to a wealth of information from **external databases** is available for a fee from commercial online services and with or without charge from many sources on the World Wide Web.
- Web sites provide an endless variety of hyperlinked pages of multimedia documents in *hypermedia databases* for you to access.

External Databases

- Data are available in the form of statistics on economic and demographic activity from *statistical* databanks, or you can view or download abstracts or complete copies of hundreds of newspapers, magazines, newsletters, research papers, and other published material and periodicals from *bibliographic* and *full-text* databases.

Hypermedia Databases

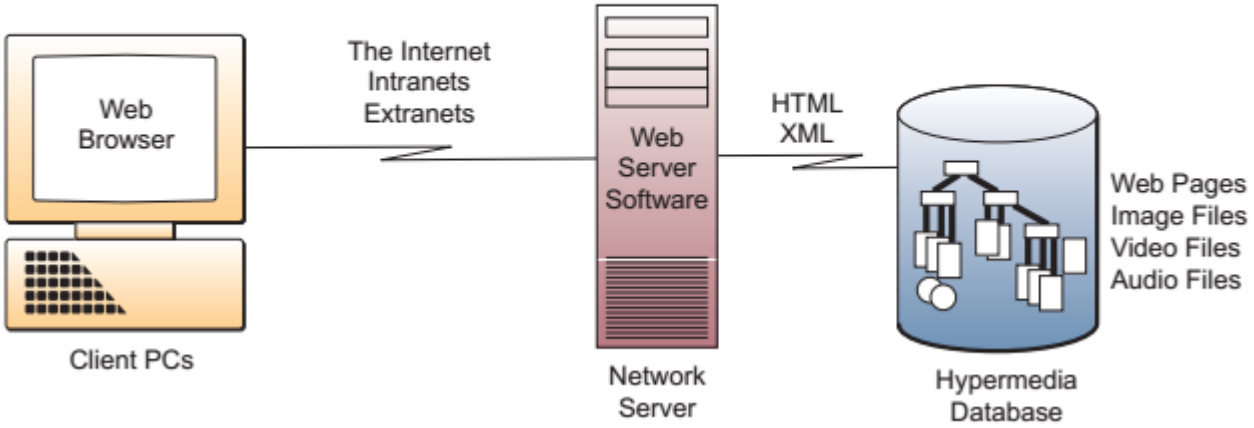
- The rapid growth of Web sites on the Internet and corporate intranets and extranets has dramatically increased the use of databases of hypertext and hypermedia documents.
- A Web site stores such information in a **hypermedia database** consisting of hyperlinked pages of multimedia (text, graphic and photographic images, video clips, audio segments, and so on).

Hypermedia Databases

- That is, from a database management point of view, the set of interconnected multimedia pages on a Web site is a database of interrelated hypermedia page elements, rather than interrelated data records.

Hypermedia Databases

FIGURE 5.16 The components of a Web-based information system include Web browsers, servers, and hypermedia databases.



