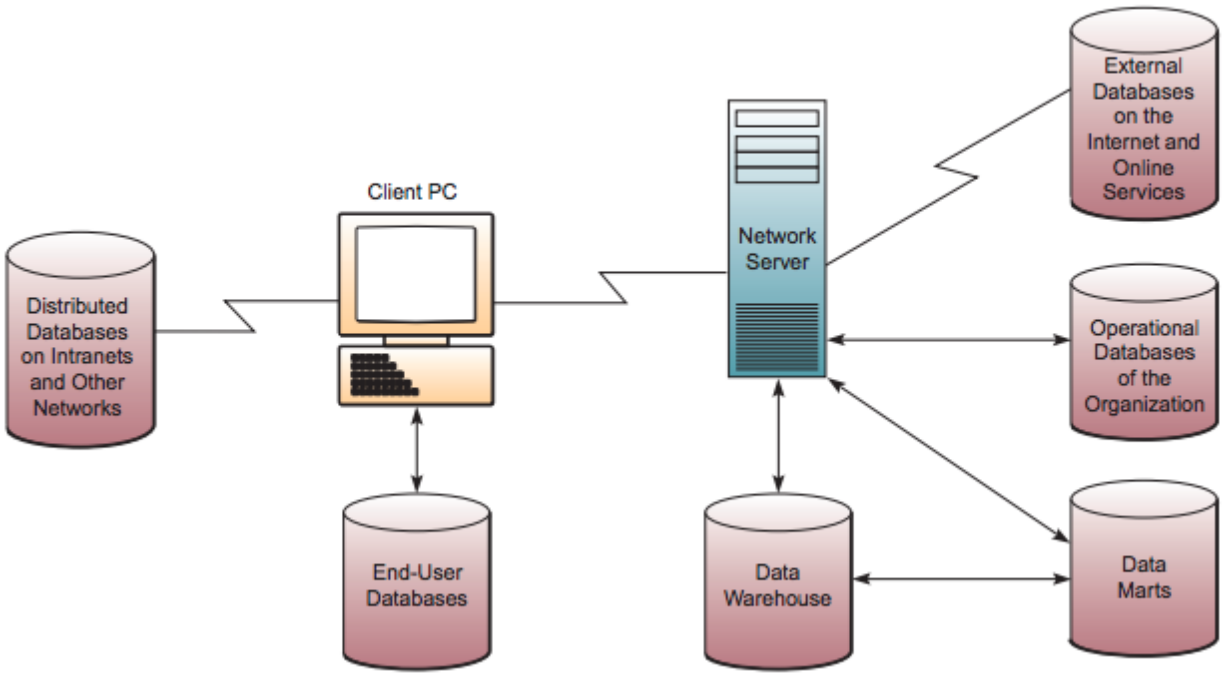
Unit 5: Managing Data Resources

Data Resource Management:

- 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.
- **Data Resource Management** is 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

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



1. **Operational Databases**
   - store detailed data needed to support the business processes and operations of a company.
   - 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.
   - For example, a human resource database 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.

2. **Distributed Databases**
   - 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 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.
   - Ensuring that the data in an organization's distributed databases are consistently and concurrently updated is a major challenge of distributed database management.

Advantages

- Protection of valuable data.
- **Storage requirements:** Often, a large database system may be distributed into smaller databases based on some logical relationship between the data and the location. For example, 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.

Challenges

- **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**.

  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.
  - 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.

  Duplication:
  - is much less complicated.

- 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.
- 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.

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

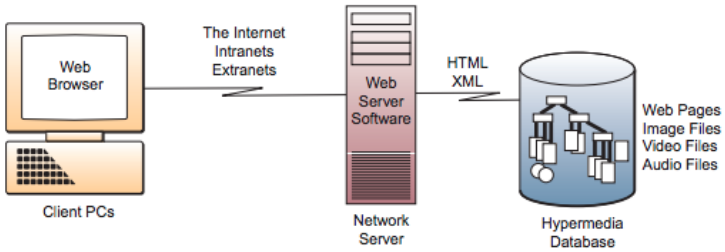
3. External Databases

- Websites provide an endless variety of hyperlinked pages of multimedia documents in hypermedia databases for us to access.
- Data are available in the form of statistics on economic and demographic activity from statistical databanks, or we 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.
- Whenever we use a search engine like Google or Yahoo to look up something on the Internet, we are using an external database - a very, very large one!