Data Warehouses

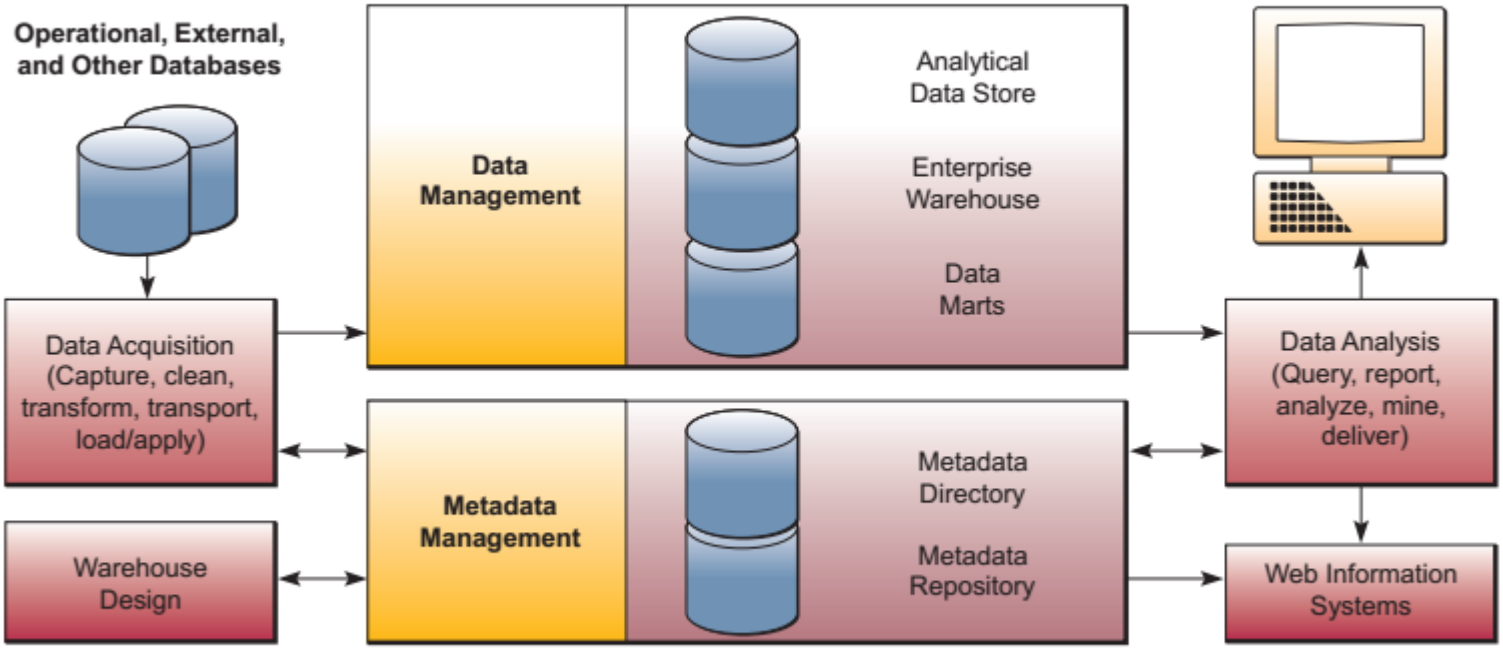
- A **data warehouse** stores data that have been extracted from the various operational, external, and other databases of an organization.
  - It is a central source of the data that have been cleaned, transformed, and cataloged so that they can be used by managers and other business professionals for data mining, online analytical processing, and other forms of business analysis, market research, and decision support.

Data Warehouses

- Data warehouses may be subdivided into **data marts,** which hold subsets of data from the warehouse that focus on specific aspects of a company, such as a department or a business process.

Data Warehouses

FIGURE 5.17 The components of a complete data warehouse system.



Source: Courtesy of Hewlett-Packard.

Data Warehouses and Data Mining

- Data acquisition process might include activities like:
  - Consolidating data from several sources
  - Filtering out unwanted data
  - Correcting incorrect data
  - Converting data to new data elements, or
  - Aggregating data into new data subsets.

Data Warehouses

- These data are then stored in the enterprise data warehouse, from which they can be moved into data marts or to an *analytical data store* that holds data in a more useful form for certain types of analysis.
- *Metadata* (data that define the data in the data warehouse) are stored in a metadata repository and cataloged by a metadata directory.

Data Warehouses

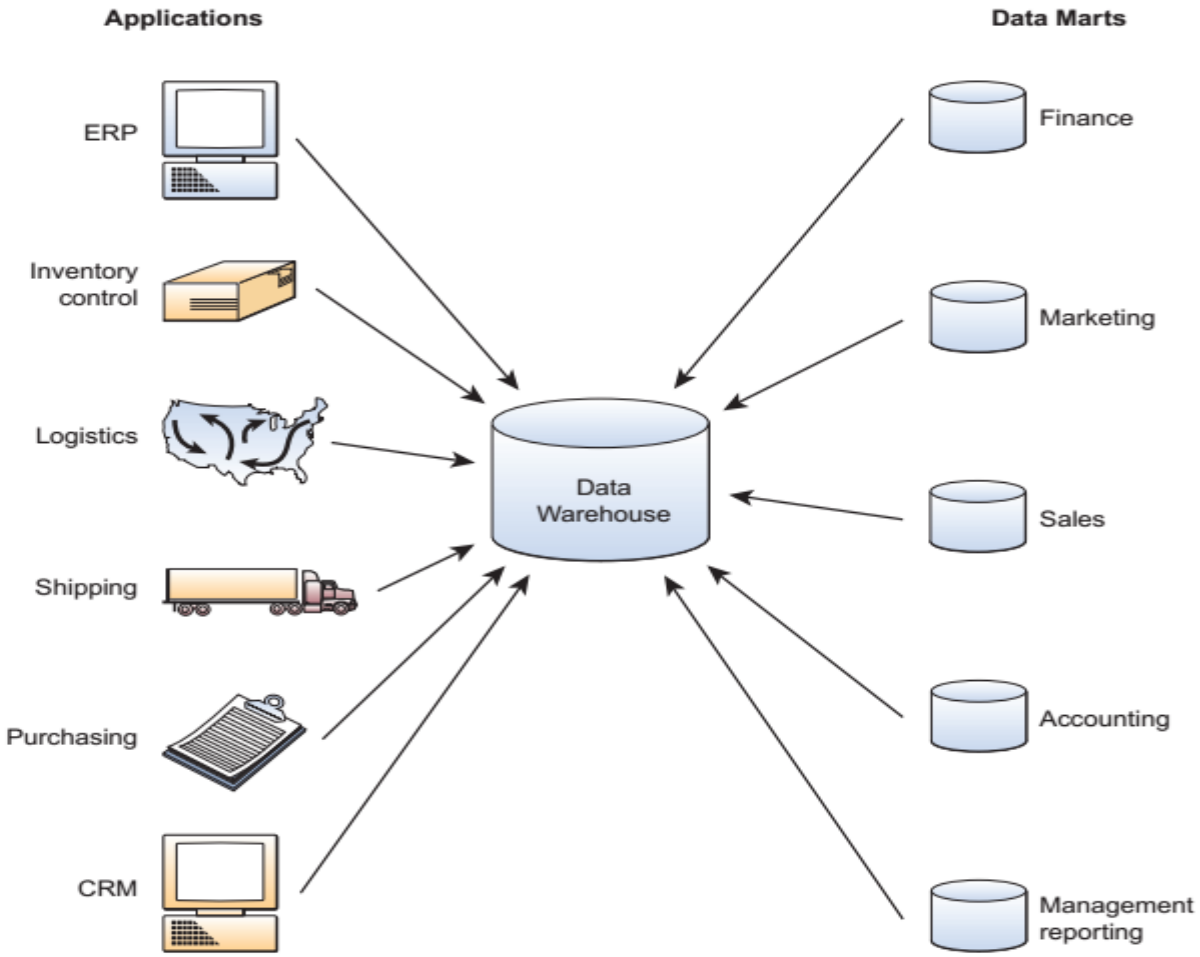
- Finally, a variety of analytical software tools can be provided to query, report, mine, and analyze the data for delivery via Internet and intranet Web systems to business end users.

Data Warehouses

- One important characteristic about the data in a data warehouse is that, unlike a typical database in which changes can occur constantly, data in a data warehouse are *static*, which means that once the data are gathered up, formatted for storage, and stored in the data warehouse, they will never change.

Data Warehouses

FIGURE 5.18
A data warehouse and its data mart subsets hold data that have been extracted from various operational databases for business analysis, market research, decision support, and data mining applications.