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

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

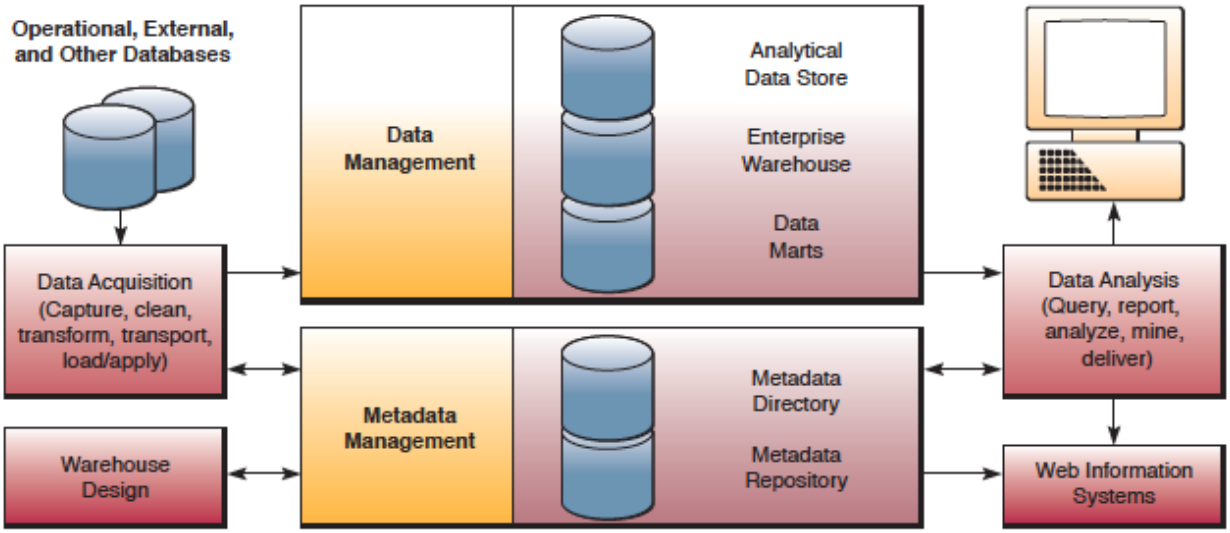


Data warehousing

- 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 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.

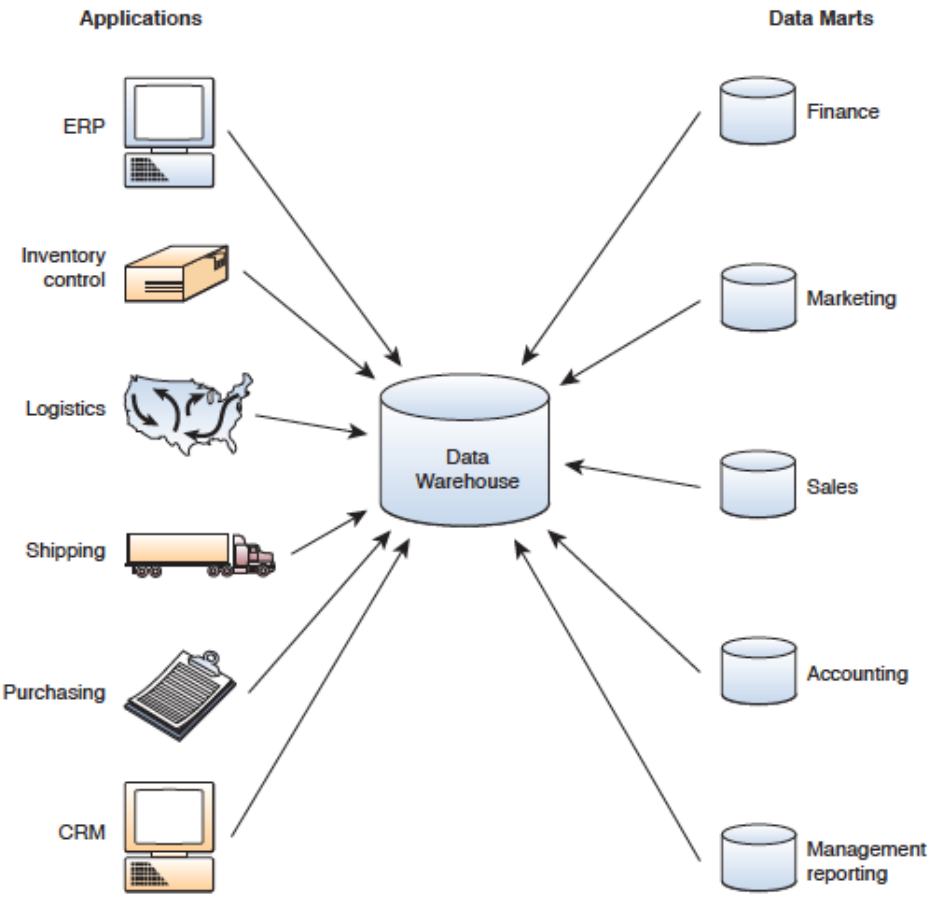
FIGURE 5.17 The components of a complete data warehouse system.