Data Warehouses

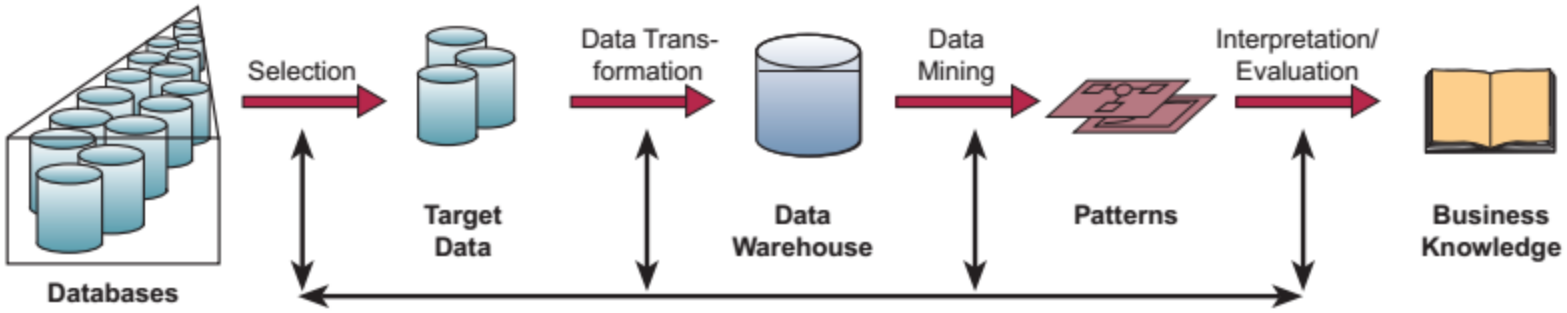
- This restriction is so that queries can be made on the data to look for complex patterns or historical trends that might otherwise go unnoticed with dynamic data that change constantly as a result of new transactions and updates.

Data Mining

- **Data mining** is a major use of data warehouse databases and the static data they contain.
  - In data mining, the data in a data warehouse are analyzed to reveal hidden patterns and trends in historical business activity.

Data Mining

FIGURE 5.19 How data mining extracts business knowledge from a data warehouse.



Data Mining

- This analysis can be used to help managers make decisions about strategic changes in business operations to gain competitive advantages in the marketplace.
- Data mining can discover new correlations, patterns, and trends in vast amounts of business data (frequently several terabytes of data) stored in data warehouses.

Data Mining

- Data mining software uses advanced pattern recognition algorithms, as well as a variety of mathematical and statistical techniques, to sift through mountains of data to extract previously unknown strategic business information.

Data Mining

- For example, many companies use data mining to:
  - Perform market-basket analysis to identify new product bundles.
  - Find root causes of quality or manufacturing problems.
  - Prevent customer attrition and acquire new customers.
  - Cross-sell to existing customers.
  - Profile customers with more accuracy.

The Database Management Approach

- The database management approach consolidates data records, formerly held in separate files, into databases that can be accessed by many different application programs.
- In addition, a *database management system* (DBMS) serves as a software interface between users and databases, which helps users easily access the data in a database.