Source: Courtesy of Hewlett-Packard.

- Data from various operational and external databases are captured, cleaned, and transformed into data that can be better used for analysis.
- This 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.
- 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.
- 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.

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.

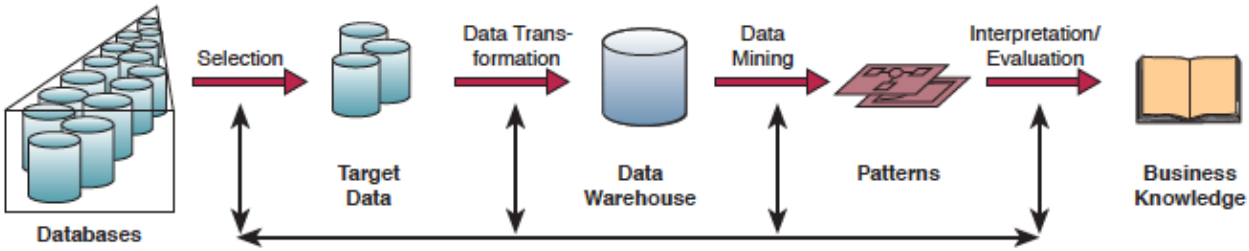


- One important characteristic about the data in a data warehouse is that, unlike a typical database in which changes can occur constantly, data in a data warehouse are static, which means that once the data are gathered up, formatted for storage, and stored in the data warehouse, they will never change.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

- In data mining, the data in a data warehouse are analyzed to reveal hidden patterns and trends in historical business activity.
- This analysis can be used to help managers make decisions about strategic changes in business operations to gain competitive advantages in the marketplace.

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



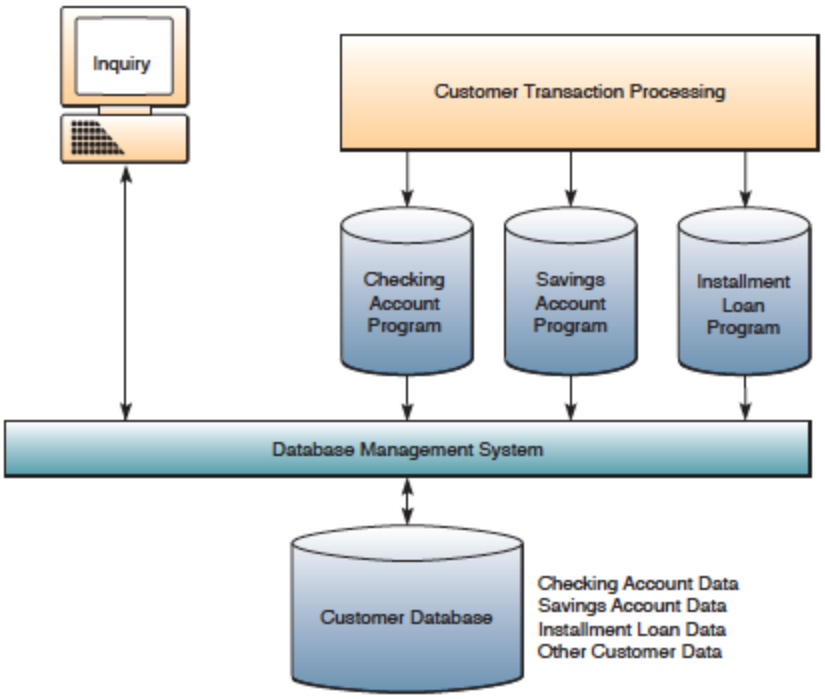
- Data mining can discover new correlations, patterns, and trends in vast amounts of business data (frequently several terabytes of data) stored in data warehouses.
- Data mining software uses advanced pattern recognition algorithms, as well as a variety of mathematical and statistical techniques, to sift through mountains of data to extract previously unknown strategic business information.
- 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.