The Database Management Approach

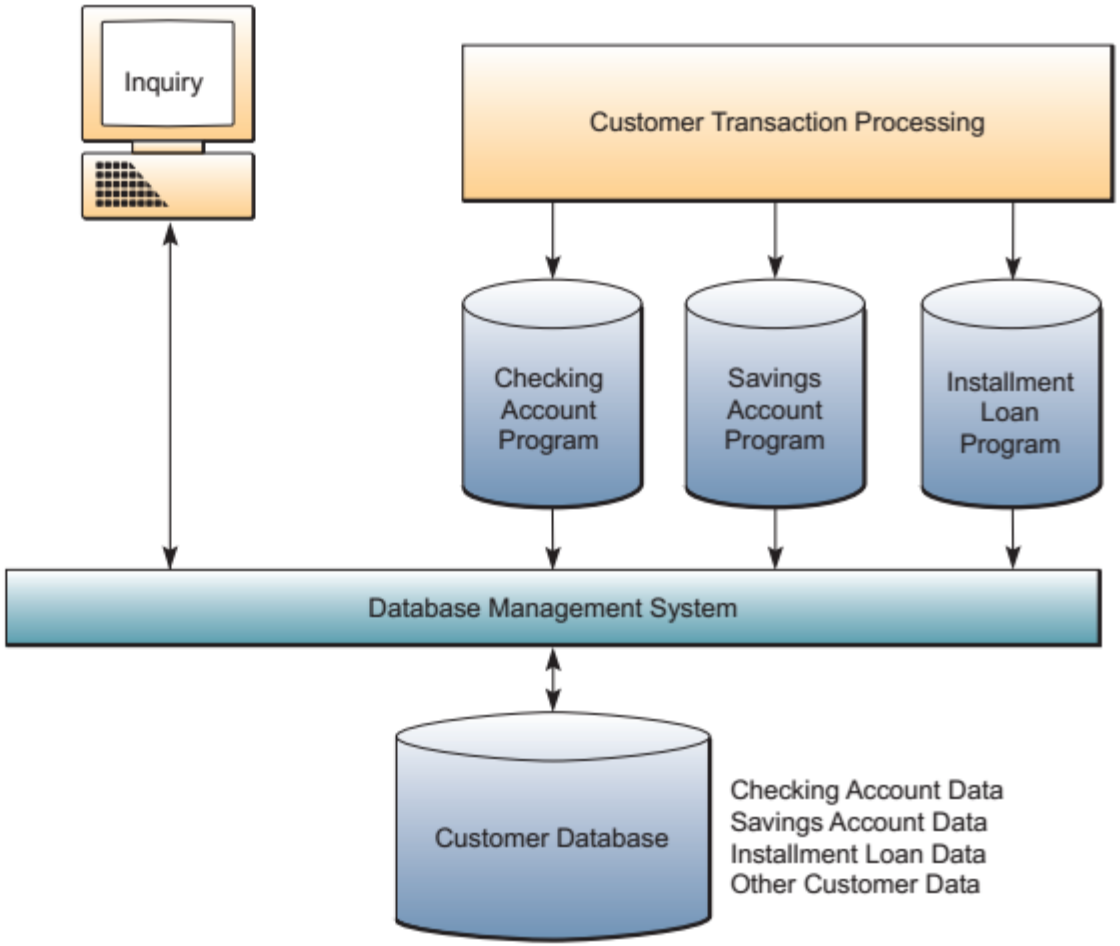
- Thus, database management involves the use of database management software to control how databases are created, interrogated, and maintained to provide information that end users need.

The Database Management Approach

- For example, customer records and other common types of data are needed for several different applications in banking, such as check processing, automated teller systems, bank credit cards, savings accounts, and installment loan accounting.
  - These data can be consolidated into a common *customer database,* rather than being kept in separate files for each of those applications.

The Database Management Approach

FIGURE 5.21
An example of a database management approach in a banking information system. Note how the savings, checking, and installment loan programs use a database management system to share a customer database. Note also that the DBMS allows a user to make direct, ad hoc interrogations of the database without using application programs.



The Database Management Approach

- A **database management system (DBMS)** is the main software tool of the database management approach because it controls the creation, maintenance, and use of the databases of an organization and its end users.

The Database Management Approach

- The three major functions of a database management system are to:
  - C*reate* new databases and database applications
  - *Maintain* the quality of the data in an organization's databases, and
  - *Use* the databases of an organization to provide the information that its end users need

Database Maintenance

- The **database maintenance** process is accomplished by *transaction processing systems* and other end-user applications, with the support of the DBMS.
- End users and information specialists can also employ various utilities provided by a DBMS for database maintenance.

Database Maintenance

- The databases of an organization need to be updated continually to reflect new business transactions (e.g., sales made, products produced, inventory shipped) and other events.
- Other miscellaneous changes also must be made to update and correct data (e.g., customer or employee name and address changes) to ensure the accuracy of the data in the databases.

End of Chapter 4