*[**attrition**: the process of reducing something's strength or effectiveness through sustained attack or pressure.]*

The Database Management Approach

- To solve the problems encountered with the file processing approach, the database management approach was conceived as the foundation of modern methods for managing organizational data.
- 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.
- 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.
- 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.

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.



Database Management System

- 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.
- Examples of popular DBMS software are IBM's DB2 Universal Database, Oracle 10g by Oracle Corp., and MySQL, a popular open-source DBMS.

FIGURE 5.23 Common software components and functions of a database management system.

Common DBMS Software Components	
• **Database Definition**	Language and graphical tools to define entities, relationships, integrity constraints, and authorization rights.
• **Nonprocedural Access**	Language and graphical tools to access data without complicated coding.
• **Application Development**	Graphical tools to develop menus, data entry forms, and reports.
• **Procedural Language Interface**	Language that combines nonprocedural access with full capabilities of a programming language.
• **Transaction Processing**	Control mechanisms to prevent interference from simultaneous users and recover lost data after a failure.
• **Database Tuning**	Tools to monitor and improve database performance.

Source: Michael V. Mannino, *Database Application Development and Design* (Burr Ridge, IL: McGraw-Hill/Irwin, 2001), p. 7.

- The three major functions of a database management system are:
    1. to create new databases and database applications,
    2. to maintain the quality of the data in an organization's databases, and
    3. to use the databases of an organization to provide the information that its end users need.

- **Database development** involves defining and organizing the content, relationships, and structure of the data needed to build a database.
- **Database application development** involves using a DBMS to develop prototypes of queries, forms, reports, and Web pages for a proposed business application.
- **Database maintenance** involves using transaction processing systems and other tools to add, delete, update, and correct the data in a database.
- The primary use of a database by end users involves employing the database interrogation capabilities of a DBMS to access the data in a database to selectively retrieve and display information and produce reports, forms, and other documents.

Database Interrogation

- A database interrogation capability is a major benefit of the database management approach.
- End users can use a DBMS by asking for information from a database using a *query feature* or a *report generator*.
- They can receive an immediate response in the form of video displays or printed reports. No difficult programming is required.
- The **query language** feature lets us easily obtain immediate responses to ad hoc data requests: in a few short inquiries - in some cases, using common sentence structures just like we would use to ask a question.
- The **report generator** feature allows us to specify a report format for information we want presented as a report.

SQL Queries

- Structured Query Language, is an international standard query language found in many DBMS packages. In most cases, SQL is the language structure used to “ask a question” that the DBMS will retrieve the data to answer.
- The basic form of a SQL query is:
  SELECT . . . FROM . . . WHERE . . .
  - After SELECT, we list the data fields we want retrieved.
  - After FROM, we list the files or tables from which the data must be retrieved.
  - After WHERE, we specify conditions that limit the search to only those data records in which we are interested.

Boolean Logic

- Boolean logic allows us to refine our searches for specific information such that only the desired information is obtained.
- Boolean logic consists of three logical operators: (1) AND, (2) OR, and (3) NOT.
- Using these operators in conjunction with the syntax of a SQL query, a database user can refine a search to ensure that only the desired data are retrieved.
- For example, to retrieve the students of BIM who have no BackPapers can be:
  Students WHERE program is BIM AND NOT BackPapers

Graphical and Natural Queries

- Many end users (and IS professionals) have difficulty correctly phrasing SQL and other database language search queries.
- So most end-user database management packages offer GUI (graphical user interface) point-and-click methods, which are easier to use and are translated by the software into SQL commands.

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.
- 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.

Application Development

- DBMS packages play a major role in application development.
- End users, systems analysts, and other application developers can use the internal 4GL programming language and built-in software development tools provided by many DBMS packages to develop custom application programs.
- For example, we can use a DBMS to develop the data entry screens, forms, reports, or Web pages of a business application that accesses a company database to find and update the data it needs.
- A DBMS also makes the job of application software developers easier, because they do not have to develop detailed data-handling procedures using conventional programming languages every time they write a program.
- Instead, they can include features such as data manipulation language (DML) statements in their software that call on the DBMS to perform necessary data-handling activities